Distance Matrix Embeddability Constraints and a Gaussian Noise Corruption Reversal Algorithm

Brian Chen
Department of Computer Science
Columbia University
New York, NY 10027
bc2924@columbia.edu

Yihan Shen
Department of Computer Science
Columbia University
New York, NY 10027
ys3524@columbia.edu

Andrew Yang
Department of Applied Physics
Columbia University
New York, NY 10027
ay2546@columbia.edu

Kevin Zhang
Department of Computer Science
Columbia University
New York, NY 10027
kyz2005@columbia.edu

Abstract

Euclidean distance matrices (EDMs) have important applications in both physical and computational science. However, the data collection process for the EDM can be noisy or error-prone, causing certain entries of the distance matrix to be incorrect or missing altogether. We call these events a corruption. In this work, we investigate corruptions on EDMs and provide a denoising algorithm in a special Gaussian noise case. First, we demonstrate corruptions to Euclidean distance matrices can preserve Euclidean embeddability. Since traditional methods to detect and reverse corruptions do not assume any model for the underlying noise model, such algorithms fail in these situations. We then consider a setting where symmetric additive Gaussian noise is applied to one row and column of the EDM and provide an algorithm to find the maximum likelihood estimate of the original distance matrix. We validate our algorithm on synthetically generated data.

1 Introduction

Distance matrices have important applications, especially in crystallography. When chemists synthesize new materials, they often are uncertain about their structures. Experiments to determine the exact positions of atom often require extreme amounts of energy and can destroy the sample. Instead, materials scientists use x-ray diffraction (XRD) [8] to obtain a pair distribution function (PDF) [9]. The PDF gives information about distances between pairs of atoms. More complex experiments exist to get the full distance matrix for atomic pairs. The distance can be any metric, but physicists are interested in the L_2 metric in Euclidean space.

However, experimental errors and limitations can give rise to “corruptions" in the matrix. Random noise in XRD experiments lead to random noise in the resulting distance matrices [13]. Insufficient energy in the experiments can also cause high-distance entries to be masked [9]. Similarly, one can consider a sensor network problem, where the distances between two sensors can only be measured if the two sensors are within a certain distance of each other. Such limitations can manifest as random noise corruption, masking of large entries, and missing distance entries altogether.

Suppose we observe a noisy realization $\tilde{D}$ of an underlying Euclidean distance matrix D. We are interested in the following questions:

COMS4774 Unsupervised Learning Final Project.

1. For which noise families is the corruption detectable?
2. If we know the location and family of the corruption, can we recover D?

2 Prior Work

Several popular algorithms for the Euclidean Distance Matrix (EDM) completion problem exist for settings with specific conditions. Most noteworthy are rank alternation [5], multidimensional unfolding (MDU) [11], and semidefinite relaxation (SMR) [12]. Rank alternation depends on the relationship between the rank of the EDM and the embedding dimension of the data. The algorithm performs singular value decomposition (SVD) and discards smaller singular values to comply with the rank constraint. On the other hands, SMR transforms the EDM completion problem into a convex optimization by relaxing the constraint on matrix rank.

Recently [15] proposed to extend the previously common rank alternation and semidefinite relaxation methods to handle multiple observations of noisy and incomplete distance matrices. Since the rank of any EDM for data in $\mathbb{R}^d$ is less than $d+2$, the authors instead formulate a Frobenius norm minimization problem with rank constraints. The results significantly improved individual and averaged approaches, especially in the cases of multiplicative noise.

Our method is fundamentally different than these approaches because we assume that the form of the corruption is known. Specifically, we study scenarios where noise is applied (symmetrically) to a single element or row and column. In the latter case, we reformulate the denoising problem as a unconstrained optimization which can be solved efficiently using numerical methods under certain conditions.

3 Embeddability Conditions

This section details a comprehensive list of conditions satisfied by Euclidean-embeddable distance matrices. Determining whether a matrix is embeddable in $\mathbb{R}^k$ can be done by checking if any condition below is violated. Though not explored further in this paper, certain conditions may be helpful in identifying which entries cause a corrupted distance matrix to fail to be Euclidean embeddable.

Definition 1. Given n distinct points $x_i \in \mathbb{R}^k$, define the distance matrix D entry-wise as $d_{ij} = ||x_i - x_j||_2^2$. We say D is embeddable in $\mathbb{R}^{k'}$ for $k' \leq k$ if there exists some $y_i \in \mathbb{R}^{k'}$ s.t. $d_{ij} = ||y_i - y_j||_2^2$.

Proposition 2. If the Gram matrix $G = -\frac{1}{2}HDH$ is symmetric positive semi-definite (PSD), the distance matrix D is Euclidean embeddable. The centering matrix is $H = I - (1/n)\mathbb{1}\mathbb{1}^T$ [6].

Definition 3. A matrix M is called conditionally negative semi-definite (CNSD) if $x^T M x \leq 0$ for all $x^T \mathbb{1} = 0$.

Corollary 4. A matrix D is CNSD iff D is a distance matrix.

Proof. Any vector v can be written as $v = p + \beta\mathbb{1}$, where $p^T\mathbb{1} = 0$. One can easily confirm $Hv = p$, meaning $v^T G v = -\frac{1}{2}(Hv)^T D(Hv) = -\frac{1}{2}p^T D p$. The proof is completed by Proposition 2. □

Definition 5. For any matrix M, the Hadamard exponent $M^{\circ p}$ is defined element-wise as $[M^{\circ p}]_{ij} m_{ij}^p$.

Lemma 6. A matrix M that is CNSD and has all entries $m_{ij} > 0$ satisfies M^p PSD for any $p < 0$ [4]. Furthermore, $\log \circ M$ is CNSD [3] where $\log \circ M$ is defined entry-wise by $[\log \circ M]_{ij} = \log(m_{ij})$. The converse of both statements is not necessarily true.

Proposition 7. Let λ_{n-2} be the second largest eigenvalue of a distance matrix D. Given only the main diagonal of D is 0, for $0 < \epsilon \leq |\lambda_{n-2}|$, $(D + \epsilon I)^{\circ p}$ is PSD for all real $p < 0$. Furthermore, $\log \circ (D + \epsilon I)$ is CNSD.

