Journal of Global Optimization **25:** 3–21, 2003.

Some Properties for the Euclidean Distance Matrix and Positive Semidefinite Matrix Completion Problems

HONG-XUAN HUANG[1], ZHI-AN LIANG[2] and PANOS M. PARDALOS[3]
[1]*Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China;*
[2]*Department of Mathematics, College of Sciences and Technology, Inner Mongolia University, Hohhot 010021, P.R. China;* [3]*Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA*

This paper is dedicated to Professor J. Ben Rosen on the occasion of his 80th birthday.

Abstract. The Euclidean distance matrix (EDM) completion problem and the positive semidefinite (PSD) matrix completion problem are considered in this paper. Approaches to determine the location of a point in a linear manifold are studied, which are based on a referential coordinate set and a distance vector whose components indicate the distances from the point to other points in the set. For a given referential coordinate set and a corresponding distance vector, sufficient and necessary conditions are presented for the existence of such a point that the distance vector can be realized. The location of the point (if it exists) given by the approaches in a linear manifold is independent of the coordinate system, and is only related to the referential coordinate set and the corresponding distance vector. An interesting phenomenon about the complexity of the EDM completion problem is described. Some properties about the uniqueness and the rigidity of the conformation for solutions to the EDM and PSD completion problems are presented.

Key words: Euclidean distance geometry problem; Euclidean distance matrix (EDM) completion problem; Positive semidefinite (PSD) matrix completion problem; Partial EDM (PSD) matrix; Linear manifold; Referential coordinate set; Complexity; Unique conformation; Rigid conformation

1. Introduction

A matrix $D=(d_{ij})\in\Re^{n\times n}$ is called a *Euclidean distance matrix* if there exist vectors $x_1,\ldots,x_n\in\Re^k$ (for some $k\geqslant 1$) such that $\|x_i-x_j\|=d_{ij}$ for all $i,j\in N=\{1,\ldots,n\}$, where $\|\cdot\|$ denotes the Euclidean norm in $\Re^k$. The set of the vectors $X=\{x_i\,|\,i\in N\}$ is called a *realization* of the Euclidean distance matrix D. In other words, a realization X of the Euclidean distance matrix D can be regarded as an embedding of a set with n atoms (Denote this set by $A=\{a_i\,|\,i\in N\}$ for simplicity) into the Euclidean space $\Re^k$ (for some $k\geqslant 1$) such that the Euclidean distance matrix of X is equal to D. Let EDM_n denote the set of all Euclidean distance matrices in $\Re^{n\times n}$. The *Euclidean distance geometry problems* are referred to as follows:

The basic Euclidean distance geometry problem:
Given a matrix $D \in \Re^{n \times n}$, is it a Euclidean distance matrix, i.e., $D \in EDM_n$?
The general Euclidean distance geometry problem:
Given an index set $I \subset \{(i, j) \mid i, j = 1, \ldots, n\}$ and two real sets $\{l_{ij} \mid (i, j) \in I\}$ and $\{u_{ij} \mid (i, j) \in I\}$, is there a Euclidean distance matrix $D \in EDM_n$ such that, for any $(i, j) \in I$, $l_{ij} \leqslant d_{ij} \leqslant u_{ij}$?

In the fields of bio-informatics and computational chemistry, many research subjects about protein structures and macro-molecular modeling are related to the reconstruction of a three-dimensional set of atoms by using information about their inter-atomic distances, where atoms may be regarded as points from the viewpoint of mathematics [3, 6]. The distances usually can be obtained via X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy and analyzed by distance geometry methods [7, 11]. Hence, this kind of reconstruction problem is called *the molecular problem* [11] or *the Euclidean distance geometry* (EDG) *problem* [7]. In addition, *the multidimensional scaling problem*, arising in statistics in order to deal with the similarity/dissimilarity among some objects, is related to a certain kind of the EDG problem [5]. Some matrix completion problems, which have received a lot of attention in the literature in recent years within the community of linear algebra, are also related to the EDG problem [2, 13, 16].

Since NMR experiments afford broad ranges of possible distances only for some atom pairs, the NMR data often have the following two features: (i) The distances obtained are not absolutely exact and have errors; (ii) Only a sparse distance matrix is available. For those pairs whose distances are not given, the lower bounds will simply be determined by the Van der Waals radii, and the upper bounds by a typical extended ranges for the related molecule. In order to deal with NMR data, people have proposed many algorithms for the EDG problems, such as the EMBED algorithm [7], the spectral gradient algorithm [9], the graph reduction algorithm [11, 12], the global continuation algorithm [17], the tabu-based pattern search method [18] and the spectral distance geometry algorithm [21]. For the molecular distance geometry problem with exact inter-atomic distances, a linear time algorithm can be found in Ref. [8]. Based on a new error function defined by the sum of the absolute difference, distance geometry problem can be reduced into a concave quadratic minimization problem, for which positive semidefinite relaxations are possible [22].

In this paper, we will study a special class of Euclidean distance geometry problems in a linear manifold, whose corresponding distance matrix has some entries specified exactly, while others may not be specified. The importance of this special class of the EDM problems is that its solution usually can be used as an iteratively approximated solution to the general EDM problem and is also related closely to the positive semidefinite matrix completion problem.

For a given set X, a mapping $d : X^2 \rightarrow \Re$ is said to be *premetric* on X if the mapping d satisfies the conditions:

(i) $d(u, v) = d(v, u),\ \forall u, v \in X$;
(ii) $d(u, v) = 0$ if and only if $u = v,\ \forall u, v \in X$.

In this case, a pair (X, d) is called a *premetric space*. When X is a finite set, denote its elements by $\{x_1, \dots, x_n\}$, the $n \times n$ matrix $D = (d(x_i, x_j))$ is called a *premetric matrix*. Furthermore, if the mapping d also satisfies the condition:

(iii) $d(u, v) \geqslant 0,\ \forall u, v \in X$,

then d is said to be a *semimetric* on X. The pair (X, d) and the corresponding matrix D are called a *semimetric space* and a *semimetric matrix*, respectively. If the mapping d satisfies (i)–(iii) and the condition:

(iv) $d(u, v) \leqslant d(u, w) + d(w, v),\ \forall u, v, w \in X$,

then d is said to be a *metric* on X. The pair (X, d) and the corresponding matrix D are called a *metric space* and a *metric matrix*, respectively [4, 7].

Given any set of points A and an embedding function $p : A \to \Re^m$, then, the embedding p of the set A will induce an m-dimensional Euclidean distance function $d_p(u, v) := \|p(u) - p(v)\|$, where $u, v \in A$ and $\|\cdot\|$ is the Euclidean norm on the vector space $\Re^m$, such that (A, d_p) becomes a Euclidean distance space associated with the set A. In particular, $(\Re^m, d_p)$ (denoted by $\Re^m$ for simplicity) is a special m-dimensional Euclidean distance space.

A necessary condition for $D \in EDM_n$ is that D must be also a premetric, semimetric and metric matrix. It is clear that there exist many realizations for a given Euclidean distance matrix D. A translation, rotation or reflection of a realization X of the matrix $D \in EDM_n$ gives another realization Y of the matrix D, but we will show that the *conformation* (i.e., the spatial structure related to points in a realization) of these realizations remains the same. A realization X of the matrix $D \in EDM_n$ is said to be *congruent* to another realization Y if X can be obtained from Y by a rigid motion (i.e., by a translation, a rotation or a reflection), or a composition of some rigid motions. Denote two congruent realizations X and Y (i.e., two congruent embeddings of the Euclidean distance space (A, D)) by $X \sim Y$.

