An equivalent nonlinear optimization model with triangular low-rank factorization for semidefinite programs

Yuya Yamakawa\textsuperscript{a}, Tetsuya Ikegami\textsuperscript{b}, Ellen H. Fukuda\textsuperscript{a} and Nobuo Yamashita\textsuperscript{a}

\textsuperscript{a}Graduate School of Informatics, Kyoto University, Kyoto, Japan; \textsuperscript{b}Data and Science Solutions Group, Technology Group, Yahoo Japan Corporation, Chiyoda-ku, Tokyo, Japan

ABSTRACT

In this paper, we propose a new nonlinear optimization model to solve semidefinite optimization problems (SDPs), providing some properties related to local optimal solutions. The proposed model is based on another nonlinear optimization model given by [S. Burer and R. Monteiro, \textit{A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization}, Math. Program. Ser. B 95 (2003), pp. 329–357], but it has several nice properties not seen in the existing one. Firstly, the decision variable of the proposed model is a triangular low-rank matrix. Secondly, the existence of a strict local optimum of the proposed model is guaranteed under some conditions, whereas the existing model has no strict local optimum. In other words, it is difficult to construct solution methods equipped with fast convergence using the existing model. We also present some numerical results, showing that the use of the proposed model allows to deliver highly accurate solutions.

ARTICLE HISTORY

Received 24 December 2021
Accepted 3 June 2023

KEYWORDS

Semidefinite optimization problems; nonlinear optimization problems; sequential quadratic programming method; triangular low-rank factorization

AMS CLASSIFICATIONS

90C22; 90C30; 90C55

1. Introduction

In this paper, we consider the following \textit{semidefinite optimization problem} (SDP):

\begin{equation}
\begin{aligned}
\text{(SDP)} \quad & \text{Minimize} & \langle C, X \rangle \\
& \text{subject to} & A(X) = b, \ X \succeq O,
\end{aligned}
\end{equation}

where $\mathbb{S}^n$ denotes the set of $n \times n$ real symmetric matrices, and the operator $A: \mathbb{S}^n \rightarrow \mathbb{R}^m$ is defined by $A(X) := [\langle A_1, X \rangle \ldots \langle A_m, X \rangle]^\top$, and the matrices $C, A_1, \ldots, A_m \in \mathbb{S}^n$ and the vector $b \in \mathbb{R}^m$ are given. Throughout this paper, we assume that the matrices $A_1, \ldots, A_m$ are linearly independent. For two matrices $P$ and $Q$ included in $\mathbb{R}^{p \times q}$, the inner product of them is defined by $\langle P, Q \rangle := \text{tr}(P^\top Q)$, where $\text{tr}(M)$ represents the trace of a square matrix $M$, and the superscript $\top$ indicates the transposition of a matrix or a vector. Let $S^+_n (\mathbb{S}^+)$ be the set of $n \times n$ real symmetric positive (semi)definite matrices. For a matrix $M \in \mathbb{S}^n$, $M \succeq O$ and $M \succ O$ mean that $M \in \mathbb{S}^+_n$ and $M \in S^+_n$, respectively.
SDPs include some classes of optimization problems, such as linear programs, quadratic programs, and second-order cone programs. Moreover, they have a wide range of application fields, such as control theory, graph theory, structural optimization, combinatorial optimization, and so forth [16,18,20]. Until now, a lot of solution methods for SDPs have been proposed by many researchers [11,13,15,16,19,22,23]. Specifically, the primal-dual interior-point method is known as one of the most popular ones, and there exist several efficient software packages in which they are implemented, such as SeDuMi [14], SDPT3 [17], and SDPA [21]. They are based on the Newton method and can solve small- or medium-scale problems very accurately. However, it may not be applicable to problems whose scale is too large. Moreover, it is well known that the Newton equation used in the primal-dual interior-point method becomes unstable in the neighbourhoods of solutions, and hence it is difficult to obtain solutions with high accuracy.

To solve large-scale SDPs whose \( m \) (the number of equality constraints) is small and \( n \) (the matrix size) is large, Burer and Monteiro have proposed an equivalent nonlinear optimization model [6]. In their proposal, the decision variable \( X \) is replaced with a matrix product \( RR^\top \), where \( R \in \mathbb{R}^{n \times r} \). Since \( RR^\top \) is positive semidefinite, the semidefinite constraint \( X \succeq O \) can be ignored, and hence the model is expressed as a usual nonlinear optimization problem. It is known that if \( r \geq \max\{r \in \mathbb{N} : \frac{r(r+1)}{2} \leq m\} \), then their model is equivalent to (SDP). Note that depending on the value \( r \), the dimension of such model’s decision variable space can be smaller than \( \frac{n(n+1)}{2} \), which is the dimension of the original (SDP). However, the model has no strict local minimum. Therefore, the second-order sufficient conditions do not hold, and it is difficult to construct solution methods equipped with fast convergence. Moreover, when \( m \) is not small enough, the dimension \( nr \) is larger than \( \frac{n(n+1)}{2} \).

In this paper, we present a new nonlinear optimization model, which overcomes the above drawbacks of Burer and Monteiro’s model, and show that the proposed model has several nice properties associated with local optima. A remarkable point of the proposed model is that its decision variable space is the set of \( n \times r \) real matrices whose upper triangular part is all zero, that is, its dimension is equal to \( nr - \frac{r(r-1)}{2} \). It is slightly smaller than \( nr \), which is the dimension of the variables of the existing model. Moreover, the dimension \( nr - \frac{r(r-1)}{2} \) in the proposed model is at most \( \frac{n(n+1)}{2} \), that is to say, it never exceeds the dimension of the original (SDP). We also show that the existence of a strict local optimum is guaranteed under some appropriate conditions. From this fact, it is expected that second-order methods, such as sequential quadratic programming (SQP) and interior-point methods, have fast convergence to solutions.

