Harmonic Retrieval Using Weighted Lifted-Structure Low-Rank Matrix Completion

Mohammad Bokaei[a,b], Saeed Razavikia[a,c], Stefano Rini[d], Arash Amini[a], Hamid Behrouzi[a]

[a]*Electrical Engineering Department, Sharif University of Technology, Iran*
[b]*Department of Electronic Systems, Aalborg university, Denmark*
[c]*School of Electrical Engineering and Computer Science, KTH university, Sweden*
[d]*Electrical and Computer Engineering Department, National Yang-Ming Chao-Tung University (NYCU), Taiwan*

Abstract

In this paper, we investigate the problem of recovering the frequency components of a mixture of K complex sinusoids from a random subset of N equally-spaced time-domain samples. Because of the random subset, the samples are effectively non-uniform. Besides, the frequency values of each of the K complex sinusoids are assumed to vary continuously within a given range. For this problem, we propose a two-step strategy: (i) we first lift the incomplete set of uniform samples (unavailable samples are treated as missing data) into a structured matrix with missing entries, which is potentially low-rank; then (ii) we complete the matrix using a weighted nuclear minimization problem. We call the method a *weighted lifted-structured (WLi) low-rank matrix recovery.* Our approach can be applied to a range of matrix structures such as Hankel and double-Hankel, among others, and provides improvement over the unweighted existing schemes such as EMaC and DEMaC. We provide theoretical guarantees for the proposed method, as well as numerical simulations in both noiseless and noisy settings. Both the theoretical and the numerical results confirm the superiority of the proposed approach.

Keywords: Hankel structure; Lifting operator; Low-rank matrix completion.

1. Introduction

Mixture of complex exponential functions are observed in many real-world applications such as medical imaging [1], astronomical imaging [2], millimeter-wave imaging [3], and artworks [4, 5]. For instance, in many imaging systems, the optical point spread function can be fairly approximated by a mixture of a few exponential functions. Therefore, the measurements observed in an imaging system are mainly a combination of a few exponential components. Even if these mixtures are of fairly high dimension, their representation in the Fourier domain consists of a small number of spikes. The spectral estimation challenge is to decompose such a mixture into exponential elements or, simply, to separate the Fourier spikes. As we observe analog signals through digital sensing devices, the sampling resolution is a limiting factor in properly detecting and separating exponential components in a mixture. There is extensive literature within the signal processing community for surpassing the limits imposed on the physical resolution using processing techniques, which are generally known as super-resolution algorithms [6]. The technique proposed in [7] is a famous grid-less example that relies on sparse recovery and compressed sensing tools. In this work, equally-spaced samples are employed to separate the components using a convex optimization in the line spectral estimation problem. The authors of [8] consider a similar scenario but with random samples. For this setting, a probabilistic guarantee for the number of required samples for perfect recovery is provided. For perfect recovery, both [7] and [8] require a minimum phase separation between the complex sinusoid components (minimum distance between the spikes). Inspired by the matrix pencil algorithm in [9], [10] proposed the EMaC method using matrix completion to reduce this minimum required separation. The method uses a Hankel lifting structure that transforms uniform samples of exponential mixtures into a low-rank matrix. Once lifted, the recovery of missing samples can be treated as completing a low-rank matrix, which is a well-studied problem [11]. This approach is further extended in [12] for the larger class of signals with finite rate of innovation. In [13], the problem of low-rank Hankel matrix recovery for random Gaussian projections is investigated, and a lower bound for sample complexity (with high probability) is derived. Besides the low-rank property of the matrix, most matrix completion techniques require the available entries to be uniformly spread within the matrix. For the case of non-uniform samples, a two-phase sampling-recovery strategy is proposed in [14]. However, the method does not work for structured matrices (such as Hankel matrices).

Contributions: In this work, we consider the problem of separating exponential signals from a mixture of non-uniform samples. The proposed spectral estimation approach is comprised of two steps. We first lift the samples/measurements to a structured low-rank matrix with missing samples. Next, we propose a general approach to enhance matrix completion by using a measurement-adaptive weighting scheme in which the weights reflect the relative significance of the samples. The lifting operation encompasses Hankel, double-Hankel, wrap-around Hankel, Hankel-block-Hankel, Toeplitz, and multi-level Toeplitz structures as special cases. While our approach bears similarity to the atomic norm minimizations (ANM) in [15, 16], we do not require the statistics of the sources. Instead, we take advantage of the the concept of *leverage scores* studied in [14] for determining the sample weights based on their informativeness. We refer to this approach as "weighted lifted-structure (WLi) low-rank matrix recovery". We show that the weighting scheme of [14] reduces the number of measurements required for estimating the exponential components. After having presented the weighted lifted-structured low-rank matrix completion in some generality, we consider Hankel and double-Hankel structures as special cases (similar to EMaC

[10], and DEMac [17], respectively), resulting in WLi-EMaC and WLi-DEMaC methods. The simulation results show that the weighted methods require much fewer samples to recover the input vector than the unweighted scheme.

Note well: This paper is the first part of a two-part submission. In this part, we mainly discuss the theoretical perspective of the weighted recovery in a lifted-structure completion problem. In Part II [18], we focus on the specific application of DOA estimation using the theoretical results derived here.

Organization: in Section 2, we introduce lifting operator. The problem formulation is provided in Section 3. The weighted matrix completion problem is described in Section 4. In Section 5, we present the theoretical guarantees for the weighted approach. Numerical simulations are provided in Section 6. Finally, we conclude the paper in Section 7. *Notations:* We use lowercase letters, lower and upper-case boldface letters to represent scalars, vectors, and matrices respectively. We further show linear operators and their adjoints by calligraphic notations such as $\mathcal{X}$ and $\mathcal{X}^\dagger$, where the superscript in the latter stands for the adjoint operator. Moreover, $\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{H}$ denote the transpose and Hermitian of a matrix $\mathbf{X}$, respectively. $\mathbf{X}\odot\mathbf{Y}$ and $\langle\mathbf{X},\mathbf{Y}\rangle$ show the Hadamard (element-wise) product and inner product of two equi-size matrices $\mathbf{X}$ and $\mathbf{Y}$, respectively. We denote the spectral, Frobenius, and nuclear norms of a matrix $\mathbf{X}$ by $\|\mathbf{X}\|$, $\|\mathbf{X}\|_\mathrm{F}$ and $\|\mathbf{X}\|_*$, respectively. Similarly, $\|\mathbf{X}\|_1$ and $\|\mathbf{X}\|_\infty$ stand for the element-wise ℓ_1 and ℓ_∞ norms of $\mathbf{X}$ (treating $\mathbf{X}$ as a vector). $\|\mathbf{X}\|_{\infty\to\infty}$ is defined as $\max_{i\in[N]}\sum_j |x_{i,j}|$ where $x_{i,j}$ denotes (i,j) element of matrix $\mathbf{X}$. Further, $\|\mathbf{X}\|_0$ is the number of non-zero elements of matrix $\mathbf{X}$. We refer to $\mathbf{e}_i^N$ as the i-th canonical basis vector in dimensional N. For an integer n, $[n]$ stands for $\{1,2,\dots,n\}$. We also define the operator $\mathrm{diag}:\mathbb{C}^N\mapsto\mathbb{C}^{N\times N}$ maps a vector $\mathbf{x}\in\mathbb{C}^N$ into a diagonal matrix with diagonal entries as in $\mathbf{X}$. The Greek letter Ω always reflects a finite set of the integers and $|\Omega|$ denotes its cardinality.

2. Lifting operator

Fundamental to our approach is the transformation of a vector into a matrix to gain degrees of freedom. We call this transformation the *lifting operator*. To properly define the class of lifting operators considered in this paper, we initially introduce the concept of the *lifting basis*.

Definition 1. *We call* $\{\mathbf{A}_n\}_{n\in[N]} \subseteq \mathbb{C}^{d_1\times d_2}$ *a* lifting basis *if*

1. *for all* $1 \leq n \leq N$ *we have that* $\|\mathbf{A}_n\|_{\mathrm{F}} = 1$,
2. *all the non-zero elements of* $\mathbf{A}_n$ *are positive, real and equal, and*
3. $\mathbf{A}_n$*s are orthogonal:*

$$\langle \mathbf{A}_{n_1}\,,\, \mathbf{A}_{n_2} \rangle = \mathrm{tr}\big(\mathbf{A}_{n_1}^{\mathsf{T}}\, \mathbf{A}_{n_2}\big) = \delta[n_1 - n_2], \tag{1}$$

4. *and each column of* $\mathbf{A}_n$ *(for* $n \in [N]$*) has at most one nonzero element, i.e.,*

$$\sum_{j\in[d_2]} \Bigg(\sum_{i\in[d_1]} [\mathbf{A}_n]_{i,j} \Bigg)^2 = 1. \tag{2}$$

