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

Jian-Feng Cai[a,1], **Xiaobo Qu**[b,2], **Weiyu Xu**[c,3], and **Gui-Bo Ye**[d]

[a]Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

[b]Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, P.O. Box 979, Xiamen 361005, China

[c]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, United States

[d]Department of Mathematics, University of Iowa, Iowa City, IA 52242, United States

Abstract

This paper explores robust recovery of a superposition of R distinct complex exponential functions with or without damping factors from a few random Gaussian projections. We assume that the signal of interest is of $2N-1$ dimensions and $R < 2N-1$. 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.

Keywords

Complex sinusoids; Random Gaussian projection; Low-rank Hankel matrix

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 [16], analog-to-digital conversion [25], 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 [19]. How to recover those superpositions 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 $\hat{x} \in \mathbb{C}^{2N-1}$ be a vector satisfying

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

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

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

$$b = A\hat{x}, \tag{2}$$

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

We will use a Hankel structure to reconstruct the signal of interest $\hat{x}$. The Hankel structure originates from the matrix pencil method [15] for harmonic retrieval for complex sinusoid. The conventional matrix pencil method assumes fully observed $\hat{x}$ as well as the model order R, which are both unknown here. Following the ideas of the matrix pencil method in [15] and enhanced matrix completion (EMaC) in [10], we construct a Hankel matrix based on signal $\hat{x}$. More specifically, define the Hankel matrix $\hat{H} \in \mathbb{C}^{N \times N}$ by

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

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

Low rank matrix recovery has been widely studied [2,5,6,20]. 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 $x \in \mathbb{C}^{2N-1}$, let $H(x) \in \mathbb{C}^{N \times N}$ be the Hankel matrix whose first row and last column is x, i.e., $[H(x)]_{jk} = x_{j+k}$. We propose to solve

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

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

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

we solve

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

where $\delta = \|\eta\|_2$ is the noise level. The reconstruction of low-rank Hankel matrices via nuclear norm minimization were also proposed in [13] for system identification and realization.

An important theoretical question is how many measurements are required to get a robust reconstruction of $\hat{H}$ via (4) or (5). For a generic unstructured $N \times N$ matrix of rank R, standard theory [6,7,9,20] indicates that $O(NR \cdot poly(\log N))$ measurements are needed for a robust reconstruction by nuclear norm minimization. This result, however, is unacceptable here since the number of parameters of $\hat{H}$ is only $2N-1$ with the actual degrees of freedom R. The main contribution of this paper is then to prove that (4) and (5) give a robust recovery of $\hat{H}$ (hence $\hat{x}$) as soon as the number of projections exceeds $O(R\ln^2 N)$ if we choose the linear operator 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 only requiring $O(R\ln^2 N)$ measurements

instead of $O(R\ln^4 N)$ in [10]. Furthermore, our theoretical result provides some guidance on how many samples to choose for the model proposed in [19] to recover NMR spectroscopy.

- **Complex sinusoids.** When $|z_k| = 1$ for $k = 1, \ldots, R$, we must have $z_k = e^{2\pi \iota f_k}$ for some frequency f_k. In this case, $\hat{x}$ is a superposition of complex sinusoids, for example, in the analog-to-digital conversion of radio signals [25]. We often encounter the problem of signal recovery from compressed linear measurements of the superposition of complex sinusoids in various applications. For example, in compressed sensing of spectrally sparse bandlimited signals [25], the random demodulator obtains linear mixing measurements of spectrally sparse bandlimited signals through matching filters. We refer the reader to Sections III and IV of [25] for details. In array signal processing for Direction of Arrival (DoA) estimation of electromagnetic waves [26], the signals received at the antennas of the antenna array are a superposition of complex sinusoids with different frequencies. Suppose that the battery-powered antenna array aims to save energy in sending the measurements to the fusion center where the DoAs are calculated, the antenna array can send linear projections of the signals received across the antenna array. Moreover, if the antenna array is a non-uniform antenna array, the observations across the antenna array are also linear non-uniform compressive sampling of the uniform antenna array.

The problem on recovering $\hat{x}$ 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, $\hat{x}$ is sparse in the discrete Fourier transform domain, and CS theory [8,12] suggests that it is possible to reconstruct $\hat{x}$ 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, the authors of [4,23] 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 $\hat{x}$ from a non-uniform sampling of very few entries of $\hat{x}$, provided that the frequencies f_k, $k = 1, \ldots, R$, 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 $\hat{x}$ can be robustly reconstructed via EMaC from its very few non-uniformly 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 in this paper does not need any separation condition on the frequencies, while achieving better or comparable bounds on the number of measurements.

- **Accelerated NMR spectroscopy.** When $z_k = e^{-\tau_k} e^{2\pi \iota f_k}$, $k = 1, \ldots, R$, $\hat{x}$ 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. [19] proposed an algorithm based on low rank Hankel matrix. In this specific application, A is a matrix that denotes the under-sampling of NMR signals in the time domain. We remark that linear non-uniform subsampling measurements of the signal can greatly speed up the NMR spectroscopy [19]. Numerical results show its efficiency in [19] for which theoretical guarantee results are still needed. 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. The performance of our algorithm is demonstrated by numerical experiments in Section 5. Finally, in Section 6, we conclude the paper and point out some possible future works.

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 $\hat{x}$ has rank R. Let $\hat{H}$ be the Hankel matrix defined by (3). Eq. (1) leads to a decomposition

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

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

$$\min_{x} \mathrm{rank}(H(x)), \quad \text{subject to} \quad Ax = b, \tag{6}$$

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

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

where $\delta = \|\boldsymbol{\eta}\|_2$ 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 [6,7,9,20], 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).

Theoretical results are desirable to guarantee the success of this Hankel matrix completion method. The results in [6,7,9,20] do not consider the Hankel structure. For generic $N \times N$ rank-R matrix, they require $O(NR \cdot \text{poly}(\log N))$ measurements for robust recovery which is too much since there are only $2N - 1$ degrees of freedom in $\boldsymbol{H}(\boldsymbol{x})$. The theorems proposed in [23] 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 [4,10,23] 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 $\boldsymbol{A}$. In this paper, we provide a theoretical result ensuring exact recovery when $\boldsymbol{A}$ is a scaled random Gaussian matrix. Our result does not assume any incoherence conditions on the original signal.

