Solving Semidefinite Programs via Nonlinear Programming Part I: Transformations and Derivatives*

Samuel Burer† Renato D.C. Monteiro‡ Yin Zhang§

September, 1999
(Revised December, 1999)

Abstract

In this paper, we introduce a transformation that converts a class of linear and nonlinear semidefinite programming (SDP) problems into nonlinear optimization problems over "orthants" of the form $\Re^n_{++} \times \Re^N$, where n is the size of the matrices involved in the problem and N is a nonnegative integer dependent upon the specific problem. For example, in the case of the SDP relaxation of a MAXCUT problem, n is the number of vertices in the underlying graph and N is zero. The class of transformable problems includes instances of SDP relaxations of combinatorial optimization problems with binary variables, as well as other important SDP problems. We also derive formulas for the first and second derivatives of the objective function of the resulting nonlinear optimization problem, hence enabling applications of existing nonlinear optimization techniques to the solution of large-scale SDP problems.

Keywords: transformation, semidefinite program, semidefinite relaxation, nonlinear programming.

AMS 1991 subject classification: 90C06, 90C27, 90C30.

*This is Technical Report TR99-17, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005.

†School of Mathematics, Georgia Tech., Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants CCR-9700448 and CCR-9902010. (E-mail: burer@math.gatech.edu).

‡School of ISyE, Georgia Tech., Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants INT-9600343, CCR-9700448 and CCR-9902010. (Email: monteiro@isye.gatech.edu).

§Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA. This author was supported in part by DOE Grant DE-FG03-97ER25331, DOE/LANL Contract 03891-99-23 and NSF Grant DMS-9973339. (Email: zhang@caam.rice.edu).

1 Introduction

This paper concerns the solution of large-scale linear and/or nonlinear semidefinite programs. In the past several years, semidefinite programming (SDP) has been one of the most active research areas in mathematical programming. There are two major factors that are responsible for this increased interest in SDP. Firstly, SDP has found numerous applications in various fields, such as statistics, structural design, electrical engineering and combinatorial optimization. Secondly, interior point methods have proven to be reliable and effective in the solution of small to moderate sized problems. It is often the case, however, that many problems arising from applications have very large sizes. This is especially true for SDP problems that arise as relaxations of combinatorial optimization problems. An SDP problem is large if either the size of the matrix variable is large, or the number of constraints is large, or both.

Generally speaking, the most popular interior-point methods are primal-dual methods. For large-scale problems, however, the capability of primal-dual interior-point methods are severely limited due to their high demand for storage and computation, despite efforts in exploiting sparsity in data matrices (for example, see [3]). Recently, Benson et al. [2] proposed a potential-reduction dual-scaling interior-point method that can better take advantage of the special structure of SDP relaxations of certain combinatorial optimization problems.

Interior-point methods rely on Newton's method as their solution engine which at each iteration requires computing and storing matrices and solving linear equations. Recently, some nonlinear programming methods have been proposed for solving large SDP problems (see [1, 6, 7]), which rely on gradient-based methods and avoid costly matrix operations and linear solves.

An approach by Helmberg and Rendl [6] is to solve a certain partial Lagrangian dual formulation, whose objective function is non-differentiable, using the bundle method for nonsmooth convex programming. On the other hand, Homer and Peinado [7] use the change of variables $X = VV^T$, $V \in \Re^{n \times n}$, where X is the primal matrix variable of the MAXCUT SDP relaxation, to transform it into a constrained nonlinear programming problem in the new variable V. More recently, Burer and Monteiro [1] proposed a variant of Homer and Peinado's method based on the constrained nonlinear programming reformulation of the MAXCUT SDP relaxation obtained by using the change of variable $X = LL^T$, where L is a lower triangular matrix. We note that the substitutions $X = VV^T$ and $X = LL^T$ can be viewed as a matrix analogy to the square-slack variable substitution for scalar inequality constraints.

The transformation proposed in this paper for converting an SDP problem into a nonlinear programming problem is also based on the same change of variables used in [1]. An important idea introduced in this paper, however, is the derivation of a fundamental mapping expressing a subset of the lower entries (say, the off-diagonal) of L as functions of the other entries (the diagonal) of L. With the aid of this mapping, we show how to transform a large class of linear and/or nonlinear SDP problems into the problem of minimizing a nonlinear function over an "orthant" of the form $\Re^n_{++} \times \Re^N$, where n is the size of L and N is an appropriate nonnegative integer. This transformation has several nice properties, including that the number of variables of the resulting nonlinear program is minimal in the sense that it equals the dimension of the feasible region of the original SDP problem and that the positive-semidefinite-matrix constraint is replaced by a positive-vector constraint.

This paper is organized as follows. In Section 2, we describe the class of SDP problems to which the transformation introduced in Section 3 applies and give several examples of linear and nonlinear SDP problems lying in this class in order to illustrate its generality. In Section 3, we show that every problem in the above class can be converted to a nonlinear optimization problem over an orthant of the form $\Re^n_{++} \times \Re^N$, where n and N are as described above. The fundamental mapping used in this conversion is also described in this section. In Section 4, we develop explicit formulas for the first and second derivatives of the objective function of the nonlinear optimization problem in terms of the first and second derivatives of the functions involved in the original problem and some simple linear operators. In Section 5, we analyze the computational requirement involved in the function and derivative evaluations for both the dense and sparse cases. Some final remarks are offered in the last section.

1.1 Preliminary notation

In this paper, $\Re$, $\Re^n$, and $\Re^{n\times n}$ denote the space of real numbers, real n-dimensional column vectors, and real $n \times n$ matrices, respectively. By $\mathcal{S}^n$ we denote the space of real $n \times n$ symmetric matrices, and we define $\mathcal{S}^n_+$ and $\mathcal{S}^n_{++}$ to be the subsets of $\mathcal{S}^n$ consisting of the positive semidefinite and positive definite matrices, respectively. We write $A \succeq 0$ and $A \succ 0$ to indicate that $A \in \mathcal{S}^n_+$ and $A \in \mathcal{S}^n_{++}$, respectively. We let $\mathrm{tr}(A)$ denote the trace of a matrix $A \in \Re^{n\times n}$, namely the sum of the diagonal elements of A. Moreover, for $A, B \in \Re^{n\times n}$, we define $A \bullet B \equiv \mathrm{tr}(A^T B)$. If $\mathcal{I}$ is a finite set, we let $|\mathcal{I}|$ denote its cardinality, that is the number of elements of $\mathcal{I}$. Given any two functions f and g such that the domain of f is the range of g, then $f \circ g$ denotes the composition of f with g, i.e., $(f \circ g)(\cdot) \equiv f(g(\cdot))$. Given any linear operator Ω, the adjoint operator of Ω is denoted by Ω^*.

This paper is heavy in notation. Instead of collecting all the symbols in one place, we will gradually develop additional notation throughout the course of the paper.

2 The Problem Class and Examples

In this section, we introduce a standard form problem which describes the class of semidefinite programming problems studied in this paper. We then present several specific examples of important linear and nonlinear SDPs which are members of this class.

Define

$$\mathcal{K} \equiv \{(i,j) : 1 \le j \le i \le n\}, \tag{1}$$

i.e., $\mathcal{K}$ consists of the ordered index pairs corresponding to the lower-triangular positions of an $n \times n$ matrix. Let $\mathcal{D} \subseteq \mathcal{K}$ denote the subset corresponding to the diagonal indices, that is, $\mathcal{D} \equiv \{(i,i) : i = 1, 2, \ldots, n\}$. In addition, for any $\mathcal{A} \subseteq \mathcal{K}$, define

$$\mathcal{S}^{\mathcal{A}} \equiv \{X \in \mathcal{S}^n : X_{ij} = 0 \text{ if } i \ge j \text{ and } (i,j) \notin \mathcal{A}\},$$

i.e., $\mathcal{S}^{\mathcal{A}}$ is the subspace of $\mathcal{S}^n$ consisting of those matrices which have zeros in the positions corresponding to $\mathcal{K} \setminus \mathcal{A}$. For instance, $\mathcal{S}^{\mathcal{D}}$ is the set of diagonal matrices.

Let $\mathcal{I}$ be a proper subset of $\mathcal{K}$ which contains all the diagonal indices, namely, $\mathcal{D} \subseteq \mathcal{I} \subset \mathcal{K}$, and let m be a nonnegative integer. Given functions $g : \mathcal{S}^n \times \Re^m \to \Re$ and $H : \Re^m \to \mathcal{S}^n$, consider the following semidefinite program (NLSDP):

$$(NLSDP) \qquad \min \left\{ g(Z^{\mathcal{I}}, y) : (Z^{\mathcal{I}}, y) \in \mathcal{F}_{sdp} \right\}, \tag{2}$$

where

$$\mathcal{F}_{sdp} \equiv \left\{ (Z^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}} \times \Re^m : Z^{\mathcal{I}} + H(y) \succeq 0 \right\}. \tag{3}$$

Note that both functions g and H can be either linear or nonlinear, and $m = 0$ means that the variable y is non-existent. In addition, since $\mathcal{I}$ contains the diagonal indices, for any $y \in \Re^m$, there exists some $Z^{\mathcal{I}} \in \mathcal{S}^{\mathcal{I}}$ such that $Z^{\mathcal{I}} + H(y) \succ 0$, implying that the interior of $\mathcal{F}_{sdp}$, which we denote as int $(\mathcal{F}_{\mathrm{sdp}})$, is nonempty.

Many SDP problems can be put in the form of (2) as the following examples demonstrate.

Example 2.1 *A Class of Dual Linear SDPs*

Let an index set $\mathcal{I}$ such that $\mathcal{D} \subseteq \mathcal{I} \subset \mathcal{K}$ and a nonnegative integer m be given. Consider the following primal SDP:

$$\max \left\{ C \bullet X : X_{ij} = B^{\mathcal{I}}_{ij}, (i,j) \in \mathcal{I};\ A_k \bullet X = b_k,\ k = 1, \ldots, m;\ X \succeq 0 \right\}, \tag{4}$$

where $C, A_1, \ldots, A_m \in \mathcal{S}^n$, $B^{\mathcal{I}} \in \mathcal{S}^{\mathcal{I}}$, and $b \in \Re^m$ are the data. Its dual is

$$\min \left\{ B^{\mathcal{I}} \bullet Z^{\mathcal{I}} + b^T y : Z^{\mathcal{I}} + \sum_{k=1}^{m} y_k A_k - C \succeq 0;\ (Z^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}} \times \Re^m \right\}, \tag{5}$$

which is in the form of (2) with $g(Z^{\mathcal{I}}, y) = B^{\mathcal{I}} \bullet Z^{\mathcal{I}} + b^T y$ and $H(y) = \sum_{k=1}^m y_k A_k - C$.

Example 2.2 *SDP Relaxations of Binary Combinatorial Optimization Problems*

In SDP relaxations of binary combinatorial optimization problems, the binary constraints $x_i = \pm 1,\ i = 1, 2, \ldots, n$, are relaxed into $X_{ii} = 1,\ i = 1, 2, \ldots, n$, (see Lovász and Schrijver [10]), resulting in primal linear SDPs in the form of (4) with $\mathcal{I} = \mathcal{D}$ and $B^{\mathcal{I}} = I$, the identity matrix. The dual form of these SDP relaxations are special instances of (5). In particular, when $m = 0$, we obtain the MAXCUT SDP relaxation introduced by Goemans and Williamson in [4]. Another example where $m > 0$ is given below.

Example 2.3 *The Lovász Theta Number*

Let a simple graph $G = (V, E)$ with vertex set $V = \{1, \ldots, n\}$ be given. For notational convenience, we assume that the edge set E is given as a subset of $\mathcal{K}$, i.e., an edge between u and v, $u \neq v$, is represented in E as either (u, v) if $u > v$ or (v, u) if $v > u$. Let e_{ijk} denote the $(n+1)$-dimensional column vector of all zeros except ones in the i-th, j-th, and k-th positions, and define $A_{ijk} \equiv e_{ijk} e_{ijk}^T$. Then the (unweighted) Lovász theta number $\vartheta(G)$ of the graph G is the optimal value of the $(n+1)$-dimensional SDP

$$\max \left\{ C \bullet X : X_{ii} = 1, i = 1, \ldots, n+1;\ A_{ij(n+1)} \bullet X = 1, (i, j) \in E;\ X \succeq 0 \right\},$$

where

$$C = \begin{bmatrix} 1/2 & & & 1/4 \\ & \ddots & & \vdots \\ & & 1/2 & 1/4 \\ 1/4 & \ldots & 1/4 & \end{bmatrix} \in \mathcal{S}^{n+1}$$

(see Laurent et al. [8]). This problem is an instance of the primal SDP (4) with $\mathcal{I} = \mathcal{D}$ and $m = |E|$. Hence, its dual is an instance of (5).

