Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 14, No. 1, pp. 77–99, 2018

AN ITERATIVE METHOD USING BOUNDARY DISTANCE FOR BOX-CONSTRAINED NONLINEAR SEMIDEFINITE PROGRAMS

Akihiko Komatsu and Makoto Yamashita

Abstract: We propose an iterative method for nonlinear semidefinite programs with box constraints. The search direction in the proposed method utilizes the distance from the current point to the boundary of a feasible set. The computation of the search direction exploits the second derivative of the objective function only in a quadratic form, and this property saves the computation cost compared to an evaluation of the whole entries of the second derivative. We compute a step length in an interval determined by a radius and we update the radius using a quadratic approximation function. In this paper, we also discuss convergence properties of the proposed method based on structures of the search direction. Numerical tests show that the proposed method solves problems in which the size of a variable matrix is larger than 5,000 and that it is faster than a feasible direction method for objective functions with strong nonlinearity.

Key words: *cnonlinear programming, semidefinite programming, box constraint*

Mathematics Subject Classification: *90C22, 90C30*

1 Introduction

This paper is concerned with a box-constrained nonlinear semidefinite problem (shortly, box-constrained SDP)

$$\min \quad f(\boldsymbol{X}) \quad \text{subject to} \quad \boldsymbol{O} \preceq \boldsymbol{X} \preceq \boldsymbol{I}. \tag{1.1}$$

The variable in this problem is $\boldsymbol{X} \in \mathbb{S}^n$, and we use $\mathbb{S}^n$ to denote the space of $n \times n$ symmetric matrices. The notation $\boldsymbol{A} \succeq \boldsymbol{B}$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n$ means that the matrix $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite. The matrix $\boldsymbol{I}$ is the identity matrix of the appropriate dimension. We assume that the objective function $f : \mathbb{S}^n \to \mathbb{R}$ is a twice continuously differentiable function on an open set containing the feasible set $\mathcal{F} := \{\boldsymbol{X} \in \mathbb{S}^n : \boldsymbol{O} \preceq \boldsymbol{X} \preceq \boldsymbol{I}\}$.

The feasible set of (1.1) can express a more general feasible set $\{\boldsymbol{X} \in \mathbb{S}^n : \boldsymbol{L} \preceq \boldsymbol{X} \preceq \boldsymbol{U}\}$ with $\boldsymbol{L}, \boldsymbol{U} \in \mathbb{S}^n$ such that $\boldsymbol{L} \preceq \boldsymbol{U}$. This type of problems appears as a sub problem in other methods [12]. We can assume that $\boldsymbol{U} - \boldsymbol{L}$ is positive definite without loss of generality [25], therefore, we use a Cholesky factorization matrix $\boldsymbol{C}$ of $\boldsymbol{U} - \boldsymbol{L}$ that satisfies $\boldsymbol{U} - \boldsymbol{L} = \boldsymbol{C}\boldsymbol{C}^T$ to convert a problem

$$\min \quad f(\boldsymbol{X}) \quad \text{subject to} \quad \boldsymbol{L} \preceq \boldsymbol{X} \preceq \boldsymbol{U}$$

into an equivalent problem

$$\min \quad f(\boldsymbol{C}\overline{\boldsymbol{X}}\boldsymbol{C}^T + \boldsymbol{L}) \quad \text{subject to} \quad \boldsymbol{O} \preceq \overline{\boldsymbol{X}} \preceq \boldsymbol{I}$$

by the relation $\overline{\boldsymbol{X}} = \boldsymbol{C}^{-1}(\boldsymbol{X} - \boldsymbol{L})(\boldsymbol{C}^{-1})^T$. In this paper, we use the superscript T to denote the transpose of a matrix.

A box-constrained nonlinear optimization problem

$$\min \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}, \boldsymbol{x} \in \mathbb{R}^n. \tag{1.2}$$

is an important case of (1.1), since if the variable matrix $\boldsymbol{X}$ in (1.1) is a diagonal matrix, (1.1) is reduced to (1.2). The problem (1.2) is a basic problem in constrained optimization and many methods are proposed. Hei *et. al.* [11] compared the performance of four active-set methods and two interior-point methods. Trust-region methods for (1.2) are also discussed in [4, 5, 24],

On the other hand, the positive semidefinite condition on a matrix ($\boldsymbol{X} \succeq \boldsymbol{O}$) is extensively studied in the context of SDP (semidefinite programs). The range of SDP applications is very wide and includes control theory [3], combinatorial optimization [9], polynomial optimization [14] and quantum chemistry [8]. Many software packages, for example [23, 26], have been developed for SDP. A number of studies on SDP can be found in the survey of Todd [22], the handbook edited by Anjos and Lassere [1] and the references therein.

For solving the box-constrained SDP (1.1), we may apply the penalty barrier method proposed in [2, 13]. Though it can handle the problem (1.1) with additional constraints, it requires the full information of the second derivative of the objective function, and it can solve the problems in practical time only when the size of variable matrix is small; $n \leq 500$.

To solve large problems with $n \geq 500$, we should discuss methods specialized for solving (1.1). Xu *et al* [25] proposed a feasible direction method for (1.1). This method is an iterative method and it searches a point which satisfies a first-order optimality condition.

We say that $\boldsymbol{X}^* \in \mathcal{F}$ satisfies a first-order optimality condition of (1.1) if

$$\langle \nabla f(\boldsymbol{X}^*) \mid \boldsymbol{X} - \boldsymbol{X}^* \rangle \geq 0 \qquad \text{for} \quad \forall \boldsymbol{X} \in \mathcal{F}. \tag{1.3}$$

Here, we use $\langle \boldsymbol{A} \mid \boldsymbol{B} \rangle$ to denote the inner-product between $\boldsymbol{A} \in \mathbb{S}^n$ and $\boldsymbol{B} \in \mathbb{S}^n$, and $\nabla f(\boldsymbol{X}^*) \in \mathbb{S}^n$ is the gradient matrix of f at $\boldsymbol{X}^*$. In particular, when $f(\boldsymbol{X})$ is a convex function, a point $\boldsymbol{X}^* \in \mathcal{F}$ that satisfies (1.3) is an optimal solution. We can derive an equivalent but more convenient condition for $\boldsymbol{X}^* \in \mathcal{F}$,

$$\underline{f}(\boldsymbol{X}^*) = 0$$

where

$$\underline{f}(\widehat{\boldsymbol{X}}) := \min \{ \langle \nabla f(\widehat{\boldsymbol{X}}) \mid \boldsymbol{X} - \widehat{\boldsymbol{X}} \rangle \ : \ \boldsymbol{X} \in \mathcal{F} \}. \tag{1.4}$$

Xu *et al* [25] proved that the feasible direction method generates an sequence $\{\boldsymbol{X}^k\} \subset \mathcal{F}$ that attains $\lim_{k \to \infty} \underline{f}(\boldsymbol{X}^k) = 0$. They conducted numerical tests on simple objective functions that involved the variable matrix $\boldsymbol{X}$ in linear or quadratic terms.

In this paper, we propose an iterative method for the box-constrained SDP (1.1) using the distance from the current point to the boundary of $\mathcal{F}$. We introduce a concept of the distance to the boundary of the feasible set from a trust-region method of Coleman and Li [4] proposed for the simple-bound problem (1.2). However, we can not directly apply the search direction of [4] to the box-constrained SDP (1.1) by copying the interval condition $\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}$ to the eigenvalue conditions $\boldsymbol{O} \preceq \boldsymbol{X} \preceq \boldsymbol{I}$, since the matrix $\boldsymbol{X}$ involves not only the eigenvalues but also the eigenvectors. In particular, it is not straightforward to guarantee a non-zero step length if we define a search direction ignoring the property that the eigenvectors are not always continuous functions on $\boldsymbol{X}$. We devise a new search direction

by taking both the eigenvalues and the eigenvectors into consideration. We give a non-zero range of the step length, and we ensure that a movement along the search direction in this range remains in $\mathcal{F}$.

We also introduce a quadratic approximation function and a radius adjustment from the trust-region methods [6, 10, 18, 21, 27]. In ordinary trust-region methods, the search direction is obtained by solving a trust-region sub-problem, and the sub-problem is usually an optimization problem that minimizes a quadratic function with a constraint where the search direction is bounded by a trust-region radius. The search direction by such a trust-region sub-problem was examined for nonlinear semidefinite complementarity programs in [15], but an evaluation of the second derivative functions required a huge computation cost and the problem size there was at most $n = 100$. In our approach, we first obtain the search direction based on the distance to the boundary, then we obtain the step length along this search direction so that the next point will stay in the region determined by a radius. In the computation of the step length, we utilize the second derivative in its quadratic form, hence the computation cost in each iteration is lower than the evaluation of whole entries of the second derivative. We update the radius for the next iteration using an deviation of the quadratic approximation function from the objective function.

In this paper, we discuss convergence properties of the generated sequence for the first-order optimality condition. Numerical tests in this paper show that the proposed method solves strongly-nonlinear functions faster than the feasible direction method. The computation cost of the proposed method in each iteration is low compared to the penalty barrier method implemented in PENLAB [7], and the proposed method can handle larger problems than the penalty barrier method. This paper is organized as follows. Section 2 discusses equivalent conditions of the first-optimality conditions. We introduce the new search direction $\boldsymbol{D}(\boldsymbol{X})$, and propose the iterative method with adaptive radius adjustment in Algorithm 2.3. Section 3 establishes the convergence properties of the proposed method. Section 4 reports numerical results on the performance comparison of the proposed method, the feasible direction, and the penalty barrier method. Finally, Section 5 gives a conclusion of this paper and discusses future directions.

1.1 Notation and preliminaries

The inner-product between $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ is defined by $\langle \boldsymbol{A}|\boldsymbol{B}\rangle := Trace(\boldsymbol{A}^T\boldsymbol{B})$. Here, $Trace(\boldsymbol{X})$ for a matrix $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ is the summation of its diagonal elements, that is, $Trace(\boldsymbol{X}) := \sum_{i=1}^{n} X_{ii}$.

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we define the Frobenius norm by $||\boldsymbol{A}||_F := \sqrt{\langle \boldsymbol{A} \mid \boldsymbol{A}\rangle}$. From the Cauchy-Schwartz inequality, it holds $|\langle \boldsymbol{A} \mid \boldsymbol{B}\rangle| \le ||\boldsymbol{A}||_F||\boldsymbol{B}||_F$ for $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$. Throughout the paper, we often use the relation $\langle \boldsymbol{A} \mid \boldsymbol{B}\rangle = \langle \boldsymbol{B} \mid \boldsymbol{A}\rangle$. In addition, we use the inequality $\langle \boldsymbol{A} \mid \boldsymbol{B}\rangle \ge 0$ for two positive semidefinite matrices $\boldsymbol{A} \succeq \boldsymbol{O}$ and $\boldsymbol{B} \succeq \boldsymbol{O}$.

The notation $diag(\kappa_1, \kappa_2, \ldots, \kappa_n)$ stands for the diagonal matrix whose diagonal elements are $\kappa_1, \kappa_2, \ldots, \kappa_n$. When $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^T$ is the eigenvalue decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ with the diagonal matrix $\boldsymbol{K} = diag(\kappa_1, \kappa_2, \ldots, \kappa_n)$, the rth power of $\boldsymbol{A}$ for $r \in \mathbb{R}$ is given by $\boldsymbol{A}^r := \boldsymbol{Q}diag(\kappa_1^r, \kappa_2^r, \ldots, \kappa_n^r)\boldsymbol{Q}^T$ and the two-norm $||\boldsymbol{A}||_2$ is given by $||\boldsymbol{A}||_2 := (\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A}))^{1/2} = \max_{i=1,\ldots,n} |\kappa_i|$. Here, $\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})$ is the largest eigenvalue of the matrix $\boldsymbol{A}^T\boldsymbol{A}$

The gradient matrix $\nabla f(\boldsymbol{X}) \in \mathbb{S}^n$ and the Hessian mapping $\nabla^2 f(\boldsymbol{X})$ at $\boldsymbol{X} \in \mathbb{S}^n$ are defined so that a Taylor expansion for $\boldsymbol{D} \in \mathbb{S}^n$ holds with

$$f(\boldsymbol{X} + \boldsymbol{D}) = f(\boldsymbol{X}) + \langle \nabla f(\boldsymbol{X}) \mid \boldsymbol{D}\rangle + \frac{1}{2}\langle \boldsymbol{D} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{D}\rangle + o(||\boldsymbol{D}||_F^2),$$

where $o(d)$ is of the higher order of d. For example, for a function $\hat{f}(\boldsymbol{X}) = \langle \boldsymbol{X} \mid \boldsymbol{X} \rangle$, we have $\nabla \hat{f}(\boldsymbol{X}) = 2\boldsymbol{X}$ and $\langle \boldsymbol{D} \mid \nabla^2 \hat{f}(\boldsymbol{X}) \mid \boldsymbol{D} \rangle = 2\langle \boldsymbol{D} \mid \boldsymbol{D} \rangle$ from the relation $\langle \boldsymbol{X} + \boldsymbol{D} \mid \boldsymbol{X} + \boldsymbol{D} \rangle = \langle \boldsymbol{X} \mid \boldsymbol{X} \rangle + 2 \langle \boldsymbol{X} \mid \boldsymbol{D} \rangle + \langle \boldsymbol{D} \mid \boldsymbol{D} \rangle$. The gradient matrix $\nabla f(\boldsymbol{X})$ corresponds to the Fréchet derivative, and we have $\langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle = \sum_{i,j,k,l=1}^{n} \frac{\partial^2 f(\boldsymbol{X})}{\partial X_{kl} \partial X_{ij}} A_{ij} B_{kl}$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n$.

