Noname manuscript No.
(will be inserted by the editor)

Low-rank Matrix Completion in a General Non-orthonormal Basis

Abiy Tasissa · Rongjie Lai

the date of receipt and acceptance should be inserted later

Abstract This paper considers theoretical analysis of recovering a low rank matrix given a few expansion coefficients with respect to any basis. The current approach generalizes the existing analysis for the standard matrix completion problem and the matrix completion problem with respect to a symmetric orthonormal basis. The analysis is based on dual certificates using a dual basis approach and does not assume the restricted isometry properrty (RIP). We introduce a condition on the basis called the correlation condition. This condition can be computed in time $O(n^3)$ and holds for many cases of deterministic basis where RIP might not hold or is NP hard to verify. If the correlation 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 expansion coefficients.

1 Introduction

The low-rank matrix recovery problem, the problem of recovering a low-rank matrix given incomplete linear observations, finds applications in many areas of interest such as computational chemistry [17, 31, 13], machine learning [28] and sensor localization [12, 2]. In addition, many interesting applied problems can be formulated as a low-rank recovery problem broadening the scope of applications of this framework. We mention a few examples. The work in [22] proposes a method to solve PDEs on data represented as incomplete distance [22] using low-rank regularization. //mention even broader applications// Solving system of quadratic equation, which is a common security protocol in cryptography [11], can be formulated as a low-rank matrix recovery problem. In recommender systems, users and items have some relation to each other implying that there is some structure to the underlying matrix. The work in [21] proposes a structured low rank-recovery framework via graphs which has been found to be effective for the recommender system problem. In [26], the authors propose the weighted trace-norm regularization and demonstrate that this regularizer works well with non-uniform sampling. The optimzation problem of minimizing the weighted trace-norm given some entries of a matrix can be written as low-rank recovery problem. The PhaseLift problem [8], the problem of recovering a signal given measurements with missing phase information, can be formulated can be modeled as a low-rank recovery problem. A further application is spectrally sparse signal reconstruction [3] using low rank regularization. The matrix completion problem, completing a low-rank matrix given a random sample of its entries, is an instance of this problem.

A natural framework to the low-rank recovery problem is rank minimization under linear constraints. However, this problem is NP hard [25] and thus motivates alternative solutions. A series of theoretical papers

This work is supported in part by NSF DMS–1522645 and an NSF Career Award DMS–1752934.

Abiy Tasissa
Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180
E-mail: tasisa@rpi.edu

Rongjie Lai
Department of Mathematics, Rensselaer Polytechnic Institute, Troy, NY 12180
E-mail: lair@rpi.edu

[6,10,18,24,25] showed that the NP hard rank minimization problem for the matrix completion problem can be obtained by solving the following convex nuclear norm minimization problem:

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

where $\|\boldsymbol{X}\|_*$ denotes the nuclear norm, the sum of the singular values of a matrix, and $\Omega \subset \{(i,j)|i,j=1,...,n\}$, $|\Omega| = m$, denotes a random set that consists of all 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 sampled measurements. The idea that the nuclear norm is a good approximation of the rank function was first argued in [14]. 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 [18] considers measurements with respect to a generic symmetric orthonormal basis and gives comparable theoretical guarantees. In particular, it shows that the true low rank matrix can be recovered with high probability from $O(nr\log^2(n))$ uniformly sampled measurements.

In this paper, the matrix completion problem with respect to any non-orthogonal basis is considered. Assume a general basis $\{\boldsymbol{w}_\alpha\}_{\alpha=1}^L$ each unit-norm, $L \leq n^2$, which spans $\mathbb{S}$, an L dimensional subspace of $\mathbb{R}^{n\times n}$. The nuclear norm minimization program for this general matrix completion problem is given by

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

We are interested in the following two problems.

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

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 some 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 [17,31,28,12,22]. Formally, consider a set of n points $\boldsymbol{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_n\}^T \in \mathbb{R}^r$. Let $\boldsymbol{D} = [d_{i,j}^2]$ denote the Euclidean distance matrix. The inner product matrix, also known as the Gram matrix, is defined as $\boldsymbol{X}_{i,j} = \langle \boldsymbol{p}_i\,,\boldsymbol{p}_j\rangle$. $\boldsymbol{X}$ is positive semidefinite matrix of rank r. A minor analysis reveals that $\boldsymbol{D}$ and $\boldsymbol{X}$ can be related in the following way: $\boldsymbol{D}_{i,j} = \boldsymbol{X}_{i,i} + \boldsymbol{X}_{j,j} - 2\boldsymbol{X}_{i,j}$. For $r \ll n$, $\boldsymbol{X}$ can be recovered using nuclear norm minimization.

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

The constraint $\boldsymbol{X}\cdot\mathbf{1} = \mathbf{0}$ fixes the translation ambiguity. The above minimization problem can be equivalently interpreted as a general matrix completion problem with respect to some operator basis $\boldsymbol{w}_\alpha$.

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

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

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 solving the convex l_1 minimization problem [7]. In [3], the authors consider the recovery of spectrally sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ of known order r where $r \ll n$. Let $\mathcal{H} : \mathbb{C}^n \to \mathbb{C}^{n_1 \times n_2}$ be the linear operator that maps a vector $\boldsymbol{z} \in \mathbb{C}^n$ to a Hankel matrix $\mathcal{H}\boldsymbol{z} \in \mathbb{C}^{n_1 \times n_2}$ with $n + 1 = n_1 + n_2$. Denote the orthonormal basis of $n_1 \times n_2$ Hankel matrices by $\{\boldsymbol{H}_\alpha\}_{\alpha=1}^{n_1 \times n_2}$. After some analysis, the reconstruction problem is formulated as low-rank Hankel matrix completion problem [3].

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

$\mathcal{P}_\Omega$ is the sampling operator defined as: $\mathcal{P}_\Omega(\boldsymbol{Z}) = \sum_{\alpha \in \Omega} \langle \boldsymbol{Z}, \boldsymbol{H}_\alpha \rangle \boldsymbol{H}_\alpha$ where Ω is the random set that consists of the sampled Hankel basis. For the case where r is unknown, one can consider the following nuclear norm minimization problem

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

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

Signal Recovery Under Quadratic Measurements Given a signal $\boldsymbol{x} \in \mathbb{R}^n$, consider quadratic measurements of the form $\boldsymbol{a}_i = \langle \boldsymbol{x}, \boldsymbol{w}_i \rangle$ where $\boldsymbol{w}_i \in \mathbb{R}^n$ is some vector. 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 [4], introduce the matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T$ which is a rank 1 positive semidefinite matrix. With this, the signal recovery problem can be written as the following nuclear norm minimization problem

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

Above $\boldsymbol{W}_\alpha = \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^*$ and Ω is the random set that consists of the indices of the sampled vectors. In the case that $\boldsymbol{a}_i$ are sampled independently and uniformly at random on the unit sphere, the above minimization problem is the well known PhaseLift problem[8]. One can consider a general case where the assumption is simply that the $\boldsymbol{w}_i$'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 [26]. An alternative is the weighted nuclear norm minimization [16,26] which promotes low rank [4,15]. With $\boldsymbol{D}$ as a weight matrix, the weighted nuclear norm minimization problem is given by

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

For simplicity, let $\boldsymbol{D}$ be a diagonal matrix. In this case, $\|\boldsymbol{D}\boldsymbol{X}\|$ can be interpreted as weighing certain rows of $\boldsymbol{X}$ more than others. Introducing the matrix $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{X}$, the above minimization problem can be rewritten as follows.

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

$\overline{\boldsymbol{M}}$ is the weighted ground matrix $\boldsymbol{M}$ defined as $\overline{\boldsymbol{M}} = \boldsymbol{D}\boldsymbol{M}$. This is a general matrix completion problem with respect to the basis $\boldsymbol{D}^{-1}\boldsymbol{w}_\alpha$. Since $\boldsymbol{D}$ is a diagonal, the set $\{\boldsymbol{D}^{-1}\boldsymbol{w}_\alpha\}$ is a basis. Of interest is the following question: What kind of choices for $\boldsymbol{D}$ lead to successful recovery algorithms?