A partial symmetric matrix $P \in \Re^{n \times n}$ is a matrix whose entries are specified only on a subset of the positions, but in such a way that p_{ji} is specified and equal to p_{ij} whenever p_{ij} is specified. *The matrix completion problem* is referred to as follows: Given a partial matrix $P \in \Re^{n \times n}$, can the unspecified entries of P be chosen such that the resulting matrix satisfies a certain property? *The Euclidean distance matrix completion problem* (EDM completion problem, for short) asks whether a given partial symmetric matrix can be completed to a Euclidean distance matrix. *The positive semidefinite matrix completion problem* (PSD completion problem, for short) asks whether a given partial symmetric matrix can be completed to a positive semidefinite matrix. Many results have been obtained for these two completion problems, which are based on certain graphic structures corresponding to the specified entries (see [14–16] and references therein). But no efficient algorithm is known for deciding whether or not a given partial EDM or PSD matrix can be completed.

A set of points $X(D) = \{x_i \mid i \in N\}$ is said to be a *realization* of a given partial Euclidean distance matrix D if it is a realization of the completed Euclidean distance matrix of D. Similarly, a realization $X(D)$ of the matrix D is said to be *congruent* to another realization $Y(D)$ if $X(D)$ can be obtained from $Y(D)$ by a rigid motion or a composition of some rigid motions. All congruent realizations of the matrix D constitute an equivalence class of point sets.

It is clear that, for the EDM completion problem, one may consider only partial matrices whose diagonal entries are all specified and equal to 0. For the PSD completion problem, one can restrict it to the case of partial matrices whose diagonal entries are all specified and equal to 1. The positive semidefinite matrix, whose diagonal entries equal to 1, is known as *the correlation matrix*. Let PSD_n denote the set of all positive semidefinite matrices in $\Re^{n\times n}$. Denote the set of all correlation matrices in $\Re^{n\times n}$ by

$$\Phi_n = PSD_n \cap \{P = (p_{ij}) \mid p_{ii} = 1,\ \forall i \in N\}\,.$$

The paper is organized as follows: In Section 2, the connection between the EDM and PSD completion problems are considered. In Section 3, approaches to determine the location of a point in a linear manifold are studied, which use the distances from the point to the ones in a referential coordinate set. For a distance vector associated with a referential coordinate set, the sufficient and necessary conditions for the existence of such a point that the distance vector can be realized, are also presented. In Section 4, an interesting phenomenon about the complexity of the EDM completion problem will be described, and properties about the unique conformation and the rigid conformations of solutions to the EDM and PSD completion problems are obtained. Some concluding remarks are given in the final section.

2. Connections between the EDM and PSD completion problems

Let $G = (V_n, E)$ be a graph with a node set $V_n = \{1, \ldots, n\}$ and an edge set E. Let K_n denote the complete graph with n nodes. Let $\Phi(G)$ and $EDM(G)$ denote the projections of Φ_n and EDM_n on the subspace $\Re^E$ indexed by the edge set of G, respectively. In particular, the sets $\Phi(K_n)$ and Φ_n are in one-to-one correspondence, as well as the sets $EDM(K_n)$ and EDM_n. Let $PSG(G)$ denote the projections of PSD_n on the subspace $\Re^{E\cup\{(i,i)\mid i\in V_n\}}$ indexed by the union of the edge set of G and the set $\{(i,i) \mid i \in V_n\}$. For the simplicity, let $E' = E \cup \{(i,i) \mid i \in V_n\}$.

The suspension graph ∇G is defined as the graph with the node set $V_{n+1} = V_n \cup \{n+1\}$ and with the edge set $E(\nabla G) = E \cup \{(i, n+1) \mid i \in V_n\}$. Given a subset $U \subseteq V_n$, let $G[U]$ denote the subgraph of G induced by U, with the node set U and with the edge set $\{(u,v)\in E\,|\,u,v\in U\}$. U is called a *clique* if $G[U]$ is a complete graph.

For the graph G and its suspension graph ∇G, let a one-to-one linear correspondence ξ between the space $\Re^{E(\nabla G)}$ and $\Re^{E'}$, which is called *covariance mapping*, be defined in the following manner:

For $D\in\Re^{E(\nabla G)}$, $P=\xi(D)=(p_{ij})\in\Re^{E'}$ such that

$$p_{ij}=\frac{1}{2}(d^2_{i,n+1}+d^2_{j,n+1}-d^2_{ij})\quad (i,j)\in E'\,. \tag{1}$$

For a partial symmetric matrix $D\in EDM(\nabla G)$, without loss of generality, denote a realization of the completed matrix of the matrix D in $\Re^m$ by $\{x_i\,|\,i=1,2,\dots,n+1\}$, it is easy to check that the completed matrix of the corresponding partial matrix P is equal to X^TX, where the matrix $X=(x_1-x_{n+1},\dots,x_n-x_{n+1})\in\Re^{m\times n}$. This indicates that $P\in PSD(G)$. In fact, the following assertion holds:

$$D\in EDM(\nabla G)\Leftrightarrow P=\xi(D)\in PSD(G) \tag{2}$$

The well-known correspondence between EDM_{n+1} and PSD_n was proved by Schoenberg [20] as follows:

$$D\in EDM_{n+1}\Leftrightarrow P=\xi(D)\in PSD_n\,. \tag{3}$$

The function $F_\lambda : t\to\exp(-\lambda t)$, where $t,\lambda\in\Re_+=\{x:x\in\Re,\ x\geqslant 0\}$, is called *the Schoenberg transform*. A connection beween EDM_n and Φ_n can be found in [16, 20].

LEMMA 2.1 [16, 20]. *Let $D\in\Re^{n\times n}$ be a symmetric matrix with an all-zero diagonal. The following assertions are equivalent:*
(i) $D\in EDM_n$.
(ii) The matrix $(F_\lambda(d^2_{ij}))\in\Phi_n$ for all $\lambda>0$.
(iii) The matrix $((1-F_\lambda(d^2_{ij}))^{1/2})\in EDM_n$ for all $\lambda>0$.

The result has been extended to the EDM and PSD completion cases as follows:

LEMMA 2.2 [15, 16]. *Let $G=(V_n,E)$ be a graph and let the matrix $D\in\Re^E$. The following assertions are equivalent:*
(i) $D\in EDM(G)$.
(ii) The matrix $(F_\lambda(d^2_{ij}))\in\Phi(G)$ for all $\lambda>0$.
(iii) The matrix $((1-F_\lambda(d^2_{ij}))^{1/2})\in EDM(G)$ for all $\lambda>0$.

By the formulae (1)–(3), Lemma 2.1 and Lemma 2.2, we know that the PSD completion problem is related closely to the EDM completion problem. Therefore, we will pay more attention to the EDM completion problem in the following sections. Based on the results for the EDM completion problem, it is easy to derive the corresponding results for the PSD completion problem.

3. Basic approaches for determing the location of a point

From properties of linear dependent and linear independent sets [1], we can obtain easily the following Lemma 3.1.

