Matrix Recovery via Symmetric Unit-Modulus Rank-One Measurements and Diagonal Information

Tongyu Zhou
Harbin Institute of Technology Shenzhen
210210629@stu.hit.edu.cn

Wei Zhang*
Harbin Institute of Technology Shenzhen
zhangwei.sz@hit.edu.cn

Abstract

Unit-modulus measurement vectors, due to their phase-only variations, exhibit favorable hardware compatibility and storage efficiency. In this paper, we study low-rank matrix recovery using symmetric unit-modulus rank-one measurements. We identify a fundamental limitation of these measurements—their inability to distinguish individual diagonal entries—and we establish recovery guarantees when the diagonal entries are known or directly measurable. Specifically, we construct a stacked operator that combines the unit-modulus measurements with diagonal information and prove that it satisfies a mixed-norm restricted isometry property (mRIP). Based on this property, we establish exact recovery in the noiseless setting and stable recovery in the presence of noise. In particular, based on the distribution of the measurement vectors, our recovery guarantees cover two classes of low-rank complex matrices: general complex matrices and transpose-symmetric complex matrices. Our work provides the theoretical foundation for matrix recovery under symmetric rank-one unit-modulus measurements. Numerical experiments further corroborate these theoretical findings.

1 Introduction

1.1 Problem Setup

The recovery of low-rank matrices has a wide range of applications, including signal processing [Candes et al., 2013], face recognition [Candès et al., 2011], recommendation systems [Koren et al., 2009], and quantum state tomography [Alquier et al., 2013]. The goal of low-rank matrix recovery is to reconstruct a low-rank matrix from a limited set of measurements. In this paper, we consider the problem of matrix recovery via symmetric rank-one projections,

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) + \mathbf{z}, \tag{1}$$

where $\mathbf{y} = (y_1, \cdots, y_m)^T \in \mathbb{C}^m$, and $\mathbf{z} = (z_1, z_2, \cdots, z_m)^T \in \mathbb{C}^m$ denotes the noise vector. The operator $\mathcal{A}$ is a linear map with random measurement matrices satisfying

$$[\mathcal{A}(\mathbf{X})]_j = \langle \mathbf{A}_j, \mathbf{X} \rangle = \langle \mathbf{u}_j \mathbf{u}_j^H, \mathbf{X} \rangle = \mathbf{u}_j^H \mathbf{X} \mathbf{u}_j, \tag{2}$$

where $\mathbf{u}_j = (u_{j,1}, u_{j,2}, \cdots, u_{j,n})^T \in \mathbb{C}^n$. Our goal is to recover $\mathbf{X}$ from these measurements. This rank-one measurement model arises in many applications, including phase retrieval [Candes et al., 2013], quantum state tomography [Kueng et al., 2017], and covariance sketching [Chen et al., 2015].

A key issue in low-rank matrix recovery is the design of measurement vectors $\mathbf{u}_j$ in (2). In this paper, we focus on the following unit-modulus measurement vectors: The entries $u_{j,k}$ are independent and identically distributed (i.i.d.) across all j, k and satisfy

$$\mathbb{E}[u_{j,k}] = 0, \quad |u_{j,k}| = 1, \quad \forall\, j, k. \tag{3}$$

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Define $c_u = |\mathbb{E}[u_{j,k}^2]|$, which satisfies $0 \leq c_u \leq 1$. Accordingly, we study low-rank matrix recovery under (1) and (3) in two regimes: (i) $0 \leq c_u < 1$ and (ii) $c_u = 1$.

1.2 Motivation

Low-rank matrix recovery has found wide applications in various fields where incomplete observations are available and the underlying data exhibit low-rank structures [Gross et al., 2010, Basri and Jacobs, 2003, Candès et al., 2011, Chang et al., 2025]. In such problems, the design of measurement schemes in (1) plays a crucial role. A key consideration is the storage cost of the measurement matrices $\mathbf{A}_j$. In the standard i.i.d. setting, where the entries of $\mathbf{A}_j$ are independently and identically distributed, storing a single measurement matrix requires $O(n^2)$ memory [Recht et al., 2010, Candes and Plan, 2011, Zhang et al., 2018], which becomes prohibitive for large-scale problems. Rank-one measurements, where $\mathbf{A}_j = \mathbf{u}_j\mathbf{v}_j^H$ (asymmetric) [Zhong et al., 2015, Foucart and Subramanian, 2019, Cai and Zhang, 2015] or $\mathbf{A}_j = \mathbf{u}_j\mathbf{u}_j^H$ (symmetric) [Qin et al., 2024, Chen et al., 2015, Kueng et al., 2017, Li et al., 2019], offer a significant advantage in storage efficiency by reducing the memory requirement from $O(n^2)$ to $O(n)$. This substantial reduction in memory footprint makes rank-one measurements particularly attractive for applications with limited computational and storage resources, such as covariance sketching for data streams [Muthukrishnan et al., 2005] and non-coherent energy measurements in communications and signal processing [Daniels and Heath, 2008], as demonstrated in [Chen et al., 2015]. Motivated by these storage advantages, this paper focuses on the matrix recovery problem under symmetric rank-one measurements $\mathbf{u}$ in (3).

Unit-modulus measurement vectors, i.e., $|u_{j,k}| = 1$, exhibit phase-only variations. This property endows them with favorable hardware compatibility and storage efficiency. In computational imaging and phase retrieval, phase-only masks in coded diffraction systems yield constant-modulus measurement vectors and enable physically realizable acquisition designs [Candès et al., 2015, Leblanc et al., 2024]. In wireless communications, analog/hybrid beamforming relies on RF phase shifters that impose constant-modulus constraints on precoders and combiners, while RIS architectures similarly enforce unit-modulus reflection coefficients, making phase-only control naturally consistent with system models [El Ayach et al., 2014, Yu et al., 2020, Pan et al., 2022, Zhang and Tay, 2021, Zhang et al., 2022]. Constant-envelope signaling further improves energy efficiency via power-amplifier-friendly transmission [Mohammed and Larsson, 2013]. Thus, rank-one sensing with unit-modulus vectors in (3) is both memory-efficient and hardware-compatible. Matrix recovery under such measurements is therefore of significant interest, as it provides theoretical insight into sensing with phase-only measurement vectors [Candès et al., 2015, Leblanc et al., 2024, El Ayach et al., 2014, Yu et al., 2020, Pan et al., 2022, Mohammed and Larsson, 2013].

1.3 Related Work

The goal of low-rank matrix recovery is to reconstruct $\mathbf{X}$ from the linear map $\mathcal{A}(\cdot)$ and measurements $\mathbf{y}$ in (1). For matrix recovery problems that do not involve rank-one measurements, various approaches have been developed, including both convex and non-convex methods [Recht et al., 2010, Jain et al., 2013], whose recovery guarantees are based on the following restricted isometry property.

Definition 1.1 (RIP-ℓ_2/ℓ_2, [Recht et al., 2010])**.** *For all $\mathbf{X}$ of rank at most r, the isometry constant δ_r of $\mathcal{A}(\cdot)$ is the smallest quantity such that*

$$(1-\delta_r)||\mathbf{X}||_F^2 \leq ||\mathcal{A}(\mathbf{X})||_2^2 \leq (1+\delta_r)||\mathbf{X}||_F^2, \tag{4}$$

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

When the entries of $\mathbf{A}_j$ are i.i.d. zero-mean complex Gaussian entries [Recht et al., 2010] or zero-mean unit-modulus entries [Zhang et al., 2018], the linear map satisfies RIP $-\ \ell_2/\ell_2$ with high probability, provided that the number of measurements is sufficiently large.

However, for rank-one measurements, the situation is fundamentally different. For rank-one measurements, since the elements in the measurement matrix $\mathbf{A}_j$ are not independent, it becomes challenging to satisfy the RIP $-\ \ell_2/\ell_2$ condition defined in Definition 1.1. In fact, it has been established in [Chen et al., 2015] and [Cai and Zhang, 2015] that, for rank-one sampling models with i.i.d. Gaussian vectors, the corresponding linear map cannot satisfy the RIP $-\ \ell_2/\ell_2$ condition at the desired $O(nr)$ sampling scale in the low-rank regime, both in the asymmetric case $\mathbf{A}_j = \mathbf{u}_j\mathbf{v}_j^H$ and in the symmetric

case $\mathbf{A}_j = \mathbf{u}_j\mathbf{u}_j^H$. Thus, the RIP-$\ell_2/\ell_2$ condition cannot provide recovery guarantees at the desired $O(nr)$ sampling scale for these rank-one measurements.

For this reason, recovery guarantees for matrix recovery under rank-one measurements are usually based on other properties of the linear operator $\mathcal{A}(\cdot)$. For example, convex algorithms mainly focus on nuclear norm minimization for matrix recovery, where exact recovery is established through the RIP $-$ ℓ_2/ℓ_1 (RUB) [Chen et al., 2015, Cai and Zhang, 2015], which is a variant of the RIP $-$ ℓ_2/ℓ_2 in Definition 1.1.

Definition 1.2 (RIP-ℓ_2/ℓ_1, [Chen et al., 2015])**.** *For all* $\mathbf{X}$ *of rank at most* r*, the RIP* $-$ ℓ_2/ℓ_1 *constants* δ_r^{lb} *and* δ_r^{ub} *of* $\mathcal{A}(\cdot)$ *are the smallest quantities such that*

$$(1-\delta_r^{lb})||\mathbf{X}||_F \le ||\mathcal{A}(\mathbf{X})||_1 \le (1+\delta_r^{ub})||\mathbf{X}||_F, \tag{5}$$

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

For the asymmetric rank-one case, Cai and Zhang [2015] study asymmetric sub-Gaussian measurement vectors and show that the associated operator satisfies the RIP $-$ ℓ_2/ℓ_1 property, which guarantees matrix recovery via convex optimization. Zhang and Wang [2024] study asymmetric rank-one measurements $\mathbf{A}_j = \mathbf{u}_j\mathbf{v}_j^H$ generated by independent unit-modulus vectors with i.i.d. uniform phases.

In contrast, for the symmetric rank-one case, Chen et al. [2015] study real symmetric matrices using real sub-Gaussian measurement vectors satisfying the following moment conditions:

$$\mathbb{E}[u_{j,k}] = 0,\ \mathbb{E}\big[|u_{j,k}|^2\big] = 1,\ \mathbb{E}\big[|u_{j,k}|^4\big] > 1,\ \forall\, j,k. \tag{6}$$

In the symmetric case, however, the operator $\mathcal{A}$ in (2) will introduce a bias in the measurements, which thus prevents $\mathcal{A}$ from satisfying RIP $-$ ℓ_2/ℓ_1. To address this issue, Chen et al. [2015] proposed a debiasing operator $\mathcal{B}(\cdot)$. The effectiveness of the convex optimization approach was then established by showing that $\mathcal{B}(\cdot)$ satisfies the RIP $-$ ℓ_2/ℓ_1 property [Chen et al., 2015].

