Small Journal Name, ?, 1–18 (199?)


# SOLVING EUCLIDEAN DISTANCE MATRIX COMPLETION PROBLEMS VIA SEMIDEFINITE PROGRAMMING

ABDO Y. ALFAKIH aalfakih@orion.math.uwaterloo.ca
*University of Waterloo, Department of Combinatorics and Optimization, Waterloo, Ontario N2L 3G1, Canada*

AMIR KHANDANI khandani@claude.uwaterloo.ca
*University of Waterloo, Department of Electrical & Computer Engineering, Waterloo, Ontario N2L 3G1, Canada*

HENRY WOLKOWICZ henry@orion.math.uwaterloo.ca.
*University of Waterloo, Department of Combinatorics and Optimization, Waterloo, Ontario N2L 3G1, Canada.* *Research supported by Natural Sciences Engineering Research Council Canada.*

**Abstract.** Given a partial symmetric matrix $A$ with only certain elements specified, the Euclidean distance matrix completion problem (EDMCP) is to find the unspecified elements of $A$ that make $A$ a Euclidean distance matrix (EDM). In this paper, we follow the successful approach in [20] and solve the EDMCP by generalizing the completion problem to allow for approximate completions. In particular, we introduce a primal-dual interior-point algorithm that solves an equivalent (quadratic objective function) semidefinite programming problem (SDP). Numerical results are included which illustrate the efficiency and robustness of our approach. Our randomly generated problems consistently resulted in low dimensional solutions when no completion existed.

**Keywords:** Euclidean Distance Matrices, Semidefinite Programming, Completion Problems, Primal-Dual Interior-Point Method.

**AMS Subject Classifications:** 90C25 90C27 15A48

Dedicated to Olvi Mangasarian

The first time that I came across Olvi's work was as a graduate student in the 70's when I studied from his book on Nonlinear Programming (now a SIAM classic) and also used the *Mangasarian-Fromovitz constraint qualification.* This is *the* constraint qualification (CQ) in nonlinear programming (NLP), since it guarantees the existence of Lagrange multipliers and is equivalent to stability of the NLP. This CQ has since been extended to various generalizations of NLP and plays a crucial role in perturbation theory.

In 1983 I was a visitor at The University of Maryland, College Park, and was teaching a course in the Business College. While walking through the halls one day I noticed the name Fromovitz on one of the doors. I could not pass this by and knocked and asked if this was *the* Fromovitz. The reply was "yes"; and, this is the story of the now famous CQ. Stan Fromovitz had just received his PhD from Stanford and was working at Shell Development

Company in the Applied Math Dept. Olvi needed a special Theorem of the Alternative for his work on a CQ. Stan went digging into the Math Library at Berkeley and came up with exactly what was needed: Motzkin's Theorem of the Alternative. The end result of this was the MF CQ.

I have followed Olvi's work very closely throughout my career. His work is marked by many beautiful and important results in various areas. Some of the ones I am aware of are: condition numbers for nonlinear programs; generalized convexity; complementarity problems; matrix splittings; solution of large scale linear programs.

It is a pleasure and an honour to be able to contribute to this special issue.

Henry

## 1. INTRODUCTION

Euclidean distance matrices have received a lot of attention in recent years both for their elegance and for their many important applications. Two other research areas of high interest currently, are semidefinite programming and interior-point methods. In this paper we solve the Euclidean distance matrix completion problem by generalizing the completion problem to allow for approximate completions, i.e. we find a weighted, closest Euclidean distance matrix. In particular, we introduce a primal-dual interior-point algorithm that solves an equivalent (quadratic objective) semidefinite programming problem.

An $n \times n$ symmetric matrix $D = (d_{ij})$ with nonnegative elements and zero diagonal is called a *pre-distance matrix* (or dissimilarity matrix). In addition, if there exist points $x^1, x^2, \ldots, x^n$ in $\Re^r$ such that

$$d_{ij} = \|x^i - x^j\|^2, \qquad i, j = 1, 2, \ldots, n, \tag{1}$$

then $D$ is called a *Euclidean distance matrix* (EDM). The smallest value of $r$ is called *the embedding dimension* of $D$. Note that $r$ is always $\leq n-1$. Given a partial symmetric matrix $A$ with certain elements specified, the Euclidean distance matrix completion problem (EDMCP) consists in finding the unspecified elements of $A$ that make $A$ a EDM. Alternatively, for the approximate EDMCP, let $A$ be a pre-distance matrix, $H$ be an $n \times n$ symmetric matrix with nonnegative elements, and let $\|A\|_F = \sqrt{\operatorname{trace} A^t A}$ denote the *Frobenius norm* of $A$. Consider the objective function

$$f(D) := \|H \circ (A - D)\|_F^2,$$

where $\circ$ denotes *Hadamard product*, e.g. [19]. The weighted, *closest Euclidean distance matrix problem* is

$$(CDM_0) \quad \mu^* := \begin{array}{cc} \min & f(D) \\ \text{subject to} & D \in \mathcal{E}, \end{array}$$

where $\mathcal{E}$ denotes the cone of EDMs.

Applications of EDMs abound: e.g. molecular conformation problems in chemistry [9], [31]; multidimensional scaling and multivariate analysis problems in statistics [25], [26]; genetics, geography, and others, [1]. Many of these applications require a low embedding dimension, e.g. $r = 3$.

Theoretical properties of EDMs can be found in e.g. [7], [11], [15], [16], [21], [24], [34]. This includes characterizations as well as graph theoretic conditions for existence of completions. More information can be found in the recent survey article by Laurent [24]. (Generalizations of EDM arise in [36].)

An interesting discussion on algorithms for EDMCP appears in [41]. The point is made that there is no definitive general algorithm for EDMCP, i.e. one cannot provide an efficient decision rule for the question of whether a completion exists or not. However, there are many algorithms that find approximate completions. In [41], [39], [40], the author presents results on finding EDM completions based on spectral decompositions. In particular, the computationally hard problem of fixing the rank (the embedding dimension) is discussed. Some work on finding the closest EDM to a given symmetric matrix appears in [12], [47], [2]. Another approach based on global optimization and allowing for intervals for the distances is given in [28], [29] and also in [48]. We build a convex (tractable) model by relaxing the constraint on the embedding dimension (rank).