We use the matrices $\boldsymbol{P}(\boldsymbol{X})$ and $\boldsymbol{\Gamma}(\boldsymbol{X})$ to denote the eigenvalue decomposition of $\nabla f(\boldsymbol{X})$ as $\nabla f(\boldsymbol{X}) = \boldsymbol{P}(\boldsymbol{X})\boldsymbol{\Gamma}(\boldsymbol{X})\boldsymbol{P}(\boldsymbol{X})^T$. The matrix $\boldsymbol{\Gamma}(\boldsymbol{X})$ is the diagonal matrix whose diagonal elements are the descending-order eigenvalues of $\nabla f(\boldsymbol{X})$, denoted by $\gamma_1(\boldsymbol{X}) \geq \gamma_2(\boldsymbol{X}) \geq \cdots \geq \gamma_n(\boldsymbol{X})$. The jth column of $\boldsymbol{P}(\boldsymbol{X})$, denoted by $\boldsymbol{p}_j(\boldsymbol{X})$, is the associated eigenvector of $\gamma_j(\boldsymbol{X})$. We use $n_+(\boldsymbol{X})$ and $n_-(\boldsymbol{X})$ to denote the number of positive and non-positive eigenvalues of $\nabla f(\boldsymbol{X})$, respectively. We divide $\boldsymbol{\Gamma}(\boldsymbol{X})$ into the two blocks, $\boldsymbol{\Gamma}_+(\boldsymbol{X}) := diag(\gamma_1(\boldsymbol{X}), \gamma_2(\boldsymbol{X}), \ldots, \gamma_{n_+(\boldsymbol{X})})$, $\boldsymbol{\Gamma}_-(\boldsymbol{X}) := diag(\gamma_{n_+(\boldsymbol{X})+1}(\boldsymbol{X}), \gamma_{n_+(\boldsymbol{X})+2}(\boldsymbol{X}), \ldots, \gamma_n)$. Note that the sizes of $\boldsymbol{\Gamma}_+(\boldsymbol{X})$ and $\boldsymbol{\Gamma}_-(\boldsymbol{X})$ can be zero, but the total is always $n_+(\boldsymbol{X}) + n_-(\boldsymbol{X}) = n$. We also divide $\boldsymbol{P}(\boldsymbol{X})$ into the two matrices $\boldsymbol{P}_+(\boldsymbol{X}), \boldsymbol{P}_-(\boldsymbol{X})$ by collecting the corresponding vectors, so the columns of $\boldsymbol{P}_+(\boldsymbol{X})$ are $\boldsymbol{p}_1(\boldsymbol{X}), \ldots, \boldsymbol{p}_{n_+(\boldsymbol{X})}(\boldsymbol{X})$ in this order. As a property of eigenvectors, we have $\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{P}_-(\boldsymbol{X}) = \boldsymbol{O}$. We also know that $\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{P}_+(\boldsymbol{X})$ is the identity matrix of dimension $n_+(\boldsymbol{X})$ and $\boldsymbol{P}_-(\boldsymbol{X})^T\boldsymbol{P}_-(\boldsymbol{X})$ is the identity matrix of dimension $n_-(\boldsymbol{X})$. Finally, we define $\gamma_{\max}(\boldsymbol{X}) := ||\nabla f(\boldsymbol{X})||_2$. From the definition of the two-norm, it holds that $\gamma_{\max}(\boldsymbol{X}) = \max\{|\gamma_1(\boldsymbol{X})|, |\gamma_n(\boldsymbol{X})|\}$.

2 An Iterative Method Using Boundary Distance Information

For the simple bound problem (1.2), Coleman and Li [4] proposed a trust-region method which measures the distance from the current feasible point $\boldsymbol{x} \in \mathbb{R}^n$ to the boundary of the feasible set ($\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}$). They defined the vector $\boldsymbol{v}(\boldsymbol{x}) \in \mathbb{R}^n$ as

$$v_i(\boldsymbol{x}) = \begin{cases} x_i - l_i & \text{if } \frac{\partial f(\boldsymbol{x})}{\partial x_i} \geq 0 \text{ and } l_i > -\infty \\ x_i - u_i & \text{if } \frac{\partial f(\boldsymbol{x})}{\partial x_i} < 0 \text{ and } u_i < \infty \\ 1 & \text{if } \frac{\partial f(\boldsymbol{x})}{\partial x_i} \geq 0 \text{ and } l_i = -\infty \\ -1 & \text{if } \frac{\partial f(\boldsymbol{x})}{\partial x_i} < 0 \text{ and } u_i = \infty. \end{cases}$$

This vector was used to control the approach to the boundary, and the key observation in the discussion of [4] was that $\boldsymbol{x}^*$ satisfies the first-order optimality condition if and only if $\frac{\partial f(\boldsymbol{x})}{\partial x_i} v_i(\boldsymbol{x}) = 0$ for each $i = 1, \ldots, n$. Though the cases $l_i = -\infty$ or $u_i = \infty$ are considered in [4], we will focus the situation $l_i > -\infty$ and $u_i < \infty$ in this paper, since the box-constrained SDP (1.1) has a bounded feasible set.

We can not directly extend the definition of $\boldsymbol{v}(\boldsymbol{x})$ to the box-constrained SDPs (1.1) using the conditions on the eigenvalue of $\boldsymbol{X}$, since the distance to the boundary of $\mathcal{F}$ relates to not only the eigenvalues but also the eigenvectors. To take the effect of eigenvectors into account, we define two positive semidefinite matrices for $\boldsymbol{X} \in \mathcal{F}$;

$$\boldsymbol{V}_+(\boldsymbol{X}) := \boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_+(\boldsymbol{X}), \quad \text{and} \quad \boldsymbol{V}_-(\boldsymbol{X}) := \boldsymbol{P}_-(\boldsymbol{X})^T(\boldsymbol{I} - \boldsymbol{X})\boldsymbol{P}_-(\boldsymbol{X}).$$

The definition of these matrices brings us other properties of the first-order optimality condition in Lemma 2.1. In the lemma, we use a matrix $\boldsymbol{D}(\boldsymbol{X}) \in \mathbb{S}^n$ and a scalar $N(\boldsymbol{X})$ defined by

$$\boldsymbol{D}(\boldsymbol{X}) := \boldsymbol{P}(\boldsymbol{X}) \begin{pmatrix} \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} & \gamma_{\max}(\boldsymbol{X})\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X}) \\ \gamma_{\max}(\boldsymbol{X})\boldsymbol{P}_-(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_+(\boldsymbol{X}) & \boldsymbol{V}_-(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/2} \end{pmatrix} \boldsymbol{P}(\boldsymbol{X})^T \tag{2.1}$$

$$N(\boldsymbol{X}) := \langle \nabla f(\boldsymbol{X}) \mid \boldsymbol{D}(\boldsymbol{X}) \rangle. \tag{2.2}$$

The definition of the matrix $\boldsymbol{D}(\boldsymbol{X})$ includes the distance information to the boundary of the feasible sets $\mathcal{F}$ via the matrices $\boldsymbol{V}_+(\boldsymbol{X})$ and $\boldsymbol{V}_-(\boldsymbol{X})$ like $\boldsymbol{v}(\boldsymbol{x})$ above. In particular, $|v_i(\boldsymbol{x})|$ measures the distance of $\boldsymbol{x}$ from the lower (upper) bound $x_i - l_i \geq 0$ ($u_i - x_i \geq 0$) by $x_i - l_i$ ($u_i - x_i$) for the case $\frac{\partial f(\boldsymbol{x})}{\partial x_i} \geq 0$ ($\frac{\partial f(\boldsymbol{x})}{\partial x_i} < 0$), respectively. The matrices $\boldsymbol{V}_+(\boldsymbol{X})$ and $\boldsymbol{V}_-(\boldsymbol{X})$ use this concept; $\boldsymbol{V}_+(\boldsymbol{X})$ ($\boldsymbol{V}_-(\boldsymbol{X})$) evaluates the distance of $\boldsymbol{X}$ from the lower (upper) bound of $\mathcal{F}$, that is $\boldsymbol{X} - \boldsymbol{O} \succeq \boldsymbol{O}$ ($\boldsymbol{I} - \boldsymbol{X} \succeq \boldsymbol{O}$), with an adjustment by the eigenvectors $\boldsymbol{P}_+(\boldsymbol{X})$ ($\boldsymbol{P}_-(\boldsymbol{X})$) that correspond to the positive (non-positive) eigenvalues of $\nabla f(\boldsymbol{X})$, respectively.

Using the relations $\nabla f(\boldsymbol{X}) = \boldsymbol{P}_+(\boldsymbol{X})\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{P}_+(\boldsymbol{X})^T + \boldsymbol{P}_-(\boldsymbol{X})\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{P}_-(\boldsymbol{X})^T$, we can compute $||\boldsymbol{D}(\boldsymbol{X})||_F^2$ and $N(\boldsymbol{X})$ as follow;

$$\begin{aligned} ||\boldsymbol{D}(\boldsymbol{X})||_F^2 &= ||\boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2}||_F^2 + ||\boldsymbol{V}_-(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/2}||_F^2 \\ &\quad + 2(\gamma_{\max}(\boldsymbol{X}))^2||\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X})||_F^2, \end{aligned} \tag{2.3}$$

$$N(\boldsymbol{X}) = ||\boldsymbol{V}_+(\boldsymbol{X})^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/4}||_F^2 + ||\boldsymbol{V}_-(\boldsymbol{X})^{1/4}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/4}||_F^2. \tag{2.4}$$

Lemma 2.1. *For a matrix $\boldsymbol{X}^* \in \mathcal{F}$, the following conditions are equivalent.*

(a) $\boldsymbol{X}^*$ *satisfies the first-order optimality condition (1.3).*

(b) $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle = 0.$

(c) $N(\boldsymbol{X}^*) = 0.$

(d) $||\boldsymbol{D}(\boldsymbol{X}^*)||_F = 0.$

Proof. [$(a) \Rightarrow (b)$] We define a matrix $\widehat{\boldsymbol{X}} := \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T\boldsymbol{X}^*\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T$. Since $\boldsymbol{X}^* \in \mathcal{F}$, we obtain $\widehat{\boldsymbol{X}} \succeq \boldsymbol{O}$ and

$$\begin{aligned} \boldsymbol{I} - \widehat{\boldsymbol{X}} &= (\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T) \\ &\quad -(\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T\boldsymbol{X}^*\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T) \\ &= \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T(\boldsymbol{I} - \boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T \succeq \boldsymbol{O}, \end{aligned}$$

hence, $\widehat{\boldsymbol{X}} \in \mathcal{F}$. Substituting $\widehat{\boldsymbol{X}} \in \mathcal{F}$ into the inequality (1.3), we have

$$\begin{aligned} & \langle \nabla f(\boldsymbol{X}^*) \mid \widehat{\boldsymbol{X}} - \boldsymbol{X}^* \rangle \\ = \; & \langle \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{\Gamma}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{\Gamma}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T \\ & \quad \mid \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T\boldsymbol{X}^*\boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T - \boldsymbol{X}^* \rangle \\ = \; & \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{I} \rangle - \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{P}_-(\boldsymbol{X}^*)^T\boldsymbol{X}^*\boldsymbol{P}_-(\boldsymbol{X}^*) \rangle = \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle \geq 0. \end{aligned}$$

Here, we used $\langle \boldsymbol{A} \mid \boldsymbol{B} \rangle = Trace(\boldsymbol{A}^T\boldsymbol{B}) = Trace(\boldsymbol{B}^T\boldsymbol{A})$, $\boldsymbol{P}_+(\boldsymbol{X}^*)^T\boldsymbol{P}_+(\boldsymbol{X}^*) = \boldsymbol{I}$ and $\boldsymbol{P}_+(\boldsymbol{X}^*)^T\boldsymbol{P}_-(\boldsymbol{X}^*) = \boldsymbol{O}$. Since $-\boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \succeq \boldsymbol{O}$ and $\boldsymbol{V}_-(\boldsymbol{X}^*) \succeq \boldsymbol{O}$, we also have $\langle -\boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle \geq 0$, so that we obtain $\langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle = 0$.

Similarly, for the matrix $\overline{\boldsymbol{X}} := \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T\boldsymbol{X}^*\boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T \succeq \boldsymbol{O}$, we can show $\boldsymbol{I} - \overline{\boldsymbol{X}} = \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T(\boldsymbol{I} - \boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T \succeq \boldsymbol{O}$, therefore we have $\overline{\boldsymbol{X}} \in \mathcal{F}$. Putting $\overline{\boldsymbol{X}} \in \mathcal{F}$ into (1.3), we have $\langle \nabla f(\boldsymbol{X}^*) \mid \overline{\boldsymbol{X}} - \boldsymbol{X}^* \rangle = -\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle \geq 0$. On the other hand, from the properties $\boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \succeq \boldsymbol{O}$ and $\boldsymbol{V}_+(\boldsymbol{X}^*) \succeq \boldsymbol{O}$, it holds $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle \geq 0$. Hence, we obtain $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = 0$.