This paper is organized as follows. In Section 2, we introduce some important concepts regarding (SDP) and the existing model proposed by Burer and Monteiro. In Section 3, we propose a new nonlinear optimization model and give some properties associated with its local optima. In Section 4, we provide a reformulation from the proposed model into some quadratic program. Section 5 reports some numerical experiments to confirm the high accuracy of the solutions obtained with the proposed model. Finally, we make some concluding remarks in Section 6.

Throughout this paper, we use the following notation. The identity matrix and the all-ones vector are represented by \( I \) and \( e \), respectively, where their dimensions are defined by each context. For a vector \( w \in \mathbb{R}^p \), \([w]_i\) denotes the \( i \)th element of \( w \), and \( \|w\| \) is the Euclidean norm of \( w \) defined by \( \|w\| := \sqrt{\langle w, w \rangle} = \sqrt{w^\top w} \). Let \( W \in \mathbb{R}^{p \times q} \). We express
the \((i,j)\)-entry of \(W\) by \([W]_{ij}\). Moreover, we write \(\|W\|_F\) and \(\|W\|_2\) for the Frobenius norm and the operator norm of \(W\), respectively, that is, \(\|W\|_F := \sqrt{\langle W, W \rangle} (= \sqrt{\text{tr}(W^T W)})\) and \(\|W\|_2 := \sup\{\|Wx\| : \|x\| = 1\}\), where \(\text{tr}(M)\) denotes the trace of a square matrix \(M\). For real numbers \(r_1, \ldots, r_d \in \mathbb{R}\) and a vector \(v \in \mathbb{R}^d\), we use the notation below:

\[
\text{diag}(r_1, \ldots, r_d) := \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_d \end{bmatrix}, \quad \text{diag}(v) := \begin{bmatrix} [v]_1 & 0 & \cdots & 0 \\ 0 & [v]_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & [v]_d \end{bmatrix}.
\]

Let \(U \in \mathbb{S}^d\) be a matrix. The minimum and the maximum eigenvalues of \(U\) are denoted by \(\lambda_{\text{min}}(U)\) and \(\lambda_{\text{max}}(U)\), respectively. Let \(\Phi\) be a mapping from \(P_1 \times P_2\) to \(P_3\), where \(P_1\) and \(P_2\) are open sets. We express the Fréchet derivative of \(\Phi\) as \(\nabla \Phi\). Moreover, we denote the Fréchet derivative of \(\Phi\) with respect to a variable \(Z \in P_1\) as \(\nabla_Z \Phi\). For a positive integer \(k \in \mathbb{N}\), we define

\[
r_k := \max \left\{ r \in \mathbb{N} : \frac{r(r+1)}{2} \leq k \right\}.
\]

### 2. Preliminaries

We give some important concepts related to (SDP). Next, we introduce the existing nonlinear model for (SDP) and provide some of its properties.

#### 2.1. Basic facts related to SDP

As it is well known, the dual of (SDP) can be written as

\[
\text{(DSDP)} \quad \begin{array}{ll}
\text{maximize} & \langle b, y \rangle \\
\text{subject to} & A^*(y) + Z = C, Z \succeq O,
\end{array}
\]

where \(A^* : \mathbb{R}^m \to \mathbb{S}^n\) is the adjoint operator of \(A\), which is defined by \(A^*(v) := \sum_{j=1}^m [v]_j A_j\) for all \(v \in \mathbb{R}^m\). Throughout this paper, we assume the existence of \((X^*, y^*, Z^*) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n\) such that

\[
A(X^*) = b, A^*(y^*) + Z^* = C, \langle X^*, Z^* \rangle = 0, X^* \succeq O, Z^* \succeq O,
\]

which are called Karush-Kuhn-Tacker (KKT) conditions of (SDP). These conditions are necessary and sufficient for optimality. The following result shows the existence of a solution with a particular limited rank.

**Theorem 2.1:** There exists an optimal solution \(X^* \in \mathbb{S}^n\) of (SDP) such that \(\text{rank}(X^*) \leq r_m\).

**Proof:** It follows from definition (1) and [6, Theorem 1] (see also [1,12]).

Finally, we recall the definitions of face of (SDP) and minimal face of (SDP). Let \(C\) be a convex set of a real vector space \(\mathcal{X}\) and let \(\mathcal{F}\) be a nonempty convex subset of \(C\). The set \(\mathcal{F}\) is called a face of \(C\) if for every \(\eta \in \mathcal{F}\) and every \(\xi, \zeta \in C\) such that \(\eta = \theta \xi + (1-\theta) \zeta\)
for some θ ∈ (0, 1), we have ξ, ζ ∈ F. Let λ ∈ C. The minimal face of C containing λ is defined as ∩{F : FCλ ∈ F}. If C is the feasible set of (SDP), then a face of C is called a face of (SDP) and the minimal face of C containing λ is referred to as the minimal face of (SDP) containing λ.

2.2. An existing nonlinear optimization model for SDP

In [6], Burer and Monteiro proposed the following low-rank SDP, which has a rank constraint on the decision variable X ∈ Sn:

\[
\text{Minimize}_{X ∈ S^n} \quad \langle C, X \rangle \\
\text{subject to} \quad A(X) = b, \quad X ⪰ O, \quad \text{rank}(X) ≤ r.
\]

Theorem 2.1 ensures that if \(r ≥ rm\), then \((LRSDP_r)\) is equivalent to (SDP). Since an arbitrary semidefinite matrix \(X ∈ S^n\) can be rewritten as \(X = RR^T\) for some \(R ∈ \mathbb{R}^{n×r}\), \((LRSDP_r)\) can be reformulated as follows:

\[
\text{(NSDP}_r\text{)} \quad \text{Minimize}_{R ∈ \mathbb{R}^{n×r}} \quad \langle C, RR^T \rangle \\
\text{subject to} \quad A(RR^T) = b.
\]

For \((NSDP_r)\), the Lagrange function is defined as

\[
L(R,v) := \langle C, RR^T \rangle − \langle v, A(RR^T) − b \rangle = \langle C − A^*(v), RR^T \rangle + \langle b, v \rangle,
\]

where \(v ∈ \mathbb{R}^m\) is the Lagrange multiplier. We say that a feasible point \(R^*\) is stationary of \((NSDP_r)\) if there exists \(v^* ∈ \mathbb{R}^m\) such that

\[
\nabla_R L(R^*, v^*) = 2(C − A^*(v^*))R^* = 0.
\]

Problem \((NSDP_r)\) has several remarkable properties. In particular, we do not have to deal directly with the semidefinite constraint, and if \(r ∈ \mathbb{N}\) is small, then the number of variables decreases considerably compared with (SDP). However, since the purpose of this paper is to obtain a solution of the original (SDP), we need to clarify the relation between the global (or local) optimal solutions of (SDP) and \((NSDP_r)\). In fact, some of these relations can be seen in the following. Note that the first proposition below is obtained by combining Proposition 2.3 and Theorem 3.4 in [7].