Theorem 1—*Let $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{D} \in \mathbb{C}^{M \times (2N-1)}$, where $\boldsymbol{B} \in \mathbb{C}^{M \times (2N-1)}$ is a random matrix whose real and imaginary parts are i.i.d. Gaussian with mean 0 and variance 1, $\boldsymbol{D} \in \mathbb{R}^{(2N-1) \times (2N-1)}$ 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 $\varepsilon > 0$, if*

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

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

    a.   *$\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}}$, where $\tilde{\boldsymbol{x}}$ is the unique solution of (4) with $\boldsymbol{b} = \boldsymbol{A}\hat{\boldsymbol{x}}$;*

    b.   *$\|\boldsymbol{D}(\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}})\|_2 \leq 2\delta/\varepsilon$, where $\tilde{\boldsymbol{x}}$ is the unique solution of (5) with $\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2 \leq \delta$.*

The scaling matrix D is introduced to preserve energy. In particular, we will introduce a variable $y = Dx$, and then the operator $\mathscr{G}$ induced by $\mathscr{G}y = H(x)$ satisfies $\|y\|_2 = \|\mathscr{G}y\|_F$; see details in (9). This energy preserving property is critical in our estimation that will be seen later.

The number of measurements required is $O(R\ln^2 N)$, which is reasonable small compared with the number of parameters in $H(x)$. 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 $\varepsilon = O(\sqrt{R}\ln N)$. With this choice of ε, the number of measurements $M = O(R\ln^2 N)$ and the error $\|D(\tilde{x}-\hat{x})\|_2 \leq O\left(\frac{\delta}{\sqrt{M}}\right)$.

Compared to results in [10,23], our theorem does not require any incoherence condition of the matrix $\hat{H}$. In particular, our proposed approach for complex sinusoid signals does not need any separation condition on frequencies f_k's for $k = 1, 2, \ldots, R$. The reason for not needing a separation condition in noiseless case may be due to the Hankel matrix reconstruction method. Our proof of this fact also depends on the assumption of Gaussian measurements, for which we have the tool of Gaussian width analysis framework. For generic low-rank matrix reconstruction, it is well known that incoherence condition is necessary for successful reconstruction if partial entries of the underlying matrix are sampled [6,7]; however, incoherence is not required if Gaussian random projections are used [9,20]. We are in the same situation except for the additional Hankel structure. Our proposed approach uses Gaussian random projections of $\hat{H}$, while the methods in [10,23] sample partial entries $\hat{H}$. However, empirically, even for non-uniform time-domain samples, we observe that Hankel matrix completion does not seem to require the separation condition between frequencies. We thus conjecture that Hankel matrix completion does not require separations-between-frequencies condition to recover missing data from noiseless measurements under non-uniform time-domain samples, for which we currently do not have a proof.

2.1. Hankel matrix completion for recovering off-the-grid frequencies

Our results also apply to recovering frequencies in superposition of complex sinusoids, instead of recovering only the superposition of complex sinusoids. We divide our discussion into two cases.

The first case is the noise-free case, where the observations are not contaminated by additive noises. In this case, since we can recover the full signal of the superposition of the underlying sinusoids, we can use the single-snapshot MUSIC algorithm [27] to recover the underlying frequencies precisely.

The second case is the noisy case, where the observation is contaminated by additive noises. For this case, we have obtained a bound on the recovery error for the superposition signal

(Theorem 1 of our paper). We can further recover the frequencies using the single-snapshot MUSIC algorithm by choosing the R smallest local minimum of surrogate criterion function $R(\omega)$ in [27]. In [27], the authors provided the stability result of recovering frequency using the single-snapshot MUSIC algorithm (Theorem 3 of [27]). Specifically, the error in surrogate criterion function $R(\omega)$ is upper bounded by the Euclidean norm of the observation noise multiplied by a constant C, where the constant C depends on the largest and smallest nonzero singular values of the involved Hankel matrix. Moreover, the recovered frequency deviates from the true frequency in the order of noise standard deviation when the noise is small (Remark 9 of [27]). We remark that this stability result from [27] is applicable without imposing separation condition on frequencies.

For frequencies satisfying a certain separation condition (Equation (23) of [27]), the authors of [27] further provide stronger and more explicit bounds on the stability of recovering frequencies from noisy data (by explicitly bounding the singular values of the involved Hankel matrix).

Moitra [28] proved that stability of recovering frequencies from noisy observations depends on the separation of frequencies. In particular, [28] shows a sharp phase transition for the relationship between the cutoff time observation index m (namely $2N - 1$ in this paper) and the frequency separation δ. If $m > 1/\delta + 1$, there is a polynomial-complexity estimator converging to the true frequencies at an inverse polynomial rate in terms of the magnitude of the noise. And conversely, when $m < (1 - \epsilon)/\delta$, no estimator can distinguish between a particular pair of δ-separated signals if the magnitude of the noise is not exponentially small.

However, the converse results in [28] are dealing with worst-case frequencies and worst-case frequency coefficients. Namely if the separation condition is not satisfied, one can always finds a worst-case pair of signals x and x' such that telling them apart requires exponentially small noise. Thus Moitra's result in [28] is not for an average-case, fixed signal x. Moreover, Moitra's results does not mean that the single-snapshot MUSIC cannot tolerate small noises in recovering frequencies. By comparison, in this paper, our stability result is an average-case stability result, where our spectrally sparse signal is a fixed signal of superposition of complex exponentials, and our stability result is obtained over the ensemble of random Gaussian measurements. Our results are especially useful when the observations are noiseless or have high SNR.

3. Proof of Theorem 1

In this section, we prove the main result Theorem 1. The most crucial factors are that i) one has an explicit formula for the subdifferential of the objective function, and ii) the Gaussian width under the current measurement model is computable.

3.1. Orthonormal basis of the N × 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 $\boldsymbol{E}_j \in \mathbb{C}^{N \times N}$, $j = 0, 1 \ldots, 2N - 2$, be the Hankel matrix satisfying

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

where $K_j = j + 1$ for $j \quad N - 1$ and $K_j = 2N - 1 - j$ for $j \quad N - 1$ is the number of non-zeros in $\boldsymbol{E}_j$. Then, it is easy to check that $\{\boldsymbol{E}_j\}_{j=0}^{2N-2}$ 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

$$\mathscr{G} : \boldsymbol{x} \in \mathbb{C}^{2N-1} \mapsto \mathscr{G}\boldsymbol{x} = \sum_{j=0}^{2N-2} x_j \boldsymbol{E}_j \in \mathbb{C}^{N \times N}. \tag{9}$$

The adjoint $\mathscr{G}^*$ of $\mathscr{G}$ is

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

Obviously, $\mathscr{G}^* \mathscr{G}$ is the identity operator in $\mathbb{C}^{2N-1}$, and $\mathscr{G}\mathscr{G}^*$ 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 $\boldsymbol{D} \in \mathbb{C}^{(2N-1) \times (2N-1)}$, the diagonal matrix with j-th diagonal $\sqrt{K_j}$. Then, by letting $\boldsymbol{y} = \boldsymbol{Dx}$, (4) is rewritten as,

$$\min_{\boldsymbol{y}} \|\mathscr{G}\boldsymbol{y}\|_* \quad \text{subject to} \quad \boldsymbol{By} = \boldsymbol{b}, \tag{10}$$

where $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{D}^{-1}$. Recall that $\mathscr{G}$ satisfies $\mathscr{G}^* \mathscr{G} = \mathscr{Q}$, which is crucial in the proceeding analysis. Similarly, for the noisy case, (5) is rearranged to

$$\min_{\boldsymbol{y}} \|\mathscr{G}\boldsymbol{y}\|_* \quad \text{subject to} \quad \|\boldsymbol{By} - \boldsymbol{b}\|_2 \leq \varepsilon. \tag{11}$$

By our assumption in Theorem 1, $\boldsymbol{B} \in \mathbb{C}^{M \times (2N-1)}$ 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 $\tilde{\boldsymbol{y}} = \boldsymbol{D}\hat{\boldsymbol{x}}$ (respectively $\|\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}}\|_2 \le 2\delta/\varepsilon$) with dominant probability for problem (10) for the noise free case (respectively (11) for the noisy case).