Semidefinite programming, SDP, is an extension of linear programming where nonnegativity constraints on vector variables is replaced by positive semidefiniteness constraints on matrix variables. This area has attracted a lot of interest recently because of the many applications as well as the elegant mathematics involved, see e.g. the survey papers: [13], [43], [42], [5]. A lot of the interest in SDP is from the interior-point community who have completed so much successful work on linear programming. At the moment, interior-point methods are the most successful algorithms for general SDP problems, see e.g. the above survey articles as well as the books [30], [46] and the recent theses [4], [1], [22], [35], [17], [32], [33]. The above references provide some evidence of the current high level of research activity in these areas.

The main contribution of this paper is a new approach to solving EDMCP. This approach is different than those in the literature in two ways. First we change the EDMCP into an approximation problem. This latter problem is a convex problem, i.e. our model is a tractable (polynomial time) model. Moreover, we further relax the original problem since we do not fix the embedding dimension. Thus, we do not solve problems that require a given dimension, though, we hope that this approach could be used as a first approximation to such problems. (A discussion on reducing the dimension is given in the conclusion, Section 5.) Our randomly generated tests consistently resulted in optimal solutions with low dimension, e.g. $r \leq 3$, when a completion did not exist. This is in contrast to the case where completions do exist, where the interior point method found the completion of maximum rank, since interior-point methods find optimal solutions in the relative interior of the optimal face. Second, we use a semidefinite programming formulation and a primal-dual interior-point algorithm to solve the approximation problem. And, we prove that the Slater constraint qualification holds for our model if and only if the graph of the matrix of weights is connected, see Corollary 1. Usually, the lack of the constraint qualification results in numerical difficulties due to unbounded optimal sets. However, in our case we can take advantage of a disconnected graph to replace the original problem by two smaller simpler problems.

As a side issue, we point out that our algorithm uses a new search direction for semidefinite programming introduced in [23]. This search direction is based on applying a Gauss-Newton approach to the optimality conditions. Our purpose is not to compare different search directions, and other public domain packages may be used for the problems that we have solved; though our tests show that our approach is comparable. We use this approach, rather than the more standard approaches already in the literature, since it is very well suited for our particular application. And, we think that having a program that is specific for this problem has many advantages for e.g. exploiting structure. Moreover, the details of the algorithm are self-contained in this paper; and, we provide a MATLAB program for those interested in our tests. Numerical results are included which illustrate the efficiency and robustness of the interior-point approach.

The paper is organized as follows. In Section 2 we introduce the basic results of EDM's. In Section 3 we present the optimality conditions for the problem. In Section 4 we derive the algorithm and the Slater constraint qualification result. We conclude with several remarks and numerical tests in Section 5. In addition, we include Section 5.1 with some technical details on the SDP algorithm.

## 2. DISTANCE GEOMETRY

It is well known, e.g. [34], [15], [16], [38], that a pre-distance matrix $D$ is a EDM if and only if $D$ is negative semidefinite on

$$M := \left\{ x \in \Re^n : x^t e = 0 \right\},$$

the orthogonal complement of $e$, where $e$ is the vector of all ones. Thus the set of all EDMs is a convex cone, which we denote by $\mathcal{E}$. We exploit this result to translate the cone $\mathcal{E}$ to the cone of semidefinite matrices in $\mathcal{S}_{n-1}$, the space of symmetric matrices of order $n-1$.

Define the $n \times n$ orthogonal matrix

$$Q := \left[ \frac{1}{\sqrt{n}} e \mid V \right], \qquad Q^t Q = I. \tag{2}$$

Thus $V^t e = 0$ and $V^t V = I$. Moreover, the subspace $M$ can be represented as the range of the $n \times (n-1)$ matrix $V$ and

$$J := VV^t = I - \frac{ee^t}{n} \tag{3}$$

is the orthogonal projection onto $M$.

Now define the *centered* and *hollow* subspaces of $\mathcal{S}_n$

$$\begin{aligned} \mathcal{S}_C &:= \{B \in \mathcal{S}_n : Be = 0\}, \\ \mathcal{S}_H &:= \{D \in \mathcal{S}_n : \mathrm{diag}(D) = 0\}, \end{aligned} \tag{4}$$

where $\mathrm{diag}(D)$ denotes the column vector formed from the diagonal of $D$. Following [10], we define the two linear operators acting on $\mathcal{S}_n$

$$\mathcal{K}(B) := \mathrm{diag}(B)\, e^t + e\, \mathrm{diag}(B)^t - 2B, \tag{5}$$

and

$$\mathcal{T}(D) := -\tfrac{1}{2} JDJ. \tag{6}$$

The operator $-2\mathcal{T}$ is an orthogonal projection onto $\mathcal{S}_C$; thus it is a self-adjoint idempotent.

THEOREM 1 *The linear operators satisfy*

$$\mathcal{K}(\mathcal{S}_C) = \mathcal{S}_H,$$
$$\mathcal{T}(\mathcal{S}_H) = \mathcal{S}_C,$$

*and* $\mathcal{K}_{|\mathcal{S}_C}$ *and* $\mathcal{T}_{|\mathcal{S}_H}$ *are inverses of each other.*

**Proof.** See [15], [21].

□

It can easily be verified that

$$\mathcal{K}^*(D) = 2\,(\mathrm{Diag}(De) - D) \tag{7}$$

is the adjoint operator of $\mathcal{K}$, where $\mathrm{Diag}(De)$ denotes the diagonal matrix formed from the vector $De$. In addition, a hollow matrix $D$ is EDM if and only if $B = \mathcal{T}(D)$ is positive semidefinite, denoted by $B \succeq 0$. (We denote positive definiteness by $B \succ 0$.) Equivalently, $D$ is EDM if and only if $D = \mathcal{K}(B)$, for some $B$ with $Be = 0$ and $B \succeq 0$. In this case the embedding dimension $r$ is given by the rank of $B$. Moreover, if $B = XX^t$, then the coordinates of the points $x^1, x^2, \ldots, x^n$ that generate $D$ are given by the rows of $X$ and, since $Be = 0$, it follows that the origin coincides with the centroid of these points. For these and other basic results on EDM see e.g. [15], [16], [21], [24], [34].

