One-step Spectral Estimation for Euclidean Distance Matrix Approximation

Yicheng Li* and Xinghua Sun*
* School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China
E-mail: liych75@mail2.sysu.edu.cn; sunxinghua@mail.sysu.edu.cn

***Abstract*—This paper proposes and analyzes a new spectral estimator for Euclidean Distance Matrix Completion (EDMC) problem under Bernoulli sample model. This new estimator can be viewed as a simplified version of the well-known SVD (Singular Value Decomposition)-MDS (Multidimensional Scaling) approach by Drineas et al. as it reduces the computational cost while maintaining almost the same performance. Its approximation accuracy, when measured by the spectral norm, is analyzed under the presence of zero-mean, entry-wise independent, sub-Gaussian additive noise. Theoretical results show that the approximation error is well-controlled up to the noise amplification term as soon as the sample complexity reaches the order of $\mathcal{O}(\log^2 n/n)$. Numerical simulations verify the above assertions.**

I. Introduction

We are interested in the reconstruction of a point set configuration from its sampled pair-wise distance measurements, known as the Euclidean Distance Matrix Completion (EDMC) problem [1]. The EDMC serves as a fundamental problem in a variety of fields, including but not limited to data visualization and dimensionality reduction in machine learning [2, Sec. 3.1.1] [3], wireless channel estimation [4], phase retrieval with unknown sensing basis [5], computational biochemistry [6] [7] [8], operating wireless sensor networks [9] [10] and articulated robots [11]. The EDMC is known to be NP-hard without prior restrictions [12], yet practical solutions with either deterministic type guarantees [9] [13] or performance bound under random point set model [10] exist. Recently, [14] proved that if the underlying point set satisfies the incoherent assumption, i.e., a prevalent measure of the degeneracy of the problem in low-rank matrix recovery context, and the distance samples are taken uniformly at random, then the EDM can be exactly recovered with near-optimal sample complexity via the nuclear norm minimization paradigm [15].

A. Mathematical Setup

Consider a ground truth point set consists of n points, and embedded in the $\mathbb{R}^d$ space, $d = 2, 3 \ll n$. Denote its configuration by $\mathbf{Y}^\star = [\mathbf{p}_1^\star, \dots, \mathbf{p}_n^\star]^T \in \mathbb{R}^{n\times d}$, and assume the Gram matrix $\mathbf{G}^\star := \mathbf{Y}^\star\mathbf{Y}^{\star T}$ has rank d. Let $d_{ij}^\star$ be the ground truth distance between the (i,j) point, then the symmetric matrix $\mathbf{D}^\star \in \mathbb{R}^{n\times n}$, which (i,j)-th entry is $(d_{ij}^\star)^2$ is the so-called EDM, i.e., $\mathbf{D}^\star \in \mathbb{EDM}^n$. Mathematically, the observed entries in the EDM satisfies

$$\mathbf{D}_{ij}^\star = (d_{ij}^\star)^2 = \mathrm{tr}(\mathbf{G}^\star\boldsymbol{\omega}_{\boldsymbol{\alpha}}), (i,j) \in \Omega, \tag{1}$$

where tr is the matrices inner product in $\mathbb{R}^{n\times n}$, i.e, $\mathrm{tr}(\mathbf{A}\mathbf{B}^T) = \langle \mathbf{A}, \mathbf{B}\rangle$. $\boldsymbol{\alpha} = (i,j)$ is a tuple, and $\boldsymbol{\omega}_{\boldsymbol{\alpha}} := (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$, where $\mathbf{e}_i \in \mathbb{R}^n$ denotes the i-th canonical Euclidean basis. Ω is the sample set and we assume that it is generated by the widely used symmetric Bernoulli rule, that is $\Omega := \{(i,j)|\delta_{\boldsymbol{\alpha}} = 1, 1 \le i < j \le n\}$, where $\delta_{\boldsymbol{\alpha}}$ is a sequence of i.i.d. 0/1 Bernoulli random variables with $\mathbb{P}(\delta_{\boldsymbol{\alpha}} = 1) = p$. And we let the operator $\mathcal{P}_\Omega$ denote the projection onto the sample set

$$\mathcal{P}_\Omega(\cdot) := \sum_{\boldsymbol{\alpha}\in\mathbb{I}} \delta_{\boldsymbol{\alpha}}(\langle \cdot, \mathbf{e}_i\mathbf{e}_j^T\rangle \mathbf{e}_i\mathbf{e}_j^T + \langle \cdot, \mathbf{e}_j\mathbf{e}_i^T\rangle \mathbf{e}_j\mathbf{e}_i^T), \tag{2}$$

where $\mathbb{I} := \{(i,j)|1 \le i < j \le n\}$[1]. This paper details a specific spectral method to approximately solve the EDMC under the setup of (1) (2) and provides a theoretical characterization of the reconstruction error of the point set. We will also demonstrate that such coarse guess can be used to initialize those non-convex successive refinements discussed in our previous work [16] and to achieve high-accuracy recovery of the point set numerically, which might shed some light on the guaranteed non-convex EDMC algorithms design.