LEMMA 3.1. *Given a set of points* $X = \{x_i \mid i = 0, 1, \ldots, k\} \subset \Re^m$. *Let* $S_X = \{\Sigma_i \lambda_i x_i \mid \Sigma_i \lambda_i = 1,\ x_i \in X,\ \lambda_i \in \Re,\ i = 0, \ldots, k\}$ *be a linear manifold in* $\Re^m$ *generated by the set X. Then the following assertions are equivalent for the set X:*

(i) The dimension $dim(S_X)$ *of the linear manifold* S_X *is k.*
(ii) The $m \times k$ *matrix* $A_0 = (x_1 - x_0, \ldots, x_k - x_0)$ *is of full rank of columns.*

DEFINITION 3.1. A finite set $X \subset \Re^m$ is referred to as a referential coordinate set if the dimension $\dim(S_X)$ of its corresponding linear manifold S_X is equal to $|X| - 1$.

For a referential coordinate set X, it can be proved easily that the linear manifold S_X is equal to $x_0 + span\{x_1 - x_0, \ldots, x_k - x_0\}$ in the m-dimensional space $\Re^m$. Now, we give a lemma, which is fundamental to determine the location of a point in a linear manifold corresponding to a given referential coordinate set and a distance vector whose components indicate the distances from the point to those in the set.

LEMMA 3.2. *Let the index set* $I_k = \{0, 1, \ldots, k\}$. *Given a referential coordinate set* $X = \{x_i \mid i \in I_k\} \subset \Re^m$, *and a vector* $y \in \Re^m$, *consider the problem*

$$\min_{x \in S_X} \|y - x\|^2 , \tag{4}$$

where S_X *is the linear manifold generated by the set X. Then, the following two conclusions hold:*

(i) There exists a unique solution $x^* \in S_X$ *to the problem (4):*
(ii) For any integer $i \in I_k$, *denote the matrix* $A_i = (x_j - x_i \mid j \in I_k,\ j \neq i)$, *the solution to the problem (4) can be represented as follows:*

$$x^* = (I - P_{A_i})x + P_{A_i} y , \tag{5}$$

where I is the identity matrix in $\Re^{m \times m}$, *x is an arbitrary point in* S_X, *and* $P_{A_i} = A_i(A_i^T A_i)^{-1} A_i^T$ *is the projective operator onto the subspace spanned by the columns of* A_i.

Proof. (i) It is clear that the objective function in problem (4) is strict convex. Since the objective function is also a coercive function (i.e., $\|y - x\|^2 \to +\infty$ as $\|x\| \to +\infty$), and the constrained domain S_X is a nonempty convex set, the solution $x^* \in S_X$ to problem (4) exists and is unique.

(ii) For any integer $i \in I_k$, let the matrix A_i be defined as above and S_i be the subspace spanned by the columns of A_i. Since X is a referential coordinate set, by

Lemma 3.1 and properties of linear independent sets, the dimension of he linear manifold S_X is k and S_i is the same as the subspace S_j spanned by the columns of A_j, where $j \in I_k$, $j \neq i$. Denote the same subspace by S_0 in the following consideration.

First, we show that the operator P_{A_i} is independent of the matrix A_i. Let $\{u_1, u_2, \ldots, u_k\}$ be any basis of the subspace S_0 and denote the matrix $U = (u_1, \ldots, u_k)$. There must exist a nonsingular transition matrix $Q \in \Re^{k\times k}$ from A_i-coordinates to U-coordinates such that $A_i = UQ$ [1]. Denote $S_0^\perp$ to be the orthogonal complementary subspace of S_0. Since

$$P_{A_i} = A_i(A_i^T A_i)^{-1} A_i^T = U(U^T U)^{-1} U^T = P_U .$$

holds, the projective operators P_{A_i} and $I - P_{A_i}$ onto the subspaces S_0 and $S_0^\perp$, respectively, are independent of the special basis $\{x_j - x_i \mid j \in I_k, j \neq i\}$ of the subspace S_0. Hence, for the given set X and the vector y, $P_{A_i} y$ is not dependent with the matrix A_i.

Next, we prove that (5) holds for any $x \in S_X$. It is easy to check that

$$(I - P_{A_i})(x_j - x_i) \equiv 0 .$$

Using the fact $S_X = x_i + S_0$, for any $x \in S_X$, we have $(I - P_{A_i})x = (I - P_{A_i})x_i$. Hence, $(I - P_{A_i})x$ is a constant vector in the linear manifold S_X. Furthermore, the vector

$$\hat{x} = (I - P_{A_i})x + P_{A_i} y$$

is independent of any vector $x \in S_X$. It is clear that $\hat{x} \in S_X$ and is independent with the matrix A_i, $i \in I_k$. Since

$$\|\hat{x} - y\|^2 = \|(I - P_{A_i})(x - y)\|^2 \leqslant \|x - y\|^2 , \quad \forall x \in S_X ,$$

holds and the solution x^* of problem (4) is unique, we have $x^* = \hat{x}$. □

THEOREM 3.1. *Given a referential coordinate set* $X = \{x_i \mid i = 0, 1, \ldots, k\} \subset \Re^m$ *and a nonnegative vector* $d \in \Re_+^{k+1} = \{x \in \Re^{k+1} \mid x \geqslant 0\}$. *Let* S_X *be the linear manifold corresponding to X. Then, one and only one of the following cases holds*:

(i) There exists a unique point $x^* \in S_X$ *such that, for any point* $x_i \in X$, *the Euclidean distance between* x^* *and* x_i *is equal to* d_i, *i.e.,* $\|x^* - x_i\| = d_i$;
(ii) For any point $x \in S_X$, *there exists a certain* $x_i \in X$ *such that* $\|x - x_i\| \neq d_i$.

Proof. Let us consider the following equations with respect to $y \in \Re^m$

$$\|y - x_i\| = d_i , \quad i = 0, 1, \ldots, k , \tag{6}$$

which are equivalent to

$$\|y - x_i\|^2 = \|y\|^2 - 2x_i^T y + \|x_i\|^2 = d_i^2 , \quad i = 0, 1, \ldots, k .$$

Subtracting the equation corresponding to $i = 0$ from the rest, we obtain

$$-2(x_i - x_0)^T y = d_i^2 - d_0^2 + \|x_0\|^2 - \|x_i\|^2\,, \quad i = 1, \ldots, k\,.$$

These linear equations with respect to y are equivalent to the following equation in matrix form

$$A^T y = b\,, \tag{7}$$

where $A = (x_1 - x_0, \ldots, x_k - x_0)$, and $b_i = (d_0^2 - d_i^2 - \|x_0\|^2 + \|x_i\|^2)/2$ for $i = 1, \ldots, k$.

Since X is a referential coordinate set, the matrix A is of full rank of columns. Hence, $A^T A$ is a symmetric positive definite matrix. For any $y \in \Re^m$, by Lemma 3.2, the linear system (7) has a solution

$$x^* = (I - P_A)x_0 + A(A^T A)^{-1} b\,, \tag{8}$$

such that $\|y - x^*\| = \min_{x \in S_X} \|y - x\|$, where $P = A(A^T A)^{-1} A^T$. It is easy to see that the equations (6) are equivalent to $\|y - x_0\| = d_0$ and the linear system (7). It is clear that, if $y \in \Re^m$ satisfies the linear system (7), then $\|y - x_i\|^2 - d_i^2 = \|y - x_0\|^2 - d_0^2$ ($i = 1, \ldots, k$). Therefore, case (i) holds if and only if $x^* \in S_X$ satisfies $\|x^* - x_0\| = d_0$.