Example 2.4 *A Second Version of the Lovász Theta Number*

Let the graph G be as in the previous example. It is well-known that the Lovász theta number $\vartheta(G)$ of G also equals the maximum value of the n-dimensional SDP

$$\max \{ J \bullet X : X_{ij} = 0, (i, j) \in E;\ \mathrm{tr}(X) = 1; X \succeq 0 \},$$

where $J = ee^T$ is the $n \times n$ matrix of all ones. This linear SDP can be reformulated as a nonlinear SDP in the form of (2) with $\mathcal{I} \equiv \mathcal{K} \setminus E$, $m = 0$, and

$$g(X^{\mathcal{I}}) = -\frac{J \bullet X^{\mathcal{I}}}{\mathrm{tr}(X^{\mathcal{I}})}.$$

Example 2.5 *The Positive Semidefinite Matrix Completion Problem*

Given an index set $\mathcal{M}$ such that $\mathcal{D} \subseteq \mathcal{M} \subset \mathcal{K}$ and a partially positive semidefinite matrix $A^{\mathcal{M}} \in \mathcal{S}^{\mathcal{M}}$ (i.e., all the principal minors of $A^{\mathcal{M}}$ within $\mathcal{M}$ are positive semidefinite), the positive semidefinite matrix completion problem is either to find a completion matrix $X^{\mathcal{N}} \in \mathcal{S}^{\mathcal{N}}$, where $\mathcal{N} \equiv \mathcal{K} \setminus \mathcal{M}$, such that $A^{\mathcal{M}} + X^{\mathcal{N}} \succeq 0$ or to determine that no such matrix exists.

The problem can be formulated as the nonlinear convex program

$$\min \left\{ \|X^{\mathcal{D}}\|_F^2 : X^{\mathcal{D}} + A^{\mathcal{M}} + X^{\mathcal{N}} \succeq 0;\ (X^{\mathcal{D}}, X^{\mathcal{N}}) \in \mathcal{S}^{\mathcal{D}} \times \mathcal{S}^{\mathcal{N}} \right\},$$

which is in the form of (2) with $\mathcal{I} = \mathcal{D} \cup \mathcal{N}$, $m = 0$, and H the constant matrix $A^{\mathcal{M}}$. Whenever the optimal value is zero, a completion matrix is found; otherwise, there exists no such completion matrix.

3 Reduction to Nonlinear Programming

In this section, we introduce functions that play key roles in transforming the semidefinite program (2) into a nonlinear optimization problem whose feasible region is the orthant $\Re^n_{++} \times \Re^N$, where $n = |\mathcal{D}|$ and $N = |\mathcal{I} \setminus \mathcal{D}| + m$. We also discuss the differentiability of these transformations and conclude with an explicit statement of the equivalent nonlinear programming formulation of (2).

3.1 Further notation

We first define the following symbols: $\mathcal{L}^n, \mathcal{L}^n_+, \mathcal{L}^n_{++}, \mathcal{L}^{\mathcal{A}}, \pi^l_{\mathcal{A}}, \pi_{\mathcal{A}}$ and $\mathcal{L}^{\mathcal{I}}_{++}$.

- Let $\mathcal{L}^n \subset \Re^{n \times n}$ denote the set of real $n \times n$ lower triangular matrices, and let $\mathcal{L}^n_+$ and $\mathcal{L}^n_{++}$ denote the subsets of $\mathcal{L}^n$ consisting of those lower triangular matrices with nonnegative and positive diagonal entries, respectively.
- For any $\mathcal{A} \subseteq \mathcal{K}$, define $\mathcal{L}^{\mathcal{A}} \equiv \{L \in \mathcal{L}^n : L_{ij} = 0 \text{ if } (i,j) \notin \mathcal{A}\}$.
- Let $\pi^l_{\mathcal{A}} : \Re^{n \times n} \to \mathcal{L}^{\mathcal{A}}$ and $\pi_{\mathcal{A}} : \Re^{n \times n} \to \mathcal{S}^{\mathcal{A}}$ denote the orthogonal projection operators from $\Re^{n \times n}$ onto the subspaces $\mathcal{L}^{\mathcal{A}}$ and $\mathcal{S}^{\mathcal{A}}$, respectively.
- For any $\mathcal{I} \subset \mathcal{K}$ containing $\mathcal{D}$, let $\mathcal{L}^{\mathcal{I}}_{++} \equiv \{L : L \in \mathcal{L}^{\mathcal{I}} \text{ and } L_{ii} > 0, i = 1, 2, \ldots, n\}$.

It is well known that for any $X \in \mathcal{S}^n_{++}$, there exists a unique $L \in \mathcal{L}^n_{++}$ such that $X = LL^T$; L is called the lower Cholesky factor of X. We denote the transformation which maps a positive definite matrix into its lower Cholesky factor by $\mathrm{chol} : \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$. Namely, for any $X \in \mathcal{S}^n_{++}$,

$$L = \mathrm{chol}(X) \quad \Longleftrightarrow \quad LL^T = X \text{ and } L \in \mathcal{L}^n_{++}.$$

For each $k = 1, \ldots, m$, let $e^k \in \Re^m$ denote the k-th coordinate vector, and for each $(i,j) \in \mathcal{K}$, let E^{ij} denote the $n \times n$ lower triangular matrix with zeros everywhere except a one in position (i,j).

3.2 The Main ideas

From now on, we assume that $\mathcal{I}$ and $\mathcal{J}$ are given and that $\mathcal{D} \subseteq \mathcal{I} \subset \mathcal{K}$ and $\mathcal{J} \equiv \mathcal{K} \setminus \mathcal{I}$. We also assume that a nonnegative integer m and functions $g : \mathcal{S}^n \times \Re^m \to \Re$ and $H : \Re^m \to \mathcal{S}^n$ are given. For the purpose of following the main ideas presented in this subsection, it would be sufficient, and indeed preferable, for a reader to assume that $\mathcal{I}$ represents the diagonal indices and $\mathcal{J}$ the off-diagonal ones.

We employ two key ideas to transform the feasibility set $\mathcal{F}_{sdp}$ of the SDP problem (2) into a certain "orthant":

1. *The square slack variable substitution* $S = LL^T$ for any $S \in \mathcal{S}^n_+$ that turns the positive-semidefinite cone constraint $Z^{\mathcal{I}} + H(y) \succeq 0$ into the equality constraint
$$Z^{\mathcal{I}} + H(y) = LL^T. \tag{6}$$
2. *An elimination scheme* that uses the equations in (6) to eliminate all the variables L_{ij} for $(i,j) \in \mathcal{J}$, and all the variable in $Z^{\mathcal{I}}$ as well.

While the meaning of first idea is quite clear, that of the second calls for an explanation. Recall that the lower triangular index set $\mathcal{K}$ has been partitioned into $\mathcal{K} = \mathcal{I} \cup \mathcal{J}$. Accordingly, we can partition the lower triangular matrix L into $L = L^{\mathcal{I}} + L^{\mathcal{J}}$, and the equations in (6) into two systems of equations,

$$\pi_{\mathcal{J}} \left[(L^{\mathcal{I}} + L^{\mathcal{J}})(L^{\mathcal{I}} + L^{\mathcal{J}})^T - H(y) \right] = 0 \tag{7}$$

corresponding to the index set $\mathcal{J}$ and containing $|\mathcal{J}|$ distinct equations, and

$$\pi_{\mathcal{I}}(Z^{\mathcal{I}}) = \pi_{\mathcal{I}} \left[(L^{\mathcal{I}} + L^{\mathcal{J}})(L^{\mathcal{I}} + L^{\mathcal{J}})^T - H(y) \right] \tag{8}$$

corresponding to the index set $\mathcal{I}$ and containing $|\mathcal{I}|$ distinct equations. Together, there are $|\mathcal{I}| + |\mathcal{J}| \equiv n(n+1)/2$ equations in (7) and (8); and there are $2|I| + |\mathcal{J}| + m$ variables: $|\mathcal{I}|$ in $L^{\mathcal{I}}$ and also in $Z^{\mathcal{I}}$, $|\mathcal{J}|$ in $L^{\mathcal{J}}$ and m in y. Our strategy is first to use the $|\mathcal{J}|$ equations in (7) to eliminate the $|\mathcal{J}|$ variables in $L^{\mathcal{J}}$, and then to use the $|\mathcal{I}|$ equations in (8) to eliminate the $|\mathcal{I}|$ variables in $Z^{\mathcal{I}}$. In other words, we will use (7) and (8) to express $Z^{\mathcal{I}}$ and $L^{\mathcal{J}}$ as functions of $L^{\mathcal{I}}$ and y.

A fundamental observation, which will be formally stated and proved in Lemma 3.1, is that given any $H(y) \in \mathcal{S}^n$ and $L^{\mathcal{I}} \in \mathcal{L}^{\mathcal{I}}_{++}$, the $|\mathcal{J}|$ variables in $L^{\mathcal{J}}$ can be uniquely determined and explicitly computed from the $|\mathcal{J}|$ equations in (7). In other words, system

(7) defines $L^{\mathcal{J}}$ as a function of $(L^{\mathcal{I}}, y)$, namely, $L^{\mathcal{J}} = L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. By substituting $L^{\mathcal{J}} = L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ into (8), we obtain $Z^{\mathcal{I}}$ as a function of $(L^{\mathcal{I}}, y)$ as well. In fact, since $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ satisfies (7), we have

$$Z^{\mathcal{I}} = (L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y))(L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y))^T - H(y) \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}}.$$

The above two functions, $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ and $Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$, allow us to construct a homeomorphism between the sets $\mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ and $\text{int}\,(\mathcal{F}_{\text{sdp}})$, which maps a point $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ into a point $(Z^{\mathcal{I}}, y) \equiv (Z^{\mathcal{I}}(L^{\mathcal{I}}, y), y) \in \text{int}\,(\mathcal{F}_{\text{sdp}})$.

3.3 Transforming the feasibility set into an orthant

In this subsection we formalize the ideas developed in the previous subsection.

Lemma 3.1 *For each pair $(L^{\mathcal{I}}, H) \in \mathcal{L}^{\mathcal{I}} \times \mathcal{S}^n$ such that $L^{\mathcal{I}}_{ii} \neq 0$ for all $i = 1, \ldots, n$, there exists a unique $L^{\mathcal{J}} \in \mathcal{L}^{\mathcal{J}}$ satisfying*

$$\pi_{\mathcal{J}}(H) = \pi_{\mathcal{J}}\left((L^{\mathcal{I}} + L^{\mathcal{J}})(L^{\mathcal{I}} + L^{\mathcal{J}})^T\right). \tag{9}$$

Proof. System (9) can be expressed as $|\mathcal{J}|$ equations, one for each $(i,j) \in \mathcal{J}$:

$$H_{ij} = \sum_{s=1}^{n}(L^{\mathcal{I}}_{is} + L^{\mathcal{J}}_{is})(L^{\mathcal{I}}_{js} + L^{\mathcal{J}}_{js}) = L^{\mathcal{J}}_{ij}L^{\mathcal{I}}_{jj} + \sum_{s=1}^{j-1}(L^{\mathcal{I}}_{is} + L^{\mathcal{J}}_{is})(L^{\mathcal{I}}_{js} + L^{\mathcal{J}}_{js}),$$

where the second equality follows from the facts that both $L^{\mathcal{I}}$ and $L^{\mathcal{J}}$ are lower triangular matrices, that $i > j$, and that $L^{\mathcal{I}}_{ij} = L^{\mathcal{J}}_{jj} = 0$. It follows that any solution $L^{\mathcal{J}}$ of (9) must satisfy

$$L^{\mathcal{J}}_{ij} = \frac{1}{L^{\mathcal{I}}_{jj}}\left(H_{ij} - \sum_{s=1}^{j-1}(L^{\mathcal{I}}_{is} + L^{\mathcal{J}}_{is})(L^{\mathcal{I}}_{js} + L^{\mathcal{J}}_{js})\right) \tag{10}$$

for all $(i,j) \in \mathcal{J}$. Notice that $L^{\mathcal{J}}_{ij}$ is expressed completely in terms of H_{ij}, $L^{\mathcal{I}}_{jj}$, and the first $j-1$ columns of $L^{\mathcal{I}}$ and $L^{\mathcal{J}}$. Letting j_0 be the smallest integer such that there exists no index pair $(i,j) \in \mathcal{J}$ with $j < j_0$, we thus see that, for each $(i, j_0) \in \mathcal{J}$,

$$L^{\mathcal{J}}_{ij_0} = \frac{1}{L^{\mathcal{I}}_{j_0j_0}}\left(H_{ij_0} - \sum_{s=1}^{j_0-1} L^{\mathcal{I}}_{is}L^{\mathcal{I}}_{j_0s}\right).$$

Hence, column j_0 of $L^{\mathcal{J}}$ has an explicit formula in terms $L^{\mathcal{I}}_{j_0j_0}$, column j_0 of H, and columns 1 though j_0-1 of $L^{\mathcal{I}}$. Now using (10) and a simple induction argument, it is straightforward to see that $L^{\mathcal{J}}$ can be explicitly, recursively, and uniquely computed in a "top-to-bottom and left-to-right" order just like in a Cholesky factorization procedure. ∎

