Restricted Isometry Property of Rank-One Measurements with Random Unit-Modulus Vectors

Wei Zhang
Harbin Institute of Technology Shenzhen

Zhenni Wang
City University of Hong Kong

Abstract

The restricted isometry property (RIP) is essential for the linear map to guarantee the successful recovery of low-rank matrices. The existing works show that the linear map generated by the measurement matrices with independent and identically distributed (i.i.d.) entries satisfies RIP with high probability. However, when dealing with non-i.i.d. measurement matrices, such as the rank-one measurements, the RIP compliance may not be guaranteed. In this paper, we show that the RIP can still be achieved with high probability, when the rank-one measurement matrix is constructed by the random unit-modulus vectors. Compared to the existing works, we first address the challenge of establishing RIP for the linear map in non-i.i.d. scenarios. As validated in the experiments, this linear map is memory-efficient, and not only satisfies the RIP but also exhibits similar recovery performance of the low-rank matrices to that of conventional i.i.d. measurement matrices.

1 INTRODUCTION

The low-rank matrix recovery is a popular topic in many fields, such as wireless communication, signal processing, and image processing (Candes et al., 2013; Chen et al., 2015; Shechtman et al., 2015; Davenport and Romberg, 2016; Zhang et al., 2018b; Chen and Chi, 2018; Chi et al., 2019; Zhang et al., 2019; Zhang and Tay, 2021; Farias et al., 2022; Tong et al., 2022). The primary objective for this problem is to reconstruct a low-rank matrix from a limited number of observations. These observations are obtained through a linear map, which consists of the measurement matrices. To be more specific, the measurements for a low-rank matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$ are give by the following

$$y_k = \frac{1}{\sqrt{K}}\langle \mathbf{A}_k, \mathbf{X}\rangle + z_k, k = 1, 2, \ldots, K, \tag{1}$$

where $\mathbf{A}_k \in \mathbb{C}^{M\times N}$ are the measurement matrices, $\langle \mathbf{A}_k, \mathbf{X}\rangle = \operatorname{tr}(\mathbf{A}_k^H\mathbf{X})$, and z_k is the noise or measurement error. The K measurement matrices $\{\mathbf{A}_k\}_{k=1}^K$ collectively define a linear map, denoted as $\mathcal{A}(\cdot) : \mathbb{C}^{M\times N} \to \mathbb{C}^{K\times 1}$, where each entry of $\mathcal{A}(\mathbf{X})$ is given by $[\mathcal{A}(\mathbf{X})]_k = \frac{1}{\sqrt{K}}\langle \mathbf{A}_k, \mathbf{X}\rangle, k = 1, 2, \ldots, K$. Thus, the measurement model in (1) is shown in a compact form

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) + \mathbf{z}, \tag{2}$$

where $\mathbf{y} = [y_1, \ldots, y_K]^T$ and $\mathbf{z} = [z_1, \ldots, z_K]^T$.

The goal of low-rank matrix recovery is to reconstruct $\mathbf{X}$ from the linear map $\mathcal{A}(\cdot)$ and measurements $\mathbf{y}$ in (2). There are various approaches, including both convex and non-convex methods, (Recht et al., 2010; Candes and Plan, 2011; Chen and Chi, 2018; Chi et al., 2019; Ma et al., 2018; Jain et al., 2013, 2010; Zheng and Lafferty, 2015; Tu et al., 2016) can be utilized to fulfill the goal. The convex methods (Recht et al., 2010; Candes and Plan, 2011; Chen and Chi, 2018) utilize the nuclear norm of $\mathbf{X}$ as a penalty term in their objective functions to promote low-rank solutions. It has been shown that, when the linear map satisfies the restricted isometry property (RIP) with the required RIP constant, these convex methods can guarantee successful recovery in noiseless scenarios or bounded reconstruction errors in the presence of noise. In addition to convex methods, non-convex techniques (Chi et al., 2019; Ma et al., 2018; Jain et al., 2013, 2010; Zheng and Lafferty, 2015; Tu et al., 2016), such as gradient-based methods and alternating minimization methods, offer greater computational efficiency . Notably, these gradient-based methods, as demonstrated in the works by Chi et al. (2019); Zheng and Lafferty (2015); Tu et al. (2016), can ensure convergence with proper initialization when the linear map $\mathcal{A}(\cdot)$ satisfies the RIP with the necessary constant. Furthermore,

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

some studies (Ge et al., 2017; Zhang et al., 2018a) have investigated the presence of spurious local minima in low-rank matrix recovery problems. When there are no spurious local minima, non-convex methods can achieve global minima. The works by Ge et al. (2017); Zhang et al. (2018a) have shown that, when the linear map $\mathcal{A}(\cdot)$ satisfies the RIP with the required constant, the low-rank matrix recovery problem formulated in this manner has no spurious local minima, thereby guaranteeing exact recovery.

As mentioned above, one can find that the RIP of the linear map $\mathcal{A}(\cdot)$ plays an important role in ensuring the low-rank matrix recovery. In particular, the definition of RIP is presented below.

Definition 1 (Standard RIP over Low-Rank Matrices (Candes and Plan, 2011))**.** *For the set of rank-r matrices, we define the RIP constant δ_r with respect to operator $\mathcal{A}(\cdot)$ as the smallest numbers such that for all $\mathbf{X}$ of rank at most r:*

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

There are many types of linear map that satisfy the defined RIP above. When the entries of $\mathbf{A}_k$ are i.i.d. complex Gaussian entries (Recht et al., 2010; Candes and Plan, 2011), or when the entries of $\mathbf{A}_k$ are i.i.d. unit-modulus (Zhang et al., 2018b), where each entry is unit-modulus and its phase follows a uniform distribution in the range $[0, 2\pi]$, then the linear map $\mathcal{A}(\cdot)$ satisfies the RIP with high probability, on the condition that the number of measurements $K \geq c(M+N)r$ for large enough constant $c > 0$.

However, in certain scenarios of low-rank matrix recovery, the entries in the measurement matrix are non-i.i.d. or do not follow the distributions mentioned above, such as the low-rank matrix completion (Candès and Recht, 2009; Candès and Tao, 2010), phase retrieval (Candes et al., 2015; Ma et al., 2018), and quadratic sensing problem (Chen et al., 2015; Cai and Zhang, 2015). In general, the associated linear map $\mathcal{A}(\cdot)$ in these scenarios does not satisfy the RIP property. To analyze the recovery performance guarantee under these non-RIP scenarios, some variants of RIP are defined, such as incoherence for matrix completion (Candès and Recht, 2009; Candès and Tao, 2010) and RIP-ℓ_2/ℓ_1 for the rank-one measurements (Chen et al., 2015; Cai and Zhang, 2015). However, many existing advancements based on RIP, such as the works in (Chi et al., 2019; Ma et al., 2018; Jain et al., 2013, 2010; Zheng and Lafferty, 2015; Tu et al., 2016) , are not applicable for these non-RIP scenarios.

In this paper, our primary focus is on the concept of rank-one measurements as introduced by Cai and Zhang (2015). We aim to demonstrate that when the measurement matrix follows the specified distribution, the associated linear map satisfies the RIP. It is worth noting that the quadratic sensing problem and phase retrieval can be considered special cases of the rank-one projection problem. For rank-one measurements (Cai and Zhang, 2015; Li et al., 2019), the measurement matrix $\mathbf{A}_k$ can be represented as an outer product of two vectors,

$$\mathbf{A}_k = \mathbf{u}_k \mathbf{v}_k^H. \tag{3}$$

However, in general cases where the measurement matrices $\mathbf{A}_k$ are defined above, it has been established in prior works (Cai and Zhang, 2015; Chi et al., 2019) that the associated linear map does not satisfy the RIP. For example, studies of Chi et al. (2019); Cai and Zhang (2015); Chen et al. (2015) have shown that when the entries of both $\mathbf{u}_k$ and $\mathbf{v}_k$ are i.i.d. Gaussian, the associated linear map $\mathcal{A}(\cdot)$ does not satisfy the RIP. In our work, we impose specific distribution on the random vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ instead of i.i.d. Gaussian in the existing works (Chi et al., 2019; Cai and Zhang, 2015; Chen et al., 2015). We show that when the entries in $\mathbf{u}_k$ and $\mathbf{v}_k$ are i.i.d. unit-modulus, the associated linear map satisfies the RIP with high probability. As far as we know, our research marks the first attempt to tackle the challenge of establishing RIP for the linear map in non-i.i.d. scenarios. Additionally, it lays the foundational framework for proving RIP in various forms of random rank-one measurements.

2 RIP ANALYSIS OF LINEAR MAP

2.1 Sufficient Condition of RIP

For arbitrary $\mathbf{u}_k$ and $\mathbf{v}_k$ in (3), whether the corresponding linear map $\mathcal{A}(\cdot)$ satisfies the RIP is challenging to check. Fortunately, the following theorem provides a sufficient condition on which the linear map $\mathcal{A}(\cdot)$ satisfies the RIP.