Notations: **Bold** face lower letters and capital letters represent vectors or matrices, respectively. $\mathbf{1}$, $\mathbf{0}$, $\mathbf{I}$ is the vector of all ones, all zeros, and the identity matrix, respectively. $\mathbf{A}_{ij}$, $\mathrm{diag}(\mathbf{A})$ is the (i,j)-th entity, and the column vector formed by the diagonal elements of $\mathbf{A}$, respectively. The adjoint of a linear transformation $\mathcal{T}$ is denoted by $\mathcal{T}^*$ while $\mathcal{I}$ denotes the identity operator. $\|\mathbf{A}\|$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_\infty$ stand for the spectral, Frobenius, and entrywise l_∞ norm of $\mathbf{A}$. $\|\mathbf{a}\|_2$ is the vector l_2 norm. C, c are positive universal constants and may differ from line to line. $a \lesssim b$ and $a \gtrsim b$ stands for $a < Cb$ and $a > Cb$, respectively. $\mathbf{A} \succcurlyeq \mathbf{0}$ means that $\mathbf{A}$ is positive semidefinite.

Organization: In Section I-B we state our main findings while in Section I-C we briefly review the most related literature and provide a few remarks on the main result. Numerical tests that support our claims and a detailed proof are referred to Section II and Section III, respectively. This paper is concluded in Section IV.

B. Main Result

We propose the following one-step spectral estimator to approximate the EDM, i.e., to recover the point set configuration

[1]We note that since any EDM is a hollow diagonal matrix, the diagonal samples can be automatically dropped out without introducing perturbation.

up to translation and rotation ambiguity

$$\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T = \hat{\mathbf{G}}_\natural = \mathcal{T}_d\left[-\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J}\right], \tag{3}$$

where $\mathcal{T}_d$ is the rank-d singular value reconstruction and $\mathbf{J} := \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. $\hat{\mathbf{D}} = \mathbf{D}^\star + \mathbf{E}$ stands for the noisy observation and $\mathbf{E}$ is a symmetric noise matrix. We assume that $\{\mathbf{E}_{ij}\}_{i>j}$ are independent, zero-mean random variables with sub-Gaussian norm [17, Sec. 2.5] bounded by σ_n. Our main result controls the distortion of $\hat{\mathbf{G}}_\natural$ from $\mathbf{G}^\star$, which depends on the standard coherence parameter [15, Def. 1.2] of the ground truth.

Definition I.1 (Coherence)**.** *Let $\mathbf{G}^\star = \mathbf{U}^\star\mathbf{\Sigma}^\star\mathbf{U}^{\star T}$ denotes the rank-d thin Singular Value Decomposition (SVD) of $\mathbf{G}^\star$. Let the coherence parameter be defined as*

$$\mu^2 := \max_{i\in[n]} \frac{n}{d}\|\mathbf{e}_i^T\mathbf{U}^\star\|_2^2.$$

It is straightforward to check that $1 \le \mu^2 \le \frac{n}{d}$.

Theorem I.1 (Main)**.** *For a fixed, self-centered[2] point set configuration $\mathbf{Y}^\star$, under the setup of* (1) (2) (3) *and Definition I.1, for some $0 < \epsilon \le 1$, we have*

$$\|\hat{\mathbf{G}}_\natural - \mathbf{G}^\star\| \lesssim \epsilon\|\mathbf{G}^\star\| + \sigma_n\sqrt{\frac{n}{p}}, \tag{4}$$

holds with probability at least $1 - n^c$ as soon as $p \gtrsim \max\{\log^2 n, \mu^4 d^2/\epsilon^2\}/n$. The proof is referred to Section III.

C. Related Works

A detailed review of prior arts for the EDMC problem is far beyond the scope of this article, thus we choose only to discuss those that shared the most resemblance with us and refer interested readers to [1] [18] for the whole picture. The low-rank property of the EDM has been known for decades and might induce the initial interest in studying the low-rank recovery models [19]. By substituting any point set $\mathbf{Y} \in \mathbb{R}^{n\times d}$ into (1), and set $\Omega = \mathbb{I}$, i.e., a full observation, one finds

$$\mathbf{D} = g(\mathbf{G}) := \mathrm{diag}(\mathbf{G})\mathbf{1}^T + \mathbf{1}\mathrm{diag}(\mathbf{G})^T - 2\mathbf{G}, \tag{5}$$

where $\mathbf{G} = \mathbf{Y}\mathbf{Y}^T$. This yields the so-called EDM forward mapping g. From (5), it is straightforward to check that any EDM has rank no larger than $d+2$ by the utilization of rank inequality. This fact gives rise to the SVD-MDS algorithm, proposed by Drineas et al. [20], which directly combines the idea from standard spectral method for Low-rank Matrix Completion [21, Sec. 3.8] (SVD, the first step) and the Multidimensional Scaling [2, Sec. 3.1] (MDS, the second step). It reads

$$\tilde{\mathbf{D}} = \mathcal{T}_{(d+2)}\left[\frac{1}{p}\mathcal{P}_\Omega\hat{\mathbf{D}}\right], \hat{\mathbf{G}} = \mathcal{T}_d\left[g^+(\tilde{\mathbf{D}})\right], \tag{6}$$

where g^+ is the pseudo-inverse of g (which will be introduced shortly). The performance guarantee of SVD-MDS was later refined in [22] by leveraging the moment-control Matrix Bernstein inequality [23, Thm. 6.2] and a perturbation bound on MDS. They also proved that the first step in (6) for EDMC is minimax suboptimal when the noise is small. While our approach removes the rank-$(d+2)$ SVD thus reduces the computation cost approximately by half when n is small[3]. Moreover, when compared with [22, Thm. 2], Theorem I.1 directly establishes the reconstruction error of the point set with near-optimal sample complexity[4], and it is more general since we do not require the noise to be i.i.d. Gaussian. Beyond Theorem I.1, we have also observed that the performance gap between (3) and (6) is small in the numerical test, i.e., a reasonable compromise for the reduced computational complexity.