We have mentioned that x^* satisfies the system (7) and $x^* \in S_X$. If case (i) does not hold, then $\|x^* - x_0\| \neq d_0$. Since the linear manifold $S_X = x_0 + S_0$, any point $x \in S_X$ can be represented as $x_0 + A\lambda$, where $\lambda \in \Re^k$. If $\|x - x_0\| \neq d_0$, then case (ii) holds. Otherwise, we claim that there must exist a $j \geqslant 1$ such that $\|x - x_j\| \neq d_j$. If $\|x - x_j\| = d_j$, $\forall j \geqslant 1$, then these formulae together with $\|x - x_0\| = d_0$ will imply that x is a solution of the linear system (7), i.e., $A^T x = b$. By Lemma 3.2, we have

$$\begin{aligned} x^* &= (I - P_A)x + A(A^T A)^{-1} b \\ &= x - A(A^T A)^{-1}(A^T x - b) \\ &= x\,. \end{aligned}$$

Hence, $\|x - x_0\| = \|x^* - x_0\| \neq d_0$. This is a contradiction to the assumption $\|x - x_0\| = d_0$. Therefore, when the case (i) does not hold, the case (ii) must hold. □

Note that the linear system (7) has a unique solution in the linear manifold S_X.

THEOREM 3.2. *Let the index set $I_k = \{0, 1, \ldots, k\}$. Given a referential coordinate set $X = \{x_i \mid i \in I_k\} \subset \Re^m$, denote D to be the Euclidean distance matrix associated with X. Then, for any realization Y of the matrix D in $\Re^m$, Y is congruent to X, i.e., the conformation corresponding to the matrix D is unique.*

Proof. We prove this theorem by induction. For two two-point sets $X^{(1)} = \{x_0, x_1\}$ and $Y^{(1)} = \{y_0, y_1\}$, if the distance between x_0 and x_1 is equal to that between y_0 and y_1, then it is clear that $X^{(1)}$ is congruent to $Y^{(1)}$ by a translation, or a rotation, or a composition of a translation and a rotation. Hence, the conclusion holds for $k = 1$. For $k = 2$, given a three-point referential coordinate set $X^{(2)} = \{x_0, x_1, x_2\}$, let

$Y^{(2)} = \{y_0, y_1, y_2\}$ be any realization of the Euclidean distance matrix $D^{(2)}$ associated with $X^{(2)}$. By the knowledge of Euclidean geometry, we know that the two triangles generated by the points in $X^{(2)}$ and $Y^{(2)}$, respectively, are congruent, i.e., $Y^{(2)} \sim X^{(2)}$. Hence, the conclusion also holds for $k = 2$.

Suppose that the conclusion holds for $k = n$. For $k = n + 1$, let $X^{(n+1)}$ be a given referential coordinate set and $D^{(n+1)}$ be the corresponding Euclidean distance matrix. Suppose that $Y^{(n+1)}$ is any realization of the matrix $D^{(n+1)}$. Next, we will show that $Y^{(n+1)} \sim X^{(n+1)}$.

Without loss of generality, suppose that the indices of points in $X^{(n+1)}$ and $Y^{(n+1)}$ are consistent with the column indices of the matrix $D^{(n+1)}$, i.e., the distance between $x_i^{(n+1)}$ and $x_j^{(n+1)}$ (between $y_i^{(n+1)}$ and $y_j^{(n+1)}$) is equal to $d_{i+1,j+1}^{(n+1)}$, where $i, j = 0, 1, \ldots, n + 1$. Denote $D_0 := (d_{ij}^{(n+1)})_{i,j \leqslant n+1}$ to be the upper left $(n + 1) \times (n + 1)$ submatrix of $D^{(n+1)}$.

Since $X_0 := \{x_i^{(n+1)} \mid 0 \leqslant i \leqslant n\}$ and $Y_0 := \{y_i^{(n+1)} \mid 0 \leqslant i \leqslant n\}$ are two realizations of the Euclidean distance matrix D_0 and the former is also a referential coordinate set, by the induction hypothesis, we get that $Y_0 \sim X_0$.

Let x^* and y^* be solutions to problems:

$$\min_{x \in S_{X_0}} \|x_{n+1}^{(n+1)} - x\| \quad \text{and} \quad \min_{x \in S_{Y_0}} \|y_{n+1}^{(n+1)} - x\| ,$$

respectively. From the proof of Theorem 3.1, it is clear that x^* and y^* are also the solutions to the linear system (7) in the linear manifolds S_{X_0} and S_{Y_0} with respect to the distances $\{d_{i,n+2}^{(n+1)} \mid i = 1, \ldots, n + 1\}$, and the referential coordinate sets X_0 and Y_0, respectively. By Lemma 3.2 and $Y_0 \sim X_0$, we know that the location of y^* with respect to Y_0 is similar to that of x^* with respect to X_0. Hence, $Y_0 \cup \{y^*\} \sim X_0 \cup \{x^*\}$. In particular, for $i = 0, 1, \ldots, n$, $\|x_i^{(n+1)} - x^*\| = \|y_i^{(n+1)} - y^*\|$.

Furthermore, $x_{n+1}^{(n+1)} - x^*$ and $y_{n+1}^{(n+1)} - y^*$ are perpendicular to linear manifolds S_{X_0} and S_{Y_0}, respectively. For $i = 0, 1, \ldots, n$, we have

$$\begin{aligned}(d_{i+1,n+2}^{(n+1)})^2 &= \|x_{n+1}^{(n+1)} - x_i^{(n+1)}\|^2 \\ &= \|x_{n+1}^{(n+1)} - x^*\|^2 + \|x_i^{(n+1)} - x^*\|^2 ,\end{aligned}$$

and

$$\begin{aligned}(d_{i+1,n+2}^{(n+1)})^2 &= \|y_{n+1}^{(n+1)} - y_i^{(n+1)}\|^2 \\ &= \|y_{n+1}^{(n+1)} - y^*\|^2 + \|y_i^{(n+1)} - y^*\|^2 .\end{aligned}$$

Hence, $\|x_{n+1}^{(n+1)} - x^*\| = \|y_{n+1}^{(n+1)} - y^*\|$. This indicates $X^{(n+1)} \cup \{x^*\} \sim Y^{(n+1)} \cup \{y^*\}$, which implies $X^{(n+1)} \sim Y^{(n+1)}$. Therefore, the conclusion still holds for $k = n + 1$. □

THEOREM 3.3. *Let the index set* $I_k = \{0, 1, \ldots, k\}$. *Given a referential coordinate set* $X = \{x_i \mid i \in I_k\} \subset \Re^m$ *and a nonnegative vector* $d \in \Re_+^{k+1} = \{x \in \Re^{k+1} \mid x \geqslant 0\}$, *denote the Euclidean distance matrix associated with* X *by* D_0 *and the matrix*

$D = \left[\begin{smallmatrix} D_0 & d \\ d^T & 0 \end{smallmatrix}\right]$. *In the linear manifold* S_X, *let* x^* *be the unique solution to the linear system (7). Then, the following conclusions hold*:

(i) There exists a constant c *such that*

$$d_i^2 - \|x^* - x_i\|^2 = c\,, \quad \forall i \in I_k\,;$$

(ii) The matrix D *is a Euclidean distance matrix if and only if the constant* $c \geqslant 0$. *In particular, when* $k < m$, $c \geqslant 0$, *or* $k = m$, $c = 0$, *there exists a realization of the Euclidean distance matrix* D *in* $\Re^m$; *when* $k = m$, $c > 0$, *there exists a realization in* $\Re^{m+1}$.