For a lifting basis $\{\mathbf{A}_n\}_{n\in[N]}$, Definition 1 implies that the non-zero elements in $\mathbf{A}_n$ are all equal to $\frac{1}{\sqrt{\|\mathbf{A}_n\|_0}}$. This further shows that

$$\|\mathbf{A}_n\| \leq \|\mathbf{A}_n\|_{\mathrm{F}} = 1. \tag{3}$$

Definition 2. *Let* $\{\mathbf{A}_n\}_{n\in[N]} \subseteq \mathbb{C}^{d_1\times d_2}$ *be a lifting basis according to Definition 1. The linear mapping* $\mathscr{L} : \mathbb{C}^N \mapsto \mathbb{C}^{d_1\times d_2}$ *defined by*

$$\mathscr{L}(\mathbf{x}) = \sum_{n\in[N]} a_n \,\langle \mathbf{e}_n^N, \mathbf{x}\rangle\, \mathbf{A}_n, \tag{4}$$

is called a lifting operator*, where* $\{a_n\}_{n\in[N]} \subseteq \mathbb{C}$ *are constants. We can check that* $\mathscr{L}^\dagger : \mathbb{C}^{d_1\times d_2} \mapsto \mathbb{C}^N$ *with*

$$\mathbf{M} \in \mathbb{C}^{d_1\times d_2} : \quad \mathscr{L}^\dagger(\mathbf{M}) = \sum_{\substack{n\in[N] \\ a_n\neq 0}} \tfrac{1}{a_n} \langle \mathbf{A}_n, \mathbf{M}\rangle\, \mathbf{e}_n^N, \tag{5}$$

defines the orthogonal back projection from $\mathbb{C}^{d_1\times d_2}$ *into* $\mathbb{C}^N$*.*

By tuning the lifting basis, one can achieve various matrix structures in the output of the lifting operator, such as Hankel, double-Hankel, wrap-around Hankel, Hankel-block-Hankel, Toeplitz, and multi-level Toeplitz. As an example, we examine the Hankel lifting operator $\mathscr{H} : \mathbb{C}^N \to \mathbb{C}^{d\times(N-d+1)}$ with $d \in [N]$:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad \Rightarrow \quad \mathscr{H}(\mathbf{x}) := \begin{bmatrix} x_1 & x_2 & \dots & x_{N-d+1} \\ x_2 & x_3 & \dots & x_{N-d+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_d & x_{d+1} & \dots & x_N \end{bmatrix}. \tag{6}$$

It is not difficult to verify that this operator corresponds to

$$\begin{aligned}
1 \le n \le d : \quad & \mathbf{A}_n^{(\mathscr{H})} := \frac{\sum_{i\in[n]} \mathbf{e}_i^d \mathbf{e}_{n-i+1}^{(N-d+1)^\mathsf{T}}}{\sqrt{n}}, && a_n^{(\mathscr{H})} := \sqrt{n}, \\
d{+}1 \le n \le N{-}d{+}1 : \quad & \mathbf{A}_n^{(\mathscr{H})} := \frac{\sum_{i\in[d]} \mathbf{e}_i^d \mathbf{e}_{n-i+1}^{(N-d+1)^\mathsf{T}}}{\sqrt{d}}, && a_n^{(\mathscr{H})} := \sqrt{d}, \\
N{-}d{+}2 \le n \le N : \quad & \mathbf{A}_n^{(\mathscr{H})} := \frac{\sum_{i=n-N+d}^{d} \mathbf{e}_i^d \mathbf{e}_{n-i+1}^{(N-d+1)^\mathsf{T}}}{\sqrt{N-n+1}}, && a_n^{(\mathscr{H})} := \sqrt{N-n+1}.
\end{aligned} \tag{7}$$

Our recovery method uses certain values associated with a lifted structure known as *leverage scores*. These scores were originally defined in [14] for an adaptive sampling scheme: if we have observed an incomplete matrix and wish to take some more samples before starting to estimate the unobserved entries, which elements are the best options that facilitate the estimation task. For this purpose, each matrix element was assigned a score in [14], which was later interpreted as the sampling probability; i.e., an unobserved entry with a larger leverage score is more likely to be sampled. It is shown in [14] that for a given recovery quality, this strategy requires fewer samples compared to the case of observing the matrix entries uniformly at random. In the matrix completion problem of [14], each of the elements of the low-rank matrix $\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}$ could be observed independently of other elements; hence, $d_1 \times d_2$ leverage scores are defined. In contrast, among $d_1 \times d_2$ elements of $\mathscr{L}(\mathbf{x})$ only N are different ($\mathbf{x} \in \mathbb{C}^N$). This means that N different leverage scores are possible. In Definition 3, we generalize the concept of leverage scores to the case of lifted structures with reduced degrees of freedom.

Definition 3. *Let $\mathscr{L}$ be a lifting operator with basis $\{\mathbf{A}_n\}_{n\in[N]}$. For each $\mathbf{x} \in \mathbb{C}^N$ we define leverage scores $\{\mu_n\}_{n\in[N]}$ as*

$$\mu_n := \frac{N}{\tilde{K}} \max\left\{\|\mathbf{U}^{\mathsf{H}}\mathbf{A}_n\|_{\mathrm{F}}^2, \|\mathbf{A}_n\mathbf{V}^{\mathsf{H}}\|_{\mathrm{F}}^2\right\}, \tag{8}$$

where $\tilde{K}$ is the rank of $\mathscr{L}(\mathbf{x})$ and $\mathbf{U}_{d_1\times\tilde{K}}\mathbf{\Sigma}_{\tilde{K}\times\tilde{K}}(\mathbf{V}_{d_2\times\tilde{K}})^{\mathsf{H}}$ represents the singular value decomposition (SVD) of $\mathscr{L}(\mathbf{x})$.

For the particular case of low-rank Hankel matrix completion, it is shown in [10, Theorem 1] that nuclear-norm minimization succeeds in recovering the matrix if the number of [random] samples exceeds a threshold that is proportional to $\max_n\{\mu_n\}$. In this paper, we provide recovery guarantees (noiseless and noisy cases) for sample sizes that scale with $\frac{\sum_n \mu_n}{N}$.

3. Signal Model

Let the signal of interest $y(t)$ be a linear mixture of K exponential components. The samples of this signal (whether available or unavailable) are

$$y_n = y(n) = \sum_{k\in[K]} b_k z_k^n, \quad n \in [N], \tag{9}$$

where $\{b_k\}_{k\in[K]} \in \mathbb{C}$ are the coefficients in the linear mixture and $\{z_k\}_{k\in[K]} \in \mathbb{C}$ are the complex basis. Using vectorial notations, we write $\mathbf{y} = [y_1, \ldots, y_N]^T \in \mathbb{C}^N$. In case we have measurement noise, we have access to the noisy samples (if available)

$$\tilde{y}_n = y_n + e_n, \tag{10}$$

where $\tilde{\mathbf{y}} = [\tilde{y}_1, \ldots, \tilde{y}_N]^T$ and $\mathbf{e} = [e_1, \ldots, e_N]^T \in \mathbb{C}^N$ stand for the measurement and the noise vectors, respectively. We assume that for each $n \in [N]$, noise amplitudes are upper-bounded as $|e_n| < \eta$ with high probability. This implies that $\|\mathbf{e}\|_2 \leq \sqrt{N}\eta$ (with high probability). Further, let $\Omega \subseteq [N]$ with $|\Omega| = M \leq N$ be the index set of available samples, i.e., y_n or $\tilde{y}_n$ is available only if $n \in \Omega$. Mathematically, we denote the vector of available samples as

$$\mathbf{y}_\Omega = \mathcal{P}_\Omega(\mathbf{y}), \text{ or } \tilde{\mathbf{y}}_\Omega = \mathcal{P}_\Omega(\tilde{\mathbf{y}}), \tag{11}$$

for the noiseless and noisy cases, respectively. Here, $\mathcal{P}_\Omega : \mathbb{C}^N \to \mathbb{C}^M$ stands for the orthogonal projection that keeps the elements with index inside Ω. Note that the energy of the noise content in $\tilde{\mathbf{y}}_\Omega$ is upperbounded by $\sqrt{|\Omega|}\eta = \sqrt{M}\,\eta$ with high probability.

The main challenge with signal model in (9), is to estimate $y(t)$, or equivalently, $\{b_k\}_{k\in[K]}$ and $\{z_k\}_{k\in[K]}$, by observing the noiseless samples $\mathbf{y}_\Omega$ or the noisy measurements $\tilde{\mathbf{y}}_\Omega$. In this work, our focus is on the case where $K \ll N$.

This challenge appears in a number of real-world applications, such as magnetic resonance imaging (MRI) and X-ray computed tomography [19], direction of arrival estimation [20], spike sorting in neural recordings [21], and super-resolution microscopy [22].

4. Low Rank interpolation

To recover the unseen samples from the available measurements in subset Ω, one can use the fact that the rank of the Hankel transform $\mathscr{H}(\mathbf{y})$ is upper-bounded by K which is usually smaller than the size of $\mathbf{y}$ [10, 5, 23, 12]. Similar structures like Toeplitz, wrap-around Hankel, and double-Hankel impose similar low-rank properties. To include all these structures in our analysis, we use the generic $\mathscr{L}$ operator defined in (4) that maps samples of exponential mixtures into low-rank matrices. By choosing $\mathscr{L}$, our next step is to recover $\mathscr{L}(\mathbf{y})$ based on the measurements $\mathbf{y}_\Omega$ (or $\tilde{\mathbf{y}}_\Omega$). In other words, the measurements within the index set Ω shall be extended to the whole set $[N]$. As $\mathscr{L}(\mathbf{y})$ is a low-rank matrix, this task can be reformulated as a matrix completion problem: the elements of $\mathscr{L}(\mathbf{y})$ associated with Ω are observed (possibly noisy), and we want to estimate the rest.