In addition to convex approaches, non-convex algorithms have also been widely studied for rank-one matrix recovery. For example, Zhong et al. [2015] studied low-rank matrix sensing from asymmetric rank-one Gaussian measurements, and proposed an alternating minimization algorithm, proving linear convergence by exploiting concentration properties of Gaussian measurements. Li et al. [2021] proposed a nonconvex gradient descent method with spectral initialization for low-rank PSD matrices from symmetric rank-one Gaussian measurements, and showed global convergence using spectral initialization, local strong convexity, and high-dimensional concentration inequalities. Foucart and Subramanian [2019] introduced a modified iterative hard thresholding algorithm for low-rank recovery from rank-one projections, with linear convergence and robustness guarantees under a RIP-ℓ_2/ℓ_1 condition. Qin et al. [2024] developed a general stochastic gradient descent method for symmetric rank-one matrix sensing with approximately orthogonal measurement vectors, where the recovery accuracy is quantified by a potential function.

Our setting differs from existing symmetric rank-one and asymmetric unit-modulus models in both measurement structure and distributional assumptions. The recovery guarantees of Chen et al. [2015] for real symmetric matrices assume real sub-Gaussian measurement vectors satisfying (6), including the strict fourth-moment condition $\mathbb{E}[|u_{j,k}|^4] > 1$. Although unit-modulus random variables are also sub-Gaussian, they satisfy $\mathbb{E}[|u_{j,k}|^4] = 1$ and hence fall outside those assumptions. Meanwhile, Zhang and Wang [2024] study asymmetric unit-modulus measurements with independent left and right vectors and uniform phases. In that model, the diagonal measurement coefficients are random products, whereas in our symmetric unit-modulus model they are identically one. Consequently, our measurements capture the diagonal only through $\mathrm{tr}(\mathbf{X})$, and the debiasing operator annihilates the entire diagonal part. We therefore incorporate known or directly measured diagonal information and develop a mixed-norm analysis combining the debiased rank-one measurements with diagonal observations. Moreover, we allow general zero-mean unit-modulus distributions characterized by $c_u = |\mathbb{E}u^2|$, including discrete distributions. We further characterize the distinct identifiability regimes for $c_u < 1$ and $c_u = 1$, with uniform phases corresponding to the special case $c_u = 0$.

1.4 Our Contributions

This work focuses on the matrix recovery problem under symmetric rank-one measurements with unit-modulus measurement vectors. Our main contributions are summarized below:

- **Identification of the fundamental limitation of symmetric unit-modulus measurements.** We theoretically establish an inherent limitation of symmetric rank-one measurements with unit-modulus vectors: such measurements can only capture the sum of diagonal entries rather than individual diagonal elements, leading to an irreducible diagonal ambiguity. To address this, we establish a matrix recovery framework under the premise that diagonal information is known or directly measurable, and introduce an auxiliary operator $\mathcal{C}(\cdot)$ to characterize the direct measurement of diagonal entries.
- **Establishment of the mixed-norm restricted isometry property.** For the unit-modulus measurement model, we construct a stacked operator $\mathcal{D}(\cdot)$ by combining the debiasing operator $\mathcal{B}(\cdot)$ and the diagonal sampling operator $\mathcal{C}(\cdot)$, and propose a novel theoretical framework termed the mixed-norm restricted isometry property (mRIP). This property measures the unit-modulus measurements and diagonal outputs in the ℓ_1- and ℓ_2-norms, respectively, avoiding the issus of directly using the ℓ_1-norm for both. We prove that $\mathcal{D}(\cdot)$ satisfies the mRIP with high probability using $m = O(nr)$ random measurements together with the diagonal information in both regimes: (i) general complex matrices $\mathbf{X} \in \mathbb{C}^{n\times n}$[2] with $0 \le c_u < 1$, and (ii) transpose-symmetric complex matrices $\mathbf{X} \in \mathbb{S}_C^n$ with $c_u = 1$.
- **Exact and stable recovery guarantees.** Leveraging the established mRIP conditions, we prove that nuclear norm minimization achieves exact recovery of matrices of rank at most r in the noiseless setting. In the noisy setting, we establish stable recovery with explicit error bounds, where the recovery error is controlled by the best rank-r approximation error and the noise level. These guarantees hold in both regimes, namely (i) $0 \le c_u < 1$ for general complex matrices and (ii) $c_u = 1$ for transpose-symmetric complex matrices, and are corroborated by numerical experiments.

2 Near Isometry of Unit-Modulus Measurements

This section analyzes the near-isometry properties of the stacked operator $\mathcal{D}(\cdot)$, formed by combining the debiasing operator $\mathcal{B}(\cdot)$ and the diagonal map $\mathcal{C}(\cdot)$. We establish high-probability mRIP for $\mathcal{D}(\cdot)$ in two regimes: $0 \le c_u < 1$ over $\mathbf{X} \in \mathbb{C}^{n\times n}$, and $c_u = 1$ over $\mathbf{X} \in \mathbb{S}_C^n$.

2.1 Limitations of Unit-Modulus Measurements

First, we observe that it is not feasible to achieve matrix recovery using only unit-modulus measurements as defined in (3). The main reason is that measurements with random unit-modulus vectors capture only the sum of the diagonal elements rather than the information of each individual diagonal element. This can be seen as follows.

In the noiseless case, for convenience, i.e., when $\mathbf{z} = 0$, the measurements can be expressed as:

$$y_j = \mathbf{u}_j^H \mathbf{X} \mathbf{u}_j = \sum_{p=1}^{n} X_{p,p} + \sum_{p\neq q} X_{p,q} u_{j,p}^* u_{j,q}, \forall\, j. \tag{7}$$

From (7), we observe that, for the diagonal elements of $\mathbf{X}$, the measurements $\mathbf{y}$ depend only on the sum of all diagonal elements $\sum_{p=1}^{n} X_{p,p}$, but not on the individual diagonal entries $X_{p,p}$ themselves. This diagonal ambiguity prevents uniform recovery, regardless of the number of measurements. For example, consider the distinct rank-one matrices $\mathbf{X}_1 = \mathbf{e}_1\mathbf{e}_1^H$ and $\mathbf{X}_2 = \mathbf{e}_2\mathbf{e}_2^H$, where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the first two standard basis vectors. One can verify that they satisfy $\mathbf{u}_j^H \mathbf{X}_1 \mathbf{u}_j = \mathbf{u}_j^H \mathbf{X}_2 \mathbf{u}_j = 1$ for every unit-modulus measurement vector $\mathbf{u}_j$. Because these two matrices produce identical measurements, they cannot be distinguished by this measurement scheme. Thus, low-rank matrices cannot be uniquely recovered in general from these measurements alone.

In many applications, diagonal information is pre-known or directly measurable. For correlation matrices, all diagonal entries equal one [Miao et al., 2016]. In Euclidean distance geometry, the diagonal of a distance matrix is zero [Liberti et al., 2014]. In quantum state tomography, the diagonal of the density matrix can be read out via simple projective measurements [Binosi et al., 2024].

Motivated by this, we study the matrix recovery under unit-modulus symmetric rank-one sensing while assuming that $\mathrm{diag}(\mathbf{X})$ is known or directly measurable. We employ $\mathbf{y}_d \in \mathbb{C}^n$ to characterize

[2]Here, "general" means an arbitrary square complex matrix, with no Hermitian constraint.

the measurements of diagonal entries,

$$\mathbf{y}_d = \mathcal{C}(\mathbf{X}) + \mathbf{z}_d, \tag{8}$$

where $\mathbf{y}_d = (y_{d,1}, \cdots, y_{d,n})^T \in \mathbb{C}^n$, $\mathbf{X} \in \mathbb{C}^{n\times n}$, $\mathbf{z}_d = (z_{d,1}, \cdots, z_{d,n})^T \in \mathbb{C}^n$ denotes the measurement error, and $\mathcal{C}(\cdot) : \mathbb{C}^{n\times n} \to \mathbb{C}^n$ is a linear map with deterministic measurement matrices obeying the following.

$$[\mathcal{C}(\mathbf{X})]_j = \mathbf{e}_j^H \mathbf{X} \mathbf{e}_j, \tag{9}$$

where $\mathbf{e}_j$ is a vector with 1 in the jth position and 0 elsewhere.

2.2 Auxiliary Operator for Unit-Modulus Measurements and Analysis of c_u

For the symmetric rank-one case, the original measurement matrix $\mathbf{A}_j = \mathbf{u}_j \mathbf{u}_j^H$ has non-zero mean, which biases the output measurements. As a consequence, $\mathcal{A}(\cdot)$ fails to satisfy the RIP $-$ ℓ_2/ℓ_1 in Definition 1.2. Inspired by [Chen et al., 2015], an auxiliary operator $\mathcal{B}(\cdot)$ is therefore introduced to study the RIP properties of symmetric rank-one measurements in (2).

Specifically, operator $\mathcal{B}(\cdot) : \mathbb{C}^{n\times n} \to \mathbb{C}^{\lfloor m/2 \rfloor}$ is defined by

$$[\mathcal{B}(\mathbf{X})]_j = \frac{[\mathcal{A}(\mathbf{X})]_{2j-1} - [\mathcal{A}(\mathbf{X})]_{2j}}{\sqrt{2}m}. \tag{10}$$

Recall that in [Chen et al., 2015], when the entries of $\mathbf{u}_j$ are i.i.d. real sub-Gaussian with $\mathbb{E}[u_{j,k}] = 0$, $\mathbb{E}[|u_{j,k}|^2] = 1$, and $\mathbb{E}[|u_{j,k}|^4] > 1$ for all j, k, the operator $\mathcal{B}(\cdot)$ satisfies the RIP $-$ ℓ_2/ℓ_1 on real symmetric matrices of rank at most r, provided that $m \geq Cnr$ for a sufficiently large constant C. However, due to the limitations of unit-modulus measurements discussed in Section 2.1, when $\mathbf{u}_j$ is the unit-modulus vector given in (3), the auxiliary operator $\mathcal{B}(\cdot)$ no longer directly satisfies the RIP $-$ ℓ_2/ℓ_1.

To address this, we construct a stacked operator $\mathcal{D}(\cdot)$ for the unit-modulus sampling model in (3).

$$\mathcal{D}(\mathbf{X}) = \begin{bmatrix} \mathcal{B}(\mathbf{X}) \\ \mathcal{C}(\mathbf{X}) \end{bmatrix}, \tag{11}$$

where the first component of the operator corresponds to $\mathcal{B}(\cdot)$, and the second component represents the measurements of the diagonal elements as defined in (9). We will prove that $\mathcal{D}(\cdot)$ satisfies the mRIP in the following subsections.

Before proceeding, recall the unit-modulus model in (3), namely $\mathbb{E}[u_{j,k}] = 0$ and $|u_{j,k}| = 1$. We find that whether $\mathcal{D}(\cdot)$ in (11) satisfies the mRIP is governed by the scalar parameter

$$c_u := \left| \mathbb{E}\left[u_{j,k}^2\right] \right|. \tag{12}$$

The property of c_u, and the structure at the boundary $c_u = 1$, are summarized in the following lemma.

