

# Exact Reconstruction of Euclidean Distance Geometry Problem Using Low-rank Matrix Completion

Abiy Tasissa, Rongjie Lai

## Abstract

The Euclidean distance geometry problem arises in a wide variety of applications, from determining molecular conformations in computational chemistry to localization in sensor networks. When the distance information is incomplete, the problem can be formulated as a nuclear norm minimization problem. In this paper, this minimization program is recast as a matrix completion problem of a low-rank  $r$  Gram matrix with respect to a suitable basis. The well known restricted isometry property can not be satisfied in this scenario. Instead, a dual basis approach is introduced to theoretically analyze the reconstruction problem. If the Gram matrix satisfies certain coherence conditions with parameter  $\nu$ , the main result shows that the underlying configuration of  $n$  points can be recovered with very high probability from  $O(nrv \log^2(n))$  uniformly random samples. Computationally, simple and fast algorithms are designed to solve the Euclidean distance geometry problem. Numerical tests on different three dimensional data and protein molecules validate effectiveness and efficiency of the proposed algorithms.

## Index Terms

Euclidean distance geometry, low-rank matrix completion, nuclear norm minimization, dual basis, random matrices, Gram matrix.

## I. INTRODUCTION

The Euclidean distance geometry problem, EDG hereafter, aims at constructing the configuration of points given partial information on pairwise inter-point distances. The problem has applications in diverse areas, such as molecular conformation in computational chemistry [1], [2], [3], dimensionality reduction in machine learning [4] and localization in sensor networks [5], [6]. For instance, in the molecular conformation problem, the goal is to determine the structure of protein given partial interatomic distances obtained from nuclear magnetic resonance (NMR) spectroscopy experiments. Since the structure of protein determines its physical and chemical properties, the molecular conformation problem is crucial for biological applications such as drug design. In recent works, novel applications of EDG like solving partial differential equations (PDEs) on manifolds represented as incomplete inter-point distance information have been explored in [7].

For the mathematical setup of the problem, consider a set of  $n$  points  $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}^T \in \mathbb{R}^{n \times r}$ . The squared distance between any two points  $\mathbf{p}_i$  and  $\mathbf{p}_j$  is given by  $d_{i,j}^2 = \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 = \mathbf{p}_i^T \mathbf{p}_i + \mathbf{p}_j^T \mathbf{p}_j - 2\mathbf{p}_i^T \mathbf{p}_j$ . The Euclidean distance matrix  $\mathbf{D} = [d_{i,j}^2]$  can be written compactly as follows

$$\mathbf{D} = \mathbf{1} \operatorname{diag}(\mathbf{P}\mathbf{P}^T)^T + \operatorname{diag}(\mathbf{P}\mathbf{P}^T)\mathbf{1}^T - 2\mathbf{P}\mathbf{P}^T \quad (1)$$

where  $\mathbf{1}$  is a column vector of ones and  $\operatorname{diag}(\cdot)$  is a column vector of diagonal entries of the matrix in consideration.  $\mathbf{X} = \mathbf{P}\mathbf{P}^T$  is the inner product matrix well known as the Gram matrix. If the complete distance information is available, the coordinates can be recovered from the eigendecomposition of the Gram matrix using classical multidimensional scaling (MDS) [8]. It should be noted that this solution is unique up to rigid transformations and translations.

In practice, the distance matrix is incomplete and finding the underlying point coordinates with partial information is in general not possible. One approach proposed in [7] is based on a matrix completion applied to the Gram matrix which we briefly summarize here. Assume that the entries of  $\mathbf{D}$  are sampled uniformly at random. By construction,  $\mathbf{X}$  is symmetric and positive semidefinite. The solution for  $\mathbf{X}$  is unique up to translations. This ambiguity is fixed by considering the constraint that the centroid of the points is located at the origin,  $\sum_{k=1}^n \mathbf{p}_k = \mathbf{0}$ , which leads to  $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ . Assuming that  $\mathbf{X}$  is low-rank,  $r \ll n$ , the work in [7] considers the following nuclear norm minimization program to recover  $\mathbf{X}$ .

$$\begin{aligned} & \text{minimize} && \|\mathbf{X}\|_* \\ & \text{subject to} && \mathbf{X}_{i,i} + \mathbf{X}_{j,j} - 2\mathbf{X}_{i,j} = \mathbf{D}_{i,j} \quad (i, j) \in \Omega \\ & && \mathbf{X} \cdot \mathbf{1} = \mathbf{0}; \mathbf{X} = \mathbf{X}^T; \mathbf{X} \geq \mathbf{0} \end{aligned} \quad (2)$$

A. Tasissa is with the Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180, (tasisa@rpi.edu).

R. Lai is with the Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180, U.S.A. (lair@rpi.edu). This work is partially supported by NSF DMS-1522645 and an NSF CAREER award, DMS-1752934.

Here  $\|\mathbf{X}\|_*$  denotes the nuclear norm and  $\Omega \subset \{(i, j) | i, j = 1, \dots, n, i < j\}$ ,  $|\Omega| = m$ , denotes the random set that consists of all the sampled indices. One characterization of the Euclidean distance matrix, due to Gower [9], states that  $\mathbf{D}$  is an Euclidean distance matrix if and only if  $\mathbf{D} = \mathbf{X}_{i,i} + \mathbf{X}_{j,j} - 2\mathbf{X}_{i,j} = \mathbf{D}_{i,j} \forall i, j$  for some positive semidefinite matrix  $\mathbf{X}$  satisfying  $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ . As such, the above nuclear norm minimization can be interpreted as a regularization of the rank with a prior assumption that the true Gram matrix is low rank. An alternative approach based on the matrix completion method is direct completion of the distance matrix [10], [11], [12]. Compared to this approach, an advantage of the above minimization program can be seen by comparing the rank of the Gram matrix  $\mathbf{X}$  and the distance matrix  $\mathbf{D}$ . Using (1), the rank of  $\mathbf{D}$  is at most  $r+2$  while the rank of  $\mathbf{X}$  is simply  $r$ . Using matrix completion theory, it can be surmised that relatively less number of samples are required for the Gram matrix completion. Numerical experiments in [7] confirm this observation. In [13], the authors consider a theoretical analysis of a specific instance of localization problem and propose an algorithm similar to (2). The paper considers a random geometric model and derives interesting results of bound of errors in reconstructing point coordinates. While the localization problem and EDG problem share a common theme, we remark that the EDG problem is different and our analysis adopts the matrix completion framework. The main task of this paper is a theoretical analysis of the above minimization problem. In particular, under appropriate conditions, we will show that the above nuclear norm minimization recovers the underlying inner product matrix.

*a) Related Work:* The EDG problem has been well studied in various contexts. Early works establish mathematical properties of Euclidean distance matrices (EDM) [9] and prove conditions for a symmetric hollow matrix to be EDM [14], [15]. Particularly, the important result due to Schoenberg states that a symmetric hollow matrix  $\mathbf{D}$  is EDM if and only if the matrix  $-\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J}$  is positive semidefinite.  $\mathbf{J}$  is the geometric centering matrix defined as  $\mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$  [14]. In follow-up theoretical papers, the EDM completion problem was considered. Given a symmetric hollow matrix with some entries specified, the EDM completion problem asks if the partial matrix can be completed to an EDM. A work of Hendrickson relates the EDM completion problem to a graph realization problem and provides necessary graph theoretic conditions on the partial matrix that ensures a unique solution [16]. Bakonyi and Johnson establish a link between EDM completion and the graph of partial distance matrix. Specifically, they show that there is a completion to a distance matrix if the graph is chordal [17]. The connection between the EDM completion problem and the positive semidefinite matrix completion was established in [18] and [19]. In [20], Alfakih establishes necessary and sufficient conditions for the solution of EDM completion problem to be unique. The emphasis in the above theoretical works is analytic characterization of EDMs and the EDM completion problem mostly employing graph theoretic methods. In the numerical side, a variety of algorithms using different optimization techniques have been developed to solve the EDM completion problem [3], [21], [22], [23], [24], [25], [26]. The above review is by no means exhaustive and we recommend the interested reader to refer to [27] and [28]. This paper adopts the low-rank matrix recovery framework. While the well-known matrix completion theory can be applied to reconstruct  $\mathbf{D}$  [10], [12], it can not be directly used to analyze (2). In particular, the measurements in (2) are with respect to the Gram matrix while the measurements in the work of [10], [12] are entry wise sampling of the distance matrix. The emphasis in this paper is on theoretical understanding of the nuclear norm minimization formulation for the EDG problem as stated in (2). In particular, the goal is to provide rigorous probabilistic guarantees that give precise estimates for the minimum number of measurements needed for a certain success probability of the recovery algorithm.

*b) Main Challenges:* The random linear constraints in (2) can be equivalently written as a linear map  $\mathcal{L}$  acting on  $\mathbf{X}$ . A first approach to showing uniqueness for the problem (2) is to check if  $\mathcal{L}$  satisfies the restricted isometry property (RIP) [12]. However, the RIP does not hold for the completion problem (2). This can be simply verified by choosing any  $(i, j) \notin \Omega$  and construct a matrix  $\mathbf{X}$  with  $\mathbf{X}_{i,j} = \mathbf{X}_{j,i} = \mathbf{X}_{i,i} = \mathbf{X}_{j,j} = 1$  and zero everywhere else. Then, it is clear that  $\mathcal{L}(\mathbf{X}) = \mathbf{0}$  implying that the RIP condition does not hold. In general, RIP based analysis is not suitable for deterministic structured measurements. Adopting the framework introduced in [29], the nuclear norm minimization problem in (2) can be written as a matrix completion problem with respect to a basis  $\{\mathbf{w}_\alpha\}$ . It, however, turns out that  $\mathbf{w}_\alpha$  is not orthogonal. With non-orthogonal basis, the measurements  $\langle \mathbf{w}_\alpha, \mathbf{X} \rangle$  are not compatible with the expansion coefficients of the true Gram matrix. A possible remedy is orthogonalization of the basis, say using the Gram-Schmidt orthogonalization. Unfortunately, the orthogonalization process does not preserve the structure of the basis  $\mathbf{w}_\alpha$ . This has the implication that the modified minimization problem (2) no longer corresponds to the original problem. As such, the lack of orthogonality is critical in this problem. In addition, it is necessary that the solution of (2) is symmetric positive semidefinite satisfying the constraint  $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ . On the basis of the above considerations, an alternative approach is considered to show that (2) admits a unique solution. The analysis presented in this paper is not based on RIP but on the dual certificate approach introduced in [10]. Our proof was inspired by the work of David Gross [29] where the author generalizes the matrix completion problem to any orthonormal basis. In the case of the EDG problem, one main challenge is that the sampling operator, an important operator in matrix completion, is no longer self-adjoint. This necessitates modifications and alternative proofs to some of the technical statements that appear in [29].

*c) Major Contributions:* In this paper, a dual basis approach is introduced to show that (2) has a unique solution under appropriate sampling conditions. First, the minimization problem in (2) is written as matrix completion problem with respect to a basis  $\mathbf{w}_\alpha$ . Second, by introducing a dual basis  $\{\mathbf{v}_\alpha\}$  to  $\{\mathbf{w}_\alpha\}$ , one can ensure that the measurements  $\langle \mathbf{X}, \mathbf{w}_\alpha \rangle$  in (2) are compatible with expansion coefficients of the true Gram matrix  $\mathbf{M}$ . The two main contributions of this paper are as follows.

- 1) A dual basis approach is introduced to address the EDG problem. Under certain assumptions, we show that the nuclear

norm minimization problem succeeds in recovering the underlying low rank solution. Precisely, the main result states that if  $|\Omega| = m \geq O(nr \log^2 n)$ , the nuclear norm minimization program (2) recovers the underlying inner product matrix with very high probability,  $1 - n^{-\beta}$  for  $\beta > 1$  (see more details in Theorem 1). The minimization problem has a positive semidefinite constraint. The proof describes how this constraint is handled. In addition to the Bernstein bound which appears in matrix completion analysis, a crucial part in the proof uses the operator Chernoff bound. We would like to emphasize that this part of the proof and the estimate using Chernoff bound is simple and could be useful in a broader setting.

- 2) We develop simple and fast algorithms to solve the EDG problem under two instances. The first instance considers the case of exact partial information. The second instance considers the more realistic setup of a noisy partial information. Numerical tests on various data show that the algorithms are accurate and efficient.

The outline of the paper is as follows. In section II, we introduce a dual basis approach and formulate a well-defined matrix completion problem for the EDG problem. We conclude the section by proposing coherence conditions for the EDG problem and explaining the sampling scheme. In section III, the proof of exact reconstruction is presented. In brief terms, the main parts are summarized as follows. From convex optimization theory, showing uniqueness of the EDG problem is equivalent to showing that there exists a dual certificate, denoted by  $\mathbf{Y}$ , satisfying certain conditions.  $\mathbf{Y}$  is constructed using the golfing scheme proposed in [29]. Next, we show that these conditions hold with very high probability by employing concentration inequalities. This implies that there is a unique solution with very high probability. In section IV, fast numerical algorithms for the EDG matrix completion problem are proposed. Section V validates the efficiency and effectiveness of the proposed numerical algorithms. Finally, we conclude the work in the last section.

d) *Notation:* To make our notation consistent in the paper, we summarize notations used in this paper in Table I.

TABLE I  
NOTATIONS

$\mathbf{x}$	Vector	$\ \mathbf{x}\ _2$	$l_2$ norm
$\mathbf{X}$	Matrix	$\ \mathbf{X}\ _F$	Frobenius norm
$\mathcal{X}$	Operator	$\ \mathcal{X}\ $	Operator norm
$\mathbf{X}^T$	Transpose	$\ \mathbf{X}\ _*$	Nuclear norm
$\text{Tr}(\mathbf{X})$	Trace	$\ \mathcal{A}\ $	$\sup_{\ \mathbf{X}\ _F=1} \ \mathcal{A}\mathbf{X}\ _F$ .
$\langle \mathbf{X}, \mathbf{Y} \rangle$	Trace( $\mathbf{X}^T \mathbf{Y}$ )	$\lambda_{\max}, \lambda_{\min}$	Maximum, Minimum eigenvalue
$\mathbf{1}$	A vector or matrix of ones	Sgn $\mathbf{X}$	$\mathbf{U}\text{sign}(\Sigma)\mathbf{V}^T$ ; here $[\mathbf{U}, \Sigma, \mathbf{V}] = \text{svd}(\mathbf{X})$
$\mathbf{0}$	A vector or matrix of zeros	$\Omega, \mathbb{I}$	Random sampled set, Universal set

## II. DUAL BASIS FORMULATION

The aim of this section is to show that the EDG problem (2) can be equivalently stated as a matrix completion problem with respect to a special designed basis. The matrix completion problem is an instance of the low rank matrix recovery problem where the goal is to recover a low rank matrix given random partial linear observations. The natural minimization problem for low rank recovery problem is rank minimization with linear constraints. Unfortunately, this problem is NP hard [12] motivating other solutions. In a series of remarkable theoretical papers [10], [12], [29], [30], [31], it was shown that, under certain conditions, the solutions of the NP hard rank minimization problem can be obtained by solving the convex nuclear norm minimization problem which is computationally tractable [32], [33]. Our theoretical analysis is inspired by the work [29] where the author extends the nuclear norm minimization formulation to recovering a low rank matrix given that a few of its coefficients in a fixed orthonormal basis are known.

a) *Dual basis:* To write the EDG problem (2) as a matrix completion problem with respect to an appropriate operator basis, let us introduce few notations. We write  $\mathbb{I} = \{\alpha = (\alpha_1, \alpha_2) \mid 1 \leq \alpha_1 < \alpha_2 \leq n\}$  and define

$$\mathbf{w}_\alpha = \mathbf{e}_{\alpha_1, \alpha_1} + \mathbf{e}_{\alpha_2, \alpha_2} - \mathbf{e}_{\alpha_1, \alpha_2} - \mathbf{e}_{\alpha_2, \alpha_1} \quad (3)$$

where  $\mathbf{e}_{\alpha_1, \alpha_2}$  is a matrix whose entries are all zero except a 1 in the  $(\alpha_1, \alpha_2)$ -th entry. It is clear that  $\{\mathbf{w}_\alpha\}$  forms a basis of the linear space  $\mathbb{S} = \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^T \& \mathbf{X} \cdot \mathbf{1} = \mathbf{0}\}$  and the number of basis is  $L = \frac{n(n-1)}{2}$ . For conciseness and ease in later analysis, we further define  $\mathbb{S}_+ = \mathbb{S} \cap \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} \geq \mathbf{0}\}$ . Therefore, for any given subsample  $\Omega$  of  $\mathbb{I}$ , the EDG problem (2) can be written as the following nuclear norm minimization problem

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{S}_+}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \langle \mathbf{w}_\alpha, \mathbf{X} \rangle = \langle \mathbf{w}_\alpha, \mathbf{M} \rangle \quad \forall \alpha \in \Omega \end{aligned} \quad (4)$$

where  $\mathbf{M}$  is the true underlying rank  $r$  Gram matrix satisfying  $\mathbf{M}_{i,i} + \mathbf{M}_{j,j} - 2\mathbf{M}_{i,j} = \mathbf{D}_{i,j} \forall (i, j)$ . The EDG problem can now be equivalently interpreted as a matrix completion problem with respect to the basis  $\mathbf{w}_\alpha$ .

By construction,  $\mathbf{w}_\alpha$  is symmetric and satisfies  $\mathbf{w}_\alpha \cdot \mathbf{1} = \mathbf{0}$ . The latter condition naturally enforces the constraint  $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ . It is clear that any  $\mathbf{X} \in \mathbb{S}$  can be expanded in the basis  $\{\mathbf{w}_\alpha\}$  as  $\mathbf{X} = \sum_{\beta \in \mathbb{I}} c_\beta \mathbf{w}_\beta$ . After minor algebra,  $c_\beta = \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle H^{\alpha, \beta}$  where we define  $H_{\alpha, \beta} = \langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle$  and  $H^{\alpha, \beta} = H_{\alpha, \beta}^{-1}$ . Note that, since  $\{\mathbf{w}_\alpha\}$  is a basis, the inverse  $H^{-1}$  is well defined. This results in the following representation of  $\mathbf{X}$ .

$$\mathbf{X} = \sum_{\beta \in \mathbb{I}} \sum_{\alpha \in \mathbb{I}} H^{\alpha, \beta} \langle \mathbf{w}_\alpha, \mathbf{X} \rangle \mathbf{w}_\beta \quad (5)$$

The crux of the dual basis approach is to simply consider (5) and rewrite it as follows

$$\mathbf{X} = \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{v}_\alpha \quad (6)$$

where  $\mathbf{v}_\alpha = \sum_{\beta} H^{\alpha, \beta} \mathbf{w}_\beta$ . It can be easily verified that  $\{\mathbf{v}_\alpha\}$  is a dual basis of  $\{\mathbf{w}_\alpha\}$  satisfying  $\langle \mathbf{v}_\alpha, \mathbf{w}_\beta \rangle = \delta_{\alpha, \beta}$ . Let  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L]$  and  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_L]$  denote the matrix of vectorized basis matrices and vectorized dual basis matrices respectively. The following basic relations are useful in later analysis,  $\mathbf{H} = \mathbf{W}^T \mathbf{W}$ ,  $\mathbf{H}^{-1} = \mathbf{V}^T \mathbf{V}$  and  $\mathbf{V} = \mathbf{W} \mathbf{H}^{-1}$ . In the context of the EDG completion problem, the dual basis approach ensures that the expansion coefficients match the measurements while preserving the condition that the matrix in consideration is symmetric with zero row sums. With this, (4) turns into a well formulated matrix completion problem with respect to the basis  $\mathbf{w}_\alpha$ . An alternative way to rewrite (4) makes use of the sampling operator defined as follows.

$$\mathcal{R}_\Omega : \mathbf{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{v}_\alpha \quad (7)$$

where we assume that  $\Omega$  are sampled uniformly at random from  $\mathbb{I}$  with replacement and  $m = |\Omega|$  is the size of  $\Omega$ . The scaling factor  $\frac{L}{m}$  is for ease in later analysis. It can be easily verified that  $\frac{m^2}{L^2} \mathcal{R}_\Omega^2 = \frac{m}{L} \mathcal{R}_\Omega$ . The adjoint operator of  $\mathcal{R}_\Omega$ ,  $\mathcal{R}_\Omega^*$ , can be simply derived and is given by

$$\mathcal{R}_\Omega^* : \mathbf{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle \mathbf{w}_\alpha \quad (8)$$

Using the sampling operator  $\mathcal{R}_\Omega$ , we can write (4) as follows

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{S}_+}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \mathcal{R}_\Omega(\mathbf{X}) = \mathcal{R}_\Omega(\mathbf{M}) \end{aligned} \quad (9)$$

In addition to the sampling operator, the restricted frame operator appears in the analysis and is defined as follows

$$\mathcal{F}_\Omega : \mathbf{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{w}_\alpha \quad (10)$$

Note that the restricted frame operator is self-adjoint and positive semidefinite.

b) *Coherence*: Pathological situations can arise in (9) when  $\mathbf{M}$  has very few non-zero coefficients. In the context of EDG, in extreme cases, this happens if only the diagonal elements of  $\mathbf{D}$  are sampled and/or there are many overlapping points. This motivates the notion of coherence first introduced in [10]. Let  $\mathbf{M} = \sum_{k=1}^r \lambda_k \mathbf{u}_k \mathbf{u}_k^T$ , where the eigenvectors have been chosen to be orthonormal and  $\lambda_k \geq 0$  as  $\mathbf{M} \in \mathbb{S}_+$ . We write  $\mathbf{U} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$  as the column space of  $\mathbf{M}$ ,  $\mathbf{U}^\perp = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_n\}$  as the orthogonal complement of  $\mathbf{U}$  and further denote  $\mathcal{P}_U$  and  $\mathcal{P}_{U^\perp}$  as the orthogonal projections onto  $\mathbf{U}$  and  $\mathbf{U}^\perp$ , respectively. Let  $\mathbb{T} = \{\mathbf{UZ}^T + \mathbf{ZU}^T : \mathbf{Z} \in \mathbb{R}^{n \times r}\}$  be the tangent space of the rank  $r$  matrix in  $\mathbb{S}_+$  at  $\mathbf{M}$ . The orthogonal projection onto  $\mathbb{T}$  is given by

$$\mathcal{P}_T \mathbf{X} = \mathcal{P}_U \mathbf{X} + \mathbf{X} \mathcal{P}_U - \mathcal{P}_U \mathbf{X} \mathcal{P}_U \quad (11)$$

It can be readily verified that  $\mathcal{P}_{T^\perp} \mathbf{X} = \mathbf{X} - \mathcal{P}_T \mathbf{X} = \mathcal{P}_{U^\perp} \mathbf{X} \mathcal{P}_{U^\perp}$ . The coherence conditions can now be defined as follows.