In a matrix completion problem, ideally, one searches for a matrix with the minimum rank that satisfies the constraints. The rank function is, however, both non-convex and non-smooth. Therefore, the exact rank minimization problem is generally NP-hard. The common alternative is to relax the rank$(\cdot)$ function with the nuclear norm [11, 24, 25]. Adopting this relaxation, we shall consider the following matrix completion problem.

$$\widehat{\mathbf{y}} = \underset{\mathbf{g}\in\mathbb{C}^N}{\operatorname{argmin}} \quad \|\mathscr{L}(\mathbf{g})\|_*, \quad \text{s.t.} \quad \mathcal{P}_\Omega(\mathbf{g}) = \mathbf{y}_\Omega, \tag{12}$$

for the noiseless, and

$$\widehat{\tilde{\mathbf{y}}} = \underset{\mathbf{g} \in \mathbb{C}^N}{\operatorname{argmin}} \quad \|\mathscr{L}(\mathbf{g})\|_*, \quad \text{s.t.} \quad \|\mathcal{P}_\Omega(\mathbf{g}) - \tilde{\mathbf{y}}_\Omega\|_2 \leq \sqrt{M}\eta, \tag{13}$$

for the noisy case, $\eta > 0$ was previously introduced as an upper bound for noise amplitudes (with high probability).

4.1. Weighted Matrix Completion

In our scenario, the samples are fixed, and we cannot sample adaptively. As a result, the conventional interpretation of leverage scores as the sampling probabilities is useless here [14]. Instead, we try to set two weight matrices $\mathbf{W}_L \in \mathbb{C}^{d_1 \times d_1}$ and $\mathbf{W}_R \in \mathbb{C}^{d_2 \times d_2}$ such that the recovery of $\mathbf{W}_L \mathscr{L}(\mathbf{g}) \mathbf{W}_R^{\mathsf{H}}$ with a non-uniform sampling strategy that is consistent with the available samples (a more rigorous definition will be provided later) requires fewer samples. The weighted lifted-structured low-rank matrix recovery is defined by incorporating left and right weight matrices into (12) as

$$\widehat{\mathbf{y}} = \underset{\mathbf{g} \in \mathbb{C}^N}{\operatorname{argmin}} \quad \|\mathbf{W}_L \mathscr{L}(\mathbf{g}) \mathbf{W}_R^{\mathsf{H}}\|_* \quad \text{s.t.} \qquad \mathcal{P}_\Omega(\mathbf{g}) = \mathbf{y}_\Omega. \tag{14}$$

Similarly, for the noisy case, we have that

$$\widehat{\mathbf{y}} = \underset{\mathbf{g} \in \mathbb{C}^N}{\operatorname{argmin}} \quad \|\mathbf{W}_L \mathscr{L}(\mathbf{g}) \mathbf{W}_R^{\mathsf{H}}\|_* \quad \text{s.t.} \qquad \|\mathcal{P}_\Omega(\mathbf{g}) - \tilde{\mathbf{y}}_\Omega\|_{\mathrm{F}} \leq \sqrt{M}\eta. \tag{15}$$

The optimization in (15) is convex and can be reformulated into a semi-definite program (SDP) using Schur complement as in [24], i.e.,

$$\begin{aligned} \widehat{\mathbf{y}} = \min_{\mathbf{P},\mathbf{Q}} \quad & \tfrac{1}{2}\mathrm{tr}(\mathbf{P}) + \tfrac{1}{2}\mathrm{tr}(\mathbf{Q}) \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{P} & \mathbf{W}_R \mathscr{L}^{\mathsf{H}}(\mathbf{g}) \mathbf{W}_L^{\mathsf{T}} \\ \mathbf{W}_L \mathscr{L}(\mathbf{g}) \mathbf{W}_R^{\mathsf{T}} & \mathbf{Q} \end{bmatrix} \succeq \mathbf{0}, \\ & \mathcal{P}_\Omega(\mathbf{g}) = \mathbf{y}_\Omega, \quad \mathbf{P}, \mathbf{Q} \succeq \mathbf{0}, \end{aligned} \tag{16}$$

where $\mathbf{P} \in \mathbb{C}^{d_2 \times d_2}$ and $\mathbf{Q} \in \mathbb{C}^{d_1 \times d_1}$ are Hermitian matrices.

We should highlight that the results in [10] are not directly applicable to the weighted problem in (14). In the next section, we analyze the weighted minimization with the general perspective of non-uniform sampling.

5. Theoretical Guarantee and Main results

In this section, we investigate the conditions under which the uniqueness of the solution can be guaranteed. Here, we present recovery guarantees for generic lifting operators $\mathscr{L}$ that transform the vector of exponential mixtures into low-rank matrices (including Hankel, double Hankel, wrap-around Hankel, Toeplitz, Hankel-block-Hankel, and multi-level Toeplitz among others).

Before proceeding further, let us define the weighted leverage scores as a generalization of Definition 3.

Definition 4. *For given weight matrices $\mathbf{W}_L, \mathbf{W}_R$ and an arbitrary vector $\mathbf{x} \in \mathbb{C}^N$, assume that the rank of $\mathbf{W}_L \mathscr{L}(\mathbf{x}) \mathbf{W}_R^{\mathsf{H}}$ is $\tilde{K}$. Further, let $\mathbf{U}_{d_1 \times \tilde{K}} \mathbf{\Sigma}_{\tilde{K} \times \tilde{K}} (\mathbf{V}_{d_2 \times \tilde{K}})^{\mathsf{H}}$ be the SVD thereof. For each $n \in [N]$, we define the weighted leverage scores $\tilde{\mu}_n$ as*

$$\tilde{\mu}_n := \frac{N}{\tilde{K}} \max\{\|\mathcal{P}_U(\mathbf{A}_n)\|_{\mathrm{F}}^2, \|\mathcal{P}_V(\mathbf{A}_n)\|_{\mathrm{F}}^2\}, \quad n \in [N], \tag{17}$$

where $\mathcal{P}_U(\mathbf{Y})$ and $\mathcal{P}_V(\mathbf{Y})$ for arbitrary $\mathbf{Y} \in \mathbb{C}^{d_1 \times d_2}$ are defines as:

$$\mathcal{P}_U(\mathbf{Y}) = \mathbf{W}_L^{\mathsf{H}}\mathbf{U}\left(\mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathbf{W}_L^{\mathsf{H}}\mathbf{U}\right)^{-1}\mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathbf{Y}, \tag{18a}$$

$$\mathcal{P}_V(\mathbf{Y}) = \mathbf{Y}\mathbf{W}_R^{\mathsf{H}}\mathbf{V}\left(\mathbf{V}^{\mathsf{H}}\mathbf{W}_R\mathbf{W}_R^{\mathsf{H}}\mathbf{V}\right)^{-1}\mathbf{V}^{\mathsf{H}}\mathbf{W}_R. \tag{18b}$$

5.1. Recovery Guarantees for Non-uniform Random Sampling

We first assume a random sampling scenario in which Ω is formed by selecting $n \in [N]$ with probability p_n independently of other elements $k \neq n$. Below, we provide a set of lower bounds on $\{p_n\}$s to guarantee perfect (or robust) recovery with a high probability of using noiseless (noisy) samples.

Theorem 1. *Let $\mathbf{y} \in \mathbb{C}^N$ be as in (9), and Ω represent a location set of size M formed by selecting each element $n \in [N]$ with probability p_n independent of other elements. We can recover $\mathbf{y}$ from the measurements $\mathbf{y}_\Omega = \mathcal{P}_\Omega(\mathbf{y})$ using the noiseless setup in (14) with probability no less than $1 - N^{3-b_1}$ if*

$$p_n \geq \min\left\{1\,,\ \frac{1}{N}\max\left\{1\,,\ \mathsf{R}_{\mathscr{L}}^2 c\,\tilde{\mu}_n \widetilde{K}^2 \log(N)\right\}\right\}, \tag{19}$$

and

$$\frac{1}{8\sqrt{\log(N)}} \leq \min_{i \in [N]}\left\{\|\mathbf{A}_i\|_0 \min\{\|\mathcal{P}_U(\mathbf{A}_i)\|_{\mathrm{F}}^2, \|\mathcal{P}_V(\mathbf{A}_i)\|_{\mathrm{F}}^2\}\right\}, \tag{20}$$

where d_1, d_2 and $\tilde{K}$ are dimensions and the rank of the lifted structure $\mathscr{L}(\mathbf{y})$, respectively. Additionally, $\tilde{\mu}_n$ is the weighted leverage score in (17), the coefficient $\mathsf{R}_{\mathscr{L}}$ is defined as

$$\mathsf{R}_{\mathscr{L}} = \sum_{n \in [N]} \left\|\mathbf{A}_n \odot \mathbf{A}_n\right\|_{\infty\to\infty}, \tag{21}$$

and $c = 192^2(b_1 + 1)$ for $b_1 \geq 3$.

Proof. The proof is provided in the appendix. □

Corollary 1. *The parameter* $\mathsf{R}_{\mathscr{L}}$ *in* (21) *in Theorem 1 is a function of the lifted-structure. For instance,* $\mathsf{R}_{\mathscr{L}} = \mathcal{O}(\log(N))$ *for Hankel, Toeplitz, and double-Hankel structures. For the wrap-around Hankel structure, however, we have* $\mathsf{R}_{\mathscr{L}} = \mathcal{O}(1)$.