Proof. For finite distance matrices $D \in \mathbb{R}^{n\times n}$, D is CNSD, meaning for all $p^T\mathbb{1} = 0$, $p^T Dp \leq 0$. As D is symmetric, is has n eigenvalues $\lambda_0 \leq \cdots \leq \lambda_{n-1}$. The Courant-Fischer theorem [1] states

$$\lambda_k = \min_{\substack{T\subseteq\mathbb{R}^n \\ \dim(T)=k+1}} \max_{\substack{v\in T \\ v\neq 0}} \frac{v^T Dv}{v^T v}.$$

As the subspace $S = \{p \in \mathbb{R}^n : p^T\mathbb{1} = 0\}$ has dimension $\dim(S) = n-1$, for all $0 \leq k \leq n-1$,

$$\lambda_k = \min_{\substack{T\subseteq\mathbb{R}^n \\ \dim(T)=k+1}} \max_{\substack{v\in T \\ v\neq 0}} \frac{v^T Dv}{v^T v} \leq \min_{\substack{T'\subseteq S \\ \dim(T')=k+1}} \max_{\substack{p\in T' \\ v\neq 0}} \frac{p^T Dp}{p^T p} \leq 0.$$

This guarantees D has $n-1$ non-positive eigenvalues $\lambda_0, \cdots, \lambda_{n-2}$. As D is non-negative, for $D \neq 0$, $\mathbb{1}^T D\mathbb{1} > 0$, so D must have one positive eigenvalue λ_{n-1}. Finally, for $k = n-2$, we know

$$\lambda_{n-2} = \min_{\substack{T\subseteq\mathbb{R}^n \\ \dim(T)=n-1}} \max_{\substack{v\in T \\ v\neq 0}} \frac{v^T Dv}{v^T v}.$$

Assume for the sake of contradiction $\mathbb{1}$ is in the minimizing T^*. As $\frac{\mathbb{1}^T D\mathbb{1}}{\mathbb{1}^T\mathbb{1}} > 0$, this implies $\lambda_{n-2} > 0$, a contradiction. Thus, $\mathbb{1} \notin T^*$, so $T^* \subseteq S$, and $T^* = S$ as T^*, S are both dimension $n-1$. We conclude

$$\max_{\substack{p\in S \\ p\neq 0}} \frac{p^T Dp}{p^T p} = \lambda_{n-2} \leq 0,$$

so

$$\max_{\substack{p\in S \\ p\neq 0}} \frac{p^T (D+\epsilon I)p}{p^T p} = \epsilon - \lambda_{n-2}.$$

For $\epsilon \leq |\lambda_{n-2}|$, $D + \epsilon I$ is CNSD. As only the main diagonal of D is 0, for $\epsilon > 0$, $D + \epsilon I$ is positive in all entries. The proof is completed by Lemma 6. □

Proposition 8. If D is a distance matrix, $D^{\circ 1/p}$ has one positive eigenvalue and all other negative eigenvalues for $p > 1$ [7].

Proposition 9. Given a matrix D, define the $n-1 \times n-1$ matrix D' element-wise by $d'_{ij} = -(d_{i,j} + d_{i+1,j+1} - d_{i,j+1} - d_{i+1,j})$. D is a distance matrix iff D' is PSD [3].

Proposition 10. Let $\exp\circ(-\alpha D)$ be defined entry-wise by $[\exp\circ(-\alpha D)]_{ij} = \exp\{-\alpha d_{ij}\}$. D is a distance matrix with no row or column with all zeros iff $\exp\circ(-\alpha D)$ is positive definite.

Proof. The proof follows from Schoenberg's theorem [16]. □

Lemma 11. The matrices G, $(D+\epsilon I)^{\circ p}$, $\log\circ(D+\epsilon I)$, $D^{\circ 1/p}$, and D' can all be computed in $O(n^2)$ from D.

4 Corruption Bounds

In this section, we derive a few bounds on the distortions that can be applied while preserving Euclidean embeddability. This implies that with sufficiently small distortion to the original matrix D, we still may have an embeddable distance matrix D'. Without a model of what noise is applied to D to produce D', traditional techniques for de-noising an EDM will return the already-embeddable D'.

First, we bound how much moving each point $x_i \in \mathbb{R}^k$ by some $\delta_i \in \mathbb{R}^k$ changes the resulting distance matrix. Given the resulting distance matrix after moving the points and the maximum distance $\delta = \max_i ||\delta_i||_2$ any point has moved, we can bound the values of the entries of the original distance matrix. Techniques exist to then recover the original distance matrix [2].

Definition 12. Given a n-point dataset $x_i \in \mathbb{R}^d$ with distance matrix D_x and another n-point dataset $y_i \in \mathbb{R}^d$ with distance matrix D_y, let $\delta_i = y_i - x_i$. Define $T \equiv D_y - D_x$ to be the distortion between distance matrice.

Lemma 13. Let D_x be a distance matrix constructed from $x_1, \cdots, x_n \in \mathbb{R}^k$ and G_x be the corresponding Gram matrix of rank k. Let $R = \max_i ||x_i - \overline{x}||_2$ where $\overline{x}$ is the center of the set $\{x_i\}_{i\in[n]}$. Then, $R \leq \sqrt{k\rho(G_x)}$, where $\rho(M)$ is the spectral radius of matrix M. A tighter bound can be constructed using $R \leq \sqrt{\sum_i \lambda_i}$, where λ_i are the eigenvalues of G.

Proof. see Appendix A. □

Lemma 14. Let G_x, G_y be Gram matrices corresponding to distance matrices D_x of points $x_i \in \mathbb{R}^n$ and D_y of points $y_i \in \mathbb{R}^n$. Let $\delta = \max_i ||\delta_i||_2$ where $\delta_i = y_i - x_i$ and $R = \max_i ||x_i - \overline{x}||_2$. Then, $R \leq \rho(G_y) + 2\delta$.

Proof. see Appendix A. □

Lemma 15. Let $R = \max_i ||x_i - \overline{x}||_2$ and $\delta = \max_i ||\delta_i||_2$. Then, $\max\{-4R^2, 4\delta^2 - 8\delta R\} \leq t_{ij} \leq 4\delta^2 + 8R\delta$, where $t_{ij} = [T]_{ij}$ for the matrix in Definition 12.