Definition 3.2 *The function $L^{\mathcal{J}} : \mathcal{L}^{\mathcal{I}}_{++} \times \mathcal{S}^n \to \mathcal{L}^{\mathcal{J}}$ is defined by (9), i.e., for any given $(L^{\mathcal{I}}, H) \in \mathcal{L}^{\mathcal{I}}_{++} \times \mathcal{S}^n$, $L^{\mathcal{J}}(L^{\mathcal{I}}, H)$ is the unique solution of (9). Moreover, for any $H : \Re^m \to \mathcal{S}^n$, we also let (with a slight abuse of notation but without confusion) $L^{\mathcal{J}}(L^{\mathcal{I}}, y) \equiv L^{\mathcal{J}}(L^{\mathcal{I}}, H(y))$, which is a mapping from $\mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ to $\mathcal{L}^{\mathcal{J}}$.*

Proposition 3.3 *The function $L^{\mathcal{J}}(L^{\mathcal{I}}, H)$ is infinitely differentiable in $\mathcal{L}^{\mathcal{I}}_{++} \times \mathcal{S}^n$. Furthermore, $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ is as smooth in $\mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ as $H(y)$ is in $\Re^m$. In particular, if $H(y)$ is analytic, then so is $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$.*

Proof. Equation (10) developed in the proof of Lemma 3.1 implies that $L^{\mathcal{J}}(L^{\mathcal{I}}, H)$ is infinitely differentiable in its domain. The statements regarding the differentiability of $L^{\mathcal{J}}(L^{\mathcal{I}}, H)$ follow directly from the differentiability of $L^{\mathcal{J}}(L^{\mathcal{I}}, H)$ and $H(y)$ and the definition of $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. ∎

Definition 3.4 (Ψ Transformation) *The function $\Psi : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \mathcal{S}^{\mathcal{I}} \times \Re^m$ is defined by $\Psi(L^{\mathcal{I}}, y) \equiv (Z^{\mathcal{I}}(L^{\mathcal{I}}, y), y)$, where the function $Z^{\mathcal{I}} : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \mathcal{S}^{\mathcal{I}}$ is given by*

$$Z^{\mathcal{I}}(L^{\mathcal{I}}, y) \equiv (L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y))(L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y))^T - H(y). \tag{11}$$

Note that $Z^{\mathcal{I}}(L^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}}$ due to the fact that (9) holds with $H = H(y)$ and $L^{\mathcal{J}} = L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. In addition, it is straightforward to verify that the function $\Psi^{-1} : \mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}}) \to \mathcal{L}^{\mathcal{I}}_{++} \times \Re^{\mathrm{m}}$ defined by

$$\Psi^{-1}(Z^{\mathcal{I}}, y) \equiv \left(\pi_{\mathcal{I}}[\mathrm{chol}(Z^{\mathcal{I}} + H(y))], y\right)$$

is in fact the inverse of Ψ, and so the following result holds.

Lemma 3.5 *The mapping Ψ is bijective between $\mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ and $\mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}})$.*

Moreover, we note that all the mappings involved in the definitions of Ψ and Ψ^{-1} are infinitely differentiable in the respective domains, except possibly for H. Therefore, both are as smooth as H is in the respective domains. We state this fact as the following proposition.

Proposition 3.6 *The mapping Ψ is as smooth as H is in $\mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and so is Ψ^{-1} in $\mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}})$.*

3.4 The nonlinear programming formulation

Consider the function $f : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \Re$ defined by $f(L^{\mathcal{I}}, y) \equiv g(\Psi(L^{\mathcal{I}}, y))$ and the following nonlinear program (NLP):

$$(NLP) \qquad \inf \left\{ f(L^{\mathcal{I}}, y) : (L^{\mathcal{I}}, y) \in \mathcal{F}_{nlp} \right\}, \tag{12}$$

where the feasible set $\mathcal{F}_{nlp}$ is the open convex set

$$\mathcal{F}_{nlp} \equiv \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m = \Psi^{-1}(\mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}})). \tag{13}$$

Theorem 3.7 *If the function g is continuous and attains it minimum in $\mathcal{F}_{sdp}$, then*

$$\min\{g(Z^{\mathcal{I}}, y) : (Z^{\mathcal{I}}, y) \in \mathcal{F}_{sdp}\} = \inf\{f(L^{\mathcal{I}}, y) : (L^{\mathcal{I}}, y) \in \mathcal{F}_{nlp}\},$$

i.e., the optimal value of (NLSDP) equals the optimal value of (NLP).

Proof. Since Ψ is a homeomorphism between $\mathcal{F}_{nlp}$ and $\mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}})$, the continuity of g implies

$$\begin{aligned}\min\{g(Z^{\mathcal{I}}, y) : (Z^{\mathcal{I}}, y) \in \mathcal{F}_{sdp}\} = \inf\{g(Z^{\mathcal{I}}, y) : (Z^{\mathcal{I}}, y) \in \mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}})\} = \\ \inf\{g(\Psi(L^{\mathcal{I}}, y)) : (L^{\mathcal{I}}, y) \in \Psi^{-1}(\mathrm{int}\,(\mathcal{F}_{\mathrm{sdp}}))\} = \inf\{f(L^{\mathcal{I}}, y) : (L^{\mathcal{I}}, y) \in \mathcal{F}_{nlp}\}.\end{aligned}$$

∎

4 The Derivative Formulas

In this section, we develop formulas for the computation of the first and second derivatives of the objective function f of (NLP). In particular, we provide expressions for the first and second derivatives of f in terms of the first and second derivatives of g and H and some simple linear operators associated with the sets $\mathcal{I}$ and $\mathcal{J}$.

Since the derivation of the derivative formulas provides little insight, for the interest of readability we will first state the formulas without derivation, and then give proofs in a separate subsection along with the necessary technical results.

4.1 Preliminaries

We assume throughout this subsection that g is differentiable on $\mathcal{F}_{sdp}$ and that H is differentiable on $\Re^m$. In addition, we define the linear operator $\mathrm{sim} : \Re^{n\times n} \to \mathcal{S}^n$ by

$$\mathrm{sim}(P) \equiv P + P^T.$$

Given subsets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{K}$ and $L \in \mathcal{L}^n$, we also define the linear operator $\Omega[L]_{\mathcal{B}\mathcal{A}} : \mathcal{L}^{\mathcal{A}} \to \mathcal{S}^{\mathcal{B}}$ by the formula

$$\Omega[L]_{\mathcal{B}\mathcal{A}}(M^{\mathcal{A}}) = \pi_{\mathcal{B}}\left(\mathrm{sim}(M^{\mathcal{A}} L^T)\right).$$

When the matrix L is understood from the context, we will denote $\Omega[L]_{\mathcal{B}\mathcal{A}}$ simply by $\Omega_{\mathcal{B}\mathcal{A}}$. Of particular interest for a given $L \in \mathcal{L}^n$ will be the four operators $\Omega_{\mathcal{I}\mathcal{I}}$, $\Omega_{\mathcal{I}\mathcal{J}}$, $\Omega_{\mathcal{J}\mathcal{I}}$, and $\Omega_{\mathcal{J}\mathcal{J}}$.

Proposition 4.1 *Let $L \in \mathcal{L}^n$ be such that $L_{ii} \neq 0$ for all $i = 1, \ldots, n$. Then $\Omega_{\mathcal{J}\mathcal{J}}$ is invertible.*

Proof. To show that $\Omega_{\mathcal{J}\mathcal{J}}$ is invertible, we must demonstrate that, for each $M^{\mathcal{J}} \in \mathcal{S}^{\mathcal{J}}$, there exists a unique $N^{\mathcal{J}} \in \mathcal{L}^{\mathcal{J}}$ such that $\Omega_{\mathcal{J}\mathcal{J}}(N^{\mathcal{J}}) = M^{\mathcal{J}}$. Using an approach similar to the one in the proof of Lemma 3.1, we note that the system $\Omega_{\mathcal{J}\mathcal{J}}(N^{\mathcal{J}}) = M^{\mathcal{J}}$ consists of $|\mathcal{J}|$ equations, one for each $(i, j) \in \mathcal{J}$:

$$\sum_{s=1}^{j} \left(N^{\mathcal{J}}_{is} L_{js} + L_{is} N^{\mathcal{J}}_{js} \right) = M^{\mathcal{J}}_{ij}.$$

Using that L and $N^{\mathcal{J}}$ are lower triangular matrices, that $i > j$, and that $N^{\mathcal{J}}_{jj} = 0$, we conclude that

$$N^{\mathcal{J}}_{ij} = \frac{1}{L_{jj}} \left(M^{\mathcal{J}}_{ij} - \sum_{s=1}^{j-1} \left(N^{\mathcal{J}}_{is} L_{js} + L_{is} N^{\mathcal{J}}_{js} \right) \right).$$

From this formula, it is clear that $N^{\mathcal{J}}$ can be computed explicitly, recursively, and uniquely in a "top-to-bottom and left-to-right" fashion, similar to the proof of Lemma 3.1. ∎

We will denote the adjoint of the linear operator $\Omega_{\mathcal{B}\mathcal{A}}$ by $\Omega^a_{\mathcal{A}\mathcal{B}}$, that is,

$$\Omega^a_{\mathcal{A}\mathcal{B}} = (\Omega_{\mathcal{B}\mathcal{A}})^*.$$

Clearly, $\Omega^a_{\mathcal{A}\mathcal{B}}$ is a map from $\mathcal{S}^{\mathcal{B}}$ into $\mathcal{L}^{\mathcal{A}}$. The following result gives an explicit expression for the operator $\Omega^a_{\mathcal{A}\mathcal{B}}$.

Proposition 4.2 *For any $L \in \mathcal{L}^n$ and pair of sets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{K}$, the adjoint $\Omega^a_{\mathcal{A}\mathcal{B}} : \mathcal{S}^{\mathcal{B}} \to \mathcal{L}^{\mathcal{A}}$ of the linear operator $\Omega_{\mathcal{B}\mathcal{A}}$ is given by*

$$\Omega^a_{\mathcal{A}\mathcal{B}}(M^{\mathcal{B}}) \equiv 2\, \pi^l_{\mathcal{A}}(M^{\mathcal{B}} L)$$

for all $M^{\mathcal{B}} \in \mathcal{S}^{\mathcal{B}}$.

Proof. To prove the proposition, we must show that, for each pair $(M^{\mathcal{A}}, M^{\mathcal{B}}) \in \mathcal{L}^{\mathcal{A}} \times \mathcal{S}^{\mathcal{B}}$,

$$M^{\mathcal{B}} \bullet \Omega_{\mathcal{B}\mathcal{A}}(M^{\mathcal{A}}) = 2\, \pi^l_{\mathcal{A}}(M^{\mathcal{B}} L) \bullet M^{\mathcal{A}},$$

which is done as follows:

$$\begin{aligned} M^{\mathcal{B}} \bullet \Omega_{\mathcal{B}\mathcal{A}}(M^{\mathcal{A}}) &= M^{\mathcal{B}} \bullet \pi_{\mathcal{B}}(\mathrm{sim}(M^{\mathcal{A}} L^T)) = M^{\mathcal{B}} \bullet \mathrm{sim}(M^{\mathcal{A}} L^T) \\ &= M^{\mathcal{B}} \bullet (M^{\mathcal{A}} L^T) + M^{\mathcal{B}} \bullet (L (M^{\mathcal{A}})^T) \\ &= 2\, \left(M^{\mathcal{B}} L\right) \bullet M^{\mathcal{A}} = 2\, \pi^l_{\mathcal{A}}(M^{\mathcal{B}} L) \bullet M^{\mathcal{A}}, \end{aligned}$$

where the second equality follows from the fact that $M^{\mathcal{B}} \in \mathcal{S}^{\mathcal{B}}$ and the last equality similarly follows from the fact that $M^{\mathcal{A}} \in \mathcal{L}^{\mathcal{A}}$. ∎

4.2 First-derivative formulas

We now state the formulas for the first derivatives (or the gradient) of $f(L^{\mathcal{I}}, y)$. Given $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}} \times \Re^m$, we let $(\partial f/\partial L^{\mathcal{I}})(L^{\mathcal{I}}, y)$ denote the matrix in $\mathcal{L}^{\mathcal{I}}$ whose (i,j)-entry is equal to the partial derivative $(\partial f/\partial L^{\mathcal{I}}_{(i,j)})(L^{\mathcal{I}}, y)$ for every $(i,j) \in \mathcal{I}$, and for a given $(Z^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}} \times \Re^m$, we let $(\partial g/\partial Z^{\mathcal{I}})(Z^{\mathcal{I}}, y)$ denote the matrix in $\mathcal{S}^{\mathcal{I}}$ whose (i,j)-entry is equal to $(\partial g/\partial Z^{\mathcal{I}}_{(i,j)})(Z^{\mathcal{I}}, y)$ for every $(i,j) \in \mathcal{I}$.