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

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

To characterize the recovery condition, we need to use the minimum value of $\frac{\|\boldsymbol{B}\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}$ for nonzero $\boldsymbol{z} \in \mathfrak{T}(\hat{\boldsymbol{y}})$. This quantity is commonly called the *minimum gain* of the measurement operator $\boldsymbol{B}$ restricted on $\mathfrak{T}(\hat{\boldsymbol{y}})$ [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 $\mathfrak{T}(\hat{\boldsymbol{y}})$ be defined by* (12). *Assume*

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

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

    b. *Let $\tilde{\boldsymbol{y}}$ be the solution of* (11) *with $\|\boldsymbol{b} - \boldsymbol{B}\hat{\boldsymbol{y}}\|_2 \le \delta$. Then $\|\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}}\|_2 \le 2\delta/\varepsilon$.*

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

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

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 $\min_{\boldsymbol{z} \in \mathfrak{T}(\hat{\boldsymbol{y}})} \frac{\|\boldsymbol{B}\boldsymbol{z}\|_2}{\|\boldsymbol{z}\|_2}$. Gordon gave a solution using Gaussian width of a set [9,14] to estimate the lower bound of minimum gain.

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

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

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 $\lambda_n = \sqrt{2}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$ 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 [14]. It gives a bound on minimum gain for a random map $\Pi: \mathbb{R}^p \mapsto \mathbb{R}^n$.

Theorem 2. (See Corollary 1.2 in [14].)—*Let Ω be a closed subset of $\{x \in \mathbb{R}^p | \|x\|_2 = 1\}$. Let $\Pi \in \mathbb{R}^{n \times p}$ be a random matrix with i.i.d. Gaussian entries with mean 0 and variance 1. Then, for any $\varepsilon > 0$,*

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

provided $\lambda_n - w(\Omega) - \varepsilon \geq 0$. 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 $\mathfrak{T}_{\mathbb{R}}(\hat{y}) \cap \mathbb{S}_{\mathbb{R}}^{4N-3}$, where $\mathfrak{T}_{\mathbb{R}}(\hat{y})$ is a cone in $\mathbb{R}^{4N-2}$ defined by

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

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

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

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

<u>Proof:</u> 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 $B = \Phi + \iota\Psi \in \mathbb{C}^{M \times (2N-1)}$, where both $\Phi \in \mathbb{R}^{M \times (2N-1)}$ and $\Psi \in \mathbb{R}^{M \times (2N-1)}$ are real-valued random matrices whose entries are i.i.d. mean-0 variance-1 Gaussian. Then, for any $z = \alpha + \iota\beta \in \mathbb{C}^{2N-1}$ with $\alpha, \beta \in \mathbb{R}^{2N-1}$,

$$\|\boldsymbol{Bz}\|_2 = \|(\boldsymbol{\Phi}+\iota\boldsymbol{\Psi})(\boldsymbol{\alpha}+\iota\boldsymbol{\beta})\|_2 = \|(\boldsymbol{\Phi}\boldsymbol{\alpha}-\boldsymbol{\Psi}\boldsymbol{\beta})+\iota(\boldsymbol{\Psi}\boldsymbol{\alpha}+\boldsymbol{\Phi}\boldsymbol{\beta})\|_2$$

$$= \left(\left\| [\ \boldsymbol{\Phi}\ \ -\boldsymbol{\Psi}\ ] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2 + \left\| [\ \boldsymbol{\Psi}\ \ \boldsymbol{\Phi}\ ] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2 \right)^{-1/2}$$

Then

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

$$(15)$$

implies

$$\min_{\boldsymbol{z}\in\mathfrak{I}(\hat{\boldsymbol{y}})\cap\mathbb{S}_c^{2N-2}} \|\boldsymbol{Bz}\|_2 \geq \varepsilon.$$

Therefore,

$$\mathrm{P}\left(\min_{\boldsymbol{z}\in\mathfrak{I}(\hat{\boldsymbol{y}})\cap\mathbb{S}_c^{2N-2}} \|\boldsymbol{Bz}\|_2 \geq \varepsilon \right) \geq \mathrm{P}\left((15)\ \mathrm{holds\ true} \right).$$

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

$$\mathrm{P}\left((15)\ \mathrm{holds\ true} \right) \geq 1 - 2e^{-\frac{1}{2}\left(\lambda_M - w(\mathfrak{I}_{\mathbb{R}}(\hat{\boldsymbol{y}})\cap\mathbb{R}^{4N-3}) - \frac{\varepsilon}{\sqrt{2}} \right)^2},$$

and therefore we get the desired result.

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

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

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

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

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

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

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

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

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

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

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

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

According to Theorem 23.4 in [21], the cone dual to the descent cone is the conic hull of subgradient, which is exactly (19).

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

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

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

where $\mathrm{E}(\|\mathscr{G}g\|_2)$ is the expectation with respect to $g \in \mathbb{C}^{2N-1}$. Here 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 $\partial\mathscr{F}(\hat{\boldsymbol{\omega}})$ and thus $\mathfrak{T}_{\mathbb{R}}^*(\hat{y})$. Let $\hat{\boldsymbol{\Omega}}_1 = \mathscr{G}\hat{\boldsymbol{\omega}}_1$ and $\hat{\boldsymbol{\Omega}}_2 = \mathscr{G}\hat{\boldsymbol{\omega}}_2$. Then $\mathscr{G}\hat{\boldsymbol{y}} = \hat{\boldsymbol{\Omega}}_1 + \iota\hat{\boldsymbol{\Omega}}_2$. Let a singular value decomposition of the rank-R matrix $\mathscr{G}\hat{\boldsymbol{y}}$ be

$$\mathscr{G}\hat{\boldsymbol{y}} = U\sum V^*, \quad \text{with} \quad U = \boldsymbol{\Theta}_1 + \iota\boldsymbol{\Theta}_2, \ \ V = \boldsymbol{\Xi}_1 + \iota\boldsymbol{\Xi}_2, \quad (20)$$

where $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\Xi}_1, \boldsymbol{\Xi}_2 \in \mathbb{R}^{N \times R}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{R \times R}$, and $U \in \mathbb{C}^{N \times R}$ and $V \in \mathbb{C}^{N \times R}$ satisfies $U^*U = V^*V = I$. Then, by direct calculation,

$$\boldsymbol{\Theta} \equiv \begin{bmatrix} \boldsymbol{\Theta}_1 & -\boldsymbol{\Theta}_2 \\ \boldsymbol{\Theta}_2 & \boldsymbol{\Theta}_1 \end{bmatrix} \in \mathbb{R}^{2N \times (2R)}, \quad \boldsymbol{\Xi} \equiv \begin{bmatrix} \boldsymbol{\Xi}_1 & -\boldsymbol{\Xi}_2 \\ \boldsymbol{\Xi}_2 & \boldsymbol{\Xi}_1 \end{bmatrix} \in \mathbb{R}^{2N \times (2R)} \quad (21)$$