We now introduce the composite operators

$$\mathcal{K}_V(X) := \mathcal{K}(VXV^t), \tag{8}$$

and

$$\mathcal{T}_V(D) := V^t \mathcal{T}(D) V = -\tfrac{1}{2} V^t D V, \tag{9}$$

where $V$ is defined in (2).

LEMMA 1

$$\mathcal{K}_V(\mathcal{S}_{n-1}) = \mathcal{S}_H,$$
$$\mathcal{T}_V(\mathcal{S}_H) = \mathcal{S}_{n-1},$$

*and $\mathcal{K}_V$ and $\mathcal{T}_V$ are inverses of each other on these two spaces.*

**Proof.** This immediately follows from Theorem 1 and the definition of $V$. □

From (7), we get that

$$\mathcal{K}_V^*(D) = V^t\mathcal{K}^*(D)V \tag{10}$$

is the adjoint operator of $\mathcal{K}_V$. The following corollary summarizes the relationships between $\mathcal{E}$, the cone of Euclidean distance matrices of order $n$, and $\mathcal{P}$, the cone of positive semidefinite matrices of order $n-1$, which we will need for our model problem.

COROLLARY 1 *Suppose that $V$ is defined as in (2). Then:*

$$\mathcal{K}_V(\mathcal{P}) = \mathcal{E},$$
$$\mathcal{T}_V(\mathcal{E}) = \mathcal{P}.$$

**Proof.** We saw earlier that $D$ is EDM if and only if $D = \mathcal{K}(B)$ with $Be = 0$ and $B \succeq 0$. Let $X = V^tBV$, then since $Be = 0$ we have $B = VXV^t$. Therefore, $VXV^t \succeq 0$ if and only if $X \succeq 0$; and the result follows using (8) and Lemma 1. □

Note that the $n \times (n-1)$ matrix $V$ as defined in (2) is not unique. In our code we use

$$V := \begin{bmatrix} y & y & \dots & y \\ 1+x & x & \dots & x \\ x & 1+x & \dots & x \\ \dots & \dots & \ddots & \dots \\ x & x & \dots & 1+x \end{bmatrix}, \tag{11}$$

where $x = \frac{-1}{n+\sqrt{n}}$ and $y = \frac{-1}{\sqrt{n}}$. With this choice, it can be easily verified that $V^te = 0$, $V^tV = I$, and $VV^t = J$ as required by (2).

### 2.1. Program Formulations

Since $\operatorname{diag}(A) = \operatorname{diag}(D) = 0$, we can assume without loss of generality that $\operatorname{diag}(H) = 0$. Note that $H_{ij} = 0$ means that $D_{ij}$ is free, while $H_{ij} > 0$ forces $D_{ij}$ to be approximately equal to $A_{ij}$. i.e., $A_{ij}$ is approximately fixed. If we want $D_{ij} = A_{ij}$ exactly, then we can add a linear constraint to the program. (See below.) Recall that the graph of $H$ is *connected* if for all indices $i \neq j$ there is a *path* of indices $i_1, i_2, ..., i_k$ such that $H_{i,i_1} \neq 0, H_{i_1,i_2} \neq 0, \ldots, H_{i_{k-1},i_k} \neq 0, H_{i_k,j} \neq 0$, see e.g. [8]. Thus, we can assume that the graph of $H$ is connected or the problem can be solved more simply as two smaller problems, see Lemma 1 below. In particular, we can assume that $H$ does not have a row (hence a column) of all zeros; otherwise the corresponding row and column in $A$ and $D$ are free (independent) and the problem can be posed in a lower dimensional space.

By abuse of notation, let the function

$$f(X) := ||H \circ (A - \mathcal{K}_V(X))||_F^2 = ||H \circ \mathcal{K}_V(B - X)||_F^2,$$

where $B = \mathcal{T}_V(A)$. We now apply Corollary 1 and get the following problem, equivalent to $(CDM_0)$.

$$(CDM) \qquad \begin{array}{rcl} \mu^* := & \min & f(X) \\ & \text{subject to} & \mathcal{A}X = b \\ & & X \succeq 0. \end{array}$$

We allow for an additional constraint using the linear operator $\mathcal{A} : \mathcal{S}_{n-1} \longrightarrow \Re^m$, i.e., $b \in \Re^m$. The addition of this linear operator could represent some of the fixed elements in the given matrix $A$, e.g. adding the constraint $(K_V(X))_{ij} = A_{ij}$ fixes the $ij$ element of $D$. Also, note that $X \in \mathcal{S}_{n-1}$. It is in this lower dimensional space that we solve the problem. We can recover the optimal distance matrix using the optimal $X$ and the relation

$$D = \mathcal{K}_V(X).$$

Using finite precision, we can never solve the approximation problem exactly. In addition, we need to calculate the embedding dimension. The following lemma shows we lose little in the objective function if we choose a small embedding dimension using a numerical rank approach, i.e. if we only discard very small eigenvalues, then the objective function changes very little.

LEMMA 2 *Suppose that $X^*$ solves (CDM). Let $\bar{X}$ be the closest symmetric matrix to $X^*$ with rank $k$, i.e. we set the smallest $n-k$ eigenvalues of $X^*$ to 0, $\lambda_{k+1} = \ldots \lambda_n = 0$. Then,*

$$\sqrt{f(\bar{X})} \leq \sqrt{f(X^*)} + 2\gamma\left(\sqrt{n}+1\right)\sqrt{\sum_{i=k+1}^{n} \lambda_i^2}, \qquad (12)$$

*where* $\gamma := \max_{ij} H_{ij}$.

**Proof.** By the Cauchy-Schwartz inequality,

$$\sqrt{f(\bar{X})} \leq \sqrt{f(X^*)} + \gamma||\mathcal{K}(V(X^* - \bar{X})V^t)||_F.$$

The result now follows from the definition of $\mathcal{K}$. More precisely, let $B = V(X^* - \bar{X})V^t$. Then

$$\begin{aligned} \|\mathcal{K}(V(X^* - \bar{X})V^t)\|_F &= \|\mathcal{K}(B)\|_F \\ &\leq 2\|B\|_F + 2\|\mathrm{diag(B)e^t}\|_\mathrm{F} \\ &\leq 2\|B\|_F + 2\sqrt{n}\|B\|_F. \end{aligned}$$

□

## 3. DUALITY and OPTIMALITY

We now derive the optimality conditions and duality theory needed for a primal-dual interior-point approach.