Theorem 4.3 *Let* $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ *and* $Z^{\mathcal{I}} \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$. *Then*

$$\begin{aligned}
\frac{\partial f}{\partial L^{\mathcal{I}}}(L^{\mathcal{I}}, y) &= \left[\Omega^a_{\mathcal{I}\mathcal{I}} - \Omega^a_{\mathcal{I}\mathcal{J}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right]\left(\frac{\partial g}{\partial Z^{\mathcal{I}}}(Z^{\mathcal{I}}, y)\right), \\
\frac{\partial f}{\partial y_k}(L^{\mathcal{I}}, y) &= \frac{\partial H(y)}{\partial y_k} \bullet \left[\left[(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}} - I\right]\left(\frac{\partial g}{\partial Z^{\mathcal{I}}}(Z^{\mathcal{I}}, y)\right)\right] + \frac{\partial g}{\partial y_k}(Z^{\mathcal{I}}, y),
\end{aligned}$$

for every $k = 1, \ldots, m$, *where* I *is the identity operator.*

Proof. See Section 4.5. ■

We find it convenient and useful to express the above first-derivative formulas in terms of a matrix-valued function $X(L^{\mathcal{I}}, y) \in \mathcal{S}^n$ defined by

$$X(L^{\mathcal{I}}, y) \equiv \left[I - (\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right](\nabla_{Z^{\mathcal{I}}}\, g), \tag{14}$$

where $\nabla_{Z^{\mathcal{I}}}\, g \equiv \partial g(Z^{\mathcal{I}}, y)/\partial Z^{\mathcal{I}}$.

Corollary 4.4 *Let* $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ *and* $X \equiv X(L^{\mathcal{I}}, y)$. *Then*

$$\nabla_{L^{\mathcal{I}}}\, f(L^{\mathcal{I}}, y) = 2\,\pi^l_{\mathcal{I}}(XL), \tag{15}$$

$$\nabla_y\, f(L^{\mathcal{I}}, y) = \nabla_y\, g - \nabla H(X), \tag{16}$$

where, given $y \in \Re^m$, $\nabla H : \mathcal{S}^n \to \Re^m$ *is the linear operator defined for every* $M \in \mathcal{S}^n$ *by*

$$[\nabla H(M)]_k = \frac{\partial H(y)}{\partial y_k} \bullet M, \quad k = 1, 2, \cdots, m. \tag{17}$$

Proof. The definition of X implies that $(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}(\nabla_{Z^{\mathcal{I}}}\, g) = \nabla_{Z^{\mathcal{I}}}\, g - X$. Substituting this into the first formula in Theorem 4.3, and using Proposition 4.2, we have

$$\begin{aligned}
\nabla_{L^{\mathcal{I}}}\, f(L^{\mathcal{I}}, y) &= \Omega^a_{\mathcal{I}\mathcal{I}}(\nabla_{Z^{\mathcal{I}}}\, g) - \Omega^a_{\mathcal{I}\mathcal{J}}(\nabla_{Z^{\mathcal{I}}}\, g - X) \\
&= 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g]L) - 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g - X]L) \\
&= 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g]L) - 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g]L - XL) \\
&= 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g]L) - 2\,\pi^l_{\mathcal{I}}([\nabla_{Z^{\mathcal{I}}}\, g]L) + 2\,\pi^l_{\mathcal{I}}(XL) \\
&= 2\,\pi^l_{\mathcal{I}}(XL),
\end{aligned}$$

which proves (15). Formula (16) follows directly from the definition of X and the second formula in Theorem 4.3. ■

The matrix $X(L^{\mathcal{I}}, y)$ has the following characterization.

Theorem 4.5 *For any $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ and $\nabla_{Z^{\mathcal{I}}} g \in \mathcal{S}^{\mathcal{I}}$, let $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. Then the matrix $X(L^{\mathcal{I}}, y)$ defined in (14) is the unique solution in $\mathcal{S}^n$ of the equations:*

$$\begin{aligned} \pi_{\mathcal{I}}(X) &= \nabla_{Z^{\mathcal{I}}} g, & (18)\\ \pi^l_{\mathcal{J}}(XL) &= 0. & (19) \end{aligned}$$

Proof. The fact that $X = X(L^{\mathcal{I}}, y)$ satisfies (18) follows directly from the definition of X. Now we show that $X = X(L^{\mathcal{I}}, y)$ also satisfies (19). Using Proposition 4.2 and (14), we obtain

$$\begin{aligned} 2\,\pi^l_{\mathcal{J}}(XL) &= \Omega^a_{\mathcal{J}\mathcal{K}}(X) = \Omega^a_{\mathcal{J}\mathcal{K}}\left(\left[I - (\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right](\nabla_{Z^{\mathcal{I}}} g)\right)\\ &= \Omega^a_{\mathcal{J}\mathcal{K}}(\nabla_{Z^{\mathcal{I}}} g) - \left[\Omega^a_{\mathcal{J}\mathcal{K}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right](\nabla_{Z^{\mathcal{I}}} g)\\ &= \Omega^a_{\mathcal{J}\mathcal{I}}(\nabla_{Z^{\mathcal{I}}} g) - \left[\Omega^a_{\mathcal{J}\mathcal{J}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right](\nabla_{Z^{\mathcal{I}}} g)\\ &= 0, \end{aligned}$$

where the next-to-last inequality follows from the fact that $[(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}](\nabla_{Z^{\mathcal{I}}} g) \in \mathcal{S}^{\mathcal{J}}$ and $\nabla_{Z^{\mathcal{I}}} g \in S^{\mathcal{I}}$, which establishes that $X(L^{\mathcal{I}}, y)$ satisfies (19).

We now give a constructive proof for the uniqueness of solution to the square linear system (18)-(19). We note that (i) (18) and (19) form a square linear system of equations for the lower entries of $X \in \mathcal{S}^n$; (ii) $L \in \mathcal{L}^n_{++}$, hence $L_{jj} > 0$ and $L_{kj} = 0$ for all $k < j$; and (iii) $\mathcal{J}$ does not contain any diagonal entries, hence $i > j$ for all $(i,j) \in \mathcal{J}$. It is already known from (18) that $X_{ij} = X_{ji} = [\nabla_{Z^{\mathcal{I}}} g]_{ij}$ for all $(i,j) \in \mathcal{I}$. We can uniquely determine $X_{ij} = X_{ji}$ for all $(i,j) \in \mathcal{J}$ from the equations in (19) by the recursive formula

$$X_{ij} = -\frac{1}{L_{jj}} \sum_{k>j}^{n} X_{ik} L_{kj}, \quad (i,j) \in \mathcal{J}, \tag{20}$$

in a "bottom-to-top and right-to-left" order (or some other workable order) such that all the $X_{ik} = X_{ki}$ terms in the right-hand side have already been determined before their appearance in the sum. This completes the proof. ■

4.3 Second-derivative formulas

To state the formulas for the second derivatives (or the Hessian) of $f(L^{\mathcal{I}}, y)$, we need to introduce two linear operators $\Gamma : \mathcal{L}^{\mathcal{I}} \to \mathcal{L}^n$, defined by

$$\Gamma(M^{\mathcal{I}}) = \left[I - \Omega^{-1}_{\mathcal{J}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{I}}\right](M^{\mathcal{I}}) \tag{21}$$

for all $M^{\mathcal{I}} \in \mathcal{L}^{\mathcal{I}}$, and $\Theta : \mathcal{S}^n \to \mathcal{L}^{\mathcal{J}}$ defined by

$$\Theta(M) = \left[\Omega^{-1}_{\mathcal{J}\mathcal{J}} \pi_{\mathcal{J}}\right](M) \tag{22}$$

for all $M \in \mathcal{S}^n$.

We let E^{ij} denote the matrix (of appropriate dimension) whose entries are all zeros except for the (i,j)-entry which equal to one. The dimension of E^{ij} will be clear from the context.

Theorem 4.6 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and define $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$, $Z^{\mathcal{I}} \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$ and $X \equiv X(L^{\mathcal{I}}, y)$. Then for all $(i,j), (p,q) \in \mathcal{I}$ and all $k, \ell \in \{1, \ldots, m\}$,*

$$
\begin{aligned}
\frac{\partial^2 f(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} &= 2\,\Gamma^*\left(\pi^l_{\mathcal{K}}\left[X\,\Gamma(E^{ij})\right]\right) \bullet E^{pq} + g''_{Z^{\mathcal{I}}Z^{\mathcal{I}}}\left(\frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{ij}}, \frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{pq}}\right), \\
\frac{\partial^2 f(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial y_k} &= 2\,\Theta^*\left(\pi^l_{\mathcal{J}}\left[X\,\Gamma(E^{ij})\right]\right) \bullet H_k + g''_{Z^{\mathcal{I}}y}\left(\frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{ij}}, e_k\right) + g''_{Z^{\mathcal{I}}Z^{\mathcal{I}}}\left(\frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{ij}}, \frac{\partial Z^{\mathcal{I}}}{\partial y_k}\right), \\
\frac{\partial^2 f(L^{\mathcal{I}}, y)}{\partial y_k \partial y_\ell} &= 2\,\Theta^*\left(\pi^l_{\mathcal{J}}\left[X\,\Theta(H_k)\right]\right) \bullet H_\ell - X \bullet H_{k\ell} + g''_{Z^{\mathcal{I}}Z^{\mathcal{I}}}\left(\frac{\partial Z^{\mathcal{I}}}{\partial y_k}, \frac{\partial Z^{\mathcal{I}}}{\partial y_l}\right) \\
&\quad + g''_{Z^{\mathcal{I}}y}\left(\frac{\partial Z^{\mathcal{I}}}{\partial y_k}, e_l\right) + g''_{Z^{\mathcal{I}}y}\left(\frac{\partial Z^{\mathcal{I}}}{\partial y_l}, e_k\right) + g''_{yy}(e_k, e_l),
\end{aligned}
$$

where $H_k \equiv \partial H(y)/\partial y_k$ and e_k is the k-th unit vector in $\Re^m$ for $k = 1, \ldots, m$, and for every $M^{\mathcal{I}}, N^{\mathcal{I}} \in \mathcal{S}^{\mathcal{I}}$ and $u, v \in \Re^m$,

$$
\begin{aligned}
g''_{Z^{\mathcal{I}}Z^{\mathcal{I}}}\left(M^{\mathcal{I}}, N^{\mathcal{I}}\right) &\equiv \sum_{(\alpha,\beta)\in\mathcal{I}} \sum_{(\gamma,\delta)\in\mathcal{I}} \frac{\partial^2 g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}_{\alpha\beta} \partial Z^{\mathcal{I}}_{\gamma\delta}} M^{\mathcal{I}}_{\alpha\beta} N^{\mathcal{I}}_{\gamma\delta}, \\
g''_{Z^{\mathcal{I}}y}\left(M^{\mathcal{I}}, u\right) &\equiv \sum_{(\alpha,\beta)\in\mathcal{I}} \sum_{\gamma=1}^{m} \frac{\partial^2 g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}_{\alpha\beta} \partial y_\gamma} M^{\mathcal{I}}_{\alpha\beta} u_\gamma, \\
g''_{yy}(u, v) &\equiv \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} \frac{\partial^2 g(Z^{\mathcal{I}}, y)}{\partial y_\alpha \partial y_\beta} u_\alpha v_\beta.
\end{aligned}
$$

Proof. See Section 4.5. ∎

4.4 Second-derivative results for linear SDP

We first specialize the second-derivative formulas to the linear SDP (5), i.e.,

$$\min\left\{B^{\mathcal{I}} \bullet Z^{\mathcal{I}} + b^T y : Z^{\mathcal{I}} + \sum_{k=1}^{m} y_k A_k - C \succeq 0;\ (Z^{\mathcal{I}}, y) \in \mathcal{S}^{\mathcal{I}} \times \Re^m\right\}.$$

It is well known that every bilinear map $B : V \times W \to \Re$, with V and W being inner product vector spaces can be identified with a linear transformation $T : W \to V$ according to the relation $B(v, w) = \langle T(w), v \rangle$ for all $v \in V$ and $w \in W$. In the result below, for any fixed $(L^{\mathcal{I}}, y)$ we view the four partial derivative bilinear maps $f''_{L^{\mathcal{I}}L^{\mathcal{I}}}(L^{\mathcal{I}}, y)$, $f''_{L^{\mathcal{I}}y}(L^{\mathcal{I}}, y)$, $f''_{yL^{\mathcal{I}}}(L^{\mathcal{I}}, y)$ and $f''_{yy}(L^{\mathcal{I}}, y)$ as linear transformations from $\mathcal{L}^{\mathcal{I}}$ to $\mathcal{L}^{\mathcal{I}}$, from $\Re^m$ to $\mathcal{S}^{\mathcal{I}}$, from $\mathcal{S}^{\mathcal{I}}$ to $\Re^m$ and from $\Re^m \to \Re^m$, respectively, and provide concise expressions for them for the case of the linear SDP (5). Let the linear operator $\mathbf{A} : \mathcal{S}^n \to \Re^m$ be defined by $[\mathbf{A}(M)]_k = A_k \bullet M$ for all $k = 1, 2, \ldots, m$ and $M \in \mathcal{S}^n$. Then in the case of the linear SDP, $H(y) = C + \mathbf{A}^*(y)$, where $\mathbf{A}^* : \Re^m \to \mathcal{S}^n$ is given by $\mathbf{A}^*(y) = \sum_{k=1}^m A_k y_k$, and hence $H'(y) = \mathbf{A}^*$, for all $y \in \Re^m$.

