

Low-rank Matrix Completion in a General Non-orthogonal Basis^{*}

Abiy Tasissa^{a,*}, Rongjie Lai^{b,*}

^a*Department of Mathematics, Tufts University, Medford, MA 02155*

^b*Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180*

Keywords: Matrix completion, nuclear norm minimization, restricted isometry property, operator Chernoff bounds, dual basis

1. Introduction

Recovering low-rank matrices from given incomplete linear measurements plays an important role in many problems such as image and video processing [1], model reduction [2], phase retrieval [3], molecular conformation [4, 5, 6], localization in sensor networks [7, 8], dimensionality reduction [9], recommender systems [10] as well as solving PDEs on manifold-structured data represented as incomplete distance [11], just to name a few. A natural framework to the low-rank recovery problem is rank minimization under linear constraints. However, this problem is NP-hard [12] and thus motivates alternative solutions. A series of theoretical papers [13, 14, 15, 16, 12] showed that the NP-hard rank minimization problem for matrix completion can be obtained by solving the following convex nuclear norm minimization problem:

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{n \times n}}{\text{minimize}} && \|\mathbf{X}\|_* \\ & \text{subject to} && \mathbf{X}_{i,j} = \mathbf{M}_{i,j} \quad (i,j) \in \Omega \end{aligned} \tag{1}$$

^{*}This work is supported in part by NSF DMS-1522645 and an NSF Career Award DMS-1752934.

^{*}Corresponding author

Email address: abiy.tasissa@tufts.edu (Abiy Tasissa)

where $\|\mathbf{X}\|_*$ denotes the nuclear norm defined as the sum of the singular values of \mathbf{X} , and $\Omega \subset \{(i, j) | i, j = 1, \dots, n\}$, $|\Omega| = m$, denotes a random set that consists of the sampled indices. The remarkable fact is that, under certain conditions, the underlying low-rank matrix can be reconstructed exactly with high probability from only $O(nr \log^2(n))$ uniformly random measurements. The idea to use the nuclear norm as an approximation of the rank function was first discussed in [2]. Loosely, minimizing the sum of singular values will likely lead to a solution with many zero singular values resulting a low rank matrix. One generalization of the matrix completion problem in [15] considers measurements with respect to a symmetric orthonormal basis and gives comparable theoretical guarantees based on the elegant dual certificate analysis. In particular, it shows that the true low rank matrix can be recovered with high probability from $O(nr \log^2(n))$ uniformly random measurements.

The starting point and inspiration for this work was our recent work in [17] which studies a matrix completion problem with respect to a specific non-orthogonal basis. In this paper, we consider the matrix completion problem with respect to any non-orthogonal basis. Given a general unit-norm basis $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$ which spans an L dimensional subspace \mathbb{S} of $\mathbb{R}^{n \times n}$, the nuclear norm minimization program for this general matrix completion problem is provided by

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{S}}{\text{minimize}} && \|\mathbf{X}\|_* \\ & \text{subject to} && \langle \mathbf{X}, \mathbf{w}_\alpha \rangle = \langle \mathbf{M}, \mathbf{w}_\alpha \rangle \quad \alpha \in \Omega \end{aligned} \tag{2}$$

where Ω denotes a random set sampling the basis indices. We are interested in the following two problems.

1. Could we obtain comparable recovery guarantees for the general matrix completion problem?
2. If the answer to (1) is affirmative, what conditions are needed on the basis $\{\mathbf{w}_\alpha\}$ and Ω ?

The main goal of this paper is a theoretical analysis of these two problems. We start by discussing few examples which show how the problem naturally arises in several applications.

Euclidean Distance Geometry Problem.. Given partial information on pairwise distances, the Euclidean distance geometry problem is concerned with constructing the configuration of points. The problem has applications in diverse areas [4, 5, 9, 7, 11]. Formally, consider a set of n points in an r -dimensional space $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}^T \in \mathbb{R}^{n \times r}$. Let $\mathbf{D} = [d_{i,j}^2]$ denote the squared Euclidean distance matrix. The inner product matrix, also known as the Gram matrix and defined as $\mathbf{X}_{i,j} = \langle \mathbf{p}_i, \mathbf{p}_j \rangle$, is a positive semidefinite matrix of rank r . A minor analysis reveals that \mathbf{D} and \mathbf{X} can be related in the following way: $\mathbf{D}_{i,j} = \mathbf{X}_{i,i} + \mathbf{X}_{j,j} - 2\mathbf{X}_{i,j}$. For $r \ll n$, consider the following nuclear norm minimization program to recover \mathbf{X} .

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{n \times n}}{\text{minimize}} && \|\mathbf{X}\|_* \\ & \text{subject to} && \mathbf{X}_{i,i} + \mathbf{X}_{j,j} - 2\mathbf{X}_{i,j} = \mathbf{D}_{i,j} \quad (i,j) \in \Omega \\ & && \mathbf{X} \cdot \mathbf{1} = \mathbf{0}; \mathbf{X} = \mathbf{X}^T; \mathbf{X} \in \mathbb{S}_+ \end{aligned} \quad (3)$$

In the above program, \mathbb{S}_+ denotes the set of positive semidefinite matrices and the constraint $\mathbf{X} \cdot \mathbf{1} = \mathbf{0}$ fixes the translation ambiguity. The above minimization problem can be equivalently interpreted as a general matrix completion problem with respect to some operator basis \mathbf{w}_α .

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

where $\mathbf{w}_\alpha = \frac{1}{2}(\mathbf{e}_{\alpha_1, \alpha_1} + \mathbf{e}_{\alpha_2, \alpha_2} - \mathbf{e}_{\alpha_1, \alpha_2} - \mathbf{e}_{\alpha_2, \alpha_1})$ and $\mathbb{S} = \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^T \text{ \& } \mathbf{X} \cdot \mathbf{1} = \mathbf{0}\} \cap \mathbb{S}_+$ with \mathbb{S}_+ denoting the set of positive semidefinite matrices. The constant $\frac{1}{2}$ is a normalization constant and $\mathbf{e}_{\alpha_1, \alpha_2}$ is a matrix whose entries are all zero except a 1 at the (α_1, α_2) -th entry. It can be verified that $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$, $L = \frac{n(n-1)}{2}$, is a non-orthogonal basis for the linear space $\{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} =$

$\mathbf{X}^T \& \mathbf{X} \cdot \mathbf{1} = \mathbf{0}\}$. Theoretical analysis of this problem was recently conducted by the authors of this paper in [17] and in fact inspired this work.

Spectrally Sparse Signal Reconstruction.. The problem of signal reconstruction has many important practical applications. When the underlying signal is assumed to be sparse, the theory of compressive sensing states that the signal can be recovered by solving the convex l_1 minimization problem [18]. In [1], the authors consider the recovery of spectrally sparse signal $\mathbf{x} \in \mathbb{R}^n$ of known order r where $r \ll n$. Let $\mathcal{H} : \mathbb{C}^n \rightarrow \mathbb{C}^{n_1 \times n_2}$ be the linear operator that maps a vector $\mathbf{z} \in \mathbb{C}^n$ to a Hankel matrix $\mathcal{H}\mathbf{z} \in \mathbb{C}^{n_1 \times n_2}$ with $n+1 = n_1 + n_2$. Denote the orthonormal basis of $n_1 \times n_2$ Hankel matrices by $\{\mathbf{H}_\alpha\}_{\alpha=1}^{n_1 \times n_2}$. After some analysis, the reconstruction problem is formulated as low-rank Hankel matrix completion problem [1].

$$\text{find } \mathcal{H}\mathbf{z} \quad \text{subject to} \quad \text{rank}(\mathcal{H}\mathbf{z}) = r \quad \mathcal{P}_\Omega(\mathcal{H}\mathbf{z}) = \mathcal{P}_\Omega(\mathcal{H}\mathbf{x})$$

\mathcal{P}_Ω is the sampling operator defined as: $\mathcal{P}_\Omega(\mathbf{Z}) = \sum_{\alpha \in \Omega} \langle \mathbf{Z}, \mathbf{H}_\alpha \rangle \mathbf{H}_\alpha$ where Ω is the random set that consists of the sampled Hankel basis. For the case where r is not specified, one can consider the following nuclear norm minimization problem.

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}}{\text{minimize}} && \|\mathbf{X}\|_* \\ & \text{subject to} && \langle \mathbf{X}, \mathbf{H}_\alpha \rangle = \langle \mathbf{M}, \mathbf{H}_\alpha \rangle \quad \forall \alpha \in \Omega \end{aligned} \quad (5)$$

where $\mathbf{X} = \mathcal{H}\mathbf{z}$ and $\mathbf{M} = \mathcal{H}\mathbf{x}$. (5) is now in the form of a general matrix completion problem.

Signal Recovery Under Quadratic Measurements.. Given a signal $\mathbf{x} \in \mathbb{R}^n$, consider quadratic measurements of the form $\mathbf{a}_i = \langle \mathbf{x}, \mathbf{w}_i \rangle$ with some vector $\mathbf{w}_i \in \mathbb{R}^n$. Given that m random measurements are available, it is of interest to determine if the underlying signal can be recovered. Using the lifting idea in [19], the signal recovery problem can be written as the following nuclear

norm minimization problem.

$$\begin{aligned}
& \underset{\mathbf{X} \in \mathbb{R}^{n \times n}}{\text{minimize}} && \|\mathbf{X}\|_* \\
& \text{subject to} && \langle \mathbf{X}, \mathbf{W}_\alpha \rangle = \langle \mathbf{M}, \mathbf{W}_\alpha \rangle \quad \forall \alpha \in \Omega \\
& && \mathbf{X} = \mathbf{X}^T; \mathbf{X} \succeq 0
\end{aligned} \tag{6}$$

Above $\mathbf{W}_\alpha = \mathbf{w}_\alpha \mathbf{w}_\alpha^*$ and Ω is the random set that consists of the indices of the sampled vectors. In the case that \mathbf{a}_i are sampled independently and uniformly at random on the unit sphere, the above minimization problem is the well known PhaseLift problem [3]. One can consider a general case where the assumption is simply that the \mathbf{w}_i 's are structured and form a basis. The framework introduced in this paper allows, under certain conditions, to state results about the uniqueness of this general recovery problem.

Weighted Nuclear Norm minimization.. The usual assumption in matrix completion is uniform random measurements. In the case of general sampling models, it has been argued that the nuclear norm minimization is not a suitable approach [20]. An alternative which has been argued to promote more accurate low rank solutions [19, 21] is the weighted nuclear norm minimization [22, 20]. The general weighted nuclear norm minimization problem is given by

$$\begin{aligned}
& \text{minimize} && \|\mathbf{R}\mathbf{X}\mathbf{C}\|_* \\
& \text{subject to} && \langle \mathbf{X}, \mathbf{w}_\alpha \rangle = \langle \mathbf{M}, \mathbf{w}_\alpha \rangle \quad \forall \alpha \in \Omega
\end{aligned} \tag{7}$$

where \mathbf{R} and \mathbf{C} , diagonal matrices with positive diagonal entries, scale the rows and columns of \mathbf{X} respectively. If the basis matrices $\{\mathbf{w}_\alpha\}$ are the canonical basis, the resulting problem is the standard matrix completion problem. In the context of the movie recommendation problem, the work in [20] demonstrates that the weighted nuclear minimization obtains better prediction accuracy than the unweighted nuclear norm minimization. For this case, the scaling matrices \mathbf{R} and \mathbf{C} reflect the non-uniform weight assumed to different users and different movies. Motivated by these considerations, we consider the general nuclear norm minimization problem where the basis matrices are sampled uniformly at

random but with a prior on uncertainty of the underlying matrix modeled by the scaling matrices \mathbf{R} and \mathbf{C} . Introducing the matrix $\mathbf{Y} = \mathbf{R}\mathbf{X}\mathbf{C}$, the above minimization problem can equivalently be rewritten as follows.

$$\begin{aligned} & \text{minimize} && \|\mathbf{Y}\|_* \\ & \text{subject to} && \langle \mathbf{Y}, \mathbf{R}^{-1}\mathbf{w}_\alpha\mathbf{C}^{-1} \rangle = \langle \overline{\mathbf{M}}, \mathbf{R}^{-1}\mathbf{w}_\alpha\mathbf{C} \rangle \quad \forall \alpha \in \Omega \end{aligned} \quad (8)$$

$\overline{\mathbf{M}}$ is the weighted ground matrix \mathbf{M} defined as $\overline{\mathbf{M}} = \mathbf{R}\mathbf{M}\mathbf{C}$. This is a general matrix completion problem with respect to the basis $\mathbf{R}^{-1}\mathbf{w}_\alpha\mathbf{C}^{-1}$. It can be verified that the set $\{\mathbf{R}^{-1}\mathbf{w}_\alpha\mathbf{C}^{-1}\}$ is a basis. Of interest is the following question: What kind of choices for \mathbf{R} and \mathbf{C} lead to successful recovery algorithms?