For $\Lambda \in \mathcal{S}_{n-1}$ and $y \in R^m$, let

$$L(X, y, \Lambda) = f(X) + \langle y, b - \mathcal{A}(X)\rangle - \operatorname{trace} \Lambda X \tag{13}$$

denote the *Lagrangian* of (CDM). It is easy to see that the primal program (CDM) is equivalent to

$$\mu^* = \min_{X} \max_{\substack{y \\ \Lambda \succeq 0}} L(X, y, \Lambda) = \min_{X \succeq 0} \max_{\substack{y \\ \Lambda \succeq 0}} L(X, y, \Lambda). \tag{14}$$

We assume that the generalized Slater's constraint qualification,

$$\exists X \succ 0 \text{ with } \mathcal{A}(X) = b,$$

holds for (CDM). Slater's condition implies that strong duality holds, i.e.,

$$\mu^* = \max_{\substack{y \\ \Lambda \succeq 0}} \min_{X} L(X, y, \Lambda) = \max_{\substack{y \\ \Lambda \succeq 0}} \min_{X \succeq 0} L(X, y, \Lambda), \tag{15}$$

and $\mu^*$ is attained for some $y$ and $\Lambda \succeq 0$, see e.g. [44]. Since the semidefinite constraint on $X$ can be treated as redundant, the inner minimization of the convex, in $X$, Lagrangian is unconstrained and we can differentiate to get the equivalent problem

$$\mu^* = \max_{\substack{\nabla f(X) - \mathcal{A}^* y = \Lambda \\ \Lambda \succeq 0}} f(X) + \langle y, b - \mathcal{A}(X)\rangle - \operatorname{trace} \Lambda X. \tag{16}$$

We can now state the dual problem

$$\begin{array}{llll} \text{(DCDM)} & \mu^* := & \max & f(X) + \langle y, b - \mathcal{A}(X)\rangle - \operatorname{trace} \Lambda X \\ & & \text{subject to} & \nabla f(X) - \mathcal{A}^* y - \Lambda = 0 \\ & & & \Lambda \succeq 0, (X \succeq 0). \end{array} \tag{17}$$

We keep the semidefinite constraint on $X$ in brackets to emphasize that it is a hidden constraint in the dual, though it is a given constraint in the primal. The above pair of primal and dual programs, (CDM) and (DCDM), provide an optimality criteria in terms of feasibility and complementary slackness. This provides the basis for many algorithms including primal-dual interior-point algorithms. In particular, we see that the duality gap, in the case of primal and dual feasibility, is given by the complementary slackness condition:

$$\operatorname{trace} X(2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)) - \mathcal{A}^*\bar{y}) = 0, \tag{18}$$

or equivalently

$$X(2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)) - \mathcal{A}^*\bar{y}) = 0,$$

where $H^{(2)} = H \circ H$.

THEOREM 1 *Suppose that Slater's condition holds. Then* $\bar{X} \succeq 0$*, and* $\bar{y}$*,* $\bar{\Lambda} \succeq 0$ *solve (CDM) and (DCDM), respectively, if and only if the following three equations hold.*

$$\begin{array}{cc} \mathcal{A}(\bar{X}) = b & \textit{primal feasibility} \\ 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)) - \mathcal{A}^*\bar{y} - \bar{\Lambda} = 0 & \textit{dual feasibility} \\ \operatorname{trace} \bar{\Lambda}\bar{X} = 0 & \textit{complementary slackness} \end{array}$$

In the case that there is no linear operator $\mathcal{A}$, we get a simplified variational principle. The statement that no feasible direction is a descent direction translates into the following characterization of an optimal solution $\bar{X}$ of (CDM):

$$\nabla f(\bar{X}) \in (\mathcal{P} - \bar{X})^+, \tag{19}$$

where $\mathcal{P}$ is the cone of positive semidefinite matrices and

$$S^+ = \{P : \operatorname{trace} QP \geq 0, \quad , \forall Q \in S\}$$

is the *polar cone* of the set $S$. This yields the following characterization of optimality.

THEOREM 2 *Suppose that (CDM) has no linear constraint* $\mathcal{A}$*. The matrix* $\bar{X} \succeq 0$ *solves (CDM) if and only if*

$$\operatorname{trace} \mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B))\,(X - \bar{X}) \geq 0, \quad \forall X \succeq 0. \tag{20}$$

**Proof.** Note that the gradient acting on the symmetric matrix $h$, is

$$\begin{aligned} \langle \nabla f(X), h \rangle &= 2 \operatorname{trace} (H^{(2)} \circ \mathcal{K}_V(\bar{X} - B))\mathcal{K}_V(h) \\ &= 2 \operatorname{trace} \mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B))h. \end{aligned} \tag{21}$$

Therefore the gradient of the objective function is

$$\nabla f(X) = 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)). \tag{22}$$

The result follows upon replacing $h$ by the direction $X - \bar{X}$, for $X \succeq 0$, and applying the so-called Pshenichnyi condition (19).

□

## 4. PRIMAL-DUAL INTERIOR-POINT ALGORITHM

We now derive a primal-dual interior point method using the log-barrier approach [18]. This is an alternative way of deriving the optimality conditions in Theorem 1. For simplicity, we treat the unconstrained problem. i.e., we consider (CDM) without the linear equality constraint $\mathcal{A}X = b$. In this case, the Slater constraint qualification holds for both primal and dual problems, as can be seen from the following lemma. (In the case that some of the elements of $A$ are definitely fixed, then one needs to rederive the algorithm and include the constraint $\mathcal{A}X = b$.)