(3) is actually inspired by [14] [24], in which they introduced the dual-basis representation method to reformulate the forward and inverse EDM mapping. To proceed, we need the explicit expression of g^+

$$\mathbf{G} = g^+(\mathbf{D}) := -\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J}, \mathbf{D} \in \mathbb{EDM}^n, \mathbf{G} \succcurlyeq \mathbf{0}. \tag{7}$$

This map is well-defined on $\{\mathbf{1}\}^\perp$ [1, Thm. 2], and we call $\mathbf{J}$ the geometry centering matrix as it forces $\mathbf{G}\mathbf{1} = \mathbf{0}$. [14] and the following up works showed that (5) (7) have alternative forms by noticing that

$$\begin{aligned}\mathcal{I}_{\{\mathbf{1}\}^\perp}(\cdot) = g^+g(\cdot) &= -\frac{1}{2}\mathbf{J}\left[\sum_{\boldsymbol{\alpha}\in\mathbb{I}}\langle\cdot,\boldsymbol{\omega}_{\boldsymbol{\alpha}}\rangle(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)\right]\mathbf{J}\\ &= \sum_{\boldsymbol{\alpha}\in\mathbb{I}}\langle\cdot,\boldsymbol{\omega}_{\boldsymbol{\alpha}}\rangle\boldsymbol{\nu}_{\boldsymbol{\alpha}},\end{aligned} \tag{8}$$

where $\mathcal{I}_{\{\mathbf{1}\}^\perp}$ is the identity operator but restricted to $\{\mathbf{1}\}^\perp$, and $\boldsymbol{\nu}_{\boldsymbol{\alpha}} := -\frac{1}{2}\mathbf{J}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T)\mathbf{J}$ is the dual-basis of $\boldsymbol{\omega}_{\boldsymbol{\alpha}}$. Some linear algebra verifies this claim by checking that

$$\mathrm{tr}(\boldsymbol{\omega}_{\boldsymbol{\alpha}}\boldsymbol{\nu}_{\boldsymbol{\beta}}) = \begin{cases}1, & \boldsymbol{\alpha} = (i,j) = (k,l) = \boldsymbol{\beta}\\ 0, & \text{otherwise}\end{cases}.$$

In other words, the forward mapping can be established by the basis $\boldsymbol{\omega}_{\boldsymbol{\alpha}}$, and the inverse mapping is encoded implicitly in $\boldsymbol{\nu}_{\boldsymbol{\alpha}}$. One interesting operator originated from (8) is its sampled version, defined as follows

$$\mathcal{R}_\Omega := g^+\mathcal{P}_\Omega g = \sum_{\boldsymbol{\alpha}\in\mathbb{I}}\delta_{\boldsymbol{\alpha}}\langle\cdot,\boldsymbol{\omega}_{\boldsymbol{\alpha}}\rangle\boldsymbol{\nu}_{\boldsymbol{\alpha}}.$$

It is straightforward to check $\mathbb{E}\frac{1}{p}\mathcal{R}_\Omega(\mathbf{G}) = \mathbf{G}$ holds for self-centered $\mathbf{G}$ due to (8). Thus, one may then conjecture that $\frac{1}{p}\mathcal{R}_\Omega(\mathbf{G})$ concentrate around its mean nicely, i.e., the rank-d SVD of $\frac{1}{p}\mathcal{R}_\Omega(\mathbf{G})$ is close to the unperturbed matrix $\mathbf{G}$. After simple linear algebra, this idea gives rise to (3).

[24] proposed another spectral estimator for the same purpose. They use SVD to truncate the matrix after applying

[2]Without loss of generality, we assume that point sets in this work are always self-centered, i.e., $\mathbf{Y}^T\mathbf{1} = \mathbf{0}$. Otherwise, one can perform the projection $\mathbf{Y}' = \mathbf{J}\mathbf{Y}$ to force this property.

[3]We note that the main computation cost in either (3) or (6) comes from the thin SVD when n is small.

[4]The $\log^2 n$ term in Theorem I.1 comes from controlling the spectral norm of $\frac{1}{p}\mathcal{P}_\Omega\mathbf{E}$. Under noiseless situation, the bound can be improved to $p \gtrsim \max\{\log n, \mu^4 d^2/\epsilon^2\}/n$, which is only sub-optimal in d.