Proof. (i) From the proof of Theorem 3.1, it is clear that the linear system (7) is equivalent to the following equations:

$$\|y - x_i\|^2 - d_i^2 = \|y - x_0\|^2 - d_0^2\,, \quad i = 1, 2, \dots, k\,.$$

Since $x^* \in S_X$ is the solution of the linear system (7), we know that the conclusion (i) holds for the constant $c = d_0^2 - \|x^* - x_0\|^2$.

(ii) ($\Leftarrow$) If $c = 0$, then, for any $k \leqslant m$, it is easy to see $\|x^* - x_i\| = d_i$, $i \in I_k$. In this case, the matrix D is a Euclidean distance matrix and $X \cup \{x^*\}$ is a realization of the matrix D.

If $c > 0$ and $k < m$, then the dimension of the linear manifold $S_X = x_0 + S_0$ (where $S_0 = span\{x_1 - x_0, \dots, x_k - x_0\}$) is less than m, and the complementary subspace $S_0^{\perp}$ of S_0 contains nonzero points. For any nonzero point $\tilde{x} \in S_0^{\perp}$, $x^* + \tilde{x}$ is also a solution of the equation (7). For $i \in I_k$, by the conclusion (i) and $x^* - x_i \in S_0$, we have

$$\begin{aligned} \|x^* + \tilde{x} - x_i\|^2 - d_i^2 &= \|\tilde{x}\|^2 + \|x^* - x_i\|^2 - d_i^2 \\ &= \|\tilde{x}\|^2 - c\,. \end{aligned}$$

This indicates that the matrix D is a Euclidean distance matrix, and $X \cup \{x^* + \tilde{x}\}$ becomes a realization of the matrix D if we choose $\tilde{x} \in S_0^{\perp}$ such that $\|\tilde{x}\|^2 = c$.

If $k = m$ and $c > 0$, then, $\forall i \in I_k$, let the vector $\hat{x}_i = \left[\begin{smallmatrix} x_i \\ 0 \end{smallmatrix}\right] \in R^{m+1}$. In this case, we consider a problem similar to the second case above, but in the $(m+1)$-dimensional space. Note that $S_X = \Re^m$ and it can be embedded into the space $\Re^{m+1}$. Denote the point $\hat{x} = \left[\begin{smallmatrix} x^* \\ \lambda \end{smallmatrix}\right] \in \Re^{m+1}$. By the conclusion (i), $\forall i \in I_k$, the Euclidean distance between $\hat{x}$ and $\hat{x}_i$ is equal to

$$\begin{aligned} \|\hat{x} - \hat{x}_i\| &= \|x^* - x_i\|^2 + \lambda^2 \\ &= d_i^2 - c + \lambda^2\,. \end{aligned}$$

Obviously, the matrix D is a Euclidean distance matrix, and $\{\hat{x}, \hat{x}_i \in \Re^{m+1} \mid i \in I_k\}$ becomes a realization of the matrix D if we choose $\lambda = c^{1/2}$.

($\Rightarrow$) If the matrix D is a Euclidean distance matrix, then there exists a realization of the matrix D, denoted by $\{v_i \mid i = 0, 1, \dots, k+1\}$, such that $\|v_i - v_j\| = D_{i+1,j+1}$. Since $k \leqslant m$, without loss of generality, suppose that $V = \{v_i \mid i \in I_k\} \subset \Re^m$ is a realization of the principal sub-matrix D_0 of the matrix D (If $v_i \in \Re^{m+1}$, we consider them in the m-dimensional linear manifold associated with V). Since X and V are two realizations of the matrix D_0, by Theorem 3.2, we know that the set V is also a referential coordinate set and $V \sim X$. Hence, the linear manifold S_X is isomorphic with the linear manifold S_V. Similar to the discussion in the proof of Theorem 3.2, we have

$$\begin{aligned} d_i^2 - \|x^* - x_i\|^2 &= d_i^2 - \|v^* - v_i\|^2 \\ &= \|v_{k+1} - v^*\|^2 , \quad i \in I_k , \end{aligned}$$

where $v^* \in S_V$ is the solution to the problem

$$\min_{v \in S_V} \|v_{k+1} - v\| .$$

This indicates that $c = \|v_{k+1} - v^*\|^2 \geqslant 0$. □

4. Complexity, unique and rigid conformations

4.1. THE COMPLEXITY OF THE EDM COMPLETION PROBLEM

Saxe [19] had prove that, embeddability of weighted graphs in k-space is strongly NP-hard. Based on this result, it is well known that, the EDM completion problem in the space $\Re$ is strongly NP-complete, and it is strongly NP-hard in $\Re^m$ for $m > 1$. An interesting phenomenon is described below:

> The instances [7, 11], which were used to prove the complexity of the EDM completion problem in $\Re^m$, can be solved easily in $\Re^{m+1}$, where $m \geqslant 1$.

We will demonstrate this interesting phenomenon for $m = 1$, and the similar idea can be extended to the instances used in Refs. [7, 11] for $m > 1$. First, we give a proof of the complexity of the EDM completion problem for $m = 1$. The idea is to reduce a well-known NP-complete integer partition problem to the EDM completion problem for $m = 1$ [7, 11, 16].

Given positive integers $a_1, \dots, a_n$, decide whether or not there exists a certain subset $S \subset N = \{1, \dots, n\}$ such that $\Sigma_{i \in S} a_i = \Sigma_{i \notin S} a_i$. Let us consider a partial matrix $D \in \Re^{n \times n}$ with the entries $d_{i,i+1} = d_{i+1,i} = a_i$ for $i \in N$ (the subscript $n + 1$ here and in the following lines is regarded as 1). We want to know whether or not there exists a realization $X = \{x_i \mid i \in N\} \subset \Re$ of the matrix D such that $|x_{i+1} - x_i| = a_i$ for all $i \in N$. Setting the set $S := \{i \mid a_i = x_{i+1} - x_i\}$, then the set X is a realization of the matrix D if and only if $\Sigma_{i \in S} a_i = \Sigma_{i \in N \setminus S} a_i$. This indicates that the EDM completion problem in $\Re$ is strongly NP-complete.

Next, we give an approach to decide if the above matrix D has a realization in $\Re^2$. Without loss of generality, let $n \geqslant 2$. Denote $A_k = \Sigma_{i \leqslant k} a_i$ for all $k \in N$. Let

$k^* = \max\{0, k \mid A_k \leqslant A_n/2\}$. If $k^* = 0$, or $k^* = n-1$ but $A_{n-1} < A_n/2$, then there is not any realization of the matrix D since $k^* = 0$ (or $k^* = n-1$ but $A_{n-1} < A_n/2$) indicates that $A_1 = a_1 > A_n/2$, i.e., $a_1 > \Sigma_{i>1} a_i$ (or $\Sigma_{i \leqslant n-1} a_i < a_n$), which is a contradiction with the property of the triangle inequality of the Euclidean distance. It is clear that there is a realization of the matrix D if $A_{k^*} = A_n/2$ for $k^* \geqslant 1$. When $1 \leqslant k^* < n-1$ and $A_{k^*} < A_n/2$. Define a symmetric matrix $\hat{D} \in \Re^{3\times 3}$ such that

$$\hat{d}_{ii} = 0\,, \quad i = 1, 2, 3\,,$$

$$\hat{d}_{12} = A_{k^*}\,,$$

$$\hat{d}_{23} = a_{k^*+1}\,,$$

$$\hat{d}_{13} = A_n - A_{k^*+1}\,.$$

There exists a realization of the matrix $\hat{D}$ in $\Re^2$, from which a realization of the matrix D can be easily obtained.