Challenges The minimization problem in (2) has random linear constraints which can be expressed as $\mathcal{L}(\boldsymbol{X})$ where $\mathcal{L}$ is the appropriate linear operator. Using the work in [25], one approach to show uniqueness of the general matrix completion problem is to check if $\mathcal{L}$ obeys the restricted isometry property, RIP in short, 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 any $(i,j) \notin \Omega$ and construct a matrix $\boldsymbol{X}$ with $\boldsymbol{X}_{i,j} = \boldsymbol{X}_{j,i} = \boldsymbol{X}_{i,i} = \boldsymbol{X}_{j,j} = 1$ and zero everywhere else. One can easily check that $\mathcal{L}(\boldsymbol{X}) = \boldsymbol{0}$ 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 [18]. However, the basis $\{\boldsymbol{w}_\alpha\}_{\alpha=1}^{L}$ in the general matrix completion problem are not necessarily orthogonal. This has the implication that the measurements $\langle \boldsymbol{w}_\alpha, \boldsymbol{X}\rangle$ are not compatible with the expansion coefficients of $\boldsymbol{M}$. 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 algorithms fail to preserve the structure of the basis which means that one is no longer solving the the original problem (the measurements are now different). As such, the lack of orthogonality mandates an alternative analysis to show that the general matrix completion problem admits a unique solution. In this paper, the analysis is based on the dual certificate approach [6]. Our proof was inspired by the work of David Gross [18] where the author generalizes the matrix completion problem to any orthonormal basis. For the case of orthonormal measurement basis, the sampling operator, an important operator for the analysis of the matrix completion problem, is self-adjoint. The main technical challenge is that the sampling operator for the general matrix completion problem is not self-adjoint necessitating alternative analysis.

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 $\boldsymbol{w}_\alpha$, denoted by $\boldsymbol{v}_\alpha$, ensures that the measurements, $\langle \boldsymbol{X}, \boldsymbol{w}_\alpha\rangle$, in (2) are compatible with expansion coefficients of $\boldsymbol{M}$. Within the framework of the dual basis approach, we construct a proof that shows that the minimization problem recovers the underlying matrix. The two main contributions of this paper are as follows.

1. A dual basis approach is used to 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. It should be remarked that this part of the proof and the estimate using Chernoff bound is simple and might find use in broad settings.
2. Two important conditions are identified that determine whether the nuclear norm minimization program succeeds for a given general matrix completion problem. For the case of deterministic measurements, the well-known RIP condition might not hold. However, the two conditions could hold and more importantly can be checked in polynomial time.

Outline The outline of the paper is as follows. Section 2 introduces the correlation parameter 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. Namely, if one can show that there exists a dual certificate $\boldsymbol{Y}$ that satisfies certain conditions, it follows that the general matrix completion problem has a unique solution. The construction of $\boldsymbol{Y}$ follows the golfing scheme proposed in [18]. 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 gives conclusion and suggests future works.

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

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.

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

Table 1 Notations

2.1 Correlation Parameter

We consider an L dimensional subspace $\mathbb{S} \subset \mathbb{R}^{n\times n}$ as the feasible solutions of the general matrix completion. In general, we allow $L < n^2$ such as the case where the feasible solutions naturally satisfy linear constraints. Given a complete set of unit-norm basis $\{\boldsymbol{w}_\alpha\}_{\alpha=1}^L$ of the subspace $\mathbb{S}$, any $\boldsymbol{X} \in \mathbb{S}$ can be determined if one specifies all the measurements defined as $\langle \boldsymbol{X}, \boldsymbol{w}_\alpha \rangle$. The problem studied in this paper considers the case where we have access to a few of these random measurements. Leaving the precise notion of "a few" for latter, we consider the following two questions: 1) Is there an advantage if one is doing measurements with respect to an orthogonal basis as opposed to non-orthogonal basis? 2) Are there non-orthogonal basis for which the matrix completion framework, learning from few measurements, still works? The intuition is that, informally speaking, there is decoupling in orthogonality which means every additional measurement is informative. However, with non-orthogonal basis, the basis might be correlated and every additional measurement might not necessarily be informative. 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 certain conditions. The first is the correlation parameter which informs, loosely, how far the nonorthogonal basis is from an orthonormal basis. The precise definition is stated below.

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

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

Intuitively, the above definition says that the operators $\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha$ and $\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^T$ 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 $\{\boldsymbol{w}_\alpha\}_{n=1}^L$ with correlation parameter μ, the following statements hold:*

a. The correlation parameter is bounded above by n, that is, $\mu \leq n$.
b. If the correlation condition holds, it follows that

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

c. If $\{\boldsymbol{w}_\alpha\}_{n=1}^{L=n^2}$ is an orthonormal basis, 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 $\boldsymbol{A}$ and $\boldsymbol{B}$, the norm inequality $\|\boldsymbol{A} - \boldsymbol{B}\| \leq \max(\|\boldsymbol{A}\|, \|\boldsymbol{B}\|)$ holds. Let $\boldsymbol{A} = \frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha$ and let $\boldsymbol{B} = \mathcal{I}$. Using the fact that $\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha$ is positive semidefinite, it follows that $\left\|\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha\right\| \leq \mathrm{Trace}(\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha) \leq n$. Therefore, $\left\|\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha - \mathcal{I}\right\| \leq \max(n, 1) = n$. An analogous argument obtains $\left\|\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^T - \mathcal{I}\right\| \leq n$. Hence, $\mu \leq n$ as desired. It should be remarked that although μ is bounded above by n, the theoretical analysis requires $\mu = O(1)$.

2. Note that $\lambda_{\max}\left(\sum_{\alpha=1}^{L} \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha\right) = \|\sum_{\alpha=1}^{L} \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha - n\mathcal{I} + n\mathcal{I}\| \leq \|\sum_{\alpha=1}^{L} \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha - n\mathcal{I}\| + \|n\mathcal{I}\| = (\mu+1)n$. A similar argument results $\lambda_{\max}\left(\sum_{\alpha=1}^{L} \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^T\right) \leq (\mu+1)n$.
3. For an orthonormal basis, the completeness relation states that $\sum_{\alpha=1}^{n^2} \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^T = \sum_{\alpha=1}^{n^2} \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha = n\mathcal{I}$. 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 $\mu = 0$. //I think you need to assume $L = n^2$ in this case, plus the proof in lemma 6 seems not correct//
4. For the Euclidean distance geometry problem, the basis are symmetric so it suffices to consider $\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^2$. A short calculation results $\frac{1}{n}\sum_\alpha \boldsymbol{w}_\alpha^2 = \frac{1}{2n}(n\mathcal{I} - \mathbf{1}\mathbf{1}^T)$. 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 $\mu = 1$.

□

Comparison with RIP condition The restricted isometry property, RIP in short, was introduced first in [9] for the compressive sensing problem. If the measurement matrix in the compressive sensing problem, denoted by $\boldsymbol{A}$, obeys the RIP condition, it implies that the underlying sparse vector can be recovered by solving a convex l_1 minimization problem. The notion of the restricted isometry property can be extended to the matrix completion problem and its definition, which first appeared in [25], is restated below.

Definition 2. *For every integer r with $1 \leq r \leq n$, a linear map $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ obeys the RIP condition with isometry constant δ_r if*

$$(1-\delta_r)\|\boldsymbol{X}\|_F^2 \leq \|\mathcal{A}\boldsymbol{X}\|_2^2 \leq (1+\delta_r)\|\boldsymbol{X}\|_F^2$$

holds for all matrices $\boldsymbol{X}$ of rank at most r.