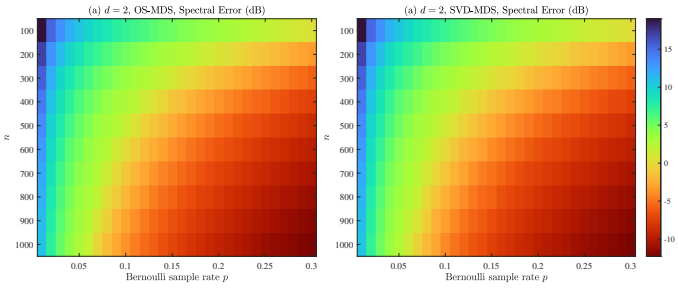

Fig. 1. The spectral error (in dB) of OS-MDS and SVD-MDS when varying Bernoulli sample rate p and the number of points n. For each pair (n, p), the result is the median value of 1000 independent trials due to the limited iteration cost when evaluating the spectral norm, i.e., to exclude outliers.

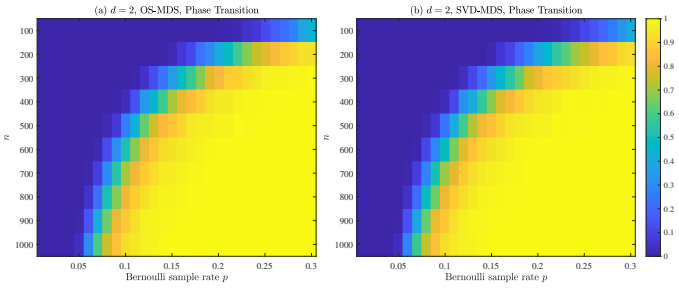

Fig. 2. The phase transition of OS-MDS and SVD-MDS when varying Bernoulli sample rate p and the number of points n. To verify Theorem I.1, we claim a success if the spectral error falls below 1. For each pair (n, p), the result is obtained by 1000 independent trials.

a self-dual operator $\mathcal{R}_\Omega^*\mathcal{R}_\Omega = g^*\mathcal{P}_\Omega g^+ g^+ \mathcal{P}_\Omega g$. The explicit expression of this scheme is given as follows

$$\hat{\mathbf{G}} = \mathcal{T}_d\left[\frac{1}{p^2} g^*\mathcal{P}_\Omega g^+ g^+ \mathcal{P}_\Omega(\hat{\mathbf{D}})\right]. \tag{9}$$

Even though (9) has almost the same computation cost as (3), the composition of $\mathcal{R}_\Omega^*\mathcal{R}_\Omega$ makes theoretical analysis challenging since g^+g^+ is not a diagonal operator. The non-orthonormality of $\boldsymbol{\nu}_{\boldsymbol{\alpha}}$ brings crossing term into $\mathcal{R}_\Omega^*\mathcal{R}_\Omega(\mathbf{G})$, and, the reuse of the sample set raises complicated dependency structure. This statement can be verified by expanding the terms, we have

$$\begin{aligned}\mathcal{R}_\Omega^*\mathcal{R}_\Omega(\mathbf{G}) = &\sum_{\boldsymbol{\alpha}=\boldsymbol{\beta}} \delta_{\boldsymbol{\alpha}}\langle \mathbf{G}, \boldsymbol{\omega}_{\boldsymbol{\alpha}}\rangle \|\boldsymbol{\nu}_{\boldsymbol{\alpha}}\|_F^2 \boldsymbol{\omega}_{\boldsymbol{\alpha}} \\ &+ \sum_{\boldsymbol{\alpha}\neq\boldsymbol{\beta}} \delta_{\boldsymbol{\alpha}}\delta_{\boldsymbol{\beta}}\langle \mathbf{G}, \boldsymbol{\omega}_{\boldsymbol{\alpha}}\rangle \langle \boldsymbol{\nu}_{\boldsymbol{\alpha}}, \boldsymbol{\nu}_{\boldsymbol{\beta}}\rangle \boldsymbol{\omega}_{\boldsymbol{\beta}},\end{aligned} \tag{10}$$

where we use the fact that $\delta_{\boldsymbol{\alpha}}^2$ has the same distribution as $\delta_{\boldsymbol{\alpha}}$. The first term in (10) introduces bias when measured on the population level, while the second term indicates that $\mathcal{R}_\Omega^*\mathcal{R}_\Omega$ has limited randomness structure and we leave this analysis to our future work.

II. Numerical Simulations

We present numerical results to verify the claimed performance of the proposed new estimator (3), which is referred to as OS-MDS (One Step-Multidimensional Scaling) in the following sections. The comparison between OS-MDS and SVD-MDS [20] is conducted in both noiseless and noisy tests over synthetic point sets, while OS-MDS initialized gradient refinement is also tested over four protein data downloaded from the Protein Data Bank [25].

A. Implementation Details

The synthetic ground truth point set $\mathbf{Y}^\star$ used in this study is generated by standard Gaussian distribution, while the sample mask Ω is forced to be symmetric, i.e., to satisfy (2). The noise $\mathbf{E}$ is a symmetric, hollow diagonal, Gaussian matrix, i.e., $\mathbf{E}_{ij} \sim \mathcal{N}(0, \sigma_n^2)$ with σ_n^2 varied for different noise intensities. The SVD truncation step is implemented by MATLAB `svds` function. If not specified otherwise, under each different scenario setup, we run these two spectral estimators for 1000 independent trials. The non-convex refinement discussed in [16], which takes the approximation result obtained by OS-MDS and SVD-MDS as initialization and perform Gradient Descent (GD) to optimize the s-stress function ((11) below) is also tests here. The GD solver comes from Manopt [26] default `steepestdescent` function, whose step size is chosen according to the widely used Armijo rule. Except the default stopping criteria, the gradient refinement is terminated if either it reaches the maximum iteration (800 times), or the Frobenius norm of the gradient is less than 10^{-6}.