Challenges. The minimization problem in (2) has random linear constraints which can be expressed as $\mathcal{L}(\mathbf{X})$ where \mathcal{L} is the appropriate linear operator. Using the work in [12], one approach to show uniqueness of the general matrix completion problem is to check if \mathcal{L} obeys the restricted isometry property (RIP) condition. Our basis are structured and deterministic and so in general the RIP condition does not hold. As an example, consider the nuclear norm minimization program in (4) for the Euclidean distance geometry problem. We choose 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. One can easily check that $\mathcal{L}(\mathbf{X}) = \mathbf{0}$ which shows that the RIP condition does not hold. The general matrix completion problem resembles the matrix completion problem with respect to a symmetric orthonormal basis first considered in [15]. However, the basis $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$ in the general matrix completion problem are not necessarily orthogonal. One solution is to employ any of the basis orthogonalization algorithms and consider the minimization problem in the new orthonormal basis. This solution however is not useful since the measurements can not be treated as independent in the new basis. As such, the lack of orthogonality mandates an alternative analysis to show that the general matrix completion problem admits a unique solution.

Related work. The compressive sensing (CS) problem considers the solution of an undetermined linear system $\mathcal{A}\mathbf{x} = \mathbf{b}$ with the prior that the underlying vec-

tor \mathbf{x} is sparse. A series of works [18, 23, 24, 25] have shown that if \mathcal{A} is well conditioned, expressed using for instance the restricted isometry property (RIP) or the coherence condition, $m = O(s \log n)$ measurements suffice to recover the sparse solution \mathbf{x} with very high probability. With the measurement vectors, rows of the matrix \mathcal{A} , sampled from some random distribution, the conventional CS problem is the isotropic CS problem where the covariance matrix of the measurements is proportional to the identity matrix. If the isotropy condition does not hold, the standard results are not directly applicable. With that, the anisotropic compressive sensing problem has been studied in [26, 27]. The general matrix completion problem with respect to a non-orthogonal basis can be seen as analogue of anisotropic compressive sensing. Another line of work that relates to ours is structured matrix completion where model priors are integrated in the underlying matrix to be recovered. In [28], the authors consider the problem of recovering a spectrally sparse signal given some time domain samples. They propose an algorithm “Enhanced Matrix Completion” that first transforms the data matrix into its enhanced form employing a Hankel structure and then applies matrix completion on the enhanced form. In [29], this approach was further generalized using a low-rank interpolation approach allowing it to be used for the recovery of a broad class of signals. In addition to theoretical results on signal recovery, the work therein shows strong empirical results that validate that the structured matrix completion model outperforms conventional CS based signal recovery algorithms. A related problem is [15] where the author generalizes the standard matrix completion problem to any symmetric orthonormal basis.

The main technical difference in this paper compared to the aforementioned previous works is that the sampling operator, central to the analysis of the matrix completion problem, is not self adjoint. In particular, a common step in the analysis of matrix completion algorithms is the isometry condition which considers a spectral norm bound on $\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega} \mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}$ where \mathcal{P} is a projection on a certain subspace \mathbb{T} and \mathcal{R}_{Ω} is the sampling operator. In the general matrix

completion setting, this spectral bound does not hold as the sampling operator is no longer self adjoint. A natural alternative is a spectral norm bound on $\mathcal{P}_{\mathbb{T}}\mathcal{R}_{\Omega}^*\mathcal{R}_{\Omega}\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}$. It turns out that standard concentration inequality based arguments obtain suboptimal probability of success for this bound and can only be made optimal under very stringent conditions on the non-orthogonal basis. The main technical novelty of this paper is replacing this bound by a minimum eigenvalue bound on the operator $\lambda_{\min}(\mathcal{P}_{\mathbb{T}}\mathcal{F}_{\Omega}\mathcal{P}_{\mathbb{T}})$ where \mathcal{F}_{Ω} is a restricted frame operator. This idea along with the dual certificate approach [13] was employed, for a specific application of the distance geometry problem, in our previous work [17]. Much inspired by this work, we were interested in generalizing the result to any non-orthogonal basis. The current paper is a culmination of this effort. To our knowledge, the first work that uses a dual basis approach is the work in [30]. Therein, the authors study the compressive sensing problem in a general transform domain and provide theoretical guarantees. The problem we study in this paper can be seen as a direct analogue of this work for matrix completion. The techniques in [30] consider the CS setting. A direct translation of these results requires a suitable bound on $\|\mathcal{P}_{\mathbb{T}}\mathcal{R}_{\Omega}^*\mathcal{R}_{\Omega}\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\|$. As remarked earlier, this bound can not be realized for the general matrix completion problem and requires an alternative approach.

Contributions. In this paper, under suitable sampling conditions, a dual basis approach is used to show that the general matrix completion problem admits a unique solution. Introducing a dual basis to \mathbf{w}_{α} , denoted by \mathbf{z}_{α} , ensures that the measurements $\langle \mathbf{X}, \mathbf{w}_{\alpha} \rangle$ in (2) are compatible with expansion coefficients of \mathbf{M} . Based on the framework of the dual basis approach, we show that the minimization problem recovers the underlying matrix under suitable conditions. Two main contributions of this paper are as follows.

1. A dual basis approach is used to prove a uniqueness result for the general matrix completion problem. The main result shows that if the number of random measurements m is of order $O(nr \log^2 n)$, under certain assumptions, the nuclear norm minimization program recovers the underlying

low-rank solution with very high probability. A key part of our proof uses the operator Chernoff bound. This part of the proof based on the Chernoff bound is simple and might find use in other problems.

2. An important condition, named the correlation condition, is introduced. This condition determines whether the nuclear norm minimization program succeeds for a given general matrix completion problem. The correlation condition could hold for deterministic measurements and more importantly can be checked in polynomial time.

Outline. The outline of the paper is as follows. Section 2 introduces the correlation condition and discusses the dual basis approach. The proof of the main result is presented in section 3. The key components of the proof can be described as follows. The general matrix completion problem is a convex minimization problem for which a sufficient condition to optimality is the KKT condition. We will use a dual certificate \mathbf{Y} , satisfying certain conditions, to certify that the general matrix completion problem admits a unique minimizer. The construction of \mathbf{Y} follows the elegant golfing scheme proposed in [15]. The bulk of the theoretical work then focuses on using this scheme and proving that the conditions hold with very high probability. The implication of the proof is that there is a unique solution to the general matrix completion problem with very high probability. Section 4 concludes the work.

Notation. The notations used in the paper are summarized in Table 1.

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

Table 1: Notations

2. Matrix Completion Problem under a Non-orthogonal Basis

In this section, we introduce a new condition referred as a *correlation condition* for low-rank matrix completion in a non-orthogonal basis. We will also discuss a dual basis formulation which plays an important role in our main result.

2.1. Correlation Parameter

We consider an L dimensional subspace $\mathbb{S} \subset \mathbb{R}^{n \times n}$ as the feasible set of the general matrix completion. We allow $L \leq n^2$ such as the case where the feasible solutions naturally satisfy linear constraints. Given a complete set of unit-norm basis $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$ of the subspace \mathbb{S} , any $\mathbf{X} \in \mathbb{S}$ can be determined if one specifies all the measurements $\{\langle \mathbf{X}, \mathbf{w}_\alpha \rangle\}_{\alpha=1}^L$. The problem studied in this paper considers the case where we have random access to a few of these measurements. Leaving the precise notion of “a few” for later, we consider the following question: Are there non-orthogonal basis for which the matrix completion framework, learning from few measurements, still works? In this paper, we note that as long as a certain condition, named correlation condition, on the basis matrices is satisfied, the sample complexity of low rank matrix completion with respect to non-orthogonal basis is of the same order as sample complexity of low rank matrix completion with respect to an orthogonal basis. In the specific case of the general matrix completion problem, the goal is to rigorously show that, under certain conditions, a low rank matrix can be recovered from few random non-orthogonal measurements. The intuitive arguments above motivate the following correlation condition which loosely informs how far the non-orthogonal basis is from an orthonormal basis.

Definition 1. *The unit-norm basis $\{\mathbf{w}_\alpha \in \mathbb{R}^{n \times n}\}_{\alpha=1}^L$ has correlation parameter μ if there is a constant $\mu \geq 0$ such that the following two equations hold*

$$\left\| \frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^T \mathbf{w}_\alpha - \mathcal{I} \right\| \leq \mu \quad \& \quad \left\| \frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha \mathbf{w}_\alpha^T - \mathcal{I} \right\| \leq \mu \quad (9)$$

Intuitively, the above definition describes that the operators $\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^T \mathbf{w}_\alpha$ and $\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha \mathbf{w}_\alpha^T$ are nearly isometric to the identity operator. To understand

the correlation condition, we first establish certain properties and follow by computing the correlation condition of certain basis matrices.

Lemma 1. *[Properties and examples of correlation condition] Given a unit-norm basis $\{\mathbf{w}_\alpha \in \mathbb{R}^{n \times n}\}_{\alpha=1}^L$ with correlation parameter μ , the following statements hold:*

a. *The correlation parameter is bounded above by n , that is, $\mu \leq \max(1, \frac{L}{n})$.*

b. *If the correlation condition holds, it follows that*

$$\lambda_{\max} \left(\sum_{\alpha=1}^L \mathbf{w}_\alpha^T \mathbf{w}_\alpha \right) \leq (\mu + 1)n \quad \& \quad \lambda_{\max} \left(\sum_{\alpha=1}^L \mathbf{w}_\alpha \mathbf{w}_\alpha^T \right) \leq (\mu + 1)n$$

c. *If $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$ is an orthonormal basis $L = n^2$, the correlation condition holds with $\mu = 0$.*

d. *For the basis matrices in the Euclidean distance geometry problem, the correlation condition holds with $\mu = 1$.*

Proof.

1. For positive semidefinite matrices \mathbf{A} and \mathbf{B} , the norm inequality $\|\mathbf{A} - \mathbf{B}\| \leq \max(\|\mathbf{A}\|, \|\mathbf{B}\|)$ holds. Let $\mathbf{A} = \frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^T \mathbf{w}_\alpha$ and let $\mathbf{B} = \mathcal{I}$. Applying the triangle inequality, it follows that $\|\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^T \mathbf{w}_\alpha\| \leq \frac{1}{n} \sum_{\alpha} \|\mathbf{w}_\alpha^T\| \|\mathbf{w}_\alpha\| \leq \frac{L}{n}$. Therefore, $\|\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^T \mathbf{w}_\alpha - \mathcal{I}\| \leq \max(n, 1) = n$. An analogous argument obtains $\|\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha \mathbf{w}_\alpha^T - \mathcal{I}\| \leq \max(1, \frac{L}{n})$. Hence, $\mu \leq \max(1, \frac{L}{n})$ as desired. It should be remarked that the theoretical analysis requires $\mu = O(1)$.
2. Note that $\lambda_{\max} \left(\sum_{\alpha=1}^L \mathbf{w}_\alpha^T \mathbf{w}_\alpha \right) = \|\sum_{\alpha=1}^L \mathbf{w}_\alpha^T \mathbf{w}_\alpha - n\mathcal{I} + n\mathcal{I}\| \leq \|\sum_{\alpha=1}^L \mathbf{w}_\alpha^T \mathbf{w}_\alpha - n\mathcal{I}\| + \|n\mathcal{I}\| = (\mu + 1)n$. A similar argument results $\lambda_{\max} \left(\sum_{\alpha=1}^L \mathbf{w}_\alpha \mathbf{w}_\alpha^T \right) \leq (\mu + 1)n$.
3. For an orthonormal basis in $\mathbb{S} = \mathbb{R}^{n \times n}$, the completeness relation states that $\sum_{\alpha=1}^{n^2} \mathbf{w}_\alpha \mathbf{w}_\alpha^T = \sum_{\alpha=1}^{n^2} \mathbf{w}_\alpha^T \mathbf{w}_\alpha = n\mathcal{I}$. The proof of this fact is standard. For ease of reference, a proof is included in Lemma 6. Using the completeness relation, it follows that the correlation condition holds with $\mu = 0$.
4. For the Euclidean distance geometry problem, the basis are symmetric so it suffices to consider $\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^2$. A short calculation results $\frac{1}{n} \sum_{\alpha} \mathbf{w}_\alpha^2 = \frac{1}{2n}(n\mathcal{I} - \mathbf{1}\mathbf{1}^T)$. The correlation condition bound amounts to finding the operator norm of $\frac{1}{2}\|\mathcal{I} + \frac{1}{n}\mathbf{1}\mathbf{1}^T\|$. Since the maximum eigenvalue of $\frac{1}{n}\mathbf{1}\mathbf{1}^T$ is 1, $\frac{1}{2}\|\mathcal{I} + \frac{1}{n}\mathbf{1}\mathbf{1}^T\| = 1$. Hence, for the Euclidean distance geometry problem, the correlation condition holds with $\mu = 1$.