Remark 1. *The expected number of observed elements* $|\Omega|$ *in Theorem 1 is no less than* $c\mathsf{R}_{\mathscr{L}}^2\big(\frac{\sum_n \tilde{\mu}_n}{N}\big)\tilde{K}^2 \log(N)$. *With Hoeffding's inequality, it is possible to check that the actual number of observed elements in this random setting concentrates around its expected value and is upper bounded by* $M \leq 2c\mathsf{R}_{\mathscr{L}}^2\big(\frac{\sum_n \tilde{\mu}_n}{N}\big)\tilde{K}^2 \log(N)$. *with high probability.*

Remark 2. *If* $\mathscr{L}$ *is the Hankel structure, we call the resulting method WLi-EMaC (because of the similarity of the technique with EMaC in [10]). The guaranteed sample size for exact recovery using WLi-EMaC and EMaC algorithms are* $\mathsf{R}_{\mathscr{L}}^2 c\big(\frac{\sum_n \tilde{\mu}_n}{N}\big)\tilde{K}^2 \log(N)$ *and* $c_1 \max\{\mu_n\}\tilde{K} \log^4(N)$, *respectively. As for the contribution of the leverage scores, we should emphasize that* $\max\{\mu_n\}$ *in EMaC is reduced to* $\frac{\sum_n \tilde{\mu}_n}{N}$ *in WLi-EMaC (the average score instead of the maximum score). This reduction is substantial when some complex basis of the input signal are similar to each other, i.e.,* $z_k \approx z_{k'}$ *for* $k, k' \in [K]$. *In such cases, few leverage scores become very large, causing a considerable gap between the maximum and the average leverage scores. Although the guaranteed sample size scales with* $\tilde{K}^2$ *in WLi-EMaC (compared to* $\tilde{K}$ *in EMaC), our numerical results in the form of the phase transition diagrams in Figure 1 show that the actual sample size scales with* K *and not* $\tilde{K}^2$. *We believe there is room for improvement in our theoretical analysis of the guarantees.*

Remark 3. *The constraint in* (20) *captures the incoherence condition in low-rank recovery problems that reflects both the characteristics of the lifting struc-*

ture – basis $\boldsymbol{A}_i$ – and the location of the frequencies – $\mathcal{P}_{\boldsymbol{U}}$ and $\mathcal{P}_{\boldsymbol{V}}$.

In Theorem 2, a linear error bound for the recovery in terms of the input noise level is established.

Theorem 2. *Let* $\mathbf{y}$ *and* Ω *be similarly defined to Theorem 1 and assume* (19) *holds. Further, let* $\tilde{\mathbf{y}}_\Omega = P_\Omega(\tilde{\mathbf{y}})$ *be the vector of observed noisy measurements where the noise term* $\mathbf{e} \in \mathbb{C}^N$ *satisfies* $\|P_\Omega(\mathbf{e})\|_2 \leq \sqrt{M}\eta$. *Then, with a probability no less than* $1 - N^{3-b_1}$ $(b_1 \geq 3)$, *any solution* $\widehat{\mathbf{y}}$ *of* (15) *satisfies*

$$\left\|\mathbf{W}_L\big(\mathscr{L}(\widehat{\mathbf{y}}) - \mathscr{L}(\mathbf{y})\big)\mathbf{W}_R^{\mathsf{H}}\right\|_{\mathrm{F}} \leq c_2\sqrt{M}\eta\,\frac{\min(d_1,d_2)}{\min_n p_n^2}, \tag{22}$$

where $c_2 < 102$ *is a fixed constant.*

Proof. The proof is provided in the supplementary material, Sec. II. □

Corollary 2. *In Definition 4, if* $\mathbf{W}_L \in \mathbb{R}_+^{d_1\times d_1}$ *and* $\mathbf{W}_R \in \mathbb{R}_+^{d_2\times d_2}$ *are restricted to non-negative-valued diagonal matrices, i.e.,*

$$\mathbf{W}_L = \mathrm{diag}\big(\sqrt{w_{L,1}}, \ldots \sqrt{w_{L,d_1}}\big), \quad \mathbf{W}_R = \mathrm{diag}\big(\sqrt{w_{R,1}}, \ldots, \sqrt{w_{R,d_2}}\big), \tag{23}$$

then, the leverage scores will be bounded by

$$\frac{\tilde{\mu}_n \tilde{K}}{N} \leq \max\left\{\frac{\|\mathbf{W}_L\mathbf{A}_n\|_{\mathrm{F}}^2}{\sum_{k=1}^{\lfloor \frac{N}{\beta\tilde{K}}\rfloor} w_{L,i_k}}, \frac{\left\|\mathbf{A}_n\mathbf{W}_R^{\mathsf{T}}\right\|_{\mathrm{F}}^2}{\sum_{k=1}^{\lfloor \frac{N}{\beta\tilde{K}}\rfloor} w_{R,j_k}}\right\}, \tag{24}$$

where $w_{L,i_1} \leq \ldots \leq w_{L,i_{d_1}}$ *and* $w_{R,j_1} \leq \ldots \leq w_{R,j_{d_2}}$ *are the sorted squared diagonal elements of* $\mathbf{W}_L$ *and* $\mathbf{W}_R$, *respectively, and* $\beta = \frac{N}{\tilde{K}} \max\left\{\frac{1}{\|\mathbf{U}^{\mathsf{H}}\|^2}, \frac{1}{\|\mathbf{V}^{\mathsf{H}}\|^2}\right\}$.

Proof. The proof is provided in the supplementary material, Sec. VII. □

5.2. Adjusting the Weight Matrices

On the one hand, Theorems 1 and 2 reveal a linear relationship between the guaranteed sample complexity for perfect recovery and the leverage scores. On the other hand, the leverage scores are upper-bounded in Corollary 2 by the weight matrices. Therefore, one can adjust the weight matrices such that the overall leverage scores are minimized; this, in turn, reduces the upper bound on the sample complexity. In the deterministic setup, the strategy is to interpret the actual Ω as a realization of a random sampling set with element-wise probabilities $\{p_n\}_{n\in[N]}$. If the probabilistic guarantee works for $\{p_n\}_{n\in[N]}$, then Ω as a realization of that random sampling is also suitable with high probability.

Next, we maximize the likelihood of the observed samples by tuning the set $\{p_n\}_{n\in[N]}$. Our approach is to determine weight matrices $\mathbf{W}_L, \mathbf{W}_R$ such that the likelihood of observing Ω attains its maximum point in one of the suitable random sampling strategies given in (19). This strategy results in

$$\mathbf{W}_L, \mathbf{W}_R = \underset{\substack{\mathbf{W}_L\in\mathbb{C}^{d_1\times d_1},\\ \mathbf{W}_R\in\mathbb{C}^{d_2\times d_2}}}{\operatorname{argmax}} -\sum_{n\notin\Omega} p_n \equiv \underset{\substack{\mathbf{W}_L\in\mathbb{C}^{d_1\times d_1},\\ \mathbf{W}_R\in\mathbb{C}^{d_2\times d_2}}}{\operatorname{argmin}} \sum_{n\notin\Omega} \tilde{\mu}_n. \tag{25}$$

By solving the aforementioned optimization problem, we obtain weight matrices that reduce the sample complicity or, alternatively, increase the reconstruction quality. We should note that in cases where prior knowledge about the subspace of the signal is available (i.e., $\mathbf{U}$ and $\mathbf{V}$ are known), one can set the weight matrices to maximize the reconstruction quality [26]. In this work, however, such information is not available. Instead, in [18, Section 4], we devise an optimization problem to solve the maximization in (25) and describe the whole procedure in detail.

6. Numerical Simulations

We consider two options for the lifting operator $\mathscr{L}$ in this work: Hankel and double Hankel. We refer to these two implementations of the proposed algorithm as WLi-EMaC and WLi-DEMaC because of their similarity to EMaC and DEMaC in [10] and [17], respectively. In the sequel, we present numerical experiments comparing WLi-EMaC and WLi-DEMaC against their non-weighted counterparts EMaC [10] and DEMaC [17]. We set $N = 59$ to be an odd integer which enables us to set the Hankel pencil parameter $d_1 = d_2 = 30$ so as to achieve a square matrix. For the same reason, in the DEMaC structure, we set $d_1 = d_2 = 40$.

In Figure 1, we plot the phase transition diagram for WLi-EMaC, WLi-DEMaC, DEMaC, and EMaC algorithms. For these numerical evaluations, all algorithms are implemented by CVX toolbox with SDPT3 solver in MATLAB [27]. For each pair of (M, K), the success rates are averaged over 100 Monte Carlo trial iterations. A trial is considered successful if it satisfies $\frac{\|\mathbf{x}-\hat{\mathbf{x}}\|_{\mathrm{F}}}{\|\mathbf{x}\|_{\mathrm{F}}} \leq 10^{-3}$, where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the ground truth signal and its estimated vector, respectively. The average success rates for each cell are depicted in Figure 1: brighter colors mean higher average success rates.

Results in Figure 1 imply that the required sample size M for exact recovery with both the Hankel and double Hankel structures is proportional to K in the noiseless setting. The dashed lines almost depict the transition boundary between the success and failure cases for weighted versus unweighted strategies.

