

Conditional Quadratic Semidefinite Programming: Examples and Methods

Hou-Duo Qi*

April 14, 2014

Abstract

The conditional Quadratic Semidefinite Programming (cQSDP) refers to a class of matrix optimization problems whose matrix variables are required to be positive semidefinite on a subspace and the objectives are quadratic. The chief purpose of this paper is to focus on two primal examples of cQSDP: the problem of matrix completion/approximation on a subspace and the Euclidean distance matrix problem. For the latter problem, we review some classical contributions and establish certain links among them. Moreover, we develop a semismooth Newton method for a special class of cQSDP and establish its quadratic convergence under the condition of constraint nondegeneracy. We also include an application in calibrating the correlation matrix in Libor market models. We hope this work will stimulate new research in cQSDP.

AMS subject classifications. 49M45, 90C25, 90C33

Keywords: Matrix optimization, conditional semidefinite programming, Euclidean distance matrix, semismooth Newton method.

1 Introduction

The chief purpose of this paper is to use as an umbrella the conditional Quadratic Semi-Definite Programming (cQSDP) to include some existing applications that are close to the well-known Semi-Definite Programming (SDP) and to stimulate new research in cQSDP. We gather some examples from existing literature and show that they naturally appear as cQSDP. We cast those examples either as of a matrix completion/approximation problem on a subspace or as of an Euclidean distance matrix problem. In particular, we try our best to trace the root how those problems are originally studied and establish some general results for them. We also develop a semismooth Newton method for a special class of cQSDP, which plays a fundamental role in a wider class of cQSDP. Moreover, we include an interesting application of cQSDP arising from calibrating the correlation matrix in Libor market model. The main message in this paper is that it is worth and often necessary to design new algorithms in order to take advantage of problem structures in cQSDP.

As indicated by its name, cQSDP is very close to SDP. The key departure is that cQSDP requires the positive semidefiniteness in a subspace, in comparison to SDP which requires the positive semidefiniteness in the whole space. We make this difference precise now. Let \mathcal{S}^n denote the space of $n \times n$ symmetric matrices with the standard inner product $\langle \cdot, \cdot \rangle$, whose induced norm is the Frobenius norm $\|X\|$ for $X \in \mathcal{S}^n$. Let \mathcal{S}_+^n denote the cone of all positive semidefinite

*School of Mathematics, The University of Southampton, Highfield, Southampton SO17 1BJ, UK. E-mail: hdqi@soton.ac.uk. This research was supported by the EPSRC grant EP/K007645/1.

matrices in \mathcal{S}^n . Let $\mathcal{V} \subset \mathbb{R}^n$ be a subspace. We say that $X \in \mathcal{S}^n$ is positive semidefinite in \mathcal{V} (write as $X \succeq 0$ in \mathcal{V}) if

$$v^T X v \geq 0 \quad \forall v \in \mathcal{V}.$$

We often say that such a matrix is conditionally positive semidefinite without explicitly referring to the subspace it is associated with. We define

$$\mathcal{K}_+^n(\mathcal{V}) := \{X \in \mathcal{S}^n \mid X \succeq 0 \text{ in } \mathcal{V}\}.$$

Therefore, cQSDP refers to any problem that can be cast in the following form:

$$\begin{aligned} \min \quad & f(X) := \langle X, \mathcal{Q}(X) \rangle + \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}(X) \in b + \mathcal{P}, \\ & X \in \mathcal{K}_+^n(\mathcal{V}), \end{aligned} \tag{1}$$

where $\mathcal{Q} : \mathcal{S}^n \mapsto \mathcal{S}^n$ is a linear mapping; $C \in \mathcal{S}^n$; $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^m$ is a linear mapping and $b \in \mathbb{R}^m$; and $\mathcal{P} = \{0\}^p \times \mathbb{R}_+^q$ with $m := p + q$. Since $\mathcal{K}_+^n(\mathcal{V})$ is a closed convex cone, cQSDP (1) is of conic optimization with special structures. For easy reference, we call $\mathcal{K}_+^n(\mathcal{V})$ the conditional positive semidefinite cone.

cQSDP can be quickly reformulated as QSDP (see Toh [44]). Let $\Pi_{\mathcal{V}}(x)$ denote the orthogonal projection of x onto \mathcal{V} . That is, $\Pi_{\mathcal{V}}(x)$ is the optimal solution of the following problem:

$$\min_y \|x - y\| \quad \text{s.t. } y \in \mathcal{V},$$

where $\|y\|$ is the Euclidean norm of $y \in \mathbb{R}^n$. Since \mathcal{V} is a subspace, $\Pi_{\mathcal{V}}(\cdot)$ can be thought of an orthogonal matrix. Then, $X \in \mathcal{K}_+^n(\mathcal{V})$ if and only if $\Pi_{\mathcal{V}} X \Pi_{\mathcal{V}} \in \mathcal{S}_+^n$. One serious drawback of this reformulation is that the matrix $\Pi_{\mathcal{V}} X \Pi_{\mathcal{V}}$ always has zero as its eigenvalue and hence the generalized Slater condition is never satisfied. Furthermore, the linear transformation $\Pi_{\mathcal{V}} X \Pi_{\mathcal{V}}$ often complicates the use of off-shelf SDP solvers and may significantly slow them down. In contrast, the generalized Slater condition is often satisfied in cQSDP for some of our examples. Fortunately, it is possible in the examples encountered below to handle the conditional positive semidefinite cone $\mathcal{K}_+^n(\mathcal{V})$ directly.

The concept of conditional (positive or negative) semidefiniteness has been used in slightly different situations before. For example, in the survey paper [24], Ikramov and Savel'eva used it for \mathcal{V} being not only a subspace but also a closed convex cone. Reams [38] used it for $\mathcal{V} = e^\perp$, the subspace orthogonal to the vector e of all ones. Earlier, Chabrilac and Crouzeix in their survey paper [7] call it the semidefiniteness of the restricted quadratic form to the space \mathcal{V} . Micchelli [29] (see also Baxter [3]) used a slightly different name for $\mathcal{V} = e^\perp$. They referred it as the almost positive definiteness, whose origin can be traced back to the seminar work [42] of Schoenberg on the Euclidean distance embeddability in the (real) Hilbert space. We will say more about Schoenberg's contribution when it comes to the Euclidean distance matrix problem. In this paper, the conditional semidefiniteness refers to the case where \mathcal{V} is a subspace. The primary reason for this reference is that efficient numerical methods such as the semismooth Newton-CG method can be readily developed. The second reason is that this conditional semidefiniteness has played an important role in a number of important applications, which are to be explained to a considerable extent accessible to a wide audience in the hope of stimulating new research. However, we will not review the vast literature on one of the most important applications to the second-order optimality conditions in the classical constrained optimization because they have been well documented, see e.g., [30, Sect. 12.4] and [4, Sect. 13.4] and the relevant references therein.

The paper is organized as follows. Section 2 contains two examples of matrix completion (approximation) on a subspace. We derive two closed-form formulas (see Thm. 2.1) of the

orthogonal projection onto the conditional positive semidefinite cone. A consequence is that the projection, viewed as a function of its argument, is strongly semismooth. This result paves the way for the semismooth Newton method to be developed. Section 3 is about problems of Euclidean Distance Matrix (EDM). We survey four important and earliest contributions, which are summarized in Thm. 3.2. We then describe two nearest EDM problems: the classical Multi-Dimensional Scaling (cMDS) and the direct MDS. We establish a new equivalence result for cMDS in Thm. 3.3. In Section 4, we develop a semismooth Newton-CG method for a special class of cQSDP and establish its quadratic convergence under the assumption of constraint nondegeneracy. In Section 5, We present a new and an important application in calibrating the correlation matrix in Libor market models and report some numerical results. We conclude the paper in Section 6.

Notation: Apart from the notations used above, we also need the following (“:=” means “define”). We write $X \succ 0$ for $X \in \mathcal{S}^n$ being positive definite. $\text{Tr}(X)$ is the trace of X . e is the vector of all ones in an Euclidean space, whose dimension is clear from the context, and e_i is the i th unit vector. $A \circ B := [A_{ij}B_{ij}]$ is the Hadamard product between two matrices A and B of same size. For subsets α, β of $\{1, \dots, n\}$, $B_{\alpha\beta}$ denotes the submatrix of B indexed by α and β (α for rows and β for columns). B_α denotes the submatrix consisting of columns of B indexed by α , and $|\alpha|$ is the cardinality of α .

2 Matrix Approximation and Completion on a Subspace

Matrix approximation plays a central role in numerical linear algebra and appears in many applications. The most important one is probably the orthogonal approximation to a given matrix by a positive semidefinite matrix. For a given matrix $A \in \mathcal{S}^n$, let $\Pi_{\mathcal{S}_+^n}(A)$ be the orthogonal projection of A onto the convex cone \mathcal{S}_+^n . That is, $\Pi_{\mathcal{S}_+^n}(A)$ is the optimal solution of the following problem:

$$\min \|X - A\| \quad \text{s.t.} \quad X \in \mathcal{S}_+^n, \quad (2)$$

where $\|\cdot\|$ is the Frobenius norm induced by the standard inner product in \mathcal{S}^n . Let A have the following spectral decomposition:

$$A = P\Lambda P^T,$$

where $\Lambda := \text{Diag}(\lambda)$ is the diagonal matrix formed by the eigenvalue vector $\lambda = (\lambda_1, \dots, \lambda_n)^T$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of A in nonincreasing order; $P^T P = I$ (the columns of P consists of the normalized eigenvectors of X corresponding to λ). It is the classical result that problem (2) has the closed form solution:

$$\Pi_{\mathcal{S}_+^n}(A) = P\Lambda_+ P^T,$$

where $\Lambda_+ := \text{Diag}(\lambda_+)$ and

$$\lambda_+ := (\max(\lambda_1, 0), \dots, \max(\lambda_n, 0))^T.$$

It is important to emphasize that, as a function, the projection $\Pi_{\mathcal{S}_+^n}(\cdot)$ is not differentiable, but is almost everywhere differentiable. Moreover, it is strongly semismooth due to Sun and Sun [43] (see also Chen et al. [6]). The strong semismoothness of $\Pi_{\mathcal{S}_+^n}(\cdot)$ is a fundamental property behind semismooth Newton-CG methods for matrix optimization problems, see Qi and Sun [34] and Zhao et al. [49]. We will use one of the semismooth Newton-CG methods in our numerical part. Those nice properties of $\Pi_{\mathcal{S}_+^n}(\cdot)$ can be extended to $\Pi_{\mathcal{K}_+^n(\mathcal{V})}(\cdot)$, which appears in the following two generalized matrix completion/approximation problems.