$[(b) \Rightarrow (a)]$ For any $\boldsymbol{X} \in \mathcal{F}$, it holds that

$$\begin{aligned} & \langle \nabla f(\boldsymbol{X}^*) \mid \boldsymbol{X} - \boldsymbol{X}^* \rangle \\ = \; & \langle \boldsymbol{P}_+(\boldsymbol{X}^*)\boldsymbol{\Gamma}_+(\boldsymbol{X}^*)\boldsymbol{P}_+(\boldsymbol{X}^*)^T + \boldsymbol{P}_-(\boldsymbol{X}^*)\boldsymbol{\Gamma}_-(\boldsymbol{X}^*)\boldsymbol{P}_-(\boldsymbol{X}^*)^T \mid \boldsymbol{X} - \boldsymbol{X}^* \rangle \\ = \; & \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{P}_+(\boldsymbol{X}^*)^T \boldsymbol{X} \boldsymbol{P}_+(\boldsymbol{X}^*) \rangle - \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle \\ & \quad - \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{P}_-(\boldsymbol{X}^*)^T (\boldsymbol{I} - \boldsymbol{X}) \boldsymbol{P}_-(\boldsymbol{X}^*) \rangle + \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle \\ = \; & \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{P}_+(\boldsymbol{X}^*)^T \boldsymbol{X} \boldsymbol{P}_+(\boldsymbol{X}^*) \rangle + \langle -\boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{P}_-(\boldsymbol{X}^*)^T (\boldsymbol{I} - \boldsymbol{X}) \boldsymbol{P}_-(\boldsymbol{X}^*) \rangle \; \geq 0. \end{aligned}$$

For the last equality, we used $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle = 0$ from (b). In addition, the last non-negativity came from $\boldsymbol{P}_+(\boldsymbol{X}^*)^T \boldsymbol{X} \boldsymbol{P}_+(\boldsymbol{X}^*) \succeq \boldsymbol{O}$ and $\boldsymbol{P}_-(\boldsymbol{X}^*)^T (\boldsymbol{I} - \boldsymbol{X}) \boldsymbol{P}_-(\boldsymbol{X}^*) \succeq \boldsymbol{O}$.

$[(b) \Rightarrow (c)]$ Since $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = Trace(\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2})$ and $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \succeq \boldsymbol{O}$, the condition $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = 0$ indicates all the eigenvalues of $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2}$ are 0, therefore, $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$. We now consider the eigenvalue decomposition $\boldsymbol{V}_+(\boldsymbol{X}^*) = \boldsymbol{Q} \boldsymbol{K} \boldsymbol{Q}^T$ such that
$\boldsymbol{K} = diag(\kappa_1, \kappa_2, \ldots, \kappa_{n_+(\boldsymbol{X}^*)})$ is the diagonal matrix with the eigenvalues of $\boldsymbol{V}_+(\boldsymbol{X}^*)$. Since $\boldsymbol{V}_+(\boldsymbol{X}^*) \succeq \boldsymbol{O}$, it holds that $\kappa_i \geq 0$ for $i = 1, \ldots, n_+(\boldsymbol{X}^*)$. We define a positive semidefinite matrix $\boldsymbol{W} := \boldsymbol{Q}^T \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{Q}$. Since the matrix $\boldsymbol{Q}$ is an orthogonal matrix, $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$ leads to $\boldsymbol{K}^{1/2} \boldsymbol{W} \boldsymbol{K}^{1/2} = \boldsymbol{O}$. By taking the diagonal elements, we know $\kappa_i^{1/2} W_{ii} \kappa_i^{1/2} = 0$ for $i = 1, \ldots, n_+(\boldsymbol{X}^*)$. Therefore, it holds that $\kappa_i^{1/4} W_{ii} \kappa_i^{1/4} = 0$. Since a matrix $\boldsymbol{K}^{1/4} \boldsymbol{W} \boldsymbol{K}^{1/4}$ is positive semidefinite and its diagonal elements are zero, we obtain $\boldsymbol{K}^{1/4} \boldsymbol{W} \boldsymbol{K}^{1/4} = \boldsymbol{O}$, hence, $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} = \boldsymbol{O}$. Similarly, the condition $\langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle = 0$ implies $\boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} = \boldsymbol{O}$. Hence, we obtain (c) by (2.4).

$[(c) \Rightarrow (b)]$ The condition $N(\boldsymbol{X}^*) = 0$ leads to $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} = \boldsymbol{O}$
and $\boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} = \boldsymbol{O}$. Hence, it holds that

$$\begin{aligned} \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle &= \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \rangle \\ &= \langle \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/4} \mid \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \rangle = 0. \end{aligned}$$

Similarly, we obtain $\langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \mid \boldsymbol{V}_-(\boldsymbol{X}^*) \rangle = 0$ from $\boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \boldsymbol{V}_-(\boldsymbol{X}^*)^{1/4} = \boldsymbol{O}$.

$[(b) \Rightarrow (d)]$ As a first step of $[(b) \Rightarrow (c)]$ above, we obtained $\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$ and
$\boldsymbol{V}_-(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \boldsymbol{V}_-(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$. Since all the eigenvalues in $\boldsymbol{\Gamma}_+(\boldsymbol{X}^*)$ are positive, the properties $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \mid \boldsymbol{V}_+(\boldsymbol{X}^*) \rangle = 0$ and $\boldsymbol{V}_+(\boldsymbol{X}^*) \succeq \boldsymbol{O}$ lead to $\boldsymbol{V}_+(\boldsymbol{X}^*) = \boldsymbol{O}$. Furthermore, the decomposition $\boldsymbol{V}_+(\boldsymbol{X}^*) = \boldsymbol{P}_+(\boldsymbol{X}^*)^T (\boldsymbol{X}^*)^{1/2} (\boldsymbol{X}^*)^{1/2} \boldsymbol{P}_+(\boldsymbol{X}^*) = \boldsymbol{O}$ implies
$\boldsymbol{P}_+(\boldsymbol{X}^*)^T (\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$. Therefore, it holds that
$\boldsymbol{P}_+(\boldsymbol{X}^*)^T \boldsymbol{X}^* \boldsymbol{P}_-(\boldsymbol{X}^*) = \boldsymbol{P}_+(\boldsymbol{X}^*)^T (\boldsymbol{X}^*)^{1/2} (\boldsymbol{X}^*)^{1/2} \boldsymbol{P}_-(\boldsymbol{X}^*) = \boldsymbol{O}$. Hence, we conclude $||\boldsymbol{D}(\boldsymbol{X}^*)||_F = 0$ from (2.3).

$[(d) \Rightarrow (b)]$ From the relation (2.3), the condition $||\boldsymbol{D}(\boldsymbol{X}^*)||_F = 0$ indicates
$\boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_+(\boldsymbol{X}^*) \boldsymbol{V}_+(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$ and $\boldsymbol{V}_-(\boldsymbol{X}^*)^{1/2} \boldsymbol{\Gamma}_-(\boldsymbol{X}^*) \boldsymbol{V}_-(\boldsymbol{X}^*)^{1/2} = \boldsymbol{O}$. By taking the traces of these matrices, we obtain (b). □

Lemma 2.1, (2.2) and (2.4) indicate that when $\boldsymbol{X}$ does not satisfy the first-order optimality condition, we can take $-\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}$ as a descent direction of $f(\boldsymbol{X})$, that is, $\langle \nabla f(\boldsymbol{X}) \mid -\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F} \rangle < 0$. Hence, we can expect that the decrease of the objective function $f(\boldsymbol{X} -$

$\alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}) < f(\boldsymbol{X})$ for a certain value $\alpha > 0$. The next lemma gives a non-zero range of α to ensure $\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F} \in \mathcal{F}$.

Lemma 2.2. *If $\boldsymbol{X} \in \mathcal{F}$ does not satisfy the first-order optimality condition, then $\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F} \in \mathcal{F}$ for $\alpha \in [0, \frac{||\boldsymbol{D}(\boldsymbol{X})||_F}{\gamma_{\max}(\boldsymbol{X})}]$.*

Proof. From the definition of $\gamma_{\max}(\boldsymbol{X})$, the matrix $\boldsymbol{I} - \frac{\boldsymbol{\Gamma}_+(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}$ is a nonnegative diagonal matrix, hence this matrix is positive semidefinite. Using $\boldsymbol{P}(\boldsymbol{X})\boldsymbol{P}(\boldsymbol{X})^T = \boldsymbol{I}$ and $-\boldsymbol{\Gamma}_-(\boldsymbol{X}) \succeq \boldsymbol{O}$, it holds

$$
\begin{aligned}
&\boldsymbol{X} - \frac{\boldsymbol{D}(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})} = \boldsymbol{P}(\boldsymbol{X})\boldsymbol{P}(\boldsymbol{X})^T\left(\boldsymbol{X} - \frac{\boldsymbol{D}(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\right)\boldsymbol{P}(\boldsymbol{X})\boldsymbol{P}(\boldsymbol{X})^T \\
&= \boldsymbol{P}(\boldsymbol{X})\left\{\boldsymbol{P}(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}(\boldsymbol{X}) - \begin{pmatrix} \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\frac{\boldsymbol{\Gamma}_+(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\boldsymbol{V}_+(\boldsymbol{X})^{1/2} & \boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X}) \\ \boldsymbol{P}_-(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_+(\boldsymbol{X}) & \boldsymbol{V}_-(\boldsymbol{X})^{1/2}\frac{\boldsymbol{\Gamma}_-(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\boldsymbol{V}_-(\boldsymbol{X})^{1/2} \end{pmatrix}\right\}\boldsymbol{P}(\boldsymbol{X})^T \\
&= \boldsymbol{P}(\boldsymbol{X})\begin{pmatrix} \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\left(\boldsymbol{I} - \frac{\boldsymbol{\Gamma}_+(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\right)\boldsymbol{V}_+(\boldsymbol{X})^{1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{P}_-(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X}) + \boldsymbol{V}_-(\boldsymbol{X})^{1/2}\frac{(-\boldsymbol{\Gamma}_-(\boldsymbol{X}))}{\gamma_{\max}(\boldsymbol{X})}\boldsymbol{V}_-(\boldsymbol{X})^{1/2} \end{pmatrix}\boldsymbol{P}(\boldsymbol{X})^T \\
&\succeq \boldsymbol{O}.
\end{aligned}
$$

In a similar way, noticing $\boldsymbol{P}_+(\boldsymbol{X})^T(\boldsymbol{I} - \boldsymbol{X})\boldsymbol{P}_-(\boldsymbol{X}) = -\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X})$ and $\boldsymbol{I} + \frac{\boldsymbol{\Gamma}_-(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})} \succeq \boldsymbol{O}$, we derive

$$
\begin{aligned}
&\boldsymbol{I} - \boldsymbol{X} + \frac{\boldsymbol{D}(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})} \\
&= \boldsymbol{P}(\boldsymbol{X})\begin{pmatrix} \boldsymbol{P}_+(\boldsymbol{X})^T(\boldsymbol{I} - \boldsymbol{X})\boldsymbol{P}_+(\boldsymbol{X}) + \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\frac{\boldsymbol{\Gamma}_+(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\boldsymbol{V}_+(\boldsymbol{X})^{1/2} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_-(\boldsymbol{X})^{1/2}\left(\boldsymbol{I} + \frac{\boldsymbol{\Gamma}_-(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\right)\boldsymbol{V}_-(\boldsymbol{X})^{1/2} \end{pmatrix}\boldsymbol{P}(\boldsymbol{X})^T \\
&\succeq \boldsymbol{O}.
\end{aligned}
$$

From two linear combinations

$$
\begin{aligned}
\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F} &= \left(1 - \alpha\frac{\gamma_{\max}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\right)\boldsymbol{X} \\
&\quad + \alpha\frac{\gamma_{\max}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\left(\boldsymbol{X} - \frac{\boldsymbol{D}(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\right) \\
\boldsymbol{I} - \left(\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\right) &= \left(1 - \alpha\frac{\gamma_{\max}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\right)(\boldsymbol{I} - \boldsymbol{X}) \\
&\quad + \alpha\frac{\gamma_{\max}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\left(\boldsymbol{I} - \boldsymbol{X} + \frac{\boldsymbol{D}(\boldsymbol{X})}{\gamma_{\max}(\boldsymbol{X})}\right),
\end{aligned}
$$

we obtain $\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F} \succeq \boldsymbol{O}$ and $\boldsymbol{I} - \left(\boldsymbol{X} - \alpha\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}\right) \succeq \boldsymbol{O}$ for $\alpha \in [0, \frac{||\boldsymbol{D}(\boldsymbol{X})||_F}{\gamma_{\max}(\boldsymbol{X})}]$. □

Based on the property that $-\frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}$ is a descent direction of $f(\boldsymbol{X})$, we can use $\boldsymbol{S}(\boldsymbol{X}) := \frac{\boldsymbol{D}(\boldsymbol{X})}{||\boldsymbol{D}(\boldsymbol{X})||_F}$ as a normalized search direction to find a minimizer.

We propose an iterative method for the box-constrained SDP (1.1) as Algorithm 2.3. In Algorithm 2.3, we use a quadratic approximation of f with the direction $-\boldsymbol{S}(X)$;

$$
q(\alpha, \boldsymbol{X}) := f(\boldsymbol{X}) - \alpha\langle\nabla f(\boldsymbol{X}) \mid \boldsymbol{S}(\boldsymbol{X})\rangle + \frac{\alpha^2}{2}\langle\boldsymbol{S}(\boldsymbol{X}) \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{S}(\boldsymbol{X})\rangle.
$$

Algorithm 2.3. An iterative method using boundary distance for box-constrained SDPs

Step 1: Choose an initial point $\boldsymbol{X}^0 \in \mathcal{F}$. Set an initial radius $\Delta_0 > 0$ and set a stopping threshold $\epsilon \geq 0$. Choose parameters $\mu_1, \mu_2, \eta_1, \eta_2$ such that $0 < \mu_1 < \mu_2 < 1$ and $0 < \eta_1 < 1 < \eta_2$. Set an iteration count $k = 0$.

Step 2: If $N(\boldsymbol{X}^k) \leq \epsilon$, output $\boldsymbol{X}^k$ as a solution and stop.

Step 3: Solve a quadratic problem with respect to α;

$$\min \quad q(\alpha, \boldsymbol{X}^k) \quad \text{subject to} \quad 0 \leq \alpha \leq \min \left\{ \frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)}, \Delta_k \right\}, \tag{2.5}$$

and let the step length α_k be the minimizer of (2.5).