**Proposition 2.2 ([7, Proposition 2.3 and Theorem 3.4]):** Suppose that \(R^*\) is a local minimum of \((NSDP_r)\) with \(r ≥ rm+1\). Suppose also that \(F\) is the minimal face of (SDP) containing \(R^*(R^*)^T\). If the dimension of \(F\) is zero, then \(R^*(R^*)^T\) is an optimal extreme point of (SDP).

**Proposition 2.3 ([6, Proposition 3]):** Suppose that \(R^*\) is a stationary point of \((NSDP_r)\), i.e. there exists \(v^* ∈ \mathbb{R}^m\) such that \(\nabla_R L(R^*, v^*) = 0\). If \(C − A^*(v^*)\) is positive semidefinite, then \(R^*(R^*)^T\) and \((C − A^*(v^*), v^*)\) are optimal solutions for (SDP) and (DSDP), respectively.

**Remark 2.1:** Let \(R\) be a local optimum of \((NSDP_r)\). Note that \(RQ(RQ)^T = RR^T\) for an arbitrary orthogonal matrix \(Q\). Then, we easily see that \(RQ\) is also another local optimum.
of (NSDP_r). Moreover, we can select the matrix Q so that RQ arbitrarily approaches to R, and hence (NSDP_r) has no strict local optimum [6]. This fact shows that it is difficult to construct fast convergent methods which solve (NSDP_r).

**Remark 2.2:** Boumal et al. [4,5] investigated conditions under which the first- and second-order necessary optimality conditions for (NSDP_r) are sufficient for its global optimality. Hence, under those conditions, if R ∈ R^{n×r} is a global optimum of (NSDP_r), then RR^T ∈ S^n is a global optimum of (SDP). In the existing papers, they dealt with (NSDP_r) as Riemannian optimization and provided general theorems regarding the sufficient conditions. Moreover, they showed that some applications, such as max-cut and generalized eigenvalue problems, satisfy the assumptions of the proposed theorems.

Not only Boumal et al. [4,5] but also Grubišić et al. [9] proposed a Riemannian optimization approach for a special case of SDP. In particular, [9] focuses on solving the nearest low-rank correlation matrix problem by exploiting the Riemannian optimization technique regarding the Cholesky manifold.

### 3. A new nonlinear optimization model for SDP

Firstly, we propose a new nonlinear optimization model for (SDP). Secondly, we provide relations between solutions of the proposed model and (SDP), and some important properties related to local optima.

To begin with, we denote by L^{n×r} the set of lower triangular matrices in R^{n×r}, i.e.

\[ L^{n×r} := \{ S ∈ R^{n×r} : [S]_{ij} = 0 \text{ if } i < j \}. \]

The result below, which shows that a symmetric positive semidefinite matrix in S^n with rank smaller than n has at least one triangular low-rank factorization, is well known. However, for the sake of completeness, we also show a proof for it.

**Proposition 3.1:** For any symmetric positive semidefinite matrix X ∈ S^n satisfying rank(X) = r < n, there exists a matrix S ∈ L^{n×r} such that X = SS^T.

**Proof:** Since X is a symmetric positive semidefinite matrix satisfying rank(X) = r < n, there exists a matrix U ∈ R^{n×r} such that X = UU^T. Let us write the matrix U as follows:

\[ U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, \quad U_1 ∈ R^{r×r}, \quad U_2 ∈ R^{(n-r)×r}. \]

From the QR factorization of the matrix U_1^T, there exist an orthogonal matrix Q ∈ R^{r×r} and an upper triangular matrix R ∈ R^{r×r} such that U_1^T = QR. Now, let S be defined by

\[ S := \begin{bmatrix} R^T \\ U_2Q \end{bmatrix}. \]
Note that $S$ is included in $\mathbb{L}^{n \times r}$ because $R^T$ is lower triangular. Hence, we have

$$X = \begin{bmatrix} R^T Q^T \\ U_2 \end{bmatrix} \begin{bmatrix} QR & U_2^T \end{bmatrix} = \begin{bmatrix} R^T R & R^T Q U_2^T \\ U_2 Q & U_2^T U_2^T \end{bmatrix} = \begin{bmatrix} R^T \\ U_2 Q \end{bmatrix} \begin{bmatrix} R & Q U_2^T \end{bmatrix} = SS^T.$$

Therefore, the assertion is proven. ■

The above proposition guarantees that (LRSDP$_r$) can be rewritten as the following new nonlinear optimization model:

$$(T-\text{NSDP}_r) \quad \begin{array}{l}
\text{Minimize} \\
\text{subject to}
\end{array} \quad \langle C, SS^T \rangle
\quad \text{for } S \in \mathbb{L}^{n \times r}
\quad \mathcal{A}(SS^T) = b.$$

The Lagrange function for (T-\text{NSDP}_r) is defined by

$$\mathcal{L}(S,w) := \langle C, SS^T \rangle - \langle w, \mathcal{A}(SS^T) - b \rangle = \langle C - \mathcal{A}^*(w), SS^T \rangle + \langle b, w \rangle,$$

where $w \in \mathbb{R}^m$ is the Lagrange multiplier.

Similarly to the existing model (NSDP$_r$), we now verify how the local (global) optimal solutions of (SDP) and (T-\text{NSDP}_r) are related to each other. The following propositions can be proven in the same way as [6,7].

