

Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction

Jian-Feng Cai*

Xiaobo Qu[†]

Weiyu Xu[‡]

Gui-Bo Ye*

Abstract

This paper explores robust recovery of a superposition of R distinct complex exponential functions from a few random Gaussian projections. We assume that the signal of interest is of $2N - 1$ dimensional and $R \ll 2N - 1$. This framework covers a large class of signals arising from real applications in biology, automation, imaging science, etc. To reconstruct such a signal, our algorithm is to seek a low-rank Hankel matrix of the signal by minimizing its nuclear norm subject to the consistency on the sampled data. Our theoretical results show that a robust recovery is possible as long as the number of projections exceeds $O(R \ln^2 N)$. No incoherence or separation condition is required in our proof. Our method can be applied to spectral compressed sensing where the signal of interest is a superposition of R complex sinusoids. Compared to existing results, our result here does not need any separation condition on the frequencies, while achieving better or comparable bounds on the number of measurements. Furthermore, our method provides theoretical guidance on how many samples are required in the state-of-the-art non-uniform sampling in NMR spectroscopy. The performance of our algorithm is further demonstrated by numerical experiments.

1 Introduction

Many practical problems involve signals that can be modeled or approximated by a superposition of a few complex exponential functions. In particular, if we choose the exponential function to be complex sinusoid, it covers signals in acceleration of medical imaging [15], analog-to-digital conversion [20], inverse scattering in seismic imaging [1], etc. Time domain signals in nuclear magnetic resonance (NMR) spectroscopy, that are widely used to analyze the compounds in chemistry and protein structures in biology, are another type of signals that can be modeled or approximated by a superposition of complex exponential functions [16]. How to recover those superposition of complex exponential functions is of primary importance in those applications.

In this paper, we will consider how to recover those complex exponentials from linear measurements of their superposition. More specifically, let $\hat{\mathbf{x}} \in \mathbb{C}^{2N-1}$ be a vector satisfying

$$\hat{x}_j = \sum_{k=1}^R c_k z_k^j, \quad j = 0, 1, \dots, 2N - 2, \quad (1)$$

where $z_k \in \mathbb{C}$, $k = 1, \dots, R$, are some unknown complex numbers. In other words, $\hat{\mathbf{x}}$ is a superposition of R exponential functions. We assume $R \ll 2N - 1$. When $|z_k| = 1$, $k = 1, \dots, R$, $\hat{\mathbf{x}}$ is a superposition of complex sinusoids. When $z_k = e^{-\tau_k} e^{2\pi i f_k}$, $k = 1, \dots, R$, $\hat{\mathbf{x}}$ models the signal in NMR spectroscopy.

*Department of Mathematics, University of Iowa, Iowa City, IA 52242. Email: {jianfeng-cai, guibo-ye}@uiowa.edu

[†]Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, State Key Laboratory of Physical Chemistry of Solid Surfaces, Xiamen University, P.O. Box 979, Xiamen 361005 (China). Email: quxiaobo@xmu.edu.cn

[‡]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242. Email: weiyu-xu@uiowa.edu

Since $R \ll 2N - 1$, the degree of freedom to determine $\hat{\mathbf{x}}$ is much less than the ambient dimension $2N - 1$. Therefore, it is possible to recover $\hat{\mathbf{x}}$ from its under-sampling [3,5,8,12]. In particular, we consider to recover $\hat{\mathbf{x}}$ from its linear measurement

$$\mathbf{b} = \mathcal{A}\hat{\mathbf{x}}, \quad (2)$$

where $\mathcal{A} \in \mathbb{C}^{M \times (2N-1)}$ with $M \ll 2N - 1$.

We will use a Hankel structure to reconstruct the signal of interest $\hat{\mathbf{x}}$. The Hankel structure originates from the matrix pencil method [14] for harmonic retrieval for complex sinusoid. The conventional matrix pencil method assumes fully observed $\hat{\mathbf{x}}$ as well as the model order R , which are both unknown here. Following the ideas of the matrix pencil method in [14] and enhanced matrix completion (EMaC) in [10], we construct a Hankel matrix based on signal $\hat{\mathbf{x}}$. More specifically, define the Hankel matrix $\hat{\mathbf{H}} \in \mathbb{C}^{N \times N}$ by

$$\hat{H}_{jk} = \hat{x}_{j+k}, \quad j, k = 0, 1, \dots, N - 1. \quad (3)$$

Throughout this paper, indices of all vectors and matrices start from 0, instead of 1 in conventional notations. It can be shown that $\hat{\mathbf{H}}$ is a matrix with rank R . Instead of reconstructing $\hat{\mathbf{x}}$ directly, we reconstruct the rank- R Hankel matrix $\hat{\mathbf{H}}$, subject to the constraint that (2) is satisfied.

Low rank matrix recovery has been widely studied [2,5,6,17]. It is well known that minimizing the nuclear norm tends to lead to a solution of low-rank matrices. Therefore, a nuclear norm minimization problem subject to the constraint (2) is proposed. More specifically, for any given $\mathbf{x} \in \mathbb{C}^{2N-1}$, let $\mathbf{H}(\mathbf{x}) \in \mathbb{C}^{N \times N}$ be the Hankel matrix whose first row and last column is \mathbf{x} , i.e., $[\mathbf{H}(\mathbf{x})]_{jk} = x_{j+k}$. We propose to solve

$$\min_{\mathbf{x}} \|\mathbf{H}(\mathbf{x})\|_*, \quad \text{subject to } \mathcal{A}\mathbf{x} = \mathbf{b}, \quad (4)$$

where $\|\cdot\|_*$ is the nuclear norm function (the sum of all singular values), and \mathcal{A} and \mathbf{b} are from the linear measurement (2). When there is noise contained in the observation, i.e.,

$$\mathbf{b} = \mathcal{A}\hat{\mathbf{x}} + \boldsymbol{\eta},$$

we solve

$$\min_{\mathbf{x}} \|\mathbf{H}(\mathbf{x})\|_*, \quad \text{subject to } \|\mathcal{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta, \quad (5)$$

where $\delta = \|\boldsymbol{\eta}\|_2$ is the noise level.

An important theoretical question is how many measurements are required to get a robust reconstruction of $\hat{\mathbf{H}}$ via (4) or (5). For a generic unstructured $N \times N$ matrix of rank R , standard theory [6, 7, 9, 17] indicates that $O(NR \cdot \text{poly}(\log N))$ measurements are needed for a robust reconstruction by nuclear norm minimization. This result, however, is unacceptable here since the number of parameters of $\hat{\mathbf{H}}$ is only $2N - 1$. The main contribution of this paper is then to prove that (4) and (5) give a robust recovery of $\hat{\mathbf{H}}$ (hence $\hat{\mathbf{x}}$) as soon as the number of projections exceeds $O(R \ln^2 N)$ if we choose the linear operator \mathcal{A} to be some scaled random Gaussian projections. This result is further extended to the robust reconstruction of low-rank Hankel or Toeplitz matrices from its few Gaussian random projections.

Our result can be applied to various signals of superposition of complex exponentials, including, but not limited to, signals of complex sinusoids and signals in accelerated NMR spectroscopy. When applied to complex sinusoids, our result here does not need any separation condition on the frequencies, while achieving better or comparable bounds on the number of required measurements. Furthermore, our theoretical result provides some guidance on how many samples to choose for the model proposed in [16] to recover NMR spectroscopy.

- **Complex sinusoids.** When $|z_k| = 1$ for $k = 1, \dots, R$, we must have $z_k = e^{2\pi i f_k}$ for some frequency f_k . In this case, $\hat{\mathbf{x}}$ is a superposition of complex sinusoids, for examples, in the analog-to-digital conversion of radio signals [20]. The problem on recovering $\hat{\mathbf{x}}$ from its as few as possible linear measurements (2) may be solved using compressed sensing (CS) [8]. One can discretize the domain of frequencies f_k by a uniform grid. When the frequencies f_k indeed fall on the grid, $\hat{\mathbf{x}}$ is sparse in the discrete Fourier

transform domain, and CS theory [8, 12] suggests that it is possible to reconstruct $\hat{\mathbf{x}}$ from its very few samples via ℓ_1 -norm minimization, provided that $R \ll 2N - 1$. Nevertheless, the frequencies f_k in our setting usually do not exactly fall on a grid. The basis mismatch between the true parameters and the grid based on discretization degenerates the performance of conventional compressed sensing [11].

To overcome this, [4, 19] proposed to recover off-the-grid complex sinusoid frequencies using total variation minimization or atomic norm [9] minimization. They proved that the total variation minimization or atomic norm minimization can have a robust reconstruction of $\hat{\mathbf{x}}$ from a nonuniform sampling of very few entries of $\hat{\mathbf{x}}$, provided that the frequencies f_k , $k = 1, \dots, R$, has a good separation. Another method for recovering off-the-grid frequencies is enhanced matrix completion (EMaC) proposed by Chen et al [10], where the Hankel structure plays a central role similar to our model. The main result in [10] is that the complex sinusoids $\hat{\mathbf{x}}$ can be robustly reconstructed via EMaC from its very few nonuniformly sampled entries. Again, the EMaC requires a separation of the frequencies, described implicitly by an incoherence condition.

When applied to complex sinusoids, compared to the aforementioned existing results, our result here does not need any separation condition on the frequencies, while achieving better or comparable bound of number of measurements.

- **Accelerated NMR spectroscopy.** When $z_k = e^{-\tau_k} e^{2\pi i f_k}$, $k = 1, \dots, R$, $\hat{\mathbf{x}}$ models the signal in NMR spectroscopy, which arises frequently in studying short-lived molecular systems, monitoring chemical reactions in real-time, high-throughput applications, etc. Recently, Qu et al [16] proposed an algorithm based on low rank Hankel matrix. In this specific application, \mathcal{A} is a matrix that denotes the under-sampling of NMR signals in the time domain. Numerical results show its efficiency in [16] while theoretical results are still needed to explain. It is vital to give some theoretical results on this model since it will give us some guidance on how many samples should be chosen to guarantee the robust recovery. Though the result in [10] applies to this problem, it needs an incoherence condition, which remains uncertain for diverse chemical and biology samples. Our result in this paper does not require any incoherence condition. Moreover, our bound is better than that in [10].