$$f(\mathbf{Y}) = \frac{1}{2}\|\mathcal{P}_\Omega(g(\mathbf{Y}\mathbf{Y}^T) - \hat{\mathbf{D}})\|_F^2. \tag{11}$$

Three metrics are considered to evaluate the performance, i.e., the spectral error (12a), EDM recovery rate (12b), and the Frobenius error (12c), defined below.

$$\mathrm{SE} := \frac{\|\hat{\mathbf{G}} - \mathbf{G}^\star\|}{\|\mathbf{G}^\star\|}, \tag{12a}$$

$$\mathrm{RE} := \frac{\|\bar{\mathbf{D}} - \mathbf{D}^\star\|_F}{\|\mathbf{D}^\star\|_F}, \tag{12b}$$

$$\mathrm{FE} := \frac{\|\hat{\mathbf{G}} - \mathbf{G}^\star\|_F}{\|\mathbf{G}^\star\|_F}, \tag{12c}$$

where $\hat{\mathbf{G}}$ is the approximation generated by either OS-MDS or SVD-MDS, and $\bar{\mathbf{D}}$ denotes the recovered EDM after GD refinement. The SNR in noisy tests is defined by

$$\mathrm{SNR} := \frac{\|\mathbf{G}^\star\|_F^2}{n^2\sigma_n^2}.$$

We say "in dB" means taking $20\log_{10}$ to the residual in (12), or $10\log_{10}$ to the SNR. All experiments are run on a dual-socket Intel Xeon Gold 6226r server with 256 GB of RAM, NVIDIA A5000, Ubuntu 20.04.4, MATLAB 2022a.

B. Noiseless Test

Fig. 1 plots the spectral error of OS-MDS and SVD-MDS when varying Bernoulli sample rate p and the number of points n. We use the median value instead of the mean of 1000 independent trials mainly due to the observation of artificial outliers when evaluating (12a). This is caused by the limited iteration budget when using power method to calculate the spectral norm. Fig. 2 shows the phase transition of the test

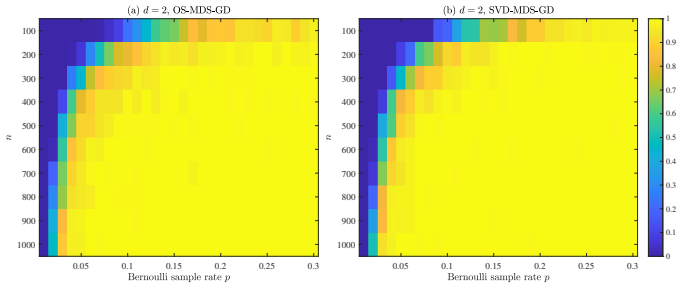

Fig. 3. The phase transition of OS-MDS and SVD-MDS initialized GD with Armijo step size, when varying Bernoulli sample rate p and the number of points n. We claim a success if the EDM recovery rate falls below 10^{-3}. For each pair (n, p), the result is obtained by 100 independent trials.

estimators. To verify Theorem I.1 with $\epsilon < 1$, we claim a success if (12a) falls below 1. The transition edge clearly shows the $p = \mathcal{O}(\log n/n)$ dependency, which matches the theoretical result[5]. By comparing Fig. 1(a)/Fig. 2(a) with Fig. 1(b)/Fig. 2(b), one finds that the performance gap between OS-MDS and SVD-MDS is ignorable.

The phase transition of GD refinement, when initialized by either OS-MDS or SVD-MDS is plotted in Fig. 3. Again, the two strategies show minor differences. It is interesting to see that GD can converge in a very less restrictive region around the ground truth. Yet a non-trivial explanation of this phenomenon, when specified to the EDMC problem, is still a mystery, please see [16] for more discussion.

C. Noisy Test

Fig. 4 plots the noise-resistance behavior, evaluated by the Frobenius error, of OS-MDS and SVD-MDS when varying Bernoulli sample rate p and the SNR. The range of p is chosen to be the period where phase transition of these two estimators happens. The result clearly shows the existence of an SNR threshold around -5 dB, i.e., below this, the noise dominates the approximation error. When the noise is ignorable, the first term in (4) starts to matter. It is surprising to see OS-MDS performs slightly better than SVD-MDS as the problem becomes well-posed, since it only uses one-step SVD reconstruction.

D. Protein Test

We perform a similar test over real protein data as in [14, Sec. V-A]. Specifically, we choose four of the tested proteins in [27, Sec. 5.3] and randomly delete inter-atom distance measurements according to the symmetric Bernoulli rule. Then the OS-MDS initialized GD is applied to solve this EDMC problem. The algorithm is run on GPU using MATLAB `gpuArray` to speed up computation. Fig. 5 presents the EDM recovery rate of this process when varying Bernoulli sample rate p and the protein type. It suggests that all four protein structures can be exactly recovered from less than 5% random distance samples by a simple, scalable, memory-friendly strategy. Yet this numerical witness does not mean

[5]Since $d = 2$ and $n \geq 100$ in all tests and we use Gaussian point set ($\nu = \mathcal{O}(1)$), the $\log n$ term in Theorem I.1 dominates.