**Proposition 3.2 ([7, Proposition 2.3 and Theorem 3.4]):** Suppose that $S^*$ is a local minimum of (T-\text{NSDP}_r) with $r \geq r_{m+1}$. Suppose also that $F$ is the minimal face of (SDP) containing $S^*(S^*)^T$. If the dimension of $F$ is zero, then $S^*(S^*)^T$ is an optimal extreme point of (SDP).

**Proposition 3.3 ([6, Proposition 3]):** Suppose that $S^*$ is a stationary point of (T-\text{NSDP}_r), i.e. there exists $w^* \in \mathbb{R}^m$ such that $\nabla_S \mathcal{L}(S^*,w^*) = 0$. If $C - \mathcal{A}^*(w^*)$ is positive semidefinite, then $S^*(S^*)^T$ and $(C - \mathcal{A}^*(w^*), w^*)$ are optimal solutions for (SDP) and (DSDP), respectively.

The proposed model (T-\text{NSDP}_r) has several advantages over the existing model (NSDP$_r$). One of them is that the number of variables in (T-\text{NSDP}_r) (more precisely, $nr - \frac{r^2 - 1}{2}$) can be smaller than (NSDP$_r$)'s (i.e. $nr$). The reduction of variables also brings another nice property. As stated in Remark 2.1, (NSDP$_r$) has no strict local optimum. On the other hand, (T-\text{NSDP}_r) has a strict local optimum under some appropriate conditions because it decreases the degree of freedom of variables compared with (NSDP$_r$). To show this fact, we recall the following well-known result.
Lemma 3.5: Suppose that \( P \in \mathbb{L}^{\ell \times \ell} \) is an arbitrary full rank matrix. Then, there exists \( \delta > 0 \) such that if \( Q \in \mathbb{L}^{\ell \times \ell} \) satisfies \( PP^T = QQ^T \), then \( P = Q \) or \( \|P - Q\|_F \geq \delta \).

Proof: We show the assertion by contradiction. Let \( F := PP^T \). Because \( \text{rank}(P) = \ell \), it is clear that \( F > 0 \), that is, \( \lambda_j(F) > 0 \) for all \( j \in \{1, \ldots, \ell\} \). Hence, we define \( \delta := \sqrt{\lambda_{\min}(F)} > 0 \). Since the assertion is not true, there exists \( Q_\delta \in \mathbb{L}^{\ell \times \ell} \) such that
\[
F = Q_\delta Q_\delta^T, \quad P \neq Q_\delta, \quad \|P - Q_\delta\|_F < \delta. \tag{3}
\]
Note that \( [P]_{jj} \neq 0 \) and \( [Q_\delta]_{jj} \neq 0 \) for all \( j \in \{1, \ldots, \ell\} \) because \( P \in \mathbb{L}^{\ell \times \ell} \), \( Q_\delta \in \mathbb{L}^{\ell \times \ell} \), and \( \ell = \text{rank}(F) = \text{rank}(P) = \text{rank}(Q_\delta) \). Now, we define diagonal matrices \( G \in \mathbb{R}^{\ell \times \ell} \) and \( H_\delta \in \mathbb{R}^{\ell \times \ell} \) such that
\[
[G]_{jj} := \begin{cases} 1 & \text{if } [P]_{jj} > 0, \\ -1 & \text{if } [P]_{jj} < 0, \end{cases} \quad [H_\delta]_{jj} := \begin{cases} 1 & \text{if } [Q_\delta]_{jj} > 0, \\ -1 & \text{if } [Q_\delta]_{jj} < 0 \end{cases} \forall j \in \{1, \ldots, \ell\}. \tag{4}
\]
Then, there exists \( \hat{P} \in \mathbb{L}^{\ell \times \ell} \) such that
\[
P = \hat{P}G, \quad [\hat{P}]_{jj} > 0 \forall j \in \{1, \ldots, \ell\}. \tag{5}
\]
Similarly, there exists \( \hat{Q}_\delta \in \mathbb{L}^{\ell \times \ell} \) such that \( Q_\delta = \hat{Q}_\delta H_\delta \) and \( [\hat{Q}_\delta]_{jj} > 0 \) for all \( j \in \{1, \ldots, \ell\} \). It follows from these results that \( F = \hat{P}\hat{P}^T = \hat{Q}_\delta \hat{Q}_\delta^T \). However, Proposition 3.4 ensures that \( \hat{P} = \hat{Q}_\delta \). As a result, we obtain
\[
P = \hat{P}G, \quad Q_\delta = \hat{P}H_\delta. \tag{6}
\]
Combining (3) and (6) yields
\[
\lambda_{\min}(F) = \delta^2 > \|P - Q_\delta\|_F^2 \\
= \|\hat{P}(G - H_\delta)\|_F^2 \\
= \text{tr}(\hat{P}^T(\hat{P}(G - H_\delta))^2) \\
\geq \lambda_{\min}(F)\text{tr}((G - H_\delta)^2), \tag{7}
\]
where the third equality follows from the fact that \( \text{tr}(AB) = \text{tr}(BA) \) for any matrices \( A \) and \( B \), and the last inequality is true because \( \text{tr}(AB) \geq \lambda_{\min}(A)\text{tr}(B) \) for \( A \in \mathbb{S}_+^\ell \) and \( B \in \mathbb{S}_+^\ell \) [2, Theorem 8.4.13]. Moreover, since \( \hat{P} \in \mathbb{L}^{\ell \times \ell} \) is nonsingular from (5), the results (3) and (6) mean that \( O \neq P - Q_\delta = \hat{P}(G - H_\delta) \), i.e. \( G \neq H_\delta \). Exploiting (4) and (7) implies \( \lambda_{\min}(F) > \lambda_{\min}(F)\text{tr}((G - H_\delta)^2) \geq 4\lambda_{\min}(F) \), that is, \( 0 \geq \lambda_{\min}(F) \). Therefore, this contradicts \( F > 0 \). \hfill \blacksquare

From now on, we provide several sufficient conditions under which \( (T-\text{NSDP}_r) \) has a strict local optimum.
Theorem 3.6: Assume that (SDP) has a unique optimal solution $X^*$ satisfying \( \text{rank}(X^*) = r \). Then, (T–NSDP$_r$) has a strict local optimum.