The rest of this paper is organized as follows. We begin with our model and our main results in Section 2. Proofs for the main result are given in Section 3. Then, in Section 4, we extend the main result to the reconstruction of generic low-rank Hankel or Toeplitz matrices. Finally, the performance of our algorithm is demonstrated by numerical experiments in Section 5.

2 Model and Main Results

Our approach is based on the observation that the Hankel matrix whose first row and last column consist of entries of $\hat{\mathbf{x}}$ has rank R . Let $\hat{\mathbf{H}}$ be the Hankel matrix defined by (3). Eq. (1) leads to a decomposition

$$\hat{\mathbf{H}} = \begin{bmatrix} 1 & \dots & 1 \\ z_1 & \dots & z_R \\ \vdots & \vdots & \vdots \\ z_1^{N-1} & \dots & z_R^{N-1} \end{bmatrix} \begin{bmatrix} c_1 & & \\ & \ddots & \\ & & c_R \end{bmatrix} \begin{bmatrix} 1 & z_1 \dots & z_1^{N-1} \\ \vdots & \vdots & \vdots \\ 1 & z_R \dots & z_R^{N-1} \end{bmatrix}$$

Therefore, the rank of $\hat{\mathbf{H}}$ is R . Similar to Enhanced Matrix Completion (EMaC) in [10], in order to reconstruct $\hat{\mathbf{x}}$, we first reconstruct the rank- R Hankel matrix $\hat{\mathbf{H}}$, subject to the constraint that (2) is satisfied. Then, $\hat{\mathbf{x}}$ is derived directly by choosing the first row and last column of $\hat{\mathbf{H}}$. More specifically, for any given $\mathbf{x} \in \mathbb{C}^{2N-1}$, let $\mathbf{H}(\mathbf{x}) \in \mathbb{C}^{N \times N}$ be the Hankel matrix whose first row and last column is \mathbf{x} , i.e., $[\mathbf{H}(\mathbf{x})]_{jk} = x_{j+k}$. We propose to solve

$$\min_{\mathbf{x}} \text{rank}(\mathbf{H}(\mathbf{x})), \quad \text{subject to } \mathcal{A}\mathbf{x} = \mathbf{b}, \quad (6)$$

where $\text{rank}(\mathbf{H}(\mathbf{x}))$ denotes the rank of $\mathbf{H}(\mathbf{x})$, and \mathcal{A} and \mathbf{b} are from the linear measurement (2). When there is noise contained in the observation, i.e, $\mathbf{b} = \mathcal{A}\hat{\mathbf{x}} + \eta$, we correspondingly solve

$$\min_{\mathbf{x}} \text{rank}(\mathbf{H}(\mathbf{x})), \quad \text{subject to} \quad \|\mathcal{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta, \quad (7)$$

where $\delta = \|\eta\|_2$ is the noise level.

These two problems are all NP hard problems and not easy to solve. Following the ideas of matrix completion and low rank matrix recovery [6, 7, 9, 17], it is possible to exactly recover the low rank Hankel matrix via nuclear norm minimization. Therefore, it is reasonable to use nuclear norm minimization for our problem and it leads to the models in (4) and (5).

Intuitively, our model is reasonable and likely to work. Theoretical results are desirable to guarantee it. The results in [6, 7, 9, 17] do not consider the Hankel structure. For generic $N \times N$ rank- R matrix, they requires $O(NR \cdot \text{poly}(\log N))$ measurements for robust recovery which is too much since there are only $2N - 1$ degrees of freedom in $\mathbf{H}(\mathbf{x})$. The theorems proposed in [19] work only for a special case where signals of interest are superpositions of complex sinusoids, which excludes, e.g., the signals in NMR spectroscopy. While the results from [10] extend to complex exponentials, the performance guarantees in [4, 10, 19] require incoherence conditions, implying the knowledge of frequency interval in spectroscopy, which are not available before the realistic sampling of diverse chemical or biological samples. This limits the applicability of these theories.

It is challenging to provide a theorem guaranteeing the exact recovery for model (4) with arbitrarily linear measurements \mathcal{A} . In this paper, we provide a theoretical result ensuring exact recovery when \mathcal{A} is a scaled random Gaussian matrix. Our result does not assume any incoherence conditions on the original signal.

Theorem 1. *Let $\mathcal{A} = \mathcal{B}\mathcal{D} \in \mathbb{C}^{M \times (2N-1)}$, where $\mathcal{B} \in \mathbb{C}^{M \times (2N-1)}$ is a random matrix whose real and imaginary parts are i.i.d. Gaussian with mean 0 and variance 1, $\mathcal{D} \in \mathbb{R}^{(2N-1) \times (2N-1)}$ is a diagonal matrix with the j -th diagonal $\sqrt{j+1}$ if $j \leq N-1$ and $\sqrt{2N-1-j}$ otherwise. Then, there exists a universal constant $C_1 > 0$ such that, for an arbitrary $\epsilon > 0$, If*

$$M \geq (C_1 \sqrt{R} \ln N + \sqrt{2}\epsilon)^2 + 1,$$

then, with probability at least $1 - 2e^{-\frac{M-1}{8}}$, we have

- (a) $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is the unique solution of (4) with $\mathbf{b} = \mathcal{A}\hat{\mathbf{x}}$;
- (b) $\|\mathcal{D}(\tilde{\mathbf{x}} - \hat{\mathbf{x}})\|_2 \leq 2\delta/\epsilon$, where $\tilde{\mathbf{x}}$ is the unique solution of (5) with $\|\mathbf{b} - \mathcal{A}\hat{\mathbf{x}}\|_2 \leq \delta$.

The number of measurements required is $O(R \ln^2 N)$, which is reasonable small compared with the number of parameters in $\mathbf{H}(\mathbf{x})$. Furthermore, there is a parameter ϵ in Theorem 1. For the noise-free case (a), the best choice of ϵ is obviously a number that is very close to 0. For the noisy case (b), we can balance the error bound and the number of measurements to get an optimal ϵ . On the one hand, according to the result in (b), in order to make the error in noisy case as small as possible, we would like ϵ to be as large as possible. On the other hand, we would like to keep the measurements M of the order of $R \ln^2 N$. Therefore, a seemingly optimal choice of ϵ is $\epsilon = O(R \ln^2 N)$. With this choice of ϵ , the number of measurements $M = O(R \ln^2 N)$ and the error $\|\mathcal{D}(\tilde{\mathbf{x}} - \hat{\mathbf{x}})\|_2 \leq O\left(\frac{\delta}{\sqrt{M}}\right)$.

3 Proof of Theorem 1

In this section, we prove the main result Theorem 1.

3.1 Orthonormal Basis of the $N \times N$ Hankel Matrices Subspace

In this subsection, we introduce an orthonormal basis of the subspace of $N \times N$ Hankel matrices and use it to define a projection from $\mathbb{C}^{N \times N}$ to the subspace of all $N \times N$ Hankel matrices.

Let $\mathbf{E}_j \in \mathbb{C}^{N \times N}$, $j = 0, 1, \dots, 2N - 2$, be the Hankel matrix satisfying

$$[\mathbf{E}_j]_{kl} = \begin{cases} 1/\sqrt{K_j}, & \text{if } k + l = j, \\ 0, & \text{otherwise,} \end{cases} \quad k, l = 0, \dots, N - 1, \quad (8)$$

where $K_j = j + 1$ for $j \leq N - 1$ and $K_j = 2N - 1 - j$ for $j \geq N - 1$ is the number of non-zeros in \mathbf{E}_j . Then, it is easy to check that $\{\mathbf{E}_j\}_{j=0}^{2N-2}$ forms an orthonormal basis of the subspace of all $N \times N$ Hankel matrices, under the standard inner product in $\mathbb{C}^{N \times N}$.

Define a linear operator

$$\mathcal{G} : \mathbf{x} \in \mathbb{C}^{2N-1} \mapsto \mathcal{G}\mathbf{x} = \sum_{j=0}^{2N-2} x_j \mathbf{E}_j \in \mathbb{C}^{N \times N}. \quad (9)$$

The adjoint \mathcal{G}^* of \mathcal{G} is

$$\mathcal{G}^* : \mathbf{X} \in \mathbb{C}^{N \times N} \mapsto \mathcal{G}^*\mathbf{X} \in \mathbb{C}^{2N-1}, \quad [\mathcal{G}^*\mathbf{X}]_j = \langle \mathbf{X}, \mathbf{E}_j \rangle.$$

Obviously, $\mathcal{G}^*\mathcal{G}$ is the identity operator in \mathbb{C}^{2N-1} , and $\mathcal{G}\mathcal{G}^*$ is the orthogonal projector onto the subspace of all Hankel matrices.

3.2 Recovery condition based on restricted minimum gain condition

First of all, let us simplify the minimization problem (4) by introducing $\mathcal{D} \in \mathbb{C}^{(2N-1) \times (2N-1)}$, the diagonal matrix with j -th diagonal $\sqrt{K_j}$. Then, by letting $\mathbf{y} = \mathcal{D}\mathbf{x}$, (4) is rewritten as,

$$\min_{\mathbf{y}} \|\mathcal{G}\mathbf{y}\|_* \quad \text{subject to} \quad \mathcal{B}\mathbf{y} = \mathbf{b}, \quad (10)$$

where $\mathcal{B} = \mathcal{A}\mathcal{D}^{-1}$. Similarly, for the noisy case, (5) is rearranged to

$$\min_{\mathbf{y}} \|\mathcal{G}\mathbf{y}\|_* \quad \text{subject to} \quad \|\mathcal{B}\mathbf{y} - \mathbf{b}\|_2 \leq \epsilon. \quad (11)$$

By our assumption in Theorem 1, $\mathcal{B} \in \mathbb{C}^{M \times (2N-1)}$ is a random matrix whose real and imaginary parts are both real-valued random matrices with i.i.d. Gaussian entries of mean 0 and variance 1. We will prove $\hat{\mathbf{y}} = \mathcal{D}\hat{\mathbf{x}}$ (respectively $\|\hat{\mathbf{y}} - \mathbf{y}\|_2 \leq 2\delta/\epsilon$) with dominant probability for problem (10) for the noise free case (respectively (11) for the noisy case).