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

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

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

$$\mathscr{F}(\hat{\boldsymbol{\omega}}) = \|\mathscr{G}\hat{\boldsymbol{y}}\|_* = \|\sum\|_* = \frac{1}{2}\|\hat{\boldsymbol{\Omega}}\|_*. \quad (23)$$

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

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

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

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

On the one hand, the adjoint $\mathscr{E}^*$ 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 $\mathbb{R}^{N \times N}$,

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

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

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

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

$$\partial \mathscr{F}(\hat{\boldsymbol{\omega}}) = \left\{ \begin{bmatrix} \mathscr{G}^*\left(\boldsymbol{\Theta}_1 \boldsymbol{\Xi}_1^T + \boldsymbol{\Theta}_2 \boldsymbol{\Xi}_2^T + \frac{\Delta_{11} + \Delta_{22}}{2}\right) \\ \mathscr{G}^*\left(\boldsymbol{\Theta}_2 \boldsymbol{\Xi}_1^T - \boldsymbol{\Theta}_1 \boldsymbol{\Xi}_2^T + \frac{\Delta_{21} - \Delta_{12}}{2}\right) \end{bmatrix} \Big| \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \ \boldsymbol{\Theta}^T \Delta = \mathbf{0}, \ \Delta \boldsymbol{\Xi} = \mathbf{0}, \ \|\Delta\|_2 \leq 1 \right\}.$$

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

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

where $\boldsymbol{U}, \boldsymbol{V}$ are in (20). Then, it can be checked that

$$\mathfrak{H} \equiv \left\{ \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \Big| \boldsymbol{\alpha} + \iota \boldsymbol{\beta} \in \mathfrak{S} \right\} \subset \partial \mathscr{F}(\hat{\boldsymbol{\omega}}). \quad (28)$$

Actually, for any $\boldsymbol{W} = {}_1 + \iota {}_2$ satisfying $\boldsymbol{U}^*\boldsymbol{W} = 0$, $\boldsymbol{WV} = 0$ and $\|\boldsymbol{W}\|_2 \quad 1$, we choose $\Delta = \begin{bmatrix} \Delta_1 & -\Delta_2 \\ \Delta_2 & \Delta_1 \end{bmatrix}$. Obviously, this choice of satisfies the constraints on in $\mathscr{F}(\hat{\boldsymbol{\omega}})$. Furthermore, $\boldsymbol{UV}^* + \boldsymbol{W} = (\boldsymbol{\Theta}_1 \boldsymbol{\Xi}_1^T + \boldsymbol{\Theta}_2 \boldsymbol{\Xi}_2^T + \Delta_1) + \iota(\boldsymbol{\Theta}_2 \boldsymbol{\Xi}_1^T + \boldsymbol{\Theta}_1 \boldsymbol{\Xi}_2^T + \Delta_2)$. Therefore, (28) holds.

With the help of (28), we get

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

We then convert the real-valued vectors to complex-valued vectors by letting $\boldsymbol{g} = \boldsymbol{\xi}_1 + \iota\boldsymbol{\xi}_2$ and $\boldsymbol{c} = \boldsymbol{\gamma}_1 + \iota\boldsymbol{\gamma}_2$, where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are the first and second half of $\boldsymbol{\xi}$ respectively and so for $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. This leads to

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

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

$$\|\boldsymbol{g} - \lambda\boldsymbol{c}\|_2 = \|\mathscr{G}\boldsymbol{g} - \lambda\mathscr{G}\boldsymbol{c}\|_F = \|\mathscr{G}\boldsymbol{g} - \lambda\mathscr{G}\mathscr{G}^*(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W})\|_F$$
$$= (\|\mathscr{G}\boldsymbol{g} - \lambda(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W})\|_F^2 - \|\lambda(\mathscr{I} - \mathscr{G}\mathscr{G}^*)(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W})\|_F^2)^{1/2}$$
$$\leq \|\mathscr{G}\boldsymbol{g} - \lambda(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W})\|_F , \tag{30}$$

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

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

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

$$\mathscr{P}_1\boldsymbol{U}\boldsymbol{V}^* = \boldsymbol{U}\boldsymbol{V}^*, \quad \mathscr{P}_2\boldsymbol{U}\boldsymbol{V}^* = 0, \quad \mathscr{P}_1\boldsymbol{W} = 0, \quad \mathscr{P}_2\boldsymbol{W} = \boldsymbol{W}, \tag{31}$$

where $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ are the same as those in (27). We choose

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

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

$$\min_{\boldsymbol{\gamma} \in \mathfrak{T}_{\mathbb{R}}^*(\hat{\boldsymbol{y}})} \|\boldsymbol{\xi} - \boldsymbol{\gamma}\|_2 \leq \|\mathscr{G}\boldsymbol{g} - \|\mathscr{P}_2(\mathscr{G}\boldsymbol{g})\|_2\boldsymbol{U}\boldsymbol{V}^* - \mathscr{P}_2(\mathscr{G}\boldsymbol{g})\|_F = \|\mathscr{P}_1(\mathscr{G}\boldsymbol{g}) - \|\mathscr{P}_2(\mathscr{G}\boldsymbol{g})\|_2\boldsymbol{U}\boldsymbol{V}^*\|_F$$
$$\leq \|\mathscr{P}_1(\mathscr{G}\boldsymbol{g})\|_F + \|\mathscr{P}_2(\mathscr{G}\boldsymbol{g})\|_2\|\boldsymbol{U}\boldsymbol{V}^*\|_F = \|\mathscr{P}_1(\mathscr{G}\boldsymbol{g})\|_F + \sqrt{R}\|\mathscr{P}_2(\mathscr{G}\boldsymbol{g})\|_2 .$$