□

An advantage of the correlation condition is its computational complexity. The computational complexity of checking the correlation condition is at most $O(n^3)$. As such, given a measurement operator, it can easily be checked whether it satisfies the correlation condition or not. If the result is positive, the analysis in this paper can be applied. If the correlation condition does not hold, no conclusion can be made.

2.2. Dual basis formulation

For the general matrix completion problem in (2), the measurements are in the form $\langle \mathbf{X}, \mathbf{w}_\alpha \rangle$. If \mathbf{w}_α is orthonormal, \mathbf{X} can be expanded as $\mathbf{X} = \sum_\alpha \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{w}_\alpha$ where $\langle \mathbf{X}, \mathbf{w}_\alpha \rangle$ are the expansion coefficients. However, for non-orthogonal basis, this expansion does not hold. Since the measurements are inherent to the problem, the ideal expansion will be of the form $\mathbf{X} = \sum_\alpha \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{z}_\alpha$. The idea of the dual basis approach is to realize this form with the implication that the expansion coefficients match the random measurements. This approach is briefly summarized below and we refer the interested reader to [30] where the dual basis approach is first considered in the context of the general compressive sensing problem. The work in [17] uses the dual basis approach in the context of the context of matrix completion problem in the context of the Euclidean distance geometry problem. Given the basis $\{\mathbf{w}_\alpha\}_{\alpha=1}^L$ of \mathbb{S} , we define the matrix \mathbf{H} as $\mathbf{H}_{\alpha,\beta} = \langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle$ and write $\mathbf{H}_{\alpha,\beta}^{-1}$ as $\mathbf{H}^{\alpha,\beta}$. After minor analysis, it is straightforward to check $\mathbf{z}_\alpha = \sum_\beta \mathbf{H}^{\alpha,\beta} \mathbf{w}_\beta$ is a dual basis satisfying $\langle \mathbf{z}_\alpha, \mathbf{w}_\beta \rangle = \delta_{\alpha,\beta}$. We note the following relations which will be used in later analysis, $\mathbf{H} = \mathbf{W}^T \mathbf{W}$, $\mathbf{H}^{-1} = \mathbf{Z}^T \mathbf{Z}$ and $\mathbf{Z} = \mathbf{W} \mathbf{H}^{-1}$ where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L]$ and $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_L]$ denote the matrix of vectorized basis matrices and vectorized dual basis matrices respectively. The sampling operator, a central operator in the analysis of the general matrix completion problem, is defined as follows.

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

The sampling model, for the set Ω , is uniform random with replacement from $\mathbb{I} = \{1, \dots, L\}$. The size of Ω is denoted by m and the scaling factor $\frac{L}{m}$ is simply for convenience of analysis. The adjoint operator of the sampling operator also appears in the analysis and has the following form.

$$\mathcal{R}_\Omega^* : \mathbf{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{z}_\alpha \rangle \mathbf{w}_\alpha \quad (11)$$

Using the sampling operator \mathcal{R}_Ω , we can write (2) as follows.

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{S}}{\text{minimize}} \quad \|\mathbf{X}\|_* \\ & \text{subject to} \quad \mathcal{R}_\Omega(\mathbf{X}) = \mathcal{R}_\Omega(\mathbf{M}) \end{aligned} \quad (12)$$

Another operator which appears in the analysis is the restricted frame operator defined as follows.

$$\mathcal{F}_\Omega : \mathbf{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{w}_\alpha \quad (13)$$

It can be readily verified that the restricted frame operator is self-adjoint and positive semidefinite.

Coherence.. One can not expect to have successful reconstruction for an arbitrary matrix \mathbf{M} , in particular, when \mathbf{M} has very few non-zero expansion coefficients. Thus, the notion of coherence is introduced in [13] to guarantee successful completion. Consider the singular value decomposition of $\mathbf{M} = \sum_{k=1}^r \lambda_k \mathbf{u}_k \mathbf{v}_k^T$. Let's write $\mathbb{U} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, $\mathbb{U}^\perp = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_n\}$ as the orthogonal complement of \mathbb{U} , $\mathbb{V} = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ and $\mathbb{V}^\perp = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ as the orthogonal complement of \mathbb{V} . Projections onto these spaces appear frequently in the analysis and can be summarized as follows. Let $\mathcal{P}_\mathbb{U}$ and $\mathcal{P}_\mathbb{V}$ denote the orthogonal projections onto \mathbb{U} and \mathbb{V} respectively. $\mathcal{P}_{\mathbb{U}^\perp}$ and $\mathcal{P}_{\mathbb{V}^\perp}$ are defined analogously. Define $\mathbb{T} = \{\mathbf{U}\Theta^T + \Theta\mathbf{V}^T : \Theta \in \mathbb{R}^{n \times r}\}$ to be the tangent space of the rank r matrix in $\mathbb{R}^{n \times n}$ at \mathbf{M} . The orthogonal projection onto \mathbb{T} is given by

$$\mathcal{P}_\mathbb{T} \mathbf{X} = \mathcal{P}_\mathbb{U} \mathbf{X} + \mathbf{X} \mathcal{P}_\mathbb{V} - \mathcal{P}_\mathbb{U} \mathbf{X} \mathcal{P}_\mathbb{V} \quad (14)$$

It then follows that $\mathcal{P}_{\mathbb{T}^\perp} \mathbf{X} = \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X} = \mathcal{P}_{\mathbb{U}^\perp} \mathbf{X} \mathcal{P}_{\mathbb{V}^\perp}$. We define a coherence condition as follows.

Definition 2. *The aforementioned rank r matrix $M \in \mathbb{R}^{n \times n}$ has coherence ν with respect to unit-norm basis $\{\mathbf{w}_\alpha\}_{\alpha \in \mathbb{I}}$ if the following estimates hold*

$$\max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle^2 \leq \nu \frac{r}{n} \quad (15)$$

$$\max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{z}_\alpha, \mathbf{w}_\beta \rangle^2 \leq c_v \nu \frac{r}{n} \quad (16)$$

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

where $\{\mathbf{z}_\alpha\}_{\alpha \in \mathbb{I}}$ is the dual basis of $\{\mathbf{w}_\alpha\}_{\alpha \in \mathbb{I}}$ and c_v is a constant satisfying $c_v \geq \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty$.

Remark 1. *Since the above coherence conditions are a central part of the analysis, we consider equivalent simplified forms. We start with a bound on $\|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2$ using (15) and Lemma 8 which results*

$$\lambda_{\min}(\mathbf{H}) \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \max_{\alpha \in \mathbb{I}} \sum_{\beta \in \mathbb{I}} \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle^2 \leq \nu \frac{r}{n} \implies \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \lambda_{\max}(\mathbf{H}^{-1}) \nu \frac{r}{n}$$

Next, using the above inequality, the following bound for $\|\mathcal{P}_{\mathbb{T}} \mathbf{z}_\alpha\|_F$ follows.

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{z}_\alpha\|_F \leq \sum_{\beta \in \mathbb{I}} \|\mathbf{H}^{\alpha, \beta} \mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F = \sum_{\beta \in \mathbb{I}} |\mathbf{H}^{\alpha, \beta}| \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\beta\|_F \leq \|\mathbf{H}^{-1}\|_\infty \sqrt{\lambda_{\max}(\mathbf{H}^{-1})} \sqrt{\frac{\nu r}{n}}$$

It should be noted that the analysis presented in this paper requires that $\|\mathbf{H}^{-1}\|_\infty$ is at most $O(1)$. Finally, we use the previous inequality and (17) to derive a bound for $\langle \mathbf{z}_\alpha, \mathbf{U} \mathbf{V}^T \rangle^2$.

$$\begin{aligned} \langle \mathbf{z}_\alpha, \mathbf{U} \mathbf{V}^T \rangle^2 &= \left\langle \sum_{\beta \in \mathbb{I}} \mathbf{H}^{\alpha, \beta} \mathbf{w}_\beta, \mathbf{U} \mathbf{V}^T \right\rangle^2 \leq \max_{\beta \in \mathbb{I}} \langle \mathbf{w}_\beta, \mathbf{U} \mathbf{V}^T \rangle^2 \left(\sum_{\beta \in \mathbb{I}} |\mathbf{H}^{\alpha, \beta}| \right)^2 \\ &\leq \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu r}{n^2} \end{aligned}$$

The coherence conditions can now be summarized as follows.

$$\max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha\|_F^2 \leq \lambda_{\max}(\mathbf{H}^{-1}) \nu \frac{r}{n} \quad (18)$$

$$\max_{\alpha \in \mathbb{I}} \|\mathcal{P}_{\mathbb{T}} \mathbf{z}_\alpha\|_F^2 \leq \|\mathbf{H}^{-1}\|_\infty^2 \lambda_{\max}(\mathbf{H}^{-1}) \nu \frac{r}{n} \quad (19)$$

$$\max_{\alpha \in \mathbb{I}} \langle \mathbf{z}_\alpha, \mathbf{U} \mathbf{V}^T \rangle^2 \leq \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu r}{n^2} \quad (20)$$

The coherence parameter is indicative of concentration of information in the ground truth matrix. If the underlying matrix has low coherence, each measurement is equally informative as the other. On the other hand, if a matrix has high coherence, it means that the information is concentrated on few measurements.

Sampling Model. For the general matrix completion problem, the basis matrices are sampled uniformly at random with replacement. The advantage of this model is that the sampling process is independent. This property is crucial since our analysis uses concentration inequalities for i.i.d matrix valued random variables. A disadvantage of this sampling process is that the same measurement could be repeated and the analysis needs to account for the number of duplicates.

3. Main Result and Proof

The main result of this paper shows that the nuclear norm minimization program for the general matrix completion problem in (12) recovers the underlying matrix with very high probability. A precise statement is stated in the theorem below.

Theorem 1. *Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be a matrix of rank r that obeys the coherence conditions (15), (16) and (17) with coherence ν and satisfies the correlation condition (9) with correlation parameter μ . Define C as follows: $C = \max\left(\lambda_{\max}(\mathbf{H}^{-1})^3, c_v, \frac{(\mu+1)\|\mathbf{H}^{-1}\|_{\infty}}{\min((\mu+1)\|\mathbf{H}^{-1}\|_{\infty}, \frac{1}{4})^2}\right)$ with parameter c_v from (16) and $\|\mathbf{H}^{-1}\|_{\infty} = O(1)$. Assume m measurements, $\{\langle \mathbf{M}, \mathbf{w}_{\alpha} \rangle\}_{\alpha \in \Omega}$, are sampled uniformly at random with replacement. For $\beta > 1$, if*

$$m \geq \log_2 \left(4\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \sqrt{r} \right) nr \left(48 \left[C\nu + \frac{n}{Lr} \right] \left[\beta \log(n) + \log \left(4 \log_2 \left(4\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \sqrt{r} \right) \right) \right] \right) \quad (21)$$

the solution to (12) is unique and equal to \mathbf{M} with probability at least $1 - n^{-\beta}$.

Since the general matrix completion problem in (12) is convex, the optimizer can be characterized using the KKT conditions. A compact and simple form of these conditions is derived in [13]. With this, a brief outline of the proof is

as follows. The proof is divided into two main parts. In the former part, we show that if the aforementioned optimality conditions hold, then \mathbf{M} is a unique solution to the minimization problem. The latter and main part of the proof is concerned with showing that, under certain assumptions, these conditions do hold with very high probability. The implication of this is that, for a suitable choice of m , \mathbf{M} is a unique solution for the general matrix completion problem.

Our proof adapts arguments from [15, 17]. Few remarks on the difference of our proof to the matrix completion proofs in [15, 16] and our previous work [17] are in order.