2.1 Generalized Matrix Approximation

A generalized matrix approximation studied in this part seeks the best candidate to A from $\mathcal{K}_+^n(\mathcal{V})$, where \mathcal{V} is a given subspace characterized by

$$\mathcal{V} = \{x \in \mathbb{R}^n \mid S^T x = 0\}, \quad (3)$$

where S is a $n \times r$ real matrix with $\text{rank}(S) = r$. In other words, the columns of S form a basis of the orthogonal subspace of \mathcal{V} (denoted by \mathcal{V}^\perp). The generalized matrix approximation can be stated as follows:

$$\min \|X - A\| \quad \text{s.t.} \quad X \in \mathcal{K}_+^n(\mathcal{V}). \quad (4)$$

This is a special case of cQSDP (1) with $f(X) = \|X - A\|$ and no linear constraints. Problem (4) was studied by Hayden and Wells [21]. Like the positive semidefiniteness case (2), it also has a closed form solution.

Let $Q \in \mathbb{R}^{n \times n}$ be the product of r Householder transformations such that

$$QS = \begin{bmatrix} 0 \\ L \end{bmatrix}, \quad (5)$$

where L is a $r \times r$ lower triangular nonsingular matrix (this can be done by the QR factorization of S). We know that $Q^T Q = I$. We have the following result.

Theorem 2.1 *Let $\bar{A} = QAQ^T$ and partition \bar{A} into the following blocks:*

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & \bar{A}_2 \\ \bar{A}_2^T & \bar{A}_0 \end{bmatrix} \quad \text{with} \quad \bar{A}_1 \in \mathcal{S}^{n-r}, \quad \bar{A}_0 \in \mathcal{S}^r, \quad \bar{A}_2 \in \mathbb{R}^{(n-r) \times r}.$$

Then, $\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A)$ can be calculated by

$$\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A) = Q^T \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-r}}(\bar{A}_1) & \bar{A}_2 \\ \bar{A}_2^T & \bar{A}_0 \end{bmatrix} Q \quad (6)$$

or by

$$\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A) = A + \Pi_{\mathcal{S}_+^n}(-J_S A J_S), \quad (7)$$

where

$$J_S := Q^T \begin{bmatrix} I_{n-r} & 0 \\ 0 & 0 \end{bmatrix} Q. \quad (8)$$

Proof. Formula (6) was proved in [21, Thm. 2.2]. Formula (7) just follows from formula (6) as we show below. We note that

$$J_S A J_S = Q^T \begin{bmatrix} \bar{A}_1 & 0 \\ 0 & 0 \end{bmatrix} Q.$$

Since Q is orthogonal, we have

$$\Pi_{\mathcal{S}_+^n}(-J_S A J_S) = Q^T \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-r}}(-\bar{A}_1) & 0 \\ 0 & 0 \end{bmatrix} Q.$$

We also note that $A = Q^T \bar{A} Q$. Therefore, we have

$$\begin{aligned}
A + \Pi_{\mathcal{S}_+^n}(-J_S A J_S) &= Q^T \begin{bmatrix} \bar{A}_1 + \Pi_{\mathcal{S}_+^{n-r}}(-\bar{A}_1) & \bar{A}_2 \\ \bar{A}_2^T & \bar{A}_0 \end{bmatrix} Q \\
&= Q^T \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-r}}(\bar{A}_1) & \bar{A}_2 \\ \bar{A}_2^T & \bar{A}_0 \end{bmatrix} Q \\
&= \Pi_{\mathcal{K}_+^n(\mathcal{V})}(A). \tag{by (6)}
\end{aligned}$$

□

An equivalent way to state (6) is the detailed description of $\mathcal{K}_+^n(\mathcal{V})$.

$$\mathcal{K}_+^n(\mathcal{V}) = \left\{ Q^T \begin{bmatrix} A_1 & A_2 \\ A_2^T & A_0 \end{bmatrix} Q \mid \begin{array}{l} A_1 \in \mathcal{S}_+^{n-r}, A_2 \in \mathbb{R}^{(n-r) \times r} \\ A_0 \in \mathcal{S}^r \end{array} \right\}. \tag{9}$$

The first consequence of formula (7) is a characterization of $\mathcal{K}_+^n(\mathcal{V})$. It is easy to see that $A \in \mathcal{K}_+^n(\mathcal{V})$ if and only if

$$\Pi_{\mathcal{S}_+^n}(-J_S A J_S) = 0,$$

which in turn implies $J_S A J_S \succeq 0$. Hence, we have

$$\mathcal{K}_+^n(\mathcal{V}) = \{X \in \mathcal{S}^n \mid J_S X J_S \succeq 0\}. \tag{10}$$

Another important consequence of formula (7) is that the projection $\Pi_{\mathcal{K}_+^n(\mathcal{V})}(\cdot)$ is strongly semismooth because it is a composition of linear functions and $\Pi_{\mathcal{S}_+^n}(\cdot)$, which has been known to be strongly semismooth. This consequence opens the possibility of developing a semismooth Newton method for problems that involve $\Pi_{\mathcal{K}_+^n(\mathcal{V})}(\cdot)$. One of such problems is the nearest Euclidean distance matrix problem to be described soon. The need of iterative algorithms comes from the fact that additional constraints are often added to problem (4) in order to increase its modelling power. One example is the positive semidefinite matrix completion on a subspace studied by Johnson and Smith [26].

2.2 Positive Semidefinite Matrix Completion on a Subspace

Let A be a $n \times n$ symmetric matrix with missing values. The known values are contained in the index set \mathcal{B} , where we assume $i \leq j$ if $(i, j) \in \mathcal{B}$. The upper part of A is defined by values b_{ij} , $(i, j) \in \mathcal{B}$.

$$A_{ij} = \begin{cases} b_{ij} & \text{if } (i, j) \in \mathcal{B} \\ \text{value missing} & \text{otherwise.} \end{cases}$$

For a given partial symmetric matrix A as above and a matrix $S \in \mathbb{R}^{n \times r}$, the problem in question is whether there exists $X \in \mathcal{K}_+^n(\mathcal{V})$ such that

$$X_{ij} = A_{ij} \quad \text{for } (i, j) \in \mathcal{B},$$

where \mathcal{V} is defined by (3). This problem is called the positive semidefinite matrix completion on \mathcal{V} [26]. It is without loss of generality to assume that all diagonal elements of A are known (i.e., $(i, i) \in \mathcal{B}$ for $i = 1, \dots, n$). It is because that any missing diagonal element can be made arbitrarily large so that the positive semidefiniteness can be reduced to that of the matrix by removing the column and the row associated with that diagonal element.

We use a simple example to illustrate this problem. Let $A \in \mathcal{S}^3$ have just one missing value as given below.

$$A = \begin{bmatrix} 1 & 2 & x \\ 2 & 0 & -1 \\ x & -1 & 1 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

It is easy to verify that any matrix A with $x \leq 4$ belongs to $\mathcal{K}_+^3(\mathcal{V})$. Therefore, there are infinitely many completions. One often chooses the solution with the least Frobenius norm via the following quadratic optimization:

$$\begin{aligned} \min \quad & \|X\|^2 \\ \text{s.t.} \quad & X_{ij} = b_{ij} \quad \forall (i, j) \in \mathcal{B} \\ & X \in \mathcal{K}_+^n(\mathcal{V}). \end{aligned} \tag{11}$$

This problem is a special case of a more general problem considered by Hu et al. [23]:

$$\min \|X - C\|^2, \quad \text{s.t.} \quad AX = B, \quad X \in \mathcal{K}_+^n(\mathcal{V}), \tag{12}$$

where A, B are $m \times n$ real matrices and $C \in \mathcal{S}^n$ is given. Sufficient conditions have been studied for the feasibility of (12). It would be interesting to see their implications to the special case (11), especially their relationships to those studied in [26].

The equality constraints may render the feasible region in (11) empty. The following matrix approximation problem gets around of this infeasibility issue:

$$\min \|H \circ (X - A)\|^2 \quad \text{s.t.} \quad X \in \mathcal{K}_+^n(\mathcal{V}), \tag{13}$$

where $H \in \mathcal{S}^n$ is defined according to the index set \mathcal{B} as follows:

$$H_{ij} = \begin{cases} 1 & \text{if } (i, j) \text{ or } (j, i) \in \mathcal{B} \\ 0 & \text{otherwise.} \end{cases}$$

Problem (13) is often referred to as the H -weighted positive semidefinite approximation problem on a subspace. We will go no further about this problem except pointing out the fact that it is significantly more difficult to solve than its counterpart (11).