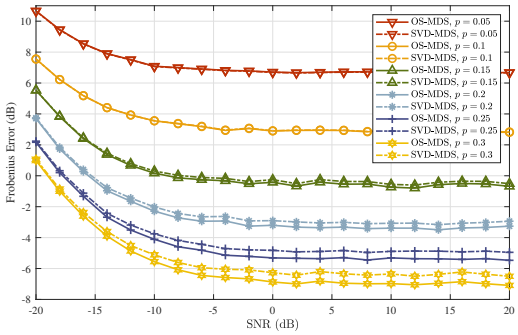

Fig. 4. The Frobenius error (in dB) of OS-MDS and SVD-MDS v.s. SNR (in dB) when varying Bernoulli sample rate p. We fix $n = 500$, $d = 2$. The results are the mean value of 1000 independent trials.

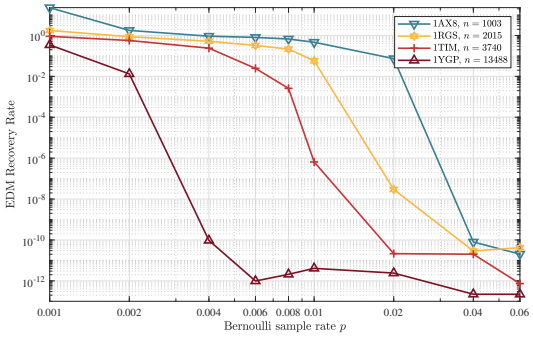

Fig. 5. The EDM recovery rate of OS-MDS initialized GD with Armijo step size when varying Bernoulli sample rate p, tested on four proteins. The proteins have different atom numbers (n) and fixed embedding dimension $d = 3$. 1AX8 and 1RGS are single-chain while the other are double-chain. The results are the mean value of five independent trials.

that the protein configuration problem [8] is ready to be handled via delicate initialized GD, since the practical distance sampling scheme, caused by the behavior of Nuclear Magnetic Resonance (NMR) techniques[6], do not follow the Bernoulli rule. Please see [16, Sec. III-C] for more discussion regarding the impact of this practical scheme for non-convex EDMC algorithm design.

III. Proof

Our proof of Theorem I.1 relies on a celebrated sharp bound by Bandeira and van Handel [28], restated here as the lemma below. It is tighter than the matrix Bernstein inequality [23, Thm. 6.1] when applied to matrix with independent random entries.

Lemma III.1 ([28], Coro. 3.12, Remark 3.13)**.** *Let* $\mathbf{X}$ *be a symmetric matrix in* $\mathbb{R}^{n\times n}$ *whose entries* $\mathbf{X}_{ij}$ *are independent, zero-mean random variables. Let*

$$\sigma_x^2 := \max_i \sum_{j=1}^{n} \mathbb{E}[\mathbf{X}_{ij}^2], \max_{ij} |\mathbf{X}_{ij}| \leq B.$$

[6]NMR can estimate the distance between two atoms if their pair-wise distance is less than 6Å [27]. A similar sample rule also appears in the sensor network localization problem [10].

Then there exist some universal constant $c > 0$, such that for every $t > 0$

$$\mathbb{P}\{\|\mathbf{X}\| > 3\sqrt{2}\sigma_x + t\} \leq n\exp(-\frac{t^2}{cB^2}).$$

One immediate corollary is that [21, Sec. 3.1], if $\mathbf{X}$ satisfies $\max_{ij}\mathbb{E}[\mathbf{X}_{ij}^2] \leq \sigma^2$, we have

$$\|\mathbf{X}\| \leq 3\sqrt{2}\sqrt{n\sigma^2} + c'B\sqrt{\beta\log n} \tag{13}$$

holds for some universal constant c' and $\beta > 1$ with probability at least $1 - n^{\beta-1}$.

Proof: We are now ready to proceed with the proof. Notice that

$$\begin{aligned}
&\| -\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J} - \mathbf{G}^\star\| \overset{(i)}{=} \|\frac{1}{p}\mathcal{R}_\Omega(\mathbf{G}^\star) - \mathbf{G}^\star + g^+(\frac{1}{p}\mathcal{P}_\Omega\mathbf{E})\| \\
&\leq \|\frac{1}{p}\mathcal{R}_\Omega(\mathbf{G}^\star) - \mathbf{G}^\star\| + \|\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\mathbf{E})\mathbf{J}\| \\
&\overset{(ii)}{\leq} \|g^+(\frac{1}{p}\mathcal{P}_\Omega\mathbf{D}^\star - \mathbf{D}^\star)\| + \frac{1}{2}\|\frac{1}{p}\mathcal{P}_\Omega\mathbf{E}\| \\
&\overset{(iii)}{\leq} \frac{1}{2}(\underbrace{\|\frac{1}{p}\mathcal{P}_\Omega\mathbf{D}^\star - \mathbf{D}^\star\|}_{I_1} + \underbrace{\|\frac{1}{p}\mathcal{P}_\Omega\mathbf{E}\|}_{I_2}).
\end{aligned} \tag{14}$$

Where (i) from our model setup in Section I-B, (ii), (iii) by noticing that $\mathbf{J}$ is an orthogonal projection thus $\|\mathbf{J}\| \leq 1$. What remains is to handle I_1 and I_2.