Step 4: Let $\overline{\boldsymbol{X}}^k := \boldsymbol{X}^k - \alpha_k \boldsymbol{S}(\boldsymbol{X}^k)$ where $\boldsymbol{S}(\boldsymbol{X}^k) := \frac{\boldsymbol{D}(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}$. Compute the ratio

$$r_k := \frac{f(\boldsymbol{X}^k) - f(\overline{\boldsymbol{X}}^k)}{f(\boldsymbol{X}^k) - q(\alpha_k, \boldsymbol{X}^k)}, \tag{2.6}$$

and set

$$\boldsymbol{X}^{k+1} = \begin{cases} \overline{\boldsymbol{X}}^k & \text{if} \quad r_k \geq \mu_1 \\ \boldsymbol{X}^k & \text{otherwise.} \end{cases}$$

Step 5: Update the radius Δ_k by

$$\Delta_{k+1} = \begin{cases} \eta_1 \Delta_k & \text{if} \quad r_k < \mu_1 \\ \Delta_k & \text{if} \quad \mu_1 \leq r_k \leq \mu_2 \\ \eta_2 \Delta_k & \text{if} \quad r_k > \mu_2. \end{cases}$$

Step 6: Set $k \leftarrow k + 1$ and return to Step 2.

We should note that the quadratic approximation function $q(\alpha, \boldsymbol{X}^k)$ requires $\nabla^2 f(\boldsymbol{X}^k)$ in only the scalar value $\langle \boldsymbol{S}(\boldsymbol{X}^k) \mid \nabla^2 f(\boldsymbol{X}^k) \mid \boldsymbol{S}(\boldsymbol{X}^k) \rangle$. Hence, we do not always need to evaluate each element of $\nabla^2 f(\boldsymbol{X}^k)$ in each iteration. For example, for a function $\hat{f}(\boldsymbol{X}) = \cos(\langle \boldsymbol{X} \mid \boldsymbol{X} \rangle)$ and a symmetric matrix $\boldsymbol{S} \in \mathbb{S}^n$, it holds $\langle \boldsymbol{S} \mid \nabla^2 \hat{f}(\boldsymbol{X}) \mid \boldsymbol{S} \rangle = -2 \sin(\langle \boldsymbol{X} \mid \boldsymbol{X} \rangle) \langle \boldsymbol{S} \mid \boldsymbol{S} \rangle - 4 \cos(\langle \boldsymbol{X} \mid \boldsymbol{X} \rangle) \langle \boldsymbol{X} \mid \boldsymbol{S} \rangle^2$. This makes each iteration of Algorithm 2.3 low cost compared to the interior-point methods or the penalty barrier method.

We note that the generated sequence by Algorithm 2.3 remains in $\mathcal{F}$, that is, $\{\boldsymbol{X}^k\} \subset \mathcal{F}$ from Lemma 2.2. In Steps 4 and 5, we adjust the radius Δ_k. This adjustment is necessary to discuss the convergence properties.

3 Convergence properties

A matrix $\boldsymbol{X}^* \in \mathcal{F}$ satisfies the first-order optimality condition (1.3) if and only if $\underline{f}(\boldsymbol{X}^*) = 0$, as noted in Section 1. In this section, we show that the sequence $\{\boldsymbol{X}^k\} \subset \mathcal{F}$ generated by Algorithm 2.3 with the stopping threshold $\epsilon = 0$ attains $\lim_{k\to\infty} \underline{f}(\boldsymbol{X}^k) = 0$. We divide the proof into two parts. The first part shows there exists a subsequence of $\{N(\boldsymbol{X}^k)\}$ that converges to zero. The second part shows $\lim_{k\to\infty} N(\boldsymbol{X}^k) = 0$ in Theorem 3.2, and finally $\lim_{k\to\infty} \underline{f}(\boldsymbol{X}_k) = 0$ in Theorem 3.3. We should remark that if we use the threshold $\epsilon = 0$ and $\boldsymbol{X}^k$ at some iteration k satisfies $N(\boldsymbol{X}^k) = 0$ exactly, we can terminate Algorithm 2.3 since $\boldsymbol{X}^k$ satisfies the first-order optimality condition from Lemma 2.1. For the convergence analysis in this section, therefore, we discuss the case where $\{\boldsymbol{X}^k\} \subset \mathcal{F}$ is an infinite sequence.

Using the matrix $\boldsymbol{D}(\boldsymbol{X})$, we can employ similar approaches to [4] for the proof of the first part. However, we can not directly apply the results of [4] to the second part. This is mainly because that the eigenvector matrices $\boldsymbol{P}_+(\boldsymbol{X})$ and $\boldsymbol{P}_-(\boldsymbol{X})$ are not always continuous functions in $\boldsymbol{X}$. Instead, our proof relies on the boundedness of $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}) \mid \boldsymbol{V}_+(\boldsymbol{X})\rangle$ and $\langle -\boldsymbol{\Gamma}_-(\boldsymbol{X}) \mid \boldsymbol{V}_-(\boldsymbol{X})\rangle$.

3.1 Convergence of subsequence

To analyze Algorithm 2.3, we introduce two constant values

$$\begin{aligned} M_1 &:= \max_{\boldsymbol{X}\in\mathcal{F}} ||\nabla f(\boldsymbol{X})||_2, \\ M_2 &:= \max\left\{ \max_{\boldsymbol{X}\in\mathcal{F}, \boldsymbol{D}\in\mathbb{S}^n, \boldsymbol{D}\neq\boldsymbol{O}} \left| \frac{\langle \boldsymbol{D} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{D}\rangle}{\langle \boldsymbol{D} \mid \boldsymbol{D}\rangle} \right|, \omega \right\}. \end{aligned}$$

Here, ω is a small positive number.

The values M_1 and M_2 are finite from the assumptions that the feasible set $\mathcal{F}$ is a bounded and closed set and that the objective function $f(\boldsymbol{X})$ is a twice continuously differentiable function on an open set containing $\mathcal{F}$. We can assume that $M_1 > 0$ without loss of generality, since, if $M_1 = 0$, then $f(\boldsymbol{X})$ is a constant function in $\mathcal{F}$ and every point $\boldsymbol{X} \in \mathcal{F}$ is optimal. We remark that $\omega > 0$ in the definition of M_2 ensures that we can take $\frac{1}{M_2}$. We do not need to determine a specific value for ω, since M_1 and M_2 will be used for only the discussions of the convergence analysis and they do not appear in Algorithm 2.3. Hence, the value of ω does not affect numerical performance.

We now evaluate the quadratic approximation function $q(\alpha_k, \boldsymbol{X}^k)$.

Lemma 3.1. *The step length α_k in Step 3 satisfies*

$$q(\alpha_k, \boldsymbol{X}^k) \leq f(\boldsymbol{X}^k) - \frac{1}{2} \min\left\{ \frac{N(\boldsymbol{X}^k)^2}{M_2 ||\boldsymbol{D}(\boldsymbol{X}^k)||_F^2}, \frac{N(\boldsymbol{X}^k)}{\gamma_{\max}(\boldsymbol{X}^k)}, \frac{\Delta_k N(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} \right\}.$$

Proof. We define a quadratic function $\phi(\alpha) := -\alpha \frac{N(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} + \frac{\alpha^2}{2} M_2$. From the definitions of $N(\boldsymbol{X}^k)$ and M_2, we have $q(\alpha, \boldsymbol{X}^k) \leq f(\boldsymbol{X}^k) + \phi(\alpha)$, hence,

$$q(\alpha_k, \boldsymbol{X}^k) \leq f(\boldsymbol{X}^k) + \min_{\alpha \in \left[0, \min\left\{ \frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)}, \Delta_k \right\}\right]} \phi(\alpha).$$

Since $N(\boldsymbol{X}^k) = \langle \nabla f(\boldsymbol{X}^k) \mid \boldsymbol{D}(\boldsymbol{X}^k) \rangle > 0$ (otherwise, Lemma 2.1 indicates that $\boldsymbol{X}^k$ already satisfies the first-order optimality condition) and $\phi(\alpha)$ is a quadratic function with respect to α, the minimum of ϕ is attained at one of the three candidates $\frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)}, \Delta_k$ or $\widehat{\alpha} := \frac{N(\boldsymbol{X}^k)}{M_2||\boldsymbol{D}(\boldsymbol{X}^k)||_F}$. Let $\alpha_{\min}$ be the minimizer of $\phi(\alpha)$ subject to $0 \leq \alpha \leq \min\left\{ \frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)}, \Delta_k \right\}$.

If $\alpha_{\min} = \widehat{\alpha}$, we have $\phi(\widehat{\alpha}) = -\frac{1}{2}\frac{N(\boldsymbol{X}^k)^2}{M_2||\boldsymbol{D}(\boldsymbol{X}^k)||_F^2}$. Next, if $\alpha_{\min} = \frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}}$, we have $\frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)} \leq \widehat{\alpha}$, therefore, $\frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F^2}{\gamma_{\max}(\boldsymbol{X}^k)} M_2 \leq N(\boldsymbol{X}^k)$. Hence, it holds that $\phi\left(\frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{\gamma_{\max}(\boldsymbol{X}^k)}\right) = -\frac{N(\boldsymbol{X}^k)}{\gamma_{\max}(\boldsymbol{X}^k)} + \frac{1}{2}\frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F^2}{\gamma_{max}(\boldsymbol{X}^k)^2} M_2 \leq -\frac{1}{2}\frac{N(\boldsymbol{X}^k)}{\gamma_{\max}(\boldsymbol{X}^k)}$. Finally, if $\alpha_{\min} = \Delta_k$, the inequality $\Delta_k \leq \widehat{\alpha}$ indicates that $\Delta_k \leq \frac{N(\boldsymbol{X}^k)}{M_2||\boldsymbol{D}(\boldsymbol{X}^k)||_F}$. Hence, it holds that $\phi(\Delta_k) = -\Delta_k \frac{N(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} + \frac{1}{2}\Delta_k^2 M_2 \leq -\Delta_k \frac{N(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} + \frac{1}{2}\Delta_k \frac{N(\boldsymbol{X}^k)}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} \leq -\frac{1}{2}\frac{\Delta_k N(\boldsymbol{X}^k)}{||D(\boldsymbol{X}^k)||_F}$.

Taking the maximum of the three cases, we obtain the inequality of this lemma. □

To simplify the inequality of Lemma 3.1, we replace $\gamma_{\max}(\boldsymbol{X}^k)$ and $||\boldsymbol{D}(\boldsymbol{X}^k)||_F$ by convenient upper bounds. Since $\gamma_{\max}(\boldsymbol{X}^k)$ is bounded by M_1, we consider an upper bound on $||\boldsymbol{D}(\boldsymbol{X}^k)||_F$.

Lemma 3.2. *For $\boldsymbol{X} \in \mathcal{F}$, it holds that $||\boldsymbol{D}(\boldsymbol{X})||_F^2 \leq N(\boldsymbol{X}) + \frac{1}{2}M_1^2 n^3$.*

Proof. Let $\boldsymbol{V}_+(\boldsymbol{X}) = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^T$ be the eigenvalue decomposition of $\boldsymbol{V}_+(\boldsymbol{X})$ such that $\boldsymbol{K} = diag(\kappa_1, \kappa_2, \ldots, \kappa_{n+(\boldsymbol{X})})$ is the diagonal matrix with the eigenvalues of $\boldsymbol{V}_+(\boldsymbol{X})$. Since $\boldsymbol{O} \preceq \boldsymbol{X} \preceq \boldsymbol{I}$, we have $\boldsymbol{O} \preceq \boldsymbol{V}_+(\boldsymbol{X}) \preceq \boldsymbol{I}$, hence, $0 \leq \kappa_i \leq 1$ for $i = 1, 2 \ldots, n_+(\boldsymbol{X})$. Using a matrix $\boldsymbol{W} := \boldsymbol{Q}^T\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{Q}$, we show $||\boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2}||_F \leq ||\boldsymbol{V}_+(\boldsymbol{X})^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/4}||_F$;

$$
\begin{aligned}
& ||\boldsymbol{V}_+(\boldsymbol{X})^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/4}||_F^2 - ||\boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2}||_F^2 \\
= \; & \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}) \mid \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} \rangle \\
& -\langle \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} \mid \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} \rangle \\
= \; & \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}) - \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} \mid \boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2} \rangle \\
= \; & \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}) - \boldsymbol{Q}\boldsymbol{K}^{1/2}\boldsymbol{Q}^T\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{Q}\boldsymbol{K}^{1/2}\boldsymbol{Q}^T \mid \boldsymbol{Q}\boldsymbol{K}^{1/2}\boldsymbol{Q}^T\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{Q}\boldsymbol{K}^{1/2}\boldsymbol{Q}^T \rangle \\
= \; & \langle \boldsymbol{W} - \boldsymbol{K}^{1/2}\boldsymbol{W}\boldsymbol{K}^{1/2} \mid \boldsymbol{K}^{1/2}\boldsymbol{W}\boldsymbol{K}^{1/2} \rangle \\
= \; & ||\boldsymbol{K}^{1/4}\boldsymbol{W}\boldsymbol{K}^{1/4}||_F^2 - ||\boldsymbol{K}^{1/2}\boldsymbol{W}\boldsymbol{K}^{1/2}||_F^2 \\
= \; & \sum_{i=1}^{n_+(\boldsymbol{X})} \sum_{j=1}^{n_+(\boldsymbol{X})} (W_{ij}\kappa_i^{1/4}\kappa_j^{1/4})^2 - \sum_{i=1}^{n_+(\boldsymbol{X})} \sum_{j=1}^{n_+(\boldsymbol{X})} (W_{ij}\kappa_i^{1/2}\kappa_j^{1/2})^2 \\
= \; & \sum_{i=1}^{n_+(\boldsymbol{X})} \sum_{j=1}^{n_+(\boldsymbol{X})} W_{ij}^2(\kappa_i^{1/2}\kappa_j^{1/2} - \kappa_i\kappa_j) \geq 0.
\end{aligned}
$$

The last inequality comes from $0 \leq \kappa_i \leq 1$ for $i = 1, \ldots, n_+(\boldsymbol{X})$. In a similar way, we also derive $||\boldsymbol{V}_-(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/2}||_F^2 \leq ||\boldsymbol{V}_-(\boldsymbol{X})^{1/4}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/4}||_F^2$. We evaluate the last term of (2.3) by a property of the Frobenius norm,

$$
||\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X})||_F^2 \leq ||\boldsymbol{P}_+(\boldsymbol{X})||_F^2 \cdot ||\boldsymbol{X}||_F^2 \cdot ||\boldsymbol{P}_-(\boldsymbol{X})||_F^2 \leq n_+(\boldsymbol{X}) \cdot n \cdot n_-(\boldsymbol{X}) \leq \frac{n^3}{4}.
$$

Here, we used $||\boldsymbol{P}_+(\boldsymbol{X})||_F^2 = Trace(\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{P}_+(\boldsymbol{X})) = n_+(\boldsymbol{X})$. In addition, we used $||\boldsymbol{X}||_F \leq \sqrt{n}||\boldsymbol{X}||_2$ from [20, (1.2.27)] and $\boldsymbol{O} \preceq \boldsymbol{X} \preceq \boldsymbol{I}$ to derive $||\boldsymbol{X}||_F^2 \leq n$, and we used the relation $n_+(\boldsymbol{X}) + n_-(\boldsymbol{X}) = n$ to derive $n_+(\boldsymbol{X}) \cdot n_-(\boldsymbol{X}) \leq \frac{n^2}{4}$.