Lemma 2.1. *Let $\{u_{j,k}\}$ be identically distributed random variables with $\mathbb{E}[u_{j,k}] = 0$ and $|u_{j,k}| = 1$, and let c_u be as in* (12)*. Then, c_u is always confined to the unit interval, i.e.,*

$$0 \leq c_u \leq 1. \tag{13}$$

Moreover, whenever $c_u = 1$, each $u_{j,k}$ is uniformly distributed on the two-point set $\{e^{\mathrm{i}\theta}, e^{\mathrm{i}(\theta+\pi)}\}$ for some fixed constant $\theta \in [0, \pi)$.

The proof of Lemma 2.1 is given in Appendix B.1. Therefore, according to the reuslts of Lemma 2.1, when $c_u = 1$, the law of each $u_{j,k}$ collapses to a two-point distribution on the unit circle. This critical boundary condition is precisely what governs whether $\mathcal{D}(\cdot)$ can satisfy the mRIP for the low-rank matrices. In the next two subsections, we develop the analysis in detail by studying the mRIP of $\mathcal{D}(\cdot)$ separately for (i) $0 \leq c_u < 1$ and (ii) $c_u = 1$.

2.3 mRIP for Unit-Modulus Measurements with $c_u < 1$

In this subsection, we analyze the isometry property of the operator $\mathcal{D}(\cdot)$ in (11) when the measurement vectors follow the unit-modulus model in (3) with $0 \leq c_u < 1$ and $\mathbf{X} \in \mathbb{C}^{n\times n}$ is a general complex matrix. Since $\mathcal{B}(\mathbf{X})$ contains no diagonal information, we write $\mathbf{X}_o$ for the matrix obtained by setting the diagonal entries of $\mathbf{X}$ to zero. The following lemma bounds $\|\mathcal{B}(\mathbf{X})\|_1$ in terms of $\|\mathbf{X}_o\|_F$ with high probability for any fixed matrix $\mathbf{X}$.

Lemma 2.2. *Let $\mathcal{A}(\cdot)$ be sampled from the unit-modulus model with $0 \leq c_u < 1$ in* (3)*, and let $\mathcal{B}(\cdot)$ be given in* (10)*. For any fixed matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$, there exist universal constants $c_1, c_2, c_3 > 0$ such that, with probability at least $1 - 2\exp(-c_3 m)$, one has*

$$c_1||\mathbf{X}_o||_F \leq ||\mathcal{B}(\mathbf{X})||_1 \leq c_2||\mathbf{X}_o||_F. \tag{14}$$

Combining Lemma 2.2 with the diagonal measurements gives the corresponding bounds for $||\mathcal{D}(\mathbf{X})||_1$ in the following corollary.

Corollary 2.1. *For any fixed matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$, under the assumptions of Lemma 2.2 and taking $c_1 \leq 1$ without loss of generality, with probability at least $1 - 2\exp(-c_3 m)$, the operator $\mathcal{D}(\cdot)$ in* (11) *satisfies the following*

$$c_1||\mathbf{X}||_F \leq ||\mathcal{D}(\mathbf{X})||_1 \leq \sqrt{2}\max(c_2, \sqrt{n})||\mathbf{X}||_F. \tag{15}$$

Proofs of Lemma 2.2 and Corollary 2.1 are provided in Appendices B.2 and B.3, respectively. We observe from (15) that the upper bound in Corollary 2.1 depends on the matrix dimension n through $||\mathcal{C}(\mathbf{X})||_1$. Thus, this bound does not provide a universal constant needed for our recovery analysis. To address this, we introduce the following mixed version of RIP.

Definition 2.1 (mRIP for low-rank matrices)**.** *Let $\mathcal{D}(\cdot)$ be the operator formed by concatenating the outputs of two operators $\mathcal{B}(\cdot)$ and $\mathcal{C}(\cdot)$, i.e.,*

$$\mathcal{D}(\mathbf{X}) = \begin{bmatrix} \mathcal{B}(\mathbf{X}) \\ \mathcal{C}(\mathbf{X}) \end{bmatrix}. \tag{16}$$

We define the mRIP constants δ_r^{lb} and δ_r^{ub} with respect to $\mathcal{D}(\cdot)$ as the smallest numbers such that the following inequalities hold for all matrices $\mathbf{X}$ of rank at most r:

$$(1 - \delta_r^{lb})||\mathbf{X}||_F \leq ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq (1 + \delta_r^{ub})||\mathbf{X}||_F. \tag{17}$$

We measure the component $\mathcal{B}(\cdot)$ in the ℓ_1-norm and the component $\mathcal{C}(\cdot)$ in the ℓ_2-norm, which removes the dependence on the matrix dimension n in the upper bound of $\|\mathcal{D}(\mathbf{X})\|_1$ in (15). Combining the concentration bound for $\mathcal{B}(\mathbf{X})$ in Lemma 2.2 with the deterministic identity $\|\mathcal{C}(\mathbf{X})\|_2 = \|\operatorname{diag}(\mathbf{X})\|_2$ yields the following bound for any fixed matrix.

Lemma 2.3. *Let $\mathcal{A}(\cdot)$ be the unit-modulus model with $0 \leq c_u < 1$ in* (3) *and $\mathcal{C}(\cdot)$ be the diagonal sampling model in* (8)*. For any fixed matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$, there exist universal constants $c_1, c_2, c_3 > 0$ such that, with probability at least $1 - 2\exp(-c_3 m)$, one has*

$$c_1||\mathbf{X}||_F \leq ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq c_2||\mathbf{X}||_F. \tag{18}$$

The proof of Lemma 2.3 is given in Appendix B.4. A covering argument extends this bound to all matrices of rank at most r, yielding the following mRIP guarantee with $m = O(nr)$.

Theorem 2.1. *Consider the unit-modulus sampling model with $0 \leq c_u < 1$ in* (3)*, the diagonal sampling model in* (8)*. The operator $\mathcal{D}(\cdot)$ in* (11) *satisfies mRIP in Definition 2.1 for all matrices $\mathbf{X} \in \mathbb{C}^{n\times n}$ of rank at most r with probability exceeding $1 - \exp(-c_3 m)$,*

$$(1 - \delta_r^{lb})||\mathbf{X}||_F \leq ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq (1 + \delta_r^{ub})||\mathbf{X}||_F \tag{19}$$

where $\delta_r^{lb}, \delta_r^{ub}$ obey

$$1 - \delta_r^{lb} \geq \frac{2}{3}c_1, 1 + \delta_r^{ub} \leq \frac{4}{3}c_2, \tag{20}$$

provided that $m > c_4 nr$, where $c_3, c_4 > 0$ are universal constants, and $c_1, c_2 > 0$ are the universal constants given in (18)*.*

Thus, we have established the mRIP for $\mathcal{D}(\cdot)$ in (11) with $0 \leq c_u < 1$. In Section 3.1, we will show that exact and stable matrix recovery can be achieved based on this property. The covering argument for Theorem 2.1 is detailed in Appendix B.5.

2.4 mRIP for Unit-Modulus Measurements with $c_u = 1$

In this subsection, we study the mRIP for $\mathcal{D}(\cdot)$ under the unit-modulus sampling model with $c_u = 1$ in (3). We first show that, when $c_u = 1$, a general complex matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ is not identifiable through the corresponding unit-moduls measurements.

Recall when $c_u = 1$, Lemma 2.1 implies that $u_{j,k}$ is uniformly distributed in $\{e^{\mathrm{i}\theta}, e^{\mathrm{i}(\theta+\pi)}\}$ for some fixed constant $\theta \in [0, \pi)$. Equivalently, each measurement vector can be written as

$$\mathbf{u}_j = e^{\mathrm{i}\theta}\mathbf{s}_j, \qquad \mathbf{s}_j \in \{\pm 1\}^n. \tag{21}$$

The *binary unit-modulus* measurements in (21) will lead to an intrinsic identifiability limitation, as shown in the following proposition.

Proposition 2.1 (Binary unit-modulus degeneracy)**.** *Let* $\mathbf{u} = e^{\mathrm{i}\theta}\mathbf{s}$*, where* $\mathbf{s} \in \{\pm 1\}^n$ *and* $\theta \in [0, \pi)$*. Then, for any* $\mathbf{X} \in \mathbb{C}^{n\times n}$*,*

$$\mathbf{u}^H\mathbf{X}\mathbf{u} = \mathbf{s}^T\mathbf{X}\mathbf{s} = \mathbf{s}^T\left(\frac{\mathbf{X} + \mathbf{X}^T}{2}\right)\mathbf{s}. \tag{22}$$

Consequently, every transpose-skew-symmetric matrix $\mathbf{K} \in \mathbb{C}^{n\times n}$ *satisfying* $\mathbf{K}^T = -\mathbf{K}$ *is annihilated by the measurement, namely* $\mathbf{u}^H\mathbf{K}\mathbf{u} = 0$*.*

Consequently, for general matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$ and any transpose-skew-symmetric matrix $\mathbf{K}$, we have $\mathcal{D}(\mathbf{X}) = \mathcal{D}(\mathbf{X} + \mathbf{K})$ when $c_u = 1$. Thus, a general complex matrix cannot be uniquely identified from the observations $\mathcal{D}(\mathbf{X})$. The proof of Proposition 2.1 is given in Appendix B.6. Therefore, when $c_u = 1$, we restrict the target matrix to the transpose-symmetric complex matrix class $\mathbb{S}_C^n$, where

$$\mathbb{S}_C^n := \{\mathbf{X} \in \mathbb{C}^{n\times n} : \mathbf{X} = \mathbf{X}^T\}. \tag{23}$$

This restriction eliminates the structural ambiguity in Proposition 2.1. Analogous to Theorem 2.1, when the unit-moduls model in (3) satisfies $c_u = 1$, the following theorem establishes the mRIP for low-rank transpose-symmetric complex matrices.

Theorem 2.2. *Consider the unit-modulus sampling model with* $c_u = 1$ *in* (3)*, the diagonal sampling model in* (8)*. The operator* $\mathcal{D}(\cdot)$ *in* (11) *satisfies mRIP in Definition 2.1 for all matrices* $\mathbf{X} \in \mathbb{S}_C^n$ *of rank at most* r *with probability exceeding* $1 - \exp(-c_3 m)$*,*

$$(1 - \delta_r^{lb})||\mathbf{X}||_F \le ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \le (1 + \delta_r^{ub})||\mathbf{X}||_F \tag{24}$$

where $\delta_r^{lb}, \delta_r^{ub}$ *obey*

$$1 - \delta_r^{lb} \ge \frac{2}{3}c_1, 1 + \delta_r^{ub} \le \frac{4}{3}c_2, \tag{25}$$

provided that $m > c_4 nr$*, where* $c_3, c_4 > 0$ *are universal constants, and* $c_1, c_2 > 0$ *are the universal constants given in Lemma B.7.*

The proof of Theorem 2.2 is given in Appendix B.7.