This reveals that our algorithm performs better than the predicted bound in Theorem 1, where the required sample size M for exact recovery scales with K^2. By comparing the phase transitions in Figure 1, we observe that the proposed weighted approach improves the completion performance for both Hankel and Double-Hankel structures. Indeed, top row charts in Figure 1 show the supe-

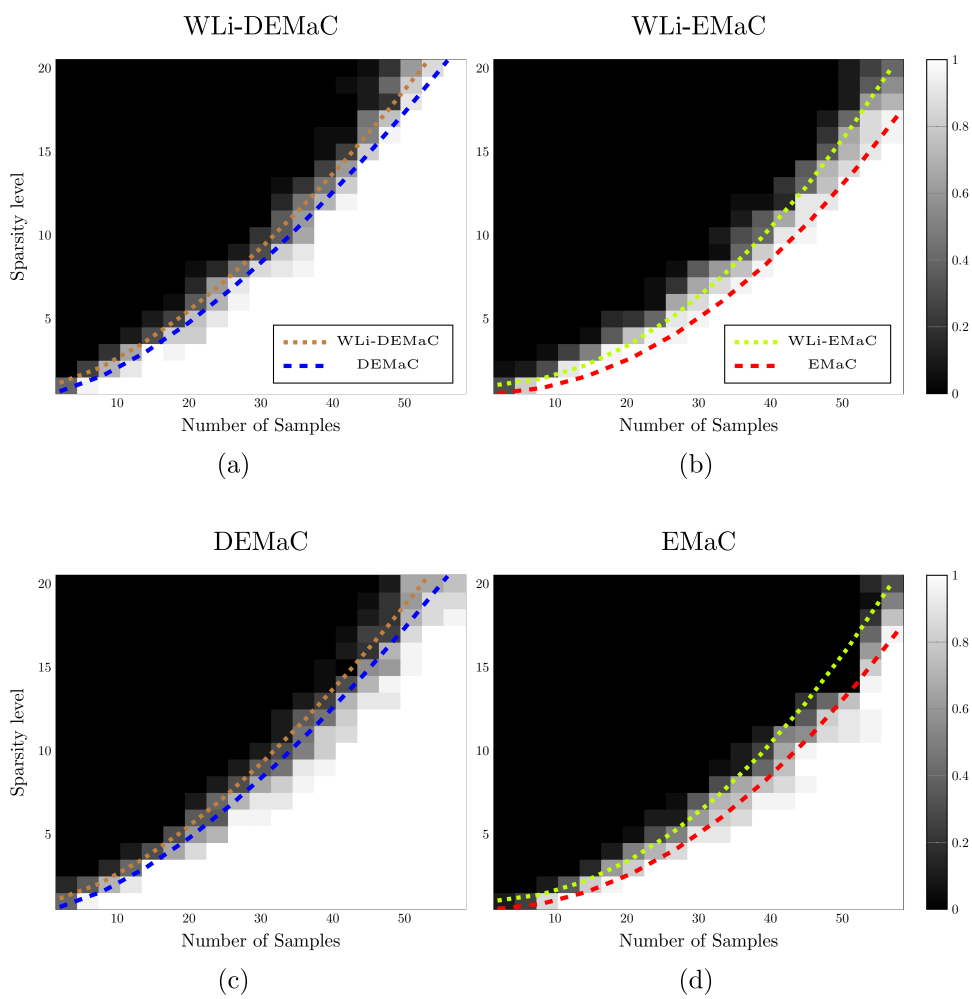

Figure 1: (a) WLi-DEMaC – (b) WLi-EMaC – (c) DEMaC – (d) EMaC phase transition diagrams in Section 6 as a function of the number of samples – x axis–, M in (11), and the sparsity level – y axis–, evaluated as number of frequency components (K) in (9). In (a) and (c), the dotted brown and the dashed blue lines approximately show the boundary of the transition between the success and failure cases for WLi-DEMaC and DEMaC, respectively. Similarly, in (b) and (d) dotted green and dashed red lines show the transition boundary between the success and failure cases for WLi-EMaC and EMaC, respectively.

riority of the WLi-DEMaC and the WLi-EMaC algorithms over the DEMaC and the EMaC algorithms, respectively. Also, Figure 1 indicates that the Double Hankel structure results in better reconstructions compared to the simple Hankel structure.

7. Conclusion

In this paper we proposed a novel approach for recovering the summation of exponential functions closely related to the line spectral estimation problem. The proposed approach comprised of three steps: 1) lifting the observed samples to a chosen structured matrix such as Hankel or Toeplitz, 2) tuning the left and right weighting matrices based on the sample informativeness, and 3) solving a weighted matrix completion problem to find the missing samples. For a given choice of the lifting structured matrix, this weighting approach generalizes other low-rank matrix completion techniques in the literature, such as EMaC (Hankel) and DEMaC (double Hankel). Both theoretical analysis and numerical results showed that the weighted lifted (WLi-) approach outperforms the case without weighting. In other words, WLi-EMaC and WLi-DEMac outperform EMaC and DEMaC in terms of NMSE.

References

[1] T. Zhang, J. M. Pauly, I. R. Levesque, Accelerating parameter mapping with a locally low rank constraint, Magnetic resonance in medicine 73 (2) (2015) 655–661.

[2] H. Pan, T. Blu, M. Vetterli, Towards generalized FRI sampling with an application to source resolution in radioastronomy, IEEE Trans. Sig. Proc. 65 (4) (2016) 821–835.

[3] D. M. Sheen, D. L. McMakin, T. E. Hall, Three-dimensional millimeter-wave imaging for concealed weapon detection, IEEE Transactions on microwave theory and techniques 49 (9) (2001) 1581–1592.

[4] M. Fatemi, A. Amini, L. Baboulaz, M. Vetterli, Shapes from pixels, IEEE Trans. Image Proc. 25 (3) (2016) 1193–1206.

[5] S. Razavikia, A. Amini, S. Daei, Reconstruction of binary shapes from blurred images via Hankel-structured low-rank matrix recovery, IEEE Trans. Image Proc. 29 (2019) 2452–2462.

[6] H. Krim, M. Viberg, Two decades of array signal processing research, IEEE Signal Process. Mag., July (1996).

[7] E. J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Communications on pure and applied Mathematics 67 (6) (2014) 906–956.

[8] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Info. Theo. 59 (11) (2013) 7465–7490.

[9] Y. Hua, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil, IEEE Trans. Sig. Proc. 40 (9) (1992) 2267–2280.

[10] Y. Chen, Y. Chi, Robust spectral compressed sensing via structured matrix completion, IEEE Trans. Info. Theo. 10 (60) (2014) 6576–6601.

[11] E. J. Candès, B. Recht, Exact matrix completion via convex optimization, Foundations of Computational Mathematics 9 (6) (2009) 717.

[12] J. C. Ye, J. M. Kim, K. H. Jin, K. Lee, Compressive sampling using annihilating filter-based low-rank interpolation, IEEE Trans. Info. Theo. 63 (2) (2016) 777–801.

[13] J.-F. Cai, X. Qu, W. Xu, G.-B. Ye, Robust recovery of complex exponential signals from random Gaussian projections via low-rank Hankel matrix reconstruction, Applied and Computational Harmonic Analysis 41 (2) (2016) 470–490.

[14] Y. Chen, S. Bhojanapalli, S. Sanghavi, R. Ward, Completing any low-rank matrix, provably, The Journal of Machine Learning Research 16 (1) (2015) 2999–3034.

[15] Y. Li, Y. Chi, Off-the-grid line spectrum denoising and estimation with multiple measurement vectors, IEEE Trans. Sig. Proc. 64 (5) (2015) 1257–1269.

[16] Z. Yang, L. Xie, On gridless sparse methods for line spectral estimation from complete and incomplete data, IEEE Trans. Sig. Proc. 63 (12) (2015) 3139–3153.

[17] Z. Yang, X. Wu, New low-rank optimization model and convex approach for robust spectral compressed sensing, arXiv preprint arXiv:2101.06433 (2021).

[18] M. Bokaei, S. Razavikia, A. Amini, S. Rini, Weighted lifted-structure low-rank matrix completion-part II: Applications for DOA estimation, Elsevier Signal Processing (2023) Submitted.

[19] K. H. Jin, D. Lee, J. C. Ye, A general framework for compressed sensing and parallel mri using annihilating filter-based low-rank hankel matrix, IEEE Trans. on Comp. Imaging 2 (4) (2016) 480–495.

[20] M. Bokaei, S. Razavikia, A. Amini, S. Rini, Two-snapshot DOA estimation via hankel-structured matrix completion, in: IEEE ICASSP, 2022, pp. 5018–5022.

[21] C. Rossant, et al., Spike sorting for large, dense electrode arrays, Nature neuroscience 19 (4) (2016) 634.

[22] J. Min, L. Carlini, M. Unser, S. Manley, J. C. Ye, Fast live cell imaging

at nanometer scale using annihilating filter-based low-rank Hankel matrix approach, Vol. 9597, SPIE, 2015, pp. 204–211.

[23] S. Razavikia, H. Zamani, A. Amini, Sampling and recovery of binary shapes via low-rank structures, in: IEEE SampTA, 2019, pp. 1–4.

[24] B. Recht, M. Fazel, P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM review 52 (3) (2010) 471–501.

[25] D. Gross, Recovering low-rank matrices from few coefficients in any basis, IEEE Trans. Info. Theo. 57 (3) (2011) 1548–1566.

[26] H. S. F. Ardakani, S. Daei, F. Haddadi, Multi-weight nuclear norm minimization for low-rank matrix recovery in presence of subspace prior information, IEEE Trans. on Sig. Proc. (2022).

[27] M. Grant, S. Boyd, CVX: MATLAB software for disciplined convex programming, version 2.1, http://cvxr.com/cvx (Mar. 2014).

Appendix A. Proof of Theorem 1

In this appendix, we shall present the proof of Theorem 1. We do so by analyzing the dual problem.