Consequently, it holds from (2.3) that

$$\begin{aligned}
||\boldsymbol{D}(\boldsymbol{X})||_F^2 &= ||\boldsymbol{V}_+(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/2}||_F^2 + ||\boldsymbol{V}_-(\boldsymbol{X})^{1/2}\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/2}||_F^2 \\
&\quad + 2\gamma_{\max}(\boldsymbol{X})^2||\boldsymbol{P}_+(\boldsymbol{X})^T\boldsymbol{X}\boldsymbol{P}_-(\boldsymbol{X})||_F^2 \\
&\leq ||\boldsymbol{V}_+^{1/4}(\boldsymbol{X})\boldsymbol{\Gamma}_+(\boldsymbol{X})\boldsymbol{V}_+(\boldsymbol{X})^{1/4}||_F^2 \\
&\quad + ||\boldsymbol{V}_-^{1/4}(\boldsymbol{X})\boldsymbol{\Gamma}_-(\boldsymbol{X})\boldsymbol{V}_-(\boldsymbol{X})^{1/4}||_F^2 + 2\gamma_{\max}(\boldsymbol{X})^2\frac{n^3}{4} \\
&\leq N(\boldsymbol{X}) + \frac{1}{2}M_1^2n^3.
\end{aligned}$$

□

We put Lemma 3.2 into Lemma 3.1 to obtain a new upper bound on $q(\alpha_k, \boldsymbol{X}^k)$;

$$\begin{aligned}
q(\alpha_k, \boldsymbol{X}^k) &\leq f(\boldsymbol{X}^k) \\
&\quad - \frac{1}{2}\min\left\{\frac{N(\boldsymbol{X}^k)^2}{M_2\left(N(\boldsymbol{X}^k) + \frac{1}{2}M_1^2n^3\right)}, \frac{N(\boldsymbol{X}^k)}{M_1}, \frac{\Delta_k N(\boldsymbol{X}^k)}{\sqrt{N(\boldsymbol{X}^k) + \frac{1}{2}M_1^2n^3}}\right\}. \qquad (3.1)
\end{aligned}$$

In Algorithm 2.3, we call the kth iteration a *successful* iteration if $\boldsymbol{X}^{k+1}$ is set as $\overline{\boldsymbol{X}}^k$ in Step 4, that is, $r_k \geq \mu_1$. Otherwise, the kth iteration is called an *unsuccessful* iteration. For a successful iteration, we obtain a decrease in the objective function

$$\begin{aligned}
f(\boldsymbol{X}^{k+1}) &\leq f(\boldsymbol{X}^k) - \mu_1(f(\boldsymbol{X}^k) - q(\alpha_k, \boldsymbol{X}^k)) \\
&\leq f(\boldsymbol{X}^k) \\
&\quad - \frac{\mu_1}{2}\min\left\{\frac{N(\boldsymbol{X}^k)^2}{M_2\left(N(\boldsymbol{X}^k) + \frac{1}{2}M_1^2n^3\right)}, \frac{N(\boldsymbol{X}^k)}{M_1}, \frac{\Delta_k N(\boldsymbol{X}^k)}{\sqrt{N(\boldsymbol{X}^k) + \frac{1}{2}M_1^2n^3}}\right\}. \qquad (3.2)
\end{aligned}$$

Since it holds $f(\boldsymbol{X}^{k+1}) = f(\boldsymbol{X}^k)$ for an unsuccessful iteration, the objective value $f(\boldsymbol{X}^k)$ is non-increasing in Algorithm 2.3.

We are now prepared to show that there exists a subsequence of $\{N(\boldsymbol{X}^k)\}$ that converges to zero.

Theorem 3.1. *When the sequence $\{\boldsymbol{X}^k\}$ generated by Algorithm 2.3 with the stopping threshold $\epsilon = 0$ is an infinite sequence, it holds that*

$$\liminf_{k\to\infty} N(\boldsymbol{X}^k) = 0.$$

Proof: We assume that there exists $\hat{\epsilon} > 0$ such that $N(\boldsymbol{X}^k) \geq \hat{\epsilon}$ for any $k \geq 0$, and we will derive a contradiction.

Let $\mathcal{K} = \{k_1, k_2, \ldots, k_i, \ldots\}$ be the successful iterations. If $\mathcal{K}$ is a finite sequence, let k_i be the last iteration of $\mathcal{K}$. Since all of the iterations after k_i are unsuccessful, the update rule of Δ_k (Step 5 of Algorithm 2.3) implies $\Delta_{k_i+j} = \eta_1^j\Delta_{k_i}$. Hence, we obtain $\lim_{j\to\infty}\Delta_j = 0$.

Next, we consider the case when $\mathcal{K}$ is an infinite sequence. The function $\frac{x^2}{x+\frac{1}{2}M_1^2n^3}$ is an increasing function for $x > 0$, so that it holds from (3.2) that for $k_i \in \mathcal{K}$,

$$f(\boldsymbol{X}^{k_i+1}) \leq f(\boldsymbol{X}^{k_i}) - \frac{\mu_1}{2}\min\left\{\frac{N(\boldsymbol{X}^{k_i})^2}{M_2\left(N(\boldsymbol{X}^{k_i})+\frac{1}{2}M_1^2n^3\right)}, \frac{N(\boldsymbol{X}^{k_i})}{M_1}, \frac{\Delta_{k_i}N(\boldsymbol{X}^{k_i})}{\sqrt{N(\boldsymbol{X}^{k_i})+\frac{1}{2}M_1^2n^3}}\right\}$$
$$\leq f(\boldsymbol{X}^{k_i}) - \frac{\mu_1}{2}\min\left\{\frac{\hat{\epsilon}^2}{M_2\left(\hat{\epsilon}+\frac{1}{2}M_1^2n^3\right)}, \frac{\hat{\epsilon}}{M_1}, \frac{\Delta_{k_i}\hat{\epsilon}}{\sqrt{\hat{\epsilon}+\frac{1}{2}M_1^2n^3}}\right\}.$$

Since f is continuous on a closed set $\mathcal{F}$ and $\boldsymbol{X}^k \in \mathcal{F}$ for each k, $f(\boldsymbol{X}^{k_i})$ is bounded below. Therefore, it holds $\lim_{i\to\infty}\Delta_{k_i} = 0$. From Step 5 of Algorithm 2.3, it holds that $\Delta_j \leq \eta_2\Delta_{k_i}$ for the unsuccessful iterations $j = k_i+1\ldots, k_{i+1}-1$. Hence, we obtain $\lim_{j\to\infty}\Delta_j = 0$, regardless of the finiteness of $\mathcal{K}$. From (3.1) and $N(\boldsymbol{X}^k) \geq \hat{\epsilon}$, it holds for sufficiently large k that

$$f(\boldsymbol{X}^k) - q(\alpha_k, \boldsymbol{X}^k) \geq \frac{1}{2}\frac{\Delta_k\hat{\epsilon}}{\sqrt{\hat{\epsilon}+\frac{1}{2}M_1^2n^3}} > 0. \tag{3.3}$$

We will take a close look at the ratio r_k. From the Taylor expansion, there exists $\xi \in (0,1)$ such that

$$f(\boldsymbol{X}^k - \alpha_k\boldsymbol{S}(\boldsymbol{X}^k)) = f(\boldsymbol{X}^k) - \alpha_k\langle\nabla f(\boldsymbol{X}^k) \mid \boldsymbol{S}(\boldsymbol{X}^k)\rangle$$
$$+ \frac{\alpha_k^2}{2}\langle\boldsymbol{S}(\boldsymbol{X}^k) \mid \nabla^2 f(\boldsymbol{X}^k - \xi\alpha_k\boldsymbol{S}(\boldsymbol{X}^k)) \mid \boldsymbol{S}(\boldsymbol{X}^k)\rangle.$$

Therefore,

$$|f(\overline{\boldsymbol{X}}^k) - q(\alpha^k, \boldsymbol{X}^k)| \leq \frac{\alpha_k^2}{2}\Big|\langle\boldsymbol{S}(\boldsymbol{X}^k) \mid \nabla^2 f(\boldsymbol{X}^k - \xi\alpha_k\boldsymbol{S}(\boldsymbol{X}^k)) \mid \boldsymbol{S}(\boldsymbol{X}^k)\rangle$$
$$-\langle\boldsymbol{S}(\boldsymbol{X}^k) \mid \nabla^2 f(\boldsymbol{X}^k) \mid \boldsymbol{S}(\boldsymbol{X}^k)\rangle\Big|$$
$$\leq \frac{\Delta_k^2}{2}(M_2 + M_2) = \Delta_k^2 M_2.$$

Using (3.3) in the denominator, the ratio r_k is evaluated by

$$|r_k - 1| = \frac{|f(\overline{\boldsymbol{X}}^k) - q(\alpha_k, \boldsymbol{X}^k)|}{|f(\boldsymbol{X}^k) - q(\alpha_k, \boldsymbol{X}^k)|} \leq \frac{\Delta_k^2 M_2}{\frac{1}{2}\frac{\Delta_k\hat{\epsilon}}{\sqrt{\hat{\epsilon}+\frac{1}{2}M_1^2n^3}}} = \Delta_k\frac{2M_2\sqrt{\hat{\epsilon}+\frac{1}{2}M_1^2n^3}}{\hat{\epsilon}}.$$

Therefore, $\lim_{k\to\infty}\Delta_k = 0$ leads to $\lim_{k\to\infty} r_k = 1 > \mu_2$. From Step 5 of Algorithm 2.3, we have $\Delta_{k+1} = \eta_2\Delta_k \geq \Delta_k$ for sufficiently large k. Thus, there exists $\hat{k}_0$ such that $\Delta_k \geq \Delta_{\hat{k}_0}$ for $\forall k \geq \hat{k}_0$, but this contradicts $\lim_{k\to\infty}\Delta_k = 0$. Hence, $\liminf_{k\to\infty} N(\boldsymbol{X}^k) = 0$. □

3.2 Convergence of the whole sequence

Using the convergence of the subsequence, we will show in Theorem 3.2 that the whole sequence of $\{N(\boldsymbol{X}^k)\}$ converges to zero. We use the following two lemmas to prove Theorem 3.2.

Lemma 3.3. *For $\boldsymbol{X} \in \mathcal{F}$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n$, we have*

$$\begin{aligned} |\langle \nabla f(\boldsymbol{X}) \mid \boldsymbol{A} \rangle| &\leq \sqrt{n} M_1 ||\boldsymbol{A}||_F \\ \left|\langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle\right| &\leq 3M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F. \end{aligned}$$

Proof. The first inequality holds by $|\langle \nabla f(\boldsymbol{X}) \mid \boldsymbol{A} \rangle| \leq ||\nabla f(\boldsymbol{X})||_F ||\boldsymbol{A}||_F$ and $||\nabla f(\boldsymbol{X})||_F \leq \sqrt{n}||\nabla f(\boldsymbol{X})||_2$ from [20, (1.2.27)].