3 Exact and Stable Matrix Recovery With Unit-Modulus Measurements

Leveraging the mRIP established in Section 2, we derive exact and stable recovery guarantees for the nuclear norm minimization estimator under symmetric rank-one unit-modulus measurements. For both regimes $0 \le c_u < 1$ over general complex matrices $\mathbf{X} \in \mathbb{C}^{n\times n}$ and $c_u = 1$ over transpose-symmetric complex matrices $\mathbf{X} \in \mathbb{S}_C^n$, we show that $O(nr)$ measurements suffice to achieve exact recovery in the noiseless setting and stable recovery with explicit error bounds in the noisy setting.

3.1 Matrix Recovery for $0 \le c_u < 1$

Consider the rank-one measurement operator $\mathcal{A}(\cdot)$ in (2), generated by unit-modulus vectors satisfying (3) with $0 \le c_u < 1$, together with the diagonal measurement operator $\mathcal{C}(\cdot)$ in (8). We aim to recover the low-rank matrix $\mathbf{X}$ via the following nuclear-norm minimization problem,

$$\hat{\mathbf{X}} = \underset{\mathbf{M}\in\mathbb{C}^{n\times n}}{\arg\min} ||\mathbf{M}||_*, \text{s. t. } ||\mathbf{y} - \mathcal{A}(\mathbf{M})||_1 \le \varepsilon_1, ||\mathbf{y}_d - \mathcal{C}(\mathbf{M})||_2 \le \varepsilon_2, \tag{26}$$

where ε_1 and ε_2 are known bounds satisfying $||\mathbf{z}||_1 \le \varepsilon_1$ and $||\mathbf{z}_d||_2 \le \varepsilon_2$ for the noise in (1) and (8). The following lemma establishes a recovery error bound under mRIP condition on the operator $\mathcal{D}(\cdot)$.

Lemma 3.1. *Assume that the noise vectors in* (1) *and* (8) *satisfy* $||\mathbf{z}||_1 \leq \varepsilon_1$ *and* $||\mathbf{z}_d||_2 \leq \varepsilon_2$. *For a given integer* $1 \leq r \leq n$, *consider any matrix* $\mathbf{X} \in \mathbb{C}^{n\times n}$, *and let* $\delta^{lb}_{2r+K_1}, \delta^{ub}_{K_1}$ *be the mRIP constants for* $\mathcal{D}(\cdot) : \mathbb{C}^{n\times n} \to \mathbb{C}^{\lfloor \frac{m}{2} \rfloor + n}$ *in* (11). *If there exists an integer* $K_1 > 2r$ *such that*

$$\frac{1-\delta^{lb}_{2r+K_1}}{\sqrt{2}} - (1+\delta^{ub}_{K_1})\sqrt{\frac{2r}{K_1}} \geq \beta_1 > 0, \tag{27}$$

holds for some numerical value β_1, *the minimizer* $\hat{\mathbf{X}}$ *to* (26) *obeys*

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \leq (\frac{C_1}{\beta_1} + C_3)\frac{||\mathbf{X}_c||_*}{\sqrt{K_1}} + \frac{C_2}{\beta_1}(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2), \tag{28}$$

where $C_1, C_2, C_3 > 0$ *are some universal constants depending only on* $\delta^{lb}_{2r+K_1}, \delta^{ub}_{K_1}$. *Here,* $\mathbf{X}_c = \mathbf{X} - \mathbf{X}_r$, *and* $\mathbf{X}_r$ *denotes the best rank-*r *approximation of* $\mathbf{X}$.

The proof of the lemma is provided in Appendix B.8. The results in Lemma 3.1 show that the recovery error of (26) is bounded when the operator $\mathcal{D}(\cdot)$ satisfies the mRIP in (27). In particular, the bound consists of three terms: the low-rank approximation error associated with $\mathbf{X}_c$ and the noise contributions from $\mathcal{A}(\cdot)$ and $\mathcal{C}(\cdot)$, controlled by ε_1 and ε_2, respectively. Moreover, it is worth noting that the lemma applies to any fixed operator $\mathcal{D}(\cdot)$ satisfying the stated mRIP condition. Consequently, if a randomly sampled operator satisfies this mRIP condition with high probability, the error bound in (28) will also hold with high probability. Formally, combining this lemma with the mRIP guarantee in Theorem 2.1 yields the following recovery theorem.

Theorem 3.1. *Consider the unit-modulus rank-one measurement with* $0 \leq c_u < 1$ *in* (3), *diagonal sampling model in* (8) *and assume that* $||\mathbf{z}||_1 \leq \varepsilon_1$,$||\mathbf{z}_d||_2 \leq \varepsilon_2$. *Then with probability exceeding* $1 - 2\exp(-c_0 m)$, *the solution* $\hat{\mathbf{X}}$ *to* (26) *satisfies*

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \leq C_1\frac{||\mathbf{X}_c||_*}{\sqrt{r}} + C_2(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2), \tag{29}$$

for all $\mathbf{X} \in \mathbb{C}^{n\times n}$, *provided that* $m > c_1 nr$. *Here,* $\mathbf{X}_c = \mathbf{X} - \mathbf{X}_r$ *with* $\mathbf{X}_r$ *denoting the best rank-*r *approximation of* $\mathbf{X}$, *and* $c_0, c_1, C_1, C_2 > 0$ *are some universal constants.*[3]

The proof of Theorem 3.1 is given in Appendix B.9. Under the unit-modulus rank-one measurement model with $0 \leq c_u < 1$ in (3) and the diagonal sampling model in (8), Theorem 3.1 shows that exact and stable matrix recovery can be achieved through the estimator in (26) when the number of measurements satisfies $m \geq Cnr$ for a sufficiently large constant C. As in Lemma 3.1, the error bound consists of three terms: the low-rank approximation error and the noise contributions from the rank-one and diagonal measurements. In particular,

- In the noiseless case, i.e., $\varepsilon_1 = \varepsilon_2 = 0$, we have $||\hat{\mathbf{X}} - \mathbf{X}||_F \leq C_1||\mathbf{X}_c||_*/\sqrt{r}$. Furthermore, when $\mathrm{rank}(\mathbf{X}) \leq r$, $\mathbf{X}_c = \mathbf{X} - \mathbf{X}_r = 0$, and $||\hat{\mathbf{X}} - \mathbf{X}||_F = 0$ implies exact matrix recovery.
- In the noisy case with $\mathrm{rank}(\mathbf{X}) \leq r$, we have $||\hat{\mathbf{X}} - \mathbf{X}||_F \leq C_2(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2)$, where the recovery error is controlled by the noise levels in the rank-one and diagonal measurements.

3.2 Matrix Recovery for $c_u = 1$

For unit-modulus measurements with $c_u = 1$, due to the degeneracy in Proposition 2.1, we assume that the target matrix $\mathbf{X}$ is a transpose-symmetric complex matrix. Accordingly, compared to (26), we impose this additional constraint and recover $\mathbf{X}$ by solving the following optimization problem:

$$\hat{\mathbf{X}} = \underset{\mathbf{M}\in\mathbb{S}^n_C}{\arg\min} ||\mathbf{M}||_*, \mathrm{s.t.}\ ||\mathbf{y} - \mathcal{A}(\mathbf{M})||_1 \leq \varepsilon_1, ||\mathbf{y}_d - \mathcal{C}(\mathbf{M})||_2 \leq \varepsilon_2, \tag{30}$$

We also obtain a recovery guarantee similar to that in Theorem 3.1 in the following.

[3] It is worth noting that, with a slight abuse of notation, we use the same symbols for these constants as in Lemma 3.1, although they are defined differently. The constants in Theorem 3.2 are likewise redefined.

Theorem 3.2. *Consider the unit-modulus rank-one measurement with $c_u = 1$ in* (3)*, diagonal sampling model in* (8) *and assume that $||\mathbf{z}||_1 \leq \varepsilon_1$,$||\mathbf{z}_d||_2 \leq \varepsilon_2$. Then with probability exceeding $1 - 2\exp(-c_0 m)$, the solution $\hat{\mathbf{X}}$ to* (30) *satisfies*

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \leq C_1 \frac{||\mathbf{X}_c||_*}{\sqrt{r}} + C_2(\frac{\varepsilon_1}{\sqrt{2m}} + \varepsilon_2), \tag{31}$$

for all $\mathbf{X} \in \mathbb{S}_C^n$, provided that $m > c_1 nr$. Here, $\mathbf{X}_c = \mathbf{X} - \mathbf{X}_r$ with $\mathbf{X}_r$ denoting the best rank-r approximation of $\mathbf{X}$, and $c_0, c_1, C_1, C_2 > 0$ are some universal constants.

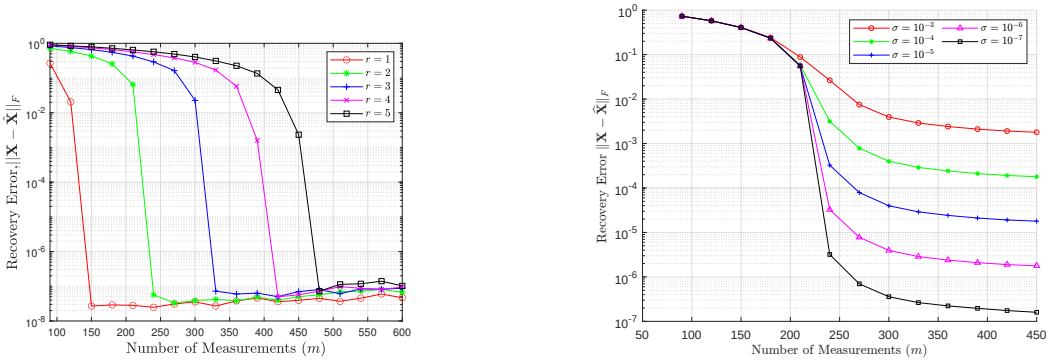
The proof of Theorem 3.2 is given in Appendix B.10. In particular, for unit-modulus rank-one measurements with $0 \leq c_u < 1$, Theorem 3.1 establishes recovery guarantees for general complex matrices and hence also for transpose-symmetric complex matrices. In other words, if the target matrix is transpose-symmetric complex, the results in Theorems 3.1 and 3.2 show that any random unit-modulus measurements (3) (i.e, $0 \leq c_u \leq 1$), together with diagonal information, can guarantee recovery via nuclear norm minimization under the corresponding sampling conditions.

In summary, Theorems 3.1 and 3.2 show that, for both $0 \leq c_u < 1$ (general complex matrices) and $c_u = 1$ (transpose-symmetric complex matrices), nuclear norm minimization achieves exact recovery of matrices with rank at most r in the noiseless setting and stable recovery with explicit error bounds in the noisy setting. These guarantees require $m = O(nr)$ random rank-one measurements together with the diagonal entries; if the diagonal is not already known, acquiring it requires n additional observations, giving $m + n = O(nr)$ total observations.

4 Numerical Experiments

In this section, we evaluate the matrix recovery performance under symmetric rank-one measurements with unit-modulus vectors by solving (26). The convex programs are modeled in CVX in MATLAB and solved using SDPT3. The experimental code is available at `https://github.com/tongyuz941/Matrix-Recovery-Experiments`. Due to space limitations, additional experiments are provided in Appendix A, including discrete unit-modulus measurements with phases drawn from $\{0, 2\pi/3, 4\pi/3\}$, a comparison with complex Gaussian measurements, and binary unit-modulus measurements for both general and transpose-symmetric complex matrices.