Proof. See Appendix A. □

Theorem 16. If we only have access to D_y and know $\delta = \max_i ||\delta_i||_2$, then we can bound $\max\{-4R^2, 4\delta^2 - 8\delta R\} \leq t_{ij} \leq 4\delta^2 + 8R\delta$, where $R = \sqrt{k(\rho(G_y) + 2\delta)}$. Therefore, $[D_y]_{ij} - (4\delta^2 + 8R\delta) \leq [D_x]_{ij} \leq [D_y]_{ij} + \max\{4R^2, 8\delta R - 4\delta^2\}$.

Proof. Follows from Lemmas 13-15. □

Next, we provide bounds on how much we can change a single entry of a Gram and distance matrix while maintaining Euclidean embeddability. Though the bound on such distortions of the distance matrix (Theorem 18) is tight, the bound on the Gram matrix (Theorem 16) has room for improvement.

Theorem 17. Let G be a Gram matrix. Let N^{ij} be defined by $[N^{ij}]_{k\ell} = \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}$ for $i \neq j$. Then, $G + \delta N^{ij}$ is PSD as long as $\delta \leq 2\lambda_1/n$, where λ_1 is the second smallest eigenvalue of G.

Proof. We know $G = -\frac{1}{2}HDH$, and $H\mathbb{1} = 0$. Thus, $G\mathbb{1} = 0$ and $\mathbb{1}G = 0$. Consider any $v \in \mathbb{R}^n$. Decompose $v = p + \beta\mathbb{1}$, where $p^T\mathbb{1} = 0$. Then,

$$\begin{aligned} v^T G v &= p^T G p \\ &\geq \lambda_1 ||p||_2^2, \end{aligned}$$

and

$$\begin{aligned} v^T N v &= 2v_i v_j \\ &= 2(p_i + \beta)(p_j + \beta) \\ &\geq 2\left(p_i - \frac{p_i + p_j}{2}\right)\left(p_j - \frac{p_i + p_j}{2}\right) \\ &= -\frac{1}{2}(p_i - p_j)^2 \\ &\geq -\frac{1}{2}||p||_1^2 \\ &\geq -\frac{n}{2}||p||_2^2. \end{aligned}$$

Thus,

$$v^T(G + \delta N)v = \lambda_1 ||p||_2^2 - \frac{n\delta}{2}||p||_2^2,$$

which is positive as long as $\delta \leq \lambda_1/n$. □

Theorem 18. Let A be any matrix. Let N be defined by $n_{k\ell} = \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}$ for $i \neq j$. Then, $A + \delta N$ is positive semi-definite as long as $\delta \leq \lambda_0/(n-1)$, where λ_0 is the smallest eigenvalue of A.

Proof. For any $v \in \mathbb{R}^n$,

$$\begin{aligned} v^T N v &= 2v_i v_j \\ &\geq ||v||_2^2 - ||v||_1^2 \\ &\geq (1-n)||v||_2^2. \end{aligned}$$

Thus,

$$v^T(G + \delta N)v \geq \lambda_0 ||v||_2^2 + (1-n)\delta ||v||_2^2,$$

which is positive as long as $\delta \leq \lambda_0/(n-1)$. □

Theorem 19. Let the matrix N^{ij} for $i \neq j$ be defined by $[N^{ij}]_{k\ell} = \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}$. Then, $D + \delta N^{ij}$ is Euclidean embeddable as long as $-2\lambda_1 \leq \delta \leq \frac{2n}{2-n}\lambda_1$, where λ_1 is the second smallest eigenvalue of the Gram matrix $G = -\frac{1}{2}HDH$.

Proof. see Appendix A. □

Finally, below are two statements that are conjectured to be true and valid for all empirical experiments, but no formal proof has been found. The first can be used to give a relation between the Gram matrix and distance matrix distortions. The second can be used to construct a rank $k+1$ distance matrix D' from a rank k matrix D while ensuring k eigenvectors of the corresponding Gram matrices G' and G are shared.

Conjecture 20. Let α_i be the i-th (zero-indexed) largest eigenvalue of D and β_i be the i-th smallest eigenvalue of G. For $i > 0$, $\alpha_i + 2\beta_i < 0$.

Conjecture 21. Define $M_D^i \in \mathbb{R}^{n\times n}$ by $[M_D^i]_{jk} = \delta_{ij}(1-\delta_{jk})m^2 + \delta_{ik}(1-\delta_{jk})m^2$. Define $M_G^i \in \mathbb{R}^{n\times n}$ by $[M_G^i]_{jk} = \delta_{ij}\delta_{ik}(m(1-1/n))^2 + (m/n)^2\delta_{jk}(1-\delta_{ij}\delta_{ik})$. Let $D \in \mathbb{R}^{n\times n}$ be a distance matrix and $G = -(1/2)HDH$ be the corresponding Gram matrix. If rank$(G) < n$, then $G + M_G^i = -(1/2)H(D + M_D^i)H$, and rank$(G) - 1 \leq$ rank$(G + M_G^i) \leq$ rank$(G) + 1$.

5 Single Entry Gaussian Noise Corruption

In this section, we study a particular class of corruptions, where exactly one column and row is affected by symmetric additive Gaussian noise. Formally, we observe a noisy realization of the underlying "true" distance matrix D defined as

$$\tilde{D}_{ij} = \begin{cases} D & \text{for } i,j < n \\ D + \sigma_{ij} & \text{otherwise} \end{cases}$$

where $\sigma_{ij} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$. This case is motivated by scenarios where measurements to exactly one atom in a crystalline structure is noisy. Our noise model explicitly assumes that last row and column of the distance matrix are affected. However, it is worth noting that our results can be extended for any single noisy row and column by relabeling the points.

Under these conditions, we derive conditions for recovering the original distance matrix D and formulate the problem as an unconstrained optimization problem. If even one entry in the row or column is known exactly, we can find the global optimum of the optimization efficiently using Newton's method.

Proposition 22. Let D be a $n \times n$ symmetric, non-negative matrix with zero diagonal and $\mathbf{1}_n$ n-dimensional vector with all entries equal to 1. Then D is an EDM if and only if $F = -L^T DL$ is a PSD matrix, where $L = (I - e_1\mathbf{1}_n^T)$. Furthermore,

$$-L^T DL = \begin{pmatrix} X & y \\ y^T & 2D_{1n} \end{pmatrix},$$

where $X_{ij} = D_{1i} + D_{1j} - D_{ij}$ and $y_i = D_{1i} + D_{1n} - D_{in}$ for $i, j = 1, \ldots, n-1$. The matrix F is called a Gram matrix of D and the rank of F is equal to the embedding dimension of D.