Let the descent cone of $\|\mathcal{G}\cdot\|_*$ at $\hat{\mathbf{y}}$ be

$$\mathfrak{T}(\hat{\mathbf{y}}) = \{\lambda\mathbf{z} \mid \lambda \geq 0, \|\mathcal{G}(\hat{\mathbf{y}} + \mathbf{z})\|_* \leq \|\mathcal{G}\hat{\mathbf{y}}\|_*\}. \quad (12)$$

To characterize the recovery condition, we need to use the minimum value of $\frac{\|\mathcal{B}\mathbf{z}\|_2}{\|\mathbf{z}\|_2}$ for nonzero $\mathbf{z} \in \mathfrak{T}(\hat{\mathbf{y}})$. This quantity is commonly called the *minimum gain* of the measurement operator \mathcal{B} restricted on $\mathfrak{T}(\hat{\mathbf{y}})$ [9]. In particular, if the minimum gain is bounded away from zero, then the exact recovery (respectively approximate recovery) for problem (10) (respectively (11)) holds.

Lemma 1. *Let $\mathfrak{T}(\hat{\mathbf{y}})$ be defined by (12). Assume*

$$\min_{\mathbf{z} \in \mathfrak{T}(\hat{\mathbf{y}})} \frac{\|\mathcal{B}\mathbf{z}\|_2}{\|\mathbf{z}\|_2} \geq \epsilon. \quad (13)$$

(a) Let $\tilde{\mathbf{y}}$ be the solution of (10) with $\mathbf{b} = \mathcal{B}\hat{\mathbf{y}}$. Then $\tilde{\mathbf{y}} = \hat{\mathbf{y}}$.

(b) Let $\tilde{\mathbf{y}}$ be the solution of (11) with $\|\mathbf{b} - \mathcal{B}\hat{\mathbf{y}}\|_2 \leq \delta$. Then $\|\tilde{\mathbf{y}} - \hat{\mathbf{y}}\|_2 \leq 2\delta/\epsilon$.

Proof. Since (a) is a special case of (b) with $\delta = 0$, we prove (b) only. The optimality of $\tilde{\mathbf{y}}$ implies $\tilde{\mathbf{y}} - \hat{\mathbf{y}} \in \mathfrak{T}(\hat{\mathbf{y}})$. By (13), we have

$$\|\tilde{\mathbf{y}} - \hat{\mathbf{y}}\|_2 \leq \frac{1}{\epsilon} \|\mathcal{B}(\tilde{\mathbf{y}} - \hat{\mathbf{y}})\|_2 \leq \frac{1}{\epsilon} (\|\mathcal{B}\tilde{\mathbf{y}} - \mathbf{b}\|_2 + \|\mathcal{B}\hat{\mathbf{y}} - \mathbf{b}\|_2) \leq 2\delta/\epsilon.$$

□

Minimum gain condition is a powerful concept and has been employed in recent recovery results via ℓ_1 norm minimization, block-sparse vector recovery, low-rank matrix reconstruction and other atomic norms [9].

3.3 Bound of minimum gain via Gaussian width

Lemma 1 requires to estimate the lower bound of $\min_{\mathbf{z} \in \mathfrak{T}(\hat{\mathbf{y}})} \frac{\|\mathcal{B}\mathbf{z}\|_2}{\|\mathbf{z}\|_2}$. Gordon gave a solution using Gaussian width of a set [9, 13] to estimate the lower bound of minimum gain.

Definition 1. The Gaussian width of a set $S \subset \mathbb{R}^p$ is defined as:

$$w(S) := \mathbb{E}_{\boldsymbol{\xi}} \left[\sup_{\boldsymbol{\gamma} \in S} \boldsymbol{\gamma}^T \boldsymbol{\xi} \right],$$

where $\boldsymbol{\xi} \in \mathbb{R}^p$ is a random vector of independent zero-mean unit-variance Gaussians.

Let λ_n denote the expected length of a n -dimensional Gaussian random vector. Then $\lambda_n = \sqrt{2}\Gamma(\frac{n+1}{2})/\Gamma(\frac{n}{2})$ and it can be tightly bounded as $\frac{n}{\sqrt{n+1}} \leq \lambda_n \leq \sqrt{n}$ [9]. The following theorem is given in Corollary 1.2 in [13]. It gives a bound on minimum gain for a random map $\mathbf{\Pi} : \mathbb{R}^p \mapsto \mathbb{R}^n$.

Theorem 2 (Corollary 1.2 in [13]). Let Ω be a closed subset of $\{\mathbf{x} \in \mathbb{R}^p \mid \|\mathbf{x}\|_2 = 1\}$. Let $\mathbf{\Pi} \in \mathbb{R}^{n \times p}$ be a random matrix with i.i.d. Gaussian entries with mean 0 and variance 1. Then, for any $\epsilon > 0$,

$$\mathbb{P} \left(\min_{\mathbf{z} \in \Omega} \|\mathbf{\Pi}\mathbf{z}\|_2 \geq \epsilon \right) \geq 1 - e^{-\frac{1}{2}(\lambda_n - w(\Omega) - \epsilon)^2},$$

provided $\lambda_n - w(\Omega) - \epsilon \geq 0$. Here $\frac{n}{\sqrt{n+1}} \leq \lambda_n \leq \sqrt{n}$, and $w(\Omega)$ is the Gaussian width of Ω .

By converting the complex setting in our problem to the real setting and using Theorem 2, we can get the bound of (13) in terms of Gaussian width of $\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3}$, where $\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}})$ is a cone in \mathbb{R}^{4N-2} defined by

$$\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) = \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \mid \boldsymbol{\alpha} + \imath\boldsymbol{\beta} \in \mathfrak{T}(\hat{\mathbf{y}}) \right\}. \quad (14)$$

Lemma 2. Let the real and imaginary parts of entries of $\mathcal{B} \in \mathbb{C}^{M \times (2N-1)}$ be i.i.d. Gaussian with mean 0 and variance 1. Let $\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}})$ be defined by (14) and $\mathbb{S}_{\mathbb{C}}^{2N-2}$ be the unit sphere in \mathbb{C}^{2N-1} . Then for any $\epsilon > 0$,

$$\mathbb{P} \left(\min_{\mathbf{z} \in \mathfrak{T}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{C}}^{2N-2}} \|\mathcal{B}\mathbf{z}\|_2 \geq \epsilon \right) \geq 1 - 2e^{-\frac{1}{2}(\lambda_M - w(\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3}) - \frac{\epsilon}{\sqrt{2}})^2},$$

where $\mathbb{S}_{\mathbb{R}}^{4N-3}$ is the unit sphere in \mathbb{R}^{4N-2} .

Proof. In order to use Theorem 2, we convert the complex setting in our problem to the real setting in Theorem 2. We will use Roman letters for vectors and matrices in complex-valued spaces, and Greek letters for real valued ones. Let $\mathcal{B} = \boldsymbol{\Phi} + \imath\boldsymbol{\Psi} \in \mathbb{C}^{M \times (2N-1)}$, where both $\boldsymbol{\Phi} \in \mathbb{R}^{M \times (2N-1)}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{M \times (2N-1)}$

are real-valued random matrices whose entries are i.i.d. mean-0 variance-1 Gaussian. Then, for any $\mathbf{z} = \boldsymbol{\alpha} + \imath\boldsymbol{\beta} \in \mathbb{C}^{2N-1}$ with $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^{2N-1}$,

$$\begin{aligned} \|\mathcal{B}\mathbf{z}\|_2 &= \|(\boldsymbol{\Phi} + \imath\boldsymbol{\Psi})(\boldsymbol{\alpha} + \imath\boldsymbol{\beta})\|_2 = \|(\boldsymbol{\Phi}\boldsymbol{\alpha} - \boldsymbol{\Psi}\boldsymbol{\beta}) + \imath(\boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\Phi}\boldsymbol{\beta})\|_2 \\ &= \left(\left\| \begin{bmatrix} \boldsymbol{\Phi} & -\boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2 + \left\| \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Phi} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2 \right)^{1/2} \end{aligned}$$

Then

$$\min_{\mathbf{z}=\boldsymbol{\alpha}+\imath\boldsymbol{\beta} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \left\| \begin{bmatrix} \boldsymbol{\Phi} & -\boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2 \geq \epsilon/\sqrt{2}, \quad \text{and} \quad \min_{\mathbf{z}=\boldsymbol{\alpha}+\imath\boldsymbol{\beta} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \left\| \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Phi} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2 \geq \epsilon/\sqrt{2} \quad (15)$$

implies

$$\min_{\mathbf{z} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \|\mathcal{B}\mathbf{z}\|_2 \geq \epsilon.$$

Therefore,

$$\mathbb{P} \left(\min_{\mathbf{z} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \|\mathcal{B}\mathbf{z}\|_2 \geq \epsilon \right) \geq \mathbb{P}((15) \text{ holds true}).$$

It is easy to see that both $\begin{bmatrix} \boldsymbol{\Phi} & -\boldsymbol{\Psi} \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Phi} \end{bmatrix}$ are real-valued random matrices with i.i.d. Gaussian entries of mean 0 and variance 1. By Theorem 2,

$$\mathbb{P}((15) \text{ holds true}) \geq 1 - 2e^{-\frac{1}{2}(\lambda_M - w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3}) - \frac{\epsilon}{\sqrt{2}})^2},$$

and therefore we get the desired result. \square

3.4 Estimation of Gaussian width $w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3})$

Denote $\mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}})$ be polar cone of $\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \in \mathbb{R}^{4N-2}$, i.e.,

$$\mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}}) = \{\boldsymbol{\delta} \in \mathbb{R}^{4N-2} \mid \boldsymbol{\gamma}^T \boldsymbol{\delta} \leq 0, \forall \boldsymbol{\gamma} \in \mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}})\}. \quad (16)$$

Following the arguments in Proposition 3.6 in [9], we obtain

$$w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3}) = \mathbb{E} \left(\sup_{\boldsymbol{\gamma} \in \mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3}} \boldsymbol{\xi}^T \boldsymbol{\gamma} \right) \leq \mathbb{E} \left(\min_{\boldsymbol{\gamma} \in \mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}})} \|\boldsymbol{\xi} - \boldsymbol{\gamma}\|_2 \right), \quad (17)$$

where $\boldsymbol{\xi} \in \mathbb{R}^{4N-2}$ is a random vector of i.i.d. Gaussian entries of mean 0 and variance 1. Hence, instead of estimating Gaussian width $w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3})$, we bound $\mathbb{E} \left(\min_{\boldsymbol{\gamma} \in \mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}})} \|\boldsymbol{\xi} - \boldsymbol{\gamma}\|_2 \right)$. For this purpose, let $\mathcal{F} : \mathbb{R}^{4N-2} \mapsto \mathbb{R}$ be defined by

$$\mathcal{F} \left(\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right) = \|\mathcal{G}(\boldsymbol{\alpha} + \imath\boldsymbol{\beta})\|_*. \quad (18)$$

The following lemma gives us a characterization of $\mathbb{E} \left(\min_{\boldsymbol{\gamma} \in \mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}})} \|\boldsymbol{\xi} - \boldsymbol{\gamma}\|_2 \right)$ in terms of the subdifferential $\partial\mathcal{F}$ of \mathcal{F} .