In particular, the analysis in [25] shows that if the RIP condition holds for the measurement operator of the general matrix completion problem, the underlying low rank matrix can be recovered by solving the convex nuclear norm minimization problem. With this, one could only consider certifying that the measurement operator satisfies RIP. For example, the RIP condition is satisfied with high probability for random measurement operators [25]. RIP condition could fail if one considers deterministic measurement operators. The intuition is that, for this case, one could construct matrices in a deterministic way which belong to the null space of $\mathcal{A}$ implying that RIP does not hold. As noted earlier, the RIP condition fails to hold for the EDG problem where the proof of this fact relied on constructing a counterexample, a sparse matrix $\boldsymbol{X}$, which violates the RIP condition. On the other hand, Lemma 1 (4) shows that the correlation condition holds for the EDG problem with correlation parameter $\mu = 1$. Additionally, one could construct counter examples that imply that the RIP condition does not hold for the standard matrix completion problem and the phase retrieval problem. Therefore, in certain deterministic settings where the RIP condition does not hold, an RPIless analysis based on correlation parameter can be employed. Another advantage of the correlation condition is its computational complexity. Given a measurement operator, checking whether RIP condition holds or not is hard. For example, for the compressive sensing problem//better to use results for the matrix completion problem//, it has been shown that [1] checking whether RIP holds or not is NP hard. To the best of the authors knowledge, there is no analogous work for the matrix RIP condition. However, using the relations of vector RIP to matrix RIP as detailed in [23], it can be anticipated that checking the matrix RIP is also NP hard. On the other hand, the computational complexity of checking the correlation condition is at most $O(n^3)$. As such, given a measurement operator, it can be checked whether it satisfies the correlation condition or not. If the result is positive, an RIPless analysis can be carried out as will be detailed in this paper. If the correlation condition does not hold, no conclusion can be made. Precise conditions under which the correlation condition holds require some analysis which is left for future work. The above comparison is meant to illustrate that, for the case of deterministic measurements, the correlation condition could be used and an RIPless analysis can be carried out for the general matrix completion problem.

2.2 Dual basis formulation

For the general matrix completion problem in (2), the measurements are in the form $\langle \boldsymbol{X}, \boldsymbol{w}_\alpha\rangle$. If $\boldsymbol{w}_\alpha$ is orthogonal, $\boldsymbol{X}$ can be expanded as $\boldsymbol{X} = \sum_\alpha \langle \boldsymbol{X}, \boldsymbol{w}_\alpha\rangle \boldsymbol{w}_\alpha$ where $\langle \boldsymbol{X}, \boldsymbol{w}_\alpha\rangle$ are the expansion coefficients. However,

for nonorthogonal basis, this expansion does not hold. Since the measurements are inherent to the problem, the ideal expansion will be of the form $\boldsymbol{X} = \sum_\alpha \langle \boldsymbol{X}\,, \boldsymbol{w}_\alpha \rangle \boldsymbol{v}_\alpha$. The idea of the dual basis approach is to realize this form which implies that the expansion coefficients match the random measurements. This approach is briefly summarized below and we refer the interested reader [27] where the dual basis approach is detailed in the context of the the Euclidean distance geometry problem. Given the basis $\{\boldsymbol{w}_\alpha\}_{\alpha=1}^L$ of $\mathbb{S}$, we define the matrix $\boldsymbol{H}$ as $\boldsymbol{H}_{\alpha,\,\beta} = \langle \boldsymbol{w}_\alpha\,, \boldsymbol{w}_\beta \rangle$ and write $\boldsymbol{H}^{-1}_{\alpha,\,\beta}$ as $\boldsymbol{H}^{\alpha,\,\beta}$. After minor analysis, it is straightforward to check $\boldsymbol{v}_\alpha = \sum_\beta \boldsymbol{H}^{\alpha,\,\beta} \boldsymbol{w}_\beta$ is a dual basis satisfying $\langle \boldsymbol{v}_\alpha\,, \boldsymbol{w}_\beta \rangle = \delta_{\alpha,\,\beta}$. We note the following relations which will be used in later analysis, $\boldsymbol{H} = \boldsymbol{W}^T\boldsymbol{W}$, $\boldsymbol{H}^{-1} = \boldsymbol{V}^T\boldsymbol{V}$ and $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{H}^{-1}$ where $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_L]$ and $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_L]$ 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.

$$\mathcal{R}_\Omega : \boldsymbol{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \boldsymbol{X}\,, \boldsymbol{w}_\alpha \rangle \boldsymbol{v}_\alpha \tag{10}$$

The sampling model, for the set Ω, is uniform random with replacement from $\mathbb{I} = \{1, \cdots, L\}$. The size of Ω is denoted by m and the scaling factor $\dfrac{L}{m}$ is simply for convenience of analysis. One useful property of the sampling operator, which can be easily checked, is that $\dfrac{m^2}{L^2}\mathcal{R}_\Omega^2 = \dfrac{m}{L}\mathcal{R}_\Omega$. The adjoint operator of the sampling operator also appears in the analysis and has the following form.

$$\mathcal{R}_\Omega^* : \boldsymbol{X} \in \mathbb{S} \longrightarrow \frac{L}{m} \sum_{\alpha \in \Omega} \langle \boldsymbol{X}\,, \boldsymbol{v}_\alpha \rangle \boldsymbol{w}_\alpha \tag{11}$$

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

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

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

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

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 $\boldsymbol{M}$, in particular, when $\boldsymbol{M}$ has very few non-zero coefficients. Thus, the notion of coherence is introduced in [6] to guarantee successful completion. Consider the singular value decomposition of $\boldsymbol{M} = \sum_{k=1}^r \lambda_k \boldsymbol{u}_k \boldsymbol{v}_k^T$. Let's write $\boldsymbol{U} = \text{span}\,\{\boldsymbol{u}_1, ..., \boldsymbol{u}_r\}$, $\boldsymbol{U}^\perp = \text{span}\{\boldsymbol{u}_{r+1}, ..., \boldsymbol{u}_n\}$ as the orthogonal complement of $\boldsymbol{U}$, $\boldsymbol{V} = \text{span}\,\{\boldsymbol{v}_1, ..., \boldsymbol{v}_r\}$ and $\boldsymbol{V}^\perp = \text{span}\{\boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_n\}$ as the orthogonal complement of $\boldsymbol{V}$. Projections onto these spaces appear frequently in the analysis and can be summarized as follows. Let $\mathcal{P}_{\boldsymbol{U}}$ and $\mathcal{P}_{\boldsymbol{V}}$ denote the orthogonal projections onto $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively. $\mathcal{P}_{\boldsymbol{U}^\perp}$ and $\mathcal{P}_{\boldsymbol{V}^\perp}$ are defined analogously. Define $\mathbb{T} = \{\boldsymbol{U}\boldsymbol{Z}^T + \boldsymbol{Z}\boldsymbol{V}^T : \boldsymbol{Z} \in \mathbb{R}^{n\times r}\}$ to be the tangent space of the rank r matrix in $\mathbb{R}^{n\times n}$ at $\boldsymbol{M}$. The orthogonal projection onto $\mathbb{T}$ is given by

$$\mathcal{P}_{\mathbb{T}}\boldsymbol{X} = \mathcal{P}_{\boldsymbol{U}}\boldsymbol{X} + \boldsymbol{X}\mathcal{P}_{\boldsymbol{V}} - \mathcal{P}_{\boldsymbol{U}}\boldsymbol{X}\mathcal{P}_{\boldsymbol{V}} \tag{14}$$

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

Definition 3. *The aforementioned rank r matrix $M \in \mathbb{R}^{n\times n}$ has coherence ν with respect to unit norm basis $\{\boldsymbol{w}_\alpha\}_{\alpha\in\mathbb{I}}$ if the following estimates hold*

$$\max_{\alpha\in\mathbb{I}} \sum_{\beta\in\mathbb{I}} \langle \mathcal{P}_{\mathbb{T}}\,\boldsymbol{w}_\alpha\,, \boldsymbol{w}_\beta \rangle^2 \le \nu \frac{r}{n} \tag{15}$$

$$\max_{\alpha\in\mathbb{I}} \sum_{\beta\in\mathbb{I}} \langle \mathcal{P}_{\mathbb{T}}\,\boldsymbol{v}_\alpha\,, \boldsymbol{w}_\beta \rangle^2 \le c_v \nu \frac{r}{n} \tag{16}$$

$$\max_{\alpha\in\mathbb{I}} \langle \boldsymbol{w}_\alpha\,, \boldsymbol{U}\boldsymbol{U}^T \rangle^2 \le \nu \frac{r}{n^2} \tag{17}$$

where $\{\boldsymbol{v}_\alpha\}_{\alpha\in\mathbb{I}}$ is the dual basis of $\{\boldsymbol{w}_\alpha\}_{\alpha\in\mathbb{I}}$ and c_v is a constant.