From the above description, we know that the exact complexity status of the general EDM completion problem remains an open problem. But there exists a special class of the EDM completion problems, whose graphic structure corresponding to the specified entries is chordal, that can be solved in polynomial time. There also exist some efficient algorithms to test if the corresponding graph has no K_4 minor, or can be obtained by means of clique sums from chordal graphs and graphs with no K_4 minor (see [16] and some references therein).

4.2. PROPERTIES FOR THE UNIQUE AND RIGID CONFORMATION

Given a partial symmetric matrix $D \in \Re^{n\times n}$, a graph $G_D = (V_n, E)$ (written by G, for short) is called *the associated graph* with the partial matrix D if it is on the vertex set $V_n = \{1, \ldots, n\}$ together with the edge set E, and an edge $(i, j) \in E$ for $i, j \in V_n$ if and only if the entry d_{ij} is specified. For the molecular problem with complete inter-atomic distances (i.e., the Euclidean matrix $D \in EDM_n$), the paper [8] gives a linear time algorithm to solve it, while the previous approaches rely on decomposing a Euclidean distance matrix or minimizing an error function and require $O(n^2)$ or $O(n^3)$ floating point operations. In this case, the associated graph with D is a complete graph and its realization is located in the three-dimensional space.

Based on the approaches for determining the location of a point in Section 3, we can obtain some properties about the minimal number of non-diagonal entries that needs to be specified in a partial Euclidean distance matrix for the EDM completion problem in order to assure the uniqueness or the rigidity of conformation. Note that similar properties can also be obtained for the PSD completion problem.

Usually, the results obtained by graph theory methods are very general for the EDM completion problem. In fact, the uniqueness and the rigidity of conformation for a partial Euclidean distance matrix are related closely to not only the topology structure in the associated graph, but also the numerical relationship among the specified entries of the matrix. The following results come from the exploration of the numerical relationship among the specified entries of a partial Euclidean distance matrix.

THEOREM 4.1. *Given a symmetric matrix* $D \in \Re^{n \times n}$ *with zero-diagonal and positive non-diagonal entries, where* $n > 1$, *then the following conclusions hold:*

(i) In order to decide whether or not the matrix D is a Euclidean distance matrix, the number of arithmetic operations needed is no more than $O(n^3)$. *In particular, if* $D \in EDM_n$, *then the number of arithmetic operations needed to find a realization of the matrix D is* $O(nk^2)$, *where k is the minimal dimension number of a linear manifold in which a realization of the matrix D is located.*

(ii) If $D \in EDM_n$, *then all of its realizations have the same conformation.*

Proof. (i) There are at least two methods that can be used to show the following assertion:

The number of arithmetic operations, which is needed to decide whether or not $D \in EDM_n$ is no more than $O(n^3)$.

One method is based on the covariance mapping (1) and the property (3). Note that computing the singular values of an $n \times n$ matrix can be done in at most $O(n^3)$ arithmetic operations [10]. By using the method of the matrix singular value decomposition, we can get a realization of the matrix D at most $O(n^3)$ arithmetic operations if $D \in EDM_n$ or claim that $D \notin EDM_n$.

Now, we present another point of view to show the above assertion. A similar idea for the molecular problem appears in [8]. From this viewpoint, if $D \in EDM_n$ and the dimension of a realization of D is bounded, then a linear time algorithm can be derived easily to find a realization of the matrix D. In addition, we can get even better result than $O(n^3)$ for the problem of deciding whether or not $D \in EDM_n$ (see Remark 4.1 below).

Based on Theorem 3.3, suppose that we have obtained two sets: one is a referential coordinate set $X_j = \{x_{i_0}, x_{i_1}, \ldots, x_{i_j}\}$ such that, for $l = 0, 1, \ldots, j$, the last $j - l$ entries of the vector x_{i_l} are zero; the other is $Y_q = \{y_{p_1}, \ldots, y_{p_q}\}$ (Y_q may be empty set at a certain stage), where each y_{p_t} belongs to the linear manifold S_{X_j} generated by the set X_j. A principal submatrix $D_{j,q}$ of the matrix D, which corresponds to the index set $\{i_l, p_t \mid l = 0, 1, \ldots, j,\ t = 1, \ldots, q\}$, is an Euclidean distance matrix. Given any index $s \neq i_l, p_t$, it is enough to perform $O(j^2 + qj)$ arithmetic operations in order to decide which one of the following assertions holds:

(i) There exists a point $x_{i_{j+1}}$ such that $X_j \cup \{x_{i_{j+1}}\}$ becomes a new referential coordinate set, or a point $x_{p_{q+1}}$ such that $x_{p_{q+1}} \in S_{X_j}$. Furthermore, the principal submatrix $D_{j+1,q}$ (or $D_{j,q+1}$) of the matrix D corresponding to the index set $\{i_l, p_t \mid l = 0, 1, \ldots, j+1,\ t = 1, \ldots, q\}$ (or $\{i_l, p_t \mid l = 0, 1, \ldots, j,\ t = 1, \ldots, q+1\}$) is a Euclidean distance matrix, where $i_{j+1} = s$ (or $p_{q+1} = s$).

(ii) The principal submatrix $D_{j+1,q}$ (or $D_{j,q+1}$) of the matrix D corresponding to the index set $\{i_l, p_t \mid l = 0, 1, \dots, j+1, t = 1, \dots, q\}$ (or $\{i_l, p_t \mid l = 0, 1, \dots, j, t = 1, \dots, q+1\}$) is not a Euclidean distance matrix. Therefore, D is not a Euclidean distance matrix.

Since $j \leqslant n-1$ and $q \leqslant n-2$, it is easy to see that, using at most $O(n^3)$ arithmetic operations, we can get a realization of the matrix D or claim that $D \not\in EDM_n$.

As for the second part of the conclusion (i), if $D \in EDM_n$, the matrix D must have a realization, we only need $O(j^2)$ arithmetic operations in order to decide which kind of points in the above first assertion exists. Let k denote the minimal dimension number of a linear manifold in which a realization of the matrix D is located. Then, the number of arithmetic operations needed to find a realization of the matrix D is just $O(k^3 + nk^2)$, i.e., $O(nk^2)$.

(ii) If the matrix $D \in EDM_n$, then, for any column index sets J of D, the principal submatrix D_J corresponding to J must also be a Euclidean distance matrix. Given a realization X of D, without loss of generality, let the number of elements in J is equal to $k+1$, where k is the dimension of a linear manifold in which the realization X is located. Hence, the subset $X_J := \{x_i \mid i \in J\}$ is a referential coordinate set. For any other realization Y of D, there must exist a subset $Y_0 \subset Y$ whose Euclidean distance matrix is the same as D_J (a proper order of elements in Y_0 may be considered in order to be consistent with D_J). By Theorem 3.2, we have $Y_0 \sim X_J$. Based on Theorem 3.1, the remaining points in $X \backslash X_J$ and $Y \backslash Y_0$ are uniquely determined by the distances from them to the points in X_J and Y_0, respectively. Therefore, $Y \sim X$. This indicates that all of realizations of the matrix D have the same conformation. □

REMARK 4.1. From the proof of Theorem 4.1, if the dimension of a realization for the matrix D (if it exists) is not greater than k, then we can decide whether or not $D \in EDM_n$ at most $O(n^2k)$ arithmetic operations. In particular, for the molecular problem, given a complete NMR (exact) distance data matrix, the task terminates at most $O(n^2)$ arithmetic operations.