Lemma 3. *Let $\mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}})$ and \mathcal{F} be defined by (16) and (18) respectively. Let $\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2 \in \mathbb{R}^{2N-1}$ be the real and imaginary parts of $\hat{\mathbf{y}}$ respectively and denote $\hat{\boldsymbol{\omega}} = \begin{bmatrix} \hat{\boldsymbol{\omega}}_1 \\ \hat{\boldsymbol{\omega}}_2 \end{bmatrix}$. Then*

$$\mathfrak{I}_\mathbb{R}^*(\hat{\mathbf{y}}) = \text{cone}(\partial\mathcal{F}(\hat{\boldsymbol{\omega}})) = \{\lambda\boldsymbol{\delta} \mid \lambda \geq 0, \mathcal{F}(\boldsymbol{\gamma} + \hat{\boldsymbol{\omega}}) \geq \mathcal{F}(\hat{\boldsymbol{\omega}}) + \boldsymbol{\gamma}^T \boldsymbol{\delta}, \forall \boldsymbol{\gamma} \in \mathbb{R}^{4N-2}\}. \quad (19)$$

Proof. It is observed that $\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}})$ in (14) is the descent cone of the function \mathcal{F}

$$\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) = \{\delta\gamma \mid \delta \geq 0, \mathcal{F}(\gamma + \hat{\omega}) \leq \mathcal{F}(\hat{\omega})\}.$$

According to Theorem 23.4 in [18], the cone dual to the descent cone is the conic hull of subgradient, which is exactly (19). \square

The following lemma gives us an estimation of Gaussian width $w(\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3})$ in terms of $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$.

Lemma 4. *Let $\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}})$ and \mathcal{G} be defined by (14) and (9) respectively. Then*

$$w(\mathfrak{T}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3}) \leq 3\sqrt{R} \cdot \mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2),$$

where $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$ is the expectation with respect to $\mathbf{g} \in \mathbb{C}^{2N-1}$. Here \mathbf{g} is a random vector whose real and imaginary parts are i.i.d. mean-0 and variance-1 Gaussian entries.

Proof. By using (17) and Lemma 3, we need to find $\partial\mathcal{F}(\hat{\omega})$ and thus $\mathfrak{T}_{\mathbb{R}}^*(\hat{\mathbf{y}})$. Let $\hat{\Omega}_1 = \mathcal{G}\hat{\omega}_1$ and $\hat{\Omega}_2 = \mathcal{G}\hat{\omega}_2$. Then $\mathcal{G}\hat{\mathbf{y}} = \hat{\Omega}_1 + \iota\hat{\Omega}_2$. Let a singular value decomposition of the rank- R matrix $\mathcal{G}\hat{\mathbf{y}}$ be

$$\mathcal{G}\hat{\mathbf{y}} = \mathbf{U}\Sigma\mathbf{V}^*, \quad \text{with } \mathbf{U} = \Theta_1 + \iota\Theta_2, \quad \mathbf{V} = \Xi_1 + \iota\Xi_2, \quad (20)$$

where $\Theta_1, \Theta_2, \Xi_1, \Xi_2 \in \mathbb{R}^{N \times R}$ and $\Sigma \in \mathbb{R}^{R \times R}$, and $\mathbf{U} \in \mathbb{C}^{N \times R}$ and $\mathbf{V} \in \mathbb{C}^{N \times R}$ satisfies $\mathbf{U}^*\mathbf{U} = \mathbf{V}^*\mathbf{V} = \mathbf{I}$. Then, by direct calculation,

$$\Theta \equiv \begin{bmatrix} \Theta_1 & -\Theta_2 \\ \Theta_2 & \Theta_1 \end{bmatrix} \in \mathbb{R}^{2N \times (2R)}, \quad \Xi \equiv \begin{bmatrix} \Xi_1 & -\Xi_2 \\ \Xi_2 & \Xi_1 \end{bmatrix} \in \mathbb{R}^{2N \times (2R)} \quad (21)$$

satisfy $\Theta^T\Theta = \Xi^T\Xi = \mathbf{I}$. Moreover, if we define $\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_1 & -\hat{\Omega}_2 \\ \hat{\Omega}_2 & \hat{\Omega}_1 \end{bmatrix}$, then

$$\hat{\Omega} = \Theta \begin{bmatrix} \Sigma & \\ & \Sigma \end{bmatrix} \Xi^T \quad (22)$$

is a singular value decomposition of the real matrix $\hat{\Omega}$, and the singular values $\hat{\Omega}$ are those of $\mathcal{G}\hat{\mathbf{y}}$, each repeated twice. Therefore,

$$\mathcal{F}(\hat{\omega}) = \|\mathcal{G}\hat{\mathbf{y}}\|_* = \|\Sigma\|_* = \frac{1}{2}\|\hat{\Omega}\|_*. \quad (23)$$

Define a linear operator $\mathcal{E} : \mathbb{R}^{4N-2} \mapsto \mathbb{R}^{2N \times 2N}$ by

$$\mathcal{E} \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right) = \begin{bmatrix} \mathcal{G}\alpha & -\mathcal{G}\beta \\ \mathcal{G}\beta & \mathcal{G}\alpha \end{bmatrix}, \quad \text{with } \alpha, \beta \in \mathbb{R}^{2N-1}.$$

By (23) and the definition of $\hat{\Omega}$, we obtain $\mathcal{F}(\hat{\omega}) = \frac{1}{2}\|\mathcal{E}\hat{\omega}\|_*$. From convex analysis theory and $\hat{\Omega} = \mathcal{E}\hat{\omega}$, the subdifferential of \mathcal{F} is given by

$$\partial\mathcal{F}(\hat{\omega}) = \frac{1}{2}\mathcal{E}^*\partial\|\hat{\Omega}\|_*. \quad (24)$$

On the one hand, the adjoint \mathcal{E}^* is given by, for any $\Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \in \mathbb{R}^{2N \times 2N}$ with each block in $\mathbb{R}^{N \times N}$,

$$\mathcal{E}^*\Delta = \begin{bmatrix} \mathcal{G}^*(\Delta_{11} + \Delta_{22}) \\ \mathcal{G}^*(\Delta_{21} - \Delta_{12}) \end{bmatrix}. \quad (25)$$

On the other hand, since (22) provides a singular value decomposition of $\hat{\Omega}$,

$$\partial\|\hat{\Omega}\|_* = \{\Theta\Xi^T + \Delta \mid \Theta^T\Delta = \mathbf{0}, \Delta\Xi = \mathbf{0}, \|\Delta\|_2 \leq 1\}. \quad (26)$$

Combining (24)(25)(26) and (21) yields the subdifferential of \mathcal{F} at $\hat{\omega}$

$$\partial\mathcal{F}(\hat{\omega}) = \left\{ \begin{bmatrix} \mathcal{G}^* (\Theta_1 \Xi_1^T + \Theta_2 \Xi_2^T + \frac{\Delta_{11} + \Delta_{22}}{2}) \\ \mathcal{G}^* (\Theta_2 \Xi_1^T - \Theta_1 \Xi_2^T + \frac{\Delta_{21} - \Delta_{12}}{2}) \end{bmatrix} \mid \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \Theta^T \Delta = \mathbf{0}, \Delta \Xi = \mathbf{0}, \|\Delta\|_2 \leq 1 \right\}.$$

We are now ready for the estimation of the Gaussian width. Let the set \mathfrak{S} be a subset of the set of complex-valued vectors

$$\mathfrak{S} = \{\mathcal{G}^*(UV^* + \mathbf{W}) \mid U^*W = \mathbf{0}, \mathbf{W}V = \mathbf{0}, \|\mathbf{W}\|_2 \leq 1\}, \quad (27)$$

where U, V are in (20). Then, it can be checked that

$$\mathfrak{H} \equiv \left\{ \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \mid \alpha + \imath\beta \in \mathfrak{S} \right\} \subset \partial\mathcal{F}(\hat{\omega}). \quad (28)$$

Actually, for any $\mathbf{W} = \Delta_1 + \imath\Delta_2$ satisfying $U^*W = 0, \mathbf{W}V = 0$ and $\|\mathbf{W}\|_2 \leq 1$, we choose $\Delta = \begin{bmatrix} \Delta_1 & -\Delta_2 \\ \Delta_2 & \Delta_1 \end{bmatrix}$. Obviously, this choice of Δ satisfies the constraints on Δ in $\partial\mathcal{F}(\hat{\omega})$. Furthermore, $UV^* + \mathbf{W} = (\Theta_1 \Xi_1^T + \Theta_2 \Xi_2^T + \Delta_1) + \imath(\Theta_2 \Xi_1^T + \Theta_1 \Xi_2^T + \Delta_2)$. Therefore, (28) holds.