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

$$\|\mathscr{P}_1(\mathscr{G}g)\|_F = \|UU^*(\mathscr{G}g) + (\mathscr{G}g)VV^* - UU^*(\mathscr{G}g)VV^*\|_F = \|UU^*(\mathscr{G}g) + (I - UU^*)(\mathscr{G}g)VV^*\|_F$$
$$\leq \|UU^*(\mathscr{G}g)\|_F + \|(I - UU^*)(\mathscr{G}g)VV^*\|_F \leq \|UU^*(\mathscr{G}g)\|_F + \|(\mathscr{G}g)VV^*\|_F$$
$$\leq 2\sqrt{R}\|\mathscr{G}g\|_2$$

where in the last line we have used the inequality

$$\|UU^*(\mathscr{G}g)\|_F \leq \|UU^*\|_F \|\mathscr{G}g\|_2 \leq \sqrt{R}\|\mathscr{G}g\|_2$$

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

$$\|\mathscr{P}_2(\mathscr{G}g)\|_2 = \|(I - UU^*)(\mathscr{G}g)(I - VV^*)\|_2 \leq \|I - UU^*\|_2 \|\mathscr{G}g\|_2 \|I - VV^*\|_2 \leq \|\mathscr{G}g\|_2.$$

Altogether, we obtain

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

which together with (17) gives

$$w(\mathfrak{T}_{\mathbb{R}}(\hat{\boldsymbol{y}}) \cap \mathbb{R}^{4N-3}) \leq 3\sqrt{R} \cdot \mathrm{E}(\|\mathscr{G}g\|_2).$$

3.5. Bound of $\mathrm{E}(\|\mathscr{G}g\|_2)$

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

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

$$\mathrm{E}(\|\mathscr{G}g\|_2) \leq C_1 \ln N,$$

where C_1 are some positive universal constants.

We will use the moment method (see Chapter 2.3 in [24] for more details) to prove Theorem 3. In order to help the reader easily understand the proof, we begin with the real case and introduce some ideas and lemmas first. Assume $g \in \mathbb{R}^{2N-1}$ has i.i.d standard Gaussian entries with mean 0 and variance 1. Notice that $\mathscr{G}g$ is symmetric. Therefore, for any even

integer k, $(\text{tr}\,(\mathscr{G}\boldsymbol{g})^k)^{1/k}$ is the k-norm of vector of singular values, which implies $\|\mathscr{G}\boldsymbol{g}\|_2$ (tr $(\mathscr{G}\boldsymbol{g})^k)^{1/k}$. This together with Jensen's inequality,

$$\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2) \leq \mathrm{E}\left(\left(\text{tr}\,(\mathscr{G}\boldsymbol{g})^k\right)^{1/k}\right) \leq \left(\mathrm{E}(\text{tr}\,(\mathscr{G}\boldsymbol{g})^k)\right)^{1/k}. \tag{32}$$

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

$$\mathrm{E}(\text{tr}(\boldsymbol{M}^k)) = \sum_{0 \leq i_1, i_2, \ldots, i_k \leq N-1} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1}). \tag{33}$$

Therefore, we only need to estimate Σ_0 $i_1, i_2,\ldots, i_k$ $N-1$ $\mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})$.

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 $i_1, i_2, \ldots, i_k$ and its equivalent edges on the graph. More specifically, we construct graph $\mathfrak{F}_{i_1, i_2,\ldots, i_k}$ with nodes to be $i_1, i_2, \ldots, i_k$ and edges to be (i_1, i_2), (i_2, i_3), ..., (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 $\mathfrak{F}_{i_1,i_2,\ldots,i_k}$. These equivalent classes are indexed by $1, 2, \ldots, p$ according to their order in the graph traversal $i_1 \rightarrow i_2 \rightarrow \ldots \rightarrow i_k \rightarrow i_1$. We associate with the graph $\mathfrak{F}_{i_1,i_2,\ldots,i_k}$ a sequence $c_1 c_2 \ldots c_k$, where c_j is the index of the equivalent class of the edge (i_j, i_{j+1}). We call $c_1 c_2 \ldots c_k$ the label for the equivalent classes of the graph $\mathfrak{F}_{i_1,i_2,\ldots,i_k}$.

The label for the equivalent classes of the graph $\mathfrak{F}_{i_1,i_2,\ldots,i_k}$ plays an important role in bounding $\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2)$. 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 $i_1 i_2 \ldots i_k$ that correspond to the same label for the equivalent classes of the corresponding graph. Let $\mathfrak{A}_{c_1 c_2 \ldots c_k}$ be the set of indices whose label of equivalent class of the corresponding graph is $c_1 c_2 \ldots c_k$, i.e.

$$\mathfrak{A}_{c_1 c_2 \ldots c_k} = \{i_1 i_2 \ldots i_k | \text{the label for the equivalent class of the graph } \mathfrak{F}_{i_1, i_2, \ldots, i_k} \text{ is } c_1 c_2 \ldots c_k\}$$

$$\tag{34}$$

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

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

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

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

Proof: We begin with finding free indices for any $i_1, i_2, \ldots, i_k$ in the set $\mathfrak{A}_{c_1 c_2 \ldots c_k}$. 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 $k_p(j_1, j_2, \ldots, j_p)$ and a free vertex j_{p+1}, and the weight is $k_p(j_1, j_2, \ldots, j_p) + j_{p+1}$. 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 $i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}$,

$$\mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1}) = \prod_{\ell=1}^{p} \frac{1}{K_{m_\ell}^{s_\ell/2}} \mathrm{E}\left(g_{m_\ell}^{s_\ell}\right), \tag{36}$$

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

$$\mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1}) = \prod_{\ell=1}^{p} \frac{(s_\ell - 1)!!}{K_{m_\ell}^{s_\ell/2}}. \tag{37}$$

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

$$\sum_{i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1}) \leq \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1} \cdots \sum_{j_p=0}^{N-1} \sum_{j_{p+1}=0}^{N-1} \prod_{\ell=1}^{p} \frac{(s_\ell - 1)!!}{K_{m_\ell}^{s_\ell/2}}$$

$$= \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1} \left(\frac{(s_1-1)!!}{K_{k_1(j_1)+j_2}^{s_1/2}} \sum_{j_3=0}^{N-1} \left(\frac{(s_2-1)!!}{K_{k_2(j_1,j_2)+j_3}^{s_2/2}} \sum_{i_4=0}^{N-1} \left(\cdots \sum_{j_{p+1}=0}^{N-1} \frac{(s_p-1)!!}{K_{k_p(j_1,\ldots,j_p)+j_{p+1}}^{s_p/2}} \right) \cdots \right) \right)$$

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.

The desired bound for $\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2)$ can be obtained if we know how many different sets of $\mathfrak{A}_{c_1 c_2 \dots c_k}$ are 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.