DEFINITION 4.1. A partial Euclidean distance matrix $D \in \Re^{n\times n} (n > 1)$ is said to have the U-property in an m-dimensional linear manifold, if every non-diagonal entry specified in the matrix D is not zero, and there exists a principal submatrix sequence $\{D^{(i)}\}_{i=1}^{n-1}$ of D such that the conformation consistent with $D^{(i)}$ is unique in the same manifold, where $D^{(1)} \in \Re^{2\times 2}$, $D^{(n-1)} = D$ and $D^{(i)}$ $(i < n-1)$ is a proper principal submatrix of $D^{(j)}$ for $j > i$.

THEOREM 4.2. *Given a partial Euclidean distance matrix $D \in \Re^{n\times n}$, let m be the minimal dimension of a linear manifold in which the conformation consistent with D can be realized and D has the U-property. Then, the minimal amount $M(m, n)$ of non-diagonal entries specified in D should satisfy*

$$L(m,n) \leqslant M(m,n) \leqslant U(m,n)\,, \tag{9}$$

where $n \geqslant m+1$ *and*

$$L(m,n) = 2n + (m+1)(m-4)/2\,, \tag{10}$$

$$U(m,n) = (m+1)n - (m+1)(m+2)/2\,. \tag{11}$$

Proof. First, we prove by induction that the conclusion holds when $m = n-1$. In this case, we have $L(n-1,n) = U(n-1,n) = n(n-1)/2$. We only need to prove $M(n-1,n) = n(n-1)/2$. It is clear that, for $n = 2, 3$, the conclusion holds.

Suppose that the conclusion holds for the case $n = k$, i.e., the minimal amount $M(k-1,k)$ of non-diagonal entries specified in a $k \times k$ Euclidean distance matrix is equal to $k(k-1)/2$ in order to assure the uniqueness of the conformation, where the dimension of the conformation is equal to $k-1$. When $n = k+1$, if the conformation consistent with a partial $(k+1) \times (k+1)$ Euclidean distance matrix D is unique and has dimension k, then, based on Theorem 3.2 and 3.3, the conformation consistent with the upper left $k \times k$ submatrix D_0 of the matrix D must be unique and has the dimension $k-1$. By the induction hypothesis, the minimal amount $M(k-1,k)$ of non-diagonal entries specified in D_0 is equal to $k(k-1)/2$. Denote a realization of the matrix D_0 by $X_0 = \{x_i \mid i = 0, 1, \dots, k-1\}$. Using Theorem 3.3 again, we know that the entries $\{d_{k+1,i+1} \mid i = 0, 1, \dots, k-1\}$ of the matrix D should be specified properly in order to assure that the conformation consistent with D is unique and has dimension k. Hence, the minimal amount $M(k, k+1)$ of non-diagonal entries specified in D is equal to $M(k-1,k) + k = k(k+1)/2$. That is, the conclusion holds when $n = k+1$.

Next, we prove the conclusion holds for the general m and n. Given any realization $X = \{x_0, x_1, \dots, x_{n-1}\}$ of the matrix D in a certain m-dimensional linear manifold, which corresponds to the unique conformation. Since D has the U-property, there exists a principal submatrix sequence $\{D^{(i)}\}_{i=1}^{n-1}$ of D such that each $D^{(i)}$ corresponds to a unique conformation. Without loss of generality, suppose that $X^* = \{x_0, x_1, \dots, x_k\}$ is a realization of $D^{(k)}$ that includes a referential coordinate set in the above m-dimensional linear manifold and $\{x_{k+1}\} \cup X^*$ corresponds to the matrix $D^{(k+1)}$. It is clear that $k \geqslant m$. Based on the set X^*, Theorem 3.1 and Theorem 3.3, the point x_{k+1} is determined uniquely if and only if the following two conditions hold simultaneously:

(i) The distances between x_{k+1} and each point in a certain referential subset $\{x_{i_1}, \dots, x_{i_j}\} \subset X^*$ are given, that is, the corresponding entries in D are specified.
(ii) The point x_{k+1} belongs to the linear manifold generated by the above set $\{x_{i_1}, \dots, x_{i_j}\}$.

Since the conformation is unique, we have $2 \leqslant j \leqslant m+1$. This indicates that the minimal number $M(m,n)$ of non-diagonal entries specified in D should satisfy

$$M(m, n) \geqslant M(m, k+1) + 2(n-k-1)\,,$$
$$M(m, n) \leqslant M(m, k+1) + (m+1)(n-k-1)\,.$$

If the set $X^*\backslash\{x_k\}$ also includes an m-dimensional referential coordinate set, then we can deal with $X^*\backslash\{x_k\}$ similarly. Now, suppose that the conformation corresponding to $X^*\backslash\{x_k\}$ has dimension $m-1$ and is unique. In order to assure the uniqueness of the conformation associated with D, by Theorem 3.3, the minimal amount of non-diagonal entries specified and related to x_k in D should be equal to m. Hence, we have $M(m, k+1) = M(m-1, k) + m$. By induction, we have

$$M(m-1, k) \geqslant L(m-1, k) = 2k + m(m-5)/2\,,$$
$$M(m-1, k) \leqslant U(m-1, k) = mk - m(m+1)/2\,.$$

Therefore, we obtain

$$\begin{aligned} M(m, n) &\geqslant 2k + m(m-5)/2 + m + 2(n-k-1) \\ &= L(m, n)\,, \\ M(m, n) &\leqslant mk - m(m+1)/2 + m + (m+1)(n-k-1) \\ &\leqslant U(m, n)\,, \quad \text{(using the fact } k \geqslant m\text{)} \end{aligned}$$

where $L(m, n)$ and $U(m, n)$ are defined as (10) and (11), respectively. □

REMARK 4.2. For the molecular problem, if n atoms has a unique two-dimensional conformation and their partial Euclidean distance matrix has the U-property, then the minimal amount of the specified distances satisfies $2n-3 \leqslant M(m, n) \leqslant 3n-6$, where $n \geqslant 3$; if these atoms has a unique three-dimensional conformation and the corresponding distance matrix has the U-property, then the minimal amount of the specified distances satisfies $2n-2 \leqslant M(m, n) \leqslant 4n-10$, where $n \geqslant 4$.

REMARK 4.3. The result in Theorem 4.2 is just a necessary condition for the uniqueness of the conformation related to a partial Euclidean distance matrix which has the U-property. Given a certain number M of the specified entries, the entries specified must have a certain graphic structure and satisfy certain numerical relations in order to assure the unique conformation.

A conformation is said to be *rigid* if it can not be continuously deformed while still satisfying the distance constraints specified; otherwise, it is said to be *flexible*. Obviously, the existence of a rigid conformation corresponding to a partial Euclidean distance matrix does not imply the existence of the unique conformation, but the uniqueness of the conformation implies the rigidity. Furthermore, the rigidity of the conformation is related closely to the dimension of the linear manifold in which the conformation is realized. For example, a rigid conformation in the two-dimensional plane may not be rigid in the three-dimensional space. Under the assumption of the existence of the rigid conformation, by using the similar approaches as above, we can prove the following theorem about the minimal amount of non-diagonal entries specified in a partial Euclidean distance matrix, whose proof is omitted.