**Definition 1.** The aforementioned rank  $r$  matrix  $\mathbf{M} \in \mathbb{S}_+$  has coherence  $\nu$  with respect to basis  $\{\mathbf{w}_\alpha\}_{\alpha \in \mathbb{I}}$  if the following estimates hold

$$\max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_T \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle^2 \leq 2\nu \frac{r}{n} \quad (12)$$

$$\max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_T \mathbf{v}_\alpha, \mathbf{w}_\beta \rangle^2 \leq 4\nu \frac{r}{n} \quad (13)$$

$$\max_{\alpha \in \mathbb{I}} \langle \mathbf{w}_\alpha, \mathbf{U} \mathbf{U}^T \rangle^2 \leq \frac{1}{4} \nu \frac{r}{n^2} \quad (14)$$

where  $\{\mathbf{v}_\alpha\}_{\alpha \in \mathbb{I}}$  is the dual basis of  $\{\mathbf{w}_\alpha\}_{\alpha \in \mathbb{I}}$ .

**Remark 1.** If  $\{\mathbf{w}_\alpha\}$  is an orthonormal basis, it follows trivially that the dual basis  $\{\mathbf{v}_\alpha\}$  is also  $\{\mathbf{w}_\alpha\}$ . For this specific case, the above coherence conditions are equivalent, up to constants, to the coherence conditions in [29]. However, in the most general setting, the coherence conditions (12) and (13) depart from their orthonormal counterparts. This is because these conditions depend on the spectrum of the correlation matrix  $\mathbf{H}$ . Since the analysis in the sequel makes repeated use of the coherence conditions, the above equations are further simplified to convenient forms as follows. First, using (12), consider a bound on  $\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2$ . Using Lemma A.4 and Lemma A.6, one obtains

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle^2 \leq 2\nu \frac{r}{n} \longrightarrow \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq 2\nu \frac{r}{n}$$

Using the above inequality and Lemma A.4, one obtains the following bound for  $\|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F$ .

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F \leq \sum_{\beta \in \mathbb{I}} \|\mathbf{H}^{\alpha, \beta} \mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F = \sum_{\beta \in \mathbb{I}} |\mathbf{H}^{\alpha, \beta}| \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F \leq 2 \sqrt{\frac{2\nu r}{n}}$$

Next, consider a bound on  $\langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2$ . Using (14) and Lemma A.4, one arrives at the following inequality.

$$\langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2 = \left( \sum_{\beta \in \mathbb{I}} \mathbf{H}^{\alpha, \beta} \mathbf{w}_\beta, \mathbf{U}\mathbf{U}^T \right)^2 \leq \max_{\beta \in \mathbb{I}} \langle \mathbf{w}_\beta, \mathbf{U}\mathbf{U}^T \rangle^2 \left( \sum_{\beta \in \mathbb{I}} |\mathbf{H}^{\alpha, \beta}| \right)^2 \leq \frac{\nu r}{n^2}$$

All in all, the simplified forms of the coherence conditions are summarized as follows.

$$\max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq 2\nu \frac{r}{n} \quad (15)$$

$$\max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F^2 \leq 8\nu \frac{r}{n} \quad (16)$$

$$\max_{\alpha \in \mathbb{I}} \langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2 \leq \frac{\nu r}{n^2} \quad (17)$$

The simplified coherence conditions presented above are the same as the standard coherence assumptions up to constants (see [29], [31]). If the matrix  $\mathbf{M}$  has coherence  $\nu$  with respect to the standard basis, comparable bounds could be derived for the above coherence conditions. This is true since the basis  $\{\mathbf{w}_\alpha\}$  is not “far” from the standard basis. For a precise statement of this fact, we refer the interested reader to Lemma A.5. The implication of this fact is that an incoherent  $\mathbf{M}$  with respect to the standard basis is also incoherent with respect to the EDG basis. Intuitively, the coherence parameter is fundamentally about the concentration of information in the underlying matrix. For a matrix with low coherence(incoherent), each measurement is equally as informative as the other. In contrast, the information is concentrated on few measurements for a coherent matrix.

**Remark 2** (Coherence and Geometry). *In the context of the EDG problem, an interesting question is the relationship, if any, between coherence and the geometry of the underlying points. One analytical task is to relate the coherence, using the compressive sensing definition, of the Gram matrix to the coherence conditions (12)-(14). Computationally, numerical tests indicate that points arising from random distributions or points which are a random sample of structured points have comparable coherence values irrespective of the underlying distribution. However, since estimating coherence accurately and efficiently for large  $n$  is computationally expensive, the numerical tests were limited to relatively small values of  $n$ . The question of the geometry of points and how it relates to coherence is important to us and future work will explore this problem through analysis and extensive numerical tests.*

c) *Sampling Model:* The sampling scheme is an important element of the matrix completion problem. For the EDG completion problem, it is assumed that the basis vectors are sampled uniformly at random with replacement. Previous works have considered uniform sampling with replacement [29], [31] and Bernoulli sampling [10]. The implication of our choice is that the sampling process is independent. This property is essential as we make use of concentration inequalities for i.i.d matrix valued random variables.

**Remark 3.** *For the EDG completion problem, we have considered uniform sampling with out replacement. In this case, the sampling process is no longer independent. The implication is that one could not simply use concentration inequalities for i.i.d matrix valued random variables. However, as noted in the classic work of Hoeffding [34, section 6], the results derived for the case of the sampling with replacement also hold true for the case of sampling without replacement. An analogous argument for matrix concentration inequalities, resulting from the operator Chernoff bound technique [35], under uniform sampling with out replacement is shown in [36]. The analysis based on this choice leads to a “slightly lower” number of measurements  $m$ . Since the gain is minimal, which will be made apparent in later analysis, we use the uniform sampling with replacement model.*

### III. PROOF OF MAIN RESULT

The main goal of this section is to show that the nuclear norm minimization problem (9) admits a unique solution. Thus, the optimization problem provides the desired reconstruction. Under certain conditions, our main result guarantees a unique solution for the minimization problem (9). More precisely, we have the following theorem.

**Theorem 1.** *Let  $M \in \mathbb{R}^{n \times n}$  be a matrix of rank  $r$  that obeys the coherence conditions (12), (13) and (14) with coherence  $v$ . Assume  $m$  measurements,  $\{\langle M, w_\alpha \rangle\}_{\alpha \in \Omega}$ , are sampled uniformly at random with replacement. For  $\beta > 1$ , if*

$$m \geq nr \log_2(4\sqrt{8Lrn}) \left[ 96 \left( v + \frac{1}{nr} \right) (\beta \log(n) + \log(4 \log_2(4L\sqrt{r}))) \right]$$

*the solution to (9), equivalently (2) and (4), is unique and equal to  $M$  with probability at least  $1 - n^{-\beta}$ .*

The optimization problem in (9) is a convex minimization problem for which the optimality of a feasible solution  $X$  is equivalent to the sufficient KKT conditions. For details, we refer the interested reader to [10] where the authors derive a compact form of these optimality conditions. The overall structure of the proof is as follows. First, we show that if certain deterministic optimality and uniqueness conditions hold, then it certifies that  $M$  is a unique solution to the minimization problem. The remaining part of the proof will focus on showing that these conditions hold with very high probability under certain conditions. This in turn will imply that  $M$  is a unique solution with very high probability for an appropriate choice of  $m$ .

The proof of Theorem 1 closely follows the approach in [29]. For ease of presentation, the proof is divided into several intermediate results. Readers familiar with matrix completion proof in [29], [31] will recall that one crucial part of the proof is bounding the spectral norm of  $\|\mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T - \mathcal{P}_T\|$ . The interpretation of this bound is that the operator  $\mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T$  is almost isometric to  $\mathcal{P}_T$ . In the case of our problem,  $\mathcal{R}_\Omega$  is no longer self-adjoint and the equivalent statement is a bound on  $\|\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T - \mathcal{P}_T\|$ . Unfortunately, the term  $\|\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T - \mathcal{P}_T\|$  is not amenable to simpler analysis as standard concentration inequalities result sub-optimal success rates. The idea is instead to consider the operator  $\mathcal{P}_T \mathcal{F}_\Omega \mathcal{P}_T$  and show that the minimum eigenvalue is bounded away from 0 with high probability using the operator Chernoff bound. The implication is that the operator  $\mathcal{P}_T \mathcal{F}_\Omega \mathcal{P}_T$  is full rank on the space  $T$ . For non orthogonal measurements,  $\mathcal{P}_T \mathcal{F}_\Omega \mathcal{P}_T$  can be interpreted as the analogue of the operator  $\mathcal{P}_T \mathcal{R}_\Omega \mathcal{P}_T$ . In [30], for the standard matrix completion problem with measurement basis  $e_{ij}$ , it was argued theoretically that a lower bound for  $m$  is  $O(nr \log n)$ . Theorem 1 requires on the order of  $nrv \log^2 n$  measurements which is  $\log(n)$  multiplicative factor away from this sharp lower bound. We remark that, despite the aforementioned technical challenges, our result is of the same order as the result in [29], [31] which consider the low rank recovery problem with any orthogonal basis and the matrix completion problem respectively. We start the proof by stating and proving Theorem 2 which rigorously shows that if appropriate conditions as discussed earlier hold, then  $M$  is a unique solution to (9).

**Theorem 2.** *Given  $X \in \mathbb{S}_+$ , we write  $\Delta = X - M$  as the deviation from the underlying low rank matrix  $M$ . Let  $\Delta_T$  and  $\Delta_{T^\perp}$  be the orthogonal projection of  $\Delta$  to  $T$  and  $T^\perp$ , respectively. For any given  $\Omega$  with  $|\Omega| = m$ , the following two statements hold.*

- (a). *If  $\|\Delta_T\|_F^2 \geq (n\sqrt{8L}\|\Delta_{T^\perp}\|_F)^2$  and  $\lambda_{\min}(\mathcal{P}_T \mathcal{F}_\Omega \mathcal{P}_T) > \frac{1}{2}$ , then  $\mathcal{R}_\Omega \Delta \neq 0$ .*
- (b). *If  $\|\Delta_T\|_F^2 < (n\sqrt{8L}\|\Delta_{T^\perp}\|_F)^2$  for  $\Delta \in \ker \mathcal{R}_\Omega$ , and there exists a  $Y \in \text{range } \mathcal{R}_\Omega^*$  satisfying,*

$$\|\mathcal{P}_T Y - \text{Sgn } M\|_F \leq \frac{1}{4n\sqrt{8L}} \quad \text{and} \quad \|\mathcal{P}_{T^\perp} Y\| \leq \frac{1}{2} \tag{18}$$

*then  $\|X\|_* = \|M + \Delta\|_* > \|M\|_*$ .*

To interpret the above theorem, note that Theorem 2(a) states that any deviation from  $M$  fails to be in the null space of the sampling operator for “large”  $\Delta_T$ . For “small”  $\Delta_T$ , Theorem 2(b) claims that any deviation from  $M$  increases the nuclear norm thereby violating the minimization condition. We would like to emphasize that this is a deterministic theorem which does not depend on the random sampling of  $\Omega$  as long as the conditions in the statement are satisfied. In fact, we will later show that these conditions will hold with very high probability under certain sampling conditions and an appropriate choice of  $m = |\Omega|$ .

#### A. Proof of Theorem 2

*Proof of Theorem 2(a).* Suppose  $\|\Delta_T\|_F^2 \geq (n\sqrt{8L}\|\Delta_{T^\perp}\|_F)^2$ . To prove  $\mathcal{R}_\Omega \Delta \neq 0$ , it suffices to show  $\|\mathcal{R}_\Omega \Delta\|_F > 0$ . Note that  $\|\mathcal{R}_\Omega \Delta\|_F = \|\mathcal{R}_\Omega \Delta_T + \mathcal{R}_\Omega \Delta_{T^\perp}\|_F \geq \|\mathcal{R}_\Omega \Delta_T\|_F - \|\mathcal{R}_\Omega \Delta_{T^\perp}\|_F$ . This motivates finding a lower bound for  $\|\mathcal{R}_\Omega \Delta_T\|_F$  and an upper bound for  $\|\mathcal{R}_\Omega \Delta_{T^\perp}\|_F$ . For any  $X$ ,  $\|\mathcal{R}_\Omega X\|_F^2$  can be bounded as follows

$$\|\mathcal{R}_\Omega X\|_F^2 = \langle \mathcal{R}_\Omega X, \mathcal{R}_\Omega X \rangle = \langle X, \mathcal{R}_\Omega^* \mathcal{R}_\Omega X \rangle = \frac{L^2}{m^2} \sum_{\beta \in \Omega} \sum_{\alpha \in \Omega} \langle X, w_\alpha \rangle \langle v_\alpha, v_\beta \rangle \langle X, w_\beta \rangle = \frac{L^2}{m^2} \sum_{\beta \in \Omega} \sum_{\alpha \in \Omega} \langle X, w_\alpha \rangle H^{\alpha, \beta} \langle X, w_\beta \rangle$$

noting that  $\langle \mathbf{v}_\alpha, \mathbf{v}_\beta \rangle = \mathbf{H}^{\alpha, \beta}$ . Using the min-max theorem in the above equation, one obtains

$$\frac{L^2}{m^2} \lambda_{\min}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \|\mathcal{R}_\Omega \mathbf{X}\|_F^2 \leq \frac{L^2}{m^2} \lambda_{\max}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \frac{L^2}{m^2} \lambda_{\max}(\mathbf{H}^{-1}) \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \quad (19)$$

Using the fact  $\lambda_{\max}(\mathbf{H}^{-1}) = 1$  (Lemma A.4) and  $\sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq 2n \|\mathbf{X}\|_F^2$  (Lemma A.6) and setting  $\mathbf{X} = \Delta_{\mathbb{T}^\perp}$  in the right of the above inequality results

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F^2 \leq \frac{2L^2}{m^2} nm \|\Delta_{\mathbb{T}^\perp}\|_F^2 \quad (20)$$

In the above inequality, the constant  $m$  bounds the maximum number of repetitions for any given measurement. Next, the lower bound for  $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F$  is considered using the left inequality in (19).

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 \geq \frac{L^2}{2m^2 n} \sum_{\alpha \in \Omega} \langle \Delta_{\mathbb{T}}, \mathbf{w}_\alpha \rangle^2 = \frac{L}{2mn} \langle \Delta_{\mathbb{T}}, \frac{L}{m} \sum_{\alpha \in \Omega} \langle \Delta_{\mathbb{T}}, \mathbf{w}_\alpha \rangle \mathbf{w}_\alpha \rangle = \frac{L}{2mn} \langle \Delta_{\mathbb{T}}, \mathcal{F}_\Omega \Delta_{\mathbb{T}} \rangle \quad (21)$$

with the fact that  $\lambda_{\min}(\mathbf{H}^{-1}) = \frac{1}{2n}$  (Lemma A.4). Using the min-max theorem on the projection onto  $\mathbb{T}$  of the restricted frame operator, we have

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 \geq \frac{L}{2mn} \langle \Delta_{\mathbb{T}}, \mathcal{F}_\Omega \Delta_{\mathbb{T}} \rangle = \frac{L}{2mn} \langle \Delta_{\mathbb{T}}, \mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} \rangle \geq \frac{L}{2mn} \lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) \|\Delta_{\mathbb{T}}\|_F^2 \quad (22)$$

Above, the first equality holds since  $\mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} = \Delta_{\mathbb{T}}$  and  $\mathcal{P}_{\mathbb{T}}$  is self-adjoint. The last inequality simply applies the min-max theorem on the self-adjoint operator  $\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}$ . Using the assumption that  $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{1}{2}$ , the inequality in (22) reduces to

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 > \frac{L}{4mn} \|\Delta_{\mathbb{T}}\|_F^2 \quad (23)$$

Combining (20) and (23), we have

$$\begin{aligned} \|\mathcal{R}_\Omega \Delta\|_F &> \sqrt{\frac{L}{4mn}} \|\Delta_{\mathbb{T}}\|_F - \frac{L}{m} \sqrt{2n} \|\Delta_{\mathbb{T}^\perp}\|_F \\ &\geq \sqrt{\frac{L}{4mn}} (n \sqrt{8L}) \|\Delta_{\mathbb{T}^\perp}\|_F - \frac{L}{m} \sqrt{2nm} \|\Delta_{\mathbb{T}^\perp}\|_F = 0 \end{aligned}$$

where the last inequality follows from the theorem's assumption. This concludes proof of Theorem 2(a).  $\square$

**Remark 4.** (1) The estimate of the upper bound of  $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F^2$  is not sharp. Specifically, the constant  $m$  can be much lowered. This can be achieved, for example, by employing standard concentration inequalities and arguing that the expected number of duplicate measurements is substantially smaller than  $m$ . Alternatively, as noted earlier in the sampling model section, uniform sampling with out replacement can be adopted which has the implication that the constant  $m$  is not necessary. Both analysis lead to a better estimate of the upper bound. However, the gain in terms of measurements,  $m$ , is minimal, resulting lower constants inside of a  $\log_2$ . We use the current estimate, albeit not sharp, for sake of more compact analysis. (2) Lower bounding the term  $\sum_{\alpha \in \Omega} \langle \Delta_{\mathbb{T}}, \mathbf{w}_\alpha \rangle^2$  is not amenable to simpler analysis. In fact, a mere application of the standard concentration inequalities result an undesirable probability of failure that increases with  $n$ . Note that  $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 = \langle \Delta_{\mathbb{T}}, \mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} \rangle + \|\mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}}\|_F^2$ . A lower bound for  $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2$  motivates an upper bound of  $\|\mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}}\|$ . This bound can be achieved, albeit long calculations, but it relies on the special structure of  $\mathbf{w}_\alpha$ . To make the technical analysis simple and more general, we use an alternative approach shown in the above proof. This is a different way to handle the lower bound estimation of  $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2$ .