Proof: The solution $X^*$ has a spectral decomposition $X^* = VDV^\top$ with

\[
D = \begin{bmatrix}
D_1 & 0 \\
0 & O
\end{bmatrix}, \quad D_1 \in S^r, \quad \text{rank}(D_1) = r,
\]

where $D_1$ is a diagonal matrix whose diagonal elements are positive eigenvalues of $X^*$. Let $U^* \in \mathbb{R}^{n \times r}$ be defined by

\[
U^* := \begin{bmatrix}
U_1^* \\
U_2^*
\end{bmatrix} := V \begin{bmatrix}
D_1^{\frac{1}{2}} \\
0
\end{bmatrix}.
\]

Note that $\text{rank}(U^*_1) = r$, $U_1^* \in \mathbb{R}^{r \times r}$, and $U_2^* \in \mathbb{R}^{(n-r) \times r}$. The QR factorization of $(U^*_1)^\top$ ensures the existence of an orthogonal matrix $Q \in \mathbb{R}^{r \times r}$ and an upper triangular matrix $R \in \mathbb{R}^{r \times r}$ such that $(U^*_1)^\top = QR$ and $\text{rank}(R) = r$. We define $S^* \in \mathbb{L}^{n \times r}$ as

\[
S^* := \begin{bmatrix}
S_1^* \\
S_2^*
\end{bmatrix}, \quad S_1^* := R^\top, \quad S_2^* := U_2^* Q.
\] (8)

In the following, we show that $S^*$ is a strict local optimum of (T–NSDP$_r$). Since $S_1^*$ has full rank, Lemma 3.5 guarantees the existence $\delta > 0$ such that

\[
Q \in \mathbb{L}^{r \times r}, \quad QQ^\top = S_1^*(S_1^*)^\top \implies Q = S_1^* \quad \text{or} \quad \|Q - S_1^*\|_F \geq \delta. \quad (9)
\]

We arbitrarily take $S \in \mathbb{L}^{n \times r}$ with

\[
\mathcal{A}(SS^\top) = b, \quad \|S - S^*\|_F < \delta, \quad S \neq S^*.
\] (10)

It is sufficient to show $SS^\top \neq S^*(S^*)^\top$ because the uniqueness of $X^* = U^*(U^*)^\top = S^*(S^*)^\top$ implies that $\langle C, SS^\top \rangle > \langle C, X^* \rangle = \langle C, S^*(S^*)^\top \rangle$ when $SS^\top \neq S^*(S^*)^\top (= X^*)$. To show this assertion, we assume that $SS^\top = S^*(S^*)^\top$. Let $S$ be expressed as

\[
S = \begin{bmatrix}
S_1 \\
S_2
\end{bmatrix}, \quad S_1 \in \mathbb{R}^{r \times r}, \quad S_2 \in \mathbb{R}^{(n-r) \times r}.
\] (11)

From (8), (11), and $SS^\top = S^*(S^*)^\top$, we obtain

\[
S_1S_1^\top = S_1^*(S_1^*)^\top, \quad S_1S_2^\top = S_1^*(S_2^*)^\top.
\] (12)

Combining (9) and the former equality of (12) derives $S_1 = S_1^*$ or $\|S_1 - S_1^*\|_F \geq \delta$. However, the second condition does not hold because $\|S_1 - S_1^*\|_F \leq \|S - S^*\|_F < \delta$ from (10), and hence $S_1 = S_1^*$. Since $S_1^*$ is a non-singular matrix, the latter equality of (12) yields $S_2 = S_2^*$. As a result, we obtain $S_1 = S_1^*$ and $S_2 = S_2^*$, that is, $S = S^*$. This contradicts the third condition of (10). Therefore, $SS^\top \neq S^*(S^*)^\top$ holds.

\[\square\]
Theorem 3.7: Assume that (SDP) has a unique optimal solution \( X^* \) satisfying \( \text{rank}(X^*) = \ell \in [1, r] \). Suppose also that \( X^* \) has the following structure:

\[
X^* = \begin{bmatrix}
X^*_1 & O \\
O & O \\
\end{bmatrix}, \quad X^*_1 \in S^\ell, \quad \text{rank}(X^*_1) = \ell.
\]

If either of the following two statements holds, then \( S^* \) is a strict local optimum of \((T–NSDP_r)\):

(i) \( S^* \in \mathbb{L}^{n \times r} \) is a local optimum of \((T–NSDP_r)\) with \( r \geq r_{m+1} \) and the minimal face of \((SDP)\) containing \( S^*(S^*)^T \) has zero–dimensional;

(ii) there exists \((S^*, w^*) \in \mathbb{L}^{n \times r} \times \mathbb{R}^m \) such that \( \nabla S \mathcal{L}(S^*, w^*) = 0 \) and \( C - A(w^*) \geq 0 \).

Proof: Firstly, we consider the case where statement (i) holds. Proposition 3.2 implies that \( X^* = S^*(S^*)^T \) is a unique optimum of \((SDP)\). Let \( S^* \in \mathbb{L}^{n \times r} \) be denoted by

\[
S^* = \begin{cases}
\begin{bmatrix}
S^*_1 & O \\
S^*_2 & S^*_3 \\
S^*_1 & S^*_2 \\
\end{bmatrix} & \text{if } 1 \leq \ell < r,
\end{cases}
\]

where \( S^*_1 \in \mathbb{L}^{\ell \times \ell} \). Since \( X^* = S^*(S^*)^T \) holds, we get

\[
\begin{bmatrix}
X^*_1 & O \\
O & O \\
\end{bmatrix} = \begin{bmatrix}
S^*_1 (S^*_1)^T & S^*_1 (S^*_2)^T \\
S^*_2 (S^*_1)^T & S^*_2 (S^*_2)^T + S^*_3 (S^*_3)^T \\
S^*_1 (S^*_3)^T & S^*_1 (S^*_2)^T \\
S^*_2 (S^*_1)^T & S^*_2 (S^*_2)^T \\
\end{bmatrix}
\]

if \( 1 \leq \ell < r \),

\[
\begin{bmatrix}
S^*_1 & O \\
S^*_2 & S^*_3 \\
S^*_1 & S^*_2 \\
\end{bmatrix}
\]

if \( \ell = r \).