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* $\|\mathcal{P}_{\mathbb{T}}\, \boldsymbol{w}_\alpha\|_F^2$ *using* (15) *and Lemma 8 which results*

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

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

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

It should be noted that the analysis presented in this paper requires that $\|\boldsymbol{H}^{-1}\|_\infty$ *is at most* $O(1)$*. The parameter* c_v *in* (16) *is a constant that is set as* $c_v \geq \lambda_{\max}\boldsymbol{H}^{-1}\|\boldsymbol{H}^{-1}\|_\infty$*. Finally, we use the previous inequality and* (17) *to derive a bound for* $\langle \boldsymbol{v}_\alpha\, , \boldsymbol{U}\boldsymbol{U}^T\rangle^2$*.*

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

The coherence conditions can now be summarized as follows.

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

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

$$\max_{\alpha\in\mathbb{I}} \langle \boldsymbol{v}_\alpha\, , \boldsymbol{U}\boldsymbol{U}^T\rangle^2 \leq \|\boldsymbol{H}^{-1}\|_\infty^2\, \frac{\nu r}{n^2} \tag{20}$$

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.

Remark 2. *A previous version of our analysis considered uniform sampling with out replacement. The motivation was to avoid duplicate measurements. However, this choice also implies that the sampling is no longer independent. This will mean that concentration inequalities for i.i.d matrix valued random variables can not be used freely. In the work of Hoeffding [20], for Hoefdding inequality, it is argued that the results derived for the case of the sampling with replacement also hold true for the case of sampling without replacement. In [19], it is shown that matrix concentration inequalities resulting from the operator Chernoff bound technique [?] also hold true for uniform sampling with out replacement. These exceptions justfify the use of the uniform sampling with out replacement model in the analysis. However, the gain in terms of the number of measurements* m *is minimal. For sake of streamlined presentation, the uniform sampling with replacement is adopted in this paper.*

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. With out loss of generality and for ease of analysis, the theorem considers square matrices. The proof for rectangular matrices follows with minor modifications.

Theorem 1. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ 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: $C = \max\left(\lambda_{\max}(\boldsymbol{H}^{-1})^3, c_v, \frac{(\mu+1)\|\boldsymbol{H}^{-1}\|_\infty}{\min(\,(\mu+1)\|\boldsymbol{H}^{-1}\|_\infty, \frac{1}{4})^2}\right)$. Assume m measurements, $\{\langle \boldsymbol{M}, \boldsymbol{w}_\alpha\rangle\}_{\alpha\in\Omega}$, are sampled uniformly at random with replacement. For $\beta > 1$, if*

$$m \geq \log_2\left(4\sqrt{2L}\frac{\lambda_{\max}(\boldsymbol{H})}{\lambda_{\min}(\boldsymbol{H})}\sqrt{r}\right) nr\left(48\big[C\nu + \frac{1}{nr}\big]\big[\,\beta\log(n) + \log\big(4\log_2(4L\sqrt{r})\big)\,\big]\right) \tag{21}$$

the solution to (12) *is unique and equal to $\boldsymbol{M}$ 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 [6]. 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 $\boldsymbol{M}$ 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, $\boldsymbol{M}$ is a unique solution for the general matrix completion problem.

Our proof adapts arguments from [18]. Few remarks on the difference of our proof to the matrix completion proofs in [18,24] are in order.

1. The operator $\mathcal{R}_\Omega$ 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 right 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 simpler 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 $\boldsymbol{H}$ and $\boldsymbol{H}^{-1}$ 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 $\boldsymbol{e}_{ij}$, the optimal lower bound of $O(nr\nu\log n)$ was established in [10]. 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 6 is also the same order as [18,24]. 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 some 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 [27]. Using the existing analysis in [27], $L = \frac{n(n-1)}{2}$, $\lambda_{\max}(\boldsymbol{H}^{-1}) \leq 4$, $\lambda_{\min}(\boldsymbol{H}^{-1}) = \frac{1}{8n}$ and $\|\boldsymbol{H}^{-1}\|_\infty = 8$. The constant c_v, which satisfies $c_v \geq \lambda_{\max}(\boldsymbol{H}^{-1})\|\boldsymbol{H}^{-1}\|_\infty$, 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 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 orthonormal of the Hankel basis implies that the $\boldsymbol{H} = \mathcal{I}$. Therefore, $\lambda_{\max}(\boldsymbol{H}^{-1}) = \lambda_{\min}(\boldsymbol{H}^{-1}) = \|\boldsymbol{H}^{-1}\|_\infty = 1$ 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 $O(nr\nu\log^2 n)$.

Weighted Nuclear Norm Minimization: As discussed earlier, the weight matrix $\boldsymbol{D}$ is diagonal. For simplicity, assume that $\sum_{i=1}^n \boldsymbol{D}_{i,i} = 1$. In this case, the size of a diagonal entry informs the proportion of weight

assigned to the corresponding row of the true matrix. For the main result in Theorem 1 to hold, certain assumptions are necessary. First, the basis $\{\boldsymbol{D}^{-1}\boldsymbol{w}_\alpha\}_{\alpha=1}^L$ needs to satisfy the correlation condition. Second, the size of the maximum and minimum eigenvalues of the matrix $\bar{\boldsymbol{H}} = (\boldsymbol{D}^{-1}\boldsymbol{W})^T(\boldsymbol{D}^{-1}\boldsymbol{W}) = \boldsymbol{W}^T(\boldsymbol{D}^{-1})^2\boldsymbol{W}$ is important. After minor analysis, with $\boldsymbol{H} = \boldsymbol{W}^T\boldsymbol{W}$, $\lambda_{\max}(\bar{\boldsymbol{H}}) \leq \lambda_{\max}\Big(\boldsymbol{H}) + \lambda_{\max}(\boldsymbol{W}^T[(\boldsymbol{D}^{-1})^2 - \mathcal{I}]\boldsymbol{W}\Big)$. A similar analysis leads to $\lambda_{\min}(\bar{\boldsymbol{H}}) \geq \lambda_{\min}(\boldsymbol{H}) + \lambda_{\min}\Big(\boldsymbol{W}^T[(\boldsymbol{D}^{-1})^2 - \mathcal{I}]\boldsymbol{W}\Big)$. Assume that, with no weighting, the original basis $\boldsymbol{w}_\alpha$ satisfies the necessary conditions for Theorem 1 to hold. For the weighted nuclear norm minimization to hold, one choice of sufficient conditions is that $\lambda_{\max}\Big(\boldsymbol{W}^T[(\boldsymbol{D}^{-1})^2 - \mathcal{I}]\boldsymbol{W}\Big) = c_1\lambda_{\max}(\bar{\boldsymbol{H}})$ and $\lambda_{\min}\Big(\boldsymbol{W}^T[(\boldsymbol{D}^{-1})^2 - \mathcal{I}]\boldsymbol{W}\Big) = c_2\lambda_{\min}(\bar{\boldsymbol{H}})$ where c_1, c_2 are dimension free constants. A given choice of $\boldsymbol{D}$ can be checked it it verifies these criterion. If the result is positive, the complexity of $O(nr\nu\log^2 n)$ from Theorem 1 can be attained. Note that the conditions above are sufficient but not necessary. Sharper and more explicit condition on $\boldsymbol{D}$ requires further analysis and is not within the scope of this paper. It can be surmised that, in practice, one is working with a fixed basis matrix $\boldsymbol{W}$ and controlling the spectrum of $\bar{\boldsymbol{H}}$ in terms of $\boldsymbol{D}$ is more amenable to analysis.

Remark 3. *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 [18] 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, $\boldsymbol{M}$ is a unique solution to (12).