1. The operator \mathcal{R}_Ω is not self-adjoint. The main implication of this is that the operator $\mathcal{P}_\mathbb{T}\mathcal{R}_\Omega\mathcal{P}_\mathbb{T}$, an important operator in matrix completion analysis, is no longer isometric to $\mathcal{P}_\mathbb{T}$. It turns out the appropriate operator to consider is $\mathcal{P}_\mathbb{T}\mathcal{R}_\Omega^*\mathcal{R}_\Omega\mathcal{P}_\mathbb{T}$ where the goal is to show that this operator is nearly isometric to $\mathcal{P}_\mathbb{T}$. However, this approach is not amenable to simple analysis. In this work, the main argument is based on showing that the minimum eigenvalue of the operator $\mathcal{P}_\mathbb{T}\mathcal{F}_\Omega\mathcal{P}_\mathbb{T}$ is bounded away from zero with very high probability. To prove this fact, the operator Chernoff bound is employed. The interpretation of this bound is that, restricted to the space \mathbb{T} , the operator $\mathcal{P}_\mathbb{T}\mathcal{F}_\Omega\mathcal{P}_\mathbb{T}$ is full rank. If the measurement basis is orthogonal, $\mathcal{F}_\Omega = \mathcal{R}_\Omega$, the implication is that the operator $\mathcal{P}_\mathbb{T}\mathcal{R}_\Omega\mathcal{P}_\mathbb{T}$ on \mathbb{T} is invertible. With this, $\mathcal{P}_\mathbb{T}\mathcal{F}_\Omega\mathcal{P}_\mathbb{T}$ can be understood as the operator analogue of $\mathcal{P}_\mathbb{T}\mathcal{R}_\Omega\mathcal{P}_\mathbb{T}$ for non-orthogonal measurements.
2. The measurement basis is non-orthogonal. Since we use the dual basis approach, the spectrum of the matrices \mathbf{H} and \mathbf{H}^{-1} become important. However, since we do not work with a fixed basis, all the constants are unknown. This is particularly relevant and presents some challenge in the use of concentration inequalities and will be apparent in later analysis.

For the matrix completion problem, with measurement basis \mathbf{e}_{ij} , the theoretical lower bound of $O(nr\nu \log n)$ was established in [14]. Note that, if $C = O(1)$ and $\mu = O(1)$, Theorem 1 requires on the order of $nr\nu \log^2 n$ measurements which

is only $\log(n)$ factor away from the optimal lower bound. The order of Theorem 1 is also the same order as those used in [15, 16]. These works consider the low rank recovery problem with any orthogonal basis and the matrix completion problem respectively. Before the proof of main result, we illustrate Theorem 1 on two of the examples discussed in the introduction.

Euclidean Distance Geometry Problem: A matrix completion formulation and theoretical analysis of the Euclidean distance geometry problem appears in [17]. Using the existing analysis in [17], $L = \frac{n(n-1)}{2}$, $\lambda_{\max}(\mathbf{H}^{-1}) \leq 4$, $\lambda_{\min}(\mathbf{H}^{-1}) = \frac{1}{8n}$ and $\|\mathbf{H}^{-1}\|_{\infty} = 8$. The constant c_v , which satisfies $c_v \geq \lambda_{\max}(\mathbf{H}^{-1})\|\mathbf{H}^{-1}\|_{\infty}$, is set to 32. Using Lemma 1(d), the correlation condition for the Euclidean Distance Geometry Problem holds with $\mu = 1$. Using Theorem 1, it can be seen that the number of samples needed to recover the underlying low rank Gram matrix for the Euclidean distance geometry problem is $O(nr\nu \log^2 n)$.

Spectrally Sparse Signal Reconstruction: For simplicity, consider the case $n_1 = n_2 = n$. For this problem, the orthonormality of the Hankel basis implies that $\mathbf{H} = \mathcal{I}$. Therefore, $\lambda_{\max}(\mathbf{H}^{-1}) = \lambda_{\min}(\mathbf{H}^{-1}) = \|\mathbf{H}^{-1}\|_{\infty} = 1$ and $L = n^2$. The correlation condition holds trivially with $\mu = 0$. The number of samples needed to cover the underlying low rank matrix is $O(nr\nu \log^2 n)$.

Remark 2. *It should be remarked that the minimum number of samples noted in Theorem 1 can be lowered, the constants could be improved, if one is working with explicit basis. The analysis presented here is generic and does not assume specific structure of the basis. Where the latter is readily available, most inequalities appearing in the technical details can be tightened lowering the sample complexity. For instance, if the basis is orthonormal as in the problem of spectrally sparse signal construction, the analysis in [15] gives tight results. In general, for explicit basis with some structure, one can adopt the analysis in this paper and improve certain bounds.*

Now we return to the main proof. For ease, the proof is structured into several intermediate results. The starting result is Theorem 2 which shows that

if certain conditions hold, \mathbf{M} is a unique solution to (12).

Theorem 2. *For any $\mathbf{X} \in \mathbb{S}$, let $\Delta = \mathbf{X} - \mathbf{M}$ denote the deviation from the true low rank matrix \mathbf{M} . $\Delta_{\mathbb{T}}$ and $\Delta_{\mathbb{T}^\perp}$ denote the orthogonal projection of Δ to \mathbb{T} and \mathbb{T}^\perp respectively. For any given Ω with $|\Omega| = m$, the following two statements hold.*

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

$$\|\mathcal{P}_{\mathbb{T}} \mathbf{Y} - \text{Sgn } \mathbf{M}\|_F \leq \frac{1}{4} \sqrt{\frac{1}{2L} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})}} \quad \text{and} \quad \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}\| \leq \frac{1}{2} \quad (22)$$

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

Theorem 2(a) states that, for “large” $\Delta_{\mathbb{T}}$, any deviation from \mathbf{M} is not in the null space of the operator. Theorem 2(b) states that, for “small” $\Delta_{\mathbb{T}}$, deviations from \mathbf{M} increase the nuclear norm. The theorem at hand is deterministic and at this stage no assumptions are made on the construction of the set Ω . As long as the assumptions of the theorem are satisfied, the theorem will hold true. After proving the theorem, we proceed to argue that the conditions in the theorem hold with very high probability. This will require certain sampling conditions and a suitable choice of $m = |\Omega|$.

3.1. Proof of Theorem 2

Proof of Theorem 2(a). First, observe that $\|\mathcal{R}_\Omega \Delta\|_F = \|\mathcal{R}_\Omega \Delta_{\mathbb{T}} + \mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F \geq \|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F - \|\mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F$. Since we want to show that $\mathcal{R}_\Omega \Delta \neq 0$, the observation leads to considering a lower bound for $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F$ and an upper bound for $\|\mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F$. For any \mathbf{X} , $\|\mathcal{R}_\Omega \mathbf{X}\|_F^2$ can be bounded as follows.

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

where the last inequality uses the fact that $\langle \mathbf{z}_\alpha, \mathbf{z}_\beta \rangle = \mathbf{H}^{\alpha, \beta}$. The min-max theorem applied to the above equation results

$$\frac{L^2}{m^2} \lambda_{\min}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \|\mathcal{R}_\Omega \mathbf{X}\|_F^2 \leq \frac{L^2}{m^2} \lambda_{\max}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \quad (23)$$

Setting $\mathbf{X} = \Delta_{\mathbb{T}^\perp}$ and using the right inequality above, we obtain

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}^\perp}\|_F^2 \leq \frac{L^2}{m^2} \lambda_{\max}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \Delta_{\mathbb{T}^\perp}, \mathbf{w}_\alpha \rangle^2 \leq m \frac{L^2}{m^2} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\Delta_{\mathbb{T}^\perp}\|_F^2 \quad (24)$$

where the last inequality uses the fact that $\sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \lambda_{\max}(\mathbf{H}) \|\mathbf{X}\|_F^2$ (Lemma 8 in Appendix A) and the constant m bounds the maximum number of repetitions for any given measurement. Analogously, setting $\mathbf{X} = \Delta_{\mathbb{T}}$ and using the left inequality in (23), we obtain

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 \geq \frac{L^2}{m^2} \lambda_{\min}(\mathbf{H}^{-1}) \sum_{\alpha \in \Omega} \langle \Delta_{\mathbb{T}}, \mathbf{w}_\alpha \rangle^2 = \frac{L}{m} \lambda_{\min}(\mathbf{H}^{-1}) \langle \Delta_{\mathbb{T}}, \mathcal{F}_\Omega \Delta_{\mathbb{T}} \rangle \quad (25)$$

The next step considers the projection onto \mathbb{T} of the restricted frame operator and applies the min-max theorem resulting the following inequality.

$$\begin{aligned} \|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 &\geq \frac{L}{m} \lambda_{\min}(\mathbf{H}^{-1}) \langle \Delta_{\mathbb{T}}, \mathcal{F}_\Omega \Delta_{\mathbb{T}} \rangle = \frac{L}{m} \lambda_{\min}(\mathbf{H}^{-1}) \langle \Delta_{\mathbb{T}}, \mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}} \Delta_{\mathbb{T}} \rangle \\ &\geq \frac{L}{m} \lambda_{\min}(\mathbf{H}^{-1}) \lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) \|\Delta_{\mathbb{T}}\|_F^2 \end{aligned} \quad (26)$$

Above, the first equality follows since $\mathcal{P}_{\mathbb{T}}$ is self adjoint and evidently $\Delta_{\mathbb{T}} \in \mathbb{T}$. The inequality in (26) can be reduced further using the assumption $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{1}{2} \lambda_{\min}(\mathbf{H})$ in the theorem resulting

$$\|\mathcal{R}_\Omega \Delta_{\mathbb{T}}\|_F^2 > \frac{L}{2m} \lambda_{\min}(\mathbf{H}^{-1}) \lambda_{\min}(\mathbf{H}) \|\Delta_{\mathbb{T}}\|_F^2 \quad (27)$$

Finally, use the inequalities in (24) and (27) and the assumption in the theorem to show that $\|\mathcal{R}_\Omega \Delta\|_F > 0$ as follows.

$$\begin{aligned} \|\mathcal{R}_\Omega \Delta\|_F &> \sqrt{\frac{L}{2m} \lambda_{\min}(\mathbf{H}^{-1}) \lambda_{\min}(\mathbf{H})} \|\Delta_{\mathbb{T}}\|_F - \frac{L}{\sqrt{m}} \sqrt{\frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})}} \|\Delta_{\mathbb{T}^\perp}\|_F \\ &\geq \sqrt{\frac{L}{2m} \lambda_{\min}(\mathbf{H}^{-1}) \lambda_{\min}(\mathbf{H})} \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \right) \|\Delta_{\mathbb{T}^\perp}\|_F - \frac{L}{\sqrt{m}} \sqrt{\frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})}} \|\Delta_{\mathbb{T}^\perp}\|_F = 0 \end{aligned}$$

This concludes proof of Theorem 2(a). \square

Proof of Theorem 2(b). Let $\mathbf{X} = \mathbf{M} + \Delta$ be a feasible solution to (12) with the condition that $\|\Delta_{\mathbb{T}}\|_F < \sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\Delta_{\mathbb{T}^\perp}\|_F$ and $\Delta \in \ker \mathcal{R}_\Omega$. The goal now is to show that, for any \mathbf{X} that satisfies these assumptions, the nuclear norm minimization is violated meaning that $\|\mathbf{X}\|_* = \|\mathbf{M} + \Delta\|_* > \|\mathbf{M}\|_*$. The proof of this fact makes use of the dual certificate approach in [15]. The idea is to endow a certain object, named a dual certificate \mathbf{Y} , with certain

conditions so as to ensure that any \mathbf{X} satisfying the earlier made assumptions is not a solution to (12). It then becomes a task to construct the certificate which satisfies the preset conditions. For ease of later reference, we start with the former task reproducing a proof, with minor changes, in section 2E of [15]. First, using the duality of the spectral norm and the nuclear norm, note that there exists a $\mathbf{\Lambda} \in \mathbb{T}^\perp$ with $\|\mathbf{\Lambda}\| = 1$ such that $\langle \mathbf{\Lambda}, \mathcal{P}_{\mathbb{T}^\perp}(\mathbf{\Delta}) \rangle = \|\mathcal{P}_{\mathbb{T}^\perp}(\mathbf{\Delta})\|_*$. Second, using the characterization of the subgradient of the nuclear norm [31], $\partial\|\mathbf{M}\|_* = \{\text{Sgn } \mathbf{M} + \mathbf{\Gamma} \mid \mathbf{\Gamma} \in \mathbb{T}^\perp \text{ \& } \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{\Gamma}\| \leq 1\}$. With this, it can be readily verified that $\text{Sgn } \mathbf{M} + \mathbf{\Lambda}$ is a subgradient of $\|\mathbf{M}\|_*$ at \mathbf{M} . Mathematically, we have $\|\mathbf{M} + \mathbf{\Delta}\|_* \geq \|\mathbf{M}\|_* + \langle \text{Sgn } \mathbf{M} + \mathbf{\Lambda}, \mathbf{\Delta} \rangle$. Using the condition $\mathbf{Y} \in \text{range } \mathcal{R}_\Omega^*$ which implies $\langle \mathbf{Y}, \mathbf{\Delta} \rangle = 0$ in the previous inequality, we obtain

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

The third equality follows using the earlier choice of $\mathbf{\Lambda}$ and the fact that $\text{Sgn } \mathbf{M} \in \mathbb{T}$ (from Lemma 7 in Appendix A). Finally, we apply the assumptions of the theorem to the last equation above to obtain

$$\begin{aligned} \|\mathbf{M} + \mathbf{\Delta}\|_* &\geq \|\mathbf{M}\|_* + \langle \text{Sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \mathbf{Y}, \mathbf{\Delta}_{\mathbb{T}} \rangle + \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_* - \langle \mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}, \mathbf{\Delta}_{\mathbb{T}^\perp} \rangle \\ &\geq \|\mathbf{M}\|_* - \frac{1}{4} \sqrt{\frac{1}{2L}} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})} \|\mathbf{\Delta}_{\mathbb{T}}\|_F + \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_* - \frac{1}{2} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_* \\ &\geq \|\mathbf{M}\|_* - \frac{1}{4} \sqrt{\frac{1}{2L}} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})} \|\mathbf{\Delta}_{\mathbb{T}}\|_F + \frac{1}{2} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \\ &> \|\mathbf{M}\|_* - \frac{1}{4} \sqrt{\frac{1}{2L}} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})} \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \right) + \frac{1}{2} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \\ &= \|\mathbf{M}\|_* + \frac{1}{4} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \end{aligned}$$

It can be concluded that $\|\mathbf{M} + \mathbf{\Delta}\|_* > \|\mathbf{M}\|_*$ as desired. \square

Next, we state and prove a corollary which shows that \mathbf{M} is a unique solution to (12) if the deterministic assumptions of Theorem 2 hold.