With the help of (28), we get

$$\min_{\gamma \in \mathfrak{S}_{\mathbb{R}}^*(\hat{\mathbf{y}})} \|\xi - \gamma\|_2 = \min_{\lambda \geq 0} \min_{\gamma \in \partial\mathcal{F}(\hat{\omega})} \|\xi - \lambda\gamma\|_2 \leq \min_{\lambda \geq 0} \min_{\gamma \in \mathfrak{H}} \|\xi - \lambda\gamma\|_2. \quad (29)$$

We then convert the real-valued vectors to complex-valued vectors by letting $\mathbf{g} = \xi_1 + \imath\xi_2$ and $\mathbf{c} = \gamma_1 + \imath\gamma_2$, where ξ_1 and ξ_2 are the first and second half of ξ respectively and so for γ_1 and γ_2 . This leads to

$$\min_{\gamma \in \mathfrak{S}_{\mathbb{R}}^*(\hat{\mathbf{y}})} \|\xi - \gamma\|_2 \leq \min_{\lambda \geq 0} \min_{\gamma \in \mathfrak{H}} \|\xi - \lambda\gamma\|_2 = \min_{\lambda \geq 0} \min_{\mathbf{c} \in \mathfrak{S}} \|\mathbf{g} - \lambda\mathbf{c}\|_2.$$

Since $\mathcal{G}^*\mathcal{G}$ is the identity operator and $\mathcal{G}\mathcal{G}^*$ is an orthogonal projector, for any $\lambda \geq 0$ and $\mathbf{c} \in \mathfrak{S}$,

$$\begin{aligned} \|\mathbf{g} - \lambda\mathbf{c}\|_2 &= \|\mathcal{G}\mathbf{g} - \lambda\mathcal{G}\mathbf{c}\|_F = \|\mathcal{G}\mathbf{g} - \lambda\mathcal{G}\mathcal{G}^*(UV^* + \mathbf{W})\|_F \\ &= (\|\mathcal{G}\mathbf{g} - \lambda(UV^* + \mathbf{W})\|_F^2 - \|\lambda(\mathcal{I} - \mathcal{G}\mathcal{G}^*)(UV^* + \mathbf{W})\|_F^2)^{1/2} \\ &\leq \|\mathcal{G}\mathbf{g} - \lambda(UV^* + \mathbf{W})\|_F, \end{aligned} \quad (30)$$

where \mathbf{W} satisfies the conditions in the definition of \mathfrak{S} in (27). Define two orthogonal projectors \mathcal{P}_1 and \mathcal{P}_2 in $\mathbb{C}^{N \times N}$ by

$$\mathcal{P}_1\mathbf{X} = \mathbf{U}\mathbf{U}^*\mathbf{X} + \mathbf{X}\mathbf{V}\mathbf{V}^* - \mathbf{U}\mathbf{U}^*\mathbf{X}\mathbf{V}\mathbf{V}^*, \quad \mathcal{P}_2\mathbf{X} = (\mathbf{I} - \mathbf{U}\mathbf{U}^*)\mathbf{X}(\mathbf{I} - \mathbf{V}\mathbf{V}^*).$$

Then, it can be easily checked that: $\mathcal{P}_1\mathbf{X}$ and $\mathcal{P}_2\mathbf{X}$ are orthogonal, $\mathbf{X} = \mathcal{P}_1\mathbf{X} + \mathcal{P}_2\mathbf{X}$, and

$$\mathcal{P}_1\mathbf{U}\mathbf{V}^* = \mathbf{U}\mathbf{V}^*, \quad \mathcal{P}_2\mathbf{W} = \mathbf{0}, \quad \mathcal{P}_1\mathbf{W} = \mathbf{0}, \quad \mathcal{P}_2\mathbf{W} = \mathbf{W}, \quad (31)$$

where U, V, W the same as those in (27). We choose

$$\lambda = \|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2, \quad \mathbf{W} = \frac{1}{\lambda}\mathcal{P}_2(\mathcal{G}\mathbf{g}).$$

Then, \mathbf{W} satisfies constraints in (27). This, together with (29)(30)(31), implies

$$\begin{aligned} \min_{\gamma \in \mathfrak{S}_{\mathbb{R}}^*(\hat{\mathbf{y}})} \|\xi - \gamma\|_2 &\leq \|\mathcal{G}\mathbf{g} - \|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2\mathbf{U}\mathbf{V}^* - \mathcal{P}_2(\mathcal{G}\mathbf{g})\|_F = \|\mathcal{P}_1(\mathcal{G}\mathbf{g}) - \|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2\mathbf{U}\mathbf{V}^*\|_F \\ &\leq \|\mathcal{P}_1(\mathcal{G}\mathbf{g})\|_F + \|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2\|\mathbf{U}\mathbf{V}^*\|_F = \|\mathcal{P}_1(\mathcal{G}\mathbf{g})\|_F + \sqrt{R}\|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2. \end{aligned}$$

We will estimate both $\|\mathcal{P}_1(\mathcal{G}\mathbf{g})\|_F$ and $\|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2$. For $\|\mathcal{P}_1(\mathcal{G}\mathbf{g})\|_F$, we have

$$\begin{aligned}\|\mathcal{P}_1(\mathcal{G}\mathbf{g})\|_F &= \|\mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g}) + (\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^* - \mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^*\|_F = \|\mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g}) + (\mathbf{I} - \mathbf{U}\mathbf{U}^*)(\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^*\|_F \\ &\leq \|\mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g})\|_F + \|(\mathbf{I} - \mathbf{U}\mathbf{U}^*)(\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^*\|_F \leq \|\mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g})\|_F + \|(\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^*\|_F \\ &\leq 2\sqrt{R}\|\mathcal{G}\mathbf{g}\|_2\end{aligned}$$

where in the last line we have used the inequality

$$\|\mathbf{U}\mathbf{U}^*(\mathcal{G}\mathbf{g})\|_F \leq \|\mathbf{U}\mathbf{U}^*\|_F \|\mathcal{G}\mathbf{g}\|_2 \leq \sqrt{R}\|\mathcal{G}\mathbf{g}\|_2$$

and similarly $\|(\mathcal{G}\mathbf{g})\mathbf{V}\mathbf{V}^*\|_F \leq \sqrt{R}\|\mathcal{G}\mathbf{g}\|_2$. For $\|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2$,

$$\|\mathcal{P}_2(\mathcal{G}\mathbf{g})\|_2 = \|(\mathbf{I} - \mathbf{U}\mathbf{U}^*)(\mathcal{G}\mathbf{g})(\mathbf{I} - \mathbf{V}\mathbf{V}^*)\|_2 \leq \|\mathbf{I} - \mathbf{U}\mathbf{U}^*\|_2 \|\mathcal{G}\mathbf{g}\|_2 \|\mathbf{I} - \mathbf{V}\mathbf{V}^*\|_2 \leq \|\mathcal{G}\mathbf{g}\|_2.$$

Altogether, we obtain

$$\min_{\boldsymbol{\gamma} \in \mathfrak{I}_{\mathbb{R}}^*(\hat{\mathbf{y}})} \|\boldsymbol{\xi} - \boldsymbol{\gamma}\|_2 \leq 3\sqrt{R}\|\mathcal{G}\mathbf{g}\|_2,$$

which together with (17) gives

$$w(\mathfrak{I}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3}) \leq 3\sqrt{R} \cdot \mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2).$$

□

3.5 Bound of $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$

The estimation of $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$ plays an important role in proving Theorem 1 since it needed to give the tight bound of the Gaussian width $w(\mathfrak{I}_{\mathbb{R}}(\hat{\mathbf{y}}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3})$. The following Theorem gives us a bound for $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$.

Theorem 3. *Let $\mathbf{g} \in \mathbb{R}^{2N-1}$ be a random vector whose entries are i.i.d. Gaussian random variables with mean 0 and variance 1, or $\mathbf{g} \in \mathbb{C}^{2N-1}$ a random vector whose real part and imaginary part have i.i.d. Gaussian random entries with mean 0 and variance 1. Then,*

$$\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2) \leq C_1 \ln N,$$

where C_1 are some positive universal constants.

The proof of Theorem 3 is relatively complicated. In order to help the reader easily understand the proof, we begin with real case and introduce some ideas and lemmas first. Assume $\mathbf{g} \in \mathbb{R}^{2N-1}$ has i.i.d standard Gaussian entries with mean 0 and variance 1. Notice that $\mathcal{G}\mathbf{g}$ is symmetric. Therefore, for any even integer k , $(\text{tr}(\mathcal{G}\mathbf{g})^k)^{1/k}$ is the k -norm of vector of singular values, which implies $\|\mathcal{G}\mathbf{g}\|_2 \leq (\text{tr}(\mathcal{G}\mathbf{g})^k)^{1/k}$. This together with Jensen's inequality,

$$\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2) \leq \mathbb{E}\left((\text{tr}(\mathcal{G}\mathbf{g})^k)^{1/k}\right) \leq \left(\mathbb{E}(\text{tr}(\mathcal{G}\mathbf{g})^k)\right)^{1/k}. \quad (32)$$

Thus, in order to get an upper bound of $\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2)$, we estimate $\mathbb{E}\left(\text{tr}\left((\mathcal{G}\mathbf{g})^k\right)\right)$. Denote $\mathbf{M} = \mathcal{G}\mathbf{g}$. It is easy to see that

$$\mathbb{E}(\text{tr}(\mathbf{M}^k)) = \sum_{0 \leq i_1, i_2, \dots, i_k \leq N-1} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}). \quad (33)$$

Therefore, we only need to estimate $\sum_{0 \leq i_1, i_2, \dots, i_k \leq N-1} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1})$.

To simplify the notation, we denote $i_{k+1} = i_1$. Notice that $M_{ij} = \frac{g_{i+j}}{\sqrt{K_{i+j}}}$, where g_{i+j} is a random Gaussian variable and K_j is defined in (8). Hence, $M_{i_\ell, i_{\ell+1}} = M_{i_{\ell'}, i_{\ell'+1}}$ if and only if $i_\ell + i_{\ell+1} = i_{\ell'} + i_{\ell'+1}$. In order to utilize this property, we would like to introduce a graph for any given index i_1, i_2, \dots, i_k and its

equivalent edges on the graph. More specifically, we construct graph $\mathfrak{F}_{i_1, i_2, \dots, i_k}$ with nodes to be i_1, i_2, \dots, i_k and edges to be $(i_1, i_2), (i_2, i_3), \dots, (i_{k-1}, i_k), (i_k, i_1)$. Let the weight for the edge $(i_\ell, i_{\ell+1})$ be $i_\ell + i_{\ell+1}$. The edges with the same weights are considered as an equivalent class. Obviously, $M_{i_\ell, i_{\ell+1}} = M_{i_{\ell'}, i_{\ell'+1}}$ if and only if $(i_\ell, i_{\ell+1})$ and $(i_{\ell'}, i_{\ell'+1})$ are in the same equivalent class. Assume there are p equivalent classes of the edges of $\mathfrak{F}_{i_1, i_2, \dots, i_k}$. These equivalent classes are indexed by $1, 2, \dots, p$ according to their order in the graph traversal $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_k \rightarrow i_1$. We associate the graph $\mathfrak{F}_{i_1, i_2, \dots, i_k}$ a sequence $c_1 c_2 \dots c_k$, where c_j is the index of the equivalent class of the edge (i_j, i_{j+1}) . We call $c_1 c_2 \dots c_k$ the label for the equivalent classes of the graph $\mathfrak{F}_{i_1, i_2, \dots, i_k}$.