Thus, we easily see that

\[
0 = \begin{cases}
\text{tr}(S^*_2 (S^*_2)^T + S^*_3 (S^*_3)^T) = \|S^*_2\|^2_F + \|S^*_3\|^2_F & \text{if } 1 \leq \ell < r, \\
\text{tr}(S^*_2 (S^*_2)^T) = \|S^*_2\|^2_F & \text{if } \ell = r.
\end{cases}
\]

As a result, we obtain \( S^*_1 \in \mathbb{L}^{\ell \times \ell} \), \( X^*_1 = S^*_1 (S^*_1)^T \), \( \text{rank}(S^*_1) = \text{rank}(X^*_1) = \ell \), and

\[
S^* = \begin{cases}
\begin{bmatrix}
S^*_1 & O \\
O & O \\
\end{bmatrix} & \text{if } 1 \leq \ell < r,
\end{cases}
\]

if \( \ell = r \). (13)

Now, it follows from Lemma 3.5 and \( \text{rank}(S^*_1) = \ell \) that there exists \( \delta > 0 \) such that

\[
Q \in \mathbb{L}^{\ell \times \ell}, \quad QQ^T = S^*_1 (S^*_1)^T \quad \implies \quad Q = S^*_1 \quad \text{or} \quad \|Q - S^*_1\|_F \geq \delta. \quad (14)
\]

Let \( S \in \mathbb{L}^{n \times r} \) be an arbitrary matrix satisfying

\[
A(SS^T) = b, \quad \|S - S^*\|_F < \delta, \quad S \neq S^*.
\]

Note that \( SS^T \neq S^*(S^*)^T \) is a sufficient condition under which \( S^* \) is a strict local optimum. Indeed, if \( SS^T \neq S^*(S^*)^T \) holds, then the uniqueness of \( X^* = S^*(S^*)^T \) implies that
\((C, SS^T) > \langle C, X^* \rangle = \langle C, S^*(S^*)^T \rangle\). Hence, we show \(SS^T \neq S^*(S^*)^T\) by contradiction. In the following, we consider the case where \(1 \leq \ell < r\). Concerning the case where \(\ell = r\), we can prove \(SS^T \neq S^*(S^*)^T\) in a similar way, and hence we omit its proof.

Let \(S\) be represented as follows:

\[
S = \begin{bmatrix} S_1 & O \\ S_2 & S_3 \end{bmatrix}, \quad S_1 \in \mathbb{L}^{\ell \times \ell}, \quad S_2 \in \mathbb{R}^{(n-\ell)\times \ell}, \quad S_3 \in \mathbb{R}^{(n-\ell)\times (r-\ell)}.
\] (16)

Combining (13), (16), and the assumption \(SS^T = S^*(S^*)^T\) yields

\[
\begin{bmatrix} S_1S_1^T & S_1S_2^T \\ S_2S_1^T & S_2S_2^T + S_3S_3^T \end{bmatrix} = SS^T = S^*(S^*)^T = \begin{bmatrix} S^*_1(S^*_1)^T & O \\ O & O \end{bmatrix}.
\] (17)

Notice that \(S_1 \in \mathbb{L}^{\ell \times \ell}\) from (16), and that \(S_1S_1^T = S^*_1(S^*_1)^T\) from (17). It then follows from (14) that \(S_1 = S^*_1\) or \(\|S_1 - S^*_1\|_F \geq \delta\). Since condition (15) leads to \(\|S_1 - S^*_1\|_F \leq \|S - S^*\|_F < \delta\), we get \(S_1 = S^*_1\). Now, recall that \(S \neq S^*\) by condition (15). Then, (13) and (16) yield \(S_2 \neq O\) or \(S_3 \neq O\). However, we have from (17) that \(\|S_2\|_F^2 + \|S_3\|_F^2 = \text{tr}(S_2^2S_2^2 + S_3S_3^2) = 0\), i.e. \(S_2 = O\) and \(S_3 = O\). Therefore, we see that \(SS^T \neq S^*(S^*)^T\).

Secondly, we assume that statement (ii) holds. Proposition 3.3 ensures that \(X^* = S^*(S^*)^T\) is a unique solution of (SDP). Thus, we can use the same arguments from the case (i) and this completes the proof.

\[\square\]

**Remark 3.1:** A result that corresponds to Theorems 3.6 and 3.7 was not considered in [6,7] because the existing model (NSDP\(_r\)) has no strict local optimum as described in Remark 2.1. By Theorems 3.6 and 3.7, it is expected that solutions can be obtained more rapidly and accurately if we apply second-order methods to (T–NSDP\(_r\)), such as SQP or interior point methods.

### 4. Reformulation of the nonlinear optimization models

Although (NSDP\(_r\)) and (T–NSDP\(_r\)) are nonlinear programming problems, their formulation may be unfamiliar because the variables are matrices. For handling (NSDP\(_r\)) and (T–NSDP\(_r\)) easier, we show that they can be recast into the following optimization problem:

\[
\begin{aligned}
\text{Minimize} & \quad f(x) := \frac{1}{2} \langle Hx, x \rangle \\
\text{subject to} & \quad g_j(x) := \frac{1}{2} \langle G_jx, x \rangle - \frac{1}{2} [b]_j = 0 \quad (j = 1, \ldots, m),
\end{aligned}
\]

(QECQP)

where matrices \(H\) and \(G_j\) \((j = 1, \ldots, m)\) have the following block structures:

\[
H = \begin{bmatrix} \tilde{H}_1 & O \\ \vdots & \ddots \\ O & \tilde{H}_r \end{bmatrix}, \quad \tilde{H}_k \in \mathbb{S}^{\ell_k} \quad (k = 1, \ldots, r),
\]
\[
G_j = \begin{bmatrix}
\tilde{G}_{j1} & \cdots & O \\
\vdots & \ddots & \vdots \\
O & \cdots & \tilde{G}_{jr}
\end{bmatrix} (j = 1, \ldots, m), \quad \tilde{G}_{jk} \in S^k (k = 1, \ldots, r),
\]

where \( s_k, t_k \in \mathbb{N}(k = 1, \ldots, r) \).