Proof. see Appendix B □

Proposition 23. Let X and y be defined as above. Then, D is an EDM if and only if

1. X is PSD,
2. $2D_{1n} - y^T X^+ y \geq 0$,
3. $XX^+ y = y$.

Proof. see Appendix B □

Since the last row and column of D are corrupted, the rank of X is equal to the embedding dimension of the remaining $n-1$ data points. By our noise model, the corrupted data point occupies the same embedding space, the rank of the entire Gram matrix F must be the same as the rank of X. This implies specific conditions on D_{1n}.

Lemma 24. Let X and y be defined as above and suppose D is an EDM. The embedding dimension of D and X are equal if and only if $2D_{1n} = y^T X^+ y$.

Proof. We can express the rank of a block matrix in terms of the matrix components:

$$\begin{aligned}\text{rank}(F) &= \text{rank}\begin{pmatrix} X & y \\ y^T & 2D_{1n} \end{pmatrix} \\ &= \text{rank}(X) + \text{rank}(2D_{1n} - y^T X^+ y).\end{aligned}$$

For the embedding dimension of D and X to be equal, their ranks must be equal by Proposition 1. This holds if and only if $2D_{1n} = y^T X^+ y$. □

Lemma 25. Suppose X is a $n \times n$ PSD matrix and let $v_1, \ldots, v_k$ be the k eigenvalues with non-zero eigenvalue. Then, $XX^+ y = y$ if and only if $y \in \text{span}(v_1, \ldots, v_k)$.

Proof. Since X is PSD, we can decompose $X = U\Sigma U^T$, where Σ is the diagonal matrix of eigenvalues in descending order and the columns of U are the corresponding eigenvectors. Then, $U^+ = U\Sigma^+ U^T$, where

$$(\Sigma^+)_{ii} = \begin{cases} 1/(\Sigma)_{ii} & \text{if } \Sigma_{ii} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Let $I_{:k}$ be the $n \times n$ matrix whose first k diagonal elements are one and all other entries are zero. Then,

$$XX^+ = (U\Sigma U^T)(U\Sigma^+ U^T) = U(\Sigma\Sigma^+)U^T = UI_{:k}U^T.$$

Let $v_{k+1}, \ldots, v_n$ be the zero eigenvectors. For any eigenvector v_i with eigenvalue λ_i, we have $XX^+ v = \mathbf{1}[\lambda > 0]v$. By decomposing y,

$$\begin{aligned} XX^+ y = y &\iff XX^+ \sum_{i=1}^n a_i v_i = \sum_{i=1}^n a_i v_i \\ &\iff \sum_{i=1}^n \mathbf{1}[\lambda_i > 0] a_i v_i = \sum_{i=1}^n a_i v_i. \end{aligned}$$

The last equality holds if and only if $a_i = 0$ when $\lambda_i = 0$. Thus, y is in the span of the non-zero eigenvectors. □

Lemma 26. Let X and y be defined as above. Then, D is an EDM with the same embedding dimension as X if and only if

1. X is PSD,

2. $y^T X^+ y = 2D_{1n}$,

3. $y = Vw$,

where $w \in \mathbb{R}^k$ and $V \in \mathbb{R}^{(n-1)\times k}$ has columns which are the k non-zero eigenvectors of X. Moreover, $D_{in} = D_{ni} = D_{1i} + \frac{1}{2}y^T X^+ y - y_i$.

Proof. The main statement follows by rewriting the conditions in Proposition 2 using Lemma 3 and 4. By the definition and symmetry of D, we have $D_{in} = D_{ni} = D_{1i} + D_{1n} - y_i$. Using the second condition, we get $D_{ni} = D_{1i} + \frac{1}{2}y^T X^+ y - y_i$ □

We are now prepared to derive the MLE estimate of the corrupted distance matrix.

Definition 27. Given $X \in \mathbb{R}^{(n-1)\times(n-1)}$, $y \in \mathbb{R}^{n-1}$ and $D_{1n} = d \in \mathbb{R}$ satisfying the conditions in Lemma 5, define Gram(X, y, d) to be the Gram matrix

$$\text{Gram}(X, y, d) = \begin{pmatrix} X & y \\ y^T & 2d \end{pmatrix},$$

and EDM(X, y, d) to be the EDM D satisfying $-L^T DL = \text{Gram}(X, y, d)$

Theorem 28. Let D be an underlying $n \times n$ EDM and X be defined as above for D. Suppose $\tilde{D}$ is a noisy realization of D where the n-th row and column are corrupted by some additive Gaussian noise $\mathcal{N}(0, \sigma^2)$ with fixed variance.

Let (λ_i, v_i) be non-zero eigenvalue-vector pairs of X. Define $\Gamma = \text{Diag}(\sqrt{2\lambda_1}, \ldots, \sqrt{2\lambda_k})$ and $M = V\Gamma$. Furthermore, define $b \in \mathbb{R}^{(n-1)}$ where $b_i = \tilde{D}_{1i} - \tilde{D}_{in}$. Then, the maximum likelihood estimate of D is EDM$(X, Mz, \|z\|^2)$, where

$$z = \arg\min_{x\in\mathbb{R}^k} \left\| b + \|x\|^2 \mathbf{1}_{n-1} - Mx \right\|^2$$

Proof. We do not need to consider D_{nn} since we know the diagonal of an EDM is zero. Since each entry in the distance matrix is corrupted i.i.d.,

$$\begin{aligned} \log \mathcal{L}(D) &\propto \log P(D \mid \tilde{D}) \\ &\propto \log \prod_{i=1}^{n-1} e^{-(D_{in}-\tilde{D}_{in})^2/\sigma^2} \propto -\sum_{i=1}^{n-1} (D_{in} - \tilde{D}_{in})^2, \end{aligned}$$

where $\mathcal{L}$ is the likelihood. The maximum likelihood estimate of D is the minimizer of

$$f(D) = \sum_{i=1}^{n-1} (D_{in} - \tilde{D}_{in})^2$$

subject to the constraints in Proposition 5. Let $y = Vw = Mx$, where $x = \Gamma^{-1}w$. Then,

$$\begin{aligned} y^T X^+ y &= (Mx)^T X^+ (Mx) \\ &= x^T \Gamma^T (V^T X^+ V)\Gamma x \\ &= x^T \Gamma^T \text{Diag}(1/\lambda_1, \ldots, 1/\lambda_k)\Gamma x \\ &= x^T (2I) x \\ &= 2\|x\|^2 \end{aligned}$$