In the experiments, we generate a 30×30 matrix $\mathbf{X}$ via $\mathbf{X} = \mathbf{U}\mathbf{V}^H/||\mathbf{U}\mathbf{V}^H||_F$, where $\mathbf{U}$ and $\mathbf{V}$ are randomly generated $30 \times r$ matrices with independent complex Gaussian entries. The measurement vectors $\mathbf{u}_j$ for all j are of the form $\mathbf{u}_j = [e^{i\theta_{j,1}}, \ldots, e^{i\theta_{j,n}}]^T$, where $\theta_{j,k}$ are i.i.d. uniformly distributed on $[0, 2\pi)$. The recovery error is measured as $||\mathbf{X} - \hat{\mathbf{X}}||_F$.



(a) noiseless case

(b) noisy case

Figure 1: The error of the reconstructed matrix via (26) in (a) the noiseless case and (b) the noisy case.

Figure 1a shows the matrix recovery performance in the noiseless case with $\varepsilon_1 = \varepsilon_2 = 0$. In this experiment, we vary the rank r from 1 to 5 and the number of measurements m from 90 to 600. For each (m, r) pair, we repeat 20 trials. At fixed n, the recovery curves in Figure 1a exhibit a trend consistent with the $O(nr)$ sample complexity in Theorem 3.1.

Figure 1b shows the matrix recovery performance in the noisy case with $r = 2$. We introduce additive bounded noise to each measurement by setting $\mathbf{z}$ and $\mathbf{z}_d$ to be generated from $\sigma\frac{\mathcal{U}[-1,1]+i\mathcal{U}[-1,1]}{\sqrt{2}}$, where $\mathcal{U}[-1,1]$ denotes the uniform distribution on $[-1,1]$ and σ is the noise level. We vary σ from 10^{-7} to 10^{-3} and the number of measurements m from 90 to 450. For each (σ, m) pair, we repeat the experiment 20 times. The estimator used in Figure 1b is (26) with noise radii $\varepsilon_1 = \sigma m$ and $\varepsilon_2 = \sigma\sqrt{n}$. For an exactly rank-$r$ target under the conditions of Theorem 3.1, (29) gives the theoretical error bound $||\hat{\mathbf{X}} - \mathbf{X}||_F \leq C_2(1/\sqrt{2} + \sqrt{n})\sigma$. As shown in Figure 1b, the recovery error increases with the noise level σ, which is consistent with our analysis.

5 Conclusion

In this paper, we have provided a comprehensive analysis of matrix recovery via symmetric rank-one measurements with random unit-modulus vectors. We identify the fundamental limitation of such measurements — the inability to recover individual diagonal entries — and resolve it by incorporating direct diagonal measurements. We construct a stacked operator $\mathcal{D}(\cdot)$ and establish that it satisfies the mRIP uniformly over matrices of rank at most r with $O(nr)$ random measurements together with diagonal information, for both $0 \leq c_u < 1$ over general complex matrices and $c_u = 1$ over transpose-symmetric complex matrices. Leveraging these conditions, we prove exact recovery of matrices of rank at most r in the noiseless setting and stable recovery with explicit error bounds in the noisy setting. Numerical experiments support these theoretical results.

Acknowledgments and Disclosure of Funding

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62401180, the Guangdong Basic and Applied Basic Research Foundation under Grant No. 2025A1515011733, and the Shenzhen Science and Technology R&D Funds under Grant No. GXWD20231130104830003.

References

Pierre Alquier, Cristina Butucea, Mohamed Hebiri, Katia Meziani, and Tomoyuki Morimae. Rank-penalized estimation of a quantum system. *Physical Review A—Atomic, Molecular, and Optical Physics*, 88(3):032113, 2013.

Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003.

Daniele Binosi, Giovanni Garberoglio, Diego Maragnano, Maurizio Dapor, and Marco Liscidini. A tailor-made quantum state tomography approach. *APL Quantum*, 1(3), 2024.

T Tony Cai and Anru Zhang. ROP: Matrix recovery via rank-one projections. *The Annals of Statistics*, 43(1):102–138, 2015. ISSN 0090-5364.

Emmanuel J Candes and Yaniv Plan. 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, 2011.

Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

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

Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015. doi: 10.1016/j.acha.2014.09.004.

Gengshuo Chang, Wei Zhang, and Lehan Zhang. Matrix completion with incomplete side information via orthogonal complement projection. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 7393–7416. PMLR, 2025.

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

Robert C Daniels and Robert W Heath. 60 GHz wireless communications: Emerging requirements and design recommendations. *IEEE Vehicular Technology Magazine*, 2(3):41–50, 2008.

Omar El Ayach, Sridhar Rajagopal, Shadi Abu-Surra, Zhouyue Pi, and Robert W. Heath. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Transactions on Wireless Communications*, 13(3):1499–1513, March 2014. doi: 10.1109/TWC.2014.011714.130846.

Simon Foucart and Srinivas Subramanian. Iterative hard thresholding for low-rank recovery from rank-one projections. *Linear Algebra and its Applications*, 572:117–134, 2019.

David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15):150401, 2010.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 665–674, 2013.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

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

Olivier Leblanc, Matthias Hofer, Siddharth Sivankutty, Hervé Rigneault, and Laurent Jacques. Interferometric lensless imaging: rank-one projections of image frequencies with speckle illuminations. *IEEE Transactions on Computational Imaging*, 10:208–222, 2024.

Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi. Nonconvex matrix factorization from rank-one measurements. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1496–1505. PMLR, 2019.

Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi. Nonconvex matrix factorization from rank-one measurements. *IEEE Transactions on Information Theory*, 67(3):1928–1950, 2021. doi: 10.1109/TIT.2021.3050427.

Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56(1):3–69, 2014.

Weimin Miao, Shaohua Pan, and Defeng Sun. A rank-corrected procedure for matrix completion with fixed basis coefficients. *Mathematical Programming*, 159(1):289–338, 2016.

Saif Khan Mohammed and Erik G. Larsson. Per-antenna constant envelope precoding for large multi-user MIMO systems. *IEEE Transactions on Communications*, 61(3):1059–1071, 2013. doi: 10.1109/TCOMM.2013.012913.110827.

Shanmugavelayutham Muthukrishnan et al. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, 1(2):117–236, 2005.

Cunhua Pan, Gui Zhou, Kangda Zhi, Sheng Hong, Tuo Wu, Yijin Pan, Hong Ren, Marco Di Renzo, A. Lee Swindlehurst, Rui Zhang, and Angela Yingjun Zhang. An overview of signal processing techniques for RIS/IRS-aided wireless systems. *IEEE Journal of Selected Topics in Signal Processing*, 16(5):883–917, 2022. doi: 10.1109/JSTSP.2022.3195671.

Lianke Qin, Zhao Song, and Ruizhe Zhang. A general algorithm for solving rank-one matrix sensing. In *International Conference on Artificial Intelligence and Statistics*, pages 757–765. PMLR, 2024.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Xianghao Yu, Dongfang Xu, Ying Sun, Derrick Wing Kwan Ng, and Robert Schober. Robust and secure wireless communications via intelligent reflecting surfaces. *IEEE Journal on Selected Areas in Communications*, 38(11):2637–2652, November 2020. doi: 10.1109/JSAC.2020.3007043.

Wei Zhang and Wee Peng Tay. Cost-efficient RIS-aided channel estimation via rank-one matrix factorization. *IEEE Wireless Communications Letters*, 10(11):2562–2566, 2021. doi: 10.1109/LWC.2021.3107547.

Wei Zhang and Zhenni Wang. Restricted isometry property of rank-one measurements with random unit-modulus vectors. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li, editors, *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1900–1908. PMLR, 02–04 May 2024.

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

Wei Zhang, Miaomiao Dong, and Taejoon Kim. MMV-based sequential AoA and AoD estimation for millimeter wave MIMO channels. *IEEE Transactions on Communications*, 70(6):4063–4077, 2022. doi: 10.1109/TCOMM.2022.3168886.

Kai Zhong, Prateek Jain, and Inderjit S Dhillon. Efficient matrix sensing using rank-1 Gaussian measurements. In *International Conference on Algorithmic Learning Theory*, pages 3–18. Springer, 2015.

A Additional Numerical Experiments

In the noisy experiments below, we set the noise radii to $\varepsilon_1 = m\sigma$ and $\varepsilon_2 = \sqrt{n}\sigma$.

A.1 Phase Transition in the Noiseless Case

We further examine the noiseless recovery performance with matrix size $n = 50$, ranks $r = 1, \ldots, 10$, and measurement counts $m = 50, 100, \ldots, 2000$. The target matrices and continuous uniform-phase measurement vectors are generated as in Section 4. We use exact diagonal information and solve (26) with $\varepsilon_1 = \varepsilon_2 = 0$. For each (m, r) pair, we conduct 10 trials and regard recovery as successful when $\|\hat{\mathbf{X}} - \mathbf{X}\|_F / \|\mathbf{X}\|_F < 10^{-3}$.

Figure 2 shows the empirical recovery success probabilities, with darker and lighter regions indicating lower and higher success rates, respectively. The blue curve indicates, for each rank r, the smallest tested measurement count m at which at least 50% of the trials succeed. As shown in Figure 2, at fixed n, the empirical 50% success threshold indicated by the blue curve increases approximately linearly with r, consistent with the $O(nr)$ sample-complexity trend in Theorem 3.1. To compare the empirical recovery thresholds with a theoretical benchmark, we also evaluate the degrees-of-freedom counting bound for each rank r. The red dashed curve shows the degrees-of-freedom counting bound $m = r(2n - r) - n$, where $r(2n - r)$ denoteds degrees of freedom of a rank-r general complex matrix, and the n is known diagonal entries provide additional observations. As shown in Figure 2, the empirical recovery thresholds shown by the blue curve are relatively close to the degrees-of-freedom shown by the red dashed curve, and both exhibit a similar increasing trend with the rank r.

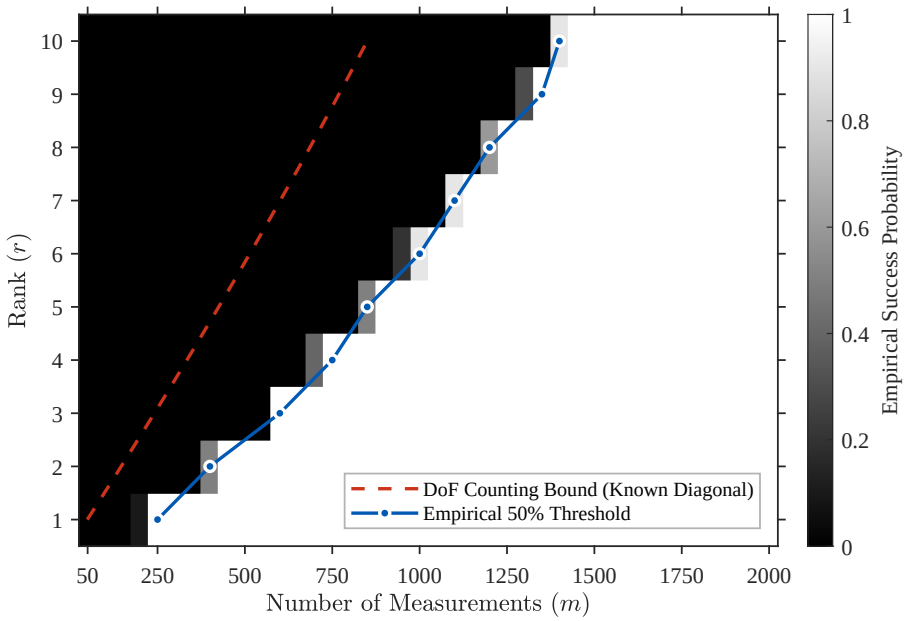


Figure 2: Empirical recovery success probabilities in the noiseless case with $n = 50$ and known diagonal entries. The red dashed curve denotes the degrees-of-freedom counting bound $m = r(2n - r) - n$. Blue curve indicates the smallest tested m attaining at least 50% empirical success for each rank.