Bound on I_1: It is straightforward to check that the matrix $\mathbf{X} := \frac{1}{p}\mathcal{P}_\Omega\mathbf{D}^\star - \mathbf{D}^\star$ satisfies all prerequisites of Lemma III.1. Specifically, $B = \|\mathbf{X}\|_\infty \leq \frac{1}{p}\|\mathbf{D}^\star\|_\infty$ and $\sigma^2 = \max_{ij}\mathbb{E}[\mathbf{X}_{ij}^2] \leq \mathbb{E}(\frac{1}{p}\delta_{\boldsymbol{\alpha}} - 1)^2\|\mathbf{D}^\star\|_\infty^2 \leq \frac{1}{p}\|\mathbf{D}^\star\|_\infty^2$. Using (13), we have

$$\begin{aligned}
I_1 &\leq c_1\sqrt{\frac{n}{p}\|\mathbf{D}^\star\|_\infty^2} + c_2\sqrt{\beta\log n}\frac{1}{p}\|\mathbf{D}^\star\|_\infty \\
&\overset{(i)}{\leq} 2c_1\sqrt{\frac{n}{p}\|\mathbf{D}^\star\|_\infty^2} \overset{(ii)}{\leq} 2c_1\sqrt{\frac{4\mu^4d^2}{np}}\|\mathbf{G}^\star\| \overset{(iii)}{\leq} \varepsilon\|\mathbf{G}^\star\|,
\end{aligned} \tag{15}$$

where c_1, c_2 are some universal constants. (i) from setting $p > (\frac{c_2}{c_1})^2\beta\log n/n$, and (iii) by forcing $p > 16c_1^2\mu^4d^2/(n\varepsilon^2)$ for some $0 < \varepsilon \leq 1/2$. (ii) from noticing that

$$\|\mathbf{D}^\star\|_\infty^2 \leq \max_{\boldsymbol{\alpha}}|\langle\boldsymbol{\omega}_{\boldsymbol{\alpha}}, \mathbf{G}^\star\rangle|^2 \overset{(a)}{\leq} 4\|\mathbf{G}^\star\|_\infty^2 \overset{(b)}{\leq} \frac{4\mu^4d^2}{n^2}\|\mathbf{G}^\star\|^2,$$

where (a) from expanding $\boldsymbol{\omega}_{\boldsymbol{\alpha}}$, i.e.,

$$\langle\boldsymbol{\omega}_{\boldsymbol{\alpha}}, \mathbf{G}^\star\rangle = \mathbf{G}^\star_{ii} + \mathbf{G}^\star_{jj} - \mathbf{G}^\star_{ij} - \mathbf{G}^\star_{ji},$$

then followed by Cauchy-Schwarz, and (b) by Definition I.1. Thus we conclude that (15) holds with probability at least $1 - n^{\beta-1}$ if $p \gtrsim \max\{\beta\log n, \mu^4d^2/\varepsilon^2\}/n$, $\beta > 1$.

Bound on I_2: This is a meanwhile standard result proved by [29, Lemma 11] using truncation argument. We briefly restate it here. By standard sub-Gaussian tail one finds $\max_{ij}|\frac{1}{p}\delta_{\boldsymbol{\alpha}}\mathbf{E}_{ij}| \leq C\sqrt{\sigma_n^2\log n}/p$ hold with probability $1 - n^{-c}$. Also $\mathbb{E}|\frac{1}{p}\delta_{\boldsymbol{\alpha}}\mathbf{E}_{ij}|^2 \leq C\sigma_n^2/p$. Using (13) and setting $p \gtrsim \frac{\beta\log^2 n}{n}$, $\beta > 1$, we have $I_2 \lesssim \sigma_n\sqrt{\frac{n}{p}}$ holds with probability at least $1 - n^{-c} - n^{\beta-1}$.

By substituting the above two estimates into (14) and using Weyl's theorem, we find

$$\begin{aligned}
\|\hat{\mathbf{G}}_\natural - \mathbf{G}^\star\| &\leq \|\mathcal{T}_d\left[-\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J}\right] + \frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J}\| \\
&\quad + \| -\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J} - \mathbf{G}^\star\| \\
&\leq 2\| -\frac{1}{2p}\mathbf{J}(\mathcal{P}_\Omega\hat{\mathbf{D}})\mathbf{J} - \mathbf{G}^\star\| \lesssim \varepsilon\|\mathbf{G}^\star\| + \sigma_n\sqrt{\frac{n}{p}},
\end{aligned}$$

where we conclude the proof after some trivial rescale of the constants. ∎

IV. Conclusion

This paper proposes a new spectral estimator, namely, the OS-MDS, for EDM approximation under Bernoulli sample model. Theoretical characterization of its behavior is conducted under the presence of additive, sub-Gaussian, and symmetric noise. Numerical results verify the claimed performance. It is of future interest to explore the theoretical guarantee of the algorithm proposed in [24], and to uncover the convergence behavior of the gradient method when applied to (11), starting from the initial point provided by OS-MDS.

References

[1] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Process Mag.*, vol. 32, no. 6, pp. 12–30, 2015. DOI: 10.1109/MSP.2015.2398954.