*Proof of Theorem 2(b).* Consider a feasible solution  $\mathbf{X} = \mathbf{M} + \Delta \in \mathbb{S}_+$  satisfying  $\|\Delta_{\mathbb{T}}\|_F < n \sqrt{8L} \|\Delta_{\mathbb{T}^\perp}\|_F$  and  $\Delta \in \ker \mathcal{R}_\Omega$ . We need to show that  $\|\mathbf{M} + \Delta\|_* > \|\mathbf{M}\|_*$  which implies that nuclear minimization is violated. The proof of this requires the construction of a dual certificate  $\mathbf{Y}$  satisfying certain conditions. The proof is similar to the proof in section 2E of [29] but it is shown below for completeness and ease of reference in later analysis. Using the pinching inequality [37],  $\|\mathbf{M} + \Delta\|_* \geq \|\mathcal{P}_{\mathbf{U}}(\mathbf{M} + \Delta)\mathcal{P}_{\mathbf{U}}\|_* + \|\mathcal{P}_{\mathbf{U}^\perp}(\mathbf{M} + \Delta)\mathcal{P}_{\mathbf{U}^\perp}\|_*$ . Noting that  $\mathcal{P}_{\mathbf{U}} \mathbf{M} \mathcal{P}_{\mathbf{U}} = \mathbf{M}$ ,  $\mathcal{P}_{\mathbf{U}^\perp} \mathbf{M} \mathcal{P}_{\mathbf{U}^\perp} = \mathbf{0}$  and  $\mathcal{P}_{\mathbf{U}^\perp} \Delta \mathcal{P}_{\mathbf{U}^\perp} = \Delta_{\mathbb{T}^\perp}$ , the above inequality reduces to

$$\|\mathbf{M} + \Delta\|_* \geq \|\mathbf{M} + \mathcal{P}_{\mathbf{U}} \Delta \mathcal{P}_{\mathbf{U}}\|_* + \|\Delta_{\mathbb{T}^\perp}\|_* \quad (24)$$

Note that  $\|\Delta_{\mathbb{T}^\perp}\|_* = \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle$ . The first term on right hand side can be lower bounded using the fact that the nuclear norm and the spectral norm are dual to one another. Stated precisely,  $\|\mathbf{X}_2\|_* = \sup_{\|\mathbf{X}_1\|_1 \leq 1} \langle \mathbf{X}_1, \mathbf{X}_2 \rangle$  for any  $\mathbf{X}_1$ . Using this inequality with  $\mathbf{X}_1 = \text{Sgn } \mathbf{M}$  and  $\mathbf{X}_2 = \mathbf{M} + \mathcal{P}_{\mathbf{U}} \Delta \mathcal{P}_{\mathbf{U}}$  in (24) results

$$\begin{aligned} \|\mathbf{M} + \mathcal{P}_{\mathbf{U}} \Delta \mathcal{P}_{\mathbf{U}}\|_* + \|\Delta_{\mathbb{T}^\perp}\|_* &\geq \langle \text{Sgn } \mathbf{M}, \mathbf{M} + \mathcal{P}_{\mathbf{U}} \Delta \mathcal{P}_{\mathbf{U}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle \\ &= \|\mathbf{M}\|_* + \langle \text{Sgn } \mathbf{M}, \mathcal{P}_{\mathbf{U}} \Delta \mathcal{P}_{\mathbf{U}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle \end{aligned}$$

Note that  $\text{Sgn } M \in \mathbb{T}$  (see Lemma A.1). Using this fact, it can be verified that  $\langle \text{Sgn } M, \mathcal{P}_U \Delta_{\mathbb{T}^\perp} \mathcal{P}_U \rangle = 0$  and  $\langle \text{Sgn } M, \mathcal{P}_U \Delta_{\mathbb{T}} \mathcal{P}_U \rangle = \langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle$ . The above inequality can now be written as

$$\|M + \mathcal{P}_U \Delta \mathcal{P}_U\|_* + \|\Delta_{\mathbb{T}^\perp}\|_* \geq \|M\|_* + \langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle \quad (25)$$

Using (24) and (25), it can be concluded that  $\|M + \Delta\|_* > \|M\|_*$  if it can be shown that  $\langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle > 0$ . Since  $\langle Y, \Delta \rangle = 0$  for any  $Y \in \text{range } \mathcal{R}_\Omega^*$ , we have

$$\begin{aligned} \langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle &= \langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle - \langle Y, \Delta \rangle \\ &= \langle \text{Sgn } M - \mathcal{P}_{\mathbb{T}} Y, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle - \langle \mathcal{P}_{\mathbb{T}^\perp} Y, \Delta_{\mathbb{T}^\perp} \rangle \end{aligned}$$

Assuming the conditions in the statement of the theorem and considering the last equation, we obtain

$$\begin{aligned} \langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle &\geq -\frac{1}{4n\sqrt{8L}} \|\Delta_{\mathbb{T}}\|_F + \|\Delta_{\mathbb{T}^\perp}\|_* - \frac{1}{2} \|\Delta_{\mathbb{T}^\perp}\|_* \\ &\geq -\frac{1}{4n\sqrt{8L}} \|\Delta_{\mathbb{T}}\|_F + \frac{1}{2} \|\Delta_{\mathbb{T}^\perp}\|_F > -\frac{1}{4n\sqrt{8L}} (n\sqrt{8L} \|\Delta_{\mathbb{T}^\perp}\|_F) + \frac{1}{2} \|\Delta_{\mathbb{T}^\perp}\|_F = \frac{1}{4} \|\Delta_{\mathbb{T}^\perp}\|_F \end{aligned}$$

Above, the first inequality follows from the duality of the spectral norm and the nuclear norm. It has been shown that  $\langle \text{Sgn } M, \Delta_{\mathbb{T}} \rangle + \langle \text{Sgn } \Delta_{\mathbb{T}^\perp}, \Delta_{\mathbb{T}^\perp} \rangle > 0$  under the assumptions of Theorem 2(b). Using (24) and (25), it follows that  $\|M + \Delta\|_* > \|M\|_*$  concluding the proof of Theorem 2(b).  $\square$

Consequently, if the deterministic conditions in Theorem 2 hold,  $M$  is a unique solution to (9). We formally write it as a corollary which can help us to prove the probabilistic statement in Theorem 1.

**Corollary 1.** *If the conditions of Theorem 2 hold,  $M$  is a unique solution to (9).*

*Proof.* For any  $X \in \mathbb{S}_+$ , define  $\Delta = M - X$ . Using Theorem 2(a),  $\mathcal{R}_\Omega \Delta \neq 0$  if  $\|\Delta_{\mathbb{T}}\|_F^2 \geq (n\sqrt{8L} \|\Delta_{\mathbb{T}^\perp}\|_F)^2$ . It then suffices to consider the case  $\|\Delta_{\mathbb{T}}\|_F^2 < (n\sqrt{8L} \|\Delta_{\mathbb{T}^\perp}\|_F)^2$  for  $\Delta \in \ker \mathcal{R}_\Omega$ . For this case, the proof of Theorem 2(b) shows that  $\|X\|_* > \|M\|_*$ . It can then be concluded that  $M$  is the unique solution to (9).  $\square$

### B. Proof of Theorem 1

It follows that certifying that the two conditions in Theorem 2 hold implies a unique solution to (9). Hence, the main task in the proof is to show that, under the assumptions of Theorem 1, these two conditions hold with very high probability. If this can be achieved, it implies that the conclusion of Theorem 1 holds true with the same high probability. The first condition in Theorem 2 is that  $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{1}{2}$ . The goal is to show that the minimum eigenvalue of the operator  $\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}$  is bounded away from zero. This will be proven in Lemma 1 to follow shortly. The lemma makes use of the matrix Chernoff bound in [38] and is restated below for convenience.

**Theorem 3.** *Let  $\{\mathcal{L}_k\}$  be a finite sequence of independent, random, self-adjoint operators satisfying*

$$\mathcal{L}_k \succeq 0 \quad \text{and} \quad \|\mathcal{L}_k\| \leq R \quad \text{almost surely}$$

*with the minimum and maximum eigenvalues of the sum of the expectations,*

$$\mu_{\min} := \lambda_{\min} \left( \sum_k E[\mathcal{L}_k] \right) \quad \text{and} \quad \mu_{\max} := \lambda_{\max} \left( \sum_k E[\mathcal{L}_k] \right)$$

*Then, we have*

$$Pr \left[ \lambda_{\min} \left( \sum_k \mathcal{L}_k \right) \leq (1 - \delta) \mu_{\min} \right] \leq n \left[ \frac{\exp(-\delta)}{(1 - \delta)^{1-\delta}} \right]^{\frac{\mu_{\min}}{R}} \quad \text{for } \delta \in [0, 1]$$

For  $\delta \in [0, 1]$ , using Taylor series of  $\log(1 - \delta)$ , note that  $(1 - \delta) \log(1 - \delta) \geq -\delta + \frac{\delta^2}{2}$ . This results the following simplified estimate

$$Pr \left[ \lambda_{\min} \left( \sum_k \mathcal{L}_k \right) \leq (1 - \delta) \mu_{\min} \right] \leq n \exp \left( -\delta^2 \frac{\mu_{\min}}{2R} \right) \quad \text{for } \delta \in [0, 1]$$

**Lemma 1.** *Given the operator  $\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}} : \mathbb{T} \rightarrow \mathbb{T}$  with  $\kappa = \frac{m}{nr}$ , the following estimate holds.*

$$Pr \left( \lambda_{\min} (\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) \leq \frac{1}{2} \right) \leq n \exp \left( -\frac{\kappa}{8\nu} \right)$$

*Proof.* Using the definition of  $\mathcal{F}_\Omega(X) = \frac{L}{m} \sum_{\alpha \in \Omega} \langle X, w_\alpha \rangle w_\alpha$ , for  $X \in \mathbb{T}$ ,  $\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}} X$  can be written as follows

$$\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}} X = \sum_{\alpha \in \Omega} \frac{L}{m} \langle X, \mathcal{P}_{\mathbb{T}} w_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha$$

Let  $\mathcal{L}_\alpha$  denote the operator in the summand,  $\mathcal{L}_\alpha = \frac{L}{m} \langle \cdot, \mathcal{P}_{\mathbb{T}} w_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha$ . It can be easily verified that  $\mathcal{L}_\alpha$  is positive semidefinite. The operator Chernoff bound requires estimate of  $R$  and  $\mu_{\min}$ .  $R$  can be estimated as follows.

$$\left\| \frac{L}{m} \langle \cdot, \mathcal{P}_{\mathbb{T}} w_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha \right\| = \frac{L}{m} \|\mathcal{P}_{\mathbb{T}} w_\alpha\|_F^2 \leq \frac{nvr}{m}$$

The last inequality follows from the coherence estimate in (15). Therefore,  $R = \frac{nvr}{m}$ . Next, we consider  $\mu_{\min}$ . The first step is to compute  $\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha]$ .

$$\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha] = \sum_{\alpha \in \Omega} \left[ \sum_{\alpha} \frac{1}{m} \langle \cdot, \mathcal{P}_{\mathbb{T}} w_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha \right] = \sum_{\alpha} \langle \cdot, \mathcal{P}_{\mathbb{T}} w_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha$$

For any  $\mathbf{X} \in \mathbb{T}$ , lower bound  $\langle \mathbf{X}, \sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha](\mathbf{X}) \rangle$  as follows.

$$\langle \mathbf{X}, \sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha](\mathbf{X}) \rangle = \sum_{\alpha} \langle \mathbf{X}, w_\alpha \rangle^2 \geq \|\mathbf{X}\|_F^2$$

The inequality results from Lemma A.4 and Lemma A.6. Noting that  $\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha]$  is a self-adjoint operator, and using the variational characterization of the minimum eigenvalue, it can be concluded that the minimum eigenvalue of  $\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha]$  is at least 1. With this, set  $\mu_{\min} = 1$ . Finally, apply the matrix Chernoff bound with  $R = \frac{nvr}{m}$  and  $\mu_{\min} = 1$ . Setting  $\delta = \frac{1}{2}$ ,  $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{1}{2}$  with probability of failure at most  $p_1$  given by

$$p_1 = n \exp\left(-\frac{\kappa}{8\nu}\right)$$

This concludes the proof.  $\square$

Under the assumptions of Theorem 1 and using Lemma 1,  $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{1}{2}$  holds with probability  $1 - p_1$  where the probability of failure is  $p_1 = n \exp\left(-\frac{\kappa}{8\nu}\right)$  with  $\kappa = \frac{m}{nr}$ .

The proof of the second condition in Theorem 2 is more technically involved and requires the construction of  $\mathbf{Y}$  satisfying the conditions in (18). The construction follows the ingenious golfing scheme introduced in [29]. Partition the basis elements in  $\Omega$  into  $l$  batches with the  $i$ -th batch  $\Omega_i$  containing  $m_i$  elements and  $\sum_{i=1}^l m_i = m$ . The restriction operator for a batch  $\Omega_i$  is defined as  $\mathcal{R}_i = \frac{L}{m_i} \sum_{\alpha \in \Omega_i} \langle \mathbf{X}, w_\alpha \rangle v_\alpha$ . Then  $\mathbf{Y} = \mathbf{Y}_l$  is constructed using the following recursive scheme

$$\mathbf{Q}_0 = \text{sgn } \mathbf{M}, \quad \mathbf{Y}_i = \sum_{j=1}^i \mathcal{R}_j^* \mathbf{Q}_{j-1}, \quad \mathbf{Q}_i = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \mathbf{Y}_i, \quad i = 1, \dots, l \quad (26)$$

Using the golfing scheme, it will be shown that the conditions in (18) hold with very high probability. The technical analysis of the scheme, particularly certifying the first condition in (18), requires a probabilistic estimate of  $\|\mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}\|_F \geq t$  for a fixed matrix  $\mathbf{X} \in \mathbb{S}$ . Using Lemma 2 to follow, a suitable estimate can be achieved. The lemma uses the vector Bernstein inequality in [29]. An easier but slightly weaker version is restated below for convenience of reference.

**Theorem 4** (Vector Bernstein inequality). *Let  $\mathbf{x}_1, \dots, \mathbf{x}_m$  be independent zero-mean vector valued random variables. Assuming that  $\max_i \|\mathbf{x}_i\|_2 \leq R$  and let  $\sum_i E[\|\mathbf{x}_i\|_2^2] \leq \sigma^2$ , then for any  $t \leq \frac{\sigma^2}{R}$ , the following inequality holds*

$$\Pr\left[\left\| \sum_{i=1}^m \mathbf{x}_i \right\|_2 \geq t\right] \leq \exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4}\right),$$

**Lemma 2.** *For a fixed matrix  $\mathbf{X} \in \mathbb{S}$ , the following estimate holds*

$$\Pr(\|\mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}\|_F \geq t \|\mathbf{X}\|_F) \leq \exp\left(-\frac{t^2 \kappa}{16\left(\nu + \frac{1}{nr}\right)} + \frac{1}{4}\right) \quad (27)$$

for all  $t \leq 1$  with  $\kappa = \frac{m}{nr}$ .

*Proof.* It is clear that  $\mathcal{P}_{\mathbb{T}} \mathbf{X} \in \mathbb{S}$  from the definition of  $\mathcal{P}_{\mathbb{T}}$ . Thus, we can expand  $\mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{P}_{\mathbb{T}} \mathbf{X}$  as follows.

$$\mathcal{P}_{\mathbb{T}} \mathcal{R}_\Omega^* \mathcal{P}_{\mathbb{T}} \mathbf{X} = \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_\alpha \rangle \mathcal{P}_{\mathbb{T}} w_\alpha \quad (28)$$

With this,  $\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}$  can be written as follows

$$\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X} = \sum_{\alpha \in \Omega} \left[ \frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} w_{\alpha} - \frac{1}{m} \mathcal{P}_{\mathbb{T}} \mathbf{X} \right] \quad (29)$$

Let  $\mathbf{X}_{\alpha} = \frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} w_{\alpha}$ . Noting that  $E[\frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} w_{\alpha}] = \frac{1}{m} \mathcal{P}_{\mathbb{T}} \mathbf{X}$ , let  $\mathbf{Y}_{\alpha} = \mathbf{X}_{\alpha} - E[\mathbf{X}_{\alpha}]$  denote the summand. It is clear that  $E[\mathbf{Y}_{\alpha}] = \mathbf{0}$  by construction. The vector Bernstein inequality requires a suitable bound for  $\|\mathbf{Y}_{\alpha}\|_F$  and  $E[\|\mathbf{Y}_{\alpha}\|_F^2]$ . Without loss of generality, assume  $\|\mathbf{X}\|_F = 1$ . The first step is to bound  $\|\mathbf{Y}_{\alpha}\|_F$ .

$$\|\mathbf{Y}_{\alpha}\|_F = \left\| \frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} w_{\alpha} - \frac{1}{m} \mathcal{P}_{\mathbb{T}} \mathbf{X} \right\|_F \leq \left\| \frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} w_{\alpha} \right\|_F + \left\| \frac{1}{m} \mathcal{P}_{\mathbb{T}} \mathbf{X} \right\|_F \quad (30)$$

$$\begin{aligned} &< \frac{n^2}{2m} \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} w_{\alpha}\|_F \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} v_{\alpha}\|_F + \frac{1}{m} \\ &\leq \frac{1}{m} (2nvr + 1) \end{aligned} \quad (31)$$

Above, the last inequality follows from the coherence estimates (15) and (16). Using the above estimate, set  $R$  in the vector Bernstein inequality as  $R = \frac{1}{m} (2nvr + 1)$ . Using the parallelogram identity and monotonicity of expectation,  $E[\|\mathbf{Y}_{\alpha}\|_F^2] \leq 2E[\|\mathbf{X}_{\alpha}\|_F^2] + 2E[\|\mathbf{X}_{\alpha}\|_F^2]$ . With this,  $E[\|\mathbf{Y}_{\alpha}\|_F^2]$  can be upper bounded as follows

$$\begin{aligned} E[\|\mathbf{Y}_{\alpha}\|_F^2] &\leq \frac{2L^2}{m^2} E[\langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle^2 \|\mathcal{P}_{\mathbb{T}} w_{\alpha}\|_F^2] + \frac{2}{m^2} \|\mathcal{P}_{\mathbb{T}} \mathbf{X}\|_F^2 \\ &\leq \frac{2L}{m^2} \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} w_{\alpha}\|_F^2 \sum_{\alpha \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, v_{\alpha} \rangle^2 + \frac{2}{m^2} \\ &\leq \frac{n^2}{m^2} \frac{2vr}{n} + \frac{2}{m^2} \leq \frac{2}{m^2} (1 + nvr) \end{aligned}$$

Above, the third inequality follows from the coherence estimates, (15) and (16), and Lemma A.6. Using the above estimate, set  $\sigma^2$  in vector Bernstein inequality as  $\sigma^2 = \frac{2}{m} (1 + nvr)$ . Finally, apply the vector Bernstein inequality with  $R = \frac{1}{m} (2nvr + 1)$  and  $\sigma^2 = \frac{2}{m} (1 + nvr)$ . For  $t \leq \frac{\sigma^2}{R} > 1$ ,

$$\Pr(\|\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}\|_F \geq t) \leq \exp\left(-\frac{t^2 \kappa}{16(v + \frac{1}{nr})} + \frac{1}{4}\right) \quad (32)$$

where  $\kappa = \frac{m}{nr}$ . This concludes the proof of Lemma 2.  $\square$

Having proved Lemma 2, the next step is to show that the golfing scheme (26) certifies the conditions in (18) with very high probability. The precise statement is stated in Lemma 3 below.

**Lemma 3.**  $\mathbf{Y}_l$  obtained from the golfing scheme (26) satisfies the conditions in (18) with failure probability at most  $p = \sum_{i=1}^l (p_2(i) + p_3(i) + p_4(i))$  where  $p_2(i) = \exp\left(\frac{-\kappa_i}{64(v + \frac{1}{nr})} + \frac{1}{4}\right)$ ,  $p_3(i) = n \exp\left(\frac{-3\kappa_i}{128v}\right)$  and  $p_4(i) = n^2 \exp\left(\frac{-3\kappa_i}{64(v + \frac{1}{nr})}\right)$  with  $\kappa_i = \frac{m_i}{nr}$ .

*Proof.* Note that  $\text{sgn } \mathbf{M}$  is symmetric. It is easy to verify that  $\mathbf{Q}_i$  is symmetric as  $\mathbf{Y}_i$  and  $\mathcal{P}_{\mathbb{T}} \mathbf{Y}_i$  are symmetric. In addition, using Lemma A.1,  $\mathbf{Q}_i$  is in  $\mathbb{T}$  since  $\text{sgn } \mathbf{M} \in \mathbb{T}$ . To show that the first condition in (18) holds, we derive a recursive formula for  $\mathbf{Q}_i$  as follows.

$$\begin{aligned} \mathbf{Q}_i &= \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \left( \sum_{j=1}^i \mathcal{R}_j^* \mathbf{Q}_{j-1} \right) = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \left( \sum_{j=1}^{i-1} \mathcal{R}_j^* \mathbf{Q}_{j-1} + \mathcal{R}_i^* \mathbf{Q}_{i-1} \right) \\ &= \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \sum_{j=1}^{i-1} \mathcal{R}_j^* \mathbf{Q}_{j-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \mathbf{Y}_{i-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} \\ &= \mathbf{Q}_{i-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} = (\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathcal{P}_{\mathbb{T}}) \mathbf{Q}_{i-1} \end{aligned} \quad (33)$$

Observe that the first condition in (18) is equivalent to a bound on  $\|\mathbf{Q}_i\|_F$ . Using the bound  $\|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathcal{P}_{\mathbb{T}}) \mathbf{Q}_{i-1}\|_F < t_{2,i} \|\mathbf{Q}_{i-1}\|_F$  with failure probability  $p_2(i)$  obtained from Lemma 2 and setting  $t_{2,i} = 1/2$ ,  $p_2(i) = \exp\left(\frac{-\kappa_i}{64(v + \frac{1}{nr})} + \frac{1}{4}\right)$  with  $\kappa_i = \frac{m_i}{nr}$ , we have the following bound of  $\|\mathbf{Q}_i\|_F$  using the above recursive formula.

$$\|\mathbf{Q}_i\|_F < \left( \prod_{k=1}^i t_{2,k} \right) \|\mathbf{Q}_0\|_F = 2^{-i} \sqrt{r} \quad (34)$$

To satisfy the first condition in (18), set  $l = \log_2(4n\sqrt{8Lr})$ . Using the union bound on the failure probabilities  $p_2(i)$ , we have  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y} - \text{Sgn } M\|_F = \|\mathbf{Q}_l\|_F \leq \sqrt{r}2^{-l} = \frac{1}{4n\sqrt{8L}}$  holds true with failure probability at most  $\sum_{i=1}^l p_2(i)$ .