Corollary 4.7 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and define $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$, $Z^{\mathcal{I}} \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$, and $X \equiv X(L^{\mathcal{I}}, y) = \left[I - (\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right](B^{\mathcal{I}})$. Let $g(Z^{\mathcal{I}}, y)$ and $H(y)$ be defined as in (5), then*

$$\begin{aligned} f''_{L^{\mathcal{I}}L^{\mathcal{I}}}(L^{\mathcal{I}}, y) &= 2\,\Gamma^* \circ \mathcal{U}_{\mathcal{K}\mathcal{K}} \circ \Gamma, \\ f''_{yL^{\mathcal{I}}}(L^{\mathcal{I}}, y) = \left[f''_{L^{\mathcal{I}}y}(L^{\mathcal{I}}, y)\right]^* &= 2\,\mathbf{A} \circ \Theta^* \circ \mathcal{U}_{\mathcal{J}\mathcal{K}} \circ \Gamma \\ f''_{yy}(L^{\mathcal{I}}, y) &= 2\,\mathbf{A} \circ \Theta^* \circ \mathcal{U}_{\mathcal{J}\mathcal{J}} \circ \Theta \circ \mathbf{A}^*, \end{aligned}$$

where, for given $\mathcal{A}, \mathcal{B} \subseteq \mathcal{K}$, the operator $\mathcal{U}_{\mathcal{B}\mathcal{A}} : \mathcal{L}^{\mathcal{A}} \to \mathcal{L}^{\mathcal{B}}$ is defined by

$$\mathcal{U}_{\mathcal{B}\mathcal{A}}(N^{\mathcal{A}}) = \pi^l_{\mathcal{B}}(XN^{\mathcal{A}})),\ \forall N^{\mathcal{A}} \in \mathcal{L}^{\mathcal{A}}.$$

Proof. It follows as an immediate consequence of Theorem 4.6. ∎

We observe that the map $\mathcal{U}_{\mathcal{B}\mathcal{A}}$ introduced in the above corollary is symmetric whenever $\mathcal{A} = \mathcal{B}$.

In what follows, we give sufficient conditions for the Hessian of $f(Z^{\mathcal{I}}, y)$ to be positive (semidefinite) definite. Let us define the following two cross-product linear operators $\mathcal{T} : \mathcal{L}^{\mathcal{I}} \times \Re^m \to \mathcal{L}^n \times \mathcal{L}^{\mathcal{J}}$ and $\mathcal{U} : \mathcal{L}^n \times \mathcal{L}^{\mathcal{J}} \to \mathcal{L}^n \times \mathcal{L}^{\mathcal{J}}$ by

$$\mathcal{T} = \begin{bmatrix} \Gamma & 0 \\ 0 & \Theta \circ \mathbf{A}^* \end{bmatrix} \quad \text{and} \quad \mathcal{U} = \begin{bmatrix} \mathcal{U}_{\mathcal{K}\mathcal{K}} & \mathcal{U}^*_{\mathcal{J}\mathcal{K}} \\ \mathcal{U}_{\mathcal{J}\mathcal{K}} & \mathcal{U}_{\mathcal{J}\mathcal{J}} \end{bmatrix},$$

where the meanings of the "operator matrices" should be self-evident. Under this notation, we clearly have the Hessian formula

$$\nabla^2 f(L^{\mathcal{I}}, y) \equiv f''(L^{\mathcal{I}}, y) = 2\,\mathcal{T}^* \circ \mathcal{U} \circ \mathcal{T}. \tag{23}$$

Theorem 4.8 *Let the conditions of Corollary 4.7 hold. Then,*

(a) $\nabla^2 f(L^{\mathcal{I}}, y)$ is positive semidefinite if $X \succeq 0$;

(b) $\nabla^2 f(L^{\mathcal{I}}, y)$ *is positive definite if* $X \succ 0$ *and* $\pi_{\mathcal{J}} \circ \mathcal{A}^*$ *is injective.*

Proof. To prove (a), it suffices to prove that the self-adjoint operator $\mathcal{U}$ is positive semidefinite, or equivalently that $P \bullet \mathcal{U}(P) \geq 0$ for every $P = (\hat{L}, \hat{L}^{\mathcal{J}}) \in \mathcal{L}^n \times \mathcal{L}^{\mathcal{J}}$, whenever $X \succeq 0$. Indeed, assume that $X \succeq 0$. Then,

$$\begin{aligned} P \bullet \mathcal{U}(P) &= \hat{L} \bullet \mathcal{U}_{\mathcal{K}\mathcal{K}}(\hat{L}) + \hat{L}^{\mathcal{J}} \bullet \mathcal{U}_{\mathcal{J}\mathcal{J}}(\hat{L}^{\mathcal{J}}) + 2\hat{L}^{\mathcal{J}} \bullet \mathcal{U}_{\mathcal{J}\mathcal{K}}(\hat{L}) \\ &= \hat{L} \bullet \pi^l_{\mathcal{K}}(X\hat{L}) + \hat{L}^{\mathcal{J}} \bullet \pi^l_{\mathcal{J}}(X\hat{L}^{\mathcal{J}}) + 2\hat{L}^{\mathcal{J}} \bullet \pi^l_{\mathcal{J}}(X\hat{L}) \\ &= \hat{L} \bullet (X\hat{L}) + \hat{L}^{\mathcal{J}} \bullet (X\hat{L}^{\mathcal{J}}) + 2\hat{L}^{\mathcal{J}} \bullet (X\hat{L}) \\ &= (\hat{L} + \hat{L}^{\mathcal{J}}) \bullet \left[X\left(\hat{L} + \hat{L}^{\mathcal{J}}\right)\right] = \|X^{1/2}(\hat{L} + \hat{L}^{\mathcal{J}})\|_F^2 \geq 0, \end{aligned}$$

where in the fourth equality we used the fact that $X \in \mathcal{S}^n$. Observe that if $X \succ 0$, then $P \bullet \mathcal{U}(P) = 0$ if and only if $\hat{L} + \hat{L}^{\mathcal{J}} = 0$.

To show that $f''(L^{\mathcal{I}}, y) = 2\mathcal{T}^* \circ \mathcal{U} \circ \mathcal{T}$ is positive definite under the stated conditions, let $Q = (N^{\mathcal{I}}, u) \in \mathcal{L}^{\mathcal{I}} \times \Re^m$ be such that $Q \bullet (\mathcal{T}^* \circ \mathcal{U} \circ \mathcal{T})(Q) = 0$. It is easy to see using the above observation and the definition of $\mathcal{T}$ that this is equivalent to $\Gamma(N^{\mathcal{I}}) + (\Theta \circ \mathbf{A}^*)(u) = 0$. Since Θ is a map from $\mathcal{S}^n$ to $\mathcal{L}^{\mathcal{J}}$, this equation implies that $\pi^l_{\mathcal{I}}\Gamma(N^{\mathcal{I}}) = 0$. Now, using the definition of Γ, it is easy to see that the operator $\pi^l_{\mathcal{I}} \circ \Gamma$ is the identity operator from $\mathcal{L}^{\mathcal{I}}$ onto itself. Hence, we conclude that $N^{\mathcal{I}} = 0$, and consequently that $(\Theta \circ \mathbf{A}^*)(u) = 0$. Using the definition of Θ, we easily see that $\pi_{\mathcal{J}} \circ \mathbf{A}^*$ is injective if and only if $\Theta \circ \mathbf{A}^*$ is injective. Hence, we conclude that $u = 0$. We have thus established that $f''(L^{\mathcal{I}}, y)$ is positive definite.
■

4.5 Proofs and technical results

In this subsection, we provide proofs for Theorems 4.3 and 4.6, together with some necessary technical results. Those readers who are not interested in the technical details may skip this subsection.

Lemma 4.9 *Let* $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ *and* $(i,j) \in \mathcal{I}$ *be given, and define* $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. *Then*

$$\frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij}} = \left[\Omega_{\mathcal{I}\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{I}}\right](E^{ij}). \tag{24}$$

Proof. In order to simplify notation, let φ'_{ij} denote $\partial L^{\mathcal{J}}(L^{\mathcal{I}}, y)/\partial L^{\mathcal{I}}_{ij}$, where $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ is defined in Definition 3.2. Differentiating both sides of equation (9) with respect to $L^{\mathcal{I}}_{ij}$, we obtain

$$\pi_{\mathcal{J}}\left(\operatorname{sim}\left(\left(E^{ij} + \varphi'_{ij}\right) L^T\right)\right) = 0, \tag{25}$$

which immediately implies

$$\varphi'_{ij} = -\left[\Omega_{\mathcal{J}\mathcal{J}}^{-1}\Omega_{\mathcal{J}\mathcal{I}}\right](E^{ij}). \tag{26}$$

Differentiating (11) with respect to $L^{\mathcal{I}}_{ij}$ and using relations (25) and (26), we obtain

$$\begin{aligned}\frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}},y)}{\partial L^{\mathcal{I}}_{ij}} &= L\left[E^{ij}+\varphi'_{ij}\right]^T + \left[E^{ij}+\varphi'_{ij}\right]L^T \\ &= \mathrm{sim}\left(\left[E^{ij}+\varphi'_{ij}\right]L^T\right) = \pi_{\mathcal{I}}\left(\mathrm{sim}\left(\left[E^{ij}+\varphi'_{ij}\right]L^T\right)\right) \\ &= \Omega_{\mathcal{I}\mathcal{J}}(\varphi'_{ij}) + \Omega_{\mathcal{I}\mathcal{I}}(E^{ij}) = \left[\Omega_{\mathcal{I}\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{J}}^{-1}\Omega_{\mathcal{J}\mathcal{I}}\right](E^{ij}).\end{aligned} \tag{27}$$

∎

Lemma 4.10 *Let* $(L^{\mathcal{I}},y) \in \mathcal{L}^{\mathcal{I}}_{++}\times\Re^m$ *and* $k\in\{1,\ldots,m\}$ *be given, and define* $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}},y)$*. Then*

$$\frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}},y)}{\partial y_k} = \left[\Omega_{\mathcal{I}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{J}}^{-1}\pi_{\mathcal{J}} - \pi_{\mathcal{I}}\right]\left(\frac{\partial H(y)}{\partial y_k}\right). \tag{28}$$

Proof. We take a similar approach as in the proof of the Lemma 4.9. Letting φ'_k denote $\partial L^{\mathcal{J}}(L^{\mathcal{I}},y)/\partial y_k$ and differentiating (9) with respect to y_k, where $H = H(y)$, we obtain

$$\pi_{\mathcal{J}}(\mathrm{sim}(\varphi'_k L^T)) = \pi_{\mathcal{J}}\left(\frac{\partial H(y)}{\partial y_k}\right), \tag{29}$$

which immediately implies that

$$\varphi'_k = \left[\Omega_{\mathcal{J}\mathcal{J}}^{-1}\pi_{\mathcal{J}}\right]\left(\frac{\partial H(y)}{\partial y_k}\right). \tag{30}$$

Differentiating (11) with respect to y_k and using relations (29) and (30), we obtain

$$\begin{aligned}\frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}},y)}{\partial y_k} &= L(\varphi'_k)^T + \varphi'_k L^T - \frac{\partial H(y)}{\partial y_k} = \mathrm{sim}(\varphi'_k L^T) - \frac{\partial H(y)}{\partial y_k} \\ &= \pi_{\mathcal{I}}(\mathrm{sim}(\varphi'_k)L^T) + \pi_{\mathcal{J}}(\mathrm{sim}(\varphi'_k)L^T) - \pi_{\mathcal{I}}\left(\frac{\partial H(y)}{\partial y_k}\right) - \pi_{\mathcal{J}}\left(\frac{\partial H(y)}{\partial y_k}\right) \\ &= \pi_{\mathcal{I}}(\mathrm{sim}(\varphi'_k)L^T) - \pi_{\mathcal{I}}\left(\frac{\partial H(y)}{\partial y_k}\right) = \Omega_{\mathcal{I}\mathcal{J}}(\varphi'_k) - \pi_{\mathcal{I}}\left(\frac{\partial H(y)}{\partial y_k}\right) \\ &= \left[\Omega_{\mathcal{I}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{J}}^{-1}\pi_{\mathcal{J}} - \pi_{\mathcal{I}}\right]\left(\frac{\partial H(y)}{\partial y_k}\right).\end{aligned}$$