By the second constraint, $\|x\|^2 = D_{1n}$. Since only the n-th row and column are corrupted, $D_{1i} = \tilde{D}_{1i}$ for $1 \le i < n$. Then,

$$D_{in} - \tilde{D}_{in} = (D_{1i} + D_{1n} - y_i) - \tilde{D}_{in} = b_i + \|x\|^2 - (Mx)_i.$$

For $D = \text{EDM}(X, y, D_{1n}) = \text{EDM}(X, Mx, \|x\|^2)$,

$$f(D) = \sum_{i=1}^{n-1} (b_i + \|x\|^2 - (Mx)_i)^2 = \left\| b + \|x\|^2 \mathbf{1}_{n-1} - Mx \right\|^2.$$

Thus, the maximum likelihood estimate of D is given by EDM$(X, Mz, \|z\|^2)$, where z is the minimizer of $\left\|b + \|x\|^2 \mathbf{1}_{n-1} - Mx\right\|^2$. □

The above problem is non-convex because of the quartic term. However, if we know even one of the distances in the corrupt row or column, we can reformulate the problem as a quadratic program with quadratic constraints (QPQC). Though QCQP problems are NP-hard, we can use a rank relaxation to reforulate this problem as a SDP with a single constraint [10], solvable with error bound epsilon ϵ with time $O(n^{3.5}\log(1/\epsilon)$ [14]. In practice, we observe at least linear global convergence using Newton's method (Section 6).

Corollary 29. Let $\tilde{D}$ be a noisy realization of an underlying EDM D corrupted in the same manner as in Theorem 7. Suppose we know exactly the distance between point 1 and n (i.e. $\tilde{D}_{1n} = D_{1n}$). Then, the maximum likelihood estimate of D is unique is equal to EDM$(X, Mz, \|z\|^2)$, where

$$z = (\Gamma^2 + \lambda I)^{-1} M^T c,$$

and λ is the unique solution to

$$\sum_{i=1}^{k} \left[\frac{(M^T c)_i}{2\lambda_i + \lambda}\right]^2 = \tilde{D}_{1n}.$$

Proof. By Theorem 7, $\|x\|^2 = D_{1n} = \tilde{D}_{1n}$ is constant. Let $c = b + \|x\|^2 \mathbf{1}_{n-1}$. The maximum likelihood estimate of D is given by EDM$(X, Mz, \|z\|^2)$, where

$$\begin{aligned} z = \arg\min_{x \in \mathbb{R}^k} & \|c - Mx\|^2 \\ \text{subject to } & \|x\|^2 = \tilde{D}_{1n}. \end{aligned}$$

We can solve the constrained optimization using Lagrange multipliers. The Lagrangian is:

$$\mathcal{L} = \|c - Mx\|^2 + \lambda(\|x\|^2 - c),$$

where λ is the Lagrange multiplier. Setting the derivatives with respect to x and λ to zero,

$$\frac{\partial \mathcal{L}}{\partial x} = 2(-M^T)(c - Mx) + 2\lambda x = 2(M^T M + \lambda I)x - 2M^T c = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = x^T x - \tilde{D}_{1n} = 0.$$

Note that $M^T M = \Gamma^T V^T V \Gamma = \Gamma^2$. Solving the first equation gives $x = (\Gamma^2 + \lambda I)^{-1} M^T c$ and substituting into the second equation gives

$$(M^T c)^T (\Gamma^2 + \lambda I)^{-2} (M^T c) = \tilde{D}_{1n}.$$

Simplifying,

$$\begin{aligned} \tilde{D}_{1n} &= (M^T c)^T \text{Diag}(2\lambda_1 + \lambda, \ldots, 2\lambda_k + \lambda)^{-2} (M^T c) \\ &= \sum_{i=1}^{k} \left[\frac{(M^T c)_i}{2\lambda_i + \lambda}\right]^2. \end{aligned}$$

The solution for λ is unique since $\lambda_i > 0$, so $(2\lambda_i + \lambda)^{-2}$ is strictly decreasing in λ. The value of λ is then used to find the minimizing z. □

We can find the exact value of λ using any root-finding method, e.g. Newton-Raphson method.

6 Experimental Results

We evaluated the convergence behavior of the maximum likelihood estimation from Theorem 7 in a synthetic setting by simulating noise along an entire row and column. Figure 1 shows the behavior of negative log likelihood during during the optimization for different dataset sizes $N \in \{10, 25, 50, 100\}$. All experiments have unit Gaussian noise added to the original EDM and use stochastic gradient descent.

Empirically we have observed an exponential convergence rate for all sample sizes, and the likelihood stabilizes within 2000 iterations. The dashed lines are the likelihood corresponding to the original EDM, and we observe that in all of our experiments the MLE estimates recovers the matrix with likelihood better than the original. Our experimental results show that stochastic gradient descent yields reliable denoising and are consistent with Theorem 28.

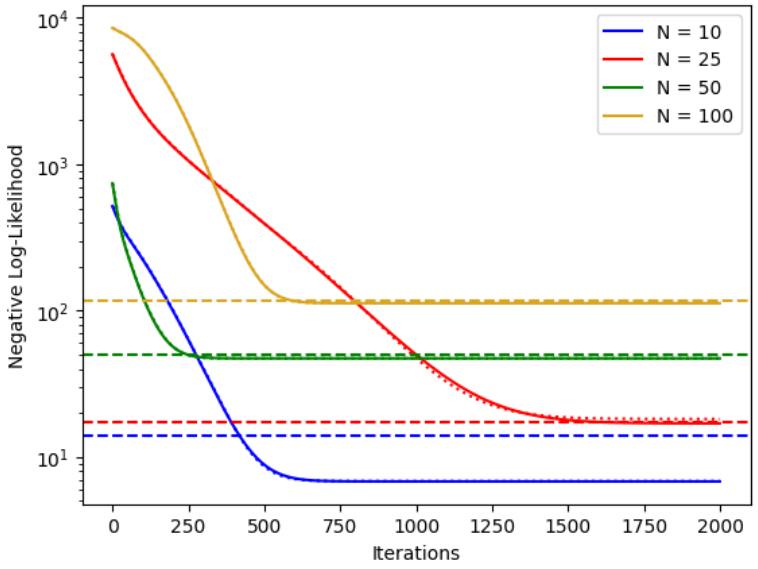

Figure 1: Convergence behavior of MLE for an EDM whose entire row and column are corrupted. Dashed lines represent the log-likelihood of the original, uncorrupted matrix. It is possible (see $N = 10$ case) that the log-likelihood of the matrix produced by our algorithm is lower than original. Dotted lines are exponential and superexponential decay fits. Fits all have > 0.99999 Pearson correlation with the original curves, providing empirical evidence of at least linear local convergence.