To complete the proof, it remains to show that  $\mathbf{Y}$  satisfies the second condition in (18). The condition is equivalent to controlling the operator norm of  $\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l$ . First, it is clear that  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l\| = \|\sum_{j=1}^l \mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\| \leq \sum_{j=1}^l \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\|$ . This motivates bounding the operator norm of  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\|$  which will be the focus of Lemma 4 below. Before proving the lemma, we start by addressing an assumption the lemma entails on the size of  $\eta(\mathbf{Q}_i) = \max_{\beta} |\langle \mathbf{Q}_i, \mathbf{v}_\beta \rangle|$ . Specifically, at the  $i$ -th stage of the golfing scheme, we need to show that the assumption  $\eta(\mathbf{Q}_i)^2 \leq \frac{\nu}{n^2} c_i^2$  with  $\|\mathbf{Q}_i\|_F^2 \leq c_i^2$  holds with very high probability. To enforce this, let  $\eta(\mathbf{Q}_i) < t_{4,i}$  with probability  $1 - p_4(i)$ . We set  $t_{4,i} = \frac{1}{2}\eta(\mathbf{Q}_{i-1})$  to obtain

$$\eta(\mathbf{Q}_i)^2 < 2^{-2}\eta(\mathbf{Q}_{i-1})^2 < 2^{-2i}\eta(\text{sgn } M)^2 \leq \frac{\nu r}{n^2}(2^{-2i}) = \frac{\nu}{n^2}(2^{-2i}r)$$

Above, the second inequality is obtained by applying the inequality  $\eta(\mathbf{Q}_i) < \frac{1}{2}\eta(\mathbf{Q}_{i-1})$  recursively and the last inequality follows from the coherence condition in (17). Using (34), at the  $i$ -th stage of the golfing scheme, the above inequality precisely enforces the condition that  $\eta(\mathbf{Q}_i)^2 \leq \frac{\nu}{n^2} c_i^2$  with  $c_i = 2^{-i} \sqrt{r}$ . Using Lemma A.8, noting that  $\eta(\mathbf{Q}_i) = \eta(\mathbf{Q}_{i-1} - \mathcal{P}_T \mathcal{R}_i^* \mathbf{Q}_{i-1})$ , the failure probability  $p_4(i)$  is given by

$$p_4(i) = n^2 \exp\left(\frac{-3\kappa_i}{64\left(\nu + \frac{1}{nr}\right)}\right) \quad \forall i \in [1, l]$$

This concludes the analysis on the assumption of the size of  $\eta(\mathbf{Q}_i)$  which will be used freely in Lemma 4. Using the bound  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\| \leq t_{3,j} c_{j-1}$ , where once again  $c_{j-1}$  satisfies  $\|\mathbf{Q}_{j-1}\|_F \leq c_{j-1} = 2^{-(j-1)}$ , with failure probability  $p_3(j)$  obtained from Lemma 4 and setting  $t_{3,j} = \frac{1}{4\sqrt{r}}$ ,  $p_3(j) = n \exp\left(\frac{-3\kappa_j}{128\nu}\right)$  with  $\kappa = \frac{m_j}{nr}$ , it follows that

$$\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l\| \leq \sum_{k=1}^l \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_k^* \mathbf{Q}_{k-1}\| \leq \sum_{k=1}^l t_{3,k} c_{k-1} = \frac{1}{4\sqrt{r}} \sum_{k=1}^l c_{k-1} = \frac{1}{4\sqrt{r}} \sum_{k=1}^l \sqrt{r}2^{-(k-1)} < \frac{1}{2}$$

where the second to the last inequality uses (34) to bound  $\|\mathbf{Q}_{k-1}\|_F$ . Using the union bound of failure probabilities,  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l\| < \frac{1}{2}$  holds true with failure probability which is at most  $\sum_{j=1}^l [p_3(j) + p_4(j)]$ . The second condition in (18) now holds with this failure probability. This completes the proof of Lemma 3.  $\square$

The proof of Lemma 4 uses the Bernstein inequality. A simplified version of the inequality was derived in [38]. Our analysis uses this simpler version and is restated below for ease of reference.

**Theorem 5** (Bernstein inequality). *Consider a finite sequence  $\{\mathbf{X}_i\}$  of independent, random, self-adjoint matrices with dimension  $n$ . Assuming that*

$$E[\mathbf{X}_i] = 0 \quad \text{and} \quad \lambda_{\max}(\mathbf{X}_i) \leq R \quad \text{almost surely.}$$

*with the norm of the total variance  $\sigma^2 = \left\| \sum_i E(\mathbf{X}_i^2) \right\|$ , then, for all  $t \geq 0$ , the following estimate holds:*

$$Pr\left[\left\| \sum_i \mathbf{X}_i \right\| > t\right] \leq \begin{cases} n \exp\left(-\frac{3t^2}{8\sigma^2}\right) & t \leq \frac{\sigma^2}{R} \\ n \exp\left(-\frac{3t}{8R}\right) & t \geq \frac{\sigma^2}{R} \end{cases} \quad (35)$$

Now we are ready to estimate  $\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\|$  formally described in the following lemma.

**Lemma 4.** *Consider a fixed matrix  $\mathbf{G} \in \mathbb{T}$ . Assume that  $\max_{\beta} \langle \mathbf{G}, \mathbf{v}_\beta \rangle^2 \leq \frac{\nu}{n^2} c^2$  with  $c$  satisfying  $\|\mathbf{G}\|_F^2 \leq c^2$ . Then, the following estimate holds for all  $t \leq \frac{1}{\sqrt{r}}$  with  $\kappa_j = \frac{m_j}{nr}$ .*

$$Pr(\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G}\| > tc) \leq n \exp\left(-\frac{3t^2\kappa_j r}{8\nu}\right)$$

*Proof of Lemma 4.* We expand  $\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G}$  in the dual basis as  $\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G} = \sum_{\alpha \in \Omega_j} \frac{L}{m_j} \langle \mathbf{G}, \mathbf{v}_\alpha \rangle \mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha$ . Let  $\mathbf{X}_\alpha$  denote the summand.

The proof makes use of Bernstein inequality (35) which mandates appropriate bound on  $\|\mathbf{X}_\alpha\|$  and  $\|E[\mathbf{X}_\alpha^2]\|$ . First consider the

bound on the latter term. Noting that  $\mathbf{X}_\alpha^2 = \frac{L^2}{m_j^2} \langle \mathbf{G}, \mathbf{v}_\alpha \rangle^2 (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2$ , we have the following bound for  $E[\mathbf{X}_\alpha]^2$  using Lemma A.7 and the fact that  $\mathbf{w}_\alpha^2$  is positive semidefinite.

$$\|E[\mathbf{X}_\alpha^2]\| \leq \frac{n^2}{2m_j^2} \max_{\|\varphi\|_2=1} \sum_{\alpha \in \mathbb{I}} \langle \mathbf{G}, \mathbf{v}_\alpha \rangle^2 \langle \varphi, \mathbf{w}_\alpha^2 \varphi \rangle \leq \frac{n^2}{2m_j^2} \max_{\alpha \in \mathbb{I}} \langle \mathbf{v}_\alpha, \mathbf{G} \rangle^2 \max_{\|\varphi\|_2=1} \langle \varphi, \left( \sum_{\alpha \in \mathbb{I}} \mathbf{w}_\alpha^2 \right) \varphi \rangle \quad (36)$$

Due to the special structure of  $\mathbf{w}_\alpha$  in the EDG problem, it is straightforward to verify that

$$\sum_{\alpha \in \mathbb{I}} \mathbf{w}_\alpha^2 = 2n\mathbf{I} - 2\mathbf{1}\mathbf{1}^T$$

It now follows that  $\lambda_{\max}(\sum_{\alpha \in \mathbb{I}} \mathbf{w}_\alpha^2) = 2n$  which implies that

$$\|E[\mathbf{X}_\alpha^2]\| \leq \frac{n^3}{m_j^2} \max_{\alpha \in \mathbb{I}} \langle \mathbf{v}_\alpha, \mathbf{G} \rangle^2 \leq \frac{n^3}{m_j^2} \frac{\nu}{n^2} c^2 = \frac{n\nu}{m_j^2} c^2$$

Next, consider a bound on  $\|\mathbf{X}_\alpha\|$  which results

$$\|\mathbf{X}_\alpha\| \leq \frac{n^2}{2m_j} \max_{\alpha \in \mathbb{I}} \langle \mathbf{v}_\alpha, \mathbf{G} \rangle \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha\| \leq \frac{n^2}{2m_j} \frac{\sqrt{\nu}}{n} \|\mathbf{w}_\alpha\| c \quad (37)$$

We consider two cases. If  $\nu \geq \frac{1}{r}$ ,  $\|\mathbf{X}_\alpha\| \leq \frac{n\nu\sqrt{r}}{m_j} c = R_1$  and if  $\nu < \frac{1}{r}$ ,  $\|\mathbf{X}_\alpha\| < \frac{n\sqrt{\nu}}{m_j} c = R_2$ . Finally, we apply the Bernstein inequality with  $R_1, R_2$  and  $\sigma^2 = \frac{n\nu}{m_j} c^2$ . It can be easily verified that  $\frac{\sigma^2}{R_1} = \frac{1}{\sqrt{r}} c$  and  $\frac{\sigma^2}{R_2} < \frac{1}{\sqrt{r}} c$ . Therefore, for all  $t \leq \frac{1}{\sqrt{r}} c$ , the following estimate holds.

$$\Pr(\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G}\| \geq t) \leq n \exp\left(-\frac{3t^2 \kappa_j r}{8\nu c^2}\right) \quad (38)$$

with  $\kappa_j = \frac{m_j}{nr}$ . This concludes the proof of Lemma 4.  $\square$

In what follows, it will be argued that  $\mathbf{M}$  is a unique solution to (9) with “very high” probability. It suffices to show all  $\mathbf{X} \in \mathbb{S}_+ - \{\mathbf{M}\}$  are solutions with a very small probability by choosing  $\Omega$  “sufficiently large”. For  $\mathbf{X} \in \mathbb{S}_+$ , write the deviation  $\Delta = \mathbf{X} - \mathbf{M}$ . We define  $\mathbb{S}_1 = \{\mathbf{X} \in \mathbb{S}_+ : \|\Delta_{\mathbb{T}}\|_F^2 \geq \left(n\sqrt{8\frac{L}{m}}\|\Delta_{\mathbb{T}^\perp}\|_F\right)^2\}$  and  $\mathbb{S}_2 = \{\mathbf{X} \in \mathbb{S}_+ : \|\Delta_{\mathbb{T}}\|_F^2 < \left(n\sqrt{8\frac{L}{m}}\|\Delta_{\mathbb{T}^\perp}\|_F\right)^2 \& \mathcal{R}_\Omega(\Delta) = \mathbf{0}\}$  be two spaces defined based on deterministic comparisons of  $\|\Delta_{\mathbb{T}}\|_F^2$  and  $\|\Delta_{\mathbb{T}^\perp}\|_F^2$ . Assuming that  $\Omega$  is sampled uniformly at random with replacement, we consider the following two cases.

- 1) Choose  $|\Omega| = m$  “sufficiently large” such that  $\Pr(\{\Omega \subset \mathbb{I} \mid |\Omega| = m, \lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > 1/2\}) \geq 1 - p_1$  based on Lemma 1. Using Theorem 2(a), all  $\mathbf{X} \in \mathbb{S}_1$  are feasible solutions to (9) with probability  $p_1$ .
- 2) Recall that using the golfing scheme  $\Pr(\{\Omega \subset \mathbb{I} \mid |\Omega| = m, \mathbf{Y} \in \text{range } \mathcal{R}_\Omega^* \& \|\mathcal{P}_{\mathbb{T}} \mathbf{Y} - \text{Sgn } \mathbf{M}\|_F \leq \frac{1}{4L} \& \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}\| \leq \frac{1}{2}\}) \geq 1 - \epsilon$  with  $\epsilon = \sum_{i=1}^l [p_2(i) + p_3(i) + p_4(i)]$  by choosing  $|\Omega| = m$  “sufficiently large”. Using Theorem 2(b), all  $\mathbf{X} \in \mathbb{S}_2$  are not solutions to (9) with probability  $\epsilon$ .

Using the union bound,  $\mathbf{X} \neq \mathbf{M}$  is a solution to the EDG nuclear minimization problem with probability which is at most  $p = p_1 + \sum_{i=1}^l [p_2(i) + p_3(i) + p_4(i)]$ . In what follows, the goal is to ensure that  $p$  is “small” with  $m$  sufficiently large implying the uniqueness of  $\mathbf{M}$  to (9) with very high probability. This is attained by a suitable choice of the parameters  $m, l$  and  $m_i$  which are detailed below.

The first failure probability is the failure that condition in Theorem 2(a) does not hold and is given by

$$p_1 = n \exp\left(\frac{-\kappa}{8\nu}\right)$$

In the proof of Lemma 3, the failure probabilities  $p_2(i)$ ,  $p_3(i)$  and  $p_4(i)$  are given by

$$p_2(i) = \exp\left(\frac{-\kappa_i}{64\left(\nu + \frac{1}{nr}\right)} + \frac{1}{4}\right) \quad ; \quad p_3(i) = n \exp\left(\frac{-3\kappa_i}{128\nu}\right) \quad ; \quad p_4(i) = n^2 \exp\left(\frac{-3\kappa_i}{64\left(\nu + \frac{1}{nr}\right)}\right) \quad \forall i \in [1, l]$$

To prove Theorem 1, it remains to specify  $\kappa_i$  and show that the total probability of failure is very “small”. In precise terms, this means that the probability of failure is bounded above by  $n^{-\beta}$  for some  $\beta > 1$ .  $\kappa_i$  is chosen in such a way that all the failure probabilities,  $p_1, p_2(i), p_3(i)$  and  $p_4(i)$ , are at most  $\frac{1}{4l}n^{-\beta}$ . With this, one appropriate choice for  $\kappa_i$  is  $\kappa_i = 96(\nu + \frac{1}{nr})(\beta \log(n) +$

$\log(4l)$ ). Using the union bound, it can be verified that the total failure is bounded above by  $n^{-\beta}$ . The number of measurement  $m = \ln rk_i$  must be at least

$$\log_2(4n\sqrt{8Lr})nr\left(96\left(v + \frac{1}{nr}\right)[\beta\log(n) + \log(4\log_2(4L\sqrt{r}))]\right) \quad (39)$$

This finishes the proof of Theorem 1 and concludes that the minimization program in (9) recovers the true inner product matrix with very high probability.

### C. Noisy EDG Completion

In a practical settings, the available partial information is not exact but noisy. For simplicity, consider an additive Gaussian noise  $Z$  with mean  $\mu$  and variance  $\sigma$ . The modified nuclear norm minimization for the EDG problem can now be written as

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{S}_+}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \|\mathcal{R}_\Omega(\mathbf{X}) - \mathcal{R}_\Omega(\mathbf{M})\|_F \leq \delta \end{aligned} \quad (40)$$

where  $\delta$  characterizes the level of noise. Unlike the noisy matrix completion problem, the noise parameters for the EDG problem can not be chosen arbitrarily. The reason is that the perturbed distance matrix needs to be non-negative. In the context of numerically solving the noisy EDG problem, details of how to set these parameters will be discussed in the next section. Under the assumption that  $\mu$  and  $\sigma$  are chosen appropriately, we posit that the theoretical guarantees for the exact case extend to noisy EDG completion. Following the analysis in [39] and using the dual basis framework, it can be surmised that Theorem 1 holds true with failure probability proportional to noise level  $\delta$ . A sketch of such a theorem is stated below.

**Theorem 6.** *Let  $\mathbf{M} \in \mathbb{R}^{n \times n}$  matrix of rank  $r$  that obeys the coherence conditions (12), (13) and (14) with coherence  $v$ . Assume  $m$  measurements  $\langle \mathbf{M}, \mathbf{w}_\alpha \rangle$ , sampled uniformly at random with replacement are corrupted with Gaussian noise of mean  $\mu$ , variance  $\sigma$  and noise level  $\delta$ . For any  $\beta > 1$ , if*

$$m \geq \log_2(4\sqrt{8Lrn})nr\left(96\left[v + \frac{1}{nr}\right][\beta\log(n) + \log(4\log_2(4L\sqrt{r}))]\right)$$

then

$$\|\mathbf{M} - \bar{\mathbf{M}}\|_F \leq f\left(n, \frac{m}{n^2}, \delta\right)$$

where  $\bar{\mathbf{M}}$  is a solution to (40) with probability at least  $1 - n^{-\beta}$ .

The interpretation of the above theorem is that the EDG problem is stable to noise. Numerical experiments confirm this consideration. However, a precise analysis mandates characterization of the level of noise and an exact specification of  $f$  in the above theorem. This is left for future work.

## IV. NUMERICAL ALGORITHMS

The theoretical analysis so far shows that the nuclear minimization approach yields a unique solution for the EDG problem. In this section, we aim at developing a practical algorithm to recover the inner product matrix  $\mathbf{X}$  given partial pairwise distance information supported on some random set  $\Omega$ . Since the available partial information might be noisy in applications, we also extend the algorithm to this case.

### A. Exact partial information

An algorithm similar to ours appears in [7] where the authors design an algorithm employing the alternative directional minimization (ADM) method to recover the Gram matrix. A crucial part of their algorithm uses the hard thresholding operator which computes eigendecompositions at every iteration and is computationally intensive. Comparatively, an advantage of our algorithm is that it does not require an eigendecomposition making it fast and suitable for tests on large data. Since the nuclear norm of a positive semidefinite matrix equates to its trace, we consider the following minimization problem.

$$\min_{\tilde{\mathbf{X}} \in \mathbb{R}^{n \times n}} \text{Trace}(\tilde{\mathbf{X}}) \quad \text{subject to} \quad \mathcal{R}_\Omega(\tilde{\mathbf{X}}) = \mathcal{R}_\Omega(\mathbf{M}), \quad \tilde{\mathbf{X}} \cdot \mathbf{1} = \mathbf{0} \text{ and } \tilde{\mathbf{X}} \succeq \mathbf{0}$$

Consider a matrix  $\mathbf{C} \in \mathbb{R}^{n \times (n-1)}$  satisfying  $\mathbf{C}^T \mathbf{C} = \mathbf{I}$  and  $\mathbf{C}^T \mathbf{1} = \mathbf{0}$ , we rewrite the above minimization problem as follows by a change of variable  $\tilde{\mathbf{X}} = \mathbf{C} \mathbf{X} \mathbf{C}^T$ .

$$\min_{\mathbf{X} \in \mathbb{R}^{(n-1) \times (n-1)}} \text{Trace}(\mathbf{C} \mathbf{X} \mathbf{C}^T) \quad \text{subject to} \quad \mathcal{R}_\Omega(\mathbf{C} \mathbf{X} \mathbf{C}^T) = \mathcal{R}_\Omega(\mathbf{M}), \quad \mathbf{X} \succeq \mathbf{0}$$

Note that the sum to one constraint drops out since  $\mathbf{C}^T \mathbf{1} = \mathbf{0}$  and  $\mathbf{C} \mathbf{X} \mathbf{C}^T \succeq \mathbf{0}$  if  $\mathbf{X} \succeq \mathbf{0}$ . To enforce that  $\mathbf{X}$  is positive semidefinite, let  $\mathbf{X} = \bar{\mathbf{P}} \bar{\mathbf{P}}^T$  where  $\bar{\mathbf{P}} \in \mathbb{R}^{(n-1) \times q}$  with  $q$  unknown a priori. Since  $\bar{\mathbf{P}} \bar{\mathbf{P}}^T$  has at most rank  $q$ , it entails a good

estimate for  $q$  which ideally should be a reasonable estimate of the rank. Due to the trace regularization, our algorithm only needs a rough guess of  $q$ . The above minimization problem now reduces to

$$\min_{\bar{\mathbf{P}} \in \mathbb{R}^{(n-1) \times q}} \text{Trace}(\mathbf{C}\bar{\mathbf{P}}\bar{\mathbf{P}}^T\mathbf{C}^T) \quad \text{subject to} \quad \mathcal{R}_\Omega(\mathbf{C}\bar{\mathbf{P}}\bar{\mathbf{P}}^T\mathbf{C}^T) = \mathcal{R}_\Omega(\mathbf{M}) \quad (41)$$

With  $\mathbf{P} = \mathbf{C}\bar{\mathbf{P}}$ , consider the simplified minimization problem.

$$\min_{\mathbf{P} \in \mathbb{R}^{n \times q}} \text{Trace}(\mathbf{P}\mathbf{P}^T) \quad \text{subject to} \quad \mathcal{R}_\Omega(\mathbf{P}\mathbf{P}^T) = \mathcal{R}_\Omega(\mathbf{M}) \quad (42)$$

The above technique, the change of variables employing  $\mathbf{C}$ , has been previously used [21], [27]. In [27], the authors remark that the reparameterization leads to numerically stable interior point algorithms. Note that  $\mathbf{C}$  is simply an orthonormal basis for the space  $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}'\mathbf{1} = 0\}$ . For the EDG problem, the goal is to find the Gram matrix  $\bar{\mathbf{X}} = \mathbf{C}\mathbf{X}\mathbf{C}^T = \mathbf{C}\mathbf{C}^T\mathbf{P}\mathbf{P}^T\mathbf{C}\mathbf{C}^T$ .  $\mathbf{C}\mathbf{C}^T$  is simply the orthogonal projection onto the aforementioned space given by  $\mathbf{C}\mathbf{C}^T = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ . Given  $\bar{\mathbf{X}}$ , classical MDS can then be used to find the point coordinates. Therefore, our focus is on solving for  $\mathbf{P}$  in (42). We employ the method of augmented Lagrangian to solve the minimization problem. The constraint  $\mathcal{R}_\Omega(\mathbf{P}\mathbf{P}^T) = \mathcal{R}_\Omega(\mathbf{M})$  can be written compactly using the linear operator  $\mathcal{A}$  defined as  $\mathcal{A} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{|\Omega|}$  with  $\mathcal{A}(\mathbf{X}) = \mathbf{f} \in \mathbb{R}^{|\Omega| \times 1}$ ,  $\mathbf{f}_i = \langle \mathbf{X}, \mathbf{w}_{\alpha_i} \rangle$  for  $\alpha_i \in \Omega$ . For latter use, the adjoint of  $\mathcal{A}$ ,  $\mathcal{A}^*$ , can be derived as follows. For  $\mathbf{y} \in \mathbb{R}^{|\Omega| \times 1}$ ,  $\langle \mathcal{A}\mathbf{X}, \mathbf{y} \rangle = \sum_i \langle \mathbf{X}, \mathbf{w}_{\alpha_i} \rangle \mathbf{y}_i = \langle \mathbf{X}, \sum_i \mathbf{y}_i \mathbf{w}_{\alpha_i} \rangle$ . It follows that  $\mathcal{A}^*\mathbf{y} = \sum_i \mathbf{y}_i \mathbf{w}_{\alpha_i}$ . Thus, we write (42) as

$$\min_{\mathbf{P}} \text{Trace}(\mathbf{P}\mathbf{P}^T) \quad \text{subject to} \quad \mathcal{A}(\mathbf{P}\mathbf{P}^T) = \mathbf{b} \quad (43)$$

The augmented Lagrangian is given by

$$\mathbf{L}(\mathbf{P}; \Lambda) = \text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{r}{2} \|\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b} + \Lambda\|_2^2$$

where  $\Lambda \in \mathbb{R}^{|\Omega| \times 1}$  denotes the Lagrangian multiplier and  $r$  is the penalty term. The augmented Lagrangian step is simply  $\mathbf{P}^k = \text{argmin } \mathbf{L}(\mathbf{P}; \Lambda^{k-1})$  followed by updating of the multiplier  $\Lambda^k$ . To solve the first problem with respect to  $\mathbf{P}$ , the Barzilai-Borwein steepest descent method [40] is employed with objective function  $\text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{r}{2} \|\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b} + \Lambda^{k-1}\|_2^2$  and gradient  $2\mathbf{P} + 2r\mathcal{A}^*(\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b} + \Lambda^{k-1})\mathbf{P}$ . The iterative scheme is summarized in Algorithm 1.