LEMMA 1 *Let $H$ be an $n \times n$ symmetric matrix with nonnegative elements and 0 diagonal such that the graph of $H$ is connected. Then*

$$\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(I)) \succ 0,$$

*where $I \in \mathcal{S}_{n-1}$ is the identity matrix.*

**Proof.** A simple calculation shows that $\mathcal{K}_V(I) = 2(ee^t - I)$. Thus $H^{(2)} \circ \mathcal{K}_V(I) = 2H^{(2)}$ and $\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(I)) = 4V^t(\mathrm{Diag}(H^{(2)}e) - H^{(2)})V$. Note that $e$ is an eigenvector of $\mathrm{Diag}(H^{(2)}e) - H^{(2)}$ with zero eigenvalue. We show that 0 is a simple eigenvalue. Assume to the contrary that there exists another nonzero eigenvector $u$, i.e. $u$ is not a multiple of $e$. Then

$$u^t(\mathrm{Diag}(H^{(2)}e) - H^{(2)})u = \sum_{i<j} H_{ij}^2 (u_i - u_j)^2 = 0.$$

From this and since the graph of $H$ is connected, it follows immediately that $u_i=$ constant for all $i$, a contradiction. Since $V$ is full column rank and $V^t e = 0$, we conclude that the columns of $V$ form a basis for the range of $(\mathrm{Diag}(H^{(2)}e) - H^{(2)})$, i.e. $4V^t(\mathrm{Diag}(H^{(2)}e) - H^{(2)})V$ is full rank.

□

Strict primal feasibility follows since every $X \succeq 0$ is feasible. Strict dual feasibility, i.e. $\Lambda \succ 0$, follows in the case of a connected graph, since $\alpha > 0$ can be chosen such that

$$X = B + \alpha I \succ 0; \tag{23}$$

we can then apply the previous Lemma 1 to get

$$\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) \succ 0. \tag{24}$$

COROLLARY 1 *Slater's constraint qualification holds for the dual problem (17) if and only if the graph of $H$ is connected.*

**Proof.** Sufficiency follows directly from the above Lemma 1.
To prove necessity, suppose that the graph is not connected. Then (CDM) can be solved as two smaller disconnected problems where the distances joining points

for these two smaller problems are completely free, i.e. the set of optimal distances matrices for ($\mathrm{CDM}_0$) is unbounded. Therefore, the set of optimal solutions $X$ of (CDM) is unbounded. This implies that Slater's condition cannot hold in the dual, see e.g. the proof of [14], Theorem 4 or [27]. In our case, it is rather simple to prove this result and we include it for completeness, i.e. we prove that Slater's condition for the dual implies that the optimal set for the primal is bounded.

Suppose that Slater's condition holds for the dual, i.e. there exists $\bar{\Lambda} \succ 0$ and $\bar{X} \succeq 0$ such that $\nabla f(\bar{X}) - \bar{\Lambda} = 0$. Equivalently, this means that $\nabla f(\bar{X}) \succ 0$. Let $X^*$ be optimal for (CDM). Then convexity of $f$ implies that

$$\left\langle \nabla f(\bar{X}), X^* - \bar{X} \right\rangle \leq 0.$$

Therefore, we get

$$\begin{array}{rcl} \left\langle \nabla f(\bar{X}), \bar{X} \right\rangle & \geq & \left\langle \nabla f(\bar{X}), X^* \right\rangle \\ & \geq & \lambda_{\min}(\nabla f(\bar{X}))\operatorname{trace} X^* \\ & \geq & \lambda_{\min}(\nabla f(\bar{X}))\lambda_{\max}(X^*), \end{array}$$

i.e. the norm of $X^*$ is bounded by $\frac{n}{\lambda_{\min}(\nabla f(\bar{X}))} \left\langle \nabla f(\bar{X}), \bar{X} \right\rangle$.

□

The log-barrier problem for (CDM) is

$$\min_{X \succ 0} B_\mu(X) := f(X) - \mu \log\det(X),$$

where $\mu \downarrow 0$. For each $\mu > 0$ we take one Newton step for solving the stationarity condition

$$\nabla B_\mu(X) = 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mu X^{-1} = 0. \tag{25}$$

Let

$$C := 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(B)) = 2\mathcal{K}_V^*(H^{(2)} \circ A). \tag{26}$$

Then the stationarity condition is equivalent to

$$\nabla B_\mu(X) = 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \mu X^{-1} = 0. \tag{27}$$

By equating $\Lambda = \mu X^{-1}$, and multiplying through by $X$, we get the optimality conditions, $F := \begin{pmatrix} F_d \\ F_c \end{pmatrix} = 0$,

$$\begin{array}{llcl} F_d & := 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \Lambda & = & 0 \ \text{ dual feas.} \\ F_c & := \Lambda X - \mu I & = & 0 \ \text{ perturbed compl. slack.,} \end{array} \tag{28}$$

and the estimate of the barrier parameter

$$\mu = \frac{1}{n-1} \operatorname{trace} \Lambda X. \tag{29}$$

Following, we present the p-d i-p framework we used. This is a common framework for both linear and semidefinite programming, see e.g. [45]. We include a centering parameter $\sigma_k$ (rather than the customary predictor-corrector approach) and let $\mathcal{F}^0$ denote the set of strictly feasible primal-dual points; $F'$ denotes the derivative of the function of optimality conditions.

ALGORITHM 1 *(p-d i-p framework:)*
**Given** $(X^0, \Lambda^0) \in \mathcal{F}^0$
**for** $k = 0, 1, 2 \ldots$
**solve** *for the search direction (in a least squares sense)*

$$F'(X^k, \Lambda^k) \begin{pmatrix} \delta X^k \\ \delta \Lambda^k \end{pmatrix} = \begin{pmatrix} -F_d \\ \Lambda^k X^k + \sigma_k \mu_k I \end{pmatrix}$$

*where* $\sigma_k$ *centering,* $\mu_k = \operatorname{trace} X^k \Lambda^k / (n-1)$

$$(X^{k+1}, \Lambda^{k+1}) = (X^k, \Lambda^k) + \alpha_k (\delta X^k, \delta \Lambda^k), \quad \alpha_k > 0,$$

*so that* $(X^{k+1}, \Lambda^{k+1}) \succ 0$
**end (for)**.

We use two approaches for solving the least squares problem. First we solve the large least squares problem; we call this the Gauss-Newton (GN) method. In the second approach, we restrict dual feasibility and substitute for $\delta\Lambda^k$ in the second equation; we call this the restricted Gauss-Newton (RGN) method. The numerical tests show that the second approach is significantly more efficient.