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

(38)

Let $\mathfrak{C}_p$ be the 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

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{M}^k)) = \sum_{0 \leq i_1, i_2, \dots, i_k \leq N-1} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1})$$
$$\leq \sum_{p=1}^{k/2} \sum_{s_1 \dots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \dots c_k \in \mathfrak{B}_{s_1 s_2 \dots s_p}} \sum_{i_1 i_2 \dots i_k \in \mathfrak{A}_{c_1 c_2 \dots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \dots M_{i_{k-1} i_k} M_{i_k i_1})$$

(39)

By bounding the cardinality of $\mathfrak{B}_{s_1 s_2 \dots s_p}$ and $\mathfrak{C}_p$, we can derive the bound $\mathrm{E}(\mathrm{tr}(\boldsymbol{M}^k))$ hence $\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2)$ 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,

$$
|\mathfrak{B}_{s_1 s_2 \ldots s_p}| \leq \binom{k-1}{s_1-1} \cdot \binom{k-s_1-1}{s_2-1} \cdot \ldots \cdot \binom{k-s_1-\ldots-s_{p-1}-1}{s_p-1}
$$

$$
= \frac{(k-1)(k-2)\ldots(k-s_1+1)}{(s_1-1)!} \frac{(k-s_1-1)\ldots(k-s_1-s_2+1)}{(s_2-1)!} \ldots 1
$$

$$
= \frac{(k-1)!}{\prod_{\ell=1}^p (s_\ell-1)! \prod_{\ell=1}^{p-1}(k-s_1-\ldots-s_\ell)},
$$

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

$$
\sum_{c_1 c_2 \ldots c_k \in \mathfrak{B}_{s_1 s_2 \ldots s_p}} \sum_{i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})
$$

$$
\leq N \frac{(k-1)!}{\prod_{\ell=1}^p (s_\ell-2)!! \prod_{\ell=1}^{p-1}(k-s_1-\ldots-s_\ell)} \prod_{\ell=1}^p B(s_\ell) \tag{40}
$$

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

$$
\sum_{s_1 \ldots s_p \in \mathfrak{C}_p} \sum_{c_1 c_2 \ldots c_k \in \mathfrak{B}_{s_1 s_2 \ldots s_p}} \sum_{i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})
$$

$$
\leq N(k-1)! \sum_{s_1 \ldots s_p \in \mathfrak{C}_p} \frac{\prod_{\ell=1}^p B(s_\ell)}{\prod_{\ell=1}^p (s_\ell-2)!! \prod_{\ell=1}^{p-1}(k-s_1-\ldots-s_\ell)} \tag{41}
$$

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

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

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

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

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

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

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

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

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

$$\sum_{s_1\ldots s_p \in \mathfrak{C}_p c_1 c_2 \ldots c_k \in \mathfrak{B}_{s_1 s_2 \ldots s_p}} \sum_{i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})$$

$$\leq N(k-1)! \sum_{s=0}^{p} \begin{pmatrix} p \\ s \end{pmatrix} \left(\frac{k}{2}\right)^{p-s} \ln^s(N+1) \left(\frac{\pi^2}{6}\right)^{p-s} \frac{1}{2^{p-s} 2^{p-1}(p-1)!}$$

$$=2N(k-1)!\frac{1}{(p-1)!} \sum_{s=0}^{p} \begin{pmatrix} p \\ s \end{pmatrix} \left(\frac{k}{2}\right)^{p-s} \ln^s(N+1) \left(\frac{\pi^2}{6}\right)^{p-s} \frac{1}{4^{p-s} 2^s}$$

$$=\frac{2N(k-1)!}{(p-1)!} \sum_{s=0}^{p} \begin{pmatrix} p \\ s \end{pmatrix} \left(\frac{\pi^2 k}{48}\right)^{p-s} \left(\frac{\ln(N+1)}{2}\right)^s$$

$$=2N(k-1)!\frac{\left(\frac{\pi^2}{48}k+\frac{\ln(N+1)}{2}\right)^p}{(p-1)!} \tag{45}$$

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

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{M}^k))= \sum_{i_1,i_2,\ldots,i_k} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})$$

$$\leq \sum_{p=1}^{k/2} \sum_{s_1\ldots s_p \in \mathfrak{C}_p c_1 c_2 \ldots c_k \in \mathfrak{B}_{s_1 s_2 \ldots s_p}} \sum_{i_1 i_2 \ldots i_k \in \mathfrak{A}_{c_1 c_2 \ldots c_k}} \mathrm{E}(M_{i_1 i_2} M_{i_2 i_3} \ldots M_{i_{k-1} i_k} M_{i_k i_1})$$

$$\leq 2N(k-1)! \sum_{p=1}^{k/2} \frac{\left(\frac{\pi^2}{48}k+\frac{\ln(N+1)}{2}\right)^p}{(p-1)!} \tag{46}$$

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

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

(46) is rearranged into

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{M}^k)) \leq 2N(k-1)! \left(\frac{\pi^2}{48}k+\frac{\ln(N+1)}{2}\right) e^{\frac{\pi^2}{48}k+\frac{\ln(N+1)}{2}}=2N\sqrt{N+1}(k-1)! \left(\frac{\pi^2}{48}k+\frac{\ln(N+1)}{2}\right) e^{\frac{\pi^2}{48}k}$$

$$\leq 2(N+1)^{\frac{3}{2}} k^k \left(\frac{\pi^2}{48}+\frac{\ln(N+1)}{2k}\right) e^{\frac{\pi^2}{48}k}.$$

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

$$\mathrm{E}(\|\boldsymbol{M}\|_2) \leq \mathrm{E}((\mathrm{tr}(\boldsymbol{M}^k))^{1/k}) \leq \left(\mathrm{E}(\mathrm{tr}(\boldsymbol{M}^k))\right)^{1/k} \leq (2(N+1)^{\frac{3}{2}})^{1/k} k \left(\frac{\pi^2}{48}+\frac{\ln(N+1)}{2k}\right)^{1/k} e^{\frac{\pi^2}{48}}$$

$$\leq 2^{\frac{\pi^2}{24\ln(N+1)}} \cdot e^{\frac{\pi^2}{16}} \cdot \frac{24}{\pi^2}\ln(N+1) \cdot \left(\frac{\pi^2}{24}\right)^{\frac{\pi^2}{24\ln(N+1)}} \cdot e^{\frac{\pi^2}{48}} \leq C_1 \ln N,$$

where the constant C_1 is some universal constant.

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

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

Therefore,

$$\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2) = \mathrm{E}(\|\mathscr{G}\boldsymbol{\xi}+\iota\mathscr{G}\boldsymbol{\eta}\|_2) \leq \mathrm{E}(\|\mathscr{G}\boldsymbol{\xi}\|_2) + \mathrm{E}(\|\mathscr{G}\boldsymbol{\eta}\|_2) \leq 2C_1 \ln N.$$

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 $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}$, we only need to prove that $\hat{\boldsymbol{y}} = \check{\boldsymbol{y}}$ for noise free data ($\|\hat{\boldsymbol{y}}-\check{\boldsymbol{y}}\|_2 \quad 2\delta/e$ for noisy data) with dominant probability. According to Lemma 1, we only need to prove (13). By Lemma 2,

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

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

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

When $M \geq \left(6C_1\sqrt{R}\ln N + \sqrt{2}\varepsilon\right)^2 + 1$, we can easily get

$\mathrm{P}\left(\min_{\boldsymbol{z} \in \mathfrak{I}(\hat{\boldsymbol{y}}) \cap \mathbb{S}_c^{2N-2}} \|\boldsymbol{B}\boldsymbol{z}\|_2 \geq \varepsilon\right) \geq 1 - 2e^{-\frac{M-1}{8}}$. We get the desired result.