Corollary 1. *If the conditions of Theorem 2 hold, \mathbf{M} is a unique solution to (12).*

Proof. Define $\mathbf{\Delta} = \mathbf{X} - \mathbf{M}$ for any $\mathbf{X} \in \mathbb{S}$. Using Theorem 2(a), $\mathcal{R}_\Omega \mathbf{\Delta} \neq \mathbf{0}$ if $\|\mathbf{\Delta}_{\mathbb{T}}\|_F^2 \geq \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \right)^2$. It then suffices to consider the case $\|\mathbf{\Delta}_{\mathbb{T}}\|_F^2 < \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\mathbf{\Delta}_{\mathbb{T}^\perp}\|_F \right)^2$ for $\mathbf{\Delta} \in \ker \mathcal{R}_\Omega$. For this case, using the proof of Theorem 2(b), $\|\mathbf{X}\|_* > \|\mathbf{M}\|_*$. Therefore, \mathbf{M} is the unique solution to (12). \square

3.2. Proof of Theorem 1

Using the corollary above, if the two conditions in Theorem 2 hold, it follows that \mathbf{M} is a unique solution to (12). The first condition in Theorem 2(a) is the assumption that $\lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_{\Omega} \mathcal{P}_{\mathbb{T}}) > \frac{1}{2} \lambda_{\min}(\mathbf{H})$. This will ensure that the minimum eigenvalue of the operator $\mathcal{P}_{\mathbb{T}} \mathcal{F}_{\Omega} \mathcal{P}_{\mathbb{T}}$ is bounded away from zero. Using the operator Chernoff bound in [32] restated below, Lemma 2 addresses the assumption.

Theorem 3 (Chernoff bound in [32]). *Consider a finite sequence $\{\mathcal{L}_k\}$ of independent, random, self-adjoint operators, acting on matrices in $\mathbb{R}^{n \times n}$, that satisfy*

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

Compute the minimum eigenvalue of the sum of the expectations,

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

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 2. *Consider the operator $\mathcal{P}_{\mathbb{T}} \mathcal{F}_{\Omega} \mathcal{P}_{\mathbb{T}} : \mathbb{T} \rightarrow \mathbb{T}$. With $\kappa = \frac{mn}{Lr}$, the following estimate holds.*

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

Proof. Recall the restricted frame operator $\mathcal{F}_{\Omega} = \sum_{\alpha \in \Omega} \frac{L}{m} \langle \mathbf{X}, \mathbf{w}_{\alpha} \rangle \mathbf{w}_{\alpha}$. For $\mathbf{X} \in \mathbb{T}$, $\mathcal{P}_{\mathbb{T}} \mathcal{F}_{\Omega} \mathcal{P}_{\mathbb{T}} \mathbf{X}$ can be equivalently represented as follows.

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

Let $\mathcal{L}_{\alpha} = \frac{L}{m} \langle \cdot, \mathcal{P}_{\mathbb{T}} \mathbf{w}_{\alpha} \rangle \mathcal{P}_{\mathbb{T}} \mathbf{w}_{\alpha}$ denote the operator in the summand. Since \mathcal{L}_{α} is positive semidefinite, the operator Chernoff bound can be used. The

bound requires estimate of an upper bound R of the spectrum norm of \mathcal{L}_α and $\mu_{\min} = \lambda_{\min}(\sum_{\beta \in \Omega} E[\mathcal{L}_\beta])$. First, we estimate R as follows.

$$\left\| \frac{L}{m} \langle \cdot, \mathcal{P}_\mathbb{T} \mathbf{w}_\alpha \rangle \mathcal{P}_\mathbb{T} \mathbf{w}_\alpha \right\| = \frac{L}{m} \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F^2 \leq \frac{L}{m} \lambda_{\max}(\mathbf{H}^{-1}) \frac{\nu r}{n}$$

The last inequality follows from the coherence estimate in (18). With this, set $R = \frac{L}{m} \lambda_{\max}(\mathbf{H}^{-1}) \frac{\nu r}{n}$. Next, we consider the estimate of μ_{\min} by first evaluating $\sum_{\beta \in \Omega} E[\mathcal{L}_\beta]$.

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

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

$$\langle \mathbf{X}, \sum_{\beta \in \Omega} E[\mathcal{L}_\beta](\mathbf{X}) \rangle = \sum_{\beta \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\beta \rangle^2 \geq \lambda_{\min}(\mathbf{H}) \|\mathbf{X}\|_F^2$$

with the last inequality following from Lemma 8 in Appendix A. The variational characterization of the minimum eigenvalue, along with the fact that $\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha]$ is a self-adjoint operator, implies that the minimum eigenvalue of $\sum_{\alpha \in \Omega} E[\mathcal{L}_\alpha]$ is at least $\lambda_{\min}(\mathbf{H})$. With this, set $\mu_{\min} = \lambda_{\min}(\mathbf{H})$. The final step is to apply the operator Chernoff bound with $R = \frac{L}{m} \lambda_{\max}(\mathbf{H}^{-1}) \frac{\nu r}{n}$ and $\mu_{\min} = \lambda_{\min}(\mathbf{H})$. Setting $\delta = \frac{1}{2}$, $\lambda_{\min}(\mathcal{P}_\mathbb{T} \mathcal{F}_\Omega \mathcal{P}_\mathbb{T}) > \frac{1}{2} \lambda_{\min}(\mathbf{H})$ with probability of failure at most p_1 given by

$$p_1 = n \exp \left(-\frac{\lambda_{\min}(\mathbf{H})^2 \kappa}{8\nu} \right)$$

This concludes the proof. \square

Lemma 2 shows that $\lambda_{\min}(\mathcal{P}_\mathbb{T} \mathcal{F}_\Omega \mathcal{P}_\mathbb{T}) > \frac{1}{2} \lambda_{\min}(\mathbf{H})$ holds with probability at least $1 - p_1$ where the probability of failure is at most $p_1 = n \exp \left(-\frac{\kappa}{8\nu} \right)$ with $\kappa = \frac{mn}{Lr}$.

In what follows, the conditions in Theorem 2(b) are analyzed. The statement there assumes the existence of a certain dual certificate \mathbf{Y} that satisfies the conditions in (22). In [15], David Gross devised a novel scheme, the golfing scheme, to construct the dual certificate \mathbf{Y} . Before showing the scheme, some notations are in order: 1) The random set Ω is partitioned into l batches. The i -th batch, denoted Ω_i , contains m_i elements with $\sum_{i=1}^l m_i = m$. 2) For a given batch, the

sampling operator can be defined as follows $\mathcal{R}_i = \frac{L}{m_i} \sum_{\alpha \in \Omega_i} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{z}_\alpha$. The inductive scheme is shown below.

$$\mathbf{Q}_0 = \text{sgn } \mathbf{M}, \quad \mathbf{Y}_i = \sum_{j=1}^i \mathcal{R}_j^* \mathbf{Q}_{j-1}, \quad \mathbf{Q}_i = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \mathbf{Y}_i, \quad i = 1, \dots, l \quad (28)$$

The main idea of the remaining analysis is to employ the golfing scheme and certify that the conditions in (22) hold with very high probability. In the analysis of the golfing scheme, the initial task is to show that the first condition in (22) holds. This 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 $\mathbf{X} \in \mathbb{S}$ and will be addressed in Lemma 3 to follow shortly. The proof of the lemma relies on the vector Bernstein inequality in [15]. We use a slightly modified version of this inequality which is stated below.

Theorem 4 (Vector Bernstein inequality). *Let $\mathbf{x}_1, \dots, \mathbf{x}_m$ be independent zero-mean vector valued random variables. Assume that $\max_i \|\mathbf{x}_i\|_2 \leq R$ and $\sum_i E[\|\mathbf{x}_i\|_2^2] \leq \sigma^2$. For any $t \leq \frac{\sigma^2}{R}$, the following estimate 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 3. *Given an arbitrary fixed $\mathbf{X} \in \mathbb{S}$, for $t \leq 1$ with $\kappa = \frac{mn}{Lr}$, the following estimate holds.*

$$\Pr(\|\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}\|_F \geq t \|\mathbf{X}\|_F) \leq \exp \left(-\frac{t^2 \kappa}{8(\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_{\infty} \nu + \frac{n}{Lr})} + \frac{1}{4} \right) \quad (29)$$

Proof. In what follows, with out loss of generality, we assume $\|\mathbf{X}\|_F = 1$. Using the dual basis expansion, $\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X}$ can be represented as follows.

$$\mathcal{P}_{\mathbb{T}} \mathcal{R}_{\Omega}^* \mathcal{P}_{\mathbb{T}} \mathbf{X} - \mathcal{P}_{\mathbb{T}} \mathbf{X} = \sum_{\alpha \in \Omega} \left[\frac{L}{m} \langle \mathcal{P}_{\mathbb{T}} \mathbf{X}, \mathbf{z}_\alpha \rangle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha - \frac{1}{m} \mathcal{P}_{\mathbb{T}} \mathbf{X} \right] \quad (30)$$

The summand, denoted \mathbf{Y}_α , can be written as $\mathbf{Y}_\alpha = \mathbf{X}_\alpha - E[\mathbf{X}_\alpha]$. Since $E[\mathbf{Y}_\alpha] = \mathbf{0}$, it satisfies the condition for the vector Bernstein inequality and we proceed to consider appropriate bounds for $\|\mathbf{Y}_\alpha\|_F$ and $E[\|\mathbf{Y}_\alpha\|_F^2]$. First, we bound $\|\mathbf{Y}_\alpha\|_F$

making use of the coherence conditions (18) and (19).

$$\begin{aligned}
\|\mathbf{Y}_\alpha\|_F &= \left\| \frac{L}{m} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle \mathcal{P}_\mathbb{T} \mathbf{w}_\alpha - \frac{1}{m} \mathcal{P}_\mathbb{T} \mathbf{X} \right\|_F \leq \left\| \frac{L}{m} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle \mathcal{P}_\mathbb{T} \mathbf{w}_\alpha \right\|_F + \left\| \frac{1}{m} \mathcal{P}_\mathbb{T} \mathbf{X} \right\|_F \\
&\leq \frac{L}{m} \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_\mathbb{T} \mathbf{z}_\alpha\|_F + \frac{1}{m} \\
&\leq \frac{1}{m} \left(\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1 \right)
\end{aligned}$$

Therefore, set $R = \frac{1}{m} \left(\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1 \right)$. To upper bound $E[\|\mathbf{Y}_\alpha\|_F^2]$, we start with the definition of $E[\|\mathbf{Y}_\alpha\|_F^2]$ and proceed as follows.

$$\begin{aligned}
E[\|\mathbf{Y}_\alpha\|_F^2] &= E \left[\frac{L^2}{m^2} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle^2 \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F^2 + \frac{1}{m^2} \|\mathcal{P}_\mathbb{T} \mathbf{X}\|_F^2 - \frac{2L}{m^2} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{w}_\alpha \rangle \right] \\
&= E \left[\frac{L^2}{m^2} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle^2 \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F^2 \right] + \frac{1}{m^2} \|\mathcal{P}_\mathbb{T} \mathbf{X}\|_F^2 - \frac{2}{m^2} \left\langle \sum_{\alpha \in \mathbb{I}} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{w}_\alpha \rangle \mathbf{z}_\alpha, \mathcal{P}_\mathbb{T} \mathbf{X} \right\rangle \\
&= E \left[\frac{L^2}{m^2} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle^2 \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F^2 \right] - \frac{1}{m^2} \|\mathcal{P}_\mathbb{T} \mathbf{X}\|_F^2 \\
&\leq \frac{L}{m^2} \max_{\alpha \in \mathbb{I}} \|\mathcal{P}_\mathbb{T} \mathbf{w}_\alpha\|_F^2 \sum_{\alpha \in \mathbb{I}} \langle \mathcal{P}_\mathbb{T} \mathbf{X}, \mathbf{z}_\alpha \rangle^2 + \frac{1}{m^2} \\
&\leq \frac{L}{m^2} \lambda_{\max}(\mathbf{H}^{-1})^2 \frac{\nu r}{n} + \frac{1}{m^2} \leq \frac{L}{m^2} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \frac{\nu r}{n} + \frac{1}{m^2}
\end{aligned}$$

Above, the second inequality results from the coherence conditions (18) and (19) and application of Lemma 8. With this, set $\sigma^2 = \frac{1}{m} \left(\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1 \right)$. To conclude the proof, we apply the vector Bernstein inequality with the specified R and σ . For $t \leq \frac{\sigma^2}{R} = 1$, with $\kappa = \frac{mn}{Lr}$, the following estimate holds.

$$\Pr(\|\mathcal{P}_\mathbb{T} \mathcal{R}_\Omega^* \mathcal{P}_\mathbb{T} \mathbf{X} - \mathcal{P}_\mathbb{T} \mathbf{X}\|_F \geq t) \leq \exp \left(- \frac{t^2 \kappa}{8 (\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu + \frac{n}{Lr})} + \frac{1}{4} \right) \quad (31)$$

□

Next, it will be argued that the golfing scheme (28) certifies the conditions in (22) with very high probability. In particular, we have the following lemma.