Theorem 1 (Candes and Plan (2011), Theorem 2.3[1])**.** *Let $\mathcal{A}(\cdot) : \mathbb{C}^{M\times N} \rightarrow \mathbb{C}^{K\times 1}$ be a linear map with random measurement matrices obeying the following condition: for any given $\mathbf{X} \in \mathbb{C}^{M\times N}$ and any fixed $0 < \alpha < 1$*

$$\Pr\left(\left|\|\mathcal{A}(\mathbf{X})\|_2^2 - \|\mathbf{X}\|_F^2\right| \geq \alpha \|\mathbf{X}\|_F^2\right) \leq C\exp(-cK) \tag{4}$$

for fixed constants $C, c > 0$ (which may depend on α). Then if $K \geq D\max\{M, N\}r$, $\mathcal{A}$ satisfies the RIP with constant $\delta_r > 0$ with probability exceeding $1 - Ce^{-dK}$ for fixed constants $D, d > 0$.

[1]It is worth noting that the original result in the work (Candes and Plan, 2011) is for the real case, i.e., $\mathcal{A} : \mathbb{R}^{M\times N} \rightarrow \mathbb{R}^{K\times 1}$ and $\mathbf{X} \in \mathbb{R}^{M\times N}$. However, the result is ready to extend to the complex case.

Without loss of generality, we assume $\|\mathbf{X}\|_F = 1$. Therefore, according to Theorem 1, in order to show the linear map $\mathcal{A}(\cdot)$ meet RIP, we need to prove the following probability

$$\Pr\left(\left|\|\mathcal{A}(\mathbf{X})\|_2^2 - 1\right| \geq \alpha\right) \tag{5}$$

is close to zero. Note that the probability is taken over the linear map $\mathcal{A}(\cdot)$ and the $\mathbf{X}$ is fixed and arbitrary.

For the linear map $\mathcal{A}(\cdot)$ where the entries in $\mathbf{A}_k$ are drawn from i.i.d. $\mathcal{CN}(0,1)$, it is easy to verify that the condition in (4) holds. Therefore, the corresponding linear map satisfies the RIP with high probability. This is also consistent with the real case where $\mathbf{A}_k$ are drawn according to i.i.d. $\mathcal{N}(0,1)$ in the following.

Remark 1 (Recht et al. (2010); Candes and Plan (2011))**.** *If the entries of $\mathbf{A}_k$ are i.i.d. Gaussian entries $\mathcal{N}(0,1)$, then $\mathcal{A}(\cdot)$ satisfies the r-RIP with RIP constant δ_r with high probability as $K \gtrsim (M+N)r/\delta_r^2$.*

Moreover, when the entries in $\mathbf{A}_k$ of the linear map are i.i.d. (Recht et al., 2010; Candes and Plan, 2011; Zhang et al., 2018b) , the central limit theorem can be applied to approximately verify the sufficient condition in (4). However, for the rank-one model in (3), the entries of measurement matrix $\mathbf{A}_k$ are dependent. Due to this dependence, the standard RIP in Definition 1 may not hold for the general $\mathbf{u}_k$ and $\mathbf{v}_k$. For example, when the entries of $\mathbf{u}_k$ and $\mathbf{v}_k$ are i.i.d. Gaussian, the RIP does not hold in this scenario because the $\mathbf{u}^H\mathbf{X}\mathbf{v}$ involves fourth moments of Gaussian variable (Cai and Zhang, 2015; Kueng et al., 2017; Candes et al., 2013).

To evaluate the concentration property of the linear map for this dependent and rank-one measurement model in (3), some alternative conditions, such as RIP-ℓ_1/ℓ_1 (Candes et al., 2013) and RIP-ℓ_2/ℓ_1 (Chen et al., 2015; Cai and Zhang, 2015), have been proposed. These studies demonstrate that the convex methods can ensure the exact recovery based on these alternative conditions. However, in the context of non-convex analyses, techniques like the alternating minimization method (Jain et al., 2013), singular value projection method (Jain et al., 2010), and other local optimal analysis (Ge et al., 2017; Chi et al., 2019; Ma et al., 2018), these variants of RIP (Candes et al., 2013; Chen et al., 2015; Cai and Zhang, 2015) are not applicable, because these analysis are based on the standard RIP. Thus, this highlights the crucial importance of standard RIP in Definition 1 compared to its variants.

It is indeed a well-established fact that the linear map $\mathcal{A}(\cdot)$ with general setting for $\mathbf{u}_k$ and $\mathbf{v}_k$ in (3) may not satisfy the standard RIP. However, in this paper, we find that if we impose some specific design for $\mathbf{u}_k$ and $\mathbf{v}_k$, it becomes possible to attain the standard RIP for the designed linear map. The main result of the paper is presented in the following theorem.

Theorem 2. *Suppose the measurement matrix $\mathbf{A}_k = \mathbf{u}_k\mathbf{v}_k^H \in \mathbb{C}^{M\times N}$, where $\mathbf{u}_k \in \mathbb{C}^{M\times 1}, \mathbf{v}_k \in \mathbb{C}^{N\times 1}$ are given in the following*

$$\begin{aligned}\mathbf{u}_k &= [e^{j\theta_{k,1}}, \ldots, e^{j\theta_{k,M}}]^T,\\ \mathbf{v}_k &= [e^{j\phi_{k,1}}, \ldots, e^{j\phi_{k,N}}]^T,\end{aligned} \tag{6}$$

with $\theta_{k,m}, \forall\, m$ and $\phi_{k,n}, \forall\, n$ being i.i.d. from a uniform distribution on $[0, 2\pi]$. For the linear map $\mathcal{A}(\cdot) : \mathbb{C}^{M\times N} \to \mathbb{C}^{K\times 1}$ generated by $\{\mathbf{A}_k\}_{k=1}^K$, where

$$\begin{aligned}[\mathcal{A}(\mathbf{X})]_k &= \frac{1}{\sqrt{K}}\langle \mathbf{A}_k, \mathbf{X}\rangle,\\ &= \frac{1}{\sqrt{K}}\mathbf{u}_k^H\mathbf{X}\mathbf{v}_k, \forall\, k = 1, 2, \ldots, K,\end{aligned}$$

it satisfies RIP with high probability as long as $K \geq D\max\{M, N\}r$ for some large enough constant D.

Intuitively, the reason that the standard RIP for the linear map (6) holds is because the entries in $\mathbf{u}_k$ and $\mathbf{v}_k$ are unit-modulus, which are bounded compared to the scenario where $\mathbf{u}_k$ and $\mathbf{v}_k$ are i.i.d. Gaussian (Chen et al., 2015; Cai and Zhang, 2015). Moreover, they experience some special symmetric statistical property compared to i.i.d. Gaussian scenario, which enables us to prove the RIP in the following sections. Before delving into details of the proof, we first discuss the applications of the measurement model outlined in (6).

Compared to the i.i.d. measurement matrix $\mathbf{A}_k$, the rank-one measurements can offer enhanced storage efficiency for the linear map, as demonstrated by (Cai and Zhang, 2015). Moreover, within the context of rank-one measurements, the designed unit-modulus setting in (6) can further save the storage of the measurement matrices, as opposed to case of i.i.d. Gaussian $\mathbf{u}_k$ and $\mathbf{v}_k$. The reason behind this efficiency is that, for the unit-modulus setting described in (6), it is necessary to store only the phases of the vectors $\mathbf{u}_k$ and $\mathbf{v}_k$. In contrast, for the i.i.d. Gaussian $\mathbf{u}_k$ and $\mathbf{v}_k$, one must preserve both the magnitudes and phases of these vectors to accurately construct the measurement matrix $\mathbf{A}_k$. Most importantly, based on the established results in Theorem 2, the proposed unit-modulus rank-one measurements are applicable for many RIP-based algorithms or analysis (Jain et al., 2013, 2010; Ge et al., 2017; Chi et al., 2019; Ma et al., 2018), making the rank-one unit-modulus measurements a promising option for the matrix recovery task. Therefore, building on the advantages highlighted earlier, the rank-one measurement model with unit-modulus vectors in (6) has widespread applications in the field of low-rank matrix recovery, especially when the configuration of the measurement matrices is applicable, such as channel estimation in communication systems (El Ayach et al., 2014; Zhang et al., 2018b, 2020), phase retrieval (Candes et al., 2013; Ma et al., 2018), covariance estimation (Chen et al., 2015), and X-ray crystallography (Shechtman et al., 2015).

2.2 Inequalities of Tail Bounds

To establish the fact that the linear map in Theorem 2 satisfies the RIP, we need to evaluate the probability of the event in (5). After straightforward manipulations, one can find that $\mathbb{E}[\|\mathbf{y}\|_2^2] = 1$. Thus, the bound in (5) is about the tail bound. When the probability is small, it means that the value of $\|\mathbf{y}\|_2^2$ is strictly concentrated around its expected value. In the following, we will evaluate the upper and lower tail bounds, respectively.