---

#### Algorithm 1 Augmented Lagrangian based scheme to solve (43)

---

- 1: **Initialization:** Set  $\Lambda^0 = \mathbf{0}$ ,  $q = 10$ ,  $\mathbf{P}^0 = \text{rand}(n, q)$ ,  $E^0 = E_{\text{Total}}^0 = 0$ . Set maxiterations, bbiterations,  $r$ , Tol
  - 2: **for**  $k = 1$ :maxiterations **do**
  - 3:   Barzilai-Borwein(BB) descent for  $\mathbf{P}^k = \text{argmin } \mathbf{L}(\mathbf{P}; \Lambda^{k-1})$ .
  - 4:    $\Lambda^k = \Lambda^{k-1} + \mathcal{A}(\mathbf{P}^k(\mathbf{P}^k)^T) - \mathbf{b}$
  - 5:    $E^k = \frac{r}{2} \|\mathcal{A}(\mathbf{P}^k(\mathbf{P}^k)^T) - \mathbf{b}\|_2^2$
  - 6:    $E_{\text{Total}}^k = \text{Trace}(\mathbf{P}^k(\mathbf{P}^k)^T) + \frac{r}{2} \|\mathcal{A}(\mathbf{P}^k(\mathbf{P}^k)^T) - \mathbf{b} + \Lambda^k\|_2^2$
  - 7:   **if**  $E^k < \text{Tol}$  &  $E_{\text{Total}}^k < \text{Tol}$  **then**
  - 8:     **break**
- 

### B. Partial information with Gaussian noise

Assume that the available partial information is noisy. Formally,  $\mathcal{R}_\Omega(\mathbf{X}) = \mathcal{R}_\Omega(\mathbf{M}) + \mathcal{R}_\Omega(\mathbf{Z})$  where  $\mathcal{R}_\Omega(\mathbf{Z})$  is an additive Gaussian noise. Proceeding analogously to the case of exact partial information, the following minimization problem is obtained.

$$\min_{\mathbf{P}} \text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{\lambda}{2} \|\mathcal{R}_\Omega(\mathbf{P}\mathbf{P}^T) - \mathcal{R}_\Omega(\mathbf{M})\|_F^2 \quad (44)$$

Using the operator  $\mathcal{A}$  introduced earlier, (44) can be rewritten as

$$\min_{\mathbf{P}} \text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{\lambda}{2} \|\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b}\|_F^2 \quad (45)$$

where  $\mathbf{b} = \mathcal{A}(\mathbf{M})$ . The augmented Lagrangian is given by

$$\mathbf{L}(\mathbf{P}; \Lambda) = \text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{\lambda}{2} \|\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b}\|_2^2$$

where  $\Lambda \in \mathbb{R}^{n \times 1}$  denotes the Lagrangian multiplier and  $r$  is a penalty term. The augmented Lagrangian step is simply  $\mathbf{P}^k = \text{argmin } \mathbf{L}(\mathbf{P}; \Lambda^{k-1})$ . As before,  $\mathbf{P}$  is computed using the Barzilai-Borwein steepest descent method with objective function  $\text{Trace}(\mathbf{P}\mathbf{P}^T) + \frac{\lambda}{2} \|\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b}\|_2^2$  and gradient  $2\mathbf{P} + 2\lambda\mathcal{A}^*(\mathcal{A}(\mathbf{P}\mathbf{P}^T) - \mathbf{b})\mathbf{P}$ . The iterative scheme is summarized in Algorithm 2.

---

**Algorithm 2** Augmented Lagrangian based scheme to solve (45)

---

```

1: Initialization: Set  $q = 10$ ,  $\mathbf{P}^0 = \text{rand}(n,q)$ ,  $E_{\text{Total}}^0 = 0$ . Set maxiterations, bbiterations,  $r$ ,  $\lambda$ , Tol
2: for  $k = 1$ :maxiterations do
3:   Barzilai-Borwein(BB) descent for  $\mathbf{P}_k$ .
4:    $E_{\text{Total}}^k = \text{Trace}(\mathbf{P}^k(\mathbf{P}^k)^T) + \frac{\lambda}{2}\|\mathcal{A}(\mathbf{P}^k(\mathbf{P}^k)^T) - \mathbf{b}\|_2^2$ 
5:   if  $E_{\text{Total}}^k < \text{Tol}$  then
6:     break

```

---

## V. NUMERICAL RESULTS

In this section, we demonstrate the efficiency and accuracy of the proposed algorithms. All the numerical experiments are ran in MATLAB on a laptop with an Intel Core I7 2.7 GHz processor and a 16GB RAM.

### A. Experiments on synthetic and real data

We first test the Euclidean distance geometry problem on different three-dimensional objects. These objects include a sphere, a cow, and a map of a subset of US cities. Given  $n$  points from these objects, the full  $n \times n$  distance matrix  $\mathbf{D}$  has  $\frac{n(n-1)}{2} = L$  degrees of freedom. The objective is to recover the full point coordinates given  $\gamma L$  entries of  $\mathbf{D}$ ,  $\gamma \in [0, 1]$ , chosen uniformly at random. Algorithm 1 and Algorithm 2 output  $\mathbf{P}$  from which  $\mathbf{X} = \mathbf{P}\mathbf{P}^T$  is constructed. Classical MDS is then used to find the global coordinates. In all of the numerical experiments, a rank estimation  $q = 10$  is used. This choice shows that the algorithms recover the ground truth despite a rough estimate of the true rank. The stopping criterion is a tolerance on the relative total energy defined as  $(E_{\text{Total}}^k - E_{\text{Total}}^{k-1})/E_{\text{Total}}^k$ . For algorithm 1, an additional stopping criterion is a tolerance on  $E^k = \frac{r}{2}\|\mathcal{A}(\bar{\mathbf{P}}_k) - \mathbf{b}\|_2^2$ . In all of the numerical experiments, the tolerance is set to  $10^{-5}$ . The maximum iteration is set to 100.

Accuracy of the reconstruction is measured using the relative error of the inner product matrix  $\frac{\|\mathbf{X} - \mathbf{M}\|_F}{\|\mathbf{M}\|_F}$  where  $\mathbf{X}$  is the numerically computed inner product matrix and  $\mathbf{M}$  is the ground truth.

For different sampling rates, Figure 1 displays the reconstructed three-dimensional objects assuming that exact partial information is provided. For all experiments, the penalty term  $r$  is set to 1.0. Table II shows the relative error of the inner product matrix for all the different cases. The algorithm provides very good reconstruction except the 1% sphere. For this specific case, the distance matrix for the sphere is comparatively small. This means that 1% provides very little information and more samples are needed for reasonable reconstruction.

TABLE II

RELATIVE ERROR OF THE INNER PRODUCT MATRIX FOR DIFFERENT THREE-DIMENSIONAL OBJECTS UNDER DIFFERENT SAMPLING RATES. THE PARTIAL INFORMATION IS EXACT. RESULTS ARE AVERAGES OVER 50 RUNS.

$\gamma$	1%	2%	3%	5%
Sphere	$3.27e - 01$	$2.10e - 03$	$7.98e - 05$	$3.21e - 05$
Cow	$1.53e - 04$	$7.08e - 05$	$5.03e - 05$	$3.65e - 05$
US Cities	$1.94e - 04$	$1.13e - 04$	$7.53e - 05$	$5.82e - 05$

For the case of partial information with the additive Gaussian noise  $N(\mu, \sigma)$ , the noisy distance matrix can be written as  $\bar{\mathbf{D}} = \mathbf{D} + N(\mu, \sigma)$ . The standard deviation  $\sigma$  is a critical parameter determining the extent to which the underlying information is noisy. In the EDG problem, the perturbed distance matrix  $\bar{\mathbf{D}}$  must have non-negative entries. Thus  $\mu$  and  $\sigma$  need to be chosen carefully to satisfy this condition. In our numerical experiments,  $\sigma$  is set to be the minimum non-zero value of the partial distance matrix and  $\mu$  is  $3\sigma$ . It can be easily verified that, with very high probability, these choices ensure a non-negative noisy distance matrix. The choice of  $\sigma$  corresponds to the case where the error of the measurement is in the order of the minimum distance. While this choice might not reflect practical measurements, the setting allows us to test the extent to which the algorithm handles a noisy data. The parameter  $\lambda$  is a penalty term which needs to be chosen carefully. It can be surmised that  $\lambda$  needs to increase with increasing sampling rate. For our numerical experiments, a simple heuristic is to set  $\lambda = 100\gamma$ . While a more sophisticated analysis might result an optimal choice of  $\lambda$ , the simple choice is found to be sufficient and effective in the numerical tests. For different sampling rates, Figure 2 displays the reconstructed three-dimensional objects assuming that a noisy partial information is provided. Table III shows the relative error of the inner product matrix for all the different cases. The algorithm results good reconstruction except the 1% sphere. As discussed earlier, this specific case requires relatively more samples for reasonable reconstruction.

We further apply the proposed algorithm to the molecular conformation problem [1]. In this problem, the aim is to determine the three-dimensional structure of a molecule given partial information on pairwise distances between the atoms. An instance of

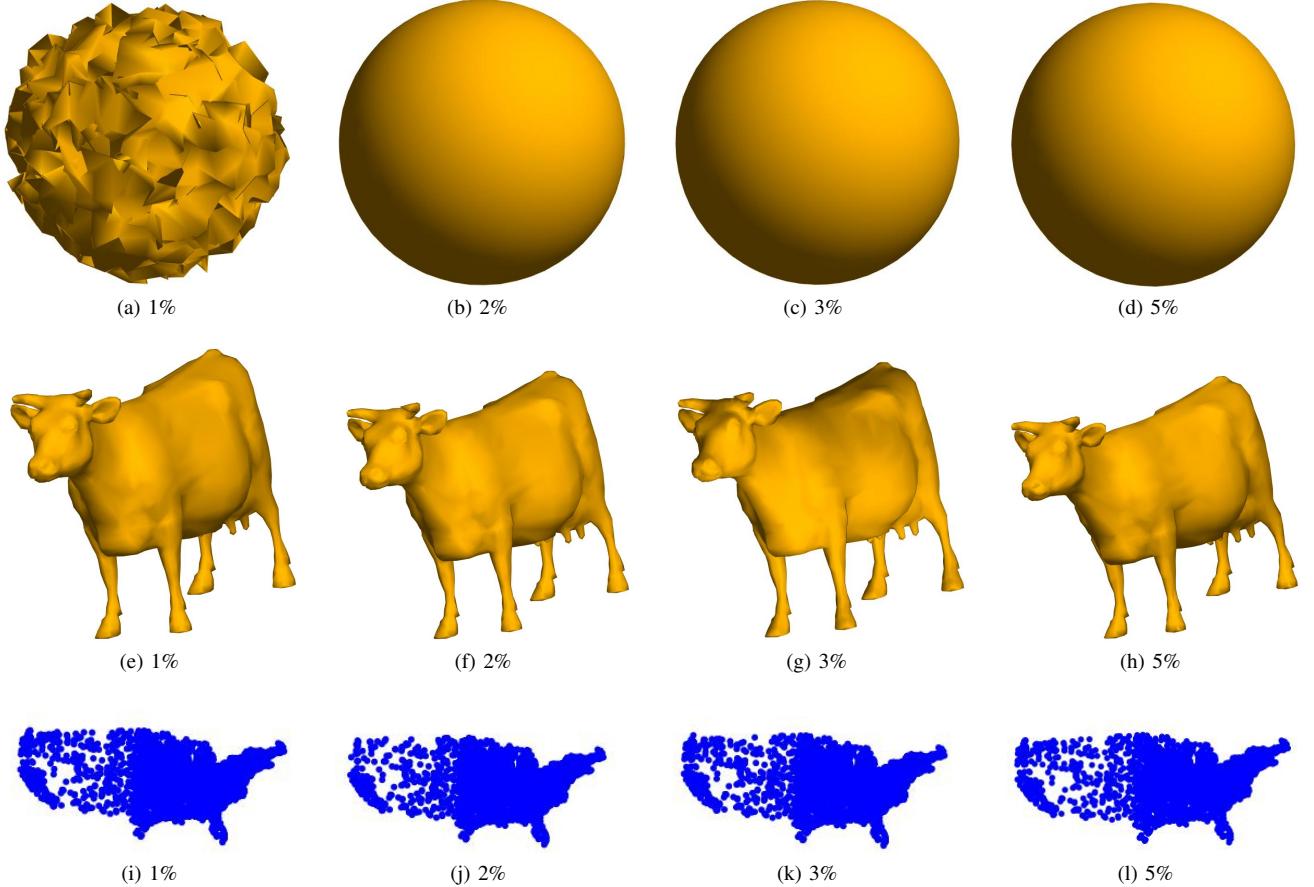


Fig. 1. Reconstruction results of different three-dimensional objects from exact partial information with sampling rates 1% (1st column), 2% (2nd column), 3% (3rd column) and 5% (4th column), respectively.

TABLE III

RELATIVE ERROR OF THE INNER PRODUCT MATRIX FOR DIFFERENT THREE-DIMENSIONAL OBJECTS UNDER DIFFERENT SAMPLING RATES. THE RELATIVE ERROR IS AN AVERAGE OF 50 RUNS AND THE PARTIAL INFORMATION IS NOISY.

	1%	2%	3%	5%
Sphere	$5.21e - 01$	$5.55e - 02$	$1.86e - 02$	$8.80e - 03$
Cow	$2.84e - 02$	$6.60e - 03$	$3.80e - 03$	$2.00e - 03$
US Cities	$2.80e - 02$	$9.00e - 03$	$5.50e - 03$	$3.10e - 03$

this problem is determining the structure of proteins using nuclear magnetic resonance (NMR) spectroscopy or X-ray diffraction experiments. The problem is challenging since the partial distance matrix obtained from experiments is sparse, non-uniform, noisy and prone to outliers [1], [41]. We test our method on a simple version of the problem to illustrate that the algorithm can also work on real data. Our numerical experiment considers two protein molecules identified as 1AX8 and 1RGS obtained from the Protein Data Bank [42]. We use a pre-processed version of the data taken from [41]. Given the full distance matrix, the partial data is a random uniform sample with 3% sampling rate. The setup of the numerical experiments is the same as before. Figure 3 displays the reconstructed three-dimensional structure of 1AX8 and 1RGS under exact partial information and noisy partial information. The results demonstrate that the algorithm provides good reconstruction of the underlying three-dimensional structures.

#### B. Computational comparisons of Algorithms

To evaluate the efficiency of the proposed algorithms, numerical tests are carried out using Algorithm 1 and Algorithm 2. For the test, the input is a random uniform sample of the exact/noisy distance matrix of one of the three-dimensional objects discussed earlier(a sphere, a cow and a map of a subset of US cities). The sampling rate is set to 5% and the algorithms are run until a stopping criterion discussed before is met. In what follows, the reported results are averages of 50 runs and the

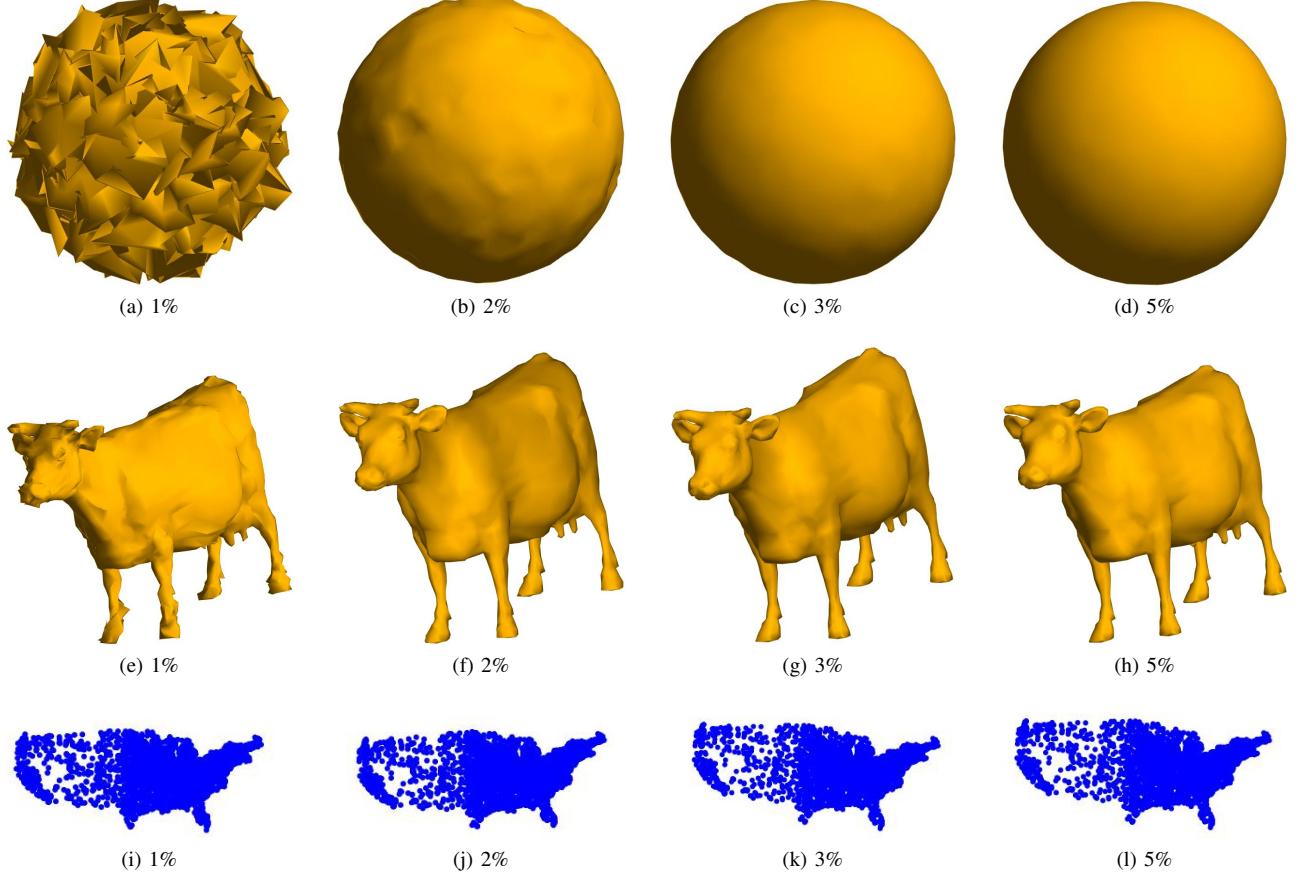


Fig. 2. Reconstruction results of different three-dimensional objects from noisy partial information with different sampling rates 1% (1st column), 2% (2nd column), 3% (3rd column) and 5% (4th column), respectively.

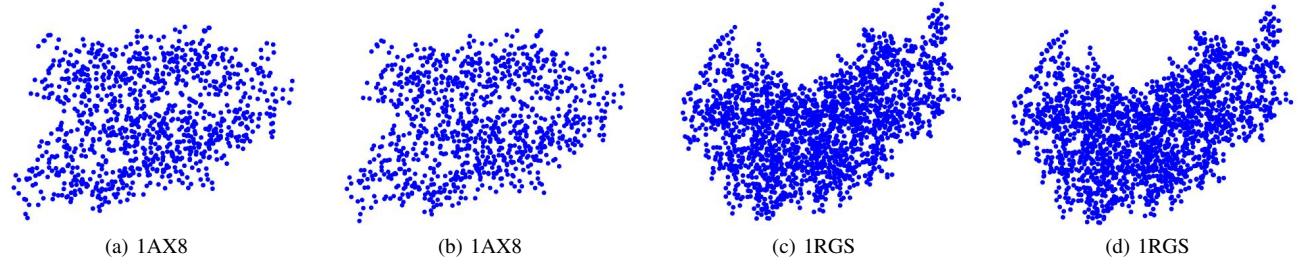


Fig. 3. Reconstructed three-dimensional structure of proteins identified as 1AX8 and 1RGS in the Protein Data Bank. Exact partial information (3% sample rate) for the first and third, noisy partial information for the second and fourth.

number of iterations is the ceiling of the average number of iterations. Tables IV summarizes the results of the computational experiments for the three-dimensional objects. It can be concluded that the algorithm is fast and converges in few iterations to the desired solution.