Lemma 4. \mathbf{Y}_l obtained from the golfing scheme (28) satisfies the conditions in

(22) with failure probability which is at most $p = \sum_{i=1}^l p_2(i) + p_3(i) + p_4(i)$ where

$$p_2(i) = \exp \left(- \frac{\kappa_i}{32 (\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu + \frac{n}{Lr})} + \frac{1}{4} \right), \quad p_3(i) = 2n \exp \left(\frac{-3 \min((\mu+1) \|\mathbf{H}^{-1}\|_\infty, \frac{1}{4})^2 \kappa_i}{8(\mu+1) \|\mathbf{H}^{-1}\|_\infty^2 \nu} \right)$$

and $p_4(i) = n^2 \exp \left(- \frac{3\kappa_i}{32 (c_v \nu + \frac{n}{Lr})} \right)$ with $k_i = \frac{m_i n}{Lr}$.

Proof. In what follows, we repeatedly make use of the fact that $\mathbf{Q}_i \in \mathbb{T}$ since $\text{sgn } \mathbf{M} \in \mathbb{T}$ (see Lemma 7 in Appendix A). The main idea for showing that the first condition in (22) holds relies on a recursive form of \mathbf{Q}_i which can be derived as follows.

$$\begin{aligned}
\mathbf{Q}_i &= \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \left(\sum_{j=1}^i \mathcal{R}_j^* \mathbf{Q}_{j-1} \right) = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \left(\sum_{j=1}^{i-1} \mathcal{R}_j^* \mathbf{Q}_{j-1} + \mathcal{R}_i^* \mathbf{Q}_{i-1} \right) \\
&= \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \sum_{j=1}^{i-1} \mathcal{R}_j^* \mathbf{Q}_{j-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} = \text{sgn } \mathbf{M} - \mathcal{P}_{\mathbb{T}} \mathbf{Y}_{i-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} \\
&= \mathbf{Q}_{i-1} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathbf{Q}_{i-1} = (\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathcal{P}_{\mathbb{T}}) \mathbf{Q}_{i-1}
\end{aligned} \tag{32}$$

The first condition in (22) is a bound on $\|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathcal{P}_{\mathbb{T}}) \mathbf{Q}_{i-1}\|_F = \|\mathbf{Q}_i\|_F$. Using Lemma 3 with $t_{2,i} = \frac{1}{2}$, $\|(\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}} \mathcal{R}_i^* \mathcal{P}_{\mathbb{T}}) \mathbf{Q}_{i-1}\| < t_{2,i} \|\mathbf{Q}_{i-1}\|_F$ holds with failure probability at most

$$p_2(i) = \exp \left(- \frac{\kappa_i}{32 (\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_{\infty} \nu + \frac{n}{Lr})} + \frac{1}{4} \right)$$

where $\kappa = \frac{m_i n}{Lr}$. Using the recursive formula in (32) repeatedly, $\|\mathbf{Q}_i\|_F$ can be upper bounded as follows.

$$\|\mathbf{Q}_i\|_F < \left(\prod_{k=1}^i t_{2,k} \right) \|\mathbf{Q}_0\|_F = 2^{-i} \sqrt{r} \tag{33}$$

Setting $l = \log_2 \left(4\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \sqrt{r} \right)$, the first condition in (22) is now satisfied. It can be concluded that, using a union bound on the failure probabilities $p_2(i)$, $\|\mathbf{Q}_l\|_F < \sqrt{r} 2^{-l} = \frac{1}{4} \sqrt{\frac{1}{2L} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})}}$ holds with failure probability that is at most $\sum_{i=1}^l p_2(i)$. This implies that the first condition in (22) also holds with the same failure probability.

Next, we consider the second condition in (22) which requires a bound on $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l\|$. First, note that $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}_l\| = \|\mathcal{P}_{\mathbb{T}^\perp} \sum_{j=1}^l \mathcal{R}_j^* \mathbf{Q}_{j-1}\| = \|\sum_{j=1}^l \mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\| \leq \sum_{j=1}^l \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\|$. As such, in what follows, the focus will be finding a suitable bound for $\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{Q}_{j-1}\|$. This will be analyzed in Lemma 5 in Appendix A. A key element in the proof of Lemma 5 is an assumption on the size of $\max_{\beta} |\langle \mathbf{Q}_i, \mathbf{z}_{\beta} \rangle|$. For ease of notation in further analysis, let $\eta(\mathbf{Q}_i)$ be defined as: $\eta(\mathbf{Q}_i) = \max_{\beta} |\langle \mathbf{Q}_i, \mathbf{z}_{\beta} \rangle|$. The assumption is that, at the i -th step of the golfing scheme, $\eta(\mathbf{Q}_i)^2 \leq \|\mathbf{H}^{-1}\|_{\infty}^2 \frac{\nu}{n^2} c_i^2$ where c_i^2 is an upper bound for $\|\mathbf{Q}_i\|_F^2$, $\|\mathbf{Q}_i\|_F^2 \leq c_i^2$. The idea is to argue that this assumption holds with very high probability. To show this, assume that $\eta(\mathbf{Q}_i) \leq t_{4,i}$ with failure probability

$p_4(i)$. Setting $t_{4,i} = \frac{1}{2}\eta(\mathbf{Q}_{i-1})$ and applying the inequality $\eta(\mathbf{Q}_i) \leq t_{4,i}$ recursively results

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

where the last inequality follows from the coherence estimate in (20). It can now be concluded that, noting (33), the inequality above ensures that $\eta(\mathbf{Q}_i)^2 \leq \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu}{n^2} c_i^2$ with $c_i = 2^{-i}\sqrt{r}$. The failure probability $p_4(i)$ follows from Lemma 9, noting that $\eta(\mathbf{Q}_i) = \eta(\mathbf{Q}_{i-1} - \mathcal{P}_\mathbb{T}\mathcal{R}_i^*\mathbf{Q}_{i-1})$, and is given by

$$p_4(i) = n^2 \exp\left(-\frac{3\kappa_j}{32\left(c_v\nu + \frac{n}{Lr}\right)}\right) \quad \forall i \in [1, l]$$

Having justified the assumption on the size of $\eta(\mathbf{Q}_i)$, a key part of Lemma 9, we now consider certifying the second condition in (22). Assume that $\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\mathbf{Q}_{j-1}\| < t_{3,j}c_{j-1}$, with $\|\mathbf{Q}_{j-1}\|_F \leq c_{j-1} = 2^{-(j-1)}$, holds with failure probability $p_3(j)$.

Fixing $t_{3,j} = \min\left(\frac{(\mu+1)\|\mathbf{H}^{-1}\|_\infty}{\sqrt{r}}, \frac{1}{4\sqrt{r}}\right)$, Lemma 5 gives

$$p_3(j) = 2n \exp\left(\frac{-3 \min\left((\mu+1)\|\mathbf{H}^{-1}\|_\infty, \frac{1}{4}\right)^2 \kappa_j}{8(\mu+1)\|\mathbf{H}^{-1}\|_\infty^2 \nu}\right)$$

where $\kappa_j = \frac{m_j n}{Lr}$. $\|\mathcal{P}_{\mathbb{T}^\perp}\mathbf{Y}_l\|$ can now be upper bounded as follows.

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

Applying the union bound over the failure probabilities, $\|\mathcal{P}_{\mathbb{T}^\perp}\mathbf{Y}_l\| < \frac{1}{2}$ holds true with failure probability which is at most $\sum_{j=1}^l [p_3(j) + p_4(j)]$. With the same failure probability, the second condition in (22) holds true. \square

Lemma 5 will be proved shortly using the Bernstein inequality in [33] which is restated below for convenience.

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

$$E[\mathbf{X}_i] = 0 \quad \text{and} \quad \|\mathbf{X}_i\| \leq R \quad \forall i$$

Let the matrix variance statistic of the sum σ^2 be defined as

$$\sigma^2 = \max\left(\left\|\sum_i E[\mathbf{X}_i^T \mathbf{X}_i]\right\|, \left\|\sum_i E[\mathbf{X}_i \mathbf{X}_i^T]\right\|\right)$$

For all $t \geq 0$,

$$\Pr \left[\left\| \sum_i \mathbf{X}_i \right\| > t \right] \leq \begin{cases} 2n \exp \left(-\frac{3t^2}{8\sigma^2} \right) & t \leq \frac{\sigma^2}{R} \\ 2n \exp \left(-\frac{3t}{8R} \right) & t \geq \frac{\sigma^2}{R} \end{cases} \quad (34)$$

Lemma 5. Consider a fixed matrix $\mathbf{G} \in \mathbb{T}$. Assume that $\max_\beta \langle \mathbf{G}, \mathbf{z}_\beta \rangle^2 \leq \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu}{n^2} c^2$ with c set as $\|\mathbf{G}\|_F^2 \leq c^2$. Then, with $\kappa_j = \frac{m_j n}{Lr}$, the following estimate holds for all $t \leq \frac{(\mu+1)\|\mathbf{H}^{-1}\|_\infty}{\sqrt{r}}$.

$$\Pr(\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G}\| \geq t c) \leq 2n \exp \left(-\frac{3t^2 \kappa_j r}{8(\mu+1)\|\mathbf{H}^{-1}\|_\infty^2 \nu} \right)$$

Proof of Lemma 5. Using the dual basis representation, $\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G} = \sum_{\alpha \in \Omega_j} \frac{L}{m_j} \langle \mathbf{G}, \mathbf{z}_\alpha \rangle \mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha$.

The summand, denoted \mathbf{X}_α , has zero expectation since $\mathbf{G} \in \mathbb{T}$. With this, the zero mean assumption for Bernstein inequality is satisfied. Next, we consider suitable estimates for R and σ^2 . The latter necessitates an estimate for $\max(\|E[\mathbf{X}_\alpha \mathbf{X}_\alpha^T]\|, \|E[\mathbf{X}_\alpha^T \mathbf{X}_\alpha]\|)$. First, we bound $\|E[\mathbf{X}_\alpha \mathbf{X}_\alpha^T]\|$. Since $\mathbf{X}_\alpha \mathbf{X}_\alpha^T = \frac{L^2}{m_j^2} \langle \mathbf{G}, \mathbf{z}_\alpha \rangle^2 (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T$, using Lemma 10 in Appendix A and the fact that $\mathbf{w}_\alpha \mathbf{w}_\alpha^T$ is positive semidefinite, $\|E[\mathbf{X}_\alpha \mathbf{X}_\alpha^T]\|$ can be upper bounded as follows.