[2] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative review," *J. Mach. Learn. Res.*, vol. 10, no. 66-71, p. 13, 2009.

[3] E. Arias-Castro, A. Javanmard, and B. Pelletier, "Perturbation bounds for procrustes, classical scaling, and trilateration, with applications to manifold learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, Article 15, 2020.

[4] P. Agostini, Z. Utkovski, and S. Stańczak, "Channel Charting: An Euclidean distance matrix completion perspective," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5010–5014. DOI: 10.1109/ICASSP40776.2020.9053639.

[5] S. Gupta, R. Gribonval, L. Daudet, and I. Dokmanić, "Don't take it lightly: Phasing optical random projections with unknown operators," *Adv. Neural Inf. Proces. Syst.*, vol. 32, 2019.

[6] G. M. Crippen, "Conformational analysis by energy embedding," *J. Comput. Chem.*, vol. 3, no. 4, pp. 471–476, 1982. DOI: 10.1002/jcc.540030404.

[7] H. Wang, J. Yang, Y. Zhang, J. Qian, and J. Wang, "Reconstruct high-resolution 3D genome structures for diverse cell-types using FLAMINGO," *Nat. Commun.*, vol. 13, no. 1, p. 2645, 2022. DOI: 10.1038/s41467-022-30270-2.

[8] B. Hendrickson, "The molecule problem: Exploiting structure in global optimization," *SIAM J. Optim.*, vol. 5, no. 4, pp. 835–857, 1995. DOI: 10.1137/0805040.

[9] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Math. Program.*, vol. 109, no. 2-3, pp. 367–384, 2006, ISSN: 0025-5610 1436-4646. DOI: 10.1007/s10107-006-0040-1.

[10] A. Javanmard and A. Montanari, "Localization from incomplete noisy distance measurements," *Found. Comput. Math.*, vol. 13, no. 3, pp. 297–345, 2013. DOI: 10.1007/s10208-012-9129-5.

[11] F. Marić, M. Giamou, A. W. Hall, S. Khoubyarian, I. Petrović, and J. Kelly, "Riemannian optimization for distance-geometric inverse kinematics," *IEEE Trans. Robot.*, vol. 38, no. 3, pp. 1703–1722, 2022. DOI: 10.1109/TRO.2021.3123841.

[12] J. B. Saxe, "Embeddability of weighted graphs in k-space is strongly np-hard," in *17th Allerton Conf. Commun. Control Comput., 1979*, 1979, pp. 480–489.

[13] A. M.-C. So and Y. Ye, "A semidefinite programming approach to tensegrity theory and realizability of graphs," in *Proc. Annu. ACM SIAM Symp. Discrete Algorithms*, 2006, pp. 766–775.

[14] A. Tasissa and R. Lai, "Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3124–3144, 2019. DOI: 10.1109/TIT.2018.2881749.

[15] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009. DOI: 10.1007/s10208-009-9045-5.

[16] Y. Li and X. Sun, "Sensor network localization via riemannian conjugate gradient and rank reduction," *IEEE Trans. Signal Process.*, vol. 72, pp. 1910–1927, 2024. DOI: 10.1109/TSP.2024.3378378.

[17] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018. DOI: 10.1017/9781108231596.

[18] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM Rev.*, vol. 56, no. 1, pp. 3–69, 2014.

[19] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proc. Am. Control Conf.*, vol. 3, 2003, pp. 2156–2162. DOI: 10.1109/ACC.2003.1243393.

[20] P. Drineas, A. Javed, M. Magdon-Ismail, G. Pandurangan, R. Virrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," in *Ann. IEEE Commun. Soc. Ad hoc Commun. Netw. Secon*, vol. 2, 2006, pp. 536–544. DOI: 10.1109/SAHCN.2006.288510.

[21] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Spectral methods for data science: A statistical perspective," *Found. Trends Mach. Learn.*, vol. 14, no. 5, pp. 566–806, 2021. DOI: 10.1561/2200000079.

[22] H. Zhang, Y. Liu, and H. Lei, "Localization from incomplete euclidean distance matrix: Performance analysis for the svd–mds approach," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2196–2209, 2019. DOI: 10.1109/TSP.2019.2904022.

[23] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, 2012. DOI: 10.1007/s10208-011-9099-z.

[24] C. M. Smith, S. P. Lichtenberg, H. Cai, and A. Tasissa, "Riemannian optimization for euclidean distance geometry," in *OPT 2023: Optimization for Machine Learning*, 2023.

[25] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–42, 2000. DOI: 10.1093/nar/28.1.235.

[26] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 42, pp. 1455–1459, 2014. [Online]. Available: https://www.manopt.org.

[27] N.-H. Z. Leung and K.-C. Toh, "An sdp-based divide-and-conquer algorithm for large-scale noisy anchor-free graph realization," *SIAM J. Sci. Comput.*, vol. 31, no. 6, pp. 4351–4372, 2010. DOI: 10.1137/080733103.

[28] A. S. Bandeira and R. van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *Ann. Probab.*, vol. 44, no. 4, pp. 2479–2506, 2016.

[29] Y. Chen and M. J. Wainwright, "Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees," *ArXiv*, vol. abs/1509.03025, 2015.