## 5. CONCLUSION and COMPUTATIONAL RESULTS

In this paper we have presented an interior-point algorithm for finding the weighted, closest Euclidean distance matrix; this provides a solution for the approximate Euclidean distance matrix completion problem. The algorithm has been extensively tested and has proven to be efficient and very robust, i.e. it has not failed on any of our test problems. In addition, an important observation is that the ranks of the optimal solutions $X$, for sparse problems where no completion existed, were typically between 1 and 3, i.e. a very small embedding dimension was obtained without any original rank restrictions.

However, when a completion existed, then typically the embedding dimension was high. In this case, we can apply the technique presented in [3] to "purify", i.e. to iteratively move to an optimal solution of smaller rank on the optimal face. (See also Pataki [32].)

We discuss more details of the Gauss-Newton approach in Section 5.1. The program was written in MATLAB. The tests were done on randomly generated problems; we used s SPARC 20 with 64 megs of RAM and SPECint 65.3, SPECfp 53.1.

The documented MATLAB code, as well as ongoing test results, can be obtained with URL (or anonymous ftp)
ftp://orion.math.uwaterloo.ca/pub/henry/software/distance.d, or
http://orion.math.uwaterloo.ca/˜hwolkowi/henry/software/distance.d.

In Table 1 we present a sample of test results. These results include problems with matrices up to dimension $n = 42$. We would like to emphasize that these are just preliminary test results. For example, we do not use a predictor-corrector approach, which has become the standard approach in interior-point methods, but rather use a centering parameter. Using the predictor-corrector approach should reduce the number of iterations from 20-30%. Tests with larger sized matrices are in progress.

*Table 1.* data for closest distance matrix: dimension; tolerance for duality gap; density of nonzeros in $H$; rank of optimal $X$; number of iterations; cpu-time for one least squares solution of the GN and restricted GN directions.

| dim | toler | $H$ dens. | $rank(X)$ | iterations | lss cpu-time | |
|---|---|---|---|---|---|---|
| | | | | | GN | RGN |
| 8 | $10^{-13}$ | .8 | 2 | 25 | .16 | .1 |
| 9 | $10^{-13}$ | .8 | 2 | 23 | .24 | .13 |
| 10 | $10^{-13}$ | .8 | 3 | 25 | .34 | .18 |
| 12 | $10^{-9}$ | .5 | 3 | 17 | .73 | .32 |
| 15 | $10^{-9}$ | .5 | 2 | 20 | 2.13 | .79 |
| 18 | $10^{-9}$ | .5 | 4 | 20 | 6.15 | 1.9 |
| 20 | $10^{-9}$ | .3 | 2 | 20 | 11.35 | 3.3 |
| 24 | $10^{-9}$ | .3 | 2 | 20 | 34.45 | 8.4 |
| 30 | $10^{-9}$ | .3 | 4 | 20 | 138. | 31.5 |
| 35 | $10^{-9}$ | .2 | 3 | 19 | 373. | 77. |
| 38 | $10^{-9}$ | .2 | 3 | 19 | 634. | 127 |
| 40 | $10^{-8}$ | .1 | 2 | 20 | 845.9 | 181.7 |
| 42 | $10^{-8}$ | .1 | 4 | 18 | 1118. | 232.02 |

We conclude with a specific example and its graph: the dimension is 11 generated with sparsity 0.5; the optimal value is 0.9256; the rank of $X$ is 3; and the number of iterations is 25 to get 13 decimals accuracy. The matrices $H$ and $A$ and the optimal distance matrix $D$ corresponding to the graph in Figure 1 are, respectively:

```
matrix H =

  0    0    0    0    0    5    2    5    0    7    0
  0    0    0    3    4    4    0    0    1    0    5
  0    0    0    0    0    0    1    0    0    0    2
  0    3    0    0    0    1    0    0    3    0    3
  0    4    0    0    0    2    7    0    3    0    0
  5    4    0    1    2    0    2    0    0    0    7
```

```
   2     0     1     0     7     2     0     6     0     1     2
   5     0     0     0     0     0     6     0     0     0     0
   0     1     0     3     3     0     0     0     0     3     0
   7     0     0     0     0     0     1     0     3     0     0
   0     5     2     3     0     7     2     0     0     0     0

-------------------------

 matrix A =

   0     0     0     0     0     4     0     0     6     6     2
   0     0     0     8     4     0     2     0     0     0     0
   0     0     0     5     6     0     6     0     0     0     7
   0     8     5     0     0     4     1     4     5     4     0
   0     4     6     0     0     0     0     0     5     0     0
   4     0     0     4     0     0     1     0     0     0     0
   0     2     6     1     0     1     0     0     0     0     3
   0     0     0     4     0     0     0     0     0     0     0
   6     0     0     5     5     0     0     0     0     0     0
   6     0     0     4     0     0     0     0     0     0     0
   2     0     7     0     0     0     3     0     0     0     0

-------------------------
    matrix D =

 Columns 1 through 7

        0    6.9200    7.2655    6.9713    0.7190    3.9912    0.5989
   6.9200         0    8.1310    6.6123    3.8224    0.6523    3.5381
   7.2655    8.1310         0    9.3511    5.9827    6.4399    6.0000
   6.9713    6.6123    9.3511         0    3.7085    3.6117    4.7459
   0.7190    3.8224    5.9827    3.7085         0    1.4994    0.0775
   3.9912    0.6523    6.4399    3.6117    1.4994         0    1.4981
   0.5989    3.5381    6.0000    4.7459    0.0775    1.4981         0
   0.1684    4.9773    6.4452    5.6419    0.2296    2.5198    0.1321
   8.9782    0.4804    9.0262    5.0302    4.8968    0.9976    4.9394
   5.9737    0.3183    7.3517    4.0355    2.7767    0.1996    2.7897
   5.2815    1.0844    7.0000    2.3417    2.1414    0.2465    2.3892

 Columns 8 through 11

   0.1684    8.9782    5.9737    5.2815
   4.9773    0.4804    0.3183    1.0844
   6.4452    9.0262    7.3517    7.0000
   5.6419    5.0302    4.0355    2.3417
   0.2296    4.8968    2.7767    2.1414
   2.5198    0.9976    0.1996    0.2465
   0.1321    4.9394    2.7897    2.3892
```

| | | | |
|---|---|---|---|
| 0 | 6.6871 | 4.1359 | 3.5984 |
| 6.6871 | 0 | 0.3050 | 0.7459 |
| 4.1359 | 0.3050 | 0 | 0.2315 |
| 3.5984 | 0.7459 | 0.2315 | 0 |