3 Euclidean Distance Matrix Problems

From the practical point of view, the Euclidean Distance Matrix (EDM) is best known in its role in the method of classical Multidimensional Scaling (cMDS), which is well documented in [10, 5]. It has also been a main subject in the recent research [11, 33, 31]. In this section, we first try to trace the earliest fundamental study of this topic. This part will be very useful for those who are not familiar with the EDM. We then discuss two nearest EDM problems, namely cMDS and dMDS (the direct MDS problem).

3.1 4 Classical Contributions

The foundation of the theory of EDM was built up in a few classical contributions, which, for the benefit of the reader, are briefly described below. A thorough study on EDM probably started from Schoenberg's seminar work [40], which answers Fréchet's question:

Let

$$a_{ik} = a_{ki} \quad (i \neq k, \quad k = 0, 1, 2, \dots, n)$$

be $\frac{1}{2}n(n+1)$ given positive quantities. What are the necessary and sufficient conditions that they be the lengths of the edges of a n -simplex $A_0A_1 \cdots A_n$? More general, what are the conditions

that they be the lengths of the edges of a n -simplex lying in a Euclidean space \mathbb{R}^r ($1 \leq r \leq n$) but not in a \mathbb{R}^{r-1} ?

The method proposed in [40] placed the first point A_0 at the origin and then defined the following $n \times n$ matrix B_0

$$(B_0)_{ij} = \frac{1}{2} (a_{0i}^2 + a_{0j}^2 - a_{ij}^2), \quad i, j = 1, \dots, n.$$

A necessary and sufficient condition for Fréchet's problem is that

$$B_0 \succeq 0 \quad \text{and} \quad \text{rank}(B_0) = r. \quad (14)$$

It then gave a construction how to find the n -simplex required in the question. It is not clear how the construction can be practically implemented. Schoenberg in the subsequent paper [42] proved that $B_0 \succeq 0$ if and only if

$$\left(-D^{(2)}\right) \succeq 0 \quad \text{on} \quad e^\perp, \quad (15)$$

where

$$D_{ij}^{(2)} = a_{ij}^2, \quad i, j = 0, 1, \dots, n.$$

This is exactly the widely adopted definition of Euclidean distance matrices. We like to point out that Schoenberg's theory goes much deeper. For example, using positive definite functions, it was proved that if $D^{(2)}$ defined above is EDM, so is any matrix $D^{(2\gamma)}$ for any $0 < \gamma \leq 1$, where

$$D_{ij}^{(2\gamma)} = a_{ij}^{2\gamma}.$$

Consequently, $D^{(1/2)}$, whose elements are a_{ij} , is also EDM (see Def. 3.1 for the formal definition of EDM).

In an apparently independent work, Young and Householder [48] also considered Fréchet's problem (but unaware it was proposed by Fréchet). They placed the last point A_n at the origin and defined the symmetric matrix B_n by

$$(B_n)_{ij} = \frac{1}{2} (a_{in}^2 + a_{jn}^2 - a_{ij}^2), \quad i, j = 0, 1, \dots, n-1.$$

It was proved that a necessary and sufficient condition for the existence of $(n+1)$ points in Fréchet's question is $B_n \succeq 0$ and the coordinates of those points come from the spectral decomposition of B_n .

$$B_n = X^T X \quad \text{and} \quad X = [x^1, \dots, x^n] \in \mathbb{R}^{r \times n}.$$

The theory of EDMs was made popular with practitioners by Torgerson [45, 46] and Gower [19]. Motivated by the work of Young and Householder [48], Torgerson emphasized that its theory is independent of the point that is to be placed at the origin and raised the question which is the best point to be placed at the origin. A heuristic argument suggested that the centroid of those points should be placed at the origin because it would mutually cancel the errors in the distance measurements. This amounts to the following necessary and sufficient condition:

$$-JD^{(2)}J \succeq 0, \quad (16)$$

where

$$J := I - \frac{1}{n+1} ee^T \quad (17)$$

and I is the identity matrix in \mathcal{S}^{n+1} and e is the vector of all ones in \mathbb{R}^{n+1} . Note that the result can be stated in \mathcal{S}^n for n points as is done in most publications, . We use \mathcal{S}^{n+1} because it has

$(n + 1)$ points in Fréchet's original question. It turns out (can be easily proved) that Torgerson condition (16) is equivalent to Schoenberg condition (15). Furthermore, the coordinates of the $(n + 1)$ points can be found through the spectral decomposition of the positive semidefinite matrix $-0.5JD^{(2)}J$. But Torgerson failed to realize that the coordinates thus produced are best from the viewpoint of principal component analysis (PCA), which was the major contribution of Gower.

Gower [19], apparently unaware of Torgerson's work, approached the problem from the viewpoint of a statistician. Given a set of distances $\{a_{ij}\}$ as in the Fréchet problem, construct the Q -matrix by

$$Q_{ij} = -\frac{1}{2}a_{ij}^2$$

and transform Q to the α -matrix (a symbol used by Gower) by

$$\alpha_{ij} = Q_{ij} - \bar{q}_i - \bar{q}_j + \bar{q},$$

where \bar{q}_i is the mean value of the i th row of Q and \bar{q} is the overall mean of Q . A necessary and sufficient condition for the distances $\{a_{ij}\}$ to be generated by a set of points $\{x^i\}$ is that the α -matrix is positive semidefinite. Moreover, those points $\{x^i\}$ can be obtained from the decomposition

$$\alpha = XX^T.$$

It is important to note that

$$\alpha = -\frac{1}{2}JD^{(2)}J.$$

Consequently, the centroid of the points is placed at the origin:

$$\sum_{i=0}^n x^i = 0.$$

Therefore, Gower rediscovered Torgerson condition (16) with more powerful optimal properties revealed from PCA. Gower termed the coordinates from α as the principal coordinates analysis.

Because J defined in (17) is the orthogonal projection matrix onto e^\perp , both the researches by Torgerson and Gower lead to Schoenberg condition (15). That is, the matrix $-D^{(2)}$ is conditionally positive semidefinite on $\mathcal{V} = e^\perp$. Another important piece of contribution is by Micchelli [29] on connections between EDMs and reproducing kernels, which play an essential role in machine learning research (e.g., support vector machines). We hope to gather examples in this field in future.

3.2 Nearest EDM problems

In this part, we will review two important classes of nearest EDM problems. To simplify notations, we deal with n points in \mathbb{R}^r rather than $(n + 1)$ points as used in Fréchet's question. Hence, the associated distance matrix is in \mathcal{S}^n . First, we give a formal definition of EDMs.

(a) On Definitions of EDM

Definition 3.1 *A symmetric matrix $D \in \mathcal{S}^n$ is called a Euclidean distance matrix if the following conditions are met:*

(i) *The diagonal elements of D are zero. That is*

$$D_{ii} = 0 \quad \forall i = 1, \dots, n.$$

(ii) The matrix $(-D)$ is positive semidefinite on $\mathcal{V} = e^\perp$.

We have following remarks regarding the use of this definition.

- (R1) This definition is essentially a restatement of Schoenberg's characterization [42, Eq.(5) and Eq. (6)] on the necessary and sufficient embedding conditions of a distance function to be embedded in the real Hilbert space. Since we are dealing with finite dimensional spaces, the embedding is in terms of Euclidean distances. Let \mathcal{S}_h^n denote the hollow subspace in \mathcal{S}^n :

$$\mathcal{S}_h^n := \{X \in \mathcal{S}^n \mid X_{ii} = 0, \ i = 1, \dots, n\}.$$

Define the almost positive semidefinite cone \mathcal{K}_+^n by

$$\mathcal{K}_+^n := \left\{ X \in \mathcal{S}^n \mid X \succeq 0 \text{ on } e^\perp \right\}.$$

Obviously, $\mathcal{K}_+^n = \mathcal{K}_+^n(\mathcal{V})$ with $\mathcal{V} = e^\perp$. Therefore, D is a EDM if and only if

$$-D \in \mathcal{S}_h^n \cap \mathcal{K}_+^n.$$

The set of all EDMs is called the Euclidean distance matrix cone, denoted by \mathcal{E}^n . The matrices in \mathcal{K}_+^n are known as almost positive matrices [12, Chp. XV] or almost positive definite matrices [29, Def. 2.2]. In this paper, we call \mathcal{K}_+^n the cone of almost positive semidefinite matrices.

We let e_i denote the i th basis vector in \mathbb{R}^n (i.e., e_i is the i th column of the identity matrix in \mathcal{S}^n). Obviously, $(e_i - e_j) \in e^\perp$ for any $i \neq j$. Hence,

$$0 \leq (e_i - e_j)^T (-D)(e_i - e_j) = -D_{ii} - D_{jj} + 2D_{ij} = 2D_{ij}.$$

In other words, the definition implicitly implies that every element in D is nonnegative. Therefore, $D_{ij} = d_{ij}^2$ for some d_{ij} . According to the basic theory on distance embedding of Schoenberg [40, 41], there exist n points $x^i \in \mathbb{R}^r$ for some r such that

$$\|x^i - x^j\| = \sqrt{D_{ij}} = d_{ij}.$$

- (R2) Based on the discussion above, an equivalent way to define a EDM is as follows. A nonnegative symmetric matrix $D \in \mathcal{S}^n$ in the form $D_{ij} = d_{ij}^2$ for $i, j = 1, \dots, n$ is a EDM if there exist n points $x^i \in \mathbb{R}^r$ for some r such that

$$\|x^i - x^j\|^2 = d_{ij}^2. \tag{18}$$

This explains why EDM is often defined as a squared distance matrix. However, Definition 3.1 is more convenient because it avoids using the term of “squared” distance. The Euclidean space \mathbb{R}^r is called the embedding space of D . Moreover, the smallest such r is called the embedding dimension of D .

From [40, 48, 46, 19] discussed in the preceding section, we know that the embedding dimension r is given by

$$r = \text{rank}(JDJ).$$

A set of such embedding points can be obtained from the spectral decomposition of the matrix $-0.5JDJ$.