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

$$M \ge \left(C_1\sqrt{R}\ln N + \sqrt{2}\varepsilon\right)^2 + 1,$$

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

    a. $\boldsymbol{H}(\hat{\boldsymbol{x}}) = \hat{\boldsymbol{H}}$, *where $\hat{\boldsymbol{x}}$ is the unique solution of*

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_* \quad subject\ to \quad \boldsymbol{Ax} = \boldsymbol{b}$$

    with $\boldsymbol{b} = \boldsymbol{A}\hat{\boldsymbol{x}}$;

    b. $\|\boldsymbol{H}(\hat{\boldsymbol{x}}) - \hat{\boldsymbol{H}}\|_F \le 2\delta/\varepsilon$, *where $\hat{\boldsymbol{x}}$ is the unique solution of*

$$\min_{\boldsymbol{x}} \|\boldsymbol{H}(\boldsymbol{x})\|_* \quad subject\ to \quad \|\boldsymbol{Ax} - \boldsymbol{b}\|_2 \le \delta$$

    with $\|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2 \le \delta$.

Moreover, Theorem 1 can be extended to the reconstruction of low-rank Toeplitz matrix from its Gaussian measurements. Let $\hat{\boldsymbol{T}} \in \mathbb{C}^{N \times N}$ be a Toeplitz matrix. Let $\hat{\boldsymbol{x}} \in \mathbb{C}^{2N-1}$ be a vector satisfying $\hat{x}_{N-1+(i-j)} = \hat{T}_{i,j}$ for $0 \le i, j \le N-1$. Let $\boldsymbol{P} \in \mathbb{C}^{N \times N}$ be an anti-diagonal matrix with anti-diagonals of 1. Then, it is easy to check that $\hat{\boldsymbol{T}} = \boldsymbol{H}(\hat{\boldsymbol{x}})\boldsymbol{P}$. Thus, we define a linear operator $\boldsymbol{T}$ that maps a vector in $\mathbb{C}^{2N-1}$ to a $N \times N$ Toeplitz matrix by $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{H}(\boldsymbol{x})\boldsymbol{P}$. Since $\boldsymbol{P}$ is a unitary matrix, one has $\|\boldsymbol{T}(\boldsymbol{x})\|_* = \|\boldsymbol{H}(\boldsymbol{x})\boldsymbol{P}\|_* = \|\boldsymbol{H}(\boldsymbol{x})\|_*$. Therefore, the above corollary can be adapted to low-rank Toeplitz matrices.

5. Numerical experiments

In this section, we use numerical experiments to demonstrate the empirical performance of our proposed approach, and compare it with other methods including those in [10,23]. We use superpositions of complex sinusoids as test signals. Note that the application of our approach is not limited to such signals but any signals that are superpositions of complex exponentials. We are going to consider two sampling schemes, namely, the random Gaussian sampling model in Theorem 1 and the non-uniform sampling of entries studied in [10,23]. For the latter sampling scheme, it randomly observes M entries of $\hat{x}$ whose locations are uniformly distributed in all M-subsets of $\{0, 1, \ldots, 2N-2\}$. We also consider two signal reconstruction algorithms, the Hankel nuclear norm minimization and the atomic norm minimization, from the given samples. Therefore, we have four different approaches to compare: the Hankel nuclear norm minimization with random Gaussian sampling (our proposed approach), the Hankel nuclear norm minimization with non-uniform sampling of entries (EMaC in [10]), the atomic norm minimization with random Gaussian sampling, and the atomic norm minimization with non-uniform sampling of entries (off-the-grid CS in [23]).

We fix $N = 64$, i.e., the dimension of the true signal $\hat{x}$ is 127. We conduct experiments under different M and R for different approaches. For each approach with a fixed M and R, we test 100 runs, where each run is executed as follows. We first generate the true signal $\hat{x} = [\hat{x}(0),$ $\hat{x}(1), \ldots, \hat{x}(126)]^T$ with $\hat{x}(t) = \sum_{k=1}^{R} c_k e^{\iota 2\pi f_k t}$ for $t = 0, 1, \ldots, 126$, where f_k are frequencies drawn from the interval $[0, 1]$ uniformly at random, and c_k are complex coefficients satisfying the model $c_k = (1+10^{0.5 m_k}) e^{\iota 2\pi \theta_k}$ with m_k and θ_k being uniformly randomly drawn from the interval $[0, 1]$. Then we get M samples of $\hat{x}$ according to the corresponding sampling scheme. Finally, a reconstruction $\tilde{x}$ is obtained by solving the corresponding reconstruction algorithm, which is numerically implemented by alternating direction method of multipliers (ADMM). If $\frac{\|\tilde{x}-\hat{x}\|_2}{\|\hat{x}\|_2} \leq 10^{-3}$, then we regard it as a successful reconstruction.

We plot in Fig. 1 the rate of successful reconstruction with respect to different M and R for different approaches. The black and white region indicate a 0% and 100% of successful reconstruction respectively, and a grey between 0% and 100%. From the figure, we see that the atomic norm minimization has similar performance under the random Gaussian sampling and the non-uniform sampling of entries. Moreover, the Hankel nuclear norm minimization also has similar performance under these two types of different sampling schemes. Compared with the atomic norm minimization, the Hankel nuclear norm minimization method is more robust when neighboring frequencies are close, despite different sampling schemes used.

6. Conclusion and future works

In this paper, we study compressed sensing of signal that is a weighted sum of R complex exponential functions with or without damping factor. The measurements are obtained by random Gaussian projections. We prove that, as long as the number of measurements is greater than $O(R\ln^2 N)$ with N the dimension of the signal, minimization (4) is guaranteed to

get a robust reconstruction of the underlying signal. Compared to results in [10,23] where partial entries of the underlying signal are observed, our proposed approach does not require any incoherence condition (i.e. does not require any separation condition on frequencies for signals without damping).

The bound $O(R\ln^2 N)$ we obtained is not optimal. There are several possible direction to improve it. Firstly, we may improve the estimation of $\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2)$, a key step in our proof. We empirically observed that $\mathrm{E}(\|\mathscr{G}\boldsymbol{g}\|_2) = O(\sqrt{R\ln N})$, which is better than the bound in Theorem 3. We would prove this bound theoretically. Secondly, we may borrow techniques from compressed sensing to get the optimal bound under Gaussian measurements. Actually, there has been recent work that yields precise bounds [17,18,22]. These results assume Gaussian measurements, but all quantities involved are reals. It is interesting to explore whether those results can be extended to our setting.