For the second inequality, we start with the following inequality derived from the definition of M_2;

$$\left|\langle \boldsymbol{D} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{D} \rangle\right| \leq M_2 ||\boldsymbol{D}||_F^2 \quad \text{for} \quad \forall \boldsymbol{D} \in \mathbb{S}^n.$$

Therefore, we get $\left|\langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{A} \rangle\right| \leq M_2 ||\boldsymbol{A}||_F^2$ and $\left|\langle \boldsymbol{B} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle\right| \leq M_2 ||\boldsymbol{B}||_F^2$. Furthermore, we put $\boldsymbol{A} - t\boldsymbol{B}$ into $\boldsymbol{D}$ to obtain the following inequality, which holds for any $t \in \mathbb{R}$;

$$\left|\langle \boldsymbol{A} - t\boldsymbol{B} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{A} - t\boldsymbol{B} \rangle\right| \leq M_2 ||\boldsymbol{A} - t\boldsymbol{B}||_F^2.$$

Therefore, the inequality

$$\begin{aligned} (M_2||\boldsymbol{B}||_F^2 - \langle \boldsymbol{B} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle) t^2 - 2(M_2 \langle \boldsymbol{A} \mid \boldsymbol{B} \rangle - \langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle) t & \\ + (M_2 ||\boldsymbol{A}||_F^2 - \langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{A} \rangle) \quad &\geq \quad 0 \end{aligned}$$

holds for any $t \in \mathbb{R}$, and we can derive

$$\begin{aligned} & (M_2 \langle \boldsymbol{A} \mid \boldsymbol{B} \rangle - \langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle)^2 \\ \leq \; & \left(M_2 ||\boldsymbol{A}||_F^2 - \langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{A} \rangle\right) \left(M_2 ||\boldsymbol{B}||_F^2 - \langle \boldsymbol{B} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle\right) \\ \leq \; & (2M_2 ||\boldsymbol{A}||_F^2)(2M_2 ||\boldsymbol{B}||_F^2). \end{aligned}$$

Consequently, it holds that

$$\begin{aligned} \langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle &\leq M_2 \langle \boldsymbol{A} \mid \boldsymbol{B} \rangle + \sqrt{(2M_2 ||\boldsymbol{A}||_F^2)(2M_2 ||\boldsymbol{B}||_F^2)} \\ &\leq M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F + 2M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F = 3M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F \end{aligned}$$

In addition, we replace $\boldsymbol{A}$ with $-\boldsymbol{A}$ to obtain

$$\langle -\boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle \leq 3M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F.$$

By combining these inequalities, we get $\left|\langle \boldsymbol{A} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{B} \rangle\right| \leq 3M_2 ||\boldsymbol{A}||_F ||\boldsymbol{B}||_F$. □

Lemma 3.4. *For $\boldsymbol{X}^k \in \mathcal{F}$, it holds that $\underline{f}(\boldsymbol{X}^k) \geq -n\sqrt{N(\boldsymbol{X}^k)}$.*

Proof. The objective function of (1.4) at $\boldsymbol{X} \in \mathcal{F}$ can be evaluated from below by

$$\begin{aligned} & \langle \nabla f(\boldsymbol{X}^k) \mid \boldsymbol{X} - \boldsymbol{X}^k \rangle \\ &= \langle \boldsymbol{P}_+(\boldsymbol{X}^k) \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \boldsymbol{P}_+(\boldsymbol{X}^k)^T \\ &\quad + \boldsymbol{P}_-(\boldsymbol{X}^k) \boldsymbol{\Gamma}_-(\boldsymbol{X}^k) \boldsymbol{P}_-(\boldsymbol{X}^k)^T \mid \boldsymbol{X} - \boldsymbol{X}^k \rangle \\ &= \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \mid \boldsymbol{P}_+(\boldsymbol{X}^k)^T \boldsymbol{X} \boldsymbol{P}_+(\boldsymbol{X}^k) \rangle - \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^k) \mid \boldsymbol{P}_-(\boldsymbol{X}^k)^T (\boldsymbol{I} - \boldsymbol{X}) \boldsymbol{P}_-(\boldsymbol{X}^k) \rangle \\ &\quad - \langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \mid \boldsymbol{V}_+(\boldsymbol{X}^k) \rangle + \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^k) \mid \boldsymbol{V}_-(\boldsymbol{X}^k) \rangle \end{aligned}$$

$$\geq -\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \mid \boldsymbol{V}_+(\boldsymbol{X}^k)\rangle + \langle \boldsymbol{\Gamma}_-(\boldsymbol{X}^k) \mid \boldsymbol{V}_-(\boldsymbol{X}^k)\rangle.$$

Furthermore, an upper bound of $\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \mid \boldsymbol{V}_+(\boldsymbol{X}^k)\rangle$ is given by

$$\begin{aligned}\langle \boldsymbol{\Gamma}_+(\boldsymbol{X}^k) \mid \boldsymbol{V}_+(\boldsymbol{X}^k)\rangle &= Trace(\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X}^k)\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4})\\ &\leq ||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X}^k)\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F\\ &\leq n_+(\boldsymbol{X}^k)||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X}^k)\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F.\end{aligned}$$

Here, we used $||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F^2 \leq n_+(\boldsymbol{X}^k)$ derived from $\boldsymbol{O} \preceq \boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4} \preceq \boldsymbol{I}$. In a similar way, it also holds $\langle -\boldsymbol{\Gamma}_-(\boldsymbol{X}^k) \mid \boldsymbol{V}_-(\boldsymbol{X}^k)\rangle \leq n_-(\boldsymbol{X}^k)||\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_-(\boldsymbol{X}^k)\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}||_F$.

Therefore, we obtain

$$\begin{aligned}\underline{f}(\boldsymbol{X}^k) &\geq -n_+(\boldsymbol{X}^k)||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X}^k)\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F\\ &\quad - n_-(\boldsymbol{X}^k)||\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_-(\boldsymbol{X}^k)\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}||_F\\ &\geq -(n_+(\boldsymbol{X}^k)+n_-(\boldsymbol{X}^k))\\ &\quad \times \sqrt{||\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_+(\boldsymbol{X}^k)\boldsymbol{V}_+(\boldsymbol{X}^k)^{1/4}||_F^2 + ||\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}\boldsymbol{\Gamma}_-(\boldsymbol{X}^k)\boldsymbol{V}_-(\boldsymbol{X}^k)^{1/4}||_F^2}\\ &= -n\sqrt{N(\boldsymbol{X}^k)}.\end{aligned}$$

For the second inequality, we used an inequality $ab+cd \leq (a+c)\sqrt{b^2+d^2}$ for $a, b, c, d \geq 0$. □

We are ready to prove the convergence of the whole sequence.

Theorem 3.2. *When the sequence $\{\boldsymbol{X}^k\}$ generated by Algorithm 2.3 with $\epsilon = 0$ is an infinite sequence, it holds that*

$$\lim_{k\to\infty} N(\boldsymbol{X}^k) = 0.$$

Proof. To derive a contradiction, we assume that there exist a positive number ϵ_1 and an infinite subsequence $\mathcal{K} := \{k_1, k_2, \ldots, k_i, \ldots\} \subset \{1, 2, \ldots\}$ such that $N(\boldsymbol{X}^{k_i}) \geq \epsilon_1$ for $\forall k_i \in \mathcal{K}$ and $0 < \epsilon_1 \leq 16n_2M_1^2$ hold.

From Theorem 3.1, we can take a subsequence $\mathcal{L} := \{l_1, l_2, \ldots, l_i, \ldots\} \subset \{1, 2, \ldots\}$ such that

$$\begin{cases} N(\boldsymbol{X}^k) & \geq & \epsilon_2^2 & \text{for} \quad k = k_i, k_i+1, \ldots, l_i-1 \\ N(\boldsymbol{X}^{l_i}) & < & \epsilon_2^2. & \end{cases}$$

where $\epsilon_2 := \frac{\epsilon_1}{4nM_1}$. Note that this is consistent with $N(\boldsymbol{X}^{k_i}) \geq \epsilon_1$, since we took $0 < \epsilon_1 \leq 16n^2M_1^2$.

If the kth iteration is a successful iteration and $k_i \leq k < l_i$, we put $N(\boldsymbol{X}^k) \geq \epsilon_2^2$ into (3.2) and obtain

$$f(\boldsymbol{X}^{k+1}) \quad \leq \quad f(\boldsymbol{X}^k) - \frac{\mu_1}{2}\min\left\{\frac{\epsilon_2^4}{M_2\left(\epsilon_2^2 + \frac{1}{2}M_1^2n^3\right)}, \frac{\epsilon_2^2}{M_1}, \frac{\Delta_k\epsilon_2^2}{\sqrt{\epsilon_2^2 + \frac{1}{2}M_1^2n^3}}\right\}.$$

Since f is bounded below and the sequence $\{f(\boldsymbol{X}^k)\}$ is non-increasing, $\lim_{k\to\infty}(f(\boldsymbol{X}^{k+1}) - f(\boldsymbol{X}^k)) = 0$. Hence, it holds that

$$0 = \lim_{k\to\infty}(f(\boldsymbol{X}^k) - f(\boldsymbol{X}^{k+1})) \geq \lim_{k\to\infty}\frac{\mu_1}{2}\min\left\{\frac{\epsilon_2^4}{M_2\left(\epsilon_2^2+\frac{1}{2}M_1^2n^3\right)}, \frac{\epsilon_2^2}{M_1}, \frac{\Delta_k\epsilon_2^2}{\sqrt{\epsilon_2^2+\frac{1}{2}M_1^2n^3}}\right\} \geq 0.$$

Since the two terms $\frac{\epsilon_2^4}{M_2\left(\epsilon_2^2+\frac{1}{2}M_1^2n^3\right)}$ and $\frac{\epsilon_2^2}{M_1}$ are constant with repect to k, we know $\lim_{k\to\infty}\Delta_k = 0$. Therefore, we have $\frac{\epsilon_2^4}{M_2\left(\epsilon_2^2+\frac{1}{2}M_1^2n^3\right)} \geq \frac{\Delta_k\epsilon_2^2}{\sqrt{\epsilon_2^2+\frac{1}{2}M_1^2n^3}}$ and $\frac{\epsilon_2^2}{M_1} \geq \frac{\Delta_k\epsilon_2^2}{\sqrt{\epsilon_2^2+\frac{1}{2}M_1^2n^3}}$ for sufficiently large k.

Hence, if k is sufficiently large, it holds that

$$f(\boldsymbol{X}^{k+1}) \quad \leq \quad f(\boldsymbol{X}^k) - \Delta_k\epsilon_3$$

where $\epsilon_3 := \frac{\mu_1}{2}\frac{\epsilon_2^2}{\sqrt{\epsilon_2^2+\frac{1}{2}M_1^2n^3}}$. We update the matrix with $\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \alpha_k\boldsymbol{S}(\boldsymbol{X}^k)$ in a successful iteration, therefore, we use $\alpha_k \leq \Delta_k$ and $||\boldsymbol{S}(\boldsymbol{X}^k)||_F = \frac{||\boldsymbol{D}(\boldsymbol{X}^k)||_F}{||\boldsymbol{D}(\boldsymbol{X}^k)||_F} = 1$ to derive

$$||\boldsymbol{X}^k - \boldsymbol{X}^{k+1}||_F \leq \Delta_k \leq \frac{f(\boldsymbol{X}^k) - f(\boldsymbol{X}^{k+1})}{\epsilon_3}.$$

The inequality $||\boldsymbol{X}^k - \boldsymbol{X}^{k+1}||_F \leq \frac{f(\boldsymbol{X}^k)-f(\boldsymbol{X}^{k+1})}{\epsilon_3}$ is also valid when the kth iteration is an unsuccessful iteration, since the matrix is updated with $\boldsymbol{X}^{k+1} = \boldsymbol{X}^k$. Hence, it holds that

$$\begin{aligned}
& ||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}||_F \\
\leq\ & ||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{k_i+1}||_F + ||\boldsymbol{X}^{k_i+1} - \boldsymbol{X}^{k_i+2}||_F \ldots + ||\boldsymbol{X}^{l_i-1} - \boldsymbol{X}^{l_i}||_F \\
\leq\ & \frac{1}{\epsilon_3}\left((f(\boldsymbol{X}^{k_i}) - f(\boldsymbol{X}^{k_i+1})) + (f(\boldsymbol{X}^{k_i+1}) - f(\boldsymbol{X}^{k_i+2})) + \cdots + (f(\boldsymbol{X}^{l_i-1}) - f(\boldsymbol{X}^{l_i}))\right) \\
=\ & \frac{f(\boldsymbol{X}^{k_i}) - f(\boldsymbol{X}^{l_i})}{\epsilon_3}.
\end{aligned}$$

Since the objective function $f(\boldsymbol{X}^k)$ is non-increasing and bounded below, this implies that $\lim_{i\to\infty}||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}||_F = 0$. Therefore, for $\epsilon_4 := \frac{\sqrt{n}\epsilon_2}{M_1+3M_2} > 0$, there exists i_0 such that $||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}||_F < \epsilon_4$ for $\forall i \geq i_0$.

Since $\boldsymbol{X}^{k_i} \in \mathcal{F}$, it holds that $-\boldsymbol{I} \preceq \boldsymbol{X} - \boldsymbol{X}^{k_i} \preceq \boldsymbol{I}$ for $\boldsymbol{X} \in \mathcal{F}$. Therefore, we have an inequality $||\boldsymbol{X} - \boldsymbol{X}^{k_i}||_F \leq \sqrt{n}$. For $\boldsymbol{X} \in \mathcal{F}$ and $i \geq i_0$, it holds that

$$\begin{aligned}
& \left|\langle\nabla f(\boldsymbol{X}^{k_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i}\rangle - \langle\nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{l_i}\rangle\right| \\
=\ & \left|\langle\nabla f(\boldsymbol{X}^{l_i} + (\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i})) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i}\rangle - \langle\nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{l_i}\rangle\right| \\
=\ & \left|\langle\nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i}\rangle + \int_0^1 \langle\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i} \mid \nabla^2 f(\boldsymbol{X}^{l_i} + \xi(\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i})) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i}\rangle d\xi\right. \\
& \left. -\langle\nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{l_i}\rangle\right| \\
=\ & \left|\int_0^1 \langle\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i} \mid \nabla^2 f(\boldsymbol{X}^{l_i} + \xi(\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i})) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i}\rangle d\xi - \langle\nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}\rangle\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_0^1 \left| \langle \boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i} \mid \nabla^2 f(\boldsymbol{X}^{l_i} + \xi(\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i})) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i} \rangle \right| d\xi \\
&\quad + \left| \langle \nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i} \rangle \right| \\
&\leq 3M_2 ||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}||_F ||\boldsymbol{X} - \boldsymbol{X}^{k_i}||_F + \sqrt{n} M_1 ||\boldsymbol{X}^{k_i} - \boldsymbol{X}^{l_i}||_F \\
&\leq 3M_2 \epsilon_4 \sqrt{n} + \sqrt{n} M_1 \epsilon_4 \\
&= \sqrt{n}(3M_2 + M_1)\epsilon_4 = n\epsilon_2.
\end{aligned}
$$

Here, we used Lemma 3.3 for the second inequality. Hence, we have

$$
\langle \nabla f(\boldsymbol{X}^{k_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{k_i} \rangle \geq \langle \nabla f(\boldsymbol{X}^{l_i}) \mid \boldsymbol{X} - \boldsymbol{X}^{l_i} \rangle - n\epsilon_2. \tag{3.4}
$$