The label for the equivalent classes of the graph $\mathfrak{F}_{i_1, i_2, \dots, i_k}$ plays an important role in bounding $\mathbb{E}(\|\mathcal{G}\|_2)$. In order to help the reader understand this concept better, we give two specific examples here. For $N = 6, k = 6, i_1 = 1, i_2 = 4, i_3 = 1, i_4 = 3, i_5 = 1, i_6 = 4$, we have a corresponding graph and its label for the equivalent classes of the graph is 112211. For $N = 6, k = 6, i_1 = 2, i_2 = 3, i_3 = 2, i_4 = 4, i_5 = 2, i_6 = 3$, the label for the equivalent classes of the corresponding graph is 112211 as well. Therefore, there may be several different index sequences $i_1 i_2 \dots i_k$ that correspond to the same label for the equivalent classes of the corresponding graph. Let $\mathfrak{A}_{c_1 c_2 \dots c_k}$ be the set of indices whose label of equivalent class of the corresponding graph is $c_1 c_2 \dots c_k$, i.e.

$$\mathfrak{A}_{c_1 c_2 \dots c_k} = \{i_1 i_2 \dots i_k \mid \text{the label for the equivalent class of the graph } \mathfrak{F}_{i_1, i_2, \dots, i_k} \text{ is } c_1 c_2 \dots c_k\} \quad (34)$$

For given $c_1 c_2 \dots c_k$, $\mathfrak{A}_{c_1 c_2 \dots c_k}$ is a subset of $\{i_1 i_2 \dots i_k \mid i_j \in \{0, 1, \dots, N-1\}, \forall j = 1, \dots, k\}$. The following lemma gives us an estimate for the bound $\sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1})$.

Lemma 5. *Let ζ be the Riemann zeta function and $\mathfrak{A}_{c_1 c_2 \dots c_k}$ be defined in (34). Define $B(s) = \ln(N+1)$ if $s = 2$ and $B(s) = \zeta(s/2) \leq \pi^2/6$ for $s \geq 4$. Then*

$$\sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \leq N \prod_{\ell=1}^p B(s_\ell) (s_\ell - 1)!! \quad (35)$$

where p is the number of equivalent classes shown in $c_1 c_2 \dots c_k$, and $s_\ell, \ell = 1, \dots, p$, is the frequency of ℓ in $c_1 c_2 \dots c_k$.

Proof. We begin with finding free indices for any i_1, i_2, \dots, i_k in the set $\mathfrak{A}_{c_1 c_2 \dots c_k}$. Let (j_1, j_2) be the first edge of the class 1. Therefore, the weight of the first class is $j_1 + j_2$. For convenience, we define $k_1(j_1) = j_1$. The first edge of the class 2 must have a vertex $k_2(j_1, j_2)$, depending on j_1 and j_2 , and a free vertex, denoted by j_3 . The weight of the second class is $k_2(j_1, j_2) + j_3$. Similarly, the first edge in class 3 has a vertex $k_3(j_1, j_2, j_3)$ and a free vertex j_4 , and the weight is $k_3(j_1, j_2, j_3) + j_4$, and so on. Finally, the first edge in class p has a vertex $k_p(j_1, j_2, \dots, j_p)$ and a free vertex j_{p+1} , and the weight is $k_p(j_1, j_2, \dots, j_p) + j_{p+1}$. Recall that the entry M_{ij} is $\frac{g_{i+j}}{\sqrt{K_{i+j}}}$, where g_{i+j} is a random Gaussian variable. Therefore, for any $i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}$,

$$\mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) = \prod_{\ell=1}^p \frac{1}{K^{s_\ell/2}} \mathbb{E}(g_{m_\ell}^{s_\ell}), \quad (36)$$

where $m_\ell = k_\ell(j_1, j_2, \dots, j_\ell) + j_{\ell+1}$. Therefore, it is non-vanishing if and only if s_1, s_2, \dots, s_p are all even. In these cases,

$$\mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) = \prod_{\ell=1}^p \frac{(s_\ell - 1)!!}{K^{s_\ell/2}}. \quad (37)$$

Summing (37) over $\mathfrak{A}_{c_1 c_2 \dots c_k}$, we obtain

$$\begin{aligned} & \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \leq \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1} \dots \sum_{j_p=0}^{N-1} \sum_{j_{p+1}=0}^{N-1} \prod_{\ell=1}^p \frac{(s_\ell - 1)!!}{K^{s_\ell/2}} \\ & = \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1} \left(\frac{(s_1 - 1)!!}{K^{s_1/2}} \sum_{j_3=0}^{N-1} \left(\frac{(s_2 - 1)!!}{K^{s_2/2}} \sum_{j_4=0}^{N-1} \left(\dots \sum_{j_{p+1}=0}^{N-1} \frac{(s_p - 1)!!}{K^{s_p/2}} \right) \dots \right) \right) \end{aligned}$$

Since, for any $0 \leq c \leq N - 1$,

$$\sum_{\ell=0}^{N-1} \frac{1}{K_{c+\ell}^{s/2}} \leq \begin{cases} 1 + 1/2 + 1/3 + \dots + 1/N \leq \ln(N+1) & s = 2, \\ 1 + 1/2^{s/2} + \dots + 1/N^{s/2} \leq \zeta(s/2) & s = 4, 6, \dots \end{cases}$$

where ζ is the Riemann zeta function. By defining $B(s) = \ln(N+1)$ if $s = 2$ and $B(s) = \zeta(s/2) \leq \pi^2/6$ for $s \geq 4$, the desired result easily follows. \square

The desired bound for $\mathbb{E}(\|\mathcal{G}\|_2)$ can be obtained if we know how many different sets of $\mathfrak{A}_{c_1 c_2 \dots c_k}$ available in the set $\{i_1 i_2 \dots i_k | i_j \in \{0, 1, \dots, N-1\}, \forall j = 1, \dots, k\}$. Let $\mathfrak{B}_{s_1 s_2 \dots s_p}$ be the set of all labels of p equivalent classes with ℓ -th class containing s_ℓ equivalent edges respectively, i.e.

$$\mathfrak{B}_{s_1 s_2 \dots s_p} = \left\{ c_1 c_2 \dots c_p \mid \begin{array}{l} c_1 c_2 \dots c_p \text{ is a valid label of equivalent classes in graph } \mathfrak{F}_{i_1 i_2 \dots i_k} \\ \text{and there are } s_\ell \text{ } \ell\text{'s in the label } c_1 c_2 \dots c_p \end{array} \right\} \quad (38)$$

Let \mathfrak{C}_p be the set of all possible set of all possible choice of p positive even numbers s_1, \dots, s_p satisfying $s_1 + s_2 + \dots + s_p = k$. Then

$$\begin{aligned} \mathbb{E}(\text{tr}(\mathbf{M}^k)) &= \sum_{0 \leq i_1, i_2, \dots, i_k \leq N-1} E(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ &\leq \sum_{p=1}^{k/2} \sum_{s_1 \dots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} E(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \end{aligned} \quad (39)$$

By bounding the cardinality of $\mathfrak{B}_{s_1 s_2 \dots s_p}$ and \mathfrak{C}_p , we can derive the bound $\mathbb{E}(\text{tr}(\mathbf{M}^k))$ hence $\mathbb{E}(\|\mathcal{G}\|_2)$ for the real case. The complex case can be proved by directly using the results for the real case. Now, we are in position to prove Theorem 3.

Proof of Theorem 3. Following (39), we need to count the cardinality of $\mathfrak{B}_{s_1 s_2 \dots s_p}$. For any $c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}$, we must have $c_1 = 1$. Therefore, there are $\binom{k-1}{s_1-1}$ choices of the positions of remaining 1's in $c_1 c_2 \dots c_k$. Once positions for 1's are fixed, the position of the first 2 has to be the first available slot, we have $\binom{k-s_1-1}{s_2-1}$ choices for the positions of remaining 2's, and so on. Thus,

$$\begin{aligned} |\mathfrak{B}_{s_1 s_2 \dots s_p}| &\leq \binom{k-1}{s_1-1} \cdot \binom{k-s_1-1}{s_2-1} \cdot \dots \cdot \binom{k-s_1-\dots-s_{p-1}-1}{s_p-1} \\ &= \frac{(k-1)(k-2)\dots(k-s_1+1)(k-s_1-1)\dots(k-s_1-s_2+1)}{(s_1-1)!(s_2-1)!} \dots 1 \\ &= \frac{(k-1)!}{\prod_{\ell=1}^p (s_\ell-1)! \prod_{\ell=1}^{p-1} (k-s_1-\dots-s_\ell)}, \end{aligned}$$

which together with (35) implies, for any $s_1 s_2 \dots s_p \in \mathfrak{C}_p$,

$$\begin{aligned} &\sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} E(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ &\leq N \frac{(k-1)!}{\prod_{\ell=1}^p (s_\ell-2)!! \prod_{\ell=1}^{p-1} (k-s_1-\dots-s_\ell)} \prod_{\ell=1}^p B(s_\ell) \end{aligned} \quad (40)$$

Summing (40) over \mathfrak{C}_p yields

$$\begin{aligned} &\sum_{s_1 \dots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} E(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ &\leq N(k-1)! \sum_{s_1 \dots s_p \in \mathfrak{C}_p} \frac{\prod_{\ell=1}^p B(s_\ell)}{\prod_{\ell=1}^p (s_\ell-2)!! \prod_{\ell=1}^{p-1} (k-s_1-\dots-s_\ell)} \end{aligned} \quad (41)$$

Let us estimate the sum in the last line. Let s be the number of 2's in $s_1 s_2 \dots s_p$. Then,

$$\prod_{\ell=1}^p B(s_\ell) \leq \ln^s(N+1) \left(\frac{\pi^2}{6}\right)^{p-s}. \quad (42)$$

Since each $s_1, \dots, s_p \geq 2$ and there are $p-s$ terms greater than 4 among them, we have

$$\prod_{\ell=1}^p (s_\ell - 2)!! \geq 2^{p-s} \quad (43)$$

and $k - s_1 - \dots - s_\ell = s_{\ell+1} + \dots + s_p \geq 2(p-\ell)$, which implies

$$\prod_{\ell=1}^{p-1} (k - s_1 - \dots - s_\ell) \geq \prod_{\ell=1}^{p-1} 2(p-\ell) = 2^{p-1}(p-1)!. \quad (44)$$