More specifically, we will construct an appropriate dual certificate; the existence of this certificate guarantees that the solution to the problem. To prove the uniqueness of the solution, we use the well-studied golfing scheme, first used in [25] to verify the existence of an approximate dual certificate. As the first step, we define the sampling operator $\mathcal{A}_n$ for any matrix $\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}$ as follows:

$$\mathcal{A}_n(\mathbf{M}) = \big\langle \mathbf{M}, \mathbf{A}_n \big\rangle \mathbf{A}_n = \mathrm{tr}\big(\mathbf{M}^{\mathsf{T}} \mathbf{A}_n\big) \mathbf{A}_n. \tag{A.1}$$

Let Ω be a random subset of $[N]$ such that the element $1 \leq n \leq N$ appears in Ω with probability p_n independent of other elements. We define the projection operator onto Ω as

$$\mathcal{A}_\Omega = \sum_{n \in [N]} \frac{\delta_n}{p_n} \mathcal{A}_n. \tag{A.2}$$

where δ_n is equal to 1 for $n \in \Omega$ and zero elsewhere and p_n is sampling probability of n-th element. We can check that $\mathbb{E}[\mathcal{A}_\Omega] = \mathcal{A}$, where $\mathcal{A}$ stands for $\sum_{n=1}^{N} \mathcal{A}_n$. It is also simple to verify that

$$\|\mathcal{A}_\Omega\| = \left\| \sum_{n \in [N]} \frac{\delta_n}{p_n} \mathcal{A}_n \right\| \leq \tfrac{1}{\min_n \{p_n\}}. \tag{A.3}$$

The projection definition of $\mathcal{A}_\Omega$ implies that for all Ω, the operator $\mathcal{A}_\Omega$ is a self-adjoint operator. Now, we reformulate the main problem in (14) as a matrix recovery problem in the lifted domain ($\mathscr{L}$). We define $\mathbf{U}$ and $\mathbf{V}$ as the left and right unitary matrices in the reduced SVD of $\mathbf{W}_L \mathscr{L}(\mathbf{M}) \mathbf{W}_R^{\mathsf{H}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathsf{H}}$. Moreover, for every $\mathbf{Y} \in \mathbb{C}^{d_1 \times d_2}$, $\mathcal{P}_U$ and $\mathcal{P}_V$ are defined as $\mathcal{P}_U(\mathbf{Y}) = \mathbf{W}_L^{\mathsf{H}} \mathbf{U} \left(\mathbf{U}^{\mathsf{H}} \mathbf{W}_L \mathbf{W}_L^{\mathsf{H}} \mathbf{U} \right)^{-1} \mathbf{U}^{\mathsf{H}} \mathbf{W}_L \mathbf{Y}$ and $\mathcal{P}_V(\mathbf{Y}) = \mathbf{Y} \mathbf{W}_R^{\mathsf{H}} \mathbf{V} \left(\mathbf{V}^{\mathsf{H}} \mathbf{W}_R \mathbf{W}_R^{\mathsf{H}} \mathbf{V} \right)^{-1} \mathbf{V}^{\mathsf{H}} \mathbf{W}_R$ respectively. A simple matrix multiplication shows that for all $\mathbf{Y}$, we have

$$\mathbf{U}^{\mathsf{H}} \mathbf{W}_L \mathcal{P}_U(\mathbf{Y}) = \mathbf{U}^{\mathsf{H}} \mathbf{W}_L \mathbf{Y}, \tag{A.4}$$

$$\mathcal{P}_V(\mathbf{Y}) \mathbf{W}_R^{\mathsf{H}} \mathbf{V} = \mathbf{Y} \mathbf{W}_R^{\mathsf{H}} \mathbf{V}. \tag{A.5}$$

We define the orthogonal operator as $\mathcal{A}^\perp = \mathcal{I} - \mathcal{A}$ where $\mathcal{I}$ is the identity operator. Then the tangent space T with respect to $\mathbf{W}_L \mathscr{L}(\mathbf{M}) \mathbf{W}_R^{\mathsf{H}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathsf{H}}$ is defined as

$$T := \{ \mathbf{W}_L^{\mathsf{H}} \mathbf{U} \mathbf{Y}_1^{\mathsf{H}} + \mathbf{Y}_2 \mathbf{V}^{\mathsf{H}} \mathbf{W}_R : \mathbf{Y}_1 \in \mathbb{C}^{d_1 \times \tilde{K}}, \mathbf{Y}_2 \in \mathbb{C}^{d_2 \times \tilde{K}} \}. \tag{A.6}$$

Also, the projection of a matrix $\mathbf{Z} \in \mathbb{C}^{d_1 \times d_2}$ onto the tangent space is denoted by $\mathcal{P}_T(\mathbf{Z})$ and we have:

$$\mathcal{P}_T(\mathbf{Z}) = \mathcal{P}_U(\mathbf{Z}) + \mathcal{P}_V(\mathbf{Z}) - \mathcal{P}_U(\mathcal{P}_V(\mathbf{Z})). \tag{A.7}$$

We can now rewrite weighted lifted-structured low-rank matrix recovery problem in (14) in form of the following general matrix completion problem:

$$\widehat{\mathbf{M}} = \underset{\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}}{\operatorname{argmin}} \quad \|\mathbf{W}_L \mathbf{M} \mathbf{W}_R^{\mathsf{H}}\|_* \quad \text{s.t.} \qquad \mathcal{Q}_\Omega(\mathbf{M}) = \mathcal{Q}_\Omega(\mathscr{L}(\mathbf{y})), \tag{A.8}$$

where $\mathcal{Q}_\Omega$ is defined as $\mathcal{Q}_\Omega = \mathcal{A}_\Omega + \mathcal{A}^\perp$. Using (A.3), we can bound $\|\mathcal{Q}_\Omega\|$ as $\|\mathcal{Q}_\Omega\| \leq \|\mathcal{A}_\Omega\| + \|\mathcal{A}^\perp\| \leq \frac{1}{\min_n p_n} + 1$.

We further have $\mathbb{E}[\mathcal{Q}_\Omega] = \mathbb{E}[\mathcal{A}_\Omega] + \mathcal{A}^\perp = \mathcal{A} + \mathcal{A}^\perp = \mathcal{I}$. As it can be seen in (A.8), scaling weight matrices does not change the problem's solution, and the matrices only need to be normalized. Hence, for simplicity of the proof, we assume $\|\mathbf{W}_L\|_{\mathrm{F}} = \|\mathbf{W}_R\|_{\mathrm{F}} = 1$.

To prove the exact recovery of the convex optimization, it suffices to produce an appropriate dual certificate, as stated in the following lemma.

Lemma 1. *For a given Ω, let the sampling operator $\mathcal{Q}_\Omega$ fulfills*

$$\|\mathcal{P}_T - \mathcal{P}_T \mathcal{Q}_\Omega \mathcal{P}_T\| \leq \frac{1}{2}, \tag{A.9}$$

if there exists a matrix $\mathbf{G}$ satisfying

$$\mathcal{Q}_\Omega^\perp(\mathbf{G}) = 0, \tag{A.10}$$

$$\|\mathcal{P}_T(\mathbf{G} - \mathbf{W}_L^{\mathsf{H}} \mathbf{U} \mathbf{V}^{\mathsf{H}} \mathbf{W}_R)\|_{\mathrm{F}} \leq \frac{1}{5\|\mathcal{Q}_\Omega\|}, \tag{A.11}$$

and

$$\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \frac{1}{2}, \tag{A.12}$$

then, $\mathbf{M}$ *is the unique solution to* (A.8)*.*

Proof. The proof is provided in the supplementary Section III. □

Lemma 1 will be satisfied, when it is sufficiently incoherent respect to the tangent space T. we bound the fluctuation of $\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T$ in the following lemma.

Lemma 2. *For a constant* $0 < \epsilon \leq \frac{1}{2}$*, if* $p_n \geq c_0\frac{\mu_n r \log(N)}{N}$ *for each* $n \in [N]$ *we have*

$$\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T\| \leq \epsilon \tag{A.13}$$

with probability exceeding $1 - N^{-b_1}$ *for sufficiently large* $c_0 \geq \frac{56}{3}(b_1 + 1)$*.*

Proof. See supplementary Section IV. □

In what follows, we show there exist a dual certificate $\mathbf{G}$ such that it satisfies conditions in (A.11) to (A.12) with high probability.

Appendix A.1. Dual Certificates Construction

We construct the dual certificate by using the golfing scheme introduced in [25]. Let $\epsilon < \frac{1}{e}$ be a small constant, and define $L := \log_{\frac{1}{\epsilon}}(N^2\|\mathcal{Q}_\Omega\|)$. Let us form L independent subsets $\{\Omega_\ell\}_{\ell=1}^L$ of $[N]$ by choosing the elements $1 \leq n \leq N$ with probability $q_n := 1 - (1 - p_n)^{\frac{1}{L}}$ independent of each other. Further, let $\overline{\Omega} = \Omega_1 \cup \cdots \cup \Omega_L$. We first check the probability that a given $1 \leq n \leq N$ belongs to $\overline{\Omega}$:

$$\mathbb{P}[n \in \overline{\Omega}] = 1 - \prod_{\ell \in [L]} (1 - p_n)^{\frac{1}{L}} = p_n. \tag{A.14}$$