∎

Proof of Theorem 4.3

Now we are ready to prove Theorem 4.3 which contains the formulas for the first derivatives.
Proof. For simplicity of notation, we let $(\partial f/\partial L^{\mathcal{I}}) = (\partial f/\partial L^{\mathcal{I}})(L^{\mathcal{I}}, y)$ and $(\partial g/\partial Z^{\mathcal{I}}) = (\partial g/\partial Z^{\mathcal{I}})(Z^{\mathcal{I}}, y)$. Using (24), the adjoint property for inner products, and the fact that $f = g \circ \Psi$, we obtain

$$
\begin{aligned}
\frac{\partial f}{\partial L^{\mathcal{I}}_{ij}} &= \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \bullet \left(\frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{ij}}\right) = \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \bullet \left\{ \left[\Omega_{\mathcal{I}\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{I}}\right](E^{ij}) \right\} \\
&= \left\{ \left[\Omega_{\mathcal{I}\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\Omega_{\mathcal{J}\mathcal{I}}\right]^* \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \right\} \bullet E^{ij} \\
&= \left\{ \left[\Omega^a_{\mathcal{I}\mathcal{I}} - \Omega^a_{\mathcal{I}\mathcal{J}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}}\right] \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \right\} \bullet E^{ij},
\end{aligned}
$$

which immediately implies (14). Similarly, using (28), we obtain

$$
\begin{aligned}
\frac{\partial f}{\partial y_k} &= \left(\frac{\partial g}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial Z^{\mathcal{I}}}{\partial y_k}\right) + \frac{\partial g}{\partial y_k} \\
&= \left\{ \frac{\partial g}{\partial Z^{\mathcal{I}}} \bullet \left[\Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\pi_{\mathcal{J}} - \pi_{\mathcal{I}}\right] \left(\frac{\partial H}{\partial y_k}\right) \right\} + \frac{\partial g}{\partial y_k} \\
&= \left\{ \left[\Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\pi_{\mathcal{J}} - \pi_{\mathcal{I}}\right]^* \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \bullet \left(\frac{\partial H}{\partial y_k}\right) \right\} + \frac{\partial g}{\partial y_k} \\
&= \left\{ \left[\pi^*_{\mathcal{J}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}} - \pi^*_{\mathcal{I}}\right] \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \bullet \left(\frac{\partial H}{\partial y_k}\right) \right\} + \frac{\partial g}{\partial y_k} \\
&= \frac{\partial H}{\partial y_k} \bullet \left[\left[(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{I}} - I\right] \left(\frac{\partial g}{\partial Z^{\mathcal{I}}}\right) \right] + \frac{\partial g}{\partial y_k},
\end{aligned}
$$

which establishes (14). Note that in the last equality we have used the fact that the adjoint of the projection operator $\pi_{\mathcal{A}} : \Re^{n\times n} \to \mathcal{S}^{\mathcal{A}}$ is the inclusion operator from $\mathcal{S}^{\mathcal{A}}$ into $\Re^{n\times n}$, that is $(\pi_{\mathcal{A}})(U^{\mathcal{A}}) = U^{\mathcal{A}}$ for all $U^{\mathcal{A}} \in \mathcal{S}^{\mathcal{A}}$. ∎

Now we work on the derivation of the second-derivative formulas.

Lemma 4.11 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ and define $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. Then, for every $(i,j), (p,q) \in \mathcal{I}$ and $k, l \in \{1, \ldots, m\}$, we have*

$$
\begin{aligned}
\frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij}\partial L^{\mathcal{I}}_{pq}} &= \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\pi_{\mathcal{J}}\right] \left(\mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T)\right), \qquad (31) \\
\frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij}\partial y_k} &= \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\pi_{\mathcal{J}}\right] \left(\mathrm{sim}\left(\Gamma(E^{ij})\,\Theta\,(H_k)^T\right)\right), \\
\frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial y_k \partial y_\ell} &= \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}}\Omega^{-1}_{\mathcal{J}\mathcal{J}}\pi_{\mathcal{J}}\right] \left(\mathrm{sim}\left(\Theta\,(H_k)\;\Theta\,(H_\ell)^T\right) - H_{kl}\right),
\end{aligned}
$$

where $H_k \equiv \partial H(y)/\partial y_k$ and $H_{kl} \equiv \partial^2 H(y)/\partial y_k \partial y_l$, for all $k, l = 1, \ldots, m$.

Proof. We prove only the first equality because the second and third can be derived in a similar manner as the first. Let φ''_{ijpq} denote $\partial^2 L^{\mathcal{J}}(L^{\mathcal{I}}, y)/\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}$. Differentiating both sides of the equation (25) obtained in the proof of Lemma 4.9 with respect to $L^{\mathcal{I}}_{pq}$, we obtain

$$\pi_{\mathcal{J}}\left(\mathrm{sim}\left(\varphi''_{ijpq} L^T\right) + \mathrm{sim}\left[(E^{ij} + \varphi'_{ij})(E^{pq} + \varphi'_{pq})^T\right]\right) = 0. \tag{32}$$

Recalling (26) and noting that a similar equation holds for (p, q), we easily see that $E^{ij} + \varphi'_{ij} = \Gamma(E^{ij})$ and $E^{pq} + \varphi'_{pq} = \Gamma(E^{pq})$. Replacing these two formulas into (32) and recalling the definition of $\Omega_{\mathcal{J}\mathcal{J}}$, we obtain

$$\varphi''_{ijpq} = -\left[\Omega^{-1}_{\mathcal{J}\mathcal{J}} \pi_{\mathcal{J}}\right]\left(\mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T)\right). \tag{33}$$

Differentiating equation (27) found in the proof of Lemma 4.9 with respect to $L^{\mathcal{I}}_{pq}$ and using (32) and (33), we see that

$$\begin{aligned}
\frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} &= \mathrm{sim}\left[\varphi''_{ijpq} L^T + (E^{ij} + \varphi'_{ij})(E^{pq} + \varphi'_{pq})^T\right] \\
&= \pi_{\mathcal{I}}\left[\mathrm{sim}\left(\varphi''_{ijpq} L^T + \Gamma(E^{ij})\Gamma(E^{pq})^T\right)\right] \\
&= \Omega_{\mathcal{I}\mathcal{J}}(\varphi''_{ijpq}) + \pi_{\mathcal{I}}\left(\mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T)\right) \\
&= \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}} \Omega^{-1}_{\mathcal{J}\mathcal{J}} \pi_{\mathcal{J}}\right](\mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T)).
\end{aligned}$$

■

Lemma 4.12 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and define $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ and $Z^{\mathcal{I}} \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$. Then, for all $(i, j), (p, q) \in \mathcal{I}$ and all $k, \ell \in \{1, \ldots, m\}$, we have*

$$\frac{\partial g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} = 2\,\Gamma^*\left(\pi^l_{\mathcal{K}}\left[X\,\Gamma(E^{ij})\right]\right) \bullet E^{pq}, \tag{34}$$

$$\frac{\partial g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial y_k} = 2\,\Theta^*\left(\pi^l_{\mathcal{J}}\left[X\,\Gamma(E^{ij})\right]\right) \bullet H_k, \tag{35}$$

$$\frac{\partial g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial y_k \partial y_\ell} = 2\,\Theta^*\left(\pi^l_{\mathcal{J}}\left[X\,\Theta(H_k)\right]\right) \bullet H_\ell - X \bullet H_{kl}, \tag{36}$$

where $H_k \equiv \partial H(y)/\partial y_k$, $H_{kl} \equiv \partial^2 H(y)/\partial y_k \partial y_l$, for all $k, l = 1, \ldots, m$, and $X \equiv X(L^{\mathcal{I}}, y)$ is given by (14).

Proof. We prove only the first equality because the second and third can be proved similarly. We write $\partial g(Z^{\mathcal{I}}, y)/\partial Z^{\mathcal{I}}$ simply as $\nabla_{Z^{\mathcal{I}}}\, g$. Using Lemma 4.11 and the standard properties of the inner product of matrices, we have

$$
\begin{aligned}
\nabla_{Z^{\mathcal{I}}} g \bullet \frac{\partial^2 Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} &= \nabla_{Z^{\mathcal{I}}} g \bullet \left\{ \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}} \Omega^{-1}_{\mathcal{J}\mathcal{J}} \pi_{\mathcal{J}} \right] \left(\mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T) \right) \right\} \\
&= \left\{ \left[\pi_{\mathcal{I}} - \Omega_{\mathcal{I}\mathcal{J}} \Omega^{-1}_{\mathcal{J}\mathcal{J}} \pi_{\mathcal{J}} \right]^* (\nabla_{Z^{\mathcal{I}}} g) \right\} \bullet \mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T) \\
&= \left\{ \left[I - (\Omega^a_{\mathcal{J}\mathcal{J}})^{-1} \Omega^a_{\mathcal{J}\mathcal{I}} \right] (\nabla_{Z^{\mathcal{I}}} g) \right\} \bullet \mathrm{sim}(\Gamma(E^{ij})\Gamma(E^{pq})^T) \\
&= 2\, X \bullet \left[\Gamma(E^{pq})\Gamma(E^{ij})^T \right] = 2\, \left(X\, \Gamma(E^{ij}) \right) \bullet \Gamma(E^{pq}) \\
&= 2\, \pi^l_{\mathcal{K}} \left(X\, \Gamma(E^{ij}) \right) \bullet \Gamma(E^{pq}) = 2\, \Gamma^* \left(\pi^l_{\mathcal{K}} \left(X\, \Gamma(E^{ij}) \right) \right) \bullet E^{pq},
\end{aligned}
$$

where in the second-to-the-last equality we used the fact that the range of Γ consists of lower triangular matrices. ■

Proof of Theorem 4.6

We are ready to prove Theorem 4.6 which contains the formulas for the second derivatives.
Proof. To prove the first equality, we note that

$$
\frac{\partial f(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij}} = \frac{\partial g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij}}.
$$

Taking the derivative of this expression and applying the chain rule gives

$$
\frac{\partial^2 f(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} = \frac{\partial g(Z^{\mathcal{I}}, y)}{\partial Z^{\mathcal{I}}} \bullet \frac{\partial Z^{\mathcal{I}}(L^{\mathcal{I}}, y)}{\partial L^{\mathcal{I}}_{ij} \partial L^{\mathcal{I}}_{pq}} + g''_{Z^{\mathcal{I}} Z^{\mathcal{I}}} \left(\frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{ij}}, \frac{\partial Z^{\mathcal{I}}}{\partial L^{\mathcal{I}}_{pq}} \right).
$$

The first identity of the theorem now follows from the above expression and (34). The other two identities can be derived in a similar way with the aid of (35) and (36). ■

5 Complexity of the Function and Derivative Evaluations

In this section, we calculate the computational time and space complexities of evaluating f at any $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ in both the dense and sparse cases. We also obtain, under the mild assumption that certain important matrices are readily available, an expression of the complexities of computing the gradient of f at any $(L^{\mathcal{I}}, y)$ in the dense and sparse cases. We will briefly mention the complexity issue for the Hessian calculation at the end.

We adopt the same convention as in Golub and Van Loan [5] for counting flops, that is a flop is a floating point operation (e.g., the inner product of two n-vectors involves $2n$ flops).

5.1 The dense case

The efficient computation of $f(L^{\mathcal{I}}, y)$ for any $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ can be arranged in the following steps:

(1) compute $H(y)$;

(2) use $H(y)$ to compute $L^{\mathcal{J}} = L^{\mathcal{J}}(L^{\mathcal{I}}, y)$;

(3) use $H(y)$ and $L^{\mathcal{J}}$ to compute $Z^{\mathcal{I}} = Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$;

(4) compute $g(Z^{\mathcal{I}}, y)$, which equals $f(L^{\mathcal{I}}, y)$.

We first consider the number of flops required to complete steps (1) through (4), and for this, we introduce two technical definitions. For any function $\xi : V \to W$ between two sets V and W, and for each $v \in V$, we denote by $\mathrm{fps}(\xi(v))$ the number of flops required to compute $\xi(v) \in W$. Moreover, we define $\mathrm{fps}(\xi) \equiv \sup\{\mathrm{fps}(\xi(v)) : v \in V\}$. We assume that both $\mathrm{fps}(g)$ and $\mathrm{fps}(H)$ are finite.

Proposition 5.1 *For each $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, the computation of $f(L^{\mathcal{I}}, y)$ takes at most $\mathcal{O}((|\mathcal{I}| + |\mathcal{J}|)n + \mathrm{fps}(g) + \mathrm{fps}(H)) = \mathcal{O}(n^3 + \mathrm{fps}(g) + \mathrm{fps}(H))$ flops.*

Proof. Steps (1) and (4) together take no more than $\mathrm{fps}(H)$+$\mathrm{fps}(g)$ flops. A straightforward analysis of the proof of Lemma 3.1 shows that step (2) takes at most $\mathcal{O}(|\mathcal{J}|n)$ flops, and it is easy to see that step (3) takes at most $\mathcal{O}(|\mathcal{I}|n)$ flops. Hence, the computation of $f(L^{\mathcal{I}}, y)$ takes at most $\mathcal{O}((|\mathcal{I}| + |\mathcal{J}|)n + \mathrm{fps}(g) + \mathrm{fps}(H))$ flops, and the equality of the proposition follows from the fact that $|\mathcal{I}| + |\mathcal{J}| = n(n+1)/2$. ∎

To consider the space requirements of the evaluation of f, let ξ be as above. Then for all $v \in V$, we denote by $\mathrm{spc}(\xi(v))$ the space required to store v and to compute and store $\xi(v) \in W$. We also define $\mathrm{spc}(\xi) \equiv \sup\{\mathrm{spc}(\xi(v)) : v \in V\}$. As with the function fps, we assume that $\mathrm{spc}(g)$ and $\mathrm{spc}(H)$ are finite.