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

(a). If $\|\boldsymbol{\Delta}_{\mathbb{T}}\|_F \geq \sqrt{2L}\dfrac{\lambda_{\max}(\boldsymbol{H}^{-1})}{\lambda_{\min}(\boldsymbol{H}^{-1})}\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F$ and $\lambda_{\min}\left(\mathcal{P}_{\mathbb{T}}\,\mathcal{F}_\Omega\,\mathcal{P}_{\mathbb{T}}\right) > \frac{1}{2}\lambda_{\min}(\boldsymbol{H})$, then $\mathcal{R}_\Omega\boldsymbol{\Delta} \neq 0$.

(b). If $\|\boldsymbol{\Delta}_{\mathbb{T}}\|_F < \sqrt{2L}\dfrac{\lambda_{\max}(\boldsymbol{H}^{-1})}{\lambda_{\min}(\boldsymbol{H}^{-1})}\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F$ for $\boldsymbol{\Delta} \in \ker\mathcal{R}_\Omega$, and there exists a $\boldsymbol{Y} \in$ range $\mathcal{R}_\Omega^*$ satisfying,

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

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

Theorem 2(a) states that, for "large" $\boldsymbol{\Delta}_{\mathbb{T}}$, any deviation from $\boldsymbol{M}$ is not in the null space of the operator. Theorem 2(b) states that, for "small" $\boldsymbol{\Delta}_{\mathbb{T}}$, deviations from $\boldsymbol{M}$ 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 $\|\mathcal{R}_\Omega\,\boldsymbol{\Delta}\|_F = \|\mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}} + \mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F \geq \|\mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}}\|_F - \|\mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F$. Since we want to show that $\mathcal{R}_\Omega\boldsymbol{\Delta} \neq 0$, the observation leads to considering a lower bound for $\|\mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}}\|_F$ and an upper bound for $\|\mathcal{R}_\Omega\,\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F$. For any $\boldsymbol{X}$, $\|\mathcal{R}_\Omega\,\boldsymbol{X}\|_F^2$ can be bounded as follows.

$$\|\mathcal{R}_\Omega\,\boldsymbol{X}\|_F^2 = \langle \boldsymbol{X}\,, \mathcal{R}_\Omega^*\,\mathcal{R}_\Omega\,\boldsymbol{X}\rangle = \frac{L^2}{m^2}\sum_{\beta\in\Omega}\sum_{\alpha\in\Omega}\langle \boldsymbol{X}\,, \boldsymbol{w}_\alpha\rangle\langle \boldsymbol{X}\,, \boldsymbol{w}_\beta\rangle\langle \boldsymbol{v}_\alpha\,, \boldsymbol{v}_\beta\rangle = \frac{L^2}{m^2}\sum_{\beta\in\Omega}\sum_{\alpha\in\Omega}\langle \boldsymbol{X}\,, \boldsymbol{w}_\alpha\rangle\langle \boldsymbol{X}\,, \boldsymbol{w}_\beta\rangle\boldsymbol{H}^{\alpha,\,\beta} \tag{23}$$

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

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

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

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

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

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

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

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

Above, the first equality follows since $\mathcal{P}_{\mathbb{T}}$ is self adjoint and evidently $\boldsymbol{\Delta}_{\mathbb{T}} \in \mathbb{T}$. The inequality in (27) can be reduced further using the assumption $\lambda_{\min}\left(\mathcal{P}_{\mathbb{T}} \, \mathcal{F}_\Omega \, \mathcal{P}_{\mathbb{T}}\right) > \frac{1}{2}\lambda_{\min}(\boldsymbol{H})$ in the theorem resulting

$$\|\mathcal{R}_\Omega \, \boldsymbol{\Delta}_{\mathbb{T}}\|_F^2 > \frac{L}{2m}\lambda_{\min}(\boldsymbol{H}^{-1})\lambda_{\min}(\boldsymbol{H}) \, \|\boldsymbol{\Delta}_T\|_F^2 \tag{27}$$

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

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

This concludes proof of Theorem 2(a).

□

Remark 4. *(a) The upper bound estimate for $||\mathcal{R}_\Omega \, \boldsymbol{\Delta}_{\mathbb{T}^\perp}||_F^2$ is not optimal. This is so since the number of times a given measurement can be duplicated is set to m which is a worst case estimate. One approach to improve the estimate is to make use of standard concentration inequalties and argue that the expected number of duplicates for a given measurement is much smaller than m. A second option, as noted in the remark on the sampling model section, is to use a uniform sampling with out replacement model. With this, the estimate can be improved since the factor m is no longer necessary. While these alternatives lead to a better estimate of the upper bound, the final gain in terms of measurement is minor as the improved estimates are inside of a $\log_2$. For these reasons and ease of presentation, the current estimate is used in the forthcoming analysis. (b)Lower bounding the term $\sum_{\alpha\in\Omega} \langle \boldsymbol{\Delta}_{\mathbb{T}} \,, \boldsymbol{w}_\alpha \rangle^2$ does not lend itself to simpler analysis. For example, the use of standard concentration inequalities results probability of failures that scale with n. Since $\|\mathcal{R}_\Omega \, \boldsymbol{\Delta}_{\mathbb{T}}\|_F^2 = \langle \boldsymbol{\Delta}_{\mathbb{T}} \,, \mathcal{P}_{\mathbb{T}}\mathcal{R}_\Omega^*\mathcal{R}_\Omega\mathcal{P}_{\mathbb{T}}\boldsymbol{\Delta}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\boldsymbol{\Delta}_{\mathbb{T}} \rangle + \|\mathcal{P}_{\mathbb{T}}\boldsymbol{\Delta}_{\mathbb{T}}\|_F^2$, equivalently, we can consider an upper bound of $\|\mathcal{P}_{\mathbb{T}}\mathcal{R}_\Omega^*\mathcal{R}_\Omega\mathcal{P}_{\mathbb{T}}\boldsymbol{\Delta}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\|$ to lower bound $\|\mathcal{R}_\Omega \, \boldsymbol{\Delta}_{\mathbb{T}}\|_F^2$. The upper bound calculations can be carried out but the calculations are involved and assume certain structure of the basis $\boldsymbol{w}_\alpha$. The simple alternative approach shown above is general since it does not make restrictive assumptions on the measurement basis.*

Proof of Theorem 2(b). Let $\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{\Delta}$ be a feasible solution to (12) with the condition that $\|\boldsymbol{\Delta}_{\mathbb{T}}\|_F < \sqrt{2L}\frac{\lambda_{\max}(\boldsymbol{H}^{-1})}{\lambda_{\min}(\boldsymbol{H}^{-1})}\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F$ and $\boldsymbol{\Delta} \in \ker \mathcal{R}_\Omega$. The goal now is to show that, for any $\boldsymbol{X}$ that satisfies these assumptions, the nuclear norm minimization is violated meaning that $\|\boldsymbol{X}\|_* = \|\boldsymbol{M} + \boldsymbol{\Delta}\|_* > \|\boldsymbol{M}\|_*$. The proof of this fact makes use of the dual certificate approach in [18]. The idea is that to endow a certain object, named a dual certificate $\boldsymbol{Y}$, with certain conditions so as to ensure that any $\boldsymbol{X}$ 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 [18]. First, using the duality of the spectral norm and the nuclear norm, note that there exists a $\boldsymbol{W} \in \mathbb{T}^\perp$ with $||\boldsymbol{W}|| = 1$ such that $\langle \boldsymbol{W}, \mathcal{P}_{\mathbb{T}^\perp}(\boldsymbol{X} - \boldsymbol{M})\rangle = ||\mathcal{P}_{\mathbb{T}^\perp}(\boldsymbol{X} - \boldsymbol{M})||_*$. Second, using the characterization of the subgradient of the nuclear norm [32], $\partial\|\boldsymbol{M}\|_* = \{\operatorname{Sgn}\boldsymbol{M} + \boldsymbol{Z} \mid \boldsymbol{Z} \in \mathbb{T}^\perp \ \& \ \|\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{Z}\| \leq 1\}$. With this, it can be readily verified that $\operatorname{Sgn}\boldsymbol{M} + \boldsymbol{W}$ is a subgradient of $||\boldsymbol{M}||_*$. Mathematically, we have $\|\boldsymbol{M} + \boldsymbol{\Delta}\|_* \geq ||\boldsymbol{M}||_* + \langle \operatorname{Sgn}\boldsymbol{M} + \boldsymbol{W}, \boldsymbol{\Delta}\rangle$. Using the condition $\boldsymbol{Y} \in \text{range}\, \mathcal{R}_\Omega^*$ which implies $\langle \boldsymbol{Y}, \boldsymbol{\Delta}\rangle = 0$, in the previous inequality, we obtain

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

The third equality follows using the earlier choice of $\boldsymbol{W}$ and the fact that $\operatorname{Sgn}\boldsymbol{M} \in \mathbb{T}$ (see Lemma 7). Finally, we apply the assumptions of the theorem to the last equation above to obtain

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

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

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

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

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

3.2 Proof of Theorem 1

Using the corollary above, if the two conditions in Theorem 2 hold, it follows that $\boldsymbol{M}$ is a unique solution to (12). The first condition in Theorem 2(a) is the assumption that $\lambda_{\min}(\mathcal{P}_{\mathbb{T}}\,\mathcal{F}_\Omega\,\mathcal{P}_{\mathbb{T}}) > \frac{1}{2}\lambda_{\min}(\boldsymbol{H})$. 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 [29], restated below, Lemma 2 addresses the assumption.