Hence, $\overline{\Omega}$ fulfils the required element-wise probabilities. Next, we construct the dual certificate matrix $\mathbf{G}$ as

$$\mathbf{G} := \sum_{\ell \in [L]} \mathcal{Q}_{\Omega_\ell}(\mathbf{F}_\ell), \tag{A.15}$$

where $\mathbf{F}_\ell = \mathcal{P}_T(\mathcal{I} - \mathcal{Q}_{\Omega_\ell})\mathcal{P}_T(\mathbf{F}_{\ell-1})$ and $\mathbf{F}_0 = \mathbf{W}_L^{\mathsf{H}}\mathbf{U}\mathbf{V}^{\mathsf{H}}\mathbf{W}_R$. Since $\mathbf{F}_\ell \in \overline{\Omega}$, we conclude that $\mathcal{Q}_{\overline{\Omega}}^{\perp}(\mathbf{G}) = 0$; i.e., $\mathbf{G}$ satisfies the first condition of Lemma 1 for $\overline{\Omega}$. In addition, we have that $\mathcal{P}_T(\mathbf{F}_\ell) = \mathbf{F}_\ell = \Big(\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_{\Omega_\ell}\mathcal{P}_T\Big)(\mathbf{F}_{\ell-1})$. Besides, from (2), we know that

$$\Big\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_{\Omega_\ell}\mathcal{P}_T\Big\| \leq \epsilon < \frac{1}{2}, \tag{A.16}$$

with a probability no less than $1 - N^{-b_1}$. To bound $\big\|\mathcal{P}_T\big(\mathbf{G} - \mathbf{F}_0\big)\big\|_{\mathrm{F}}$, we use a similar technique as in [10] to obtain $\mathcal{P}_T(\mathbf{G} - \mathbf{F}_0) = -\mathcal{P}_T(\mathbf{F}_L)$. The latter holds due to $q_\ell \geq \frac{p_\ell}{L} \geq c_0 \mathsf{R}_{\mathscr{L}}^2 \frac{\tilde{\mu}_\ell \tilde{K}^2}{N}$. Now, we are able to write

$$\|\mathcal{P}_T\big(\mathbf{G} - \mathbf{F}_0\big)\|_{\mathrm{F}} = \|\mathcal{P}_T\big(\mathbf{F}_L\big)\|_{\mathrm{F}} \leq \epsilon^L \|\mathcal{P}_T(\mathbf{F}_0)\|_{\mathrm{F}} < \frac{1}{5\|\mathcal{Q}_\Omega\|}, \tag{A.17}$$

with a probability no less than $1 - LN^{-b_1}$. This shows that $\mathbf{G}$ satisfies condition (A.11) of Lemma 1 with high probability.

Appendix A.2. Some relevant lemmas

We begin by defining the following two norms for the matrix $\mathbf{M} \in \mathbb{C}^{d_1 \times d_2}$ for a given set of lifting basis $\{\mathbf{A}_n\}_{n \in [N]}$:

$$\|\mathbf{M}\|_{\mathcal{A},\infty} := \max_{n\in[N]} \left| \frac{N\langle \mathbf{A}_n, \mathbf{M}\rangle}{\tilde{K}\tilde{\mu}_n\sqrt{\omega_n}} \right|, \tag{A.18a}$$

$$\|\mathbf{M}\|_{\mathcal{A},2} := \sqrt{\sum_{n\in[N]} \frac{|N\langle \mathbf{A}_n, \mathbf{M}\rangle|^2}{\tilde{K}\tilde{\mu}_n\omega_n}}, \tag{A.18b}$$

where we have defined $\omega_n := \|\mathbf{A}_n\|_0$.

We now state three inequalities regarding the norms in (A.18): Lemma 3, Lemma 4, and Lemma 5. Generally, these proofs rely on matrix concentration inequalities in [10, Appendix A].

Lemma 3. *Suppose* $\mathbf{M}$ *is a complex-valued* $d_1 \times d_2$ *matrix. If* $p_n \geq c_0 \frac{\tilde{\mu}_n \tilde{K}^2 \log(N)}{N}$ *for all* $n \in [N]$*, then*

$$\left\|\left(\mathcal{Q}_\Omega - \mathcal{I}\right)\mathbf{M}\right\| \leq \sqrt{\tfrac{2(b_2+1)}{c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}}\|\mathbf{M}\|_{\mathcal{A},2} + \tfrac{2(b_2+1)}{3c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}\|\mathbf{M}\|_{\mathcal{A},\infty},$$

holds with a probability at least $1 - N^{-b_2}$*, where* $c_0 \geq 2(b_2+1)$ *and* $\mathsf{R}_{\mathscr{L}} = \sum_{n\in[N]} \left\|\mathbf{A}_n \odot \mathbf{A}_n\right\|_{\infty\to\infty}$.

Proof. See supplementary Section V. □

We further control $\|.\|_{\mathcal{A},2}$ and $\|.\|_{\mathcal{A},\infty}$ norms of $\left(\mathcal{P}_T\mathcal{Q}_\Omega - \mathcal{P}_T\right)$ in the next two lemmas.

Lemma 4. *For* $c_0 \geq 16(b_3+1)$ *and arbitrary* $\mathbf{M} \in \mathbb{C}^{d_1\times d_2}$*, we have*

$$\left\|\left(\mathcal{P}_T\mathcal{Q}_\Omega - \mathcal{P}_T\right)(\mathbf{M})\right\|_{\mathcal{A},2} \leq 2\left(\sqrt{\tfrac{2(b_3+1)}{c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}}\|\mathbf{M}\|_{\mathcal{A},2} + \tfrac{2(b_3+1)}{3c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}\|\mathbf{M}\|_{\mathcal{A},\infty}\right),$$

with a probability no less than $1 - N^{-b_3}$*, given that* $p_n \geq c_0 \frac{\tilde{\mu}_n \mathsf{R}_{\mathscr{L}} \tilde{K}^2}{N} \log(N)$ *for* $n \in [N]$.

Proof. See Supplementary Section VI. □

Lemma 5. *Suppose we have that*

$$\frac{1}{8\sqrt{\log(N)}} \leq \min_{i\in[N]} \Big\{ \|\mathbf{A}_i\|_0 \min\{\|\mathcal{P}_U(\mathbf{A}_i)\|_{\mathrm{F}}^2, \|\mathcal{P}_V(\mathbf{A}_i)\|_{\mathrm{F}}^2\} \Big\}. \tag{A.19}$$

Then, for $c_0 \geq 144(b_4+1)$ and arbitrary $\mathbf{M} \in T$, we have

$$\Big\| \Big(\mathcal{P}_T \mathcal{Q}_\Omega - \mathcal{P}_T\Big)(\mathbf{M}) \Big\|_{\mathcal{A},\infty} \leq \sqrt{72}\Big(\sqrt{\tfrac{2(b_4+1)}{c_0}}\, \|\mathbf{M}\|_{\mathcal{A},2} + \tfrac{2(b_4+1)}{3c_0}\, \|\mathbf{M}\|_{\mathcal{A},\infty} \Big),$$

with probability at least $1 - N^{-b_4+1}$, given that $p_n \geq c_0 \frac{\tilde{\mu}_n \tilde{K}^2}{N} \log(N)$ for $n \in [N]$.

Proof. The proof is provided in Supplementary Sec. 7 of this document. □

In the next subsection we find the upper bound on $\|\mathcal{P}_{T^\perp}(\mathbf{G})\|$ by using the stated lemmas and defined norms in (A.18).

Appendix A.3. Upper bound derivation

Having introduced the norms in (A.18) and the lemmas in Sec. Appendix A.2, we can now return to our main objective: upper bounding $\|\mathcal{P}_{T^\perp}(\mathbf{G})\|$. Note that all lemmas in Sec. Appendix A.2 hold for some universal constant c_0. For convenience, in the remainder of the proof we shall take the value of c_0 for which all bounds in these lemmas hold, that is, we take

$$c_0 \geq \max\big\{2(b_2+1)\,,\, 16(b_3+1)\,,\, 144(b_4+1)\,,\, \tfrac{53}{3}(b_1+1)\big\}. \tag{A.20}$$

Recalling (60) (in the main manuscript), we have

$$\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \sum_{\ell\in[L]} \|\mathcal{P}_{T^\perp} \mathcal{Q}_{\Omega_\ell} \mathcal{P}_T(\mathbf{F}_{\ell-1})\|. \tag{A.21}$$

Next, we bound each term in the right hand summation of (A.21) as

$$
\begin{aligned}
&\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_\ell}\mathcal{P}_T(\mathbf{F}_{\ell-1})\| = \left\|\big(\mathcal{P}_{T^\perp}(\mathcal{Q}_{\Omega_\ell}-\mathcal{I})\mathcal{P}_T\big)(\mathbf{F}_{\ell-1})\right\| \\
&\leq \left\|\big((\mathcal{Q}_{\Omega_\ell}-\mathcal{I})\mathcal{P}_T\big)(\mathbf{F}_{\ell-1})\right\| = \|(\mathcal{Q}_{\Omega_\ell}-\mathcal{I})\,(\mathbf{F}_{\ell-1})\| \\
&\overset{(a)}{\leq} \sqrt{\tfrac{2(b_2+1)}{c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}}\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},2} + \tfrac{2(b_2+1)}{3c_0\tilde{K}\mathsf{R}_{\mathscr{L}}}\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},\infty} \\
&\leq \frac{\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},2}+\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},\infty}}{c_1\sqrt{\tilde{K}\mathsf{R}_{\mathscr{L}}}},
\end{aligned}
\tag{A.22}
$$

with probability $1-N^{-b_2}$ where in (a) we use Lemma 3 and for

$$
c_1 = \min\left\{\tfrac{3c_0\sqrt{\tilde{K}\mathsf{R}_{\mathscr{L}}}}{2(b_2+1)}, \sqrt{\tfrac{c_0}{2(b_2+1)}}\right\}.
$$