7 Conclusions and Future Work

In this paper, we have shown an algorithm to recover the original Euclidean distance matrix of a set of points (or a close estimate), given a corrupted version of the matrix. Under minor conditions on the location and type of noise, we can formulate an optimization problem for the maximum likelihood estimate of the original distance matrix. Such algorithms have applications in crystallography, where measurement error may lead to distance matrices corrupted in an analogous manner.

Future work can include progress on corruption identification, e.g. detecting the location(s) of a corruption given a distance matrix, correcting different types of corruptions beyond additive Gaussian noise, and correcting multiple corruptions within a single distance matrix. We might also consider augmentations to the problem setup, such as the unassigned distance problem, that may benefit from similar likelihood-based recovery techniques.

Acknowledgment

The authors would like to thank Prof. Nakul Verma, Edward Ri, and Noah Bergam for their helpful feedback throughout the course on this project.

References

[1] Haim Avron, Esmond Ng, and Sivan Toledo. *A Generalized Courant-Fischer Minimax Theorem*. 2008. URL: https://www.osti.gov/servlets/purl/1165117.

[2] Andrés David Báez Sánchez and Carlile Lavor. "On the estimation of unknown distances for a class of Euclidean distance matrix completion problems with interval data". In: *Linear Algebra and its Applications* 592 (2020), pp. 287–305. ISSN: 0024-3795. DOI: https://doi.org/10.1016/j.laa.2020.01.036. URL: https://www.sciencedirect.com/science/article/pii/S002437952030046X.

[3] Rajendra Bhatia. "Infinitely Divisible Matrices". In: *The American Mathematical Monthly* 113.3 (2006), pp. 221–235. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/27641890 (visited on 12/16/2024).

[4] Rajendra Bhatia and Tanvi Jain. *Mean matrices and conditional negativity*. URL: https://doi.org/10.13001/1081-3810.3256.

[5] Ivan Dokmanic et al. "Euclidean Distance Matrices: A Short Walk Through Theory, Algorithms and Applications". In: *CoRR* abs/1502.07541 (2015). arXiv: 1502.07541. URL: http://arxiv.org/abs/1502.07541.

[6] Ivan Dokmanic et al. "Euclidean Distance Matrices: Essential theory, algorithms, and applications". In: *IEEE Signal Processing Magazine* 32.6 (2015), pp. 12–30. DOI: 10.1109/MSP.2015.2398954.

[7] Nira Dyn, Timothy Goodman, and Charles A. Micchelli. "Positive powers of certain conditionally negative definite matrices". In: *Indagationes Mathematicae (Proceedings)* 89.2 (1986), pp. 163–178. ISSN: 1385-7258. DOI: https://doi.org/10.1016/S1385-7258(86)80004-X. URL: https://www.sciencedirect.com/science/article/pii/S138572588680004X.

[8] Khadija El Bourakadi, Rachid Bouhfid, and Abou el Kacem Qaiss. "Chapter 2 - Characterization techniques for hybrid nanocomposites based on cellulose nanocrystals/nanofibrils and nanoparticles". In: *Cellulose Nanocrystal/Nanoparticles Hybrid Nanocomposites*. Ed. by Denis Rodrigue, Abou el Kacem Qaiss, and Rachid Bouhfid. Woodhead Publishing Series in Composites Science and Engineering. Woodhead Publishing, 2021, pp. 27–64. ISBN: 978-0-12-822906-4. DOI: https://doi.org/10.1016/B978-0-12-822906-4.00010-4. URL: https://www.sciencedirect.com/science/article/pii/B9780128229064000104.

[9] Christopher L. Farrow and Simon J. L. Billinge. "Relationship between the atomic pair distribution function and small-angle scattering: implications for modeling of nanoparticles". In: *Acta Crystallographica Section A* 65.3 (May 2009), pp. 232–239. DOI: 10.1107/S0108767309009714. URL: https://doi.org/10.1107/S0108767309009714.

[10] Xinyue Huo and Ran Gu. *A Vectorized Positive Semidefinite Penalty Method for Unconstrained Binary Quadratic Programming*. 2024. arXiv: 2408.04875 [math.OC]. URL: https://arxiv.org/abs/2408.04875.

[11] Paul Hursky. "Multi-dimensional scaling (MDS), Euclidean distance matrices (EDMs), matrix completion and outlier identification in array calibration and beamforming". In: *The Journal of the Acoustical Society of America* 146 (Oct. 2019), pp. 3016–3016. DOI: 10.1121/1.5137452.

[12] Nathan Krislock. "Semidefinite Facial Reduction for Low-Rank Euclidean Distance Matrix Completion". PhD thesis. Jan. 2010.

[13] Long V. Le et al. "Noise reduction and peak detection in x-ray diffraction data by linear and nonlinear methods". In: *Journal of Vacuum Science & Technology B* 41.4 (June 2023), p. 044004. ISSN: 2166-2746. DOI: 10.1116/6.0002526. eprint: https://pubs.aip.org/avs/jvb/article-pdf/doi/10.1116/6.0002526/17969989/044004_1_6.0002526.pdf. URL: https://doi.org/10.1116/6.0002526.

[14] Zhi-quan Luo et al. "Semidefinite Relaxation of Quadratic Optimization Problems". In: *IEEE Signal Processing Magazine* 27.3 (2010), pp. 20–34. DOI: 10.1109/MSP.2010.936019.