- (R3) An important benefit of Def. 3.1 is that it does not depend on the actual embedding points. It would benefit us in understanding the following result: If D is EDM, then the matrix $D^{(\gamma)}$ defined by

$$D_{ij}^{(\gamma)} = D_{ij}^\gamma$$

is also EDM for $0 < \gamma < 1$ (see [42]). A direct consequence is that if D is a squared distance matrix in terms of $D_{ij} = d_{ij}^2$, then the matrix $D^{(1/2)}$, whose elements are d_{ij} , is also EDM. This means that there exist n points $z^i \in \mathbb{R}^m$ for some positive integer m such that

$$\|z^i - z^j\|^2 = d_{ij}.$$

Unlike the case in (18), it is not clear how such a set of points $\{z^i\}$ can be efficiently obtained.

We summarize the main result about EDM below.

Theorem 3.2 [40, 48, 45, 19] *Let $D \in \mathcal{S}^n$ be a EDM. Let J be the orthogonal projection matrix onto e^\perp :*

$$J := I - \frac{1}{n}ee^T. \quad (19)$$

The following hold.

(i) $-JDJ \succeq 0$ and the embedding dimension of D is $r = \text{rank}(JDJ)$.

(ii) Suppose

$$-\frac{1}{2}JDJ = X^T X \quad \text{with } X := [x^1, x^2, \dots, x^n], \quad (20)$$

where $x^i \in \mathbb{R}^r$ are called the embedding points of D . It holds that

$$\|x^i - x^j\|^2 = D_{ij}, \quad i, j = 1, \dots, n$$

and

$$\sum_{i=1}^n x^i = 0.$$

(b) Nearest EDM Problems. We say that $D \in \mathcal{S}^n$ is a pre-distance matrix if $D \in \mathcal{S}_h^n$ and $D_{ij} \geq 0$ for all $i, j = 1, \dots, n$. If D is a true EDM, then its embedding points can be obtained by the decomposition in (20). However, D often contains noises in practice and consequently it is not a true EDM. Naturally, one wishes to rectify it to get a true EDM. One would wish that the rectification is as small as possible or as simple as possible. Early methods popular in multidimensional scaling include the additive constant corrections and partial additive constant corrections [46, 10]. For more references on the two methods, see [36]. The constant correction method is very simple by its nature. In this part, we introduce two models that aim for the smallest changes for D to become a EDM.

The first model is a matrix nearness problem on the positive semidefinite cone \mathcal{S}^n . The intention was to find a best EDM Y such that

$$\nu_{cmds} := \min \|JYJ - 0.5JDJ\|^2, \quad \text{s.t. } Y \in \mathcal{S}_h^n, -JYJ \succeq 0. \quad (21)$$

There are two reasons for this model. One reason is that if D is a true EDM, then optimal solution would be D itself and the embedding points would come from decomposition of $-0.5JDJ$. The second reason is that problem (21) has a closed-form solution, as we see below. The solution method based on (21) is known as the classical multidimensional scaling (cMDS) [10, 5] and has

been used widely with some other advanced optimization techniques, see e.g., [47, 20] and the references therein.

When we replace $-JYJ$ by a positive semidefinite matrix and ignore the constraint $Y \in \mathcal{S}_h^n$, we get a relaxation problem

$$\nu_{cmds}^r := \min \|Z + 0.5JDJ\|^2, \quad \text{s.t. } Z \succeq 0. \quad (22)$$

We have the following result, which requires the mapping $\mathcal{D} : \mathcal{S}_+^n \mapsto \mathcal{E}^n$ given by

$$\mathcal{D}(Z) := \text{eddiag}(Z)^T + \text{diag}(Z)e^T - 2Z, \quad \forall Z \in \mathcal{S}_+^n.$$

We emphasize that $\mathcal{D}(Z)$ is just the EDM, whose embedding points are given by the spectral decomposition of Z . Furthermore,

$$JZJ = -\frac{1}{2}J\mathcal{D}(Z)J \quad \forall Z \in \mathcal{S}_+^n. \quad (23)$$

Theorem 3.3 *The relaxation problem (22) is exact in the following sense. Suppose Z^* is the optimal solution of (22). Then $Y^* \in \mathcal{S}^n$ defined by*

$$Y^* = \frac{1}{2}\mathcal{D}(Z^*)$$

is the unique optimal solution of (21) and the optimal objectives of the two problems are equal.

Proof. We first note that $-JYJ \succeq 0$ if and only if $-Y$ belongs to the almost positive semidefinite cone \mathcal{K}_+^n , which is $\mathcal{K}_+^n(e^\perp)$. Let Q be the Householder transformation that translates e to $[0, \dots, 0, -\sqrt{n}]^T \in \mathbb{R}^n$ (Q satisfies (5) with $S = e^\perp$). Hence,

$$Q = I - \frac{2}{v^T v} vv^T \quad \text{with } v = [1, 1, \dots, 1, \sqrt{n} + 1]^T \in \mathbb{R}^n.$$

The orthogonal matrix Q and the projection matrix J in (19) have the following relationship:

$$J = Q \begin{bmatrix} I_{n-1} & 0 \\ 0 & 0 \end{bmatrix} Q. \quad (24)$$

Now we return to the two optimization problems. The optimal solution of (22) is just the projection of $(-JDJ)$ onto \mathcal{S}_+^n :

$$Z^* = \Pi_{\mathcal{S}_+^n}(-0.5JDJ).$$

Denote

$$-0.5QDQ =: \begin{bmatrix} \widehat{D}_1 & \widehat{d} \\ \widehat{d}^T & \widehat{d}_0 \end{bmatrix} \quad \text{with } \widehat{D}_1 \in \mathcal{S}^{n-1}, \widehat{d} \in \mathbb{R}^{n-1}, \widehat{d}_0 \in \mathbb{R}.$$

It follows from (24) that

$$-0.5JDJ = Q \begin{bmatrix} \widehat{D}_1 & 0 \\ 0 & 0 \end{bmatrix} Q$$

and (because $Q^2 = I$)

$$Z^* = Q \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-1}}(\widehat{D}_1) & 0 \\ 0 & 0 \end{bmatrix} Q.$$

Furthermore,

$$\begin{aligned}
JZ^*J &= Q \begin{bmatrix} I_{n-1} & 0 \\ 0 & 0 \end{bmatrix} Q^2 \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-1}}(\widehat{D}_1) & 0 \\ 0 & 0 \end{bmatrix} Q^2 \begin{bmatrix} I_{n-1} & 0 \\ 0 & 0 \end{bmatrix} Q \\
&= Q \begin{bmatrix} \Pi_{\mathcal{S}_+^{n-1}}(\widehat{D}_1) & 0 \\ 0 & 0 \end{bmatrix} Q \\
&= Z^*.
\end{aligned}$$

Suppose $Z^* = X^T X$ with $X = [x^1, \dots, x^n]$ and $x^i \in \mathbb{R}^r$. Then $\mathcal{D}(Z^*) = (\|x^i - x^j\|^2)$. It follows from (23) that

$$0 \preceq Z^* = JZ^*J = -\frac{1}{2}JD(Z^*)J = -JY^*J.$$

Moreover, $Y^* \in \mathcal{S}_h^n$. Hence, Y^* is a feasible solution of (21). In the meantime, given that (22) is a relaxation of (21) we have

$$\nu_{cmds} \geq \nu_{cmds}^r = \|Z^* + 0.5JDJ\| = \|-JY^*J + 0.5JDJ\| = \|JY^*J - 0.5JDJ\| \geq \nu_{cmds}.$$

This proved that the relaxation is exact.

Now we prove Y^* thus defined is the unique optimal solution of (21). Define

$$\|Y\|^{(e)} := \|JYJ\|, \quad \forall Y \in \mathcal{S}^n.$$

It is easy to see that $\|Y\|^{(e)}$ is not a norm on \mathcal{S}^n . However, Mathar [28, Lemma 1] proved (by letting $s = (1/n)e$ therein) that it is a norm on the subspace \mathcal{S}_h^n . Now suppose \bar{Y} is another solution. Since (21) is convex, $Y^* + \tau(\bar{Y} - Y^*)$ is also optimal for $0 \leq \tau \leq 1$. We have

$$\|J(Y^* + \tau(\bar{Y} - Y^*))J\|^2 = \|JY^*J\|^2 + 2\tau\langle JY^*J, J(\bar{Y} - Y^*)J \rangle + \tau^2\|J(\bar{Y} - Y^*)J\|^2,$$

which implies

$$2\tau\langle JY^*J, J(\bar{Y} - Y^*)J \rangle + \tau^2\|J(\bar{Y} - Y^*)J\|^2 = 0.$$

Dividing τ on both sides leads to

$$2\langle JY^*J, J(\bar{Y} - Y^*)J \rangle + \tau\|J(\bar{Y} - Y^*)J\|^2 = 0$$

for all $0 \leq \tau \leq 1$. We must have $\|J(\bar{Y} - Y^*)J\| = 0$. Since $\|\cdot\|^{(e)}$ is a norm on \mathcal{S}_h^n and $\bar{Y} - Y^* \in \mathcal{S}_h^n$, we have $Y^* = \bar{Y}$. This proves the uniqueness of the optimal solution. \square

In literature, model (22) is also often referred to as the cMDS, which intends to solve model (21). Thm. 3.3 shows that the two models are equivalent. We are not aware of any formal proof for this equivalence. The nearness between Y^* and D in (21) is measured in terms of the norm of $J(Y^* - D)J$. A more natural nearness measure is the difference between the two matrices in terms of $\|Y^* - D\|$. This leads to our second model, which directly measure the nearness between the two matrices themselves using the norm $\|Y - D\|$. We refer the model as the direct MDS (dMDS) model for easy reference and it is stated as follows.

$$\min \|Y - D\|^2, \quad \text{s.t. } Y \in \mathcal{S}_h^n \text{ and } -Y \in \mathcal{K}_+^n. \quad (25)$$

Unlike cMDS model (21), this problem has no closed-form solution anymore. Hence, iterative algorithms have to be used. Note that the feasible region is the intersection of two convex sets and the objective is the orthogonal projection of D onto the feasible region. Therefore,

alternating projection methods are natural choices (see [13, 16]) and they have been used in molecular conformation [17, 18]. Problem (25) can also be reformulated as SDP, see [1] and [44]. A semismooth Newton-CG method has been studied in [33] and seems to outperform the alternating projection method and the interior-point methods for the SDP reformulations. Since those methods have been well documented in the above references, there is no need to describe them here. We would like to point it out that, in practice, fixed element constraints or box constraints

$$Y_{ij} = b_{ij} \quad \text{or} \quad \ell_{ij} \leq Y_{ij} \leq u_{ij} \quad (26)$$

are often added to problem (21) or problem (25). Those constraints can be readily dealt with in numerical methods for (21) and (25). To complete this section, we emphasize that both the cMDS model and the dMDS model are special cases of cQSDP of (1).