Thus,

$$
\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \tfrac{1}{c_1\sqrt{\tilde{K}\mathsf{R}_{\mathscr{L}}}}\sum_{\ell\in[L]}\big(\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},2}+\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},\infty}\big) \tag{A.23}
$$

holds with probability no less than $1-LN^{-b_2}$. Since $\mathbf{F}_\ell = \Big(\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_{\Omega_\ell}\Big)(\mathbf{F}_{\ell-1})$, we use Lemmas 4 and 5 to recursively bound $\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_\ell}\mathcal{P}_T(\mathbf{F}_{\ell-1})\|$ as

$$
\|\mathbf{F}_\ell\|_{\mathcal{A},2}+\|\mathbf{F}_\ell\|_{\mathcal{A},\infty} \leq \big(\sqrt{\tfrac{8(b_3+1)}{c_0}}+\sqrt{\tfrac{144(b_4+1)}{c_0}}\big)\,\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},2} \tag{A.24a}
$$

$$
+\big(\tfrac{4(b_3+1)}{3c_0}+\tfrac{2\sqrt{72}(b_4+1)}{3c_0}\big)\,\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},\infty} \leq \frac{\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},2}+\|\mathbf{F}_{\ell-1}\|_{\mathcal{A},\infty}}{c_2}, \tag{A.24b}
$$

with probability no less than $1-N^{-b_3}-N^{-b_4+1}$, where

$$
c_2 = \min\left\{\frac{1}{\sqrt{\frac{8(b_3+1)}{c_0}}+\sqrt{\frac{144(b_4+1)}{c_0}}}, \frac{1}{\frac{4(b_3+1)}{3c_0}+\frac{2\sqrt{72}(b_4+1)}{3c_0}}\right\}.
$$

By applying (A.24a) multiple times, we conclude that

$$\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \frac{\|\mathbf{F}_0\|_{\mathcal{A},2} + \|\mathbf{F}_0\|_{\mathcal{A},\infty}}{c_1\sqrt{\tilde{K}\mathsf{R}_{\mathscr{L}}}} \sum_{\ell\in[L]} c_2^{1-\ell}, \tag{A.25}$$

with probability no less than $1 - LN\sum_{i=2}^{4} N^{-b_i}$. We further bound $\|\mathbf{F}_0\|_{\mathcal{A},\infty}$ and $\|\mathbf{F}_0\|_{\mathcal{A},2}$ to simplify (A.25). To bound $\|\mathbf{F}_0\|_{\mathcal{A},\infty}$, we first recall that

$$\|\mathbf{F}_0\|_{\mathcal{A},\infty} = \max_{n\in[N]} \left| \frac{\langle \mathbf{A}_n\,, \overbrace{\mathbf{W}_L^{\mathsf{H}}\mathbf{U}\mathbf{V}^{\mathsf{H}}\mathbf{W}_R}^{\mathbf{F}_0}\rangle N}{\sqrt{\omega_n}\tilde{\mu}_n\tilde{K}} \right|. \tag{A.26}$$

In addition, we know

$$\begin{aligned}
\langle \mathbf{A}_n\,, \mathbf{W}_L^{\mathsf{H}}\mathbf{U}\mathbf{V}^{\mathsf{H}}\mathbf{W}_R\rangle &= \langle \mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathbf{A}_n\,,\, \mathbf{V}^{\mathsf{H}}\mathbf{W}_R\rangle \\
&= \sqrt{\omega_n}\big\langle \mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathbf{A}_n\,,\, \mathbf{V}^{\mathsf{H}}\mathbf{W}_R\mathbf{A}_n^R\big\rangle \\
&= \sqrt{\omega_n}\big\langle \mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathcal{P}_U(\mathbf{A}_n)\,,\, (\mathcal{P}_V(\mathbf{A}_n^R)\mathbf{W}_R^{\mathsf{H}}\mathbf{V})^{\mathsf{H}}\big\rangle.
\end{aligned} \tag{A.27}$$

where $\mathbf{A}_n^R$ is $d_2 \times d_2$ right diagonal version of $\mathbf{A}_n$, in which for each column its diagonal element is equal to norm-one of that column. $\mathbf{A}_n^R$ comes from the fact that $\mathbf{A}_n$s are orthonormal basis.

$$\big\|\mathbf{U}^{\mathsf{H}}\mathbf{W}_L\mathcal{P}_U(\mathbf{A}_n)\big\|_{\mathrm{F}} \leq \underbrace{\|\mathbf{U}^{\mathsf{H}}\|}_{\leq 1}\underbrace{\|\mathbf{W}_L\|_{\mathrm{F}}}_{=1}\underbrace{\|\mathcal{P}_U(\mathbf{A}_n)\|_{\mathrm{F}}}_{\leq\sqrt{\frac{\mu_k\tilde{K}}{N}}}, \tag{A.28}$$

and

$$\big\|(\mathcal{P}_V(\mathbf{A}_n^R)\mathbf{W}_R^{\mathsf{H}}\mathbf{V})^{\mathsf{H}}\big\|_{\mathrm{F}} \leq \underbrace{\|\mathbf{V}^{\mathsf{H}}\|}_{\leq 1}\underbrace{\|\mathbf{W}_R\|_{\mathrm{F}}}_{=1}\underbrace{\|\mathcal{P}_V(\mathbf{A}_n^R)\|_{\mathrm{F}}}_{\leq\sqrt{\frac{\mu_k\tilde{K}}{N}}}. \tag{A.29}$$

Now, if we apply the Cauchy-Schwartz inequality in (A.27) by using (A.28) and

(A.29), we obtain

$$\left|\langle \mathbf{A}_n\,, \mathbf{W}_L^{\mathsf{H}} \mathbf{U} \mathbf{V}^{\mathsf{H}} \mathbf{W}_R \rangle\right| \leq \frac{\sqrt{\omega_k}\mu_k \tilde{K}}{N}. \tag{A.30}$$

By plugging this result into (A.26), we get $\|\mathbf{F}_0\|_{\mathcal{A},\infty} \leq 1$. For bounding $\|\mathbf{F}_0\|_{\mathcal{A},2}$, we use

$$\begin{aligned} \|\mathbf{F}_0\|_{\mathcal{A},2}^2 &= \sum_{n\in[N]} \tfrac{N\,|\langle \mathbf{A}_n, \mathbf{F}_0\rangle|^2}{\omega_n \tilde{\mu}_n \tilde{K}} = \sum_{n\in[N]} \tfrac{\tilde{\mu}_n \tilde{K}}{N} \left(\tfrac{N\,|\langle \mathbf{A}_n, \mathbf{F}_0\rangle|}{\sqrt{\omega_n} \tilde{\mu}_n \tilde{K}} \right)^2 \\ &\leq \sum_{n\in[N]} \frac{\tilde{\mu}_n \tilde{K}}{N} \leq \sum_{n\in[N]} \|\mathcal{P}_U(\mathbf{A}_n)\|_{\mathrm{F}}^2 + \|\mathcal{P}_V(\mathbf{A}_n)\|_{\mathrm{F}}^2 \end{aligned} \tag{A.31}$$

By invoking (24), we further bound $\sum_{n\in[N]} \|\mathcal{P}_U(\mathbf{A}_n)\|_{\mathrm{F}}^2$ as

$$\begin{aligned} \sum_{n\in[N]} \|\mathcal{P}_U(\mathbf{A}_n)\|_{\mathrm{F}}^2 &\leq \Big\| \sum_{n\in[N]} (\mathbf{A}_n \odot \mathbf{A}_n) \Big\|_{\infty\to\infty} \|\mathcal{P}_U(\mathbf{1})\|_{\mathrm{F}}^2 \\ &\leq \underbrace{\sum_{n\in[N]} \|\mathbf{A}_n \odot \mathbf{A}_n)\|_{\infty\to\infty}}_{\mathsf{R}_{\mathscr{L}}} \tilde{K} \leq \mathsf{R}_{\mathscr{L}} \tilde{K}. \end{aligned} \tag{A.32}$$

A similar approach shows that $\sum_{n\in[N]} \|\mathcal{P}_V(\mathbf{A}_n)\|_{\mathrm{F}}^2 \leq \tilde{K}\mathsf{R}_{\mathscr{L}}$. Hence, $\sum_{n\in[N]} \frac{|\langle \mathbf{A}_n, \mathbf{F}_0\rangle|^2 N}{\omega_n \tilde{\mu}_n \tilde{K}} \leq 2\tilde{K}\mathsf{R}_{\mathscr{L}}$, or $\|\mathbf{F}_0\|_{\mathcal{A},2}^2 \leq 2\tilde{K}\mathsf{R}_{\mathscr{L}}$. We now get back to (A.25):

$$\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \frac{\sqrt{2\tilde{K}\mathsf{R}_{\mathscr{L}}} + 1}{c_1 \sqrt{\tilde{K}\mathsf{R}_{\mathscr{L}}}} \sum_{\ell\in[L]} c_2^{1-\ell} \leq \frac{2\sqrt{2}}{c_1} \sum_{\ell\in[L]} c_2^{1-\ell} \tag{A.33}$$

for $q_n \geq c_0 \mathsf{R}_{\mathscr{L}}^2 \frac{\tilde{\mu}_n}{N} \tilde{K}^2$, or equivalently $p_n \geq c_0 \mathsf{R}_{\mathscr{L}}^2 \frac{\tilde{\mu}_n}{N} \tilde{K}^2 \log(N)$. For $c_2 \geq 2$ and $c_1 \geq 12$, we can conclude that

$$\|\mathcal{P}_{T^\perp}(\mathbf{G})\| \leq \tfrac{2\sqrt{2}}{c_1} \Big(1 + \sum_{\ell=1}^{\infty} (\tfrac{1}{2})^\ell \Big) \leq \tfrac{4\sqrt{2}}{c_1} \leq \frac{1}{2}, \tag{A.34}$$

with probability at least $1 - LN \sum_{i=2}^{4} N^{-b_i}$. Therefore, if $p_n \geq c_0 \mathsf{R}_{\mathscr{L}}^2 \frac{\tilde{\mu}_n}{N} \tilde{K}^2 \log(N)$

for $n \in [N]$, with probability no less than $1 - LN\sum_{i=1}^{4} N^{-b_i}$, matrix $\mathbf{G}$ is a valid dual certificate. As a result, from Lemma 1, the solution of weighted lifted-structured low-rank matrix recovery problem is exact and unique, with high probability.