For the upper tail bound, due to the independence among the entries of $\mathcal{A}(\mathbf{X})$, one can apply Chernoff bound for any $h > 0$,

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \\
&= \Pr(hK\|\mathcal{A}(\mathbf{X})\|_2^2 \geq hK(1+\alpha)) \\
&= \Pr(e^{hK\|\mathcal{A}(\mathbf{X})\|_2^2} \geq e^{hK(1+\alpha)}) \\
&\leq \mathbb{E}[e^{hK\|\mathcal{A}(\mathbf{X})\|_2^2}]e^{-hK(1+\alpha)}.
\end{aligned}$$

Since the second part of the right-hand side (R.H.S.) of the inequality above, i.e., $e^{-hK(1+\alpha)}$, goes to zero as K goes to infinity, we focus on the first part,

$$\begin{aligned}
\mathbb{E}[e^{hK\|\mathcal{A}(\mathbf{X})\|_2^2)}] &= \mathbb{E}[e^{h\sum_{k=1}|\langle \mathbf{A}_k,\mathbf{X}\rangle|^2}] \\
&= \left(\mathbb{E}[e^{h|\langle \mathbf{A}_1,\mathbf{X}\rangle|^2}]\right)^K \\
&= \left(\mathbb{E}[e^{h|\mathbf{u}_1^H\mathbf{X}\mathbf{v}_1|^2}]\right)^K.
\end{aligned}$$

For convenience, we ignore the subscript the subscripts of $\mathbf{u}$ and $\mathbf{v}$. Using Taylor series for $e^{hK|\langle \mathbf{A}_1,\mathbf{X}\rangle|^2}$ gives

$$\begin{aligned}
\mathbb{E}[e^{h|\mathbf{u}^H\mathbf{X}\mathbf{v}|^2}] &= \mathbb{E}\left[\sum_{t=0}^{\infty}\frac{h^t}{t!}|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] \\
&= \sum_{t=0}^{\infty}\frac{h^t}{t!}\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]. \qquad (7)
\end{aligned}$$

For the lower tail bound in (5), we have

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \leq (1-\alpha)) \\
&= \Pr(-hK\|\mathcal{A}(\mathbf{X})\|_2^2 \geq -hK(1-\alpha)) \\
&\leq \mathbb{E}[e^{-hK\|\mathcal{A}(\mathbf{X})\|_2^2}]e^{hK(1-\alpha)} \\
&= \left(\mathbb{E}[e^{-h|\mathbf{u}^H\mathbf{X}\mathbf{v}|^2}]\right)^K e^{hK(1-\alpha)}.
\end{aligned}$$

Similarly, using the Taylor series yields

$$\begin{aligned}
\mathbb{E}[e^{-h|\mathbf{u}^H\mathbf{X}\mathbf{v}|^2}] &= \mathbb{E}\left[\sum_{t=0}^{\infty}\frac{(-h)^t}{t!}|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] \\
&= \sum_{t=0}^{\infty}\frac{(-h)^t}{t!}\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]. \qquad (8)
\end{aligned}$$

In summary, we have the following upper tail probability bound

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \leq \\
&\left(\sum_{t=0}^{\infty}\frac{h^t}{t!}\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]\right)^K e^{-hK(1+\alpha)}, \qquad (9)
\end{aligned}$$

and the lower tail probability bound

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \leq (1-\alpha)) \leq \\
&\left(\sum_{t=0}^{\infty}\frac{(-h)^t}{t!}\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]\right)^K e^{hK(1-\alpha)}. \qquad (10)
\end{aligned}$$

To check whether the sufficient condition in (4) holds for the linear map $\mathcal{A}(\cdot)$, we need to evaluate the values in (9) and (10). By observing the R.H.S. of (9) and (10), one can find that the key is to calculate $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$.

3 CONNECTION WITH THE ALL-ONE MATRIX

The value of $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ depends on the realizations of $\mathbf{X}$, which is challenging to manipulate. To handle this, we first focus on a specific $\mathbf{X}$, where $\mathbf{X} = \frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T$, then establish a relationship between $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ and $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$. In particular, when $\mathbf{X} = \frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathbf{u}^H\frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T\mathbf{v}\right|^{2t}\right] = \\
&\frac{1}{M^tN^t}\mathbb{E}\left[\left|\sum_{m=1}\mathbf{u}_m^*\right|^{2t}\right]\mathbb{E}\left[\left|\sum_{n=1}\mathbf{v}_n\right|^{2t}\right]. \qquad (11)
\end{aligned}$$

Compared to $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$, the value in (11) only depends on the random vector $\mathbf{u}$ and $\mathbf{v}$, which is applicable to derive a bound for it. We first provide some preliminaries, and all their proofs are attached in Section B of the supplementary materials.

First of all, the following lemma is about the maximization of summation of combinations.

Lemma 1. *Suppose $0 \le p \le 1$ and $n \ge 0$, the summation below*

$$s_n(p) = \sum_{k=0}^{n} \binom{n}{k}^2 p^k (1-p)^{n-k}$$

is maximized when $p = 1/2$.

Since the summation term $p+(1-p)=1$, the Lemma 1 shows that the average value, i.e., $p=1/2$, achieves the maximum of $s_n(p)$. The results in Lemma 1 will be utilized to compare the value of $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ with $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$. As a straightforward extension, when $p+(T-p)=T$, the following Corollary holds.

Corollary 1. *Suppose $0 \le p \le T$ and $n \ge 0$, the summation below*

$$s_n(p) = \sum_{k=0}^{n} \binom{n}{k}^2 p^k (T-p)^{n-k}$$

is maximized when $p = 0.5T$.

Before comparing the values of $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ and $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$, we start from the simplified case where $\mathbf{x} \in \mathbb{C}^{N\times 1}$ and disregard the vector $\mathbf{u}$. Specifically, we evaluate the values of $\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right]$ and $\mathbb{E}\left[|\mathbf{1}^T\mathbf{v}|^{2m}\right], \forall m$. By doing so, we lay the foundation for extending these findings to a more general settingby incorporating additional considerations

Theorem 3. *Suppose $\mathbf{v} = [e^{j\phi_1}, \ldots, e^{j\phi_N}]^T$ with $\phi_n, \forall n$ being i.i.d. from a uniform distribution on $[0, 2\pi]$. If $\mathbf{x} \in \mathbb{C}^{N\times 1}$ with $\|\mathbf{x}\|_2 = 1$, then for any $m \ge 0$, $\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right]$ is maximized when $x_n = \frac{1}{\sqrt{N}}, \forall n$.*

Thus, according to Theorem 3, for any $\|\mathbf{x}\|_2 = 1$, one has the following inequality,

$$\mathbb{E}[|\mathbf{x}^T\mathbf{v}|^{2m}] \le \frac{1}{N^m}\mathbb{E}[|\mathbf{1}^T\mathbf{v}|^{2m}], \forall m.$$

Furthermore, in order to extend the vector-case results in Theorem 3 to a more general setting, we need the following Lemmas 2 to 4 as preliminaries.

Lemma 2. *Suppose $t > 0$ and $k_m \ge 0$ are integers, and non-negative $c_m \in \mathbb{R}, \forall m = 1, 2, \ldots, M$, with $\sum_{m=1}^{M} c_m^2 = 1$, then the following holds*

$$\sum_{k_1+\ldots+k_M=t} \binom{t}{k_1, \ldots, k_M}^2 \prod_{m=1}^{M} c_m^{2k_m} \le \sum_{k_1+\ldots+k_M=t} M^{-t} \binom{t}{k_1, \ldots, k_M}^2, \tag{12}$$

where the equality holds when $c_m = 1/\sqrt{M}$.

The results in Lemma 2 mean that the summation about the combinatorial expression is maximized when the values of $c_m, \forall m$, are equivalent, which is the extension of result in the one-variable case of Corollary 1.

Lemma 3. *Suppose non-negative random variables X_1, X_2, for any $k_1, k_2 \ge 0$ and $1/p + 1/q = 1$ with $p, q \in [1, +\infty)$, then we have*

$$\mathbb{E}\left[X_1^{k_1} X_2^{k_2}\right] \le (\mathbb{E}[X_1^{k_1 p}])^{1/p} (\mathbb{E}[X_2^{k_2 q}])^{1/q}. \tag{13}$$

In particular,

$$\mathbb{E}[X_1^{k_1} X_2^{k_2}] \le \max\{\mathbb{E}[X_1^{k_1+k_2}], \mathbb{E}[X_2^{k_1+k_2}]\}. \tag{14}$$

The results in Lemma 3 are to bound $\mathbb{E}[X_1^{k_1} X_2^{k_2}]$ by the product of expectations, where the latter is more applicable to handle. Then, the following lemma is an extension of the results in Lemma 3, where there are N random variables.

Lemma 4. *Suppose non-negative random variables $X_1, X_2, \ldots, X_N$, for any $k_n \ge 0$, then the following inequality about expectation holds*

$$\begin{aligned}\mathbb{E}\left[\prod_{n=1}^{N} X_n^{k_n}\right] &\le \prod_{n=1}^{N} \left(\mathbb{E}\left[X_n^t\right]\right)^{k_n/t} \\ &\le \max_n \mathbb{E}\left[X_n^t\right],\end{aligned} \tag{15}$$

where $t = \sum_{n=1}^{N} k_n$.