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

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

//no need to use $\mu_{\max}$// *Then*

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

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

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

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

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

Let $\mathcal{L}_{\boldsymbol{\alpha}} = \frac{L}{m}\langle \cdot\,, \mathcal{P}_{\mathbb{T}}\,\boldsymbol{w}_{\boldsymbol{\alpha}}\rangle\,\mathcal{P}_{\mathbb{T}}\,\boldsymbol{w}_{\boldsymbol{\alpha}}$ denote the operator in the summand. Since $\mathcal{L}_{\boldsymbol{\alpha}}$ is positive semidefinite, the operator Chernoff bound can be used. The bound requires estimate of an upper bound R of the spectrum norm of $\mathcal{L}_{\boldsymbol{\alpha}}$ and $\mu_{\min} = \lambda_{\min}(\sum_{\boldsymbol{\alpha}\in\Omega} E[\mathcal{L}_{\boldsymbol{\alpha}}])$. First, we estimate R as follows.

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

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

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

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

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

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

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

This concludes the proof. □

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

In what follows, the conditions in Theorem 2(b) are analyzed. The statement there assumes the existence of a certain dual certificate $\boldsymbol{Y}$ that satisfies the conditions in (22)). In [18], David Gross devised a novel scheme, the golfing scheme, to construct the dual certificate $\boldsymbol{Y}$. 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 $\mathcal{R}_i = \frac{L}{m_i}\sum_{\alpha\in\Omega_i}\langle \boldsymbol{X}\,,\boldsymbol{w}_\alpha\rangle\boldsymbol{v}_\alpha$. The inductive scheme is shown below.

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

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 $\|\mathcal{P}_{\mathbb{T}}\mathcal{R}_{\Omega}^*\mathcal{P}_{\mathbb{T}}\boldsymbol{X} - \mathcal{P}_{\mathbb{T}}\boldsymbol{X}\|_F \geq t$ for a fixed matrix $\boldsymbol{X}\in\mathbb{S}$ and will be addressed in Lemma 3 to follow shortly. The proof of the lemma relies on the vector Bernstein inequality in [18]. We use a slightly modified version of this inequality which is stated below.

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

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

Lemma 3. *Given an arbitrary fixed $\boldsymbol{X}\in\mathbb{S}$, for $t\leq 1$ with $\kappa = \frac{mn}{Lr}$, the following estimate holds.*

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

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

$$\mathcal{P}_{\mathbb{T}}\mathcal{R}_{\Omega}^*\mathcal{P}_{\mathbb{T}}\boldsymbol{X} - \mathcal{P}_{\mathbb{T}}\boldsymbol{X} = \sum_{\alpha\in\Omega}\left[\frac{L}{m}\langle\mathcal{P}_{\mathbb{T}}\boldsymbol{X}\,,\boldsymbol{v}_\alpha\rangle P_{\mathbb{T}}\,\boldsymbol{w}_\alpha - \frac{1}{m}\mathcal{P}_{\mathbb{T}}\boldsymbol{X}\right] \tag{30}$$

The summand, denoted $\boldsymbol{Y}_\alpha$, can be written as $\boldsymbol{Y}_\alpha = \boldsymbol{X}_\alpha - E[\boldsymbol{X}_\alpha]$. Since $E[\boldsymbol{Y}_\alpha] = \boldsymbol{0}$, it satisfies the condition for the vector Bernstein inequality and we proceed to consider appropriate bounds for $\|\boldsymbol{Y}_\alpha\|_F$ and $E[\|\boldsymbol{Y}_\alpha\|_F^2]$. First, we bound $\|\boldsymbol{Y}_\alpha\|_F$ making use of the coherence conditions (18) and (19).

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

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

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

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

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

□

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. *$\boldsymbol{Y}_l$ 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* $p_2(i) = \exp\left(-\frac{\kappa_i}{32\left(\lambda_{\max}(\boldsymbol{H}^{-1})\|\boldsymbol{H}^{-1}\|_\infty\nu + \frac{n}{Lr}\right)} + \frac{1}{4}\right)$, $p_3(i) = 4n\exp\left(\frac{-3\min\left((\mu+1)\|\boldsymbol{H}^{-1}\|_\infty, \frac{1}{4}\right)^2\kappa_i}{8(\mu+1)\|\boldsymbol{H}^{-1}\|_\infty^2\nu}\right)$ *and* $p_4(i) = n^2\exp\left(-\frac{3\kappa_i}{32\left(c_v\nu + \frac{n}{Lr}\right)}\right)$ *with* $k_i = \frac{m_i n}{Lr}$.

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

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

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

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

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

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

Setting $= \log_2\left(4\sqrt{2L}\frac{\lambda_{\max}(\boldsymbol{H})}{\lambda_{\min}(\boldsymbol{H})}\sqrt{r}\right)$, the first condition in (22) is now satisfied. It can be concluded that, using a union bound on the failure probabilities $p_2(i)$, $\|\boldsymbol{Q}_l\|_F \leq \sqrt{r}2^{-l} = \frac{1}{4}\sqrt{\frac{1}{2L}}\frac{\lambda_{\min}(\boldsymbol{H}^{-1})}{\lambda_{\max}(\boldsymbol{H}^{-1})}$ 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 $\|\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{Y}_l\|$. First, note that $\|\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{Y}_l\| = \|\mathcal{P}_{\mathbb{T}^\perp}\sum_{j=1}^{l}\mathcal{R}_j^*\boldsymbol{Q}_{j-1}\| = \|\sum_{j=1}^{l}\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\boldsymbol{Q}_{j-1}\| \leq \sum_{j=1}^{l}\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\boldsymbol{Q}_{j-1}\|$. As such, in what follows, the focus will be finding a suitable bound for $\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\boldsymbol{Q}_{j-1}\|$. This will be analyzed in Lemma 5. A key element in the proof of Lemma 5 is an assumption on the size of $\max_\beta |\langle \boldsymbol{Q}_i, \boldsymbol{v}_\beta\rangle|$. For ease of notation in further analysis, let $\eta(\boldsymbol{Q}_i)$ be defined as: $\eta(\boldsymbol{Q}_i) = \max_\beta |\langle \boldsymbol{Q}_i, \boldsymbol{v}_\beta\rangle|$. The assumption is that, at the i-th step of the golfing scheme, $\eta(\boldsymbol{Q}_i)^2 \leq \|\boldsymbol{H}^{-1}\|_\infty^2\frac{\nu}{n^2}c_i^2$ with $\|\boldsymbol{Q}_i\|_F^2 \leq c_i^2$. It needs to be argued that this assumption holds with very high probability. To show this, assume that $\eta(\boldsymbol{Q}_i) \leq t_{4,i}$ with failure probability $p_4(i)$. Setting $t_{4,i} = \frac{1}{2}\eta(\boldsymbol{Q}_{i-1})$ and applying the inequality $\eta(\boldsymbol{Q}_i) \leq t_{4,i}$ recursively results

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

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

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

Having justified the assumption on the size of $\eta(\boldsymbol{Q}_i)$, a key part of Lemma 9, we now consider certifying the second condition in (22). Assume that $\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\boldsymbol{Q}_{j-1}\| < t_{3,j}c_{j-1}$, with $\|\boldsymbol{Q}_{j-1}\|_F \leq c_{j-1} = 2^{-(j-1)}$, holds with failure probability $p_3(j)$. Fixing $t_{3,j} = \min\left(\frac{(\mu+1)\|\boldsymbol{H}^{-1}\|_\infty}{\sqrt{r}}, \frac{1}{4\sqrt{r}}\right)$, Lemma 5 gives

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

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

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

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

□

Lemma 5 will be proved shortly using the Bernstein inequality in [30] which is restated below for convenience.