To show this reformulation, we consider converting the decision variable matrix \( R \in \mathbb{R}^{n \times r} \) of (NSDP\(_r\)) into a vector. Assume that \( R \) can be written as \( R = [u_1 \cdots u_r] \), where \( u_k \in \mathbb{R}^n(k = 1, \ldots, r) \). For any \( M \in \mathbb{R}^{n \times n} \), we have

\[
\langle M, RR^\top \rangle = \text{tr}(MRR^\top)
= \text{tr}\left(\begin{bmatrix}
\vdots \\
u_1^\top \\
\vdots \\
u_r^\top
\end{bmatrix} \begin{bmatrix} M u_1 & \cdots & M u_r \end{bmatrix}\right)
= \sum_{j=1}^r u_j^\top M u_j
= \begin{bmatrix} u_1^\top & \cdots & u_r^\top \end{bmatrix} \begin{bmatrix}
M & O \\
O & M
\end{bmatrix} \begin{bmatrix} u_1 \\
\vdots \\
u_r
\end{bmatrix}. \tag{18}
\]

By using (18), it can be verified that (NSDP\(_r\)) is equivalent to (QECQP) with \( d = nr \),

\[
H = \begin{bmatrix}
C & O \\
\vdots & \ddots \\
O & C
\end{bmatrix}, \quad G_j = \begin{bmatrix}
A_j & O \\
O & A_j
\end{bmatrix} (j = 1, \ldots, m).
\]

In a similar way to the reformulation of \( R \in \mathbb{R}^{n \times r} \), we consider converting the decision variable matrix \( S \in \mathbb{L}^{n \times r} \) of (T–NSDP\(_r\)) into a vector. Let \( v_j \in \mathbb{R}^{n-k+1} (k = 1, \ldots, r) \) be the column of \( S \) excluding the upper diagonal elements, i.e.

\[
S = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
v_1 & v_2 & \cdots & v_r
\end{bmatrix} \in \mathbb{L}^{n \times r}.
\]
Then, we also see that \((T-\text{NSDP}_r)\) can be reformulated as \((\text{QECQP})\) with 
\[d = nr - \frac{r(r-1)}{2},\]
where \(C_k, A_{1k}, \ldots, A_{mk}\) \((k = 1, \ldots, r)\) are matrices obtained by removing the first \(k-1\) rows and columns of \(C, A_1, \ldots, A_m\), respectively.

**Remark 4.1:** The existing papers [6,7] proposed an augmented Lagrangian method for \((\text{NSDP}_r)\) and provided some discussion regarding the exploitation of sparsity when evaluating the augmented Lagrangian. Similarly, we can exploit sparsity of the proposed model \((\text{QECQP})\). Indeed, the matrices \(H\) and \(G_j\) \((j = 1, \ldots, m)\) have block diagonal structures, and their diagonal blocks are only constructed by \(C\) and \(A_j\) \((j = 1, \ldots, m)\), respectively. This means that the memory needed for implementing \((\text{QECQP})\) is essentially the same as the case of \((\text{SDP})\). Moreover, the Hessian of the Lagrange function of \((\text{QECQP})\), which is required for second-order methods, is a sparse matrix because of the block diagonal structures of the matrices \(H\) and \(G_j\) \((j = 1, \ldots, m)\). If a Newton-type method is used for \((\text{QECQP})\), then we can efficiently compute the Newton equation via some approaches that exploit sparsity, such as the incomplete Cholesky factorization, etc.

**5. Numerical experiments**

In this section, we report some numerical experiments. All the programs were implemented with MATLAB R2022a and ran on a machine with Intel Core i9-9900k 3.60GHz CPU and 128GB of RAM. The following four problems, which are provided in SDPLIB [3], were used in the experiments:

(P1) The control and system problem,
(P2) the graph partition problem,
(P3) the max-cut problem,
(P4) the truss topology design problem.

These problems are originally provided in [8,10]. In the experiments, we compare the performance of \((T-\text{NSDP}_r)\) and \((\text{NSDP}_r)\) when using a second-order method, where \(r = \lceil(\sqrt{8m} + 9 - 1)/2 \rceil\) which is derived from \(r \geq r_{m+1}\) seen in Proposition 3.2. To this end, \((T-\text{NSDP}_r)\) and \((\text{NSDP}_r)\) were reformulated as \((\text{QECQP})\) and they were solved by the interior-point method implemented in the MATLAB solver \texttt{fmincon} with the initial point \(x_0 := e/\|e\|\). Throughout the experiments, we want to check how exactly the problems can be solved, and hence \texttt{OptimalityTolerance} and \texttt{MaxIterations} were respectively set to 0 and 400, and we compared the results of \((T-\text{NSDP}_r)\) and \((\text{NSDP}_r)\) via the following accuracy measure \(E:\)

\[
E(X,y) := \max \left\{ \frac{\|A(X) - b\|}{1 + \|b\|}, \frac{|(C - A^*(y), X)|}{1 + |(C, X)| + |(b, y)|}, \lambda_{\min}(C - A^*(y)) \right\}.
\]

Tables 1–4 respectively show the numerical results of \((T-\text{NSDP}_r)\) and \((\text{NSDP}_r)\) with respect to problems (P1)–(P4). Although the results reported in Tables 1 and 4 seem to
suggest that the two models do not differ a lot, the results presented in Tables 2 and 3 indicate that (T–NSDP$_r$) seems to find more accurate solutions than (NSDP$_r$). In particular, (T–NSDP$_r$) is clearly superior to (NSDP$_r$) for (P3) regarding the accuracy of the obtained solutions.
6. Conclusion