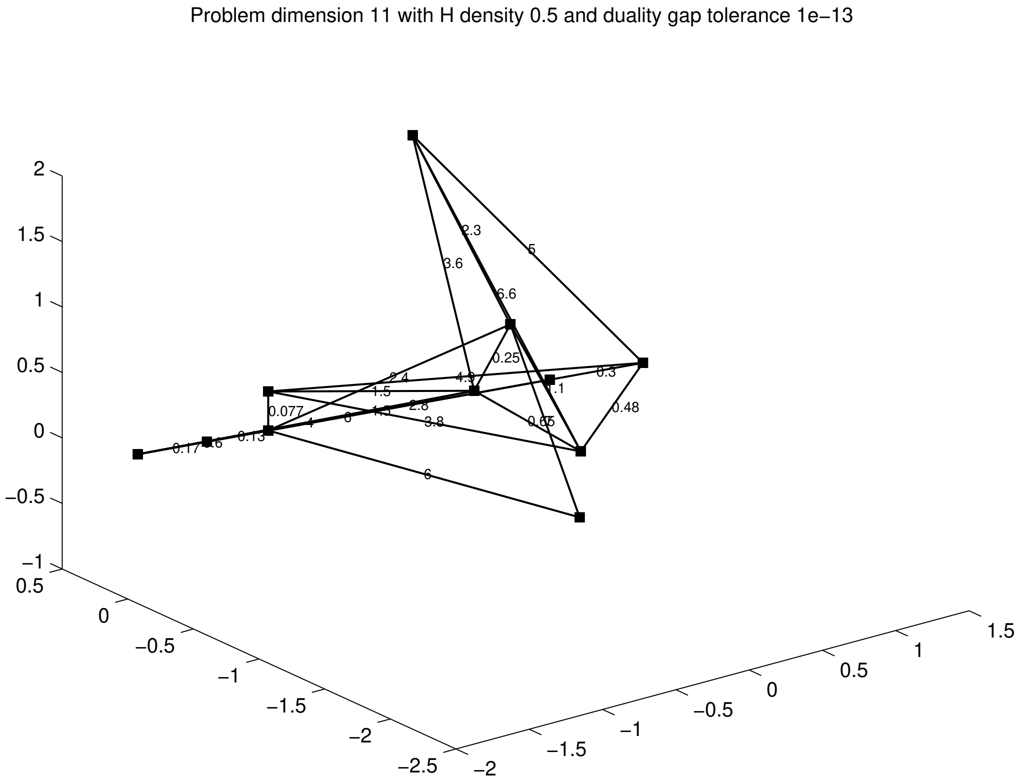


*Figure 1.* Approximate Completion Problem

### 5.1. Gauss-Newton Direction

The linear system for the search direction in Algorithm 1 is overdetermined. Therefore, it is not clear what is meant by solving this system. There are many different search directions that have been used for SDP, see e.g. [37]. Our problem is not a standard SDP since it has a quadratic objective and no constraints. Therefore, the standard public domain packages do not apply directly. [2] In our approach we use the Gauss-Newton direction introduced in [23], i.e. we linearize (28) and find the least squares solution of the resulting overdetermined linear system. This direction has many good properties, e.g.: it always exists; the linear systems for both GN and RGN are nonsingular at each iteration and in the limit; quadratic convergence steps are expected since the optimal value of the nonlinear least squares problems are 0. (For the details of these and other properties see [23].)

## Notes

1. This report is available by anonymous ftp at orion.math.uwaterloo.ca in the directory pub/henry/reports or with URLs: ftp://orion.uwaterloo.ca/pub/henry/reports/distmat.ps.gz or http://orion.math.uwaterloo.ca:80/~hwolkowi/henry/reports/distmat.ps.gz.
2. It has been pointed out to us that the package SDPpack, [6], will solve our problem if we replace our original objective function with the variable $t$ and add a constraint $||W||_F \leq t$, where $W$ represents the appropriate quantity in our original objective function. The numerical tests with this approach appear to be comparable to our approach. We thank Mike Overton for this observation.

## References

1. S. AL-HOMIDAN. *Hybrid methods for optimization problems with positive semidefinite matrix constraints.* PhD thesis, University of Dundee, 1993.
2. S. AL-HOMIDAN and R. FLETCHER. Hybrid methods for finding the nearest Euclidean distance matrix. In *Recent advances in nonsmooth optimization*, pages 1–17. World Sci. Publishing, River Edge, NJ, 1995.
3. A. ALFAKIH and H. WOLKOWICZ. On the embeddability of weighted graphs in Euclidean spaces. Technical Report CORR Report 98-12, University of Waterloo, 1998. submitted - URL: ftp://orion.uwaterloo.ca/pub/henry/reports/embedEDM.ps.gz.
4. F. ALIZADEH. *Combinatorial optimization with interior point methods and semidefinite matrices.* PhD thesis, University of Minnesota, 1991.
5. F. ALIZADEH. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5:13–51, 1995.
6. F. ALIZADEH, J.-P. HAEBERLY, M. V. NAYAKKANKUPPAM, and M. L. OVERTON. Sdppack user's guide – version 0.8 beta. Technical Report TR1997–734, Courant Institute of Mathematical Sciences, NYU, New York, NY, March 1997.
7. M. BAKONYI and C. R. JOHNSON. The Euclidean distance matrix completion problem. *SIAM J. Matrix Anal. Appl.*, 16(2):646–654, 1995.
8. R. A. BRUALDI and H. J. RYSER. *Combinatorial Matrix Theory.* Cambridge University Press, New York, 1991.
9. G. M. CRIPPEN and T.F. HAVEL. *Distance Geometry and Molecular Conformation.* Wiley, New York, 1988.
10. F. CRITCHLEY. On certain linear mappings between inner-product and squared distance matrices. *Linear Algebra Appl.*, 105:91–107, 1988.
11. R. W. FAREBROTHER. Three theorems with applications to Euclidean distance matrices. *Linear Algebra Appl.*, 95:11–16, 1987.
12. 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(4):589–600, 1990.
13. M. X. GOEMANS. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79:143–162, 1997.
14. E.G. GOL'STEIN. *Theory of Convex Programming.* American Mathematical Society, Providence , RI, 1972.