Proposition 5.2 *For each $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, the computation of $f(L^{\mathcal{I}}, y)$ requires at most $\mathcal{O}(|\mathcal{I}| + |\mathcal{J}| + spc(g) + spc(H)) = \mathcal{O}(n^2 + spc(g) + spc(H))$ space.*

Proof. Steps (1) and (4) together require no more than $\mathrm{spc}(H)$+$\mathrm{spc}(g)$ space, and assuming that $L^{\mathcal{I}}$ and $H(z)$ are easily accessible in the compuer's memory, step (2) takes no more than $|\mathcal{J}|$ units of additional space. Similarly, $|\mathcal{I}|$ space is needed for step (3). Hence, the result of the proposition follows from the equality $|\mathcal{I}| + |\mathcal{J}| = n(n+1)/2$. ∎

We remark that, in the dense case, $\mathrm{spc}(H)$ is at least $\mathcal{O}(n^2)$. Hence, the additonal space requirements incurred due to the transformation from (NLSDP) to (NLP) are not significantly greater than the requirements for evaluating g.

We now turn to the computation of the gradient of f at any $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and for this, we assume that $L \equiv L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ and $Z^{\mathcal{I}} \equiv Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$ have already been computed and stored. Using Corollary 4.4, we can arrange the computation of the gradient of f at $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ as follows:

(1) use $Z^{\mathcal{I}}$ and y to compute $\nabla_{Z^{\mathcal{I}}}\, g$, $\nabla_y\, g$ and $\partial H/\partial y_k$ for $k = 1, \ldots, m$;

(2) let $\pi_{\mathcal{I}}(X) = \nabla_{Z^{\mathcal{I}}}\, g$, and then use L to compute $\pi_{\mathcal{J}}(X)$ via (20);

(3) use X and L to compute $\nabla_{L^{\mathcal{I}}}\, f(L^{\mathcal{I}}, y) = 2\pi^l_{\mathcal{I}}(XL)$;

(4) use $\nabla_y\, g$, ∇H, and X to compute $\nabla_y\, f(L^{\mathcal{I}}, y) = \nabla_y\, g - \nabla H(X)$,

where the operation $\nabla H(X)$ is defined in (17).

We define

$$\text{fps}(\nabla g) \equiv \text{fps}(\nabla_{Z^{\mathcal{I}}}\, g) + \text{fps}(\nabla_y\, g) \quad \text{and} \quad \text{fps}(\nabla H) \equiv \sum_{k=1}^{m} \text{fps}\left(\frac{\partial H}{\partial y_k}\right)$$

and also define $\text{spc}(\nabla g)$ and $\text{spc}(\nabla H)$ similarly. In addition, we define

$$\text{nz}(\nabla H) \equiv \max\left\{\sum_{k=1}^{m} \text{nz}\left(\frac{\partial H(y)}{\partial y_k}\right) : y \in \Re^m\right\},$$

where, for each $M \in \mathcal{S}^n$, $\text{nz}(M)$ is the number of nonzero entries in the lower triangular part of M. Note that $\text{nz}(\nabla H) = \mathcal{O}(mn^2)$ in the dense case and also that $\text{nz}(\nabla H)$ differs from $\text{spc}(\nabla H)$ in that $\text{nz}(\nabla H)$ only includes the amount of space needed to store the results of ∇H, i.e., it does not include the space needed to compute ∇H. We have the following proposition.

Proposition 5.3 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, and suppose that $L^{\mathcal{I}} + L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ and $Z^{\mathcal{I}}(L^{\mathcal{I}}, y)$ have already been computed and stored. Then the number of flops required for the the computation of $\nabla f(L^{\mathcal{I}}, y)$ is at most*

$$\begin{aligned}&\mathcal{O}\left((|\mathcal{I}| + |\mathcal{J}|)n + nz(\nabla H) + \text{fps}(\nabla g) + \text{fps}(\nabla H)\right)\\ &= \mathcal{O}(n^3 + mn^2 + \text{fps}(\nabla g) + \text{fps}(\nabla H)),\end{aligned}$$

and the amount of space required is at most

$$\mathcal{O}\left(|\mathcal{I}| + |\mathcal{J}| + spc(\nabla g) + spc(\nabla H)\right) = \mathcal{O}(n^2 + spc(\nabla g) + spc(\nabla H))$$

Proof. We first address the number of flops required to compute the gradient. Step (1) clearly takes no more than $\text{fps}(\nabla g) + \text{fps}(\nabla H)$ flops, and it is easy to see using the recursive formula for X described in (20) that step (2) takes at most $\mathcal{O}(|\mathcal{J}|n)$ flops. Step (3) requires $\mathcal{O}(|\mathcal{I}|n)$ flops, and step (4) takes at most $\mathrm{nz}(\nabla H)$ flops, as can be seen from Corollary 4.4. Noting that $|\mathcal{I}| + |\mathcal{J}| = n(n+1)/2$ and that, in the dense case, $\mathrm{nz}(\nabla H) = \mathcal{O}(mn^2)$, we obtain the first result of the proposition.

We now establish the space requirements of the gradient computation. Step (1) requires $\mathrm{spc}(\nabla g) + \mathrm{spc}(\nabla H)$ space, and steps (2) and (3) need $|\mathcal{I}| + |\mathcal{J}|$ units of additional space. Finally, step (4) takes no additional space, since the computation of $\nabla_y f(L^{\mathcal{I}}, y)$ simply involves m matrix dot products involving matrices already computed and stored and since the contents of $\nabla_y g$ can clearly be overwritten with $\nabla_y f(L^{\mathcal{I}}, y)$. Noting once again that $|\mathcal{I}| + |\mathcal{J}| = n(n+1)/2$, we have the second result of the proposition. ∎

5.2 The sparse case

We first note that the sparsity characteristics of the feasible set $\mathcal{F}_{sdp}$ of (NLSDP) are directly derived from the sparsity characteristics of $H(\Re^m) \subseteq \mathcal{S}^n$, that is, the image of $\Re^m$ under the map H. These sparsity characteristics, which we will loosely refer to as the sparsity of H, have obvious consequences for the values $\mathrm{fps}(H)$ and $\mathrm{fps}(\nabla H)$ introduced in the previous subsection, but the sparsity of H also plays a significant role in determining the sparsity characteristics of $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. In what follows, we relate the sparsity of H to the time and space complexities of the function and gradient evaluations.

Via the Cholesky factorization , it is well-known that, in general, the greater the sparsity of $H(y)$ is, the greater the sparsity of $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$ will be. For each $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, define $\mathcal{Z}(L^{\mathcal{I}}, y) \subseteq \mathcal{J}$ to be the index subset of $\mathcal{J}$ of zero entries of $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$, i.e.,

$$\mathcal{Z}(L^{\mathcal{I}}, y) \equiv \{(i,j) \in \mathcal{J} : [L^{\mathcal{J}}(L^{\mathcal{I}}, y)]_{ij} = 0\},$$

and define

$$\begin{aligned} \mathcal{Z} &\equiv \cap\{\mathcal{Z}(L^{\mathcal{I}}, y) : (L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m\}, \\ \widehat{\mathcal{J}} &\equiv \mathcal{J} \setminus \mathcal{Z}. \end{aligned}$$

We now define a new function $L^{\widehat{\mathcal{J}}} : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \mathcal{L}^{\widehat{\mathcal{J}}}$ implicitly by $L^{\widehat{\mathcal{J}}} \equiv L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$, where $L^{\widehat{\mathcal{J}}} \in \mathcal{L}^{\widehat{\mathcal{J}}}$ is the unique solution of the equation

$$\pi_{\widehat{\mathcal{J}}}((L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}})(L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}})^T) = \pi_{\widehat{\mathcal{J}}}(H(y)).$$

(The uniqueness of $L^{\widehat{\mathcal{J}}}$ follows from an argument similar to proof of Lemma 3.1.) The following proposition shows that the operators $L^{\widehat{\mathcal{J}}}$ and $L^{\mathcal{J}}$ are essentially identical.

Proposition 5.4 *For all $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, $L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y) = L^{\mathcal{J}}(L^{\mathcal{I}}, y)$.*

Proof. Let $L^{\mathcal{J}} \equiv L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. The definition of $\mathcal{Z}$ implies that $L^{\mathcal{J}}_{ij} = 0$ for all $(i,j) \in \mathcal{Z}$, and so $L^{\mathcal{J}} \in \mathcal{L}^{\widehat{\mathcal{J}}}$. Since $L^{\mathcal{J}}$ satisfies (9) (with $H = H(y)$) and since $\widehat{\mathcal{J}} \subseteq \mathcal{J}$, we have that

$$\pi_{\widehat{\mathcal{J}}}((L^{\mathcal{I}} + L^{\mathcal{J}})(L^{\mathcal{I}} + L^{\mathcal{J}})^T) = \pi_{\widehat{\mathcal{J}}}(H(y)).$$

It thus follows from the definition of $L^{\widehat{\mathcal{J}}}$ that $L^{\mathcal{J}} = L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$. ∎

It now follows immediately from Proposition 5.4 that an equivalent formulation of (NLSDP) is formed by simply substituting $L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$ for $L^{\mathcal{J}}(L^{\mathcal{I}}, y)$. In particular, we have the following definitions and results which reflect this replacement.

Definition 5.5 *The function $\widehat{\Psi} : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \mathcal{S}^{\mathcal{I}} \times \Re^m$ is defined by the equation $\widehat{\Psi}(L^{\mathcal{I}}, y) \equiv (\widehat{Z}^{\mathcal{I}}(L^{\mathcal{I}}, y), y)$, where the function $\widehat{Z}^{\mathcal{I}} : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \mathcal{S}^{\mathcal{I}}$ is given by*

$$\widehat{Z}^{\mathcal{I}}(L^{\mathcal{I}}, y) \equiv (L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y))(L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y))^T - H(y).$$

In addition, the function $\widehat{f} : \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m \to \Re$ is defined by $\widehat{f}(L^{\mathcal{I}}, y) \equiv g(\widehat{\Psi}(L^{\mathcal{I}}, y))$.

Theorem 5.6 *Proposition 5.4 implies that $\widehat{\Psi} = \Psi$, $\widehat{Z}^{\mathcal{I}} = Z^{\mathcal{I}}$ and $\widehat{f} = f$. Hence, if the function g is continuous and attains it minimum in $\mathcal{F}_{sdp}$, then*

$$\min\{g(Z^{\mathcal{I}}, y) : (Z^{\mathcal{I}}, y) \in \mathcal{F}_{sdp}\} = \inf\{\widehat{f}(L^{\mathcal{I}}, y) : (L^{\mathcal{I}}, y) \in \mathcal{F}_{nlp}\}.$$

Theorem 5.7 *For any $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$ and $\nabla_{Z^{\mathcal{I}}}\, g \in \mathcal{S}^{\mathcal{I}}$, let $L \equiv L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$. Then there exists a matrix $\widehat{X} \equiv \widehat{X}(L^I, y)$ which is the unique solution in $\mathcal{S}^{\mathcal{I} \cup \widehat{\mathcal{J}}}$ of the equations:*

$$\begin{aligned} \pi_{\mathcal{I}}(\widehat{X}) &= \nabla_{Z^{\mathcal{I}}}\, g, \\ \pi^l_{\widehat{\mathcal{J}}}(\widehat{X} L) &= 0. \end{aligned}$$

Theorem 5.8 *Let $(L^{\mathcal{I}}, y) \in \mathcal{L}^{\mathcal{I}}_{++} \times \Re^m$, $L \equiv L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$ and $\widehat{X} \equiv \widehat{X}(L^{\mathcal{I}}, y)$. Then*

$$\begin{aligned} \nabla_{L^{\mathcal{I}}} \widehat{f}(L^{\mathcal{I}}, y) &= 2\, \pi^l_{\mathcal{I}}(\widehat{X} L), \\ \nabla_y \widehat{f}(L^{\mathcal{I}}, y) &= \nabla_y\, g - \nabla H(\widehat{X}). \end{aligned}$$

We have omitted the proofs of Theorems 5.6, 5.7 and 5.8 since their proofs are very similar to the corresponding results in the previous section, but a few comments regarding the relationship between $X(L^I, y)$ and $\widehat{X}(L^{\mathcal{I}}, y)$ are in order. First of all, these two functions are not equal. The defining equations for X found in Theorem 4.5 ensure that $X(L^I, y)$ is, in general, a dense symmetric matrix, while the equations in Theorem 5.7 guarantee that $\widehat{X}(L^{\mathcal{I}}, y)$ has the same sparsity pattern as $L \equiv L^{\mathcal{I}} + L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}}, y)$. Secondly, it is not difficult to see that, like $X(L^{\mathcal{I}}, y)$, $\widehat{X}(L^{\mathcal{I}}, y)$ can be computed by a recursive formula similar to (20), but for $\widehat{X}$, the computation can be done in $\mathcal{O}((|\mathcal{I}|+|\widehat{\mathcal{J}}|)n)$ flops and in $\mathcal{O}(|\mathcal{I}|+|\widehat{\mathcal{J}}|)$ space. Finally, the equality $\widehat{f}=f$ shows that the gradient of f can be computed by either using $X(L^{\mathcal{I}},y)$ or $\widehat{X}(L^{\mathcal{I}},y)$.