A.2 Stability under Joint Measurement Noise

We examine the effect of noise in both the random rank-one measurements and the diagonal observations with $n = 50$, $r = 3$, and $m = 700$. The target matrices and the uniform-phase measurement vectors are generated as in Section 4, with $\|\mathbf{X}\|_F = 1$. We add independent complex-valued noise to each entry of the rank-one measurements $\mathcal{A}(\mathbf{X})$ in (1) and the diagonal observations $\mathcal{C}(\mathbf{X})$ in (8). Each noise entry is generated as $\sigma(a + ib)/\sqrt{2}$, where a and b are independently drawn from the uniform distribution on $[-1, 1]$. We consider noise levels $\sigma \in \{10^{-7}, 10^{-6}, \ldots, 10^{-2}\}$. For each

noise level, we solve (26), setting the noise upper bounds to $\varepsilon_1 = m\sigma$ and $\varepsilon_2 = \sqrt{n}\sigma$. The recovery errors are averaged over 10 trials at each noise level.

For an exactly rank-r target, Theorem 3.1 gives the bound $\|\hat{\mathbf{X}} - \mathbf{X}\|_F \leq C_2(1/\sqrt{2} + \sqrt{n})\sigma$. The blue curve in Figure 3 shows how the mean recovery error varies with the noise level σ. The red line is a fit to the mean recovery errors in log-log coordinates, with a slope of 1.032, indicating that the error increases approximately linearly with σ over the tested range. This behavior is consistent with the linear dependence on noise in the theoretical stability bound of Theorem 3.1.

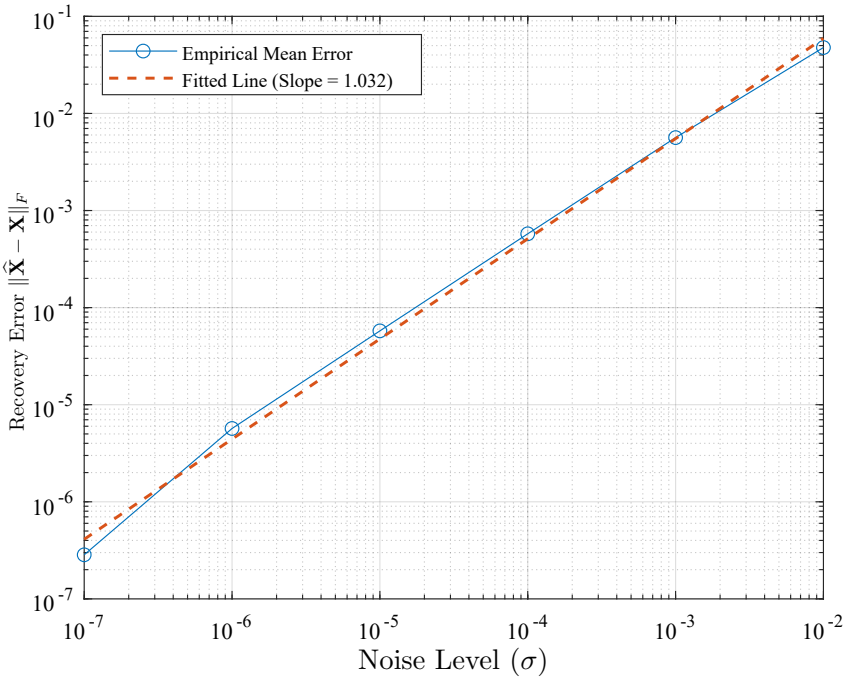


Figure 3: Mean recovery error under joint random-measurement and diagonal noise with $n = 50$, $r = 3$, and $m = 700$.

A.3 Experiments with Discrete Unit-Modulus Measurements

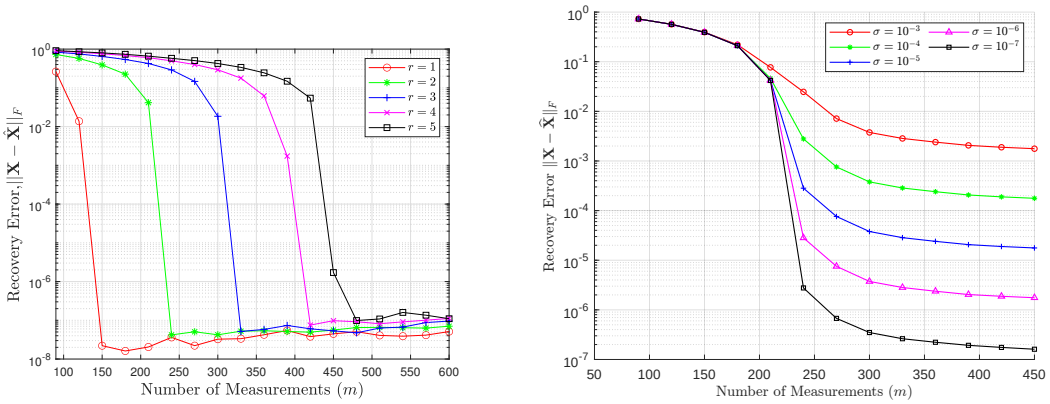
In this part, we conduct experiments for the discrete unit-modulus measurement model, where the phase of each entry is drawn uniformly from $\{0, 2\pi/3, 4\pi/3\}$. The experimental setup follows the same setting as Section 4, with matrix size 30×30 and the estimator in (26).

Figure 4a shows the matrix recovery performance in the noiseless case with $\varepsilon_1 = \varepsilon_2 = 0$, where the rank r varies from 1 to 5 and the number of measurements m varies from 90 to 600. At fixed n, the recovery curves exhibit a trend consistent with the $O(nr)$ sample complexity in Theorem 3.1.

Figure 4b shows the matrix recovery performance in the noisy case with $r = 2$, where the noise level σ varies from 10^{-7} to 10^{-3} and the number of measurements m varies from 90 to 450. The recovery error increases as the noise level σ grows, which is consistent with our theoretical result in Theorem 3.1.

A.4 Comparison with Gaussian Measurements

To further validate the effectiveness of unit-modulus measurements, we compare three types of measurement vectors: complex Gaussian vectors with i.i.d. entries, uniform unit-modulus vectors with phases drawn uniformly from $[0, 2\pi)$, and discrete unit-modulus vectors with phases drawn uniformly from $\{0, 2\pi/3, 4\pi/3\}$. The experimental setup uses rank $r = 3$ and matrix size 50×50. All three schemes use noiseless observations and exact diagonal information, and solve (26). The curves show mean recovery errors over 20 trials for each $m = 500, 510, \ldots, 650$. In this experiment, Gaussian measurements attain small recovery errors at slightly lower sample counts, while all three schemes



(a) noiseless case

(b) noisy case

Figure 4: The error of the reconstructed matrix via (26) with discrete unit-modulus measurements in (a) noiseless case and (b) noisy case

exhibit comparable recovery transitions. Unit-modulus measurements offer storage advantages: uniform unit-modulus vectors require storing only phase values rather than complex entries, and discrete unit-modulus vectors require storing only indices into a three-element set $\{0, 2\pi/3, 4\pi/3\}$, substantially reducing storage overhead compared to complex Gaussian measurements.

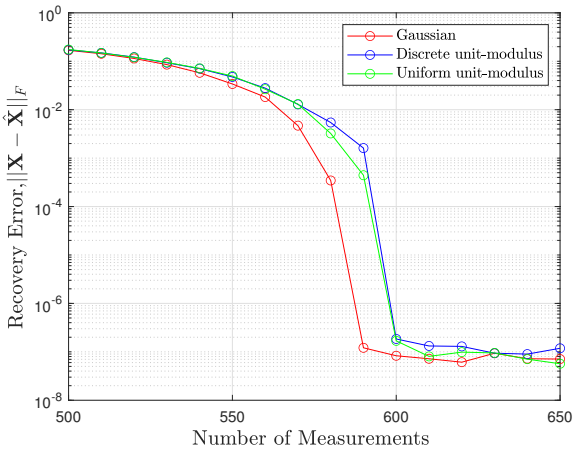


Figure 5: Comparison with Gaussian Measurements

A.5 Failure Case for Binary Unit-Modulus Measurements with General Complex Matrices

In this part, we conduct an experiment to validate the theoretical finding in Section 2.4 that binary unit-modulus measurements in (21) cannot recover general complex matrices. The measurement vectors are generated with phases drawn uniformly from $\{0, \pi\}$, corresponding to the case $c_u = 1$. The experimental setup follows the same setting as the noiseless case in Section 4, with $\mathbf{X} \in \mathbb{C}^{n\times n}$ being a general complex matrix and rank $r = 1$. As shown in Figure 6, the recovery error remains large over the tested measurement counts, consistent with the non-identifiability of general complex matrices established in Section 2.4.

A.6 Experiments with Binary Unit-Modulus Measurements for Transpose-Symmetric Complex Matrices

In this part, we validate Theorem 3.2 by conducting experiments for the binary unit-modulus measurement model with $c_u = 1$, where the phases are drawn uniformly from $\{0, \pi\}$, corresponding

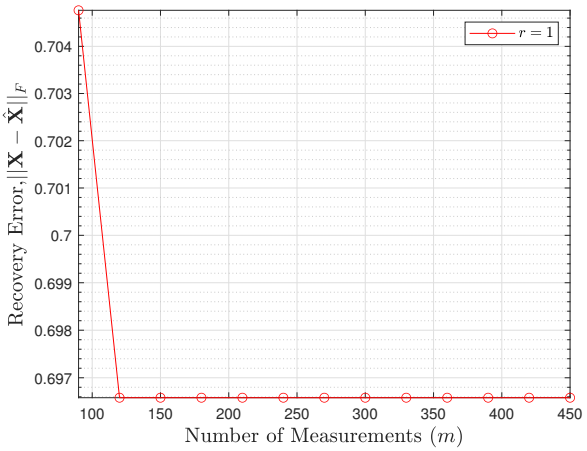
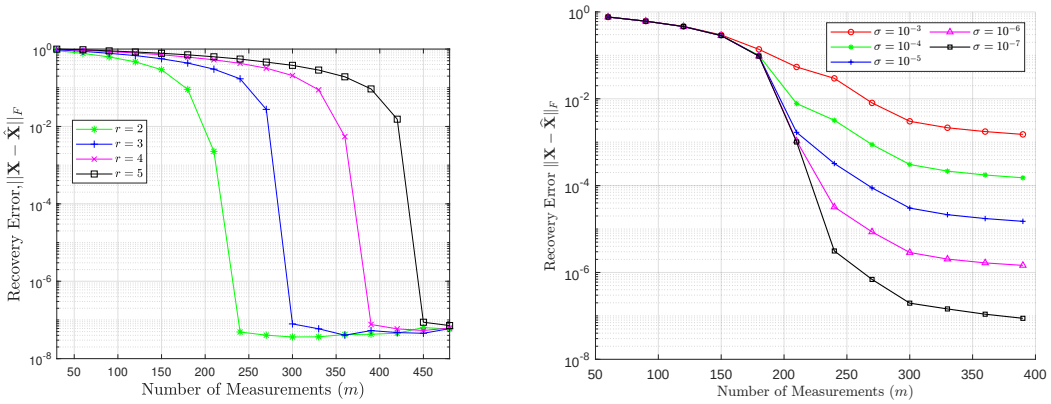


Figure 6: The error of the reconstructed matrix via (26) with binary unit-modulus measurements in noiseless case.