15. J. C. GOWER. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra Appl.*, 67:81–97, 1985.
16. T. L. HAYDEN, J. WELLS, W-M. LIU, and P. TARAZAGA. The cone of distance matrices. *Linear Algebra Appl.*, 144:153–169, 1991.
17. C. HELMBERG. *An interior point method for semidefinite programming and max-cut bounds.* PhD thesis, Graz University of Technology, Austria, 1994.
18. C. HELMBERG, F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. An interior point method for semidefinite programming. *SIAM Journal on Optimization*, pages 342–361, 1996. URL: ftp://orion.uwaterloo.ca/pub/henry/reports/sdp.ps.gz.
19. R.A. HORN and C.R. JOHNSON. *Matrix Analysis.* Cambridge University Press, New York, 1985.
20. C. JOHNSON, B. KROSCHEL, and H. WOLKOWICZ. An interior-point method for approximate positive semidefinite completions. *Computational Optimization and Applications*, 9(2):175–190, 1998.
21. C.R. JOHNSON and P. TARAZAGA. Connections between the real positive semidefinite and distance matrix completion problems. *Linear Algebra Appl.*, 223/224:375–391, 1995.
22. E. DE KLERK. *Interior point methods for semidefinite programming.* PhD thesis, Delft University, 1997.
23. S. KRUK, M. MURAMATSU, F. RENDL, R.J. VANDERBEI, and H. WOLKOWICZ. The Gauss-Newton direction in linear and semidefinite programming. Technical Report CORR 98-16, University of Waterloo, Waterloo, Canada, 1998. Detailed Web Version at URL: ftp://orion.uwaterloo.ca/pub/henry/reports/gnsdplong.ps.gz.
24. M. LAURENT. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *The Fields Institute for Research in Mathematical Sciences, Communications Series*, Providence, Rhode Island, 1998. American Mathematical Society.
25. J. DE LEEUW and W. HEISER. Theory of multidimensional scaling. In P. R. KRISHNAIAH and L. N. KANAL, editors, *Handbook of Statistics*, volume 2, pages 285–316. North-Holland, 1982.
26. S. LELE. Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Math. Geol.*, 25(5):573–602, 1993.
27. F. LEMPIO and H. MAURER. Differential stability in infinite-dimensional nonlinear programming. *Appl. Math. Optim.*, 6:139–152, 1980.
28. J.J. MORÉ and Z. Wu. Global continuation for distance geometry problems. Technical Report MCS-P505-0395, Applied Mathematics Division, Argonne National Labs, Chicago, Il, 1996.
29. J.J. MORÉ and Z. Wu. Distance geometry optimization for protein structures. Technical Report MCS-P628-1296, Applied Mathematics Division, Argonne National Labs, Chicago, Il, 1997.
30. Y. E. NESTEROV and A. S. NEMIROVSKI. *Interior Point Polynomial Algorithms in Convex Programming.* SIAM Publications. SIAM, Philadelphia, USA, 1994.
31. P. M. PARDALOS, D. SHALLOWAY, and G. XUE, editors. *Global minimization of nonconvex energy functions: molecular conformation and protein folding*, volume 23 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science.* American Mathematical Society, Providence, RI, 1996. Papers from the DIMACS Workshop held as part of the DIMACS Special Year on Mathematical Support for Molecular Biology at Rutgers University, New Brunswick, New Jersey, March 20–21, 1995.
32. G. PATAKI. *Cone programming and eigenvalue optimization: geometry and algorithms.* PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 1996.
33. M.V. RAMANA. *An algorithmic analysis of multiquadratic and semidefinite programming problems.* PhD thesis, Johns Hopkins University, Baltimore, Md, 1993.
34. I.J. SCHOENBERG. Remarks to Maurice Frechet's article: Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert. *Ann. Math.*, 36:724–732, 1935.
35. J. STURM. *Primal-dual interior point approach to semidefinite programming.* PhD thesis, Erasmus University Rotterdam, 1997.

36. P. TARAZAGA, T. L. HAYDEN, and JIM WELLS. Circum-Euclidean distance matrices and faces. *Linear Algebra Appl.*, 232:77–96, 1996.
37. M. TODD. On search directions in interior-point methods for semidefinite programming. Technical Report TR1205, School of OR and IE, Cornell University, Ithaca, NY, 1997.
38. WARREN S. TORGERSON. Multidimensional scaling. I. Theory and method. *Psychometrika*, 17:401–419, 1952.
39. M.W. TROSSET. Applications of multidimensional scaling to molecular conformation. Technical report, Rice Unversity, Houston, Texas, 1997.
40. M.W. TROSSET. Computing distances between convex sets and subsets of the positive semidefinite matrices. Technical report, Rice Unversity, Houston, Texas, 1997.
41. M.W. TROSSET. Distance matrix completion by numerical optimization. Technical report, Rice Unversity, Houston, Texas, 1997.
42. L. VANDENBERGHE and S. BOYD. Positive definite programming. In *Mathematical Programming: State of the Art, 1994*, pages 276–308. The University of Michigan, 1994.
43. L. VANDENBERGHE and S. BOYD. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.
44. H. WOLKOWICZ. Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.
45. S. WRIGHT. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, Pa, 1996.
46. Y. YE. *Interior Point Algorithms: Theory and Analysis.* Wiley-Interscience series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 1997.
47. F. Z. ZHANG. On the best Euclidean fit to a distance matrix. *Beijing Shifan Daxue Xuebao*, 4:21–24, 1987.
48. Z. ZOU, R.H. BYRD, and R.B. SCHNABEL. A stochastic/perturbation global optimization algorithm for distance geometry problems. Technical report, Dept. of Computer Science, University of Colorado, Boulder, Co, 1996.