We are now prepared to state the time and space complexities for the function and gradient evaluations in the sparse case.

Proposition 5.9 *For each $(L^{\mathcal{I}},y)\in\mathcal{L}^{\mathcal{I}}_{++}\times\Re^m$, assume that $L^{\mathcal{I}}+L^{\widehat{\mathcal{J}}}(L^{\mathcal{I}},y)$ and $\widehat{Z}^{\mathcal{I}}(L^{\mathcal{I}},y)$ have already been computed. Then the numbers of flops required for the computation of $\widehat{f}(L^{\mathcal{I}},y)$ and $\nabla\widehat{f}(L^{\mathcal{I}},y)$ are, respectively, at most*

- $\mathcal{O}((|\mathcal{I}|+|\widehat{\mathcal{J}}|)n+\mathrm{fps}(g)+\mathrm{fps}(H))$, *and*
- $\mathcal{O}((|\mathcal{I}|+|\widehat{\mathcal{J}}|)n+nz(\nabla H)+\mathrm{fps}(\nabla g)+\mathrm{fps}(\nabla H))$.

In addition, the amount of space required for these operations is, respectively, at most

- $\mathcal{O}(|\mathcal{I}|+|\widehat{\mathcal{J}}|+spc(g)+spc(H))$, *and*
- $\mathcal{O}(|\mathcal{I}|+|\widehat{\mathcal{J}}|+spc(\nabla g)+spc(\nabla H))$.

Proof. In the previous subsection, we showed that the evaluations of the function and gradient took $\mathcal{O}((|\mathcal{I}|+|\mathcal{J}|)n+\mathrm{fps}(g)+\mathrm{fps}(H))$ and $\mathcal{O}((|\mathcal{I}|+|\mathcal{J}|)n+\mathrm{nz}(\nabla H)+\mathrm{fps}(\nabla g)+\mathrm{fps}(\nabla H)$ flops, respectively, and that they required $\mathcal{O}(|\mathcal{I}|+|\mathcal{J}|+\mathrm{spc}(g)+\mathrm{spc}(H))$ and $\mathcal{O}(|\mathcal{I}|+|\mathcal{J}|+\mathrm{spc}(\nabla g)+\mathrm{spc}(\nabla H))$, respectively. In this subsection, the same arguments are valid, except that $|\mathcal{J}|$ is replaced by $|\widehat{\mathcal{J}}|$. Hence, the proposition follows. ∎

We remark that, in many special cases (e.g., the case of the linear SDP), the set $\mathcal{Z}$ defined above can be determined in polynomial time using the well-known symbolic factorization procedure. In addition, a reordering procedure can be performed in order to reformulate (NLSDP) so that the resultant size of $\mathcal{Z}$ is usually larger than the original size of $\mathcal{Z}$.

5.3 Complexity for the linear SDP

In the case of the linear SDP (Example 2.1), where $g(Z^{\mathcal{I}},y)=B^{\mathcal{I}}\bullet Z^{\mathcal{I}}+b^Ty$ and $H(y)=\sum_{k=1}^m y_kA_k-C$, we clearly have $\mathrm{fps}(\nabla g)=\mathrm{fps}(\nabla H)=0$ due to the linearity of g and H. Moreover, $\mathrm{fps}(g)=\mathcal{O}(|\mathcal{I}|+m)$, $\mathrm{fps}(H)=\mathcal{O}(\mathrm{nz}(A_1,\ldots,A_m,C))$ and $\mathrm{nz}(\nabla H)=\mathcal{O}(\mathrm{nz}(A_1,\ldots,A_m))$, where $\mathrm{nz}(\cdot)$ here represents the total number of nonzero elements of the given matrices. Notice the only difference between $\mathrm{fps}(H)$ and $\mathrm{nz}(\nabla H)$ is the appearance of the constant matrix C in $\mathrm{fps}(H)$.

Using these facts and the inequalities $n\le|\mathcal{I}|$ and $m\le O(n^2)$, Proposition 5.9 implies that the time complexities for the gradient and function evaluations are, respectively,

$$\mathcal{O}(n\,\mathrm{nz}(L)+\mathrm{nz}(A_1,\ldots,A_m,C)),$$
$$\mathcal{O}(n\,\mathrm{nz}(L)+\mathrm{nz}(A_1,\ldots,A_m)),$$

where $\mathrm{nz}(L) = |\mathcal{I}| + |\widehat{\mathcal{J}}|$ denotes the number of nonzeros in the lower Cholesky factor L.

In Table 1, we specialize the complexities above to the three cases: (i) both L and $\{A_1, \ldots, A_m, C\}$ are sparse; more specifically, we assume that $\mathrm{nz}(A_1, \ldots, A_m, C) \le \mathcal{O}(n^2)$; (ii) L is dense but $\{A_1, \ldots, A_m, C\}$ is relatively sparse whenever $m \gg n$; more specifically, we assume that $\mathrm{nz}(A_1, \ldots, A_m, C) \le \mathcal{O}(n^3)$; (iii) both L and $\{A_1, \ldots, A_m, C\}$ are dense. (Note that the fourth case never occurs, i.e., if $\{A_1, \ldots, A_m, C\}$ is dense, then L cannot be generally sparse.)

Table 1: Number of Flops Required for f and ∇f Evaluations

Sparse Case	Mixed Case	Dense Case
$\mathcal{O}(n\,\mathrm{nz}(L))$	$\mathcal{O}(n^3)$	$\mathcal{O}((n+m)n^2)$

The space requirements for evaluating f and ∇f can be considered similarly. Noting that $\mathrm{spc}(\nabla g) = \mathrm{spc}(\nabla H) = 0$ and that $\mathrm{spc}(g) = \mathcal{O}(|\mathcal{I}| + m)$, Proposition 5.9 shows that the space complexities for the gradient and function evaluations are, respectively,

$$\mathcal{O}(\mathrm{nz}(L) + \mathrm{spc}(H)),$$
$$\mathcal{O}(\mathrm{nz}(L)).$$

If we make the reasonable assumption that the matrices $\{A_1, \ldots, A_m, C\}$ are stored in memory for use throughout all iterations, then it is easy to see that H can be computed in the space of the nonzeros of the slack matrix S, i.e.,

$$\mathrm{spc}(H) = \mathcal{O}(\mathrm{nz}(S)) \le |\mathcal{I}| + |\mathcal{Z}(A_1) \cup \cdots \cup \mathcal{Z}(A_m) \cup \mathcal{Z}(C)|,$$

where, for $M \in \mathcal{S}^n$, $\mathcal{Z}(M)$ here is the index subset of nonzero entries of M. It is clear, however, that $\mathrm{nz}(L)$ is an upperbound on $\mathrm{nz}(S)$, and so we can further refine the above space complexity results. The resulting complexities for the sparse and dense cases are given in the Table 2.

Table 2: Space Required for f and ∇f Evaluations

Sparse Case	Dense Case
$\mathcal{O}(\mathrm{nz}(L))$	$\mathcal{O}(n^2)$

5.4 The Hessian calculation

For general cases, the computational complexity of evaluating the Hessian of $f(L^{\mathcal{I}}, y)$ is closely tied with the functional forms of $g(Z^{\mathcal{I}}, y)$ and $H(y)$. To simplify the discussion, we consider only the case of the dual linear SDP (5) where the first derivatives of g and H are constant matrices. Furthermore, we only consider the dense case where all matrices involved are completely dense.

A straightforward but tedious verification reveals that the computational complexity of evaluating the Hessian matrix using the expressions developed in Section 4 is at most $\mathcal{O}(n^4 + mn^3)$ flops. For large-scale problems, this cost could be prohibitively high. On the other hand, the cost of computing the product of the Hessian matrix times a vector is at most $\mathcal{O}(n^3 + mn^2)$ flops, which is of the same order as a gradient evaluation.

6 Final Remarks

In the case of linear SDPs, the new formulation operates in the interior of the dual feasibility set. Hence, it always maintains a bound on the primal objective function. This feature is important in the applications of SDPs in solving combinatorial optimization problems.

Since $LL^T \equiv (LQ)(LQ)^T$ for orthogonal matrices Q including diagonal matrices with plus and/or minus ones on the diagonal, the restriction $L^{\mathcal{I}} \in \mathcal{L}^{\mathcal{I}}_{++}$ on our transformations is more of theoretical convenience than practical necessity. The transformations are well defined throughout $\mathcal{L}^{\mathcal{I}}$ except on the boundary of $\mathcal{L}^{\mathcal{I}}_{++}$ where the optimal solutions reside in the case of linear SDPs.

One can apply our transformations to more general SDP problems than the standard form (2), though the resulting nonlinear optimization problems are no longer as simple as (12) is. For example, consider the dual linear SDP in the standard form

$$\min \left\{ b^T y : \sum_{k=1}^{m} y_k A_k - C \succeq 0, \ y \in \Re^m \right\}, \tag{37}$$

where $C, A_1, \ldots, A_m \in \mathcal{S}^n$ and $b \in \Re^m$ are the data. Without changing the optimal solution set we can add an artificial variable $Z^{\mathcal{D}} \in \mathcal{S}^{\mathcal{D}}$ to obtain the following equivalent problem

$$\min \left\{ b^T y : Z^{\mathcal{D}} + \sum_{k=1}^{m} y_k A_k - C \succeq 0, \ Z^{\mathcal{D}} \in \mathcal{S}^{\mathcal{D}}, \ y \in \Re^m, \ Z^{\mathcal{D}} \leq 0 \right\},$$

where by convention $Z^D \leq 0$ means that each (diagonal) entry of Z^D is non-positive. This new problem is strictly feasible as long as the original problem (37) is. It is almost in the standard form (2) except for the non-positivity constraint on $Z^{\mathcal{D}}$. Applying our transformations to this problem, we obtain the following nonlinear optimization problem

$$\inf \left\{ b^T y : Z^{\mathcal{D}}(L^{\mathcal{D}}, y) \leq 0, \ (L^{\mathcal{D}}, y) \in \mathcal{L}^{\mathcal{D}}_{++} \times \Re^m \right\},$$

which has $n + m$ variables and n nonlinear inequality constraints. We also note that the inequalities can be equivalently replaced by equalities without changing the optimal solution set. Another approach to converting (37) into the standard form (2) is by adding an artificial variable $Z^{\mathcal{D}}$ to the constraint along with an exact penalty term in the objective. For example, the following convex program

$$\min \left\{ b^T y + \rho \|\mathrm{diag}(Z^{\mathcal{D}})\|_1 : Z^{\mathcal{D}} + \sum_{k=1}^{m} y_k A_k - C \succeq 0;\ Z^{\mathcal{D}} \in \mathcal{S}^{\mathcal{D}}, y \in \Re^m \right\}.$$

is in the standard form (2) and is equivalent to (37) for ρ sufficiently large, though the objective function is non-differentiable at any $Z^{\mathcal{D}}$ with a zero component.

The present paper is the first part of a study on the proposed transformations and formulations for solving large-scale semidefinite programs. Further theoretical, algorithmic and computational results will be reported in subsequent papers.

References

[1] S. Burer and R. D. C. Monteiro. An Efficient Algorithm for Solving the MAXCUT SDP Relaxation. Working paper, School of ISyE, Georgia Tech., USA, December 1998.

[2] S. Benson, Y. Ye, and X. Zhang. Solving large-scale sparse semidefinite programs for combinatorial optimization. Research Report, Department of Management Science, University of Iowa, Iowa, 1998.

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

[4] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of ACM*, pages 1115–1145, 42 (1995).

[5] G. H. Golub and C. E. Van Loan. *Matrix Computations: Second Edition.* The Johns Hopkins University Press, Baltimore, MD 21211, 1989.

[6] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. Manuscript, Konrad-Zuse-Zentrum fuer Informationstechnik Berlin, Takustrasse 7, D-14195 Berlin, Germany, 1997.

[7] S. Homer and M. Peinado. Design and performance of parallel and distributed approximation algorithms for maxcut. manuscript, Dept. of Computer Science and Center for Computational Science, Boston University, 111 Cummington Street, Boston, MA, 02215, USA, 1995.

[8] M. Laurent, S. Poljak, and F. Rendl. Connections between semidefinite relaxations of the max-cut and stable set problems. *Mathematical Programming*, 77:225–246, 1997.

[9] L. Lovász. On the Shannon Capacity of a graph. *IEEE Transactions of Information Theory*, IT-25(1):1–7, January 1979.

[10] L. Lovász and A. Schrijver. Cones of matrices and set functions, and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[11] Y. Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software*, 9:141–160, 1998.

[12] S. Poljak, F. Rendl, and H. Wolkowicz. A recipe for semidefinite relaxation for 0-1 quadratic programming. *Journal of Global Optimization*, 7:51–73, 1995.

[13] S. Poljak and Z. Tuza. The max-cut problem — a survey. In W. Cook, L. Lovász, and P. Seymour, editors, *Special Year on Combinatorial Optimization*, DIMACS series in Discrete Mathematics and Theoretical Computer Science. American Mathematical Society. To be published.

[14] N. Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Science*, 25:1–11, 1987. Originally published in Tekhnicheskaya Kibernetika, No. 1, 1987, pp. 128–139.