With the results in Lemmas 2 to 4, we are now ready to compare the values of $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ and $\frac{1}{M^t N^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$ in the following theorem, which is proved in Section A of the supplementary materials.

Theorem 4. *For any matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$ with $\|\mathbf{X}\|_F = 1$, the following*

$$\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}] \le \frac{1}{M^t N^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$$

holds for any non-negative integer t. Here, $\mathbf{u} \in \mathbb{C}^{M\times 1}$ and $\mathbf{v} \in \mathbb{C}^{N\times 1}$ are random vectors given by

$$\mathbf{u} = [e^{j\theta_1}, \ldots, e^{j\theta_M}]^T, \ \mathbf{v} = [e^{j\phi_1}, \ldots, e^{j\phi_N}]^T$$

with $\theta_m, \forall m$ and $\phi_n, \forall n$ being i.i.d in $[0, 2\pi]$.

Thus, the results in Theorem 4 establish a relationship between $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ and $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$, where the latter the only depends on the random vector $\mathbf{u}$ and $\mathbf{v}$. Therefore, to further proceed with Theorem 4 and obtain a valid bound for $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ in (9) and (10), it necessary to assess the value of $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$. Due to the independence between $\mathbf{u}$ and $\mathbf{v}$, one can express this as $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}] = \mathbb{E}[|\mathbf{u}^H\mathbf{1}|^{2t}]\mathbb{E}[\mathbf{1}^T\mathbf{v}|^{2t}]$. In this context, the following proposition provides a valuable bound for both $\mathbb{E}[|\mathbf{u}^H\mathbf{1}|^{2t}]$ and $\mathbb{E}[\mathbf{1}^T\mathbf{v}|^{2t}]$.

Proposition 1. *Since the entries in $\mathbf{u}$ and $\mathbf{v}$ are i.i.d., one can check that*

$$\begin{aligned}\mathbb{E}\left[\left|\mathbf{u}^H\mathbf{1}\right|^{2t}\right] &= \mathbb{E}\left[(u_1+, \ldots, +u_M)^t (u_1^*+, \ldots, +u_M^*)^t\right] \\ &= \sum_{t_1+\ldots+t_M=t} \binom{t}{t_1, \ldots, t_M}^2.\end{aligned}$$

One can find that the value above is the number of abelian squares of length $2t$ over an alphabet with M letters (Richmond and Shallit, 2008), denoted as $g(t, M)$. Similarly, the value of $\mathbb{E}\left[|\mathbf{1}^T\mathbf{v}|^{2t}\right]$ is given by

$$\mathbb{E}\left[|\mathbf{1}^T\mathbf{v}|^{2t}\right] = \sum_{t_1+\ldots+t_N=t} \binom{t}{t_1,\ldots,t_N}^2 = g(t, N).$$

Now, the bounds of $g(t, N)$ and $g(t, M)$ are of interest. Based on the results by Richmond and Shallit (2008), the values of $g(t, M)$ and $g(t, N)$ are bounded by

$$g(t, M) \leq C_1 M^{2t} t^{(1-M)/2},$$
$$g(t, N) \leq C_2 N^{2t} t^{(1-N)/2},$$

where C_1 and C_2 are two constants. Therefore, the value of $\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}/\sqrt{MN}|^{2t}]$ is upper bounded by

$$\mathbb{E}\left[\left|\mathbf{u}^H\frac{\mathbf{1}\mathbf{1}^T}{\sqrt{MN}}\mathbf{v}\right|^{2t}\right] \leq CM^tN^t t^{1-\frac{M+N}{2}}, \tag{16}$$

where $C = C_1C_2$ is a constant.

4 RIP OF RANK-ONE UNIT-MODULUS MEASUREMENTS

In this section, we prove the RIP of rank-one measurements with unit-modulus vectors in Theorem 2. Recall that it is essential to evaluate the tail bounds in (9) and (10) to show they satisfy the sufficient condition provided in Theorem 1. Based on the established results in Section 3, we can simplify the upper and lower tail bounds in (9) and (10), respectively, as shown in the following theorem.

Theorem 5. *For the linear map $\mathcal{A}(\cdot)$ defined in Theorem 2, we have the following upper tail probability bound for any $\mathbf{X} \in \mathbb{C}^{M\times N}$ with $\|\mathbf{X}\|_F = 1$,*

$$\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \leq e^{-c_1K}, \tag{17}$$

where $c_1 > 0$ is a constant depending on α. In addition, the lower tail probability bound is given by

$$\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \leq (1-\alpha)) \leq e^{-c_2K}, \tag{18}$$

where $c_2 > 0$ is also a constant depending on α.

According to the results in Theorem 5, it is evident that both the upper and lower tail bounds exhibit an exponential decrease as the number of measurements K. This observation leads us to verify the sufficient condition for the RIP outlined in Theorem 1. Consequently, the linear map associated with random unit-modulus vectors satisfies the RIP with high probability, as shown in Theorem 2. In the following, we provide a comprehensive proof of Theorem 5.

Proof of Theorem 5. To prove the upper tail bound in (17), we combine the results in (9) and Theorem 4,

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \\
&\leq \left(\sum_{t=0}^{\infty}\frac{h^t}{t!}\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]\right)^K e^{-hK(1+\alpha)} \\
&\leq \left(\sum_{t=0}^{\infty}\frac{h^t}{t!}\mathbb{E}\left[\left|\mathbf{u}^H\frac{\mathbf{1}\mathbf{1}^T}{\sqrt{MN}}\mathbf{v}\right|^{2t}\right]\right)^K e^{-hK(1+\alpha)}.
\end{aligned}$$

Then from Proposition 1, we have

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \\
&\leq\Bigg(1+h+2h^2-\frac{2M+2N-1}{2MN}h^2+ \\
&\qquad \sum_{t=3}^{\infty}\frac{h^t}{t!}CM^tN^tt^{1-\frac{M+N}{2}}\Bigg)^K e^{-hK(1+\alpha)}.
\end{aligned} \tag{19}$$

In the following, we need to choose a h which makes the bound above tight. Note that there exists $Z > 0$, which makes the following hold for any $0 < h \leq Z\alpha$,

$$\begin{aligned}
&1+h+2h^2-\frac{2M+2N-1}{2MN}h^2+ \\
&\sum_{t=3}^{\infty}\frac{h^t}{t!}CM^tN^tt^{1-\frac{M+N}{2}} \leq 1+h+2h^2.
\end{aligned} \tag{20}$$

For convenience, we define

$$f(h) = (1+h+2h^2)e^{-h(1+\alpha)}.$$

Thus, according to (19) and (20), for any $0 < h \leq Z\alpha$, the tail probability in (19) is bounded as follows

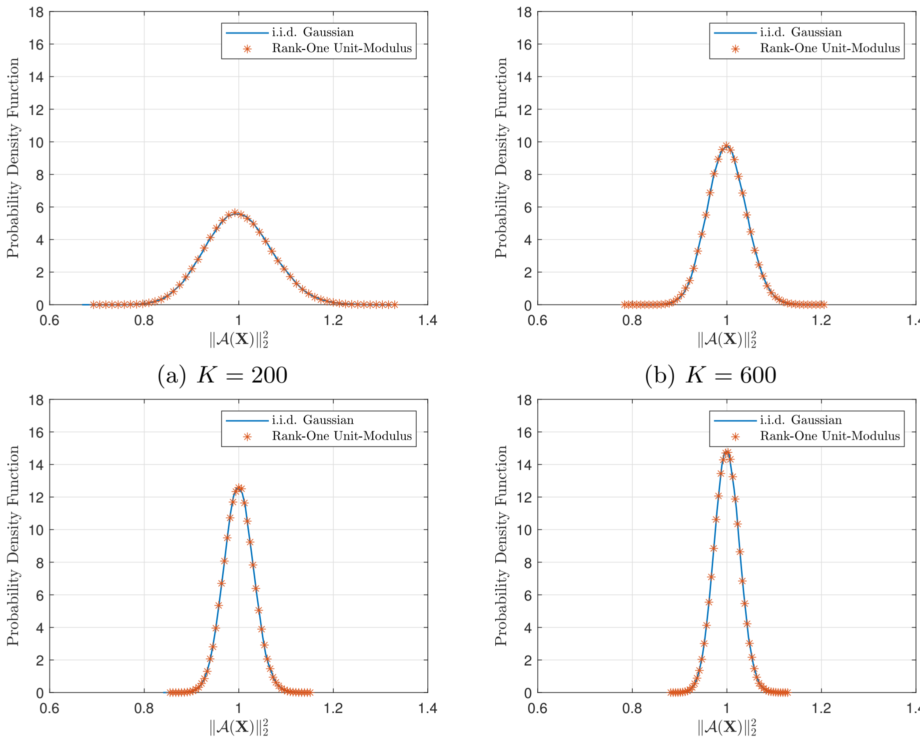
$$\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) \leq (f(h))^K. \tag{21}$$

Comparing (21) with the sufficient condition in (4), we need to prove that the above expression has the form of $C\exp(-cK)$. In other words, we need to show there exists a $h_1 \in (0, Z\alpha]$ such that $f(h_1) < 1$, then $(f(h_1))^K$ converges to zero as K goes to infinity.