4 Semismooth Newton-CG Method

Over the past few years, semismooth Newton-CG based methods have proved to be extremely important for some class of matrix optimization problems, including the nearest correlation matrix problem [34], the SDP [49], the convex SDP [25] and the nearest Euclidean distance matrix problem [33], to just name a few. This section is devoted to a similar development, but is restricted to the following basic model of cQSDP:

$$\min \frac{1}{2} \|Y - A\|^2, \quad \text{s.t.} \quad \mathcal{A}(Y) = b, \quad Y \in \mathcal{K}_+^n(\mathcal{V}), \quad (27)$$

where $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^m$ is a linear mapping and $b \in \mathbb{R}^m$. In this basic model, we restrict the objective function f in (1) to be the distance function and we do not include inequalities here.

There are two reasons for focusing on this basic model. One is that it is adequate to show the essential components in a semismooth Newton-CG method and it is simple enough to lessen the complicated linear algebra involved. The second reason is that it plays an important role in numerical methods for a wider class of cQSDPs in the sense that it can be used to solve subproblems arising from those methods (e.g. the augmented Lagrangian method). We omit the detail in substantiating the second reason. The semismooth Newton-CG method is actually for the dual problem of (27). We study the dual in the following subsection.

4.1 The Lagrangian Dual Problem

We first derive the Lagrangian dual of problem (27). Define the Lagrangian function by

$$L(Y, y) := \frac{1}{2} \|Y - A\|^2 - \langle \mathcal{A}(Y), y \rangle, \quad (Y, y) \in \mathcal{S}^n \times \mathbb{R}^m.$$

Then the Lagrangian dual problem is defined to be

$$\max_{y \in \mathbb{R}^m} \widehat{\theta}(y), \quad (28)$$

where $\widehat{\theta}(y)$ is defined by

$$\widehat{\theta}(y) := \min_{Y \in \mathcal{K}_+^n(\mathcal{V})} L(Y, y).$$

Note that

$$L(Y, y) = \frac{1}{2} \|Y - (A + \mathcal{A}^*(y))\|^2 - \frac{1}{2} \|A + \mathcal{A}^*(y)\|^2 + \frac{1}{2} \|A\|^2 + \langle b, y \rangle,$$

where \mathcal{A}^* is the adjoint of \mathcal{A} . The optimal solution Y^* in $\widehat{\theta}(y)$ is

$$Y^* = \Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y)).$$

Hence,

$$\begin{aligned}\widehat{\theta}(y) &= L(Y^*, y) \\ &= \frac{1}{2} \|\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y)) - (A + \mathcal{A}^*(y))\|^2 - \frac{1}{2} \|A + \mathcal{A}^*(y)\|^2 + \frac{1}{2} \|A\|^2 + \langle b, y \rangle \\ &= -\frac{1}{2} \|\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y))\|^2 + \langle b, y \rangle + \frac{1}{2} \|A\|^2,\end{aligned}$$

where we used the fact that

$$Y = \Pi_{\mathcal{K}_+^n(\mathcal{V})}(Y) - \Pi_{(\mathcal{K}_+^n(\mathcal{V}))^*}(Y)$$

and the dual cone $(\mathcal{K}_+^n(\mathcal{V}))^*$ is defined by

$$(\mathcal{K}_+^n(\mathcal{V}))^* := \{Z \in \mathcal{S}^n \mid \langle Z, Y \rangle \geq 0, \forall Y \in \mathcal{K}_+^n(\mathcal{V})\}.$$

It follows from (9) that

$$(\mathcal{K}_+^n(\mathcal{V}))^* := \left\{ Z \in \mathcal{S}^n \mid Z = Q^T \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} Q, \quad B \in \mathcal{S}_+^{n-r} \right\}.$$

Note that $\mathcal{K}_+^n(\mathcal{V})$ is not self-dual.

We write the dual problem (28) in the minimization form.

$$\min_{y \in \mathbb{R}^m} \theta(y) := -\widehat{\theta}(y) = \frac{1}{2} \|\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y))\|^2 - \langle b, y \rangle - \frac{1}{2} \|A\|^2. \quad (29)$$

We need the strong Slater condition, which means that the mapping $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^m$ is surjective and there exists $Y^0 \in \text{int } \mathcal{K}_+^n(\mathcal{V})$ such that $\mathcal{A}(Y^0) = b$. It follows from [14, Prop. 2.20] (problem (27) can be cast as conic programming problem treated in [14, (2.71)]) that the following result holds.

Proposition 4.1 *If the strong Slater condition holds for problem (27), then for any $0 \neq y \in \mathbb{R}^m$ satisfying $\mathcal{A}^*(y) \in -(\mathcal{K}_+^n(\mathcal{V}))^*$, we must have*

$$\langle b, y \rangle < 0.$$

Proposition 4.2 *Assume that the strong Slater condition holds for problem (27), then the level set $L_c := \{y \mid \theta(y) \leq c\}$ is bounded for any constant c*

Proof. The proof can be patterned after that of [14, Prop. 4.16], which deals with the positive semidefinite cone. All technical parts including Prop. 4.1 in the proof can be carried over from \mathcal{S}_+^n to $\mathcal{K}_+^n(\mathcal{V})$. We omit the details. \square

Therefore, the dual problem (29) is guaranteed to have an optimal solution provided that the strong Slater condition holds. If there exist a point Y^0 such that $\mathcal{A}(Y^0) = b$ and $Y^0 \in \text{int } \mathcal{K}_+^n(\mathcal{V})$ (known as generalized Slater condition), then there is no duality gap between the primal problem (27) and the dual problem (29). Suppose \bar{y} is an optimal solution of (29), then

$$\bar{Y} := \Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(\bar{y}))$$

is the optimal solution of (27). Hence, it is enough to solve the dual problem and it is relatively easy to solve as it is defined in \mathbb{R}^m rather than in the significantly larger space \mathcal{S}^n .

Since $\theta(\cdot)$ is convex and continuously differentiable (because the projection in $\theta(y)$ is onto a close convex cone), the optimal solution is global and the first-order optimality condition is

$$F(y) := \nabla\theta(y) = \mathcal{A}\left(\Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y))\right) - b = 0. \quad (30)$$

Moreover, F is strongly semismooth because $\Pi_{\mathcal{K}_+^n(\mathcal{V})}(\cdot)$ is strongly semismooth. Equation (30) is a strongly semismooth equation and semismooth Newton method (see, e.g. [37]) can be readily developed. We will study it in the next subsection.

4.2 A Semismooth Newton-CG Method

The local version of the semismooth Newton method can be stated as follows. Given $y^0 \in \mathbb{R}^m$ and let $k := 0$. Compute $V_k \in \partial F(y^k)$ and

$$y^{k+1} = y^k - V_k^{-1}F(y^k), \quad k = 0, 1, 2, \dots \quad (31)$$

Since F is the gradient of θ , the generalized Jacobian ∂F in the sense of Clarke [9, Sect. 2.6] is often called the generalized Hessian of θ .

There are a few question for the semismooth Newton method (31). (i) How to calculate $V \in \partial F(y)$?; (ii) When is V nonsingular?; (iii) How to solve the linear equation in (31)?; and (iv) How to globalize the method?. Similar issues have been dealt with for the nearest correlation matrix problem by Qi and Sun [34] and for the nearest Euclidean distance matrix problem by Qi [33]. The difference here is that we have the conditional positive semidefinite cone $\mathcal{K}_+^n(\mathcal{V})$ instead of the positive semidefinite cone \mathcal{S}_+^n in [34] and the almost positive semidefinite cone \mathcal{K}_+^n in [33].

We first address question (iv). Since $\theta(\cdot)$ is convex and continuously differentiable, an Armijo line search can be introduced to the local method (31) to get a globally convergent method. This globalization strategy has been used by Qi and Sun [34] to globalize a semismooth Newton method for the nearest correlation matrix problem. We extend it to our problem.

Algorithm 4.3 (Newton-CG Method)

Step 0. Given $y^0 \in \mathbb{R}^m$, $\eta \in (0, 1)$, $\rho, \sigma \in (0, 1/2)$. $k := 0$.