Acknowledgments

References

1. Borcea L, Papanicolaou G, Tsogka C, Berryman J. Imaging and time reversal in random media. Inverse Probl. 2002; 18:1247–1279.

2. Cai JF, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J Optim. 2010; 20:1956–1982.

3. Candes E, Li X, Ma Y, Wright J. Robust principal component analysis? J ACM. 2011:1–37.

4. Candes E, Fernandez-Granda C. Towards a mathematical theory of super-resolution. Comm Pure Appl Math. 2014; 67(6):906–956.

5. Candes E, Plan Y. Matrix completion with noise. Proc IEEE. 2009; 98:925–936.

6. Candes E, Recht B. Exact matrix completion via convex optimization. Found Comput Math. 2009; 9:717–772.

7. Candes E, Tao T. The power of convex relaxation: near-optimal matrix completion. IEEE Trans Inform Theory. 2010; 56:2053–2080.

8. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inform Theory. 2006; 52:489–509.

9. Chandrasekaran V, Recht B, Parrilo PA, Willsky AS. The convex geometry of linear inverse problems. Found Comput Math. 2012; 12:805–849.

10. Chen Y, Chi Y. Robust spectral compressed sensing via structured matrix completion. IEEE Trans Inform Theory. 2014; 60:6576–6601.

11. Chi Y, Scharf LL, Pezeshki A, Calderbank AR. Sensitivity to basis mismatch in compressed sensing. IEEE Trans Signal Process. 2011; 59:2182–2195.

12. Donoho DL. Compressed sensing. IEEE Trans Inform Theory. 2006; 52:1289–1306.

13. Fazel M, Fazel Maryam, Pong TK, Sun D, Tseng P. Hankel matrix rank minimization with applications to system identification and realization. SIAM J Matrix Anal Appl. 2013; 34:946–977.

14. Gordon, Y. Geometric Aspects of Functional Analysis, 1986/87. Vol. 1317. Springer; Berlin: 1988. On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$; p. 84-106.Lecture Notes in Math

15. Hua Y, Sarkar TK. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. IEEE Trans Acoust Speech Signal Process. 1990; 38:814–824.

16. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med. 2007; 58:1182–1195. [PubMed: 17969013]

17. Oymak, S., Thrampoulidis, C., Hassibi, B. The squared-error of generalized LASSO: a precise analysis. 51st Annual Allerton Conference on Communication, Control, and Computing, Allerton; IEEE; 2013. p. 1002-1009.

18. Thrampoulidis, C., Oymak, S., Hassibi, B. Simple error bounds for regularized noisy linear inverse problems. IEEE International Symposium on Information Theory; IEEE; 2014. p. 3007-3011.

19. Qu X, Mayzel M, Cai JF, Chen Z, Orekhov V. Accelerated NMR spectroscopy with low-rank reconstruction. Angew Chem, Int Ed. 2015; 54:852–854.

20. Recht B, Fazel M, Parrilo P. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 2010; 52:471–501.

21. Rockafellar, RT. Princeton Landmarks in Mathematics. Princeton University Press; Princeton, NJ: 1997. Convex analysis. Reprint of the 1970 original, Princeton Paperbacks

22. Stojnic M. A framework to characterize performance of lasso algorithms. 2013 arXiv:1303.7291.

23. Tang G, Bhaskar BN, Shah P, Recht B. Compressive sensing off the grid. IEEE Trans Inform Theory. 2013; 59:7465–7490.

24. Tao, T. Topics in Random Matrix Theory. American Mathematical Society; Providence, Rhode Island: 2012.

25. Tropp JA, Laska JN, Duarte MF, Romberg JK, Baraniuk RG. Beyond Nyquist: efficient sampling of sparse bandlimited signals. IEEE Trans Inform Theory. 2010; 56:520–544.

26. Rao, Richard, Kailath, Thomas. ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans Acoust Speech Signal Process. 1989; 37(7):984–995.

27. Liao W, Fannjiang A. MUSIC for single-snapshot spectral estimation: stability and super-resolution. arXiv:1404.1484.

28. Moitra, A. Super-resolution, extremal functions and the condition number of Vandermonde matrices. Proceedings of the 47th Annual ACM Symposium on Theory of Computing; STOC; 2015.
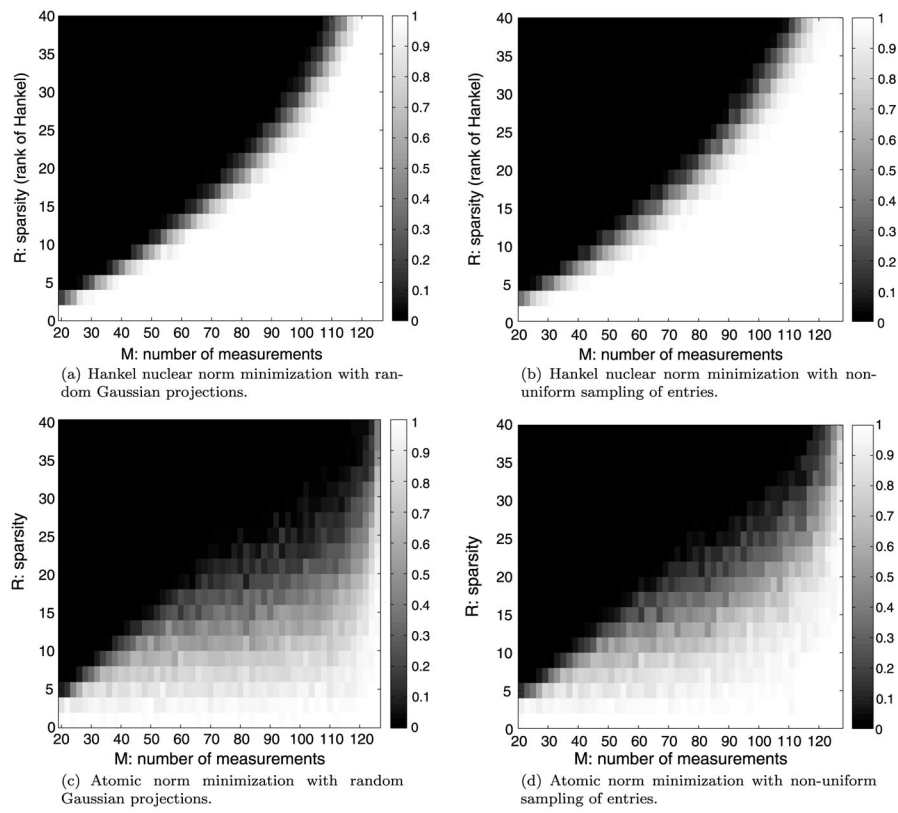
(a) Hankel nuclear norm minimization with random Gaussian projections.

(b) Hankel nuclear norm minimization with non-uniform sampling of entries.

(c) Atomic norm minimization with random Gaussian projections.

(d) Atomic norm minimization with non-uniform sampling of entries.

Fig. 1.
Numerical results.