We also conduct comparisons with the algorithm for the EDG problem proposed in [7] that also employs the augmented Lagrangian method. As remarked earlier, the main difference between the two algorithms is on the way the positive semidefinite condition is imposed on the inner product matrix  $\mathbf{X}$ . In [7], the positive semidefinite condition is imposed on  $\mathbf{X}$  directly while algorithm 1 uses the factorization  $\mathbf{P}\mathbf{P}^T$  to enforce this condition. To compare the two algorithms, we use the same input of data which is a random uniform sample of the distance matrix of one of the three-dimensional objects. The sampling rate is set to 5%. For both algorithms, the stopping criterion is the relative error of the total energy and the tolerance is set  $10^{-5}$ . Table V summarizes the comparison of these two algorithms. The reported results are averages of 10 runs. We see that our algorithm converges to the desired solution faster and with significantly less number of iterations.

Finally, we consider the molecular conformation problem and compare our algorithm to the DISCO algorithm proposed in [41]. The DISCO algorithm is an SDP based divide and conquer algorithm. In [41], the authors demonstrate that the algorithm

is effective on sparse and highly noisy molecular conformation problems. For the numerical tests, the input for both algorithms is a protein molecule. The sampling rate is set to 5% and it is assumed that the underlying partial information is exact. We downloaded the DISCO code, a MATLAB mex code version 1.4, from <http://www.math.nus.edu.sg/~mattohkc/disco.html> which provides an input file and executables. For our algorithm, the stopping criterion is maximum number of iterations set to 200. We emphasize that the rank estimate using our method for all experiments is 5. This means our method has 5/3 times variables to the method used in DISCO which compute molecular coordinates directly (i.e. rank number is 3). The DISCO algorithm has a radius parameter which implies that the input distance matrix consists pairwise distances less than or equal to the radius. For consistent comparison with the EDG problem and our algorithm, the radius is set large. In lack of a source code for DISCO, under the above setups, we run the algorithm as it is. Table VI summarizes the comparison of these two algorithms. The reported results are averages of 20 runs. We see that our algorithm attains a relative error of the same order as DISCO but is faster on all the tests. Some caveats about the comparison are the assumption on the radius and a very sparse partial exact information. In [41], the radius is set to 6 Å since NMR measurements have a limited range of validity estimated to be 6 Å. With this choice, the problem departs from the EDG problem since there is localization. For this localization problem with a noisy input data and relatively sparse input (20% of distance within the radius), we note that DISCO results in excellent reconstruction of the protein molecules. The above comparison is meant to illustrate that, for a simplistic setup, our algorithm is very fast and has the potential to handle tests on large protein molecules.

TABLE IV

COMPUTATIONAL SUMMARY OF ALGORITHM 1 FROM 50 RUNS. THE DATA ARE THE DIFFERENT THREE-DIMENSIONAL OBJECTS. THE SAMPLING RATE IS 5%.

	3D Object	Number of Points	Computational Time(Sec)	Number of iterations
Exact	Sphere	1002	3.27	9
	Cow	2601	68.08	26
	US Cities	2920	101.92	34
Noisy	Sphere	1002	10.56	19
	Cow	2601	62.63	17
	US Cities	2920	62.33	18

TABLE V

COMPARISON OF THE PROPOSED ALGORITHM 1 AND THE ALGORITHM IN [7] WITH 10 RUNS. THE DATA ARE THE DIFFERENT THREE-DIMENSIONAL OBJECTS. IT IS ASSUMED THAT THE PARTIAL INFORMATION IS EXACT. THE SAMPLING RATE IS 5%.

3D Object	Number of Points	Computational Time(Sec)		Number of iterations	
		Alg. 1	Alg. [7]	Alg. 1	Alg. [7]
Sphere	1002	3.06	101.85	9	204
Cow	2601	51.40	1387.80	20	257
US Cities	2920	72.51	2417.30	23	250

TABLE VI

COMPARISON OF THE PROPOSED ALGORITHM 1 AND THE ALGORITHM IN [41] WITH 20 RUNS. THE DATA ARE THE DIFFERENT PROTEIN MOLECULES. IT IS ASSUMED THAT THE PARTIAL INFORMATION IS EXACT. THE SAMPLING RATE IS 5%.

Protein Molecule	Number of Points	Computational Time(Sec)		Relative error of the inner product matrix	
		Alg. 1	Alg. [41]	Alg. 1	Alg. [41]
1PTQ	402	3.38	11.97	$8.03e - 07$	$1.63e - 07$
1AX8	1003	9.44	45.94	$1.02e - 08$	$8.01e - 07$
1RGS	2015	52.62	214.30	$5.41e - 09$	$1.62e - 07$
1KDH	2923	104.60	438.04	$2.86e - 09$	$8.16e - 09$
1BPM	3672	223.14	392.29	$3.46e - 09$	$3.27e - 08$

### C. Phase transition of Algorithm 1

Last but not least, we numerically investigate the optimality of the proposed algorithm by plotting the phase transition. Given the number of points and the underlying rank, the theory provides the sampling rate which leads to successful recovery with very high probability (see Theorem 1). To investigate the optimality of Algorithm 1, the following numerical experiment was

carried out. Consider sampling rates ranging from 1% to 100% and rank ranging from 1 to 40. For each pair, Algorithm 1 is run 50 times. Successful recovery refers to the case where the relative error of the inner product matrix is within tolerance. As remarked earlier, the tolerance is set to  $10^{-5}$ . Out of the 50 runs, the number of times the algorithm succeeds provides us with a probability of success. This procedure is repeated for all combination of sampling rate and rank. Figure 4 shows the optimality result of Algorithm 1. Namely, for a large portion in the sampling rate-rank domain, the proposed algorithm can provide successful reconstruction.

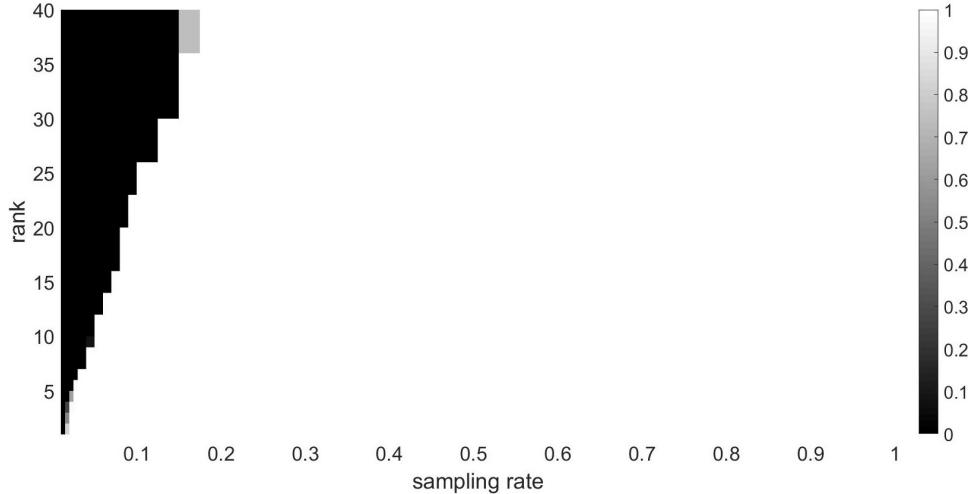


Fig. 4. Success probability of Algorithm 1 given sampling rate and rank. For each sampling rate, Algorithm 1 is run for ranks ranging from 1 to 40. The average success probability, of 50 runs, is shown above. White indicates perfect recovery and black indicates failure of the algorithm.

## VI. CONCLUSION

In this paper, we formulate the Euclidean distance geometry (EDG) problem as a low rank matrix recovery problem. Adopting the matrix completion framework, our approach can be viewed as completing the Gram matrix with respect to a suitable basis given few uniformly random distance samples. However, the existing analysis based on the restricted isometry property (RIP) does not hold for our problem. Alternatively, we conduct analysis by introducing the dual basis approach to formulate the EDG problem. Our main result shows that the underlying configuration of points can be recovered with very high probability from  $O(nrv \log^2(n))$  measurements if the underlying Gram matrix obeys the coherence condition with parameter  $v$ . Numerical algorithms are designed to solve the EDG problem under two scenarios, exact and noisy partial information. Numerical experiments on various test data demonstrate that the algorithms are simple, fast and accurate. The technique in this paper is not specifically limited to the EDG problem. The extension of this result to the low rank recovery of a matrix given few measurements with respect to any non-orthogonal basis is a work in preparation.

## VII. ACKNOWLEDGMENT

The authors would like to thank Dr. Jia Li for his discussions in the early stage of this project. Abiy Tasissa would like to thank Professor David Gross for correspondence over email regarding the work in [29]. Particularly, the proof of Lemma A.7 is a personal communication from Professor David Gross. The authors also would like to thank Professor Peter Kramer and Professor Alex Gittens for their comments and suggestions. The authors' gratitude is further extended to the anonymous reviewers for their valuable feedback which has improved the manuscript.

## REFERENCES

- [1] W. Glunt, T. Hayden, and M. Raydan, "Molecular conformations from distance matrices," *Journal of Computational Chemistry*, vol. 14, no. 1, pp. 114–120, 1993.
- [2] M. W. Trosset, *Applications of multidimensional scaling to molecular conformation*. Citeseer, 1997.
- [3] X. Fang and K.-C. Toh, "Using a distributed sdp approach to solve simulated protein molecular conformation problems," in *Distance Geometry*. Springer, 2013, pp. 351–376.
- [4] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [5] Y. Ding, N. Krislock, J. Qian, and H. Wolkowicz, "Sensor network localization, euclidean distance matrix completions, and graph realization," *Optimization and Engineering*, vol. 11, no. 1, pp. 45–66, 2010.
- [6] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Transactions on Sensor Networks (TOSN)*, vol. 2, no. 2, pp. 188–220, 2006.
- [7] R. Lai and J. Li, "Solving partial differential equations on manifolds from incomplete interpoint distance," *SIAM Journal on Scientific Computing*, vol. 39, no. 5, pp. A2231–A2256, 2017.

- [8] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.
- [9] J. C. Gower, "Properties of euclidean and non-euclidean distance matrices," *Linear Algebra and its Applications*, vol. 67, pp. 81–97, 1985.
- [10] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [11] N. Moreira, L. Duarte, C. Lavor, and C. Torezzan, "A novel low-rank matrix completion approach to estimate missing entries in euclidean distance matrices," *arXiv preprint arXiv:1711.06182*, 2017.
- [12] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.
- [13] A. Javanmard and A. Montanari, "Localization from incomplete noisy distance measurements," *Foundations of Computational Mathematics*, vol. 13, no. 3, pp. 297–345, 2013.
- [14] I. J. Schoenberg, "Remarks to maurice frechet's article“sur la definition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace de hilbert,” *Annals of Mathematics*, pp. 724–732, 1935.
- [15] G. Young and A. S. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, no. 1, pp. 19–22, 1938.
- [16] B. Hendrickson, "Conditions for unique graph realizations," *SIAM journal on computing*, vol. 21, no. 1, pp. 65–84, 1992.
- [17] M. Bakonyi and C. R. Johnson, "The euclidian distance matrix completion problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 16, no. 2, pp. 646–654, 1995.
- [18] C. R. Johnson and P. Tarazaga, "Connections between the real positive semidefinite and distance matrix completion problems," *Linear Algebra and its Applications*, vol. 223, pp. 375–391, 1995.
- [19] M. Laurent, "A connection between positive semidefinite and euclidean distance matrix completion problems," *Linear Algebra and its applications*, vol. 273, no. 1-3, pp. 9–22, 1998.
- [20] A. Y. Alfakih, "On the uniqueness of euclidean distance matrix completions," *Linear Algebra and its Applications*, vol. 370, pp. 1–14, 2003.
- [21] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving euclidean distance matrix completion problems via semidefinite programming," *Computational optimization and applications*, vol. 12, no. 1-3, pp. 13–30, 1999.
- [22] H.-r. Fang and D. P. O'Leary, "Euclidean distance matrix completion problems," *Optimization Methods and Software*, vol. 27, no. 4-5, pp. 695–717, 2012.
- [23] B. Hendrickson, "The molecule problem: Exploiting structure in global optimization," *SIAM Journal on Optimization*, vol. 5, no. 4, pp. 835–857, 1995.
- [24] J. J. Moré and Z. Wu, "Global continuation for distance geometry problems," *SIAM Journal on Optimization*, vol. 7, no. 3, pp. 814–836, 1997.
- [25] M. W. Trosset, "Distance matrix completion by numerical optimization," *Computational Optimization and Applications*, vol. 17, no. 1, pp. 11–22, 2000.
- [26] Z. Zou, R. H. Bird, and R. B. Schnabel, "A stochastic/perturbation global optimization algorithm for distance geometry problems," *Journal of Global Optimization*, vol. 11, no. 1, pp. 91–105, 1997.
- [27] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, 2015.
- [28] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *Siam Review*, vol. 56, no. 1, pp. 3–69, 2014.
- [29] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [30] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.
- [31] B. Recht, "A simpler approach to matrix completion," *The Journal of Machine Learning Research*, vol. 12, pp. 3413–3430, 2011.
- [32] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [33] A. Carmi, L. Mihaylova, and S. Godsill, *Compressed Sensing & Sparse Filtering*, ser. Signals and Communication Technology. Springer Berlin Heidelberg, 2013. [Online]. Available: <https://books.google.com/books?id=EAO4BAAQBAJ>
- [34] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, 1963.
- [35] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 569–579, 2002.
- [36] D. Gross and V. Nesme, "Note on sampling without replacing from a finite collection of matrices," *arXiv preprint arXiv:1001.2738*, 2010.
- [37] R. Bhatia, *Matrix Analysis*, ser. Graduate Texts in Mathematics. Springer New York, 2013. [Online]. Available: <https://books.google.com/books?id=lh4BCAAQBAJ>
- [38] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of computational mathematics*, vol. 12, no. 4, pp. 389–434, 2012.
- [39] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [40] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.
- [41] N.-H. Z. Leung and K.-C. Toh, "An sdp-based divide-and-conquer algorithm for large-scale noisy anchor-free graph realization," *SIAM Journal on Scientific Computing*, vol. 31, no. 6, pp. 4351–4372, 2009.
- [42] 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 research*, vol. 28, no. 1, pp. 235–242, 2000.
- [43] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1990. [Online]. Available: <https://books.google.com/books?id=PIYQN0ypTwEC>

## APPENDIX

**Lemma A.1.** If  $\mathbf{X} \in \mathbb{S}$ ,  $\text{Sgn } \mathbf{X} \in \mathbb{S}$ . If  $\mathbf{X} \in \mathbb{T}$ ,  $\text{Sgn } \mathbf{X} \in \mathbb{T}$

*Proof.* Using the eigenvalue decomposition of  $\mathbf{X}$ ,  $\mathbf{X} = \mathbf{U}\Sigma\mathbf{U}^T$ .  $\text{Sgn } \mathbf{X}$  is simply  $\text{Sgn } \mathbf{X} = \mathbf{U}(\text{Sgn } \Sigma)\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{U}^T$  where  $\mathbf{D}$  is the diagonal matrix resulting from applying the sign function to  $\Sigma$ . To show that  $\text{Sgn } \mathbf{X} \in \mathbb{S}$ , we need to verify  $\text{Sgn } \mathbf{X} = (\text{Sgn } \mathbf{X})^T$  and  $\text{Sgn } \mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ . Symmetry of  $\text{Sgn } \mathbf{X}$  is apparent from its definition,  $(\text{Sgn } \mathbf{X})^T = \mathbf{U}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{U}^T = \text{Sgn } \mathbf{X}$ . To show that  $\text{Sgn } \mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ , consider  $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$  using the spectral decomposition of  $\mathbf{X}$ .

$$\left( \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^T \right) \cdot \mathbf{1} = \mathbf{0} \longrightarrow \mathbf{u}_j^T \left( \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^T \right) \cdot \mathbf{1} = \mathbf{0} \longrightarrow (\lambda_j \mathbf{u}_j^T) \cdot \mathbf{1} = \mathbf{0}$$

The implication is that  $\lambda_j = 0$  or  $\mathbf{u}_j^T \cdot \mathbf{1} = 0$ . With this, consider the spectral decomposition of the symmetric matrix  $\text{Sgn } \mathbf{X}$ .

$$\text{Sgn } \mathbf{X} = \sum_j \text{sgn}(\lambda_j) \mathbf{u}_j \mathbf{u}_j^T$$

$\text{Sgn } \mathbf{X} \cdot \mathbf{1} = \mathbf{0}$  follows in the following way. If  $\lambda_j = 0$ ,  $\text{sgn}(\lambda_j) = 0$ . Otherwise, from above,  $\mathbf{u}_j^T \cdot \mathbf{1} = \mathbf{0}$ . It can now be concluded that  $\text{Sgn } \mathbf{X} \in \mathbb{S}$ . Next, we show that for  $\mathbf{X} \in \mathbb{T}$ ,  $\text{Sgn } \mathbf{X} \in \mathbb{T}$ . Using the eigenvalue decomposition of  $\text{Sgn } \mathbf{X}$ , consider  $\mathcal{P}_{\mathbb{T}^\perp} \text{Sgn } \mathbf{X}$ .

$$\begin{aligned}\mathcal{P}_{\mathbb{T}^\perp} \text{Sgn } \mathbf{X} &= \text{Sgn } \mathbf{X} - \mathcal{P}_{\mathbb{T}} \text{Sgn } \mathbf{X} \\ &= \mathbf{U} \mathbf{D} \mathbf{U}^T - [\mathcal{P}_{\mathbf{U}} \text{Sgn } \mathbf{X} + \text{Sgn } \mathbf{X} \mathcal{P}_{\mathbf{U}} - \mathcal{P}_{\mathbf{U}} \text{Sgn } \mathbf{X} \mathcal{P}_{\mathbf{U}}] \\ &= \mathbf{U} \mathbf{D} \mathbf{U}^T - [\mathbf{U} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{U}^T + \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{U} \mathbf{U}^T - \mathbf{U} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{U} \mathbf{U}^T] = \mathbf{0}\end{aligned}$$

where the last step simply follows from the fact that  $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ . This confirms that  $\text{sgn } \mathbf{X} \in \mathbb{T}$  and concludes the proof.  $\square$

In the next two lemmas, we state and prove some facts about  $\mathbf{H}$  and  $\mathbf{H}^{-1}$ . We start with the latter first deriving an explicit form of  $\mathbf{H}^{-1}$ . This will be the focus of Lemma A.3 to follow shortly. In the proof of Lemma A.3, a certain form of the basis  $\mathbf{v}_\alpha$  is used. The form is conjectured by inspection of the basis  $\mathbf{v}_\alpha$  for the case when  $n$  is small. We start by stating and proving this form.

**Lemma A.2.** *Given an index  $(i, j)$  with  $1 \leq i < j \leq n$ , the matrix  $\mathbf{v}_{i,j}$  has the following form.*

$$\mathbf{v}_{i,j} = \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}$$

The matrices  $\tilde{\mathbf{w}}_{i,j}$ ,  $\mathbf{p}_{i,j}$  and  $\mathbf{q}_{i,j}$  are respectively defined as follows.

$$\tilde{\mathbf{w}}_{i,j} = \frac{n-1}{n^2} \mathbf{e}_{i,i} + \frac{n-1}{n^2} \mathbf{e}_{j,j} - \frac{(n-1)^2+1}{2n^2} \mathbf{e}_{i,j} - \frac{(n-1)^2+1}{2n^2} \mathbf{e}_{j,i}$$

where  $\mathbf{e}_{\alpha_1, \alpha_2}$  is a matrix of zeros except a 1 at the location  $(\alpha_1, \alpha_2)$ .  $\mathbf{p}_{i,j}$  has the following form.

$$\mathbf{p}_{i,j} = \sum_{t, t \neq i, t \neq j} \frac{2n-4}{4n^2} \mathbf{e}_{i,t} + \sum_{s, s \neq i, s \neq j} \frac{2n-4}{4n^2} \mathbf{e}_{s,j}$$

$\mathbf{q}_{i,j}$  is defined as follows.

$$\mathbf{q}_{i,j} = \sum_{(t,s), (t,s) \neq (i,j)} -\frac{1}{n^2} \mathbf{e}_{s,t}$$

where  $(t, s) \neq (i, j)$  is defined as  $\{t, s\} \cap \{i, j\} = \emptyset$ .