Step 1. Select an element $V_k \in \partial F(y^k)$ and apply the conjugate gradient (CG) method of Hestenes and Stiefel [22] to find an approximate solution d^k to

$$\nabla\theta(y^k) + V_k d = 0 \quad (32)$$

such that

$$\|\nabla\theta(y^k) + V_k d^k\| \leq \eta_k \|\nabla\theta(y^k)\|, \quad (33)$$

where $\eta_k := \min\{\eta, \|\nabla\theta(y^k)\|\}$. If (33) is not achievable or if the condition

$$\nabla\theta(y^k)^T d^k \leq -\eta_k \|d^k\|^2$$

is not satisfied, let $d^k := -B_k^{-1}\nabla\theta(y^k)$, where B_k is any symmetric positive definite matrix in \mathcal{S}^n .

Step 2. Let m_k be the smallest nonnegative integer m such that

$$\theta(y^k + \rho^m d^k) - \theta(y^k) \leq \sigma \rho^m \nabla\theta(y^k)^T d^k.$$

Set $t_k := \rho^{m_k}$ and $y^{k+1} := y^k + t_k d^k$.

Step 3. Replace k by $k + 1$ and go to Step 1.

The global convergence of Alg. 4.3 and its quadratic convergence rate can be similarly proved as in [34, Thm. 5.3]. We omit the details. The positive definite matrix B_k can be chosen to be the identity matrix. The choice of $V_k \in \partial F(y^k)$ is dealt with in the next subsection. We emphasize that it is extremely important to use CG (or its variants) to solve the linear equation (32) because CG only needs evaluations of products of matrix V and vector d . Otherwise, it would have to use the full matrix V , which is very expensive to compute as seen from next subsection.

4.3 Characterization of $\partial F(y)$

The purpose in this subsection is to characterize any matrix $V \in \partial F(y)$. It is very challenging to characterize every element in V . But it is much less so if we just need to calculate Vh for any $h \in \mathbb{R}^m$. That is, we can compute the product of the matrix V and an arbitrary vector h . This is adequate for us because we are going to solve the linear equation (32) by a conjugate gradient method, which only uses products between matrices and vectors.

For given $y \in \mathbb{R}^m$, we let

$$Y := -J_S(A + \mathcal{A}^*(y))J_S.$$

Denote

$$\begin{bmatrix} Z_1 & Z_2 \\ Z_2^T & Z_0 \end{bmatrix} := -Q(A + \mathcal{A}^*(y))Q^T \quad \text{with} \quad Z_1 \in \mathcal{S}^{n-r}, \quad Z_2 \in \mathbb{R}^{(n-r) \times r}, \quad Z_0 \in \mathcal{S}^r.$$

We have from (8) that

$$Y = Q^T \begin{bmatrix} Z_1 & 0 \\ 0 & 0 \end{bmatrix} Q.$$

Now assume that Z_1 has the following spectral decomposition.

$$Z_1 = W\Lambda W^T,$$

with $\Lambda := \text{Diag}(\lambda_1, \dots, \lambda_{n-r})$, $\lambda_1 \geq \dots \geq \lambda_{n-r}$ are the eigenvalues of Z_1 , and $WW^T = I_{n-r}$. Define

$$\alpha := \{i \mid \lambda_i > 0\}, \quad \beta := \{i \mid \lambda_i = 0\}, \quad \gamma := \{i \mid \lambda_i < 0\}.$$

Define $\Omega \in \mathcal{S}^{n-r}$ by

$$\Omega_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\lambda_i| + |\lambda_j|}, \quad i, j = 1, \dots, n-r$$

where $0/0$ is defined to be 1. Let

$$W = [W_\alpha, W_\beta, W_\gamma] \quad \text{and} \quad \overline{W} := \begin{bmatrix} W_\alpha & W_\beta & 0 & W_\gamma \\ 0 & 0 & I_r & 0 \end{bmatrix} \in \mathcal{S}^n. \quad (34)$$

For any $h \in \mathbb{R}^m$, we partition $Q(\mathcal{A}^*(h))Q^T$ by

$$\begin{bmatrix} H_1 & H_2 \\ H_2^T & H_0 \end{bmatrix} := Q(\mathcal{A}^*(h))Q^T \quad \text{with} \quad H_1 \in \mathcal{S}^{n-r}, \quad H_2 \in \mathbb{R}^{(n-r) \times r}, \quad H_0 \in \mathcal{S}^r. \quad (35)$$

We have the following result, whose proof can be similarly constructed as that of [33, Prop. 3.2]

Proposition 4.4 *For any given $h \in \mathbb{R}^n$. Let $Q(\mathcal{A}^*(h))Q^T$ have the partition in (35). For every matrix $V \in \partial F(y)$, there exists $\tilde{V} \in \partial \Pi_{\mathcal{S}^{|\beta|+r}_+}(0)$ such that*

$$Vh = \mathcal{A}(\mathcal{A}^*(h)) - \mathcal{A}(P\mathcal{W}_hP^T), \quad (36)$$

where $P := Q\overline{W}$ and

$$\mathcal{W}_h := \begin{bmatrix} W_\alpha^T H_1 W_\alpha & [W_\alpha^T H_1 W_\beta & 0] & \Omega_{\alpha\gamma} \circ W_\alpha^T H_1 W_\gamma \\ \begin{bmatrix} W_\beta^T H_1 W_\alpha \\ 0 \end{bmatrix} & \tilde{V} \left(\begin{bmatrix} W_\beta^T H_1 W_\beta & 0 \\ 0 & 0 \end{bmatrix} \right) & 0 \\ \Omega_{\alpha\gamma}^T \circ W_\gamma^T H_1 W_\alpha & 0 & 0 \end{bmatrix}.$$

According to (36), there are two natural choices for $V_k \in \partial F(y^k)$ in Alg. 4.3. One corresponds to $\tilde{V} = 0$ and the other corresponds to $\tilde{V} = I$ (the identity matrix). The next question is when V is nonsingular so that the linear equation (32) has a unique solution. We study this issue below.

4.4 Nonsingularity of $\partial F(y)$

An important concept that ensures the nonsingularity is constraint nondegeneracy for problem (27). We refer to [8] for more references, results, and comments on its application to SDP. We describe it below.

Let $\overline{A} \in \mathcal{K}_+^n(\mathcal{V})$ have the following representation (see (9))

$$\overline{A} = Q^T \begin{bmatrix} \overline{Z}_1 & \overline{Z}_2 \\ \overline{Z}_2^T & \overline{Z}_0 \end{bmatrix} Q, \quad \text{with } \overline{Z}_1 \in \mathcal{S}^{n-r}, \overline{Z}_2 \in \mathbb{R}^{(n-r) \times r}, \overline{Z}_0 \in \mathcal{S}^r. \quad (37)$$

Let $\ell := \text{rank}(\overline{Z}_1)$ and $\overline{\lambda}_1 \geq \overline{\lambda}_2 \geq \dots \geq \overline{\lambda}_\ell > 0$ be the positive eigenvalues of \overline{Z}_1 . We assume that \overline{Z}_1 admits the spectral decomposition

$$\overline{Z}_1 = U \begin{bmatrix} \overline{\Lambda} & \\ & 0_{n-r-\ell} \end{bmatrix} U^T, \quad (38)$$

where $U^T U = I_{n-r}$. Let

$$\overline{\ell} := \{1, 2, \dots, \ell\}.$$

It follows from (9) that the tangent cone of $\mathcal{K}_+^n(\mathcal{V})$ at \overline{A} can be described as follows

$$\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\overline{A}) = \left\{ Q^T \begin{bmatrix} U \begin{bmatrix} M_1 & M_2 \\ M_2^T & M_0 \end{bmatrix} U^T & B \\ B^T & B_0 \end{bmatrix} Q \mid \begin{array}{l} M_1 \in \mathcal{S}^\ell, \quad M_0 \in \mathcal{S}_+^{n-r-\ell} \\ M_2 \in \mathbb{R}^{\ell \times (n-r-\ell)} \\ B \in \mathbb{R}^{(n-r) \times r}, \quad B_0 \in \mathcal{S}^r \end{array} \right\}.$$

The largest subspace contained in $\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\overline{A})$ is then given by

$$\text{lin}(\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\overline{A})) = \left\{ Q^T \begin{bmatrix} U \begin{bmatrix} M_1 & M_2 \\ M_2^T & 0 \end{bmatrix} U^T & B \\ B^T & B_0 \end{bmatrix} Q \mid \begin{array}{l} M_1 \in \mathcal{S}^\ell \\ M_2 \in \mathbb{R}^{\ell \times (n-r-\ell)} \\ B \in \mathbb{R}^{(n-r) \times r}, \quad B_0 \in \mathcal{S}^r \end{array} \right\}. \quad (39)$$

We say that constraint nondegeneracy holds at a feasible point \overline{A} of problem (27) if

$$\mathcal{A}(\text{lin}(\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\overline{A}))) = \mathbb{R}^m. \quad (40)$$

In other words, constraint nondegeneracy holds if the image space of $\text{lin}(\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\overline{A}))$ under the linear mapping \mathcal{A} is \mathbb{R}^m . We have the following characterization of constraint nondegeneracy.