Note that $f(0) = 1$. Thus, in order to show there exists h_1 such that $f(h_1) < 1$, it is sufficient to show the first derivative of f at 0 is negative, i.e., $f'(0) < 0$, which is obviously true. Therefore, we choose $h = h_1$, the expression in (21) is rewritten as

$$\begin{aligned}
\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \geq (1+\alpha)) &\leq (1+h_1+2h_1^2)^K e^{-h_1K(1+\alpha)} \\
&= e^{-c_1K},
\end{aligned}$$

where the constant $c_1 = -\ln(f(h_1)) = -\ln((1+h_1+2h_1^2)e^{-h_1(1+\alpha)}) > 0$. Thus, the bound for the upper tail probability (17) is proved.



(a) $K = 200$ (b) $K = 600$

(c) $K = 1000$ (d) $K = 1400$

Fig. 1: Probability Density Function of $\|\mathcal{A}(\mathbf{X})\|_2^2$ with Different Number of Measurements

For the lower tail probability in (18), we have

$$\begin{aligned}
&\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \le (1-\alpha)) \\
&\le \sum_{t=0}^{\infty} \frac{(-h)^t}{t!} \mathbb{E}[|\mathbf{u}^H \mathbf{X} \mathbf{v}|^{2t}] e^{hK(1-\alpha)} \\
&\le \left(1 - h + 2h^2 \mathbb{E}[|\mathbf{u}^H \mathbf{X} \mathbf{v}|^4]\right)^K e^{hK(1-\alpha)} \\
&\le \left(1 - h + 2h^2 \mathbb{E}\left[\left|\mathbf{u}^H \frac{\mathbf{1}\mathbf{1}^T}{\sqrt{MN}} \mathbf{v}\right|^4\right]\right)^K e^{hK(1-\alpha)} \\
&\le (1 - h + 2h^2)^K e^{hK(1-\alpha)}.
\end{aligned} \tag{22}$$

Similarly, we need to prove that the above expression has the form of $C\exp(-cK)$ in Theorem 1, where C and c are constants. The straightforward method is to minimize the value above with respect to h. Here, for simplicity, we just let $h = \frac{1}{4}$. Then, the lower tail probability in (22) is bounded by

$$\Pr(\|\mathcal{A}(\mathbf{X})\|_2^2 \le (1-\alpha)) \le 0.875^K e^{\frac{1}{4}K(1-\alpha)} \le e^{-c_2 K},$$

where $c_2 = 1.83 + (1/4)\alpha > 0$. Therefore, the bound for the lower tail probability (18) is proved. $\square$

Based on the results of Theorem 5, we finally provide the proof of the Theorem 2, and show that the linear map associated with the random unit-modulus vectors satisfies the RIP with high probability.

Proof of Theorem 2. By utilizing the union bound for the probability in (5), we have

$$\begin{aligned}
&\Pr\left(\left|\|\mathcal{A}(\mathbf{X})\|_2^2 - 1\right| \ge \alpha\right) \\
&= \Pr\left(\|\mathcal{A}(\mathbf{X})\|_2^2 \ge 1 + \alpha\right) + \Pr\left(\|\mathcal{A}(\mathbf{X})\|_2^2 \le 1 - \alpha\right).
\end{aligned}$$

According to the results inTheorem 5, combining the upper and tail bounds together gives

$$\begin{aligned}
\Pr(\left|\|\mathcal{A}(\mathbf{X})\|_2^2 - 1\right| \le \alpha) &\le e^{-c_1 K} + e^{-c_2 K} \\
&\le 2e^{-cK},
\end{aligned}$$

where $c = \min(c_1, c_2)$. Furthermore, according to Theorem 1, if the number of measurements $K \ge D\max\{M, N\}r$, the linear map $\mathcal{A}(\cdot)$ satisfies the RIP with isometry constant $\delta_r > 0$ with probability exceeding $1 - 2e^{-dK}$ for fixed constants $D, d > 0$. $\square$

5 NUMERICAL EXPERIMENTS

In this section, we verify that the linear map generated by the random unit-modulus satisfies the RIP. Subsequently, we assess the recovery performance of low-rank matrices by employing this linear map.

Directly validating the RIP of the linear map is known to be NP-hard. Hence, we opt to demonstrate that the sufficient condition for RIP as outlined in Theorem 1 holds true. This condition suggests that the

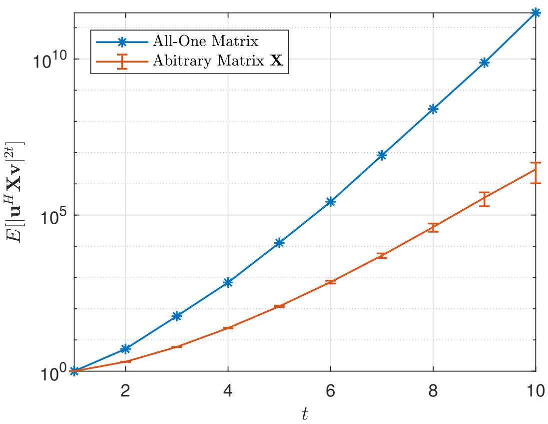


Fig. 2: Comparison of Two Expectations: All-One Matrix Scenario $\frac{1}{M^tN^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$ and Arbitrary Matrix Scenario $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$

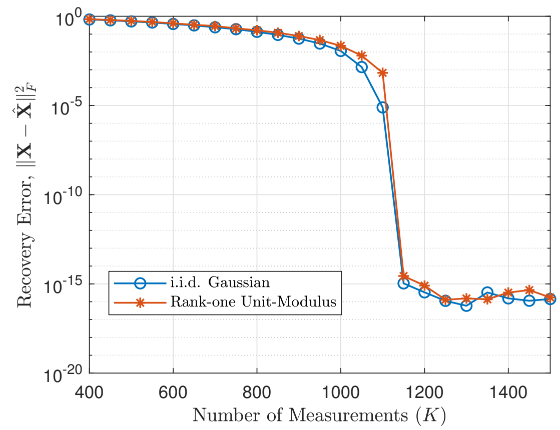


Fig. 3: Recovery Error of Low-Rank Matrix by Using Rank-One Unit-Modulus and i.i.d. Gaussian Measurements with Different Number of Measurements

value of $\|\mathcal{A}(\mathbf{X})\|_F^2$ associated with unit-modulus vectors is highly concentrated around its expected value, which is $\|\mathbf{X}\|_F^2$. To illustrate this, we conduct an experiment as presented in Fig. 1. In this experiment, we randomly generate a fixed $\mathbf{X} \in \mathbb{C}^{40\times 80}$ with $\|\mathbf{X}\|_F = 1$ and examine the probability density functions of $\|\mathcal{A}(\mathbf{X})\|_F^2$ by using two types of linear map. The first is generated by the rank-one model using random unit-modulus vectors, while the second serves as the benchmark and is based on i.i.d. Gaussian $\mathbf{A}_k$.

As depicted in Fig. 1, it is evident that the value of $\|\mathcal{A}(\mathbf{X})\|_F^2$ generated by rank-one measurements with random unit-modulus vectors is concentrated around its expectation $\|\mathbf{X}\|_F^2$ for different number of measurements. Comparing this outcome to the scenario where the entries of measurement matrix $\mathbf{A}_k$ are i.i.d. Gaussian, we observe that the probability density function gap between these two scenarios is notably narrow. Therefore, according to Theorem 1, the linear map employing the random rank-one unit-modulus measurements satisfies the RIP with high probability, similar to the scenario using i.i.d. Gaussian $\mathbf{A}_k$. Furthermore, as the number of measurements K increases, both the Gaussian and rank-one unit-modulus curves become increasingly tightly concentrated around the expected value $\|\mathbf{X}\|_F^2$. This is consistent with the results in Theorem 5, where the tail bound exponentially decreases with the number of measurements K.

Since the results in Theorem 4 are essential in the derivation of RIP analysis, we conduct an experiment in Fig. 2 to confirm the correctness of Theorem 4. Specifically, in Theorem 4, we have established the fact: under the constraint $\|\mathbf{X}\|_F = 1$, the random variable $|\mathbf{u}^H\mathbf{X}\mathbf{v}|$ achieves the largest $(2t)^{th}$ moment when $\mathbf{X} = \frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T$. In other words, we conclude that $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}] \le \frac{1}{M^tN^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}], \forall t$.

For the experiment in Fig. 2, we randomly generate 100 matrices $\mathbf{X} \in \mathbb{C}^{40\times 80}$ and empirically calculate $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ for each $\mathbf{X}$ and $t = 1, 2, \ldots, 10$. The blue line represents the scenario of the all-one matrix, $\frac{1}{M^tN^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}]$. The red curve illustrates the mean and standard deviation for the 100 realizations of arbitrary $\mathbf{X}$. Upon reviewing Fig. 2, we can readily observe that $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}] \le 1/(M^tN^t)\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}], \forall t$. Therefore, the value of $\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}]$ is maximized when the matrix $\mathbf{X}$ has the form of the all-one matrix, i.e., $\mathbf{X} = \frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T$. In conclusion, these experimental results are in perfect alignment with our analysis in Theorem 4.