[15] Sai Sumanth Natva and Santosh Nannuru. "Denoising and Completion of Euclidean Distance Matrix from Multiple Observations". In: *2024 National Conference on Communications (NCC)*. IEEE. 2024, pp. 1–6.

[16] T.R.L. Phillips, K.M. Schmidt, and A. Zhigljavsky. “Extension of the Schoenberg theorem to integrally conditionally positive definite functions”. In: *Journal of Mathematical Analysis and Applications* 470.1 (2019), pp. 659–678. ISSN: 0022-247X. DOI: https://doi.org/10.1016/j.jmaa.2018.10.032. URL: https://www.sciencedirect.com/science/article/pii/S0022247X18308552.

A Appendix A

A.1 Proof of Lemma 13

Proof. Given G_x is rank k, We know $G_x = X^T X$ for $X = [x_1 - \overline{x} \ \cdots \ x_n - \overline{x}]$ [6]. As G_x is PSD and symmetric, we can find n orthonormal eigenvectors $v_1, \cdots, v_n$ with corresponding non-negative eigenvalues $\lambda_1, \cdots, \lambda_n$. We can eigendecompose $G_x = V \Lambda V^T$ where $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$ and $V = [v_0 \ \cdots \ v_{n-1}]$. As $\lambda_i \geq 0$, $\Lambda = (\Lambda^{\circ 1/2})^T \Lambda^{\circ 1/2}$

$$X^T X = G_x = (\Lambda^{\circ 1/2} V^T)^T (\Lambda^{\circ 1/2} V^T).$$

There exists some arrangement of eigenvectors such that $X = \Lambda^{\circ 1/2} V^T$. Thus, for any $i \in [n]$,

$$\begin{aligned} \sum_{j=1}^{n} [x_j - \overline{x}]_i^2 &= \sum_{j=1}^{n} \sqrt{\lambda_i}^2 [u_i]_j^2 \\ &= \lambda_k. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^{n} ||x_i - \overline{x}||_2^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} [x_i - \overline{x}]_j^2 \\ &= \sum_{i=1}^{n} \lambda_i, \end{aligned}$$

so

$$R^2 = \max_i ||x_i - \overline{x}||_2^2 \leq \sum_{i=1}^{n} \lambda_i.$$

As each $\lambda_i \leq \rho(G_x)$ and G_x is rank k,

$$\sum_{i=1}^{n} \lambda_i \leq k \rho(G_x).$$

□

A.2 Proof of Lemma 14

Proof. We know $\delta_i = y_i - x_i$, so $\delta_i = (y_i - \overline{y}) - (x_i - \overline{x}) + (\overline{y} - \overline{x})$. Thus,

$$\begin{aligned} ||x_i - \overline{x}||_2 &= ||(y_i - \overline{y} - \delta_i + (\overline{y} - \overline{x})||_2 \\ &\leq ||y_i - \overline{y}||_2 + ||\delta_i||_2 + ||\overline{y} - \overline{x}||_2 \\ &= ||y_i - \overline{y}||_2 + ||\delta_i||_2 + \left|\left| \frac{1}{n} \sum_{j=1}^{n} \delta_j \right|\right|_2 \\ &\leq ||y_i - \overline{y}||_2 + ||\delta_i||_2 + \frac{1}{n} \sum_{j=1}^{n} ||\delta_j||_2 \end{aligned}$$

As $||\delta_i||_2 \leq \delta$ by definition and $||y_i - \overline{y}|| \leq \rho(G_y)$ by Lemma 13, $R = \max_i ||x_i - \overline{x}||_2 \leq \rho(G_y) + 2\delta$.

□

A.3 Proof of Lemma 15

Proof.

$$\begin{aligned} [D_y]_{ij} &= ||y_i - y_j||^2 \\ &= ||(x_i + \delta_i) - (x_j + \delta_j)||^2 \\ &= ||(x_i - x_j) + (\delta_i - \delta_j)||^2 \\ &= [D_x]_{ij} + ||\delta_i - \delta_j||^2 + 2\langle x_i - x_j, \delta_i - \delta_j \rangle, \end{aligned}$$

so

$$t_{ij} = ||\delta_i - \delta_j||^2 + 2\langle x_i - x_j, \delta_i - \delta_j\rangle.$$

Upper-bounding t_{ij} is simple since for $v < \nu$, by Cauchy-Schwartz,

$$\begin{aligned} ||v_i - v_j||^2 &= ||v_i||^2 + ||v_j||^2 + 2\langle v_i, v_j\rangle \\ &\leq 4\nu^2. \end{aligned}$$

Thus,

$$\begin{aligned} t_{ij} &\leq 4\delta^2 + 2||x_i - x_j||||\delta_i - \delta_j|| \\ &\leq 4\delta^2 + 8\delta R. \end{aligned}$$

Lower-bounding can be done by defining $a = x_i - x_j$ and $b = \delta_i - \delta_j$, where we have shown above that $\alpha \equiv ||a|| \leq 2R$ and $\beta \equiv ||b|| \leq 2\delta$. Then,

$$\begin{aligned} t_{ij} &= b^T b + 2a^T b \\ &= (a+b)^T(a+b) - a^T a. \end{aligned}$$

Thus, by reverse triangle inequality,

$$\begin{aligned} \min_{x_i, x_j, \delta_i, \delta_j} t_{ij} &= \min_{a,b} (a+b)^T(a+b) - a^T a \\ &= \min_{\alpha,\beta} (\alpha - \beta)^2 - \alpha^2 \\ &= \min_{\alpha,\beta} \beta(\beta - 2\alpha). \end{aligned}$$

Recall that our system is constrained by $\alpha \leq 2R$ and $\beta \leq 2\delta$. Letting $\gamma \equiv -\beta(\beta - \alpha)$,

$$\alpha = \frac{\beta}{2} + \frac{\gamma}{2\beta}.$$

Thus, for all β, $\max_\alpha \gamma$ occurs when α is maximized; therefore $\alpha = 2R$ maximizes γ and minimizes $\beta(\beta - 2\alpha)$. We can therefore eliminate α from the minimization to get

$$\begin{aligned} \min_{0\leq\alpha\leq 2R, 0\leq\beta\leq 2\delta} \beta(\beta - 2\alpha) &= \min_{0\leq\beta\leq 2\delta} \beta^2 - 4R\beta \\ &= \min_{0\leq\beta\leq 2\delta} (\beta - 2R)^2 - 4R^2, \end{aligned}$$

so $\beta = \min\{2\delta, 2R\}$, and

$$\begin{aligned} t_{ij} &\leq \max\{-4R^2, (2\delta - 2R)^2 - 4R^2\} \\ &= \max\{-4R^2, 4\delta^2 - 8\delta R\}. \end{aligned}$$