Proposition 4.5 *Let $h \in \mathbb{R}^m$ be given and $Q^T(\mathcal{A}^*(h))Q$ have the representation of (35). Let $\bar{A} \in \mathcal{K}_+^n(\mathcal{V})$ be decomposed as in (37) and the resulting \bar{Z}_1 have the spectral decomposition (38). Constraint nondegeneracy holds at \bar{A} if and only if the following implication holds*

$$\left. \begin{array}{l} U_\ell^T H_1 = 0 \\ H_2 = 0 \\ H_0 = 0 \end{array} \right\} \implies h = 0. \quad (41)$$

Proof. By (40), constraint nondegeneracy holds at \bar{A} if and only if

$$h \in \left\{ \mathcal{A} \left(\text{lin}(\mathcal{T}_{\mathcal{K}_+^{n+1}}(\bar{A})) \right) \right\}^\perp \implies h = 0. \quad (42)$$

It follows from (39) that

$$\left\{ QXQ : X \in \text{lin}(\mathcal{T}_{\mathcal{K}_+^n(\mathcal{V})}(\bar{A})) \right\} = \left\{ \left[\begin{array}{cc} U \begin{bmatrix} M_1 & M_2 \\ M_2^T & 0 \end{bmatrix} U^T & B \\ B^T & B_0 \end{array} \right] : \begin{array}{l} M_1 \in \mathcal{S}^\ell \\ M_2 \in \mathbb{R}^{\ell \times (n-r-\ell)} \\ B \in \mathbb{R}^{(n-r) \times r}, B_0 \in \mathcal{S}^r \end{array} \right\}.$$

The left-hand side of (42) is equivalent to, for any $X \in \text{lin}(\mathcal{T}_{\mathcal{K}_+^{n+1}}(\bar{A}))$,

$$\begin{aligned} 0 &= \langle h, \mathcal{A}(X) \rangle = \langle \mathcal{A}^*(h), X \rangle \\ &= \langle Q\mathcal{A}^*(h)Q, QXQ \rangle \quad (\text{because } Q^2 = I) \\ &= 2\text{Tr}(B^T H_2) + \text{Tr}(B_0 H_0) + \text{Tr} \left(U^T \underline{H}_1 U \begin{bmatrix} M_1 & M_2 \\ M_2^T & 0 \end{bmatrix} \right). \end{aligned}$$

The above identities are for any $B \in \mathbb{R}^{(n-r) \times r}$, $B_0 \in \mathcal{S}^r$, $M_1 \in \mathcal{S}^\ell$ and $M_2 \in \mathbb{R}^{\ell \times (n-r-\ell)}$. Hence, we must have

$$H_2 = 0, \quad H_0 = 0 \quad \text{and} \quad U_\ell^T H_1 U = 0.$$

Because of the nonsingularity of U , the above condition is equivalent to

$$H_2 = 0, \quad H_0 = 0 \quad \text{and} \quad U_\ell^T H_1 = 0.$$

Therefore, (42) holds if and only if (41) holds. \square

We have the following major result, whose proof can be constructed by using Prop. 4.5 and by following the steps in [2, Thm. 3.9], which deals with the nonsingularity issue of a semismooth Newton method for a Euclidean distance matrix problem with spherical constraints. We omit the details.

Theorem 4.6 *Let y be an optimal solution of the dual problem (29). Let $\bar{A} := \Pi_{\mathcal{K}_+^n(\mathcal{V})}(A + \mathcal{A}^*(y))$. We assume that constraint nondegeneracy holds at \bar{A} . Then every matrix $M \in \partial F(y)$ is positive definite. Consequently, the Newton method (31) is quadratically convergent provided that the initial point y^0 is close to y .*

Before we move on to give a new application, we would like to point out the difference between this work and our previous work [33, 2], which made most of the subspace $\mathcal{V} = e^\perp$. In contrast, our analysis here is for general subspace \mathcal{V} and the assumptions are more general. For example, in the nearest Euclidean distance matrix problem [33], both the strong Slater condition and constraint nondegeneracy hold. For the spherical EDM problem in [2], the strong Slater condition holds while constraint nondegeneracy is satisfied under some (weak) condition.

5 A New Application

In this part, we present a new application of the direct MDS model (25) to calibrating correlations in Libor Market Models (LMM correlations). The need in calibrating LMM correlations has long been known (see e.g. [39, Chapters 19, 20]). In order to understand the strategies used in calibrating LMM correlations, let us restate a result of Micchelli.

Theorem 5.1 [29, Cor. 2.1(c)] $A \in (-\mathcal{K}_+^n)$ if and only if the matrix $M \in \mathcal{S}^n$ defined by

$$M_{ij} := \exp(-\alpha A_{ij}), \quad i, j = 1, \dots, n$$

is positive semidefinite for all $\alpha > 0$. Moreover, it is positive definite if and only if

$$A_{ij} > \frac{1}{2}(A_{ii} + A_{jj}) \quad \forall i, j = 1, \dots, n.$$

Regarding this result, we have the following remarks.

(R1) In the original statement of [29, Cor. 2.1(c)], it used “positive definite” to refer to “positive semidefinite” and used “strictly positive definite” to refer to “positive definite”.

(R2) If A is a EDM, then $A \in (-\mathcal{K}_+^n)$. We know that $A_{ii} = 0$ for all i . Then the matrix

$$M = \left(\exp(-\alpha A_{ij}) \right)$$

is a correlation matrix (because $M \succeq 0$ and it has unit diagonals). Moreover, if $A_{ij} > 0$ for all $i \neq j$ (this means that the embedding points of A are distinct), then M has full rank (i.e., M is positive definite).

(R3) If A is a EDM, so is $A^{(\gamma)}$ for $0 < \gamma < 1$ according to Schoenberg [42], where

$$A_{ij}^{(\gamma)} = (A_{ij})^\gamma.$$

Consequently,

$$M^{(\gamma)} := \left(\exp(-\alpha (A_{ij})^\gamma) \right), \quad 0 < \gamma < 1, \quad \alpha > 0 \quad (43)$$

are all correlation matrices. This fact seems to have been overlooked in [32], from which we will take some LMM correlation specifications.

The standard and basic LMM correlation specification is (see [32, Eq. (0.2)])

$$\rho_{ij} := \rho_\infty + (1 - \rho_\infty) \exp(-\lambda |T_i - T_j|), \quad -1 < \rho_\infty < 1 \quad \text{and} \quad \lambda > 0$$

where $\{T_i\}$ is a tenor structure (think them as of discretization of time) and $T_i \in \mathbb{R}$, $i = 1, \dots, n$. Since

$$A = \left((T_i - T_j)^2 \right) \in \mathcal{E}^n$$

and T_i , $i = 1, \dots, n$ can be regarded as the embedding points, we have

$$A^{(1/2)} = \left(|T_i - T_j| \right) \in \mathcal{E}^n.$$

It follows from (43) that the correlation matrix ρ is a legitimate correlation matrix, where

$$\rho := \rho_\infty E + (1 - \rho_\infty) \exp(-\lambda A^{(1/2)}), \quad -1 < \rho_\infty < 1 \quad \text{and} \quad \lambda > 0.$$

It follows from Schoenberg [42] that there exist $u^i \in \mathbb{R}^r$, $i = 1, \dots, n$ such that

$$\|u^i - u^j\|^2 = |T_i - T_j|.$$

This resolves one of the concerns of Piterbarg, who questions the existence of such embedding points.

A more general LMM specification [32] is as follows

$$\rho(T_i, T_j) = \rho_\infty + (1 - \rho_\infty) \exp(-\lambda(T_i, T_j)|T_i - T_j|),$$

where $\lambda(\cdot, \cdot)$ is some function. One of the interesting choices is from Rebonato [39]

$$\lambda(T_i, T_j) = \beta \exp\{-\alpha \max\{i, j\}\}, \quad \beta > 0 \text{ and } \alpha \in \mathbb{R}. \quad (44)$$

This choice of $\lambda(\cdot, \cdot)$ makes the resulting matrix ρ satisfy some monotonic properties among the correlations (we omit the details). However, a key question is whether the resulting matrix is a true correlation matrix. For example, when $n = 10$, $T_i = i$ for $i = 1, \dots, n$ and $\rho_\infty = 0$, $\alpha = \beta = 1$, the resulting matrix ρ has one negative eigenvalue (-0.1074). Therefore, for such generated matrix, one needs to perform calibration before it can be used as a true correlation matrix.

We propose using the direct MDS (25) to perform the calibration.

Algorithm 5.2 (*Calibration via dMDS*)

S.1 Construct the pre-distance matrix $D \in \mathcal{S}^n$ by

$$D_{ij} := \lambda(T_i, T_j)|T_i - T_j|. \quad (45)$$

S.2 Perform the dMDS optimization (25) to get the nearest EDM Y from D .

S.3 Construct the correlation matrix ρ^{new} by

$$\rho_{ij}^{new} = \rho_\infty + (1 - \rho_\infty) \exp(-Y_{ij}).$$

It follows from Thm. 5.1 that $\exp(-Y)$ (and hence ρ^{new}) is a true correlation matrix. We will test the following problem of various sizes.

Example 5.3 (*Test problem*). It computes the pre-distance matrix D in (45) with $\lambda(T_i, T_j)$ defined by (44) and

$$T_i = i, \quad i = 1, \dots, n \quad \text{and} \quad \alpha = \beta = 1.$$

All tests were carried out using the 64-bit version of MATLAB R2013a on a Windows 7 desktop with 64-bit operating system having Intel(R) Core(TM) 2 Duo CPU of 3.16GHz, and 4.0 GB of RAM. The dMDS optimization problem (25) was solved by Alg. 4.3 and the stopping criterion used is

$$\text{Res} := \|\nabla \theta(y^k)\| \leq 10^{-6}.$$

In the following table, we also include: **cpu** (the cpu time in seconds); **Iter** (the total number of iterations before termination); and **feval** (the total number of function evaluations of $\theta(y)$). We note that **feval** is also the number of spectral decomposition, which is used to compute $V_k d$ in (32). The starting point y^0 is set to be zero.

It is easy to see that the semismooth Newton-CG method is satisfactory from a practical point of view. It only used about 10 iterations to reach the required accuracy and it is very fast. For example, for $n = 1000$, it only used about 4 seconds and it used about 1 and half minutes for $n = 3000$.