In Fig. 3, we evaluate the low-rank matrix recovery performance by using the linear map of the rank-one unit-modulus measurements and i.i.d. measurement matrix $\mathbf{A}_k$. We randomly generate the target complex low-rank matrix $\mathbf{X}$ with dimension of 40×80 of rank $r = 5$. We let the number of measurements K vary from 400 to 1500. For fair comparison, we utilize the well-known nuclear norm minimization (Recht et al., 2010) to recover the low-rank matrix from the measurements. The optimization problem is given by

$$\begin{aligned} &\min_{\mathbf{X}} \|\mathbf{X}\|_* \\ &\text{subject to } \mathcal{A}(\mathbf{X}) = \mathbf{y}. \end{aligned}$$

We can refer to the findings in (Recht et al., 2010) that the recovery error by using i.i.d. $\mathbf{A}_k$ decreases with the number of measurements K. As evident in Fig. 3, the recovery error of the designed linear map exhibits a similar trend as the case of i.i.d. Gaussian. Moreover, observe that there is a sharp transition to near zero error at around 1100 measurements for these two scenarios, which is consistent with the results in (Recht et al., 2010). Overall, the observations in Fig. 3 suggest that rank-one measurements with unit-modulus vectors achieves a recovery performance similar to that of i.i.d. Gaussian $\mathbf{A}_k$. Additional numiercal experiments about the recovery performance by using non-convex matrix recovery algorithms are attached in Section C of the supplementary materials.

Furthermore, the use of unit-modulus vectors allows for more efficient memory storage for the linear map, reducing the hardware costs associated with the linear map. As we have analyzed, this type of linear map not only satisfies the RIP but also exhibits similar recovery performance as the case of i.i.d. measurement matrices, as demonstrated in our experiments. Therefore, these advantages position our designed rank-one measurements with unit-modulus vectors as a promising linear map in low-rank matrix recovery.

6 CONCLUSION

In this paper, we have conducted a comprehensive RIP analysis for rank-one measurements with random unit-modulus vectors. The symmetric statistical properties of unit-modulus vectors have allowed us to derive the tail bound which exponentially decrease with the number of measurements . In comparison to the scenario of i.i.d. measurement matrices, the linear map generated by unit-modulus vectors not only satisfies the RIP but also offers the high memory efficiency. These advantageous properties show the potential of rank-one unit-modulus measurements as a highly effective linear map in the field of low-rank matrix recovery.

Acknowledgements

This work was supported in part by the Shenzhen Science and Technology R&D Funds under Grant GXWD20231130104830003.

References

Cai, T. T. and Zhang, A. (2015). ROP: Matrix recovery via rank-one projections. *The Annals of Statistics*, 43(1):102–138.

Candes, E. J., Li, X., and Soltanolkotabi, M. (2015). Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007.

Candes, E. J. and Plan, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772.

Candes, E. J., Strohmer, T., and Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080.

Chen, Y. and Chi, Y. (2018). Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization. *IEEE Signal Processing Magazine*, 35(4):14–31.

Chen, Y., Chi, Y., and Goldsmith, A. J. (2015). Exact and stable covariance estimation from quadratic sampling via convex programming. *IEEE Transactions on Information Theory*, 61(7):4034–4059.

Chi, Y., Lu, Y. M., and Chen, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269.

Davenport, M. A. and Romberg, J. (2016). An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622.

El Ayach, O., Rajagopal, S., Abu-Surra, S., Pi, Z., and Heath, R. W. (2014). Spatially sparse precoding in millimeter wave mimo systems. *IEEE Transactions on Wireless Communications*, 13(3):1499–1513.

Farias, V., Li, A. A., and Peng, T. (2022). Uncertainty quantification for low-rank matrix completion with heterogeneous and sub-exponential noise. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, pages 1179–1189. PMLR.

Ge, R., Jin, C., and Zheng, Y. (2017). No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1233–1242. PMLR.

Jain, P., Meka, R., and Dhillon, I. (2010). Guaranteed rank minimization via singular value projection. *Advances in Neural Information Processing Systems*, 23.

Jain, P., Netrapalli, P., and Sanghavi, S. (2013). Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 665–674.

Kueng, R., Rauhut, H., and Terstiege, U. (2017). Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116.

Li, Y., Ma, C., Chen, Y., and Chi, Y. (2019). Nonconvex matrix factorization from rank-one measurements. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pages 1496–1505. PMLR.

Ma, C., Wang, K., Chi, Y., and Chen, Y. (2018). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3345–3354. PMLR.

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

Richmond, L. B. and Shallit, J. (2008). Counting abelian squares. *arXiv preprint arXiv:0807.5028*.

Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J., and Segev, M. (2015). Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109.

Tong, T., Ma, C., Prater-Bennette, A., Tripp, E., and Chi, Y. (2022). Scaling and scalability: Provable nonconvex low-rank tensor completion. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, pages 2607–2617. PMLR.

Tu, S., Boczar, R., Simchowitz, M., Soltanolkotabi, M., and Recht, B. (2016). Low-rank solutions of linear matrix equations via procrustes flow. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 964–973. PMLR.

Zhang, R., Josz, C., Sojoudi, S., and Lavaei, J. (2018a). How much restricted isometry is needed in nonconvex matrix recovery? *Advances in Neural Information Processing Systems*, 31.

Zhang, W., Kim, T., and Leung, S.-H. (2020). A sequential subspace method for millimeter wave MIMO channel estimation. *IEEE Transactions on Vehicular Technology*, 69(5):5355–5368.

Zhang, W., Kim, T., Love, D. J., and Perrins, E. (2018b). Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems. *IEEE Transactions on Communications*, 66(11):5814–5827.

Zhang, W., Kim, T., Xiong, G., and Leung, S.-H. (2019). Leveraging subspace information for low-rank matrix reconstruction. *Signal Processing*, 163:123–131.

Zhang, W. and Tay, W. P. (2021). Cost-efficient RIS-aided channel estimation via rank-one matrix factorization. *IEEE Wireless Communications Letters*, 10(11):2562–2566.

Zheng, Q. and Lafferty, J. (2015). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. *Advances in Neural Information Processing Systems*, 28.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

The supplementary materials contain detailed proofs of the results that are missing in the main paper, and additional experiments are provided as well.

A PROOF OF THEOREM 4

After standard manipulations, one can find that

$$\mathbb{E}[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}] \leq \mathbb{E}[|\mathbf{u}^H\bar{\mathbf{X}}\mathbf{v}|^{2t}],$$

where each entry in $\bar{\mathbf{X}}$ is equal to the absolute value of the corresponding entry in $\mathbf{X}$. Therefore, it remains to show that

$$\mathbb{E}[|\mathbf{u}^H\bar{\mathbf{X}}\mathbf{v}|^{2t}] \leq \frac{1}{M^tN^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}].$$

For the easy notation, we let $\mathbf{X} = \bar{\mathbf{X}}$. Denoting $\mathbf{X}_{m,:}$ as the mth row of $\mathbf{X}$, we have

$$\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] = \mathbb{E}\left[\left(\sum_m u_m^*\mathbf{X}_{m,:}\mathbf{v}\right)^t\left(\sum_m u_m\mathbf{X}_{m,:}\mathbf{v}^*\right)^t\right].$$

Then, according to the multinomial theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right]\\
&= \mathbb{E}\left[\sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}\prod(u_m^*\mathbf{X}_{1,:}\mathbf{v})^{k_m}\sum_{k_1'+\ldots+k_M'=t}\binom{t}{k_1',\ldots,k_M'}\prod(u_m\mathbf{X}_{m,:}\mathbf{v}^*)^{k_m'}\right]\\
&= \mathbb{E}\left[\sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\prod|\mathbf{X}_{m,:}\mathbf{v}|^{2k_m}\right].
\end{aligned} \tag{23}$$

For the concise proof, we express $\mathbf{X}_{m,:} = c_m\mathbf{x}_m^T$ where $\|\mathbf{x}_m\|_2 = 1$. Without loss generality, we assume $c_m \geq 0, \forall\, m$. Then we rewrite the expression (23) above as

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\prod|\mathbf{X}_{m,:}\mathbf{v}|^{2k_m}\right]\\
&= \mathbb{E}\left[\sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\prod|\mathbf{x}_m^T\mathbf{v}|^{2k_m}\prod_{i=1}^M c_m^{2k_m}\right].
\end{aligned}$$

Then, according to Lemma 4, we have

$$\mathbb{E}\left[\prod|\mathbf{x}_m^T\mathbf{v}|^{2k_m}\right] \leq \max_m\mathbb{E}\left[|\mathbf{x}_m^T\mathbf{v}|^{2t}\right].$$

Thus, the value of $\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right]$ in (23) is upper bounded by