□

A.4 Proof of Theorem 18

Proof. Let

$$G' = -\frac{1}{2}H(D + \delta N)H.$$

For any v, decompose $v = p + \beta\mathbb{1}$, where $p^T\mathbb{1} = 0$. Then,

$$\begin{aligned} v^T G' v &= v^T \left(-\frac{1}{2}H(D + \delta N)H\right) v \\ &= v^T G v + \delta v^T \left(-\frac{1}{2}HNH\right) v. \end{aligned}$$

As $Hv = p$,

$$p^T G' p = p^T G p + \delta p^T \left(-\frac{1}{2} HNH \right) p.$$

For positive definiteness to hold, we require $0 \leq p^T G' p$. We know $G\mathbb{1} = 0$, so $p^T G p \geq \lambda_1 p^T p$. Therefore, if we restrict

$$\delta p^T \left(-\frac{1}{2} HNH \right) p \geq -\lambda_1 p^T p,$$

G' is PSD.

WLOG, let $(i, j) = (1, 2)$. Then, $n_{k\ell} = \delta_{1k}\delta_{2\ell} + \delta_{2k}\delta_{1\ell}$, and

$$-\frac{1}{2} HN^{ij} H = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix},$$

where

$$A = \begin{pmatrix} \frac{n-1}{n^2} & \frac{n-1}{n^2} - \frac{1}{2} \\ \frac{n-1}{n^2} - \frac{1}{2} & \frac{n-1}{n^2} \end{pmatrix},$$

and $B \in \mathbb{R}^{(n-2)\times 2}$ and $C \in \mathbb{R}^{(n-2)\times(n-2)}$ are matrices defined by

$$b_{ij} = \frac{n-2}{2n^2},\ c_{ij} = -\frac{1}{n^2}.$$

There is exactly one negative eigenvalue $\lambda_- = (2-n)/(2n)$ and exactly one positive eigenvalue $\lambda_+ = 1/2$. The corresponding eigenvectors are

$$v_- = \begin{pmatrix} 1 \\ 1 \\ -\frac{2}{n-2} \\ \vdots \\ -\frac{2}{n-2} \end{pmatrix},\ v_+ = \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Therefore,

$$\frac{2-n}{2n} \leq p^T \left(-\frac{1}{2} HN^{ij} H \right) p \leq \frac{1}{2}.$$

Thus,

$$-2\lambda_1 \leq \delta \leq \frac{2n}{2-n}\lambda_1$$

ensures G' is PSD. □

B Appendix B

B.1 Proof of Proposition 22

Proof. We provide a more generic result on characterization of EDMs:

$$D \text{ is EDM } \iff \exists s,\ s^T \mathbf{1}_n = 1, (I - \mathbf{1}s^T) \left(-\frac{1}{2} D \right) (I - s\mathbf{1}^T) \text{ is PSD.}$$

The $\implies$ direction is a well known result from [2].

For the $\impliedby$ direction, in fact, if there exists one such s that makes the right-hand side above PSD, then it must hold for any s.

To show this, we first claim that $\forall A \in \mathbf{R}^{n\times n}$, A is PSD implies $\forall B \in \mathbf{R}^{n\times n}$, B^TAB is PSD, which follows from definition of PSD: for any vector z, we have $z^TB^TABz = (Bz)^TA(Bz) \geq 0$.

Then, we observe that

$$\forall y, \bar{y}, y^T\mathbf{1} = 1 \text{ implies } (I - \mathbf{1}y^T)(I - \mathbf{1}\bar{y}^T) = I - \mathbf{1}\bar{y}^T - \mathbf{1}y^T + \mathbf{1}y^T\mathbf{1}\bar{y}^T = I - \mathbf{1}y^T$$

Therefore, if

$$\exists \bar{s}, \bar{s}^T\mathbf{1} = 1, (I - \mathbf{1}\bar{s}^T)\left(-\frac{1}{2}D\right)(I - \bar{s}\mathbf{1}^T) \text{ is PSD}$$

we can consider any $s, s^T\mathbf{1}_n = 1$, we have

$$(I - \mathbf{1}s^T)(\frac{-1}{2}D)(I - s\mathbf{1}^T) = (I - \mathbf{1}s^T)(I - \mathbf{1}\bar{s}^T)\left(-\frac{1}{2}D\right)(I - \bar{s}\mathbf{1}^T)(I - s\mathbf{1}^T),$$

which must also be PSD given our first claim. □

B.2 Proof of Proposition 23

Proof. First, we show the "only if" direction. Given D is EDM, from our proposition 1, we have matrix

$$F = -L^TDL = \begin{pmatrix} X & y \\ y^T & 2D_{1n} \end{pmatrix},$$

is PSD. Since PSD matrices have all of the principle minors greater or equal to zero, we have X must be PSD. Further, since F must have a non-negative determinant, and from Schur's formula on block matrices, we have det(F) = det(X) $\cdot(2D_{1n} - y^TX^+y) \geq 0$, which implies the second condition. Lastly, to show that $XX^+y = y$, we need to show that y is in the column space of X. Since D is EDM, we can write $D_{ij} = ||a_i - a_j||^2, \forall i, j$ for some $a_1, a_2 ... a_n$, WLOG, we can further place a_1 at the origin(i.e a_1 is the zero vector). Then, from our definition of X and y in Proposition 1, $X_{ij} = D_{1i} + D_{1j} - D_{ij} = 2{a_i}^Ta_j$ and $y_i = D_{1i} + D_{1n} - D_{in} = 2{a_i}^Ta_n$. Since we know the the rank of an EDM corresponding to points embedded in dimension d has at most rank d+2, we can express a_n in terms of $a_2 \dots a_{n-1}$.

For the forward direction, we use the inverse direction of proposition 1. Given that X is PSD, and $(2D_{1n} - y^TX^+y) \geq 0$, we have F must also be a PSD, which implies that D is EDM.

□