Problem size (n)	cpu(s)	Iter	feval	Res
500	1	9	12	4.85E-07
800	2	8	13	2.10E-07
1000	4	9	13	9.36E-07
1200	6	9	13	9.83E-07
1500	11	9	12	8.55E-07
2000	23	9	11	9.35E-07
2500	47	10	11	7.59E-07
3000	89	11	12	1.13E-07

Table 1: Semismooth Newton-CG method for problems in Example 5.3

6 Conclusion

In this paper, we used cQSDP as a framework to include a number of quadratic matrix optimization problems. Two primal examples are the matrix completion (approximation) on a subspace and the problem of Euclidean distance matrices. We review some fundamental properties of those problems in terms of cQSDP. In particular, we focused on the orthogonal projection onto the conditional positive semidefinite cone $\mathcal{K}_+^n(\mathcal{V})$. For the EDM problems, the cone is known as the almost positive semidefinite cone \mathcal{K}_+^n . The projection function enjoys the property of strong semismoothness and therefore a semismooth Newton method is naturally an ideal numerical method for such problems. We studied the essential issues in developing the semismooth Newton method. We also included a new application in calibrating LMM correlations and presented some supporting numerical results that show the semismooth Newton method worked well.

There are a few topics worth further investigation. One is related to the special case when $Q = 0$ in (1), which results in the problem of conditional SDP (cSDP). It would be interesting to have interior-point methods for such cSDP. Moreover, examples in cSDP should be collected. Another topic is when there are box constraints in (26). New techniques (such as the smoothing technique in [15]) should be developed into the semismooth Newton method studied in Section 4 in order for it to be efficient. We leave it to our future research.

Acknowledgment

I would like to thank Dr Vladimir Lucic at Barclays Capital (London) for the reference [32] and for the useful discussions on EDM.

References

- [1] A.Y. ALFAKIH, A. KHANDANI, AND H. WOLKOWICZ, *Solving Euclidean distance matrix completion problems via semidefinite programming*, Comput. Optim. Appl., 12 (1999), pp. 13–30.
- [2] S. BAI, H.-D. QI, AND N. XIU, *Constrained best Euclidean distance embedding on a sphere: a matrix optimization approach*, Tech. Report, School of Mathematics, University of Southampton, December 2013.
- [3] B.J.C. BAXTER, *Conditionally positive functions and p-norm distance matrices*, Constr. Approx. 7 (1991), pp. 427–440.

- [4] J.F. BONNANS, J.C. GILBERT, C. LEMARÉCHAL, AND C.A. SAGASTIÁBAL, *Numerical Optimization*, Springer-Verlag, Berlin 2006.
- [5] I. BORG AND P.J.F. GROENEN, *Modern Multidimensional Scaling: Theory and Applications* (2nd ed.) Springer Series in Statistics, Springer, 2005.
- [6] X. CHEN, H.-D. QI, AND P. TSENG, *Analysis of nonsmooth symmetric matrix valued functions with applications to semidefinite complementarity problems*, SIAM J. Optim. 13 (2003), pp. 960–985.
- [7] Y. CHABRILLAC AND J.-P. CROUZEIX, *Definiteness and semidefiniteness of quadratic forms revisited*, Linear Algebra Appl., 63 (1984), pp. 283–292.
- [8] Z.X. CHAN AND D.F. SUN, *Constraint nondegeneracy, strong regularity and nonsingularity in semidefinite programming*, SIAM J. Optim. 19, 370–396 (2008).
- [9] F.H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.
- [10] T.F. COX AND M.A.A. COX, *Multidimensional Scaling*, 2nd Ed, Chapman and Hall/CRC, 2001.
- [11] J. DATTORRO, *Convex Optimization and Euclidean Distance Geometry*, Meboo Publishing USA, 2005.
- [12] W.F. DONOGHUE, *Monotone Matrix Functions and Analytic Continuation*, Springer-Verlag, Berlin 1974.
- [13] N. GAFFKE AND R. MATHAR, *A cyclic projection algorithm via duality*, Metrika, 36 (1989), pp. 29–54.
- [14] Y. GAO, *Structured Low Rank Matrix Optimization Problems: a Penalty Approach*, PhD Thesis (2010), National University of Singapore.
- [15] Y. GAO AND D.F. SUN, *Calibrating least squares covariance matrix problems with equality and inequality constraints*, SIAM J. Matrix Anal. Appl. 31 (2009), pp. 1432–1457
- [16] W. GLUNT, T.L. HAYDEN, S. HONG, AND J. WELLS, *An alternating projection algorithm for computing the nearest Euclidean distance matrix*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 589–600.
- [17] W. GLUNT, T.L. HAYDEN, AND W.-M. LIU, *The embedding problem for predistance matrices*, Bulletin of Mathematical Biology, 53 (1991), pp. 769–796.
- [18] W. GLUNT, T.L. HAYDEN, AND R. RAYDAN, *Molecular conformations from distance matrices*, J. Computational Chemistry, 14 (1993), pp. 114–120.
- [19] J.C. GOWER, *Some distance properties of latent root and vector methods in multivariate analysis*, Biometrika, 53 (1966), pp. 315–328.
- [20] I.G. GROOMS, R.M. LEWIS, AND M.W. TROSSET, *Molecular mbedding via a second order dissimilarity parameterized approach*, SIAM J. Sci. Comput. 31 (2009), pp. 2733–2756.
- [21] T.L. HAYDEN AND J. WELLS, *Approximation by matrices positive semidefinite on a subspace*, Linear Algebra Appl., 109 (1988), pp. 115–130.

- [22] M.R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand. 49 (1952), pp. 409–436.
- [23] X. HU, L. ZHANG, AND W. DU, *The solvability conditions for the inverse problem of matrices positive semidefinite on a subspace*, J. Comput. Maths. 12 (1994), pp. 78–87.
- [24] KH.D. IKRAMOV AND N.V. SAVEL’EVA, *Conditionally definite matrices*, J. Mathematical Sciences 98 (2000), pp. 1–50. Translated from Itogi Nauki i Tekhniki, Seriya Sovremennaya Matematika i Ee Prilozheniya, Tematicheskie Obzory. Vol. 52, Algebra-9, 1998.
- [25] K. JIANG, D.F. SUN, AND K.-C. TOH, *An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP*, SIAM J.Optim. 22 (2012), pp. 1042–1064.
- [26] C.R. JOHNSON AND R.L. SMITH, *The positive definite completion problem relative to a subspace*, Linear Algebra Appl. 307 (2000), pp. 1–14.
- [27] V. LUCIC, *Private communication*, January 2014.
- [28] R. MATHAR, *The best Euclidean fit to a given distance matrix in prescribed dimensions*, Linear Algebra and Its Applications 67 (1985), pp. 1–6.
- [29] C.A. MICCHELLI, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx. 2 (1986), pp. 11–22.
- [30] J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 2000.
- [31] R. PARHIZKAR, *Euclidean Distance Matrices: Properties, Algorithms and Applications*, PhD Thesis (2013), École Polytechnique Fédérale Lausanne.
- [32] V.V. PITERBARG *LMM correlation parameterization guaranteed to be positive semidefinite*, Technical Report, 2009.
- [33] H.-D. QI, *A semismooth Newton method for the nearest Euclidean distance matrix problem*, SIAM J. Matrix Anal. Appl. 34 (2013), pp. 67–93.
- [34] H.-D. QI AND D.F. SUN, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM J. Matrix Anal. Appl. 28 (2006), pp. 360–385.
- [35] H.-D. QI AND D.F. SUN, *An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem*, IMA J. Numer. Anal. 31 (2011), pp. 491–511.
- [36] H.-D. QI AND N. XIU, *A convex quadratic semidefinite programming approach to the partial additive constant problem in multidimensional scaling*, Journal of Statistical Computation and Simulation, 82 (2012), 1317–1336.
- [37] L. QI AND J. SUN, *A nonsmooth version of Newton’s method*, Math. Programming 58 (1993), pp. 353–367.
- [38] R. REAMS, *Hadamard inverse, square roots and products of almost semidefinite matrices*, Linear Algebra Appl. 288 (1999), pp. 35–43.
- [39] R. REBONATO, *Volatility and Correlation*, 2nd Eds (2004), John Wiley and Sons, Ltd.
- [40] I.J. SCHOENBERG, *Remarks to Maurice Fréchet’s article “Sur la définition axiomatique d’une classe d’espaces vectoriels distanciés applicables vectoriellement sur l’espace de Hilbert”*, Ann. Math. 36 (1935), pp. 724–732.

- [41] I.J. SCHOENBERG, *On certain metric spaces arising from Euclidean spaces by a change of metric and their embedding in Hilbert space*, Ann. Math. (2), 38 (1937), pp. 787–793.
- [42] I.J. SCHOENBERG, *Metric spaces and positive definite functions*, Trans. Amer. Maths. Soc. 44 (1938), pp. 522–536.
- [43] D.F. SUN AND J. SUN, *Semismooth matrix valued functions*, Math. Oper. Res. 27 (2002), pp. 150–169.
- [44] K.C. TOH, *An inexact path-following algorithm for convex quadratic SDP*, Mathematical Programming, 112 (2008), pp. 221–254.
- [45] W.S. TORGERSON, *Multidimensional scaling: I. theory and method*, Psychometrika 17 (1952), pp. 401–419.
- [46] W.S. TORGERSON, *Theory and Methods of Scaling*, John Wiley & Sons, Inc. 1958.
- [47] M.W. TROSSET, *Distance matrix completion by numerical optimization*, Comput. Optim. Appl., 17 (2000), pp. 11–22.
- [48] G. YOUNG AND A.S. HOUSEHOLDER, *Discussion of a set of points in terms of their mutual distances*, Psychometrika 3 (1938), pp. 19–22.
- [49] X. ZHAO, D.F. SUN, AND K.-C. TOH, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim. 20 (2010), 1737–1765.