DEFINITION 4.2. A partial Euclidean distance matrix $D \in \Re^{n\times n}(n > 1)$ is said to have the R-property in an m-dimensional linear manifold, if every non-diagonal entry specified in the matrix D is not zero, and there exists a principal submatrix sequence $\{D^{(i)}\}_{i=1}^{n-1}$ of D such that the conformation consistent with $D^{(i)}$ is rigid in the same manifold, where $D^{(1)} \in \Re^{2\times 2}$, $D^{(n-1)} = D$ and $D^{(i)}$ $(i < n-1)$ is a proper principal submatrix of $D^{(j)}$ for $j > i$.

THEOREM 4.3. *Given a partial Euclidean distance matrix $D \in \Re^{n\times n}$, let m be the minimal dimension of a linear manifold in which the conformation consistent with D can be realized and D has the R-property. Then, the minimal amount $\tilde{M}(m, n)$ of non-diagonal entries specified in D should satisfy*

$$\tilde{L}(m, n) \leqslant \tilde{M}(m, n) \leqslant \tilde{U}(m, n)\,, \tag{12}$$

where $n \geqslant m + 1$ and

$$\tilde{L}(m, n) = \begin{cases} n-1\,, & \text{if } m = 1\,, \\ 2n + (m+1)(m-4)/2\,, & \text{if } m \geqslant 2\,, \end{cases} \tag{13}$$

$$\tilde{U}(m, n) = \begin{cases} n-1\,, & \text{if } m = 1\,, \\ mn - m(m+1)/2\,, & \text{if } m \geqslant 2\,. \end{cases} \tag{14}$$

Finally, based on Theorem 4.2 and the covariance mapping defined in (1), we give a corollary about the unique completeness of a partial correlation matrix, whose proof is easily obtained and is omitted.

COROLLARY 4.1. *Given a partial correlation matrix $P \in \Re^{n\times n}$, where $n > 1$, suppose that every non-diagonal entry specified in the matrix P is not one, and the matrix P can be completed uniquely. If the corresponding partial Euclidean distance matrix of P under the covariance mapping (1) can be realized in a linear manifold with the minimal dimension m $(n \geqslant m)$ and have the U-property, then, the minimal amount of non-diagonal entries specified in P should be equal to $M(m, n+1) - n$, where the function M satisfies (9)–(11).*

5. Concluding remarks

The Euclidean distance matrix (EDM) completion problem and positive semidefinite (PSD) matrix completion problem are considered in this paper. Approaches to determine the location of a point in a linear manifold are studied, which are based on the distances from the point to other points in a referential coordinate set. The location of a point (if it exists) in a linear manifold is independent of the coordinate system, and is only related to a certain referential coordinate set and the corresponding distances. Sufficient and necessary conditions for the existence of such a point are presented. An interesting phenomenon about the complexity of the EDM completion problem is described. Some properties about the uniqueness and rigidity of conformation, which are related to minimal amount of non-diagonal entries specified in a partial Euclidean distance matrix (or a partial PSD matrix), are also presented. Based on our results, we note that the difficulty of the EDM and PSD completion problems arises from data incompleteness, or from uncertainties between the graphic structure and numerical assignment about the non-diagonal entries specified in a partial EDM and PSD matrix, respectively.

Acknowledgements

The research of the first two authors was supported by the China Scholarship Council while the authors visited the Department of Industrial and Systems Engineering, University of Florida, Gainesville, USA. The research of the first author was also partially supported by Scientific Foundation of Tsinghua University under Grants 032701031 and 091101020. The research of the third author was partially supported by NSF under Grants DBI 9808210 and EIA 9872509.

References

1. Andrilli, S. and Hecker, D. (1993), *Elementary Linear Algebra*, *PWS-Kent Publishing Company*, *Boston*, *MA*.
2. Bakonyi, M. and Johnson, C.R. (1995), The Euclidean distance matrix completion problem. *SIAM Journal on Matrix Analysis and Applications* 16(2), 646–654.
3. Blaney, J.M. and Dixon, J.S. (1994), Distance geometry in molecular modeling. In: Lipkowitz, K.B. and Boyd, D.B. (eds.), *Reviews in Computational Chemistry*, VCH Publishers, New York, Vol. 5, pp. 299–335.
4. Blumenthal, C.M. (1970), *Theory and Applications of Distance Geometry*, Chelsea Publishing Co., Bronx, New York.
5. Borg, I. and Groenen, P. (1997), *Modern Multidimensional Scaling: Theory and Applications*, Springer, New York.
6. Brünger, A.T. and Nilges, M. (1993), Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy, *Quarterly Reviews of Biophysics* 26, 49–125.
7. Crippen, G.M. and Havel, T.F. (1988), *Distance Geometry and Molecular Conformation*, Research Studies Press, England; John Wiley and Sons, New York.
8. Dong, Q.F. and Wu, Z.J. (2001), A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances, to appear in *Journal of Global Optimization*.
9. Glunt, W., Hayden, T.L. and Raydan, M. (1993), Molecular conformations from distance matrices. *Journal of Computational Chemistry* 14(1), 114–120.
10. Golub, G.H. and Van Loan, C.F. (1989), *Matrix Computations*, 2nd ed. Johns Hopkins University Press, Baltimore, MD.

11. Hendrickson, B. (1990), *The Molecular Problem: Determining Conformation from Pairwise Distances*, Ph.D. thesis, Department of Computer Science, Cornell University.
12. Hendrickson, B. (1995), The molecule problem: exploiting structure in global optimization. *SIAM Journal of Optimization* 5(4), 835–857.
13. Johnson, C.R. (1990), Matrix completion problems: a survey. In: Johnson, C.R. (ed.), *Matrix Theory and Applications*, American Mathematical Society, Providence, RI, pp. 171–198.
14. Johnson, C.R. and Tarazaga, P. (1995), Connections between the real positive semidefinite and distance matrix completion problems. *Linear Algebra and Its Applications* 223/224, 375–391.
15. Laurent, M. (1998), A connection between positive semidefinite and Euclidean distance matrix completion problems. *Linear Algebra and Its Applications* 273, 9–22.
16. Laurent, M. (1998), A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In: Pardalos, P.M. and Wolkowitz, H. (eds.), *Topics in Semidefinite and Interior-Point Methods*, Fields Institute Communications, Vol. 18, American Mathematical Society, pp. 51–76.
17. Moré, J.J. and Wu, Z.J. (1997), Global continuation for distance geometry problems. *SIAM Journal on Optimization* 7(3), 814–836.
18. Pardalos, P.M. and Liu, X. (1998), A tabu based pattern search method for the distance geometry problem. In: Giannessi, F., Komlosi, S. and Rapcsak, T. (eds.), *New Trends in Mathematical Programming*, Kluwer Academic Publishers, Boston, MA, pp. 223–234.
19. Saxe, J.B. (1979), Embeddability of weighted graphs in *k*-space is strongly *NP*-hard. *Proceedings of the 17th Allerton Conference in Communications*, *Control and Computing*, pp. 480–489.
20. Schoenberg, I.J. (1938), Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* 44, 522–536.
21. Wells, C., Glunt, W. and Hayden, T.L. (1994), Searching conformational space with the spectral distance geometry algorithm. *Journal of Molecular Structure* 308, 263–271.
22. Yajima, Y. (2001), Positive semidefinite relaxations for distance geometry problems, *manuscript*.