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] &\leq \sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\max_m\mathbb{E}\left[|\mathbf{x}_m^T\mathbf{v}|^{2t}\right]\prod_{i=1}^M c_m^{2k_m}\\
&= \mathbb{E}\left[|\mathbf{u}^H\mathbf{X}^*\mathbf{v}|^{2t}\right],
\end{aligned}$$

where $\mathbf{X}^* = [c_1\mathbf{x}_{m^*}, c_2\mathbf{x}_{m^*}, \cdots, c_M\mathbf{x}_{m^*}]^T$ and $m^* = \arg\max_m\mathbb{E}\left[|\mathbf{X}_{m,:}\mathbf{v}|^{2t}\right]$. In other words,

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] &\leq \mathbb{E}\left[|\mathbf{u}^H\mathbf{X}^*\mathbf{v}|^{2t}\right]\\
&= \sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\mathbb{E}\left[|\mathbf{x}_{m^*}^T\mathbf{v}|^{2t}\right]\prod_{i=1}^M c_m^{2k_m}\\
&= \underbrace{\mathbb{E}\left[|\mathbf{x}_{m^*}^T\mathbf{v}|^{2t}\right]}_{\text{first part}}\underbrace{\sum_{k_1+\ldots+k_M=t}\binom{t}{k_1,\ldots,k_M}^2\prod_{i=1}^M c_m^{2k_m}}_{\text{second part}}.
\end{aligned} \tag{24}$$

By Theorem 3, the first part of value in (24) is maximized when the entries in $\mathbf{x}_{m^*}$ are equivalent. According to Lemma 2, the second part is maximized when $c_m = \sqrt{\frac{1}{M}}$. Recall that $\mathbf{X}^* = [c_1\mathbf{x}_{m^*}, c_2\mathbf{x}_{m^*}, \cdots, c_M\mathbf{x}_{m^*}]^T$, after substituting the setting of $\mathbf{x}_{m^*}$ and c_m into the expression of $\mathbf{X}^*$, one can easily check the corresponding matrix is expressed as $\mathbf{X}^* = \frac{1}{\sqrt{MN}}\mathbf{1}\mathbf{1}^T$. Thus, we have

$$\mathbb{E}\left[|\mathbf{u}^H\mathbf{X}\mathbf{v}|^{2t}\right] \leq \frac{1}{M^tN^t}\mathbb{E}[|\mathbf{u}^H\mathbf{1}\mathbf{1}^T\mathbf{v}|^{2t}].$$

This concludes the proof. □

B PROOF OF PRELIMINARIES

B.1 Proof of Lemma 1

Note that the Legendre polynomials $P_n(x)$ is expressed as

$$P_n(x) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}^2(x-1)^{n-k}(x+1)^k.$$

After simple manipulations, the expression of $s_n(p)$ is expressed as

$$s_n(p) = (2p-1)^nP_n\left(\frac{1}{2p-1}\right).$$

After performing first-order derivative, one can check that the value of s_n is maximized when $p = 1/2$. □

B.2 Proof of Theorem 3

Without loss of generality, we can assume that entries in $\mathbf{x}$ are non-negative, i.e., $x_n \geq 0, \forall n$. This is because $\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right] \leq \mathbb{E}\left[|\bar{\mathbf{x}}^T\mathbf{v}|^{2m}\right]$. Then, the expression of $\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right]$ is

$$\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right] = \mathbb{E}\left[(x_1v_1+\ldots+x_Nv_N)^m(x_1v_1^*+\ldots+x_Nv_N^*)^m\right]. \tag{25}$$

Without loss of generality, we assume $x_1 \neq x_2$. To complete the proof, we will show that by letting $x_1 = x_2 = \sqrt{\frac{x_1^2+x_2^2}{2}}$, the expectation in (25) will increase. Here, we denote $y = \sum_{n=3}^N x_nv_n$, and $t = \sum_{n=3}^N x_n^2$. Therefore, we have $x_1^2 + x_2^2 = 1 - t$. Then, according to the multinomial theorem, we have

$$\begin{aligned}\mathbb{E}\left[|\mathbf{x}^T\mathbf{v}|^{2m}\right] &= \mathbb{E}\left[(x_1v_1+x_2v_2+y)^m(x_1v_1^*+x_2v_2^*+y^*)^m\right]\\ &= \mathbb{E}\Bigg[\sum_{k_1+k_2+k_3=m}\binom{m}{k_1,k_2,k_3}(x_1v_1)^{k_1}(x_2v_2)^{k_2}y^{k_3}\\ &\qquad \sum_{k_1'+k_2'+k_3'=m}\binom{m}{k_1',k_2',k_3'}(x_1v_1^*)^{k_1'}(x_2v_2^*)^{k_2'}y^{k_3'}\Bigg].\end{aligned}$$

Because of the property of expectation, the expression above can be written in

$$\begin{aligned}&\sum_{k_1+k_2+k_3=m}\binom{m}{k_1,k_2,k_3}^2(x_1)^{2k_1}(x_2)^{2k_2}\mathbb{E}\left[|y|^{2k_3}\right]\\ &= \sum_{k_3=0}^{m}\sum_{k_1+k_2=m-k_3}\binom{m}{k_1,k_2,k_3}^2(x_1)^{2k_1}(x_2)^{2k_2}\mathbb{E}\left[|y|^{2k_3}\right]\\ &= \sum_{k_3=0}^{m}\sum_{k_1+k_2=m-k_3}\binom{m}{k_3,m-k_3}^2\binom{m-k_3}{k_1,k_2}^2(x_1)^{2k_1}(x_2)^{2k_2}\mathbb{E}\left[|y|^{2k_3}\right]\\ &= \sum_{k_3=0}^{m}\binom{m}{k_3,m-k_3}^2\mathbb{E}\left[|y|^{2k_3}\right]\sum_{k_1+k_2=m-k_3}\binom{m-k_3}{k_1,k_2}^2(x_1)^{2k_1}(x_2)^{2k_2}.\end{aligned}$$

Based on Corollary 1, the value of $\sum_{k_1+k_2=m-k_3} \binom{m-k_3}{k_1,k_2}^2 (x_1)^{2k_1}(x_2)^{2k_2}$ is maximized when $x_1^2 = x_2^2 = \frac{1-t}{2}$. Because we assume the positiveness, we will have that $x_1 = x_2 = \sqrt{\frac{1-t}{2}}$. This concludes the proof. $\square$

B.3 Proof of Lemma 2

Without loss of generality, we assume $c_1 \neq c_2$ while $c_1^2 + c_2^2$ is a constant. Similar to the proof in Section B.2, it is sufficient to prove that making $c_1 = c_2$ can increase the value in (12). Then, the general result in Lemma 2 can be obtained by induction. First of all, separating c_1, c_2 with $c_m, \forall\, m \geq 3$ gives the following,

$$\begin{aligned}
&\sum_{k_1+\ldots+k_M=t} \binom{t}{k_1,\ldots,k_M}^2 \prod_{m=1}^{M} c_m^{2k_m} \\
&= \sum_{k_1+\ldots+k_M=t} \binom{t}{k_1,\ldots,k_M}^2 c_1^{2k_1} c_2^{2k_2} \prod_{m=3}^{M} c_m^{2k_m} \\
&= \sum_{t_0=0}^{t} \sum_{k_1+k_2=t_0} \sum_{k_3+,\ldots,k_M=t-t_0} \binom{t}{k_1,\ldots,k_M}^2 c_1^{2k_1} c_2^{2k_2} \prod_{m=3}^{M} c_m^{2k_m} \\
&= \sum_{t_0=0}^{t} \sum_{k_1+k_2=t_0} \sum_{k_3+,\ldots,k_M=t-t_0} \binom{t}{t_0, t-t_0}^2 \binom{t_0}{k_1,k_2}^2 \binom{t-t_0}{k_3,\ldots,k_m}^2 c_1^{2k_1} c_2^{2k_2} \prod_{m=3}^{M} c_m^{2k_m}.
\end{aligned}$$

Then, according to Corollary 1, we have the following inquality,

$$\sum_{k_1+k_2=t_0} \binom{t_0}{k_1,k_2}^2 c_1^{2k_1} c_2^{2k_2} \leq \sum_{k_1+k_2=t_0} \binom{t_0}{k_1,k_2}^2 \left(\frac{c_1^2+c_2^2}{2}\right)^{t_0},$$

where the equality holds when $c_1 = c_2 = \sqrt{\frac{c_1^2+c_2^2}{2}}$. This concludes the proof.

B.4 Proof of Lemma 3

We mainly use the Hölder's inequality for the proof of Lemma 3. Specifically, for positive random variables X and Y, if $1/p + 1/q = 1$, then Hölder's inequality shows that

$$\mathbb{E}[XY] \leq (\mathbb{E}[X^p])^{1/p} (\mathbb{E}[Y^q])^{1/q}.$$

For the posted problem, we let $X = X_1^{k_1}$ and $Y = X_2^{k_2}$, then we have

$$\mathbb{E}[X_1^{k_1} X_2^{k_2}] \leq (\mathbb{E}[X_1^{k_1 p}])^{1/p} (\mathbb{E}[X_2^{k_2 q}])^{1/q}.$$

which concludes the proof of (13).