If $\gamma_{\max}(\boldsymbol{X}^{k_i}) = 0$, then $\nabla f(\boldsymbol{X}^{k_i}) = \boldsymbol{O}$, and this results in $N(\boldsymbol{X}^{k_i}) = 0$ from (2.2). Therefore, from the assumption $N(\boldsymbol{X}^{k_i}) \geq \epsilon_2^2$ we know that $\gamma_{\max}(\boldsymbol{X}^{k_i}) > 0$. Since $\boldsymbol{X}^{k_i} - \frac{\boldsymbol{D}(\boldsymbol{X}^{k_i})}{\gamma_{\max}(\boldsymbol{X}^{k_i})} \in \mathcal{F}$ from Lemma 2.2, we can put $\boldsymbol{X} = \boldsymbol{X}^{k_i} - \frac{\boldsymbol{D}(\boldsymbol{X}^{k_i})}{\gamma_{\max}(\boldsymbol{X}^{k_i})}$ into (3.4) to get

$$
\begin{aligned}
\left\langle \nabla f(\boldsymbol{X}^{k_i}) \mid -\frac{\boldsymbol{D}(\boldsymbol{X}^{k_i})}{\gamma_{\max}(\boldsymbol{X}^{k_i})} \right\rangle &\geq \left\langle \nabla f(\boldsymbol{X}^{l_i}) \mid \left(\boldsymbol{X}^{k_i} - \frac{\boldsymbol{D}(\boldsymbol{X}^{k_i})}{\gamma_{\max}(\boldsymbol{X}^{k_i})} \right) - \boldsymbol{X}^{l_j} \right\rangle - n\epsilon_2 \\
&\geq \underline{f}(\boldsymbol{X}^{l_j}) - n\epsilon_2.
\end{aligned}
$$

With Lemma 3.4 and $N(\boldsymbol{X}^{l_i}) < \epsilon_2^2$, we have an upper bound on $N(\boldsymbol{X}^{k_i})$;

$$
\begin{aligned}
N(\boldsymbol{X}^{k_i}) &= \langle \nabla f(\boldsymbol{X}^{k_i}) \mid \boldsymbol{D}(\boldsymbol{X}^{k_i}) \rangle \leq \gamma_{\max}(\boldsymbol{X}^{k_i})(-\underline{f}(\boldsymbol{X}^{l_j}) + n\epsilon_2) \\
&\leq \gamma_{\max}(\boldsymbol{X}^{k_i})(n\sqrt{N(\boldsymbol{X}^{l_j})} + n\epsilon_2) \leq M_1(n\epsilon_2 + n\epsilon_2) = 2nM_1\epsilon_2.
\end{aligned}
$$

Therefore, we obtain the contradiction;

$$
\epsilon_1 \leq N(\boldsymbol{X}^{k_i}) \leq 2nM_1\epsilon_2 = \frac{1}{2}\epsilon_1 < \epsilon_1.
$$

Hence, $\lim_{k\to\infty} N(\boldsymbol{X}^k) = 0$. □

Combining Lemma 3.4 and Theorem 3.2, we derive the property for the first-order optimality condition.

Theorem 3.3. *When the sequence $\{\boldsymbol{X}^k\}$ generated by Algorithm 2.3 with $\epsilon = 0$ is an infinite sequence, it holds that*

$$
\lim_{k\to\infty} \underline{f}(\boldsymbol{X}^k) = 0.
$$

Proof. From Lemma 3.4, we know that $-n\sqrt{N(\boldsymbol{X}^k)} \leq \underline{f}(\boldsymbol{X}^k) \leq 0$. Hence, Theorem 3.2 indicates $\lim_{k\to\infty} \underline{f}(\boldsymbol{X}^k) = 0$. □

Using Theorem 3.3, we can show an additional result on the convergence. To make the generated sequence $\{\boldsymbol{X}^k\}$ itself converge, we need a stronger assumption on the objective function, for example, strong convexity.

Corollary 3.5. *If the objective function f is strongly convex, that is, there exists $\nu > 0$ such that*

$$f(\boldsymbol{Y}) \geq f(\boldsymbol{X}) + \langle \nabla f(\boldsymbol{X}) \mid \boldsymbol{Y} - \boldsymbol{X} \rangle + \frac{\nu}{2} ||\boldsymbol{Y} - \boldsymbol{X}||_F^2 \quad \textit{for} \quad \forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathcal{F},$$

and the sequence $\{\boldsymbol{X}^k\}$ generated by Algorithm 2.3 with $\epsilon = 0$ is an infinite sequence, then $\{\boldsymbol{X}^k\}$ converges. Furthermore, the accumulation point $\boldsymbol{X}^* := \lim_{k\to\infty} \boldsymbol{X}^k$ is an optimal solution.

Proof. From $\boldsymbol{X}^k \in \mathcal{F}$ and the definition of $\underline{f}(\boldsymbol{X}^j)$ for $\boldsymbol{X}^j \in \mathcal{F}$, we have an inequality $\underline{f}(\boldsymbol{X}^j) \leq \langle \nabla f(\boldsymbol{X}^j) \mid \boldsymbol{X}^k - \boldsymbol{X}^j \rangle$. By swapping $\boldsymbol{X}^k$ and $\boldsymbol{X}^j$, we also obtain the inequality $\underline{f}(\boldsymbol{X}^k) \leq \langle \nabla f(\boldsymbol{X}^k) \mid \boldsymbol{X}^j - \boldsymbol{X}^k \rangle$. The addition of these two inequalities results in

$$\langle \nabla f(\boldsymbol{X}^k) - \nabla f(\boldsymbol{X}^j) \mid \boldsymbol{X}^k - \boldsymbol{X}^j \rangle \leq -\underline{f}(\boldsymbol{X}^k) - \underline{f}(\boldsymbol{X}^j).$$

Theorem 2.1.9 of [17] gives equivalent conditions of strong convexity, and one of them is

$$\langle \nabla f(\boldsymbol{Y}) - \nabla f(\boldsymbol{X}) \mid \boldsymbol{Y} - \boldsymbol{X} \rangle \geq \nu ||\boldsymbol{Y} - \boldsymbol{X}||_F^2 \quad \forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathcal{F}.$$

Due to this inequality, we get

$$||\boldsymbol{X}^k - \boldsymbol{X}^j||_F \leq \frac{1}{\nu} \sqrt{-\underline{f}(\boldsymbol{X}^k) - \underline{f}(\boldsymbol{X}^j)}.$$

Theorem 3.3 implies that the sequence $\{\boldsymbol{X}^k\}$ is a Cauchy sequence. Since $\{\boldsymbol{X}^k\}$ is generated in the closed and bounded set $\mathcal{F}$, it converges to a point of $\mathcal{F}$. Hence, the accumulation point $\boldsymbol{X}^* = \lim_{k\to\infty} \boldsymbol{X}^k$ satisfies the first-order optimality condition. From the assumption that the objective function is strongly convex, $\boldsymbol{X}^*$ is an optimal solution. □

4 Numerical Results

To evaluate the performance of the proposed method, we conducted a numerical test. The computing environment was Debian Linux run on AMD Opteron Processor 4386 (3 GHz) and 128 GB of memory space, and we used Matlab R2014a.

The test functions used are listed below and they are classified into the two groups. The functions of Group I were selected from [25], and we added new functions as Group II. Function 5 and 6 are an extension of Generalized Rosenbrock function [16] and its variant with cosine functions, respectively.

Group I: Function 1. $f(\boldsymbol{X}) = -2\langle \boldsymbol{C}_1 \mid \boldsymbol{X} \rangle + \langle \boldsymbol{X} \mid \boldsymbol{X} \rangle$;

Function 2. $f(\boldsymbol{X}) = 3\cos(\langle \boldsymbol{X} \mid \boldsymbol{X} \rangle) + \sin(\langle \boldsymbol{X} + \boldsymbol{C}_1 \mid \boldsymbol{X} + \boldsymbol{C}_1 \rangle)$;

Function 3. $f(\boldsymbol{X}) = \log(\langle \boldsymbol{X} \mid \boldsymbol{X} \rangle + 1) + 5\langle \boldsymbol{C}_1 \mid \boldsymbol{X} \rangle$;

Group II: Function 4. $f(\boldsymbol{X}) = 1 + 2\frac{\langle \boldsymbol{X} - \boldsymbol{C}_1 \mid \boldsymbol{X} - \boldsymbol{C}_1 \rangle^3}{n^3}$;

Function 5. $f(\boldsymbol{X}) = 1 + \sum_{i=1}^{n} \sum_{j=i}^{n} (A_{ij} - X_{ij})^2$
$+100 \sum_{i=1}^{n-1} \sum_{j=i}^{n-1} \left(\frac{A_{ij}^2}{A_{i,j+1}} X_{i,j+1} - X_{ij}^2 \right)^2$
$+100 \sum_{i=1}^{n-1} \left(\frac{A_{in}^2}{A_{i+1,i+1}} X_{i+1,i+1} - X_{i,n}^2 \right)^2$;

$$\text{Function 6. } f(\boldsymbol{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1,j\neq i}^{n}\frac{X_{ij}}{A_{ij}} - (n-1)\frac{X_{ii}^2}{A_{ii}^2}\right)^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\cos((X_{ij}-A_{ij})^2);$$

$$\text{Function 7. } f(\boldsymbol{X}) = \langle \boldsymbol{C}_1 \mid \boldsymbol{X}\rangle - \log\det(\boldsymbol{X}+\bar{\epsilon}\boldsymbol{I}) - \log\det((1+\bar{\epsilon})\boldsymbol{I}-\boldsymbol{X});$$

To generate the matrix $\boldsymbol{C}_1$ in Functions 1, 4, and 7, we chose the eigenvalues $\kappa_1, \ldots, \kappa_n$ randomly from the interval $[-1, 2]$ and multiply a randomly-generated orthogonal matrix $\boldsymbol{Q}$, namely, $\boldsymbol{C}_1 := \boldsymbol{Q}diag(\kappa_1, \ldots, \kappa_n)\boldsymbol{Q}^T$. The elements A_{ij} in Functions 5 and 6 were set as $A_{ii} = \frac{1}{2}$ for $i = 1, \ldots, n$ and $A_{ij} = \frac{1}{2(n-1)}$ for $i \neq j$. The parameter $\bar{\epsilon}$ in Function 7 was set as $\bar{\epsilon} = 0.02$.

We compared the performance of three methods, PIM (the proposed iterative method, Algorithm 2.3), FEAS (the feasible direction method of Xu *et. al.* [25]), and PEN (the penalty barrier method [2, 13] implemented in PENLAB [7]). We started PIM and FEAS with the initial point $\boldsymbol{X}^0 = \frac{1}{2}\boldsymbol{I}$, while PEN automatically chose its initial point. For PIM, we used the parameters $\mu_1 = 0.25, \mu_2 = 0.75, \eta_1 = 0.5, \eta_2 = 2.0$, and $\Delta^0 = 1$. These parameters were chosen from preliminary experiments on Functions 1 and 4. We used the following condition as the stopping criterion;

PIM	$N(\boldsymbol{X}^k) < 10^{-7}$ or $\frac{\lvert f(\boldsymbol{X}^k) - f(\boldsymbol{X}^{k-1})\rvert}{\max\{\lvert f(\boldsymbol{X}^k)\rvert, 1\}} < 10^{-6}$
FEAS	$\lvert Trace(\boldsymbol{\Gamma}_-(\boldsymbol{X}^k)) - \langle f(\boldsymbol{X}^k) \mid \boldsymbol{X}^k\rangle\rvert < 10^{-6}$ or $\frac{\lvert f(\boldsymbol{X}^k) - f(\boldsymbol{X}^{k-1})\rvert}{\max\{\lvert f(\boldsymbol{X}^k)\rvert, 1\}} < 10^{-6}$
PEN	the default parameter of PENLAB.

For details of the stopping criterion on FEAS and PEN, refer to [25] and [7], respectively. We also stopped the computation when the computation time exceeded 24 hours.

Tables 1 and 2 show the numerical results of Group I and Group II, respectively. The first column is the function type, and the second column n is the size of the matrix $\boldsymbol{X}$. The third column indicates the method we applied, and the fourth column is the objective value. The fifth column is the number of main iterations, and the six column is the computation time in seconds. The last three columns correspond to the evaluation count of the function value $f(\boldsymbol{X})$, the gradient matrix $\nabla f(\boldsymbol{X})$, and the Hessian mapping $\nabla^2 f(\boldsymbol{X})$.

From these tables, PEN was much slow compared to PIM and FEAS. We did not include the results of PEN for large problems $n \geq 500$, since PEN did not finish the computation for $n = 500$ in 24 hours. Though it attained better solution for Function 5, PENLAB [7] handled the symmetric matrix $\boldsymbol{X}$ as $n(n+1)/2$ independent variables (X_{11}, X_{12}, ..., X_{1n}, X_{22}, ..., X_{2n}, ..., X_{nn}), and it stored all the elements of the Hessian mapping $\nabla^2 f(\boldsymbol{X})$, therefore, the computation cost was estimated as $\mathcal{O}(n^4)$ from [13]. This heavy cost restricted PENLAB to the small sizes. PIM also used the information of the Hessian mapping, but in only the scalar value $\langle \boldsymbol{S} \mid \nabla^2 f(\boldsymbol{X}) \mid \boldsymbol{S}\rangle$. Hence, the computation cost of each iteration in PIM is much lower than PEN, and this low cost is the key to handling large problems as noted at the end of Section 2.