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

Using the correlation condition, in particular Lemma 1(b), $\|\sum_{\alpha \in \mathbb{I}} \mathbf{w}_\alpha \mathbf{w}_\alpha^T\| \leq (\mu+1)n$, we obtain

$$\|E[\mathbf{X}_\alpha \mathbf{X}_\alpha^T]\| \leq \frac{L(\mu+1)n}{m_j^2} \max_{\alpha \in \mathbb{I}} \langle \mathbf{z}_\alpha, \mathbf{G} \rangle^2 \leq \frac{L(\mu+1)n}{m_j^2} \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu}{n^2} c^2$$

An analogous calculation and reasoning as above yields $\|E[\mathbf{X}_\alpha^T \mathbf{X}_\alpha]\| \leq \frac{L(\mu+1)n}{m_j^2} \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu}{n^2} c^2$. Therefore, using triangle inequality, set

$$\sigma^2 = \frac{L(\mu+1)}{m_j} \|\mathbf{H}^{-1}\|_\infty^2 \frac{\nu}{n} c^2$$

To complete the proof, it remains to estimate R .

$$\|\mathbf{X}_\alpha\| \leq \frac{L}{m_j} \max_{\alpha \in \mathbb{I}} \langle \mathbf{z}_\alpha, \mathbf{G} \rangle \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha\| \leq \frac{L}{m_j} \|\mathbf{H}^{-1}\|_\infty \frac{\sqrt{\nu}}{n} c \quad (35)$$

Two pertinent cases have to be considered. If $\nu \geq \frac{1}{r}$, $\|\mathbf{X}_\alpha\| \leq \frac{L}{m_j} \|\mathbf{H}^{-1}\|_\infty \frac{\nu\sqrt{r}}{n} c = R_1$ and if $\nu < \frac{1}{r}$, $\|\mathbf{X}_\alpha\| \leq \frac{L}{m_j} \|\mathbf{H}^{-1}\|_\infty \frac{\sqrt{\nu}}{n} c = R_2$. Note that $\frac{\sigma^2}{R_1} = \frac{(\mu+1)\|\mathbf{H}^{-1}\|_\infty c}{\sqrt{r}}$ and $\frac{\sigma^2}{R_2} = \frac{(\mu+1)\|\mathbf{H}^{-1}\|_\infty c}{\sqrt{r}}$. To conclude the proof, we apply the Bernstein inequality. An application of the Bernstein inequality results the following estimate.

$$\Pr(\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{R}_j^* \mathbf{G}\| \geq t) \leq 2n \exp\left(-\frac{3t^2 \kappa_j r}{8(\mu+1)\|\mathbf{H}^{-1}\|_\infty^2 \nu c^2}\right) \quad (36)$$

for all $t \leq \frac{(\mu+1)\|\mathbf{H}^{-1}\|_\infty c}{\sqrt{r}}$ with $\kappa_j = \frac{m_j n}{Lr}$. This concludes the proof of Lemma 5. \square

Next, it will be argued that \mathbf{M} is a unique solution to (12) with very high probability. The argument considers two separate cases based on comparing $\|\Delta_{\mathbb{T}}\|_F^2$ and $\|\Delta_{\mathbb{T}^\perp}\|_F^2$. This motivates us to define the following two sets: $\mathbb{S}_1 = \left\{ \mathbf{X} \in \mathbb{S} : \|\Delta_{\mathbb{T}}\|_F^2 \geq \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\Delta_{\mathbb{T}^\perp}\|_F \right)^2 \right\}$ and $\mathbb{S}_2 = \left\{ \mathbf{X} \in \mathbb{S} : \|\Delta_{\mathbb{T}}\|_F^2 < \left(\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H}^{-1})}{\lambda_{\min}(\mathbf{H}^{-1})} \|\Delta_{\mathbb{T}^\perp}\|_F \right)^2 \& \mathcal{R}_\Omega(\Delta) = \mathbf{0} \right\}$. With this, assuming that Ω is sampled uniformly at random with replacement, the two cases are as follows.

1. For all $\mathbf{X} \in \mathbb{S}_1$, set $|\Omega| = m$ “sufficiently large” such that $\Pr\left(\left\{ \Omega \subset \mathbb{I} \mid |\Omega| = m, \lambda_{\min}(\mathcal{P}_{\mathbb{T}} \mathcal{F}_\Omega \mathcal{P}_{\mathbb{T}}) > \frac{\lambda_{\min}(\mathbf{H})}{2} \right\} \geq 1 - p_1\right)$ based on Lemma 2. Therefore, all $\mathbf{X} \in \mathbb{S}_1$ are feasible solutions to (12) with probability at most p_1 from Theorem 2(a).
2. For all $\mathbf{X} \in \mathbb{S}_2$, we obtain

$$\Pr\left(\left\{ \Omega \subset \mathbb{I} \mid |\Omega| = m, \mathbf{Y} \in \text{range } \mathcal{R}_\Omega^* \& \|\mathcal{P}_{\mathbb{T}} \mathbf{Y} - \text{Sgn } \mathbf{M}\|_F \leq \frac{1}{4} \sqrt{\frac{1}{2L}} \frac{\lambda_{\min}(\mathbf{H}^{-1})}{\lambda_{\max}(\mathbf{H}^{-1})} \& \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{Y}\| \leq \frac{1}{2} \right\} \geq \epsilon\right) \geq$$

with $\epsilon = \sum_{i=1}^l [p_2(i) + p_3(i) + p_4(i)]$ by setting $|\Omega| = m$ “sufficiently large” based on Lemma 4. Then, the probability of all $\mathbf{X} \in \mathbb{S}_2$ being solutions to (12) is at most ϵ from Theorem 2(b).

Using the above two cases and employing the union bound, any $\mathbf{X} \in \mathbb{S}$ different from \mathbf{M} is a solution to the general matrix completion problem with

probability at most $p = p_1 + \sum_{i=1}^l [p_2(i) + p_3(i) + p_4(i)]$. In the arguments above, we have used the terms “sufficiently large”, “small probability” and “very high probability” without being precise. The goal now is to set everything explicit. First, define very high probability as a probability of at least $1 - n^{-\beta}$ for $\beta > 1$. Analogously, define small failure probability as a probability of at most $n^{-\beta}$ for some $\beta > 1$. To recover the underlying matrix with high probability, the idea is to carefully set the remaining free parameters m , l and m_i so that the $p \leq n^{-\beta}$ for $\beta > 1$. This necessitates revisiting all the failure probabilities in the analysis. p_1 , the first failure probability, is the probability that the condition in Theorem 2(a) does not hold. In the construction of the dual certificate \mathbf{Y} via the golfing scheme, three failure probabilities, $p_2(i)$, $p_3(i)$ and $p_4(i)$, $\forall i \in [1, l]$, appear. With this, all the failure probabilities are noted below.

$$p_1 = n \exp \left(-\frac{\lambda_{\min}(\mathbf{H})^2 \kappa}{8\nu} \right) \quad ; \quad p_2(i) = \exp \left(-\frac{\kappa_i}{32 (\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_{\infty} \nu + \frac{n}{Lr})} + \frac{1}{4} \right)$$

$$p_3(i) = 2n \exp \left(\frac{-3 \min((\mu+1) \|\mathbf{H}^{-1}\|_{\infty}, \frac{1}{4})^2 \kappa_i}{8(\mu+1) \|\mathbf{H}^{-1}\|_{\infty}^2 \nu} \right) \quad ; \quad p_4(i) = n^2 \exp \left(-\frac{3\kappa_i}{32 (c_v \nu + \frac{n}{Lr})} \right)$$

To suitably set m_i , since $k_i = \frac{m_i n}{Lr}$, we can set k_i with the condition that all the failure probabilities are at most $\frac{1}{4l} n^{-\beta}$ for $\beta > 1$. A minor calculation results one suitable choice of k_i , $k_i = 48(C\nu + \frac{1}{nr})(\beta \log(n) + \log(4l))$, with C defined as

$$C = \max \left(\lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_{\infty}, c_v, \frac{(\mu+1) \|\mathbf{H}^{-1}\|_{\infty}}{\min((\mu+1) \|\mathbf{H}^{-1}\|_{\infty}, \frac{1}{4})^2} \right) \quad (37)$$

The total failure probability, applying the union bound, is bounded above by $n^{-\beta}$. The number of measurements, $m = l r n k_i$, is at least

$$\log_2 \left(4\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \sqrt{r} \right) n r \left(48 \left[C\nu + \frac{n}{Lr} \right] \left[\beta \log(n) + \log \left(4 \log_2 \left(4\sqrt{2L} \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})} \sqrt{r} \right) \right) \right] \right) \quad (38)$$

This finishes the proof of Theorem 1. It can be concluded that the minimization program in (12) recovers the underlying matrix with very high probability.

4. Conclusion

In this paper, we study the problem of recovering a low rank matrix given a few of its expansion coefficients with respect to any basis. The considered problem generalizes existing analysis for the standard matrix completion problem and low rank recovery problem with respect to an orthonormal basis. The main analysis uses the dual basis approach and is based on dual certificates. An important assumption in the analysis is a proposed sufficient condition on the basis matrices named as the correlation condition. This condition can be checked in $O(n^3)$ computational time and holds in many cases of deterministic basis matrices. If this condition holds and the underlying low rank matrix obeys the coherence condition with parameter ν , under additional mild assumptions, our main result shows that the true matrix can be recovered with very high probability from $O(nr\nu\log^2 n)$ uniformly random sampled coefficients. Future research will consider a detailed analysis of the correlation condition and evaluating the effectiveness of the framework of the general matrix completion in certain applications.

5. Acknowledgment

Abiy Tasissa would like to thank Professor David Gross for correspondence over email regarding the work in [15]. Particularly, the proof of Lemma 10 is a personal communication from Professor David Gross. We also thank anonymous reviewers for their insightful suggestions and comments which led to an improved manuscript.

References

- [1] J.-F. Cai, T. Wang, K. Wei, Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion, *Applied and Computational Harmonic Analysis*.
- [2] M. Fazel, H. Hindi, S. P. Boyd, A rank minimization heuristic with application to minimum order system approximation, in: *American Control*

- Conference, 2001. Proceedings of the 2001, Vol. 6, IEEE, 2001, pp. 4734–4739.
- [3] E. J. Candes, T. Strohmer, V. Voroninski, Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming, *Communications on Pure and Applied Mathematics* 66 (8) (2013) 1241–1274.
 - [4] W. Glunt, T. Hayden, M. Raydan, Molecular conformations from distance matrices, *Journal of Computational Chemistry* 14 (1) (1993) 114–120.
 - [5] M. W. Trosset, Applications of multidimensional scaling to molecular conformation.
 - [6] X. Fang, K.-C. Toh, Using a distributed sdp approach to solve simulated protein molecular conformation problems, in: *Distance Geometry*, Springer, 2013, pp. 351–376.
 - [7] Y. Ding, N. Krislock, J. Qian, H. Wolkowicz, Sensor network localization, euclidean distance matrix completions, and graph realization, *Optimization and Engineering* 11 (1) (2010) 45–66.
 - [8] P. Biswas, T.-C. Lian, T.-C. Wang, Y. Ye, Semidefinite programming based algorithms for sensor network localization, *ACM Transactions on Sensor Networks (TOSN)* 2 (2) (2006) 188–220.
 - [9] J. B. Tenenbaum, V. De Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *science* 290 (5500) (2000) 2319–2323.
 - [10] V. Kalofolias, X. Bresson, M. Bronstein, P. Vandergheynst, Matrix completion on graphs, *arXiv preprint arXiv:1408.1717*.
 - [11] R. Lai, J. Li, Solving partial differential equations on manifolds from incomplete interpoint distance, *SIAM Journal on Scientific Computing* 39 (5) (2017) A2231–A2256.

- [12] B. Recht, M. Fazel, P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM review* 52 (3) (2010) 471–501.
- [13] E. J. Candès, B. Recht, Exact matrix completion via convex optimization, *Foundations of Computational mathematics* 9 (6) (2009) 717–772.
- [14] E. J. Candès, T. Tao, The power of convex relaxation: Near-optimal matrix completion, *IEEE Transactions on Information Theory* 56 (5) (2010) 2053–2080.
- [15] D. Gross, Recovering low-rank matrices from few coefficients in any basis, *Information Theory, IEEE Transactions on* 57 (3) (2011) 1548–1566.
- [16] B. Recht, A simpler approach to matrix completion, *The Journal of Machine Learning Research* 12 (2011) 3413–3430.
- [17] A. Tasissa, R. Lai, Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion, To appear, *IEEE Transaction on Information Theory*, arXiv preprint arXiv:1804.04310.
- [18] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on information theory* 52 (2) (2006) 489–509.
- [19] E. J. Candès, Y. C. Eldar, T. Strohmer, V. Voroninski, Phase retrieval via matrix completion, *SIAM review* 57 (2) (2015) 225–251.
- [20] N. Srebro, R. R. Salakhutdinov, Collaborative filtering in a non-uniform world: Learning with the weighted trace norm, in: *Advances in Neural Information Processing Systems*, 2010, pp. 2056–2064.
- [21] M. Fazel, H. Hindi, S. P. Boyd, Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices, in: *American Control Conference, 2003. Proceedings of the 2003, Vol. 3, IEEE*, 2003, pp. 2156–2162.