There are $\binom{p}{s}$ choices of the positions of the s 2's. Moreover, once the s 2's in $s_1 s_2 \dots s_p$ are chosen, there are at most

$$\left(\frac{k}{2} - s\right) \cdot \left(\frac{k}{2} - s - 1\right) \cdot \dots \cdot \left(\frac{k}{2} - s - (p-s+1)\right) \leq \left(\frac{k}{2}\right)^{p-s}$$

choices of the remaining $p-s$ s_j 's. Altogether,

$$\begin{aligned} & \sum_{s_1 \dots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ & \leq N(k-1)! \sum_{s=0}^p \binom{p}{s} \left(\frac{k}{2}\right)^{p-s} \ln^s(N+1) \left(\frac{\pi^2}{6}\right)^{p-s} \frac{1}{2^{p-s} 2^{p-1} (p-1)!} \\ & = 2N(k-1)! \frac{1}{(p-1)!} \sum_{s=0}^p \binom{p}{s} \left(\frac{k}{2}\right)^{p-s} \ln^s(N+1) \left(\frac{\pi^2}{6}\right)^{p-s} \frac{1}{4^{p-s} 2^s} \\ & = \frac{2N(k-1)!}{(p-1)!} \sum_{s=0}^p \binom{p}{s} \left(\frac{\pi^2 k}{48}\right)^{p-s} \left(\frac{\ln(N+1)}{2}\right)^s \\ & = 2N(k-1)! \frac{\left(\frac{\pi^2}{48} k + \frac{\ln(N+1)}{2}\right)^p}{(p-1)!} \end{aligned} \quad (45)$$

Finally, (45) is summed over all possible p and we obtain

$$\begin{aligned} \mathbb{E}(\text{tr}(\mathbf{M}^k)) &= \sum_{i_1, i_2, \dots, i_k} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ &\leq \sum_{p=1}^{k/2} \sum_{s_1 \dots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathbb{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1}) \\ &\leq 2N(k-1)! \sum_{p=1}^{k/2} \frac{\left(\frac{\pi^2}{48} k + \frac{\ln(N+1)}{2}\right)^p}{(p-1)!} \end{aligned} \quad (46)$$

By using the fact that, for any $A > 0$,

$$\sum_{p=1}^{k/2} \frac{A^p}{(p-1)!} = A \left(1 + A + \frac{A^2}{2!} + \dots + \frac{A^{k/2-1}}{(k/2-1)!}\right) \leq A e^A,$$

(46) is rearranged into

$$\begin{aligned} \mathbb{E}(\text{tr}(\mathbf{M}^k)) &\leq 2N(k-1)! \left(\frac{\pi^2}{48}k + \frac{\ln(N+1)}{2} \right) e^{\frac{\pi^2}{48}k + \frac{\ln(N+1)}{2}} = 2N\sqrt{N+1}(k-1)! \left(\frac{\pi^2}{48}k + \frac{\ln(N+1)}{2} \right) e^{\frac{\pi^2}{48}k} \\ &\leq 2(N+1)^{\frac{3}{2}}k^k \left(\frac{\pi^2}{48} + \frac{\ln(N+1)}{2k} \right) e^{\frac{\pi^2}{48}k}. \end{aligned}$$

Let k be the smallest even integer greater than $\frac{24}{\pi^2} \ln(N+1)$. Then using $\|\mathbf{M}\|_2 \leq (\text{tr}(\mathbf{M}^k))^{1/k}$ lead to

$$\begin{aligned} \mathbb{E}(\|\mathbf{M}\|_2) &\leq \mathbb{E}((\text{tr}(\mathbf{M}^k))^{1/k}) \leq (\mathbb{E}(\text{tr}(\mathbf{M}^k)))^{1/k} \leq (2(N+1)^{\frac{3}{2}})^{1/k} k \left(\frac{\pi^2}{48} + \frac{\ln(N+1)}{2k} \right)^{1/k} e^{\frac{\pi^2}{48}} \\ &\leq 2^{\frac{24}{\pi^2} \ln(N+1)} \cdot e^{\frac{\pi^2}{16}} \cdot \frac{24}{\pi^2} \ln(N+1) \cdot \left(\frac{\pi^2}{24} \right)^{\frac{24}{24 \ln(N+1)}} \cdot e^{\frac{\pi^2}{48}} \leq C_1 \ln N, \end{aligned}$$

where the constant C_1 is some universal constant.

Next, we estimate the complex case. In this case, $\mathbf{g} \in \mathbb{C}^{2N-1}$, where both its real part and imaginary part have i.i.d. Gaussian entries. Write $\mathbf{g} = \boldsymbol{\xi} + \iota\boldsymbol{\eta}$, where $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^{2N-1}$ are real-valued random Gaussian vectors. From the real-valued case above, we derive

$$\mathbb{E}(\|\mathcal{G}\boldsymbol{\xi}\|_2) \leq C_1 \ln N, \quad \mathbb{E}(\|\mathcal{G}\boldsymbol{\eta}\|_2) \leq C_1 \ln N.$$

Therefore,

$$\mathbb{E}(\|\mathcal{G}\mathbf{g}\|_2) = \mathbb{E}(\|\mathcal{G}\boldsymbol{\xi} + \iota\mathcal{G}\boldsymbol{\eta}\|_2) \leq \mathbb{E}(\|\mathcal{G}\boldsymbol{\xi}\|_2) + \mathbb{E}(\|\mathcal{G}\boldsymbol{\eta}\|_2) \leq 2C_1 \ln N. \quad \square$$

3.6 Proof of Theorem 1

With Lemmas 1, 2, 4, and Theorem 3 in hand, we are in position to prove Theorem 1.

Proof of Theorem 1. Since (10) is equivalent to (4) by the relation $\mathbf{y} = \mathcal{D}\mathbf{x}$, we only need to prove that $\hat{\mathbf{y}} = \tilde{\mathbf{y}}$ for noise free data ($\|\hat{\mathbf{y}} - \tilde{\mathbf{y}}\|_2 \leq 2\delta/\epsilon$ for noisy data) with dominant probability. According to Lemma 1, we only need to prove (13). By Lemma 2,

$$\mathbb{P} \left(\min_{\mathbf{z} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \|\mathcal{B}\mathbf{z}\|_2 \geq \epsilon \right) \geq 1 - 2e^{-\frac{1}{2} \left(\lambda_M - w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3}) - \frac{\epsilon}{\sqrt{2}} \right)^2}.$$

Lemma 4, Theorem 3, and the inequality $\lambda_M \geq \frac{M}{\sqrt{M+1}}$ imply that

$$\lambda_M - w(\mathfrak{I}_\mathbb{R}(\hat{\mathbf{y}}) \cap \mathbb{S}_\mathbb{R}^{4N-3}) - \frac{\epsilon}{\sqrt{2}} \geq \frac{M}{\sqrt{M+1}} - 3C_1\sqrt{R}\ln N - \frac{\epsilon}{\sqrt{2}} \geq \sqrt{M-1} - 3C_1\sqrt{R}\ln N - \frac{\epsilon}{\sqrt{2}}.$$

When $M \geq (6C_1\sqrt{R}\ln N + \sqrt{2}\epsilon)^2 + 1$, we can easily get $\mathbb{P} \left(\min_{\mathbf{z} \in \mathfrak{I}(\hat{\mathbf{y}}) \cap \mathbb{S}_c^{2N-2}} \|\mathcal{B}\mathbf{z}\|_2 \geq \epsilon \right) \geq 1 - 2e^{-\frac{M-1}{8}}$. We get the desired result. \square

4 Extension to Structured Low-Rank Matrix Reconstruction

In this section, we extend our results to low-rank Hankel matrix reconstruction and low-rank Toeplitz matrix reconstruction from their Gaussian measurements.

Since the proof of Theorem 1 does not use the specific property that $\hat{\mathbf{y}}$ is an exponential signal, Theorem 1 holds true for any low-rank Hankel matrices. We have the following corollary, which reads that any Hankel matrix of size $N \times N$ and rank R can be recovered exactly from its $O(R \ln^2 N)$ Gaussian measurements, and this reconstruction is robust to noise.

Corollary 1 (Low-Rank Hankel Matrix Reconstruction). *Let $\hat{\mathbf{H}} \in \mathbb{C}^{N \times N}$ be a given Hankel matrix with rank R . Let $\hat{\mathbf{x}} \in \mathbb{C}^{2N-1}$ be satisfying $\hat{x}_{i+j} = \hat{H}_{ij}$ for $0 \leq i, j \leq N-1$. Let $\mathcal{A} = \mathcal{B}\mathcal{D} \in \mathbb{C}^{M \times (2N-1)}$, where $\mathcal{B} \in \mathbb{C}^{M \times (2N-1)}$ is a random matrix whose real and imaginary parts are i.i.d. Gaussian with mean 0 and variance 1, $\mathcal{D} \in \mathbb{R}^{(2N-1) \times (2N-1)}$ is the same as defined in Theorem 1. Then, there exists a universal constant $C_1 > 0$ such that, for any $\epsilon > 0$, if*

$$M \geq (C_1 \sqrt{R} \ln N + \sqrt{2}\epsilon)^2 + 1,$$

then, with probability at least $1 - 2e^{-\frac{M-1}{8}}$, we have

(a) $\mathbf{H}(\tilde{\mathbf{x}}) = \hat{\mathbf{H}}$, where $\tilde{\mathbf{x}}$ is the unique solution of

$$\min_{\mathbf{x}} \|\mathbf{H}(\mathbf{x})\|_* \quad \text{subject to} \quad \mathcal{A}\mathbf{x} = \mathbf{b}$$

with $\mathbf{b} = \mathcal{A}\hat{\mathbf{x}}$;

(b) $\|\mathbf{H}(\tilde{\mathbf{x}}) - \hat{\mathbf{H}}\|_F \leq 2\delta/\epsilon$, where $\tilde{\mathbf{x}}$ is the unique solution of

$$\min_{\mathbf{x}} \|\mathbf{H}(\mathbf{x})\|_* \quad \text{subject to} \quad \|\mathcal{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta$$

with $\|\mathbf{b} - \mathcal{A}\hat{\mathbf{x}}\|_2 \leq \delta$.