In the comparison between PIM and FEAS, the computation time of FEAS was shorter than PIM in Table 1, but longer in Table 2. The functions in Group I involved the variable matrix $\boldsymbol{X}$ in the linear form $\langle \boldsymbol{C}_1 \mid \boldsymbol{X}\rangle$ or the quadratic form $\langle \boldsymbol{X} \mid \boldsymbol{X}\rangle$, and this simple structure was favorable for the feasible direction method, which was based on a steepest descent direction. In contrast, the functions in Group II have stronger nonlinearity than Group I. The evaluation count with respect to the function value (co.f) implies that this stronger nonlinearity demanded FEAS have a large number of back-step loop. In particular, FEAS needed many iterations for Rosenbrock-type functions (Functions 5 and 6). PIM reduced the number of iterations by the properties of the search direction $D(\boldsymbol{X})$ and the

Table 1: Numerical results on Group I.

type	n	method	obj	iter	cpu	co.f	co.∇f	co.$\nabla^2 f$
1	50	PIM	-3.631×10^1	48	0.08	95	48	48
1	50	FEAS	-3.633×10^1	36	0.04	215	36	0
1	50	PEN	-3.633×10^1	22	323.70	62	31	22
1	100	PIM	-7.930×10^1	67	0.30	133	67	67
1	100	FEAS	-7.932×10^1	36	0.11	239	36	0
1	100	PEN	-7.932×10^1	23	5554.30	64	32	23
1	500	PIM	-3.572×10^2	81	7.70	161	81	81
1	500	FEAS	-3.574×10^2	37	2.24	250	37	0
1	1000	PIM	-8.648×10^2	64	30.16	127	64	64
1	1000	FEAS	-8.651×10^2	32	9.62	204	32	0
1	5000	PIM	-3.861×10^3	80	3497.86	159	80	80
1	5000	FEAS	-3.862×10^3	36	1111.89	232	36	0
1	10000	PIM	-7.731×10^3	73	24730.04	145	73	73
1	10000	FEAS	-7.734×10^3	34	7782.18	213	34	0
2	50	PIM	-4.000	23	0.04	45	23	23
2	50	FEAS	-4.000	31	0.04	293	31	0
2	50	PEN	-4.000	115	1808.54	1857	124	116
2	100	PIM	-4.000	40	0.19	79	40	40
2	100	FEAS	-4.000	13	0.05	122	13	0
2	100	PEN	-3.985	15	4581.35	114	21	18
2	500	PIM	-4.000	26	2.40	51	26	26
2	500	FEAS	-4.000	17	1.31	183	17	0
2	1000	PIM	-4.000	17	6.32	33	17	17
2	1000	FEAS	-4.000	10	4.30	134	10	0
2	5000	PIM	-4.000	28	1205.24	55	28	28
2	5000	FEAS	-4.000	13	449.59	161	13	0
2	10000	PIM	-4.000	27	8461.01	53	27	27
2	10000	FEAS	-3.951	8	2009.73	73	8	0
3	50	PIM	-3.756×10^1	201	0.35	401	201	201
3	50	FEAS	-3.756×10^1	2	0.01	3	2	0
3	50	PEN	-3.756×10^1	28	418.41	76	36	28
3	100	PIM	-7.418×10^1	208	0.93	415	208	208
3	100	FEAS	-7.419×10^1	7	0.02	24	7	0
3	100	PEN	-7.419×10^1	30	7316.99	81	37	30
3	500	PIM	-3.625×10^2	257	25.62	513	257	257
3	500	FEAS	-3.625×10^2	2	0.13	3	2	0
3	1000	PIM	-7.739×10^2	269	128.23	537	269	269
3	1000	FEAS	-7.741×10^2	2	0.65	3	2	0
3	5000	PIM	-4.129×10^3	257	11996.45	513	257	257
3	5000	FEAS	-4.129×10^3	2	74.63	3	2	0
3	10000	PIM	-8.294×10^3	256	92901.29	511	256	256
3	10000	FEAS	-8.295×10^3	2	575.84	3	2	0

Table 2: Numerical results on Group II.

type	n	method	obj	iter	cpu	co.f	co.∇f	co.$\nabla^2 f$
4	50	PIM	1.041	26	0.07	51	26	26
4	50	FEAS	1.041	16	0.03	77	16	0
4	50	PEN	1.041	23	328.07	75	37	23
4	100	PIM	1.039	42	0.19	83	42	42
4	100	FEAS	1.039	24	0.07	138	24	0
4	100	PEN	1.039	25	5824.27	83	40	25
4	500	PIM	1.024	21	2.08	41	21	21
4	500	FEAS	1.024	23	1.35	123	23	0
4	1000	PIM	1.023	14	6.77	27	14	14
4	1000	FEAS	1.023	25	7.35	142	25	0
4	5000	PIM	1.024	12	517.62	23	12	12
4	5000	FEAS	1.024	25	715.68	134	25	0
4	10000	PIM	1.025	12	4140.26	23	12	12
4	10000	FEAS	1.025	21	4866.45	109	21	0
5	50	PIM	1.122	4	0.01	7	4	4
5	50	FEAS	1.126	19	0.05	252	19	0
5	50	PEN	1.000	20	294.58	61	30	20
5	100	PIM	1.117	6	0.06	11	6	6
5	100	FEAS	1.125	16	0.15	226	16	0
5	100	PEN	1.000	20	4814.74	61	30	20
5	500	PIM	1.004	4	0.82	7	4	4
5	500	FEAS	1.125	16	4.28	286	16	0
5	1000	PIM	1.008	4	4.02	7	4	4
5	1000	FEAS	1.125	18	26.96	352	18	0
5	5000	PIM	1.002	4	192.25	7	4	4
5	5000	FEAS	1.125	90	6345.17	2279	90	0
5	10000	PIM	1.013	4	1332.14	7	4	4
5	10000	FEAS	1.124	122	51611.04	3285	122	0
6	50	PIM	-1.000	20	0.11	39	20	20
6	50	FEAS	-1.000	12	0.10	92	12	0
6	50	PEN	-1.000	300	4577.01	915	1218	300
6	100	PIM	-1.000	20	0.36	39	20	20
6	100	FEAS	-1.000	16	0.56	150	16	0
6	100	PEN	-9.997×10^{-1}	300	73262.02	1005	1308	300
6	500	PIM	-1.000	18	10.00	35	18	18
6	500	FEAS	-1.000	12	9.36	110	12	0
6	1000	PIM	-1.000	4	9.42	7	4	4
6	1000	FEAS	-1.000	12	56.33	110	12	0
6	5000	PIM	-1.000	4	406.01	7	4	4
6	5000	FEAS	-1.000	13	2046.55	130	13	0
6	10000	PIM	-1.000	3	2076.17	5	3	3
6	10000	FEAS	-1.000	14	10416.77	130	14	0
7	50	PIM	7.817×10^{1}	10	0.03	19	10	10
7	50	FEAS	7.817×10^{1}	15	0.06	108	15	0
7	50	PEN	7.817×10^{1}	13	195.41	38	19	13
7	100	PIM	1.583×10^{2}	10	0.11	19	10	10
7	100	FEAS	1.583×10^{2}	17	0.30	13	17	0
7	100	PEN	1.583×10^{2}	14	3427.86	40	20	14
7	500	PIM	7.825×10^{2}	10	2.73	19	10	10
7	500	FEAS	7.825×10^{2}	16	6.22	116	16	0
7	1000	PIM	1.556×10^{3}	10	12.02	19	10	10
7	1000	FEAS	1.556×10^{3}	10	15.79	60	10	0
7	5000	PIM	7.707×10^{3}	11	1708.11	21	11	11
7	5000	FEAS	7.707×10^{3}	16	4931.70	115	16	0
7	10000	PIM	1.533×10^{4}	11	13379.96	21	11	11
7	10000	FEAS	1.533×10^{4}	14	32643.49	94	14	0

quadratic approximation with the Hessian mapping. In particular, $\boldsymbol{D}(\boldsymbol{X})$ encompassed the information of the distance to the boundary to the box-constraints as $\boldsymbol{V}_+(\boldsymbol{X})$ and $\boldsymbol{V}_-(\boldsymbol{X})$. Therefore, PIM was faster than FEAS for the functions of Group II.

5 Conclusions and Future Directions

In this paper, we proposed an iterative method for box-constrained SDPs. The search direction $\boldsymbol{D}(\boldsymbol{X})$ studied in Section 2 enabled us to include the information of the distance from the current point to the boundary of the feasible set $\mathcal{F}$. We discussed the convergence property of the generated sequence. The numerical tests in Section 4 showed that the proposed method was more favorable for functions with strong nonlinearity than the feasible direction method, mainly due to the distance information included in $\boldsymbol{D}(\boldsymbol{X})$. In addition, the proposed method handled the larger problems than the penalty barrier method, since our method did not hold the Hessian mapping in memory space.

One of future researches would be the combination of the feasible direction and the proposed method, since the feasible direction method fits simple functions. For such a combination, we should extend the convergence analysis from this paper. Another point is the convergence for a second-order optimality condition, as proven in [4] for box-constrained problem (1.2). The proof in [4] required further stronger assumptions than this paper and the second-order optimality condition for nonlinear semidefinite programs involves not only the Hessian mapping but also an additional mapping [19], so we remain it as a matter to be discussed further. In this paper, we fixed the parameters $\mu_1, \mu_2, \eta_1, \eta_2$ in the convergence analysis and the numerical tests. By adjusting these parameters along with the iteration progress, there is a possibility that we can improve the numerical performance. Such adjustments, however, will need careful update of the radius Δ_k, and a further investigation should be conducted.

References

[1] M.F. Anjos and J.B. Lasserre (eds.), *Handbook on Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*, Springer, NY, USA, 2012.

[2] A. Ben-tal and M. Zibulevsky, Penalty/barrier multiplier methods for convex programming problems, *SIAM J. Optim.* 7 (1997) 347–366.

[3] S. Boyd, L. El Ghaoui, E. Feron and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, PA, USA, 2004.

[4] T.F. Coleman and Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optim.* 6 (1996) 418–445.

[5] A.R. Conn, N.I.M. Gould and P.L. Toint, Global convergence of a class of trust region algorithms for optimization with simple bounds, *SIAM J. Numeric. Anal.* 25 (1988) 433–460.

[6] A.R. Conn, N.I.M. Gould and P.L. Toint, *Trust-Region Methods*, SIAM, PA, USA, 2000.

[7] J. Fiala, M. Ko***v***cvara and M. Stingl, PENLAB - a solver for nonlinear semidefinite programming, http://web.mat.bham.ac.uk/kocvara/penlab/, October 2013.

[8] M. Fukuda, B.J. Braams, M. Nakata, M.L. Overton, J.K. Percus, M. Yamashita and Z. Zhao, Large-scale semidefinite programs in electronic structure calculation, *Math. Program. Series B* 109 (2007) 553–580.

[9] M.X. Goemans and D.P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comput. Mach.* 42 (1995) 1115–1145.

[10] I. Griva, S.G. Nash and A. Sofer, *Linear and Nonlinear Optimization*, SIAM, PA, USA, 2009.

[11] L. Hei, J. Nocedal and R.A. Waltz, A numerical study of active-set and interior-point methods for bound constrained optimization, in *Modeling, Simulation and Optimization of Complex Processes*, H.G. Bock, E. Kostina, H.X. Phu, and R. Rannacher (eds.), Springer, Berlin, Germany, 2008, pp. 273–292.

[12] Y. Ji, T. Wang and Y. Li, A continuous trust-region-type method for solving nonlinear semidefinite complementarity problem, *Abstr. Appl. Anal.* 2014 (2014) ArticleID 589731.

[13] M. Kočvara and M. Stingl, PENNON: A code for convex nonlinear and semidefinite programming, *Optim. Methods Softw.* 18 (2003) 317–333.

[14] J. B. Lasserre, Global optimization with polynomials and the problems of moments, *SIAM J. Optim.* 11 (2001) 796–817.

[15] F. Leibfritz and E. Mostafa, An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems, *SIAM J. Optim.* 12 (2002) 1048–1074.

[16] S. Nash, Newton-type minimization via the lanczos method, *SIAM J. Numer. Anal.* 21 (1984) 770–788.

[17] Y.E. Nesterov, *Introductory Lectures on Convex Optimization: A Basice Course*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[18] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, NY, USA, 2006.

[19] A. Shapiro, First and second order analysis of nonlinear semidefinite programs, *Math. Program.* 77 (1997) 301–320.

[20] W. Sun and Y.X. Yuan, *Optimization Theory and Methods*, Springer, NY, USA, 2006.

[21] M. Tang, Trust region algorithm with two subproblems for bound constrained problems, *Appl. Math. and Comput.* 242 (2014) 778–789.

[22] M.J. Todd. Semidefinite optimization. *Acta Numerica* 10 (2001) 515–560.

[23] K.C. Toh, M.J. Todd and R.H. Tütüncü, SDPT3 – a MATLAB software package for semidefinite programming, version 1.3, *Optim. Methods Softw.* 11 & 12 (1999) 545–581.

[24] X. Wang and Y.X. Yuan, A trust region method based on a new affine scaling technique for simple bounded optimization, *Optim. Methods and Softw.* 28 (2013) 871–888.

[25] Y. Xu, W. Sun and L. Qi, A feasible direction method for the semidefinite program with box constraints, *Appl. Math. Lett.* 24 (2011) 1874–1881.

[26] M. Yamashita, K. Fujisawa, M. Fukuda, K. Nakata and M. Nakata, Algorithm 925: Parallel solver for semidefinite programming problem having sparse Schur complement matrix, *ACM Trans. Math. Softw.* 39 (2012) Article No.6.

[27] Q. Zhou, F. Zhou and F. Cao, A nonmonotone trust region method based on simple conic models for unconstrained optimization, *Appl. Math. Comput.* 225 (2013) 295–305.

Manuscript received 14 May 2015
revised 8 January 2016
accepted for publication 25 November 2016

Akihiko Komatsu
IT Development, Tokyo Stock Exchange, INC.
E-mail address: a-komatsu@jpx.co.jp

Makoto Yamashita
Department of Mathematical and Computing Science
Tokyo Institute of Technology
E-mail address: Makoto.Yamashita@is.titech.ac.jp