- [22] R. Foygel, O. Shamir, N. Srebro, R. R. Salakhutdinov, Learning with the weighted trace-norm under arbitrary sampling distributions, in: *Advances in Neural Information Processing Systems*, 2011, pp. 2133–2141.
- [23] E. J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE transactions on information theory* 52 (12) (2006) 5406–5425.
- [24] D. L. Donoho, Compressed sensing, *IEEE Transactions on information theory* 52 (4) (2006) 1289–1306.
- [25] E. J. Candes, Y. Plan, A probabilistic and ripless theory of compressed sensing, *IEEE Transactions on Information Theory* 57 (11) (2011) 7235–7254.
- [26] M. Rudelson, S. Zhou, Reconstruction from anisotropic random measurements, in: *Conference on Learning Theory*, 2012, pp. 10–1.
- [27] R. Kueng, D. Gross, Ripless compressed sensing from anisotropic measurements, *Linear Algebra and its Applications* 441 (2014) 110–123.
- [28] Y. Chen, Y. Chi, Robust spectral compressed sensing via structured matrix completion, *IEEE Transactions on Information Theory* 10 (60) (2014) 6576–6601.
- [29] J. C. Ye, J. M. Kim, K. H. Jin, K. Lee, Compressive sampling using annihilating filter-based low-rank interpolation, *IEEE Transactions on Information Theory* 63 (2) (2016) 777–801.
- [30] K. Lee, Y. Li, K. H. Jin, J. C. Ye, Unified theory for recovery of sparse signals in a general transform domain, *IEEE Transactions on Information Theory* 64 (8) (2018) 5457–5477.
- [31] G. A. Watson, Characterization of the subdifferential of some matrix norms, *Linear algebra and its applications* 170 (1992) 33–45.

- [32] J. A. Tropp, User-friendly tail bounds for sums of random matrices, Foundations of computational mathematics 12 (4) (2012) 389–434.
- [33] J. A. Tropp, et al., An introduction to matrix concentration inequalities, Foundations and Trends® in Machine Learning 8 (1-2) (2015) 1–230.

Appendix A. Appendix A

Lemma 6. *Given an orthonormal basis $\{\mathbf{C}_\alpha\}_{\alpha=1}^{L=n^2}$, we have $\sum_\alpha \mathbf{C}_\alpha^T \mathbf{C}_\alpha = \sum_\alpha \mathbf{C}_\alpha \mathbf{C}_\alpha^T = n\mathcal{I}$.*

Proof. Consider a matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ expanded in the orthonormal basis as $\mathbf{X} = \sum_\alpha \langle \mathbf{X}, \mathbf{C}_\alpha \rangle \mathbf{C}_\alpha$. With \mathbf{x} and \mathbf{c}_α as column-wise vectorized forms of \mathbf{X} and \mathbf{C}_α respectively, $\mathbf{x} = \sum_\alpha \langle \mathbf{x}, \mathbf{c}_\alpha \rangle \mathbf{c}_\alpha = (\sum_\alpha \mathbf{c}_\alpha \mathbf{c}_\alpha^T) \mathbf{x}$. This implies that $\sum_\alpha \mathbf{c}_\alpha \mathbf{c}_\alpha^T = \mathcal{I}_{n^2}$ where the subscript denotes the size of the identity matrix. This is the standard completeness relation. The implication of this relation is that $\sum_\alpha \mathbf{C}_\alpha(i, j) \mathbf{C}_\alpha(s, t) = \delta_{s,t}^{i,j}$. Next, consider $\sum_\alpha \mathbf{C}_\alpha \mathbf{C}_\alpha^T$. The (i, j) -th entry of this sum is given by

$$\left(\sum_\alpha \mathbf{C}_\alpha \mathbf{C}_\alpha^T \right)_{i,j} = \sum_\alpha \sum_{s=1}^n \mathbf{C}_\alpha(i, s) \mathbf{C}_\alpha(j, s) = \sum_{s=1}^n \sum_\alpha \mathbf{C}_\alpha(i, s) \mathbf{C}_\alpha(j, s) = \sum_{s=1}^n \delta_{j,s}^{i,s} = n\delta_{i,j}$$

It follows that $\sum_\alpha \mathbf{C}_\alpha \mathbf{C}_\alpha^T = n\mathcal{I}$. An analogous calculation results $\sum_\alpha \mathbf{C}_\alpha^T \mathbf{C}_\alpha = n\mathcal{I}$. □

Lemma 7. *If $\mathbf{X} \in \mathbb{T}$, $\text{Sgn } \mathbf{X} \in \mathbb{T}$.*

Proof. Consider the singular value decomposition of \mathbf{X} as $\mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T$. $\text{Sgn } \mathbf{X}$ is simply $\text{Sgn } \mathbf{X} = \mathbf{U} (\text{Sgn } \Sigma) \mathbf{V}^T = \mathbf{U} \mathbf{D} \mathbf{V}^T$ where \mathbf{D} is the diagonal matrix resulting from applying the sign function to Σ . Using this decomposition, we consider $\mathcal{P}_{\mathbb{T}^\perp} \text{sgn } \mathbf{X}$.

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

Above, the last step follows from the fact that $\mathbf{U}^T \mathbf{U} = \mathcal{I}$. It can be concluded that $\text{sgn } \mathbf{X} \in \mathbb{T}$. □

Lemma 8. *Given any $\mathbf{X} \in \mathbb{R}^{n \times n}$, the following norm inequalities hold.*

$$\lambda_{\min}(\mathbf{H}) \|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \lambda_{\max}(\mathbf{H}) \|\mathbf{X}\|_F^2 ; \quad \lambda_{\min}(\mathbf{H}^{-1}) \|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{z}_\alpha \rangle^2 \leq \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{X}\|_F^2$$

Proof. Vectorize the matrix \mathbf{X} and each dual basis \mathbf{z}_α . It follows that

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

Orthogonalize \mathbf{Z} with $\overline{\mathbf{Z}} = \mathbf{Z}(\sqrt{\mathbf{H}^{-1}})^{-1}$. Since $\sum_{\beta \in \mathbb{I}} \langle \mathbf{X}, \mathbf{z}_\beta \rangle^2 = \mathbf{x}^T \overline{\mathbf{Z}} \mathbf{H}^{-1} \overline{\mathbf{Z}}^T \mathbf{x}$, we obtain

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

The above result follows from a simple application of the min-max theorem. An analogous argument gives

$$\lambda_{\min}(\mathbf{H}) \|\mathbf{X}\|_F^2 \leq \sum_{\alpha \in \mathbb{I}} \langle \mathbf{X}, \mathbf{w}_\alpha \rangle^2 \leq \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{X}\|_F^2$$

This concludes the proof. \square

Lemma 9. Define $\eta(\mathbf{X}) = \max_{\beta \in \mathbb{I}} |\langle \mathbf{X}, \mathbf{z}_\beta \rangle|$. For a fixed \mathbf{X} in \mathbb{T} , with $\kappa_j = \frac{m_j n}{Lr}$, the following estimate holds for all $t \leq \eta(\mathbf{X})$.

$$\Pr(\max_{\beta \in \mathbb{I}} |\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{z}_\beta \rangle| \geq t) \leq n^2 \exp\left(-\frac{3t^2 \kappa_j}{8\eta(\mathbf{X})^2 (c_v \nu + \frac{n}{Lr})}\right) \quad (\text{A.1})$$

Proof. $\langle \mathcal{P}_{\mathbb{T}} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{z}_\beta \rangle$, for some β , can be represented in the dual basis as follows.

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

The summand, denoted Y_α , is of the form $X_\alpha - E[X_\alpha]$ and automatically satisfies $E[Y_\alpha] = 0$. Bernstein inequality can now be applied with appropriate bound on $|Y_\alpha|$ and $|E[Y_\alpha^2]|$. First, we bound $|Y_\alpha|$ making use of the coherence conditions (18) and (19).

$$\begin{aligned} |Y_\alpha| &= \left| \frac{L}{m_j} \langle \mathbf{X}, \mathbf{z}_\alpha \rangle \langle \mathcal{P}_{\mathbb{T}} \mathbf{w}_\alpha, \mathbf{z}_\beta \rangle - \frac{1}{m_j} \langle \mathbf{X}, \mathbf{z}_\beta \rangle \right| \leq \frac{L}{m_j} \eta(\mathbf{X}) \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \frac{\nu r}{n} + \frac{1}{m_j} \eta(\mathbf{X}) \\ &= \frac{1}{m_j} \eta(\mathbf{X}) \left(\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1 \right) \end{aligned}$$

To bound $E[Y_\alpha^2]$, noting that $E[Y_\alpha^2] = E[X_\alpha^2] - E[X_\alpha]^2$, it follows that

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

The last inequality results from the coherence condition in (16). To conclude the proof, we apply the Bernstein inequality with $|Y_\alpha| \leq R = \frac{1}{m_j} \eta(\mathbf{X}) \left(\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1 \right)$ and $\sigma^2 = \frac{1}{m_j} \eta(\mathbf{X})^2 \left(\frac{L}{n} c_v \nu r + 1 \right)$. With $k_j = \frac{m_j n}{Lr}$, for $t \leq \frac{\sigma^2}{R} = \eta(\mathbf{X})^2 \frac{\frac{L}{n} c_v \nu r + 1}{\frac{L}{n} \lambda_{\max}(\mathbf{H}^{-1}) \|\mathbf{H}^{-1}\|_\infty \nu r + 1} \geq \eta(\mathbf{X})$, it holds that

$$\Pr(|\langle \mathcal{P}_\mathbb{T} \mathcal{R}_j^* \mathbf{X} - \mathbf{X}, \mathbf{z}_\beta \rangle| \geq t) \leq \exp\left(-\frac{3t^2 k_j}{8\eta(\mathbf{X})^2 (c_v \nu + \frac{n}{Lr})}\right) \quad (\text{A.2})$$

Lemma 9 now follows from applying a union bound over all elements of the dual basis. \square

Lemma 10. *Let $\mathbf{c}_\alpha \geq 0$. Then, the following two inequalities hold.*

$$\begin{aligned}
\left\| \sum_{\alpha} \mathbf{c}_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T \right\| &\leq \left\| \sum_{\alpha} \mathbf{c}_\alpha \mathbf{w}_\alpha \mathbf{w}_\alpha^T \right\| \\
\left\| \sum_{\alpha} \mathbf{c}_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) \right\| &\leq \left\| \sum_{\alpha} \mathbf{c}_\alpha \mathbf{w}_\alpha^T \mathbf{w}_\alpha \right\|
\end{aligned}$$

Proof. We start with the first statement. Using the definition of $\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha$, $\left\| \sum_{\alpha} \mathbf{c}_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T \right\|$ can be written as follows.

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

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} , $\left\| \sum_{\alpha} \mathbf{c}_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T \right\|$ can be upper bounded as follows

$$\begin{aligned}
\left\| \sum_{\alpha} \mathbf{c}_\alpha (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha) (\mathcal{P}_{\mathbb{T}^\perp} \mathbf{w}_\alpha)^T \right\| &\leq \left\| \sum_{\alpha} \mathbf{c}_\alpha \mathbf{w}_\alpha \mathcal{P}_{V^\perp} \mathbf{w}_\alpha^T \right\| = \left\| \sum_{\alpha} \mathbf{c}_\alpha \mathbf{w}_\alpha (\mathbf{w}_\alpha^T - \mathcal{P}_V \mathbf{w}_\alpha^T) \right\| \\
&= \left\| \sum_{\alpha} [\mathbf{c}_\alpha \mathbf{w}_\alpha \mathbf{w}_\alpha^T - \mathbf{c}_\alpha \mathbf{w}_\alpha \mathcal{P}_V \mathbf{w}_\alpha^T] \right\|
\end{aligned}$$

where the first equality follows from the relation $\mathcal{P}_{\mathbf{V}^\perp} = \mathcal{I} - \mathcal{P}_{\mathbf{V}}$. Since $\mathbf{c}_\alpha \geq 0$, $\sum_\alpha \mathbf{c}_\alpha \mathbf{w}_\alpha \mathbf{w}_\alpha^T$ is positive semidefinite. Using the relation $\mathcal{P}_{\mathbf{V}}^2 = \mathcal{P}_{\mathbf{V}}$ and the assumption that $\mathbf{c}_\alpha \geq 0$, $\sum_\alpha \mathbf{c}_\alpha \mathbf{w}_\alpha \mathcal{P}_{\mathbf{V}} \mathbf{w}_\alpha^T = \sum_\alpha \mathbf{c}_\alpha \mathbf{w}_\alpha \mathcal{P}_{\mathbf{V}} \mathcal{P}_{\mathbf{V}} \mathbf{w}_\alpha^T$ is also positive semidefinite. A similar argument concludes that $\sum_\alpha \mathbf{c}_\alpha \mathbf{w}_\alpha \mathcal{P}_{\mathbf{V}^\perp} \mathbf{w}_\alpha^T$ is also positive semidefinite. Finally, using the norm inequality, $\|\mathbf{A} + \mathbf{B}\| \geq \max(\|\mathbf{A}\|, \|\mathbf{B}\|)$, for positive semidefinite matrices \mathbf{A} and \mathbf{B} , it can be seen the first statement holds. An analogous proof as above yields the second statement concluding the proof. \square