(a) noiseless case

(b) noisy case

Figure 7: The error of the reconstructed matrix via (30) with binary unit-modulus measurements in (a) noiseless case and (b) noisy case

to measurement vectors with entries taking values in $\{+1, -1\}$. The transpose-symmetric complex matrix $\mathbf{X} \in \mathbb{S}_C^n$ is generated via $\mathbf{X} = \mathbf{U}\mathbf{U}^T/||\mathbf{U}\mathbf{U}^T||_F$, where $\mathbf{U}$ is a randomly generated $50 \times r$ matrix with independent complex Gaussian entries. The estimator in (30) is used for recovery.

Figure 7a shows the matrix recovery performance in the noiseless case with $\varepsilon_1 = \varepsilon_2 = 0$, where the rank r varies from 2 to 5 and the number of measurements m varies from 30 to 480 with step size 30. At fixed n, the recovery curves exhibit a trend consistent with the $O(nr)$ sample complexity in Theorem 3.2.

Figure 7b shows the matrix recovery performance in the noisy case with $r = 2$, where the noise is generated by $\sigma \frac{\mathcal{U}[-1,1]+i\mathcal{U}[-1,1]}{\sqrt{2}}$, and $\mathcal{U}[-1, 1]$ denotes the uniform distribution on $[-1, 1]$. The noise level σ varies from 10^{-7} to 10^{-3} and the number of measurements m varies from 60 to 390. The recovery error increases as the noise level σ grows, which is consistent with the stable recovery guarantee in Theorem 3.2.

B Proof of Main Result

B.1 Proof of Lemma 2.1

By the unit-modulus assumption,

$$c_u = |\mathbb{E}[u_{j,k}^2]| \le \mathbb{E}[|u_{j,k}^2|] = \mathbb{E}[|u_{j,k}|^2] = 1. \tag{32}$$

Next, suppose $c_u = 1$ and set $\theta_u = \arg(\mathbb{E}[u_{j,k}^2]) \in [0, 2\pi)$. Then $\mathbb{E}[u_{j,k}^2] = e^{\mathrm{i}\theta_u}$, so

$$\mathbb{E}\left[1 - \mathrm{Re}(e^{-\mathrm{i}\theta_u} u_{j,k}^2)\right] = 0. \tag{33}$$

Since $\mathrm{Re}(e^{-\mathrm{i}\theta_u} u_{j,k}^2) \le |e^{-\mathrm{i}\theta_u} u_{j,k}^2| = 1$, (33) gives $\mathrm{Re}(e^{-\mathrm{i}\theta_u} u_{j,k}^2) = 1$. Also,

$$\left[\mathrm{Im}(e^{-\mathrm{i}\theta_u} u_{j,k}^2)\right]^2 = |e^{-\mathrm{i}\theta_u} u_{j,k}^2|^2 - \left[\mathrm{Re}(e^{-\mathrm{i}\theta_u} u_{j,k}^2)\right]^2 = 1 - 1 = 0.$$

Therefore,

$$u_{j,k}^2 = e^{\mathrm{i}\theta_u}. \tag{34}$$

Writing $u_{j,k} = e^{\mathrm{i}\theta_{j,k}}$, we therefore have

$$u_{j,k}^2 = e^{2\mathrm{i}\theta_{j,k}} = e^{\mathrm{i}\theta_u}. \tag{35}$$

Let $\theta_1, \theta_2 \in [0, 2\pi)$ be two distinct possible values of $\theta_{j,k}$, with $\theta_1 < \theta_2$. Since they have the same square by (35),

$$2\theta_2 - 2\theta_1 = 2\pi, \tag{36}$$

and hence $\theta_2 - \theta_1 = \pi$. There is no third distinct phase in $[0, 2\pi)$ with the same square. Thus, $\theta_{j,k}$ takes values in $\{\theta_1, \theta_1 + \pi\}$, and its distribution has the form

$$\theta_{j,k} = \begin{cases} \theta_1, & \text{with probability } p, \\ \theta_1 + \pi, & \text{with probability } 1 - p, \end{cases} \tag{37}$$

where $\theta_1 \in [0, \pi)$ is fixed. The zero-mean condition $\mathbb{E}[u_{j,k}] = 0$ forces $p = \frac{1}{2}$.

Thus, we have completed the proof of Lemma 2.1. . □

B.2 Proof of Lemma B.1 and Lemma 2.2

Lemma B.1. *Let $m \ge 2$, let $\mathcal{A}(\cdot)$ be sampled from the unit-modulus model with $0 \le c_u < 1$ in* (3)*, and let $\mathcal{B}(\cdot)$ be given in* (10)*. For any matrix $\mathbf{X} \in \mathbb{C}^{n\times n}$, there exist universal constants $c_1, c_2 > 0$ such that*

$$c_1 ||\mathbf{X}_o||_F \le \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 \le c_2 ||\mathbf{X}_o||_F. \tag{38}$$

Before proving Lemmas 2.2 and B.1, we establish the mean and second-moment bounds for $\mathcal{B}(\mathbf{X})$.

Lemma B.2. *Let $\mathcal{A}$ follow the unit-modulus model in* (3) *. For any fixed $\mathbf{X} \in \mathbb{C}^{n\times n}$ and $1 \le j \le \lfloor m/2 \rfloor$,*

$$\begin{aligned} \mathbb{E}\{[\mathcal{B}(\mathbf{X})]_j\} &= 0 \\ \frac{1}{m^2}(1 - c_u^2)\|\mathbf{X}_o\|_F^2 \le &\mathbb{E}\left\{\left|\left[\mathcal{B}(\mathbf{X})\right]_j\right|^2\right\} \le \frac{1}{m^2}(1 + c_u^2)\|\mathbf{X}_o\|_F^2. \end{aligned} \tag{39}$$

Proof. First, we compute the mean of the centered measurement:

$$\mathbb{E}\{[\mathcal{A}(\mathbf{X})]_j - \mathrm{tr}(\mathbf{X})\} = \mathbb{E}\left\{\sum_{p=1}^{n}\sum_{q=1}^{n} u_{j,p}^* u_{j,q} X_{p,q} - \mathrm{tr}(\mathbf{X})\right\} = 0. \tag{40}$$

Using $|u_{j,p}| = 1$, we have

$$[\mathcal{A}(\mathbf{X})]_j - \mathrm{tr}(\mathbf{X}) = \sum_{p\ne q} u_{j,p}^* u_{j,q} X_{p,q}. \tag{41}$$

Next, we compute its second moment by expanding the squared modulus:

$$\begin{aligned}
&\mathbb{E}\,|[\mathcal{A}(\mathbf{X})]_j - \operatorname{tr}(\mathbf{X})|^2 \\
&\quad= \sum_{p\neq q}\sum_{a\neq b} X_{p,q}X^*_{a,b}\mathbb{E}[u^*_{j,p}u_{j,q}u_{j,a}u^*_{j,b}] \\
&\quad= \sum_{p\neq q}|X_{p,q}|^2 + c_u^2\sum_{p\neq q}X_{p,q}X^*_{q,p} \\
&\quad= ||\mathbf{X}_o||_F^2 + c_u^2\sum_{p\neq q}X_{q,p}X^*_{p,q}.
\end{aligned} \tag{42}$$

Thus,

$$\mathbb{E}|[\mathcal{A}(\mathbf{X})]_j|^2 = |\operatorname{tr}(\mathbf{X})|^2 + ||\mathbf{X}_o||_F^2 + c_u^2\sum_{p\neq q}X_{q,p}X^*_{p,q}. \tag{43}$$

Finally, (10) gives

$$\mathbb{E}\left[\left[\mathcal{B}(\mathbf{X})\right]_j\right] = \mathbb{E}\left\{\frac{\left[\mathcal{A}(\mathbf{X})\right]_{2j-1} - \operatorname{tr}(\mathbf{X}) - \left\{\left[\mathcal{A}(\mathbf{X})\right]_{2j} - \operatorname{tr}(\mathbf{X})\right\}}{\sqrt{2}m}\right\} = 0 \tag{44}$$

and

$$\begin{aligned}
\mathbb{E}\left\{\left|\left[\mathcal{B}(\mathbf{X})\right]_j\right|^2\right\} &= \frac{\mathbb{E}\left\{\left|\left[\mathcal{A}(\mathbf{X})\right]_{2j-1} - \operatorname{tr}(\mathbf{X})\right|^2\right\} + \mathbb{E}\left\{\left|\left[\mathcal{A}(\mathbf{X})\right]_{2j} - \operatorname{tr}(\mathbf{X})\right|^2\right\}}{2m^2} \\
&= \frac{1}{m^2}\|\mathbf{X}_o\|_F^2 + \frac{1}{m^2}c_u^2\sum_{p\neq q}X_{q,p}X^*_{p,q}
\end{aligned} \tag{45}$$

To bound the cross term in (45), use

$$-(|X_{p,q}|^2 + |X_{q,p}|^2) \le X_{q,p}X^*_{p,q} + X^*_{q,p}X_{p,q} \le |X_{p,q}|^2 + |X_{q,p}|^2. \tag{46}$$

Summing over $p \neq q$ yields

$$\frac{1}{m^2}(1-c_u^2)\|\mathbf{X}_o\|_F^2 \le \mathbb{E}\left\{\left|\left[\mathcal{B}(\mathbf{X})\right]_j\right|^2\right\} \le \frac{1}{m^2}(1+c_u^2)\|\mathbf{X}_o\|_F^2. \tag{47}$$

This proves Lemma B.2. □

Lemma B.3 (Sub-exponential bound for complex unit-modulus measurements)**.** *Let $\mathcal{A}$ be sampled from the unit-modulus model in* (3)*, and let $\mathcal{B}$ be defined by* (10)*. For any fixed $\mathbf{X} \in \mathbb{C}^{n\times n}$ and $1 \le j \le \lfloor m/2 \rfloor$, there exists an absolute constant $c > 0$ such that*

$$||[\mathcal{B}(\mathbf{X})]_j||_{\psi_1} \le \frac{c}{m}||\mathbf{X}_o||_F. \tag{48}$$

Here $\mathbf{X}_o$ denotes the off-diagonal part of $\mathbf{X}$. This bound holds for all $0 \le c_u \le 1$.