*Proof.* To make the proposed form concrete, before proceeding with the proof, consider the following example.

a) *Example 1: The form of  $\mathbf{v}_{1,2}$ :* Consider the case where  $n = 5$ . Using the proposed form, the matrix  $\mathbf{v}_{1,2}$  can be written as follows

$$\mathbf{v}_{1,2} = \tilde{\mathbf{w}}_{1,2} + \mathbf{p}_{1,2} + \mathbf{q}_{1,2}$$

where

$$\begin{aligned}\tilde{\mathbf{w}}_{1,2} &= \frac{1}{25} \begin{pmatrix} 4 & -\frac{17}{2} & 0 & 0 & 0 \\ -\frac{17}{2} & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{p}_{1,2} &= \frac{1}{25} \begin{pmatrix} 0 & 0 & 1.5 & 1.5 & 1.5 \\ 0 & 0 & 1.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 1.5 & 0 & 0 & 0 \\ 1.5 & 1.5 & 0 & 0 & 0 \end{pmatrix}\end{aligned}$$

and

$$\mathbf{q}_{1,2} = \frac{1}{25} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & -1 & -1 & -1 \end{pmatrix}$$

Therefore,  $\mathbf{v}_{1,2}$  has the following explicit form

$$\mathbf{v}_{1,2} = \frac{1}{25} \begin{pmatrix} 4 & -\frac{17}{2} & 1.5 & 1.5 & 1.5 \\ -\frac{17}{2} & 4 & 1.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & -1 & -1 & -1 \\ 1.5 & 1.5 & -1 & -1 & -1 \\ 1.5 & 1.5 & -1 & -1 & -1 \end{pmatrix}$$

As a first step simple check, consider  $\langle \mathbf{v}_{1,2}, \mathbf{w}_{1,2} \rangle$  which results  $\frac{1}{25}(8 + 17) = 1$  as desired.

The proof of Lemma 1 relies on showing that  $\mathbf{v}_{i,j}$  is dual to  $\mathbf{w}_{i,j}$ . Since the dual basis is unique, establishing duality will conclude the proof. The result will be shown considering different cases.

Case 1:  $\langle \mathbf{v}_{i,j}, \mathbf{w}_{i,j} \rangle$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{i,j} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{i,j} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{i,j} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{i,j} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{i,j} \rangle \\ &= \frac{2(n-1)}{n^2} + \frac{2(n-1)^2 + 2}{2n^2} + 0 + 0 = 1\end{aligned}$$

The second to last equality follows since  $\langle \mathbf{p}_{i,j}, \mathbf{w}_{i,j} \rangle = 0$ . and  $\langle \mathbf{q}_{i,j}, \mathbf{w}_{i,j} \rangle = 0$ .

Case 2:  $\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle$  for  $(i, j) \neq (\alpha, \beta)$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle \\ &= 0 + 0 + 0 = 0\end{aligned}$$

Case 3:  $\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle$  for  $\{i, j\} \cap \{\alpha, \beta\} \neq \emptyset$

A.  $i = \alpha$  and  $j \neq \beta$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle \\ &= \frac{n-1}{n^2} - \frac{2n-4}{2n^2} - \frac{1}{n^2} = 0\end{aligned}$$

B.  $i \neq \alpha$  and  $j = \beta$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle \\ &= \frac{n-1}{n^2} - \frac{2n-4}{2n^2} - \frac{1}{n^2} = 0\end{aligned}$$

C.  $i = \beta$  and  $j \neq \alpha$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle \\ &= \frac{n-1}{n^2} - \frac{2n-4}{2n^2} - \frac{1}{n^2} = 0\end{aligned}$$

D.  $i \neq \beta$  and  $j = \alpha$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle = \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{w}_{\alpha,\beta} \rangle \\ &= \frac{n-1}{n^2} - \frac{2n-4}{2n^2} - \frac{1}{n^2} = 0\end{aligned}$$

With this, it can be concluded that the basis  $\{\mathbf{v}_{i,j}\}$  is dual to  $\{\mathbf{w}_{i,j}\}$  and the proposed form is established.  $\square$

The next Lemma uses the result of the above Lemma to derive an explicit form of the matrix  $\mathbf{H}^{-1}$ .

**Lemma A.3.** *The inverse of the matrix  $\mathbf{H}$ ,  $\mathbf{H}^{-1}$ , has the following explicit form*

$$\mathbf{H}^{\alpha,\beta} = \langle \mathbf{v}_\alpha, \mathbf{v}_\beta \rangle = \begin{cases} \frac{(n-1)^2 + 1}{2n^2} & \text{if } \alpha_1 = \beta_1 \& \alpha_2 = \beta_2 \\ \frac{1}{n^2} & \text{if } \alpha_1 \neq \beta_1 \& \alpha_1 \neq \beta_2 \& \alpha_2 \neq \beta_1 \& \alpha_2 \neq \beta_2 \\ \frac{4-2n}{4n^2} & \text{otherwise} \end{cases}$$

*Proof.* We consider three different cases.

Case 1:  $\langle \mathbf{v}_{i,j}, \mathbf{v}_{i,j} \rangle$

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{v}_{i,j} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j} \rangle \\ &= \langle \tilde{\mathbf{w}}_{i,j}, \tilde{\mathbf{w}}_{i,j} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{p}_{i,j} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{q}_{i,j} \rangle \\ &= 2\frac{(n-1)^2}{n^4} + 2\left[\frac{(n-1)^2 + 1}{2n^2}\right]^2 + \left(\frac{2n-4}{4n^2}\right)^2 4(n-2) + \frac{1}{n^4}(n-2)^2 \\ &= \frac{2n^2 - 4n + 2}{n^4} + 2\left(\frac{n^4 - 4n^3 + 8n^2 - 8n + 4}{4n^4}\right) + \frac{16n^3 - 96n^2 + 192n - 128}{16n^4} + \frac{n^2 - 4n + 4}{n^4} \\ &= \frac{1}{2} - \frac{1}{n} + \frac{1}{n^2} = \frac{(n-1)^2 + 1}{2n^2}\end{aligned}$$

The second equality uses the fact that  $\langle \tilde{w}_{i,j}, p_{i,j} \rangle = 0$ ,  $\langle p_{i,j}, q_{i,j} \rangle = 0$  and  $\tilde{w}_{i,j}, q_{i,j} = 0$ . The third inequality simply uses the definition of  $\tilde{w}_{i,j}$ ,  $p_{i,j}$  and  $q_{i,j}$ . The second to last result follows after some algebraic calculations.

Case 2:  $\langle v_{i,j}, v_{\alpha,\beta} \rangle$  for  $(i, j) \neq (\alpha, \beta)$ ,  $\{i, j\} \cap \{\alpha, \beta\} = \emptyset$

$$\begin{aligned}\langle v_{i,j}, v_{\alpha,\beta} \rangle &= \langle \tilde{w}_{i,j} + p_{i,j} + q_{i,j}, \tilde{w}_{\alpha,\beta} + p_{\alpha,\beta} + q_{\alpha,\beta} \rangle \\ &= \langle \tilde{w}_{i,j}, \tilde{w}_{\alpha,\beta} \rangle + \langle \tilde{w}_{i,j}, p_{\alpha,\beta} \rangle + \langle \tilde{w}_{i,j}, q_{\alpha,\beta} \rangle + \\ &\quad \langle p_{i,j}, \tilde{w}_{\alpha,\beta} \rangle + \langle p_{i,j}, p_{\alpha,\beta} \rangle + \langle p_{i,j}, q_{\alpha,\beta} \rangle + \langle q_{i,j}, \tilde{w}_{\alpha,\beta} \rangle + \langle q_{i,j}, p_{\alpha,\beta} \rangle + \langle q_{i,j}, q_{\alpha,\beta} \rangle\end{aligned}$$

Each term will be evaluated separately.

1.  $\langle \tilde{w}_{i,j}, \tilde{w}_{\alpha,\beta} \rangle = 0$  since for  $(i, j) \neq (\alpha, \beta)$ ,  $\tilde{w}_{i,j}$  and  $\tilde{w}_{\alpha,\beta}$  have disjoint supports.

2.  $\langle \tilde{w}_{i,j}, p_{\alpha,\beta} \rangle = 0$  since for  $(i, j) \neq (\alpha, \beta)$ ,  $\tilde{w}_{i,j}$  and  $p_{\alpha,\beta}$  have disjoint supports.

3. Consider  $\langle \tilde{w}_{i,j}, q_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle \tilde{w}_{i,j}, q_{\alpha,\beta} \rangle &= [\tilde{w}_{i,j}]_{i,i}[q_{\alpha,\beta}]_{i,i} + [\tilde{w}_{i,j}]_{j,j}[q_{\alpha,\beta}]_{j,j} + [\tilde{w}_{i,j}]_{i,j}[q_{\alpha,\beta}]_{i,j} + [\tilde{w}_{i,j}]_{j,i}[q_{\alpha,\beta}]_{j,i} \\ &= -2 \frac{(n-1)}{n^4} + \frac{(n-1)^2 + 1}{n^4}\end{aligned}$$

4.  $\langle p_{i,j}, \tilde{w}_{\alpha,\beta} \rangle = 0$  follows by a similar argument as 2.

5. Consider  $\langle p_{i,j}, p_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle p_{i,j}, p_{\alpha,\beta} \rangle &= \sum_{s,s \neq i,t \neq j} [p_{i,j}]_{i,s}[p_{\alpha,\beta}]_{s,i} + \sum_{t,t \neq i,t \neq j} [p_{i,j}]_{j,t}[p_{\alpha,\beta}]_{t,j} + \sum_{s,s \neq i,s \neq j} [p_{i,j}]_{s,i}[p_{\alpha,\beta}]_{s,i} + \sum_{t,t \neq i,t \neq j} [p_{i,j}]_{t,j}[p_{\alpha,\beta}]_{t,j} \\ &= [p_{i,j}]_{i,\alpha}[p_{\alpha,\beta}]_{i,\alpha} + [p_{i,j}]_{i,\beta}[p_{\alpha,\beta}]_{i,\beta} + [p_{i,j}]_{j,\alpha}[p_{\alpha,\beta}]_{j,\alpha} + [p_{i,j}]_{j,\beta}[p_{\alpha,\beta}]_{j,\beta} + \\ &\quad [p_{i,j}]_{\alpha,i}[p_{\alpha,\beta}]_{\alpha,i} + [p_{i,j}]_{\beta,i}[p_{\alpha,\beta}]_{\beta,i} + [p_{i,j}]_{\alpha,j}[p_{\alpha,\beta}]_{\alpha,j} + [p_{i,j}]_{\beta,j}[p_{\alpha,\beta}]_{\beta,j} \\ &= 8 \left( \frac{2n-4}{4n^2} \right)^2 = \frac{2(n-2)^2}{n^4}\end{aligned}$$

6. Consider  $\langle p_{i,j}, q_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle p_{i,j}, q_{\alpha,\beta} \rangle &= \sum_{t,t \neq i,t \neq j} [p_{i,j}]_{i,t}[q_{\alpha,\beta}]_{i,t} + \sum_{s,s \neq i,s \neq j} [p_{i,j}]_{j,s}[q_{\alpha,\beta}]_{j,s} + \sum_{t,t \neq i,t \neq j} [p_{i,j}]_{t,i}[q_{\alpha,\beta}]_{t,i} + \sum_{s,s \neq i,s \neq j} [p_{i,j}]_{s,j}[q_{\alpha,\beta}]_{s,j} \\ &= (4n-16) \frac{2n-4}{4n^2} \left( -\frac{1}{n^2} \right) = -\frac{2(n-4)(n-2)}{n^4}\end{aligned}$$

The first equality follows since it suffices to consider  $\langle p_{i,j}, q_{\alpha,\beta} \rangle$  on the support of  $p_{i,j}$ . The second and last equality result from the following analysis. Restricted to the support of  $p_{i,j}$ , the matrix  $q_{\alpha,\beta}$  is non-zero except at the entries  $[q_{\alpha,\beta}]_{i,\alpha}$ ,  $[q_{\alpha,\beta}]_{i,\beta}$ ,  $[q_{\alpha,\beta}]_{\alpha,i}$ ,  $[q_{\alpha,\beta}]_{\beta,i}$ ,  $[q_{\alpha,\beta}]_{j,\alpha}$ ,  $[q_{\alpha,\beta}]_{j,\beta}$ ,  $[q_{\alpha,\beta}]_{\alpha,j}$  and  $[q_{\alpha,\beta}]_{\beta,j}$ . Using this and the fact that  $p_{i,j}$  has  $4(n-2)$  entries,  $|\text{supp}(p_{i,j}) \cap \text{supp}(q_{\alpha,\beta})| = 4(n-2) - 8 = 4n - 16$ . With this, the final form above follows.

7.  $\langle q_{i,j}, \tilde{w}_{\alpha,\beta} \rangle = -2 \frac{(n-1)}{n^4} + \frac{(n-1)^2 + 1}{n^4}$  follows by a similar argument as 3.

8.  $\langle q_{i,j}, p_{\alpha,\beta} \rangle = -\frac{2(n-4)(n-2)}{n^4}$  follows by a similar argument as 6.

9. Consider  $\langle q_{i,j}, q_{\alpha,\beta} \rangle$ . Since  $(i, j) \neq (\alpha, \beta)$ , both  $q_{i,j}$  and  $q_{\alpha,\beta}$  have non-zero entry at  $(s, t)$  if and only if  $s \neq i, s \neq j, s \neq \alpha, s \neq \beta, t \neq i, t \neq j, t \neq \alpha$  and  $t \neq \beta$ . Therefore,  $|\text{supp}(q_{i,j}) \cap \text{supp}(q_{\alpha,\beta})| = (n-4)(n-4) = (n-4)^2$  given the  $n-4$  choices for  $s$  and  $n-4$  choices for  $t$ .  $\langle q_{i,j}, q_{\alpha,\beta} \rangle$  can now be written as follows

$$\langle q_{i,j}, q_{\alpha,\beta} \rangle = \sum_{(s,t) \in \text{supp}(q_{i,j}) \cap \text{supp}(q_{\alpha,\beta})} [q_{i,j}]_{s,t}[q_{\alpha,\beta}]_{s,t} = (n-4)^2 \left( -\frac{1}{n^2} \right) \left( -\frac{1}{n^2} \right) = \frac{(n-4)^2}{n^4}$$

Therefore,  $\langle v_{i,j}, v_{\alpha,\beta} \rangle$  is the sum of the above terms.

$$\langle v_{i,j}, v_{\alpha,\beta} \rangle = -4 \frac{n-1}{n^4} + 2 \frac{(n-1)^2 + 1}{n^4} + \frac{2(n-2)^2}{n^4} - \frac{4(n-4)(n-2)}{n^4} + \frac{(n-4)^2}{n^4} = \frac{1}{n^2}$$

The last equality follows after some algebraic manipulations.

Case 3:  $\langle \mathbf{v}_{i,j}, \mathbf{v}_{\alpha,\beta} \rangle$  for  $\{i, j\} \cap \{\alpha, \beta\} \neq \emptyset$

A.  $i = \alpha$  and  $j \neq \beta$ .

$$\begin{aligned}\langle \mathbf{v}_{i,j}, \mathbf{v}_{\alpha,\beta} \rangle &= \langle \tilde{\mathbf{w}}_{i,j} + \mathbf{p}_{i,j} + \mathbf{q}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} + \mathbf{p}_{\alpha,\beta} + \mathbf{q}_{\alpha,\beta} \rangle \\ &= \langle \tilde{\mathbf{w}}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle + \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle + \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle + \\ &\quad \langle \mathbf{p}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle + \langle \mathbf{p}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle + \langle \mathbf{q}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle\end{aligned}$$

Each term will be evaluated separately.

$$1. \langle \tilde{\mathbf{w}}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle = [\tilde{\mathbf{w}}_{i,j}]_{i,i} [\tilde{\mathbf{w}}_{\alpha,\beta}]_{i,i} = \frac{(n-1)^2}{n^4}.$$

2. Consider  $\langle \tilde{\mathbf{w}}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle \tilde{\mathbf{w}}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle &= [\tilde{\mathbf{w}}_{i,j}]_{i,j} [\mathbf{p}_{i,j}]_{i,j} + [\tilde{\mathbf{w}}_{i,j}]_{j,i} [\mathbf{p}_{i,j}]_{j,i} = -\left(\frac{2(n-1)^2+2}{2n^2}\right)\left(\frac{2n-4}{4n^2}\right) \\ &= -\frac{(n-2)[(n-1)^2+1]}{2n^4}\end{aligned}$$

$$3. \langle \tilde{\mathbf{w}}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle = [\tilde{\mathbf{w}}_{i,j}]_{j,j} [\mathbf{q}_{\alpha,\beta}]_{j,j} = \frac{n-1}{n^2} \left(-\frac{1}{n^2}\right) = -\frac{n-1}{n^4}.$$

$$4. \langle \mathbf{p}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle = -\frac{(n-2)[(n-1)^2+1]}{2n^4} \text{ follows by a similar argument as 2.}$$

5. Consider  $\langle \mathbf{p}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle \mathbf{p}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle &= \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{i,s} [\mathbf{p}_{\alpha,\beta}]_{i,s} + \sum_{t,t \neq i,t \neq j} [\mathbf{p}_{i,j}]_{j,t} [\mathbf{p}_{\alpha,\beta}]_{j,t} + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{s,i} [\mathbf{p}_{\alpha,\beta}]_{s,i} + \sum_{t,t \neq i,t \neq j} [\mathbf{p}_{i,j}]_{t,j} [\mathbf{p}_{\alpha,\beta}]_{t,j} \\ &= \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{i,s} [\mathbf{p}_{\alpha,\beta}]_{i,s} + [\mathbf{p}_{i,j}]_{j,\beta} [\mathbf{p}_{\alpha,\beta}]_{j,\beta} + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{s,i} [\mathbf{p}_{\alpha,\beta}]_{s,i} + [\mathbf{p}_{i,j}]_{t,\beta} [\mathbf{p}_{\alpha,\beta}]_{t,\beta} \\ &= (n-3) \left(\frac{2n-4}{4n^2}\right)^2 + \frac{2n-4}{4n^2} + (n-3) \left(\frac{2n-4}{4n^2}\right)^2 + \frac{2n-4}{4n^2} \\ &= 2(n-2)(n-3) \left(\frac{2n-4}{4n^2}\right)^2 = \frac{1}{2} \frac{(n-2)^3}{n^4}\end{aligned}$$

The second line follows since  $[\mathbf{p}_{i,j}]_{j,t} = 0$  for all  $t$  except  $t = \beta$  and  $t = i$  and  $[\mathbf{p}_{i,j}]_{t,j} = 0$  for all  $t$  except  $t = \beta$  and  $t = i$ . The third line uses the fact that  $[\mathbf{p}_{i,j}]_{i,s} \neq 0$  for all  $t$  except  $t = i$  and  $t = \beta$  and  $[\mathbf{p}_{i,j}]_{s,i} \neq 0$  for all  $t$  except  $t = i$  and  $t = \beta$ . The final equality results after some algebraic manipulations.

6. Consider  $\langle \mathbf{p}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle$ .

$$\begin{aligned}\langle \mathbf{p}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle &= \sum_{t,t \neq i,t \neq j} [\mathbf{p}_{i,j}]_{i,t} [\mathbf{q}_{\alpha,\beta}]_{i,t} + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{j,s} [\mathbf{q}_{\alpha,\beta}]_{j,s} + \sum_{t,t \neq i,t \neq j} [\mathbf{p}_{i,j}]_{t,i} [\mathbf{q}_{\alpha,\beta}]_{t,i} + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{s,j} [\mathbf{q}_{\alpha,\beta}]_{s,j} \\ &= 0 + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{j,s} [\mathbf{q}_{\alpha,\beta}]_{j,s} + 0 + \sum_{s,s \neq i,s \neq j} [\mathbf{p}_{i,j}]_{s,j} [\mathbf{q}_{\alpha,\beta}]_{s,j} \\ &= 2(n-3) \frac{2n-4}{4n^2} \left(-\frac{1}{n^2}\right) = -\frac{(n-3)(n-2)}{n^4}\end{aligned}$$

The first equality follow since it suffices to consider  $\langle \mathbf{p}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle$  on the support of  $\mathbf{p}_{i,j}$ . The second equality results since  $[\mathbf{q}_{i,j}]_{i,t} = [\mathbf{q}_{i,j}]_{t,i} = 0$  for any  $t$ . The third and last equality result from the following analysis. Restricted to the support of  $\mathbf{p}_{i,j}$ , the matrix  $\mathbf{q}_{i,j}$  is zero except at the entries  $[\mathbf{q}_{i,j}]_{j,s}$  and  $[\mathbf{q}_{i,j}]_{s,j}$  for all  $j \neq \beta$ . With this,  $|\text{supp}(\mathbf{p}_{i,j}) \cap \text{supp}(\mathbf{q}_{\alpha,\beta})| = 2[(n-2)-1] = 2(n-3)$  and the final form follows.

$$7. \langle \mathbf{q}_{i,j}, \tilde{\mathbf{w}}_{\alpha,\beta} \rangle = -\frac{n-1}{n^4} \text{ follows by a similar argument as 3.}$$

$$8. \langle \mathbf{q}_{i,j}, \mathbf{p}_{\alpha,\beta} \rangle = -\frac{(n-3)(n-2)}{n^4} \text{ follows by a similar argument as 6.}$$

9. Consider  $\langle \mathbf{q}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle$ . Since  $i = \alpha$  and  $j \neq (\alpha, \beta)$ , both  $\mathbf{q}_{i,j}$  and  $\mathbf{q}_{\alpha,\beta}$  have a non-zero entry at  $(s, t)$  if and only if  $s \neq i, s \neq \alpha, s \neq \beta, t \neq i, t \neq j$ , and  $t \neq \beta$ . Therefore,  $|\text{supp}(\mathbf{q}_{i,j}) \cap \text{supp}(\mathbf{q}_{\alpha,\beta})| = (n-3)(n-3) = (n-3)^2$  given the  $n-3$  choices for  $s$  and  $n-3$  choices for  $t$ .  $\langle \mathbf{q}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle$  can now be written as follows

$$\langle \mathbf{q}_{i,j}, \mathbf{q}_{\alpha,\beta} \rangle = \sum_{(s,t) \in \text{supp}(\mathbf{q}_{i,j}) \cap \text{supp}(\mathbf{q}_{\alpha,\beta})} [\mathbf{q}_{i,j}]_{s,t} [\mathbf{q}_{\alpha,\beta}]_{s,t} = (n-3)^2 \left(-\frac{1}{n^2}\right) \left(-\frac{1}{n^2}\right) = \frac{(n-3)^2}{n^4}$$

Therefore,  $\langle \mathbf{v}_{i,j}, \mathbf{v}_{\alpha,\beta} \rangle$  is the sum of the above terms.

$$\begin{aligned} \langle \mathbf{v}_{i,j}, \mathbf{v}_{\alpha,\beta} \rangle &= \frac{(n-1)^2}{n^4} + -\frac{(n-2)[(n-1)^2 + 1]}{n^4} - 2\frac{n-1}{n^4} + \frac{1}{2}\frac{(n-2)^3}{n^4} - 2\frac{(n-3)(n-2)}{n^4} + \frac{(n-3)^2}{n^4} \\ &= \frac{2-n}{2n^2} \end{aligned}$$

The last equality follows after some algebraic manipulations.