When $k_1 = 0$ or $k_2 = 0$, the inequality in (14) holds trivially. Here, without loss of generality, we assume $k_1, k_2 \neq 0$. By letting $p = \frac{k_1+k_2}{k_1}$ and $q = \frac{k_1+k_2}{k_2}$ in the expression (13), one can check that $1/p + 1/q = 1$. Then, the following inequality holds due to the Hölder's inequality,

$$\begin{aligned}
\mathbb{E}[X_1^{k_1} X_2^{k_2}] &\leq (\mathbb{E}[X_1^{k_1+k_2}])^{\frac{k_1}{k_1+k_2}} (\mathbb{E}[X_2^{k_1+k_2}])^{\frac{k_2}{k_1+k_2}} \\
&\leq \max\{\mathbb{E}[X_1^{k_1+k_2}], \mathbb{E}[X_2^{k_1+k_2}]\}.
\end{aligned}$$

Thus, this concludes the proof of the result in (14). $\square$

B.5 Proof of Lemma 4

Without loss of generality, we assume $k_n > 0, \forall\, n$. Using the Hölder's inequality, we have

$$\mathbb{E}\left[X_1^{k_1} \cdots X_{N-1}^{k_{N-1}} X_N^{k_N}\right] \leq \left(\mathbb{E}\left[\left(X_1^{k_1} \cdots X_{N-1}^{k_{N-1}}\right)^{\frac{t}{t-k_N}}\right]\right)^{\frac{t-k_N}{t}} \left(\mathbb{E}\left[(X_N^{k_N})^{t/k_N}\right]\right)^{k_N/t}, \tag{26}$$

where $p = t/(t-k_N)$ and $q = t/k_N$. Simplifying the R.H.S. in (26) yields

$$\mathbb{E}\left[X_1^{k_1}\cdots X_{N-1}^{k_{N-1}}X_N^{k_N}\right] \le \left(\mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-1}^{k_{N-1}}\right)^{t/(t-k_N)}\right]\right)^{(t-k_N)/t}\left(\mathbb{E}\left[X_N^t\right]\right)^{k_N/t}. \tag{27}$$

Again, using Hölder's inequality for the first part of R.H.S. of (27) with $p = \frac{t-k_N}{t-k_N-k_{N-1}}$ and $q = \frac{t-k_N}{k_{N-1}}$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-1}^{k_{N-1}}\right)^{t/(t-k_N)}\right] \\
&= \mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-2}^{k_{N-2}}\right)^{t/(t-k_N)} X_{N-1}^{k_{N-1}\frac{t}{t-k_N}}\right] \\
&\le \left(\mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-2}^{k_{N-2}}\right)^{\frac{t}{t-k_N}\frac{t-k_N}{t-k_N-k_{N-1}}}\right]\right)^{\frac{t-k_N-k_{N-1}}{t-k_N}}\left(\mathbb{E}\left[\left(X_{N-1}^{k_{N-1}\frac{t}{t-k_N}}\right)^{\frac{t-k_N}{k_{N-1}}}\right]\right)^{\frac{k_{N-1}}{t-k_N}} \\
&= \left(\mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-2}^{k_{N-2}}\right)^{\frac{t}{t-k_N-k_{N-1}}}\right]\right)^{\frac{t-k_N-k_{N-1}}{t-k_N}}\left(\mathbb{E}\left[X_{N-1}^t\right]\right)^{\frac{k_{N-1}}{t-k_N}}.
\end{aligned}$$

Substituting the above equation into (27) gives

$$\begin{aligned}
&\mathbb{E}\left[X_1^{k_1}\cdots X_{N-1}^{k_{N-1}}X_N^{k_N}\right] \\
&\le \underbrace{\left(\mathbb{E}\left[\left(X_1^{k_1}\cdots X_{N-2}^{k_{N-2}}\right)^{\frac{t}{t-k_N-k_{N-1}}}\right]\right)^{\frac{t-k_N-k_{N-1}}{t-k_N}}}_{\text{first part}}\underbrace{\left(\mathbb{E}\left[X_{N-1}^t\right]\right)^{\frac{k_{N-1}}{t}}\left(\mathbb{E}\left[X_N^t\right]\right)^{\frac{k_N}{t}}}_{\text{remaining parts}}.
\end{aligned} \tag{28}$$

For the first part in (28), we iteratively utilize the Hölder's inequality, and finally obtain the following,

$$\begin{aligned}
\mathbb{E}\left[\prod_{n=1}^N X_n^{k_n}\right] &\le \prod_{n=1}^N \left(\mathbb{E}\left[X_n^t\right]\right)^{k_n/t} \\
&\le \max_n \mathbb{E}\left[X_n^t\right].
\end{aligned}$$

This concludes the proof of Lemma 4. □

C ADDITIONAL EXPERIMENTS

In this section, we evaluate the recovery performance of non-convex low-rank matrix recovery algorithms by using the rank-one unit-modulus measurements and i.i.d. Gaussian measurements. Similar to the settings in Fig. 3, we randomly generate the target low-rank matrix $\mathbf{X} \in \mathbb{C}^{40\times 80}$ with a rank of $r = 5$, and the number of measurements K ranges from 500 to 1500. There are two typical non-convex algorithms considered in the experiments. Specifically, in Fig. 4 , we utilize the alternating minimization method (Jain et al., 2013) to recover the low-rank matrix from the rank-one unit-modulus measurements or the i.i.d. Gaussian measurements. In Fig. 5, we employ the gradient-based method (Chi et al., 2019) for matrix recovery task.

As we can see in Figs. 4 and 5, for both typical non-convex methods, i.e., alternating minimization method and gradient-based method, the designed rank-one measurements with random unit-modulus vectors achieves a recovery performance similar to that of i.i.d. Gaussian measurements. Therefore, by integrating the results of experiment in Fig. 3, regardless of whether convex or non-convex optimization algorithms are utilized, the rank-one unit-modulus measurements always exhibit a similar matrix recovery performance to that of i.i.d. measurements, which is attributed to the proven RIP results.

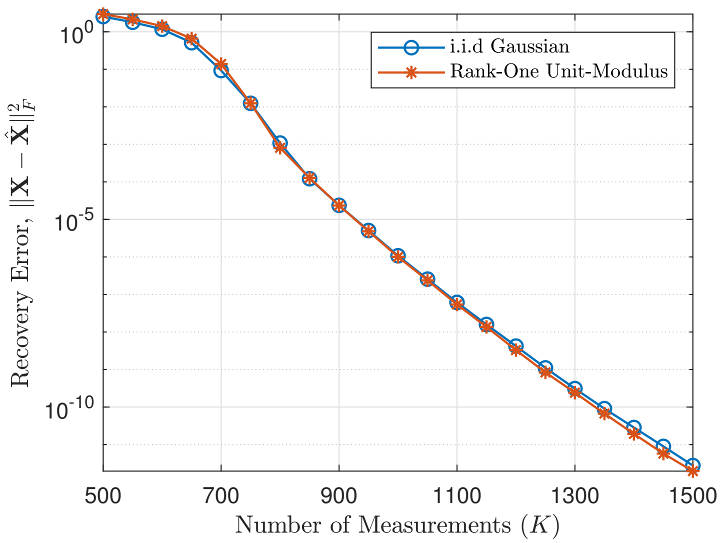


Fig. 4: Recovery Error of Alternating Minimization Method by Using Rank-One Unit-Modulus and i.i.d. Gaussian Measurements with Different Number of Measurements

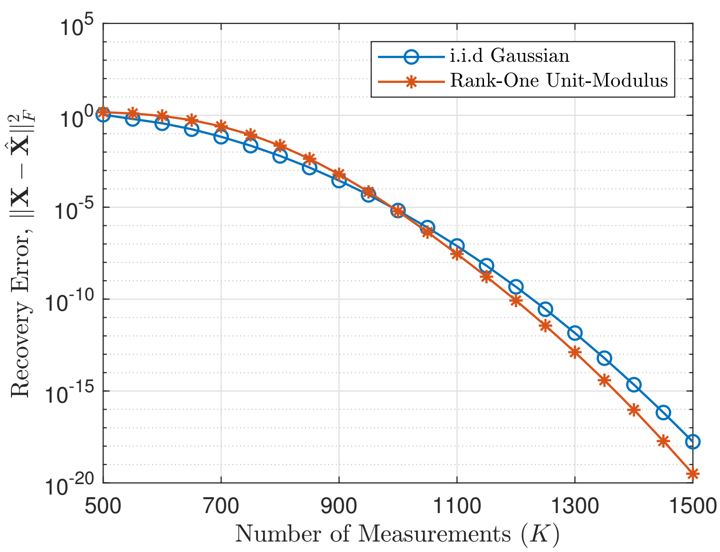


Fig. 5: Recovery Error of Gradient-Based Method by Using Rank-One Unit-Modulus and i.i.d. Gaussian Measurements with Different Number of Measurements