Moreover, Theorem 1 can be extended to the reconstruction of low-rank Toeplitz matrix from its Gaussian measurements. Let $\hat{\mathbf{T}} \in \mathbb{C}^{N \times N}$ be a Toeplitz matrix. Let $\hat{\mathbf{x}} \in \mathbb{C}^{2N-1}$ be a vector satisfying $\hat{x}_{N-1+(i-j)} = \hat{T}_{i,j}$ for $0 \leq i, j \leq N-1$. Let $\mathbf{P} \in \mathbb{C}^{N \times N}$ be an anti-diagonal matrix with anti-diagonals of 1. Then, it is easy to check that $\hat{\mathbf{T}} = \mathbf{H}(\hat{\mathbf{x}})\mathbf{P}$. Thus, we define a linear operator \mathbf{T} that maps a vector in \mathbb{C}^{2N-1} to a $N \times N$ Toeplitz matrix by $\mathbf{T}(\mathbf{x}) = \mathbf{H}(\mathbf{x})\mathbf{P}$. Since \mathbf{P} is a unitary matrix, one has $\|\mathbf{T}(\mathbf{x})\|_* = \|\mathbf{H}(\mathbf{x})\mathbf{P}\|_* = \|\mathbf{H}(\mathbf{x})\|_*$. Therefore, the above corollary can be adapted to low-rank Toeplitz matrices. We obtain the following corollary, which states that any Toeplitz matrix of size $N \times N$ and rank R can be recovered exactly from its $O(R \ln^2 N)$ Gaussian measurements, and this reconstruction is robust to noise.

Corollary 2 (Low-Rank Toeplitz Matrix Reconstruction). *Let $\hat{\mathbf{T}} \in \mathbb{C}^{N \times N}$ be a given Toeplitz matrix with rank R . Let $\hat{\mathbf{x}} \in \mathbb{C}^{2N-1}$ be the vector satisfying $\hat{x}_{N-1+(i-j)} = \hat{T}_{i,j}$ for $0 \leq i, j \leq N-1$. Let $\mathcal{A} = \mathcal{B}\mathcal{D} \in \mathbb{C}^{M \times (2N-1)}$, where $\mathcal{B} \in \mathbb{C}^{M \times (2N-1)}$ is a random matrix whose real and imaginary parts are i.i.d. Gaussian with mean 0 and variance 1, $\mathcal{D} \in \mathbb{R}^{(2N-1) \times (2N-1)}$ is the same as defined in Theorem 1. Then, there exists a universal constant $C_1 > 0$ such that, for any $\epsilon > 0$, if*

$$M \geq (C_1 \sqrt{R} \ln N + \sqrt{2}\epsilon)^2 + 1,$$

then, with probability at least $1 - 2e^{-\frac{M-1}{8}}$, we have

(a) $\mathbf{T}(\tilde{\mathbf{x}}) = \hat{\mathbf{T}}$, where $\tilde{\mathbf{x}}$ is the unique solution of

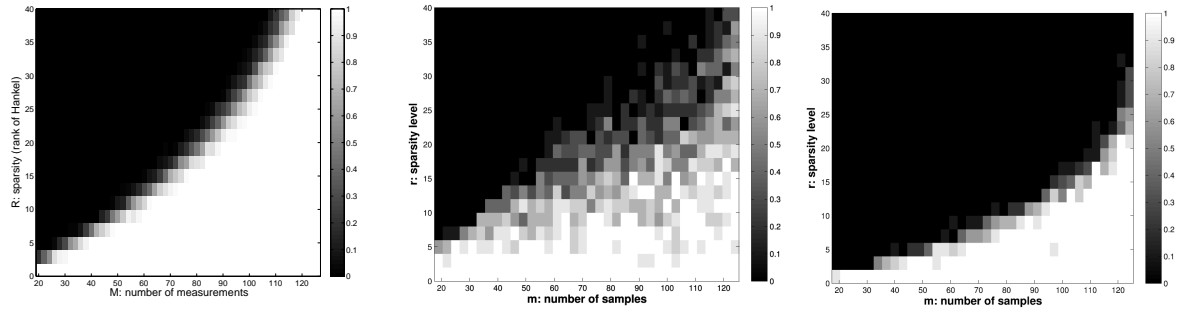
$$\min_{\mathbf{x}} \|\mathbf{T}(\mathbf{x})\|_* \quad \text{subject to} \quad \mathcal{A}\mathbf{x} = \mathbf{b}$$

with $\mathbf{b} = \mathcal{A}\hat{\mathbf{x}}$;

(b) $\|\mathbf{T}(\tilde{\mathbf{x}}) - \hat{\mathbf{T}}\|_F \leq 2\delta/\epsilon$, where $\tilde{\mathbf{x}}$ is the unique solution of

$$\min_{\mathbf{x}} \|\mathbf{T}(\mathbf{x})\|_* \quad \text{subject to} \quad \|\mathcal{A}\mathbf{x} - \mathbf{b}\|_2 \leq \delta$$

with $\|\mathbf{b} - \mathcal{A}\hat{\mathbf{x}}\|_2 \leq \delta$.



(a) Our method: Hankel nuclear norm minimization with random Gaussian projections. (b) Off-the-grid CS [19]: Atomic norm minimization with non-uniform sampling of entries. (c) EMaC [10]: Hankel nuclear norm minimization with non-uniform sampling of entries.

Figure 1: Numerical Results.

5 Numerical Experiments

In this section, we use numerical experiments to demonstrate our result and its performance improvement, compared with the results in [10, 19]. In the numerical experiments, we use superpositions of complex sinusoids as test signals. Note that the application of our result is not limited to such signals but any signals that are superpositions of complex exponentials.

The true signal $\hat{\mathbf{x}}$ is generated as follows. We choose $N = 64$, i.e., the dimension of $\hat{\mathbf{x}}$ is 127. The frequencies f_k , $k = 1, \dots, R$, are uniformly randomly drawn from the interval $[0, 1]$. The arguments of the coefficients c_k , $k = 1, \dots, R$, are from the interval $[0, 2\pi]$ uniformly at random, and their amplitudes are generated by $|c_i| = 1 + 10^{0.5m_i}$ where m_i follows the uniform distribution on $[0, 1]$. Then, we synthesize the true signal $\hat{\mathbf{x}}$ by $\hat{x}_t = \sum_{k=1}^R c_k e^{i2\pi f_k t}$ for $t = 0, 1, \dots, 126$. For each fixed M and R , we test 100 runs. We plot in Fig. 1(a) the rate of successful reconstruction by (4), which is solved by alternating direction method of multipliers (ADMM). We see from Fig. 1(a) that the phase transition of our method is very sharp.

For comparison, we plot the phase transitions of off-the-grid CS [19] and EMaC [10] in Fig. 1(b) and Fig. 1(c) respectively. These figures are from [10] under the same setting as ours. We observe that, for the same R , our method generally needs smaller M than off-the-grid CS and EMaC to achieve a high successful reconstruction rate. This illustrates that empirically our method requires fewer measurements than both off-the-grid CS and EMaC for the exact reconstruction of complex sinusoid signals. Finally and importantly, our method does not need a separation condition of frequencies to guarantee a successful recovery.

References

- [1] L. BORCEA, G. PAPANICOLAOU, C. TSOGKA, AND J. BERRYMAN, *Imaging and time reversal in random media*, *Inverse Problems*, 18 (2002), pp. 1247–1279.
- [2] J.-F. CAI, E. J. CANDÈS, AND Z. SHEN, *A singular value thresholding algorithm for matrix completion*, *SIAM J. Optimiz.*, 20 (2010), pp. 1956–1982.
- [3] E. CANDÈS, X. LI, Y. MA, AND J. WRIGHT, *Robust principal component analysis?*, *Journal of ACM*, (2011), pp. 1–37.
- [4] E. CANDÈS AND C. FERNANDEZ-GRANDA, *Towards a mathematical theory of super-resolution*, *Communications on Pure and Applied Mathematics* 67(6), 906–956.
- [5] E. CANDÈS AND Y. PLAN, *Matrix completion with noise*, *Proceedings of the IEEE*, (2009).

- [6] E. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717–772.
- [7] E. CANDÈS AND T. TAO, *The power of convex relaxation: Near-optimal matrix completion*, arXiv, 903 (2009).
- [8] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- [9] V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, AND A. S. WILLSKY, *The convex geometry of linear inverse problems*, Foundations of Computational Mathematics, 12 (2012), pp. 805–849.
- [10] Y. CHEN AND Y. CHI, *Robust spectral compressed sensing via structured matrix completion*, IEEE Transactions on Information Theory, 60 (2014), pp. 6576 – 6601.
- [11] Y. CHI, L. L. SCHARF, A. PEZESHKI, AND A. R. CALDERBANK, *Sensitivity to basis mismatch in compressed sensing*, IEEE Trans. Signal Process., 59 (2011), pp. 2182–2195.
- [12] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [13] Y. GORDON, *On Milman’s inequality and random subspaces which escape through a mesh in \mathbf{R}^n* , in Geometric aspects of functional analysis (1986/87), vol. 1317 of Lecture Notes in Math., Springer, Berlin, 1988, pp. 84–106.
- [14] Y. HUA AND T. K. SARKAR, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, Acoustics, Speech and Signal Processing, IEEE Transactions on, 38 (1990), pp. 814–824.
- [15] M. LUSTIG, D. DONOHO, AND J. M. PAULY, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195.
- [16] X. QU, M. MAYZEL, J.-F. CAI, Z. CHEN, AND V. OREKHOV, *Accelerated nmr spectroscopy with low-rank reconstruction*, Angewandte Chemie International Edition, 54 (2015), pp. 852–854.
- [17] B. RECHT, M. FAZEL, AND P. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
- [18] R. T. ROCKAFELLAR, *Convex analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.
- [19] G. TANG, B. N. BHASKAR, P. SHAH, AND B. RECHT, *Compressive sensing off the grid*, IEEE Transactions on Information Theory, 59 (2013), pp. 7465–7490.
- [20] J. A. TROPP, J. N. LASKA, M. F. DUARTE, J. K. ROMBERG, AND R. G. BARANIUK, *Beyond nyquist: Efficient sampling of sparse bandlimited signals*, Information Theory, IEEE Transactions on, 56 (2010), pp. 520–544.