Now the following three cases remain:  $i \neq \alpha$  and  $j = \beta$ ,  $i = \beta$  and  $j \neq \alpha$  and  $i \neq \beta$  and  $j = \alpha$ . However, since  $\mathbf{v}_{i,j}$  is symmetric the above argument can be adapted to these three cases by interchanging indices as appropriate. This concludes the proof.  $\square$

**Remark 5.** A short proof of the form of  $\mathbf{H}^{-1}$  might be plausible. The main technical challenge has been the locations of 1 and 0's in  $\mathbf{H}$  which does not lend itself to simple analysis.

Next, in Lemma A.4, we state and prove the spectral properties of the matrix  $\mathbf{H}$ .

**Lemma A.4.** The matrix  $\mathbf{H} \in \mathbb{R}^{L \times L}$  is symmetric and positive definite. The minimum eigenvalue of  $\mathbf{H}$  is at least 1 and the maximum eigenvalue is  $2n$ . The absolute sum of each row of  $\mathbf{H}^{-1}$  is given by  $2 - \frac{15}{2n} + \frac{8}{n^2}$ .

*Proof.* The symmetry of  $\mathbf{H}$  is trivial as  $\mathbf{H}_{\alpha,\beta} = \langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle = \langle \mathbf{w}_\beta, \mathbf{w}_\alpha \rangle = \mathbf{H}_{\beta,\alpha}$ .  $\mathbf{H}$  is positive definite since  $\mathbf{H} = \mathbf{W}^T \mathbf{W}$  and  $\mathbf{W}$  has linearly independent columns. With  $\alpha = (\alpha_1, \alpha_2)$  and  $\beta = (\beta_1, \beta_2)$ ,  $\mathbf{H}$  is defined as follows

$$\mathbf{H}_{\alpha,\beta} = \langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle = \begin{cases} 4 & \text{if } \alpha_1 = \beta_1 \& \alpha_2 = \beta_2 \\ 0 & \text{if } \alpha_1 \neq \beta_1 \& \alpha_1 \neq \beta_2 \& \alpha_2 \neq \beta_1 \& \alpha_2 \neq \beta_2 \\ 1 & \text{otherwise} \end{cases}$$

After minor analysis of  $\mathbf{H}$ , the number of zeros in any given row is given by  $\frac{(n-2)(n-3)}{2}$ . The number of ones is simply

$$L - \# \text{ number of zeros} - \# \text{ number of fours} = \frac{n(n-1)}{2} - \frac{(n-2)(n-3)}{2} - 1 = 2n - 4$$

To find the maximum eigenvalue of  $\mathbf{H}$ , note that  $\mathbf{1}$  is an eigenvector of  $\mathbf{H}$  with eigenvalue  $2n$ . From Gershgorin theorem, the upper bound for an eigenvalue is simply  $4 + (2n-4) = 2n$ . It follows that the maximum eigenvalue of  $\mathbf{H}$  is  $2n$ . Using the form of  $\mathbf{H}^{-1}$  from Lemma A.3, consider the absolute sum of any row of  $\mathbf{H}^{-1}$ ,  $\sum_i |\mathbf{H}_{ij}^{-1}|$ .

$$\sum_i |\mathbf{H}_{ij}^{-1}| = \frac{1}{n^2} \left[ \frac{(n-1)^2 + 1}{2} \right] + \frac{1}{n^2} \left[ \frac{(n-2)(n-3)}{2} \right] + \frac{1}{n^2} (2n-4) \left| 1 - \frac{1}{2}n \right| = 2 - \frac{15}{2n} + \frac{8}{n^2}$$

Finally, we make use of a variant of Gershgorin's theorem to bound the maximum eigenvalue of  $\mathbf{H}^{-1}$ . If  $\mathbf{P}$  is invertible,  $\mathbf{P}^{-1} \mathbf{H}^{-1} \mathbf{P}$  and  $\mathbf{H}^{-1}$  have the same eigenvalues. For simplicity, let  $\mathbf{P} = \text{diag}(d_1, d_2, \dots, d_L)$ . In [43, p. 347], this fact was used to state the following variant of Gershgorin's theorem. Below, we restate this result, albeit minor changes, for ease of reference.

**Corollary 2** ([43, p. 347]). Let  $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{L \times L}$  and let  $d_1, d_2, \dots, d_L$  be positive real numbers. Then all eigenvalues of  $\mathbf{A}$  lie in the region

$$\bigcup_{i=1}^L \left\{ z \in \mathbb{R} : |z - a_{ii}| \leq \frac{1}{d_i} \sum_{j=1, j \neq i}^L d_j |a_{ij}| \right\}$$

Using the corollary, set  $d_1 = 9$  and  $d_2 = d_3 = \dots = d_L = 3$ . Apply the corollary on the matrix  $\mathbf{H}^{-1}$ . After minor calculation, we have  $\lambda_{\max}(\mathbf{H}^{-1}) \leq 1$ . We remark here that with suitable choice of  $(d_1, d_2, \dots, d_L)$ , the bound can be tightened to  $\lambda_{\max}(\mathbf{H}^{-1}) = \frac{1}{2}$  but the current bound is sufficient for our analysis.  $\square$

Assume that the underlying inner product matrix  $\mathbf{M}$  has coherence  $v$  with respect to the standard basis. Let  $e_i \in \mathbb{R}^n$  be the standard vector, a vector of zeros except a 1 in the  $i$ th position. For all  $i$ ,  $1 \leq i \leq n$ , the coherence definition [10] states that

$$\|\mathcal{P}_{\mathbf{U}} e_i\|_2^2 \leq \frac{vr}{n} \quad (46)$$

It suffices to consider the condition on  $\mathbf{U}$  since  $\mathbf{M}$  is symmetric. Given this, could one derive coherence conditions for the EDG problem? The answer is affirmative and is given in Lemma A.5 below.

**Lemma A.5.** *If the underlying inner product matrix  $\mathbf{M}$  has coherence  $v$  with respect to the standard basis, i.e.  $\mathbf{M}$  satisfies (46), then the following coherence conditions hold for the EDG problem.*

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \frac{8vr}{n} \quad ; \quad \|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F^2 \leq \frac{8vr}{n}$$

*Proof.* Using the definition of  $\mathcal{P}_{\mathbb{T}}$  and the fact that  $\mathbf{w}_\alpha$  is symmetric for any  $\alpha$ , we have

$$\begin{aligned} \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 &= \langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle + \langle \mathbf{w}_\alpha, \mathbf{w}_\alpha \mathbf{U}\mathbf{U}^T \rangle - \langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \mathbf{U}\mathbf{U}^T \rangle \\ &= 2\langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle - \langle \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \mathbf{U}\mathbf{U}^T, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \mathbf{U}\mathbf{U}^T \rangle \leq 2\langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle \end{aligned}$$

Note that  $\langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle = \langle \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle \geq 0$ . Using the definition of  $\langle \mathbf{X}, \mathbf{w}_\alpha \rangle$  and the fact that  $\mathbf{w}_\alpha^2 = 2\mathbf{w}_\alpha$  for any  $\alpha$ ,  $\langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle = 2[(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2} - 2(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_2}] \leq 2[(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2} + 2|(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_2}|] \leq 4[(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2}]$ . The last inequality holds since  $\mathbf{U}\mathbf{U}^T$  is positive semidefinite. This motivates a bound on  $\max_{ij} |\mathbf{U}\mathbf{U}^T|_{i,j}$ .

$$|\mathbf{U}\mathbf{U}^T|_{i,j} = \left| \sum_{k=1}^r \mathbf{U}_{i,k} \mathbf{U}_{j,k} \right| \leq \sum_{k=1}^r |\mathbf{U}_{i,k}| |\mathbf{U}_{j,k}| \leq \sqrt{\sum_{k=1}^r \mathbf{U}_{i,k}^2} \sqrt{\sum_{k=1}^r \mathbf{U}_{j,k}^2} \leq \sqrt{\frac{vr}{n}} \sqrt{\frac{vr}{n}} = \frac{vr}{n}$$

Using the above bound,  $\langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \mathbf{w}_\alpha \rangle \leq 8\frac{vr}{n}$  resulting the following bound for  $\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2$ .

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \frac{16vr}{n}$$

To bound  $\|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F^2$ , note that  $\|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F \leq \sum_{\beta \in \mathbb{I}} \|\mathbf{H}^{\alpha, \beta} \mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F = \sum_{\beta \in \mathbb{I}} \|\mathbf{H}^{\alpha, \beta}\| \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F$ . Using Lemma A.4 and the bound for  $\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F$ ,

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{v}_\alpha\|_F^2 \leq \frac{64vr}{n}$$

□

In addition, the standard matrix completion analysis assumes that  $\max_{ij} |(\mathbf{U}\mathbf{U}^T)_{i,j}| \leq \frac{\mu_1 \sqrt{r}}{n}$  for some constant  $\mu_1$  [10]. If this holds, for some  $\alpha$ , it follows that

$$\begin{aligned} \langle \mathbf{U}\mathbf{U}^T, \mathbf{w}_\alpha \rangle &= (\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2} - 2(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_2} \\ &\leq (\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2} + 2|(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_2}| \\ &\leq 2[(\mathbf{U}\mathbf{U}^T)_{\alpha_1, \alpha_1} + (\mathbf{U}\mathbf{U}^T)_{\alpha_2, \alpha_2}] \leq 4 \frac{\mu_1 \sqrt{r}}{n} \end{aligned}$$

With this, it can be concluded that

$$\max_{\alpha \in \mathbb{I}} \langle \mathbf{w}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2 \leq 16 \frac{\mu_1^2 r}{n^2} \quad (47)$$

Another short calculation results a bound on  $\langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2$

$$\langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2 = \langle \sum_{\beta \in \mathbb{I}} \mathbf{H}^{\alpha, \beta} \mathbf{w}_\beta, \mathbf{U}\mathbf{U}^T \rangle^2 \leq \max_{\beta \in \mathbb{I}} \langle \mathbf{w}_\beta, \mathbf{U}\mathbf{U}^T \rangle^2 \left( \sum_{\beta \in \mathbb{I}} |\mathbf{H}^{\alpha, \beta}| \right)^2 \leq 64 \frac{\mu_1 r}{n^2}$$

The last inequality uses (47) and Lemma A.4. From the above inequality, it follows that

$$\max_{\alpha \in \mathbb{I}} \langle \mathbf{v}_\alpha, \mathbf{U}\mathbf{U}^T \rangle^2 \leq 64 \frac{\mu_1^2 r}{n^2}$$

Lemma A.5 and the discussion above show that the coherence conditions with respect to standard basis lead to comparable EDG coherence conditions. Specifically, we obtain conditions equivalent up to constants to (15), (16) and (17). We remark here that the condition (13) does not simply follow from the coherence conditions with respect to the standard basis. We speculate that the equivalence is possible under certain assumptions but a rigorous analysis is left for future work.

**Lemma A.6.** Given any  $\mathbf{X} \in \mathbb{S}$ , the following estimates hold.

$$\frac{1}{2n} \|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle^2 \leq \|\mathbf{X}\|_F^2 \quad ; \quad \|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq 2n \|\mathbf{X}\|_F^2$$

*Proof.* Vectorize the matrix  $\mathbf{X}$  and each dual basis  $\mathbf{v}_\alpha$ . It follows that

$$\sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle^2 = \sum_{\alpha \in \mathbb{I}} \mathbf{x}^T \mathbf{v}_\alpha \mathbf{v}_\alpha^T \mathbf{x} = \mathbf{x}^T \mathbf{V} \mathbf{V}^T \mathbf{x}$$

Orthogonalize  $\mathbf{V}$ ,  $\bar{\mathbf{V}} = \mathbf{V}(\sqrt{\mathbf{H}^{-1}})^{-1}$ . With this,  $\sum_{\beta \in \mathbb{I}} \langle \mathbf{X}, \mathbf{v}_\beta \rangle^2 = \mathbf{x}^T \bar{\mathbf{V}} \mathbf{H}^{-1} \bar{\mathbf{V}}^T \mathbf{x}$ . It follows that

$$\lambda_{\min}(\mathbf{H}^{-1}) \|\mathbf{x}\|_2^2 = \frac{1}{2n} \|\mathbf{x}\|_2^2 \leq \sum_{\beta \in \mathbb{I}} \langle \mathbf{X}, \mathbf{v}_\beta \rangle^2 \leq \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_2^2$$

where the above result follows from the min-max theorem and Lemma A.4. Proceeding analogously as above, noting that  $\lambda_{\max}(\mathbf{H}) = n$  and  $\lambda_{\min}(\mathbf{H}) \geq 1$  from Lemma A.4, we obtain

$$\|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq 2n \|\mathbf{X}\|_F^2$$

This concludes the proof.  $\square$

**Lemma A.7.** Let  $c_\alpha \geq 0$ . Then,

$$\left\| \sum_{\alpha} c_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2 \right\| \leq \left\| \sum_{\alpha} c_\alpha \mathbf{w}_\alpha^2 \right\|$$

*Proof.* Using the definition of  $\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha$ ,  $\left\| \sum_{\alpha} c_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2 \right\|$  can be written as follows.

$$\left\| \sum_{\alpha} c_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2 \right\| = \left\| \sum_{\alpha} c_\alpha \mathcal{P}_{U^\perp} \mathbf{w}_\alpha \mathcal{P}_{U^\perp} \mathbf{w}_\alpha \mathcal{P}_{U^\perp} \right\| = \left\| \mathcal{P}_{U^\perp} \left( \sum_{\alpha} c_\alpha \mathbf{w}_\alpha \mathcal{P}_{U^\perp} \mathbf{w}_\alpha \right) \mathcal{P}_{U^\perp} \right\|$$

Using the fact that the operator norm is unitarily invariant and  $\|\mathcal{P} \mathbf{X} \mathcal{P}\| \leq \|\mathbf{X}\|$  for any  $\mathbf{X}$  and a projection  $\mathcal{P}$ ,  $\left\| \sum_{\alpha} c_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2 \right\|$  can be upper bounded as follows

$$\left\| \sum_{\alpha} c_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^2 \right\| \leq \left\| \sum_{\alpha} c_\alpha \mathbf{w}_\alpha \mathcal{P}_{U^\perp} \mathbf{w}_\alpha \right\| = \left\| \sum_{\alpha} c_\alpha \mathbf{w}_\alpha (\mathbf{w}_\alpha - \mathcal{P}_U \mathbf{w}_\alpha) \right\| = \left\| \sum_{\alpha} [c_\alpha \mathbf{w}_\alpha^2 - c_\alpha \mathbf{w}_\alpha \mathcal{P}_U \mathbf{w}_\alpha] \right\|$$

where the first equality follows from the relation  $\mathcal{P}_{U^\perp} = \mathcal{I} - \mathcal{P}_U$ . Since  $\mathbf{w}_\alpha = \mathbf{w}_\alpha^T$  and  $c_\alpha \geq 0$ ,  $\sum_{\alpha} c_\alpha \mathbf{w}_\alpha^2$  is positive semidefinite. Using the relation  $\mathcal{P}_U^2 = \mathcal{P}_U$  and the assumption that  $c_\alpha \geq 0$ ,  $\sum_{\alpha} c_\alpha \mathbf{w}_\alpha \mathcal{P}_U \mathbf{w}_\alpha = \sum_{\alpha} c_\alpha \mathbf{w}_\alpha \mathcal{P}_U \mathcal{P}_U \mathbf{w}_\alpha$  is also positive semidefinite. Repeating the same argument,  $\sum_{\alpha} c_\alpha \mathbf{w}_\alpha \mathcal{P}_{U^\perp} \mathbf{w}_\alpha$  is also positive semidefinite. Finally, the norm inequality,  $\|A + B\| \geq \max(\|A\|, \|B\|)$ , for positive semidefinite matrices  $A$  and  $B$ , concludes the proof.  $\square$

**Lemma A.8.** Define  $\eta(\mathbf{X}) = \max_{\beta \in \mathbb{I}} |\langle \mathbf{X}, \mathbf{v}_\beta \rangle|$ . For a fixed  $\mathbf{X} \in \mathbb{T}$ , the following estimate holds.

$$Pr(\max_{\beta \in \mathbb{I}} |\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{v}_\beta \rangle| \geq t) \leq n^2 \exp\left(-\frac{3t^2 \kappa_j}{16\eta(\mathbf{X})^2 \left(v + \frac{1}{nr}\right)}\right) \quad (48)$$

for all  $t \leq \eta(\mathbf{X})$  with  $\kappa_j = \frac{m_j}{nr}$ .

*Proof.* For some  $\mathbf{v}_\beta$ , expand  $\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{v}_\beta \rangle$  in the following way:

$$\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{v}_\beta \rangle = \left\langle \sum_{\alpha \in \Omega_j} \frac{L}{m_j} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha - \mathbf{X}, \mathbf{v}_\beta \right\rangle = \sum_{\alpha \in \Omega_j} \left( \frac{L}{m_j} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle - \frac{1}{m_j} \langle \mathbf{X}, \mathbf{v}_\beta \rangle \right)$$

Note that the summand can be written as  $Y_\alpha = X_\alpha - E[X_\alpha]$ . By construction,  $E[Y_\alpha] = 0$ . To apply Bernstein inequality, it remains to compute a suitable bound for  $|Y_\alpha|$  and  $|E[Y_\alpha^2]|$ .  $|Y_\alpha|$  is bounded as follows.

$$\begin{aligned} |Y_\alpha| &= \left| \frac{L}{m_j} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle - \frac{1}{m_j} \langle \mathbf{X}, \mathbf{v}_\beta \rangle \right| \leq \frac{L}{m_j} \eta(\mathbf{X}) 4 \frac{vr}{n} + \frac{1}{m_j} \eta(\mathbf{X}) \\ &< \frac{1}{m_j} \eta(\mathbf{X}) (1 + 2nr v) \end{aligned}$$

Above, the first inequality follows from the coherence estimates (15) and (16). Next, we consider a bound  $|E[Y_\alpha^2]|$ . Since  $E[Y_\alpha^2] = E[X_\alpha^2] - (E[X_\alpha])^2$ ,  $E[Y_\alpha^2]$  can be upper bounded as follows

$$\begin{aligned} E[Y_\alpha^2] &\leq E\left[\frac{L^2}{m_j^2}\langle \mathbf{X}, \mathbf{v}_\alpha \rangle^2 \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle^2\right] + \frac{1}{m_j^2}\langle \mathbf{X}, \mathbf{v}_\beta \rangle^2 \\ &= \frac{L}{m_j^2} \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{v}_\alpha \rangle^2 \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle^2 + \frac{1}{m_j^2}\langle \mathbf{X}, \mathbf{v}_\beta \rangle^2 \\ &\leq \eta(\mathbf{X})^2 \frac{n^2}{2m_j^2} \sum_{\alpha \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle^2 + \frac{1}{m_j^2}\eta(\mathbf{X})^2 \\ &\leq \eta(\mathbf{X})^2 \frac{2nrv}{m_j^2} + \frac{2}{m_j^2}\eta(\mathbf{X})^2 = \frac{2\eta(\mathbf{X})^2}{m_j^2}(1 + nrv) \end{aligned}$$

Above, the last inequality follows from the coherence estimate (13). Finally, apply the scalar Bernstein inequality with  $R = \frac{\eta(\mathbf{X})(1 + 2nrv)}{m_j}$  and  $\sigma^2 = \frac{2\eta(\mathbf{X})^2(1 + nrv)}{m_j}$ . For  $t \leq \frac{\sigma^2}{R} > \eta(\mathbf{X})$ ,

$$\Pr(|\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{v}_\beta \rangle| \geq t) \leq \exp\left(-\frac{3t^2\kappa_j}{16\eta(\mathbf{X})^2\left(v + \frac{1}{nr}\right)}\right) \quad (49)$$

The proof of Lemma A.8 concludes by simply applying the union bound.  $\square$