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

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

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

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

For all $t \geq 0$,

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

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

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

Proof of Lemma 5. Using the dual basis representation, $\mathcal{P}_{\mathbb{T}^\perp}\mathcal{R}_j^*\boldsymbol{G} = \sum_{\alpha\in\Omega_j} \dfrac{L}{m_j}\langle \boldsymbol{G}, \boldsymbol{v}_\alpha\rangle \mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{w}_\alpha$. The summand, denoted $\boldsymbol{X}_\alpha$, has zero expectation since $\boldsymbol{G} \in \mathbb{T}$. 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 $\max(\|E[\boldsymbol{X}_\alpha\boldsymbol{X}_\alpha^T]\|, \|E[\boldsymbol{X}_\alpha^T\boldsymbol{X}_\alpha]\|)$. First, we bound $\|E[\boldsymbol{X}_\alpha\boldsymbol{X}_\alpha^T]\|$. Since $\boldsymbol{X}_\alpha\boldsymbol{X}_\alpha^T = \dfrac{L^2}{m_j^2}\langle \boldsymbol{G}, \boldsymbol{v}_\alpha\rangle^2(\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{w}_\alpha)(\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{w}_\alpha)^T$, using Lemma 10 and the fact that $\boldsymbol{w}_\alpha\boldsymbol{w}_\alpha^T$ is positive semidefinite, $\|E[\boldsymbol{X}_\alpha\boldsymbol{X}_\alpha^T]\|$ can be upper bounded as follows.

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

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

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

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

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

To complete the proof, it remains to estimate R.

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

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

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

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

Next, it will be argued that $\boldsymbol{M}$ is a unique solution to (12) with very high probability. This is achieved by showing that $\Pr\left(\boldsymbol{X} \in \mathbb{S} | \mathcal{R}_\Omega\boldsymbol{X} = \mathcal{R}_\Omega\boldsymbol{M}\,\&\,\boldsymbol{X} \neq \boldsymbol{M}, \|\boldsymbol{X}\|_* < \|\boldsymbol{M}\|_*\right) \leq p$ where p is a "small" failure probability. The argument considers two separate cases based on comparing $\|\boldsymbol{\Delta}_\mathbb{T}\|_F^2$ and $\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F^2$. This motivates defining the following two sets: $\mathbb{S}_1 = \left\{\boldsymbol{X} \in \mathbb{S} : \|\boldsymbol{\Delta}_\mathbb{T}\|_F^2 \geq \left(\sqrt{2\frac{L}{m}}\frac{\lambda_{\max}(\boldsymbol{H}^{-1})}{\lambda_{\min}(\boldsymbol{H}^{-1})}\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F\right)^2\right\}$ and $\mathbb{S}_2 = \left\{\boldsymbol{X} \in \mathbb{S} : \|\boldsymbol{\Delta}_\mathbb{T}\|_F^2 <\right.$

$\left(\sqrt{2\frac{L}{m}}\frac{\lambda_{\max}(\boldsymbol{H}^{-1})}{\lambda_{\min}(\boldsymbol{H}^{-1})}\|\boldsymbol{\Delta}_{\mathbb{T}^\perp}\|_F\right)^2 \;\&\; \mathcal{R}_\Omega(\boldsymbol{\Delta}) = \boldsymbol{0}\Big\}$. With this, assuming that Ω is sampled uniformly at random with replacement, the two cases are as follows.

1. For all $\boldsymbol{X} \in \mathbb{S}_1$, set $|\Omega| = m$ "sufficiently large" such that $\Pr\left(\left\{\Omega \subset \mathbb{I} \,\middle|\, |\Omega| = m, \lambda_{\min}\left(\mathcal{P}_{\mathbb{T}}\,\mathcal{F}_\Omega\,\mathcal{P}_{\mathbb{T}}\right) > \frac{\lambda_{\min}(\boldsymbol{H})}{2}\right\}\right) = 1 - p_1$ based on Lemma 2. Therefore, all $\boldsymbol{X} \in \mathbb{S}_1$ are not feasible solutions to (12) with probability $1 - p_1$ from Theorem 2(a).
2. For all $\boldsymbol{X} \in \mathbb{S}_2$, we obtain $\Pr\left(\left\{\Omega \subset \mathbb{I} \,\middle|\, |\Omega| = m, \boldsymbol{Y} \in \text{range}\,\mathcal{R}_\Omega^* \;\&\; \|\mathcal{P}_{\mathbb{T}}\boldsymbol{Y} - \text{Sgn}\,\boldsymbol{M}\|_F \le \frac{1}{4L} \;\&\; \|\mathcal{P}_{\mathbb{T}^\perp}\boldsymbol{Y}\| \le \frac{1}{2}\right\}\right) = 1 - \epsilon$ 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 $\boldsymbol{X} \in \mathbb{S}_2$ being solutions to (12) is ϵ from Theorem 2(b).

Using the above two cases and employing the union bound, any $\boldsymbol{X} \in \mathbb{S}$ different from $\boldsymbol{M}$ 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" with out 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 $p \le n^{-\beta}$ 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 $\boldsymbol{Y}$ 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.

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

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

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 k_i, $k_i = 48\big(C\nu + \frac{1}{nr}\big)\big(\beta\log(n) + \log(4l)\big)$ where C is given by

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

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

$$\log_2\left(4\sqrt{2L}\frac{\lambda_{\max}(\boldsymbol{H})}{\lambda_{\min}(\boldsymbol{H})}\sqrt{r}\right)nr\left(48\big[C\nu + \frac{1}{nr}\big]\big[\,\beta\log(n) + \log\left(4\log_2(4L\sqrt{r})\right)\big]\right) \tag{38}$$

This finishes the proof of Theorem 6. It can be concluded that the minimization program in (12) recovers the underlying matrix with very high probability.

3.3 Noisy General Matrix Completion

In practical applications, the measurements in the general matrix completion problem are prone to noise. This motivates the analysis of the robustness of the nuclear norm minimization program for the general matrix completion problem. In particular, consider the following noisy general matrix completion problem where noise is modeled as Gaussian noise with mean μ and variance σ.

$$\begin{aligned} \underset{\boldsymbol{X}\in\mathbb{S}}{\text{minimize}} \quad & \|\boldsymbol{X}\|_* \\ \text{subject to} \quad & \|\mathcal{R}_\Omega(\boldsymbol{X}) - \mathcal{R}_\Omega(\boldsymbol{M})\| \le \delta \end{aligned} \tag{39}$$

where δ characterizes the level of noise. In [5], under certain assumptions, the authors show the robustness of the nuclear norm minimization algorithm for the matrix completion problem. Following this existing analysis and the dual basis framework, we expect that a robustness result, such as the one below, can be attained.

Theorem 6. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ 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: $C = \max\left(\lambda_{\max}(\boldsymbol{H}^{-1})\|\boldsymbol{H}^{-1}\|_\infty, c_v, \dfrac{(\mu+1)\|\boldsymbol{H}^{-1}\|_\infty}{\min\left((\mu+1)\|\boldsymbol{H}^{-1}\|_\infty, \frac{1}{4}\right)^2}\right)$. Assume m measurements $\langle \boldsymbol{M}, \boldsymbol{w_\alpha}\rangle$, sampled uniformly at random with replacement are corrupted with Gaussian noise of mean μ, variance σ and noise level δ. For $\beta > 1$, if*

$$m \geq \log_2\left(4\sqrt{2L}\frac{\lambda_{\max}(\boldsymbol{H})}{\lambda_{\min}(\boldsymbol{H})}\sqrt{r}\right) nr \left(48\left[C\nu + \frac{1}{nr}\right]\left[\beta\log(n) + \log\left(4\log_2(4L\sqrt{r})\right)\right]\right) \tag{40}$$

then

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

where $\bar{\boldsymbol{M}}$ is a solution to (39) *with probability at least $1 - n^{-\beta}$.*

Remark: A proof of the robustness theorem above requires specifying the level of noise and making the function f explicit. We leave this as a future work.