In this paper, we have proposed a new nonlinear optimization model \((T–NSDP_r)\) for \((SDP)\), which can overcome the drawbacks of the existing model \((NSDP_r)\) presented by Burer and Monteiro [6,7]. Since the decision variable spaces of \((NSDP_r)\) and \((T–NSDP_r)\) are respectively \(\mathbb{R}^{n \times r}\) and \(\mathbb{L}^{n \times r} = \{ S \in \mathbb{R}^{n \times r} : [S]_{ij} = 0 \text{ if } i < j \}\), the proposed model is \(\frac{r(r-1)}{2}\)-dimensions smaller than the existing one. Moreover, this dimensional reduction produces a beneficial result that \((T–NSDP_r)\) has a strict local optimum under some appropriate conditions, whereas \((NSDP_r)\) has no strict local minimizers. Hence, we can expect that second-order methods for \((T–NSDP_r)\) can quickly obtain solutions with high accuracy. Furthermore, we have conducted some numerical experiments which show that \((T–NSDP_r)\) can in fact find highly accurate solutions compared with \((NSDP_r)\).

As one of the future works, it can be considered providing sufficient conditions under which second-order sufficient optimality conditions of \((T–NSDP_r)\) are satisfied. Moreover, there is space for improvement on how to determine a smaller dimension \(r\), and hence this is also another future work.

Acknowledgments

We would like to express our sincere gratitude to the editor and anonymous referees for their constructive comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Japan Society for the Promotion of Science KAKENHI for 19K11840, 21K11769, and 21K17709.

Notes on contributors

Yuya Yamakawa is an assistant professor at Graduate School of Informatics, Kyoto University. His current research interests include nonlinear semidefinite programming problems, Riemannian optimization problems, optimal control problems, and their optimization methods, such as primal dual interior point methods, sequential quadratic programming methods, and augmented Lagrangian methods.

Tetsuya Ikegami received a master’s degree from Graduate School of Informatics, Kyoto University in 2017 and is a researcher at Data and Science Solutions Group, Technology Group, Yahoo Japan Corporation.

Ellen H. Fukuda received bachelor’s and master’s degrees in computer science and a Ph.D. degree in applied mathematics from the University of São Paulo, Brazil, in 2011. She is currently an Associate Professor at Graduate School of Informatics, Kyoto University, Japan. Her research interests include the development of methods for continuous optimization, in particular, conic optimization, semidefinite optimization, and multiobjective optimization, as well as applications of optimization in wireless communications.

Nobuo Yamashita is a full professor at Graduate School of Informatics, Kyoto University. His research interests include nonlinear optimization and equilibrium problems, and their applications.
References

[1] A. Barvinok, Problems of distance geometry and convex properties of quadratic maps, Disc. Comput. Geom., 13 (1995), pp. 189–202.
[2] D. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas, 2nd ed., Princeton University Press, Princeton, 2009.
[3] B. Borchers, SDPLIB 1.2, a library of semidefinite programming test problems, Optim. Methods Softw. 11 (1999), pp. 683–690.
[4] N. Boumal, V. Voroninski, and A. Bandeira, The non-convex Burer-Monteiro approach works on smooth semidefinite programs, in Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2765–2773.
[5] N. Boumal, V. Voroninski, and A. Bandeira, Deterministic guarantees for Burer–Monteiro factorizations of smooth semidefinite programs, Commun. Pure Appl. Math. 73 (2020), pp. 581–608.
[6] S. Burer and R. Monteiro, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization, Math. Program., 95 (2003), pp. 329–357.
[7] S. Burer and R. Monteiro, Local minima and convergence in low-rank semidefinite programming, Math. Program. 103 (2005), pp. 427–444.
[8] K. Fujisawa, M. Fukuda, M. Kojima, and K. Nakata, Numerical evaluation of sdpa (semidefinite programming algorithm), Technical Report B–330, Tokyo Institute of Technology, 1997.
[9] I. Grubišić and R. Pietersz, Efficient rank reduction of correlation matrices, Linear Algebra Appl. 422 (2007), pp. 629–653.
[10] C. Helmberg and F. Rendl, A spectral bundle method for semidefinite programming, SIAM J. Optim. 10 (2000), pp. 673–696.
[11] R. Monteiro, C. Ortiz, and B. Svaiter, Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems, Comput. Optim. Appl. 57 (2014), pp. 45–69.
[12] G. Pataki, On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues, Math. Oper. Res. 23 (1998), pp. 339–358.
[13] J. Povh, F. Rendl, and A. Wiegele, A boundary point method to solve semidefinite programs, Computing 78 (2006), pp. 277–286.
[14] J. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optim. Methods Softw. 11 (1999), pp. 625–653.
[15] D. Sun, K. Toh, and L. Yang, An efficient inexact abcd method for least squares semidefinite programming, SIAM J. Optim. 26 (2016), pp. 1072–1100.
[16] M. Todd, Semidefinite optimization, Acta Numer. 10 (2001), pp. 515–560.
[17] K. Toh, M. Todd, and R. Tütüncü, SDPT3 — A MATLAB software package for semidefinite programming, version 1.3, Optim. Methods Softw. 11 (1999), pp. 545–581.
[18] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Rev. 38 (1996), pp. 49–95.
[19] Z. Wen, D. Goldfarb, and W. Yin, Alternating direction augmented lagrangian methods for semidefinite programming, Math. Program. Comput. 2 (2010), pp. 203–230.
[20] H. Wolkowicz, R. Saigal, and L. Vandenberghe, Handbook of Semidefinite Programming: Theory, Algorithms and Applications, Kluwer International Series in Operations Research and Management Science, Boston, 2000.
[21] M. Yamashita, K. Fujisawa, and M. Kojima, Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0), Optim. Methods Softw. 18 (2003), pp. 491–505.
[22] L. Yang, D. Sun, and K. Toh, SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, Math. Program. Comput. 7 (2015), pp. 331–366.
[23] X. Zhao, D. Sun, and K. Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, SIAM J. Optim. 20 (2010), pp. 1737–1765.