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 where the restricted isometry property (RIP) condition might not hold or is NP hard to verify. 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 evalauting the effectiveness of the framework of the general matrix completion in certain applications.

References

1. Bandeira, A.S., Dobriban, E., Mixon, D.G., Sawin, W.F.: Certifying the restricted isometry property is hard. IEEE transactions on information theory **59**(6), 3448–3450 (2013)
2. Biswas, P., Lian, T.C., Wang, T.C., Ye, Y.: Semidefinite programming based algorithms for sensor network localization. ACM Transactions on Sensor Networks (TOSN) **2**(2), 188–220 (2006)
3. Cai, J.F., Wang, T., Wei, K.: Fast and provable algorithms for spectrally sparse signal reconstruction via low-rank hankel matrix completion. Applied and Computational Harmonic Analysis (2017)
4. Candes, E.J., Eldar, Y.C., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM review **57**(2), 225–251 (2015)
5. Candes, E.J., Plan, Y.: Matrix completion with noise. Proceedings of the IEEE **98**(6), 925–936 (2010)
6. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Foundations of Computational mathematics **9**(6), 717–772 (2009)
7. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory **52**(2), 489–509 (2006)
8. Candes, E.J., Strohmer, T., Voroninski, V.: Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. Communications on Pure and Applied Mathematics **66**(8), 1241–1274 (2013)
9. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE transactions on information theory **51**(12), 4203–4215 (2005)
10. Candès, E.J., Tao, T.: The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory **56**(5), 2053–2080 (2010)
11. Courtois, N., Klimov, A., Patarin, J., Shamir, A.: Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In: International Conference on the Theory and Applications of Cryptographic Techniques, pp. 392–407. Springer (2000)

12. Ding, Y., Krislock, N., Qian, J., Wolkowicz, H.: Sensor network localization, euclidean distance matrix completions, and graph realization. Optimization and Engineering **11**(1), 45–66 (2010)
13. Fang, X., Toh, K.C.: Using a distributed sdp approach to solve simulated protein molecular conformation problems. In: Distance Geometry, pp. 351–376. Springer (2013)
14. Fazel, M., Hindi, H., Boyd, S.P.: A rank minimization heuristic with application to minimum order system approximation. In: American Control Conference, 2001. Proceedings of the 2001, vol. 6, pp. 4734–4739. IEEE (2001)
15. Fazel, M., Hindi, H., Boyd, S.P.: 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, pp. 2156–2162. IEEE (2003)
16. Foygel, R., Shamir, O., Srebro, N., Salakhutdinov, R.R.: Learning with the weighted trace-norm under arbitrary sampling distributions. In: Advances in Neural Information Processing Systems, pp. 2133–2141 (2011)
17. Glunt, W., Hayden, T., Raydan, M.: Molecular conformations from distance matrices. Journal of Computational Chemistry **14**(1), 114–120 (1993)
18. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. Information Theory, IEEE Transactions on **57**(3), 1548–1566 (2011)
19. Gross, D., Nesme, V.: Note on sampling without replacing from a finite collection of matrices. arXiv preprint arXiv:1001.2738 (2010)
20. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American statistical association **58**(301), 13–30 (1963)
21. Kalofolias, V., Bresson, X., Bronstein, M., Vandergheynst, P.: Matrix completion on graphs. arXiv preprint arXiv:1408.1717 (2014)
22. Lai, R., Li, J.: Solving partial differential equations on manifolds from incomplete interpoint distance. SIAM Journal on Scientific Computing **39**(5), A2231–A2256 (2017)
23. Oymak, S., Mohan, K., Fazel, M., Hassibi, B.: A simplified approach to recovery conditions for low rank matrices. In: Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, pp. 2318–2322. IEEE (2011)
24. Recht, B.: A simpler approach to matrix completion. The Journal of Machine Learning Research **12**, 3413–3430 (2011)
25. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM review **52**(3), 471–501 (2010)
26. Srebro, N., Salakhutdinov, R.R.: Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In: Advances in Neural Information Processing Systems, pp. 2056–2064 (2010)
27. Tasissa, A., Lai, R.: Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion. arXiv preprint arXiv:1804.04310 (2018)
28. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. science **290**(5500), 2319–2323 (2000)
29. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Foundations of computational mathematics **12**(4), 389–434 (2012)
30. Tropp, J.A., et al.: An introduction to matrix concentration inequalities. Foundations and Trends® in Machine Learning **8**(1-2), 1–230 (2015)
31. Trosset, M.W.: Applications of multidimensional scaling to molecular conformation (1997)
32. Watson, G.A.: Characterization of the subdifferential of some matrix norms. Linear algebra and its applications **170**, 33–45 (1992)

A Appendix A

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

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

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

It follows that $\sum_\alpha \boldsymbol{C}_\alpha \boldsymbol{C}_\alpha^T = n\mathcal{I}$. An analogous calculation, using row-wise vectorization of $\boldsymbol{c}_\alpha$, results $\sum_\alpha \boldsymbol{C}_\alpha^T \boldsymbol{C}_\alpha = n\mathcal{I}$. //matrix product is different from the product after vectorization. The proof seems not correct//

□

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

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

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

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

□

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

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

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

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

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

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

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

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

This concludes the proof. □

Lemma 9. *Define $\eta(\boldsymbol{X}) = \max_{\beta\in\mathbb{I}} |\langle \boldsymbol{X}\,, \boldsymbol{v}_\beta\rangle|$. For $\boldsymbol{X}$ in $\mathbb{T}$,*

$$Pr(\max_{\beta\in\mathbb{I}}\;|\langle \mathcal{P}_{\mathbb{T}}\mathcal{R}_j^*\boldsymbol{X} - \boldsymbol{X}\,, \boldsymbol{v}_\beta\rangle| \ge t) \le n^2\exp\left(-\frac{3t^2\kappa_j}{8\eta(\boldsymbol{X})^2\left(c_v\nu + \frac{n}{Lr}\right)}\right) \tag{41}$$

for all $t \le \eta(\boldsymbol{X})$ with $\kappa_j = \dfrac{m_j n}{Lr}$.

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

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

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).

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

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

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

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

$$\Pr(|\langle \mathcal{P}_{\mathbb{T}}\mathcal{R}_j^*\boldsymbol{X} - \boldsymbol{X}\,, \boldsymbol{v}_\beta\rangle| \ge t) \le \exp\left(-\frac{3t^2\kappa_j}{8\eta(\boldsymbol{X})^2\left(c_v\nu + \frac{n}{Lr}\right)}\right) \tag{42}$$

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

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

$$\left\| \sum_\alpha \boldsymbol{c}_\alpha \, (\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha)(\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha)^T \right\| \leq \left\| \sum_\alpha \boldsymbol{c}_\alpha \, \boldsymbol{w}_\alpha \boldsymbol{w}_\alpha^T \right\|$$

$$\left\| \sum_\alpha \boldsymbol{c}_\alpha \, (\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha)^T (\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha) \right\| \leq \left\| \sum_\alpha \boldsymbol{c}_\alpha \, \boldsymbol{w}_\alpha^T \boldsymbol{w}_\alpha \right\|$$

Proof. We start with the first statement. Using the definition of $\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_{\boldsymbol{\alpha}}^2$, $\left\| \sum_\alpha \boldsymbol{c}_\alpha \, (\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha)(\mathcal{P}_{\mathbb{T}^\perp} \, \boldsymbol{w}_\alpha)^T \right\|$ can be written as follows.

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

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

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

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