Proof. For $\mathbf{u}^H\mathbf{X}_o\mathbf{u}$, let δ_q be independent Bernoulli selectors with parameter $1/2$, independent of $\mathbf{u}$, and set $S = \{q : \delta_q = 1\}$. Since $\mathbb{E}_\delta[\delta_p(1-\delta_q)] = 1/4$ for $p \neq q$,

$$\mathbf{u}^H\mathbf{X}_o\mathbf{u} = 4\mathbb{E}_\delta\sum_{p\in S,\, q\notin S}u^*_pX_{p,q}u_q. \tag{49}$$

Write $\|Z\|_{L^s} = (\mathbb{E}|Z|^s)^{1/s}$. For fixed complex coefficients a_q, the real and imaginary parts of a_qu_q are centered and bounded by $|a_q|$. Hoeffding's inequality gives

$$\left\|\sum_q a_qu_q\right\|_{L^s} \le C\sqrt{s}\left(\sum_q|a_q|^2\right)^{1/2}, \qquad s \ge 2. \tag{50}$$

Fix S and condition on $\{u_q : q \notin S\}$. Apply (50) to $\{u_p^* : p \in S\}$, then use Minkowski's inequality in $L^{s/2}$ and apply the same bound to each row:

$$\begin{aligned}\left\| \sum_{p\in S,\, q\notin S} u_p^* X_{p,q} u_q \right\|_{L^s} &\le C\sqrt{s} \left\| \left(\sum_{p\in S} \left| \sum_{q\notin S} X_{p,q} u_q \right|^2 \right)^{1/2} \right\|_{L^s} \\ &\le C\sqrt{s} \left(\sum_{p\in S} \left\| \sum_{q\notin S} X_{p,q} u_q \right\|_{L^s}^2 \right)^{1/2} \\ &\le Cs \left(\sum_{p\in S,\, q\notin S} |X_{p,q}|^2 \right)^{1/2} \le Cs \|\mathbf{X}_o\|_F.\end{aligned} \tag{51}$$

By (49) and (51) and Jensen's inequality, $\|\mathbf{u}^H \mathbf{X}_o \mathbf{u}\|_{L^s} \le Cs\|\mathbf{X}_o\|_F$. The diagonal terms cancel in (10); applying the triangle inequality to the difference of two independent copies gives

$$\left(\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^s\right)^{1/s} \le \frac{cs}{m} ||\mathbf{X}_o||_F, \qquad s \ge 2. \tag{52}$$

For $1 \le s < 2$, monotonicity of the moment norms gives, with a larger c,

$$\|[\mathcal{B}(\mathbf{X})]_j\|_{L^s} \le \|[\mathcal{B}(\mathbf{X})]_j\|_{L^2} \le \frac{c}{m}\|\mathbf{X}_o\|_F \le \frac{cs}{m}\|\mathbf{X}_o\|_F. \tag{53}$$

Taking the supremum of $s^{-1}\|[\mathcal{B}(\mathbf{X})]_j\|_{L^s}$ over $s \ge 1$ proves (48). The argument uses zero mean, unit modulus, and independence across entries; hence c is independent of c_u. Independence between an entry's real and imaginary parts is not needed. □

Proof of Lemma B.1. We follow Chen et al. [2015, Appendix A]. If $\mathbf{X}_o = 0$, then $\mathcal{B}(\mathbf{X}) = 0$. For $\mathbf{X}_o \neq 0$, Lemma B.2 and Cauchy–Schwarz give

$$\{\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|\}^2 \le \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^2 \le \frac{1}{m^2}(1 + c_u^2)||\mathbf{X}_o||_F^2. \tag{54}$$

Thus,

$$\begin{aligned}\mathbb{E}||\mathcal{B}(\mathbf{X})||_1 &= \lfloor m/2 \rfloor \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j| \\ &\le \frac{\lfloor m/2 \rfloor}{m} \sqrt{1 + c_u^2}||\mathbf{X}_o||_F \le \sqrt{1 + c_u^2}||\mathbf{X}_o||_F.\end{aligned} \tag{55}$$

For the lower bound, write $[\mathcal{B}(\mathbf{X})]_j$ as the quadratic form

$$[\mathcal{B}(\mathbf{X})]_j = \frac{1}{\sqrt{2}m} [\mathbf{u}_{2j-1}^H, \mathbf{u}_{2j}^H] \begin{bmatrix} \mathbf{X}_o & 0 \\ 0 & -\mathbf{X}_o \end{bmatrix} \begin{bmatrix} \mathbf{u}_{2j-1} \\ \mathbf{u}_{2j} \end{bmatrix}. \tag{56}$$

For all $u_{j,k}$, we have $\mathbb{E}[u_{j,k}] = 0$ and

$$||u_{j,k}||_{\psi_2} = \sup_{p \ge 1} p^{-1/2} (\mathbb{E}|u_{j,k}|^p)^{1/p} = \sup_{p \ge 1} p^{-1/2} = 1. \tag{57}$$

By Lemma B.3, $[\mathcal{B}(\mathbf{X})]_j$ is sub-exponential and satisfies

$$||[\mathcal{B}(\mathbf{X})]_j||_{\psi_1} \le \frac{c_1}{m}||\mathbf{X}_o||_F. \tag{58}$$

By the moment bound for sub-exponential random variables [Vershynin, 2010, Section 5.2.4],

$$\begin{aligned}\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^4 &\le (4||[\mathcal{B}(\mathbf{X})]_j||_{\psi_1})^4 \\ &\le \left(\frac{4c_1}{m}||\mathbf{X}_o||_F \right)^4 = \frac{c_2}{m^4}||\mathbf{X}_o||_F^4.\end{aligned} \tag{59}$$

For a nonnegative random variable ξ , two applications of Cauchy–Schwarz give

$$(\mathbb{E}[\xi^2])^2 \le \mathbb{E}[\xi]\mathbb{E}[\xi^3] \le \mathbb{E}[\xi]\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\xi^4]}, \tag{60}$$

Therefore,

$$\mathbb{E}[\xi] \geq \sqrt{\frac{(\mathbb{E}[\xi^2])^3}{\mathbb{E}[\xi^4]}}. \tag{61}$$

Setting $\xi = |[\mathcal{B}(\mathbf{X})]_j|$ and using the second- and fourth-moment bounds gives

$$\begin{aligned} \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 &= \lfloor m/2 \rfloor \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j| \\ &\geq \frac{\lfloor m/2 \rfloor}{m} \sqrt{\frac{(1-c_u^2)^3 ||\mathbf{X}_o||_F^6}{c_2 ||\mathbf{X}_o||_F^4}} \geq c_3 ||\mathbf{X}_o||_F. \end{aligned} \tag{62}$$

Since $\lfloor m/2 \rfloor / m \geq 1/3$ for $m \geq 2$, we have $c_3 > 0$ for each fixed $c_u < 1$. Combining this lower bound with (55), and setting $c_4 = \sqrt{1+c_u^2}$, gives

$$c_3 ||\mathbf{X}_o||_F \leq \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 \leq c_4 ||\mathbf{X}_o||_F. \tag{63}$$

Renaming the universal constants yields Lemma B.1. □

Proof of Lemma 2.2. If $\mathbf{X}_o = 0$, the claim follows directly. Otherwise, set $Z_j = |[\mathcal{B}(\mathbf{X})]_j| - \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|$ for $1 \leq j \leq \lfloor m/2 \rfloor$. By (58), the independent, centered variables Z_j satisfy $\|Z_j\|_{\psi_1} \leq C\|\mathbf{X}_o\|_F/m$. Bernstein's inequality [Vershynin, 2010, Proposition 5.16] gives, for $t > 0$,

$$\mathbb{P}\left\{ \left| \sum_{j=1}^{\lfloor m/2 \rfloor} Z_j \right| > t \right\} \leq 2\exp\left[-c \min\left\{ \frac{m^2 t^2}{\lfloor m/2 \rfloor \|\mathbf{X}_o\|_F^2}, \frac{mt}{\|\mathbf{X}_o\|_F} \right\} \right].$$

Taking $t = \epsilon \|\mathbf{X}_o\|_F$ and using $\lfloor m/2 \rfloor \leq m/2$ gives

$$\mathbb{P}\left\{ |\|\mathcal{B}(\mathbf{X})\|_1 - \mathbb{E}\|\mathcal{B}(\mathbf{X})\|_1| > \epsilon \|\mathbf{X}_o\|_F \right\} \leq 2\exp[-cm\min\{\epsilon^2, \epsilon\}]. \tag{64}$$

Thus, with probability at least $1 - 2\exp[-cm\min\{\epsilon^2, \epsilon\}]$,

$$|\|\mathcal{B}(\mathbf{X})\|_1 - \mathbb{E}\|\mathcal{B}(\mathbf{X})\|_1| \leq \epsilon \|\mathbf{X}_o\|_F. \tag{65}$$

Combining this with Lemma B.1 gives

$$\|\mathcal{B}(\mathbf{X})\|_1 \leq \mathbb{E}\|\mathcal{B}(\mathbf{X})\|_1 + \epsilon \|\mathbf{X}_o\|_F \leq (c_2 + \epsilon)\|\mathbf{X}_o\|_F, \tag{66}$$

and

$$\|\mathcal{B}(\mathbf{X})\|_1 \geq \mathbb{E}\|\mathcal{B}(\mathbf{X})\|_1 - \epsilon \|\mathbf{X}_o\|_F \geq (c_1 - \epsilon)\|\mathbf{X}_o\|_F. \tag{67}$$

Choosing $\epsilon = c_1/2$ therefore yields

$$\frac{c_1}{2}\|\mathbf{X}_o\|_F \leq \|\mathcal{B}(\mathbf{X})\|_1 \leq \left(c_2 + \frac{c_1}{2}\right)\|\mathbf{X}_o\|_F.$$

With $\epsilon = c_1/2$, the failure probability is at most $2\exp[-cm\min\{(c_1/2)^2, c_1/2\}]$. Choose $0 < c_3 \leq \min\{c\min\{(c_1/2)^2, c_1/2\}, \log 2\}$. Then the stated bounds hold with probability at least $1 - 2\exp(-c_3 m)$. Renaming the constants in the lower and upper bounds completes the proof. □

B.3 Proof of Corollary 2.1

For the lower bound, apply Lemma 2.2 with $c_1 \leq 1$:

$$\begin{aligned} ||\mathcal{D}(\mathbf{X})||_1 = ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_1 &\geq c_1 ||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_1 \\ &\geq c_1 ||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2 \\ &\geq c_1 ||\mathbf{X}||_F. \end{aligned} \tag{68}$$

For the upper bound, Lemma 2.2 gives

$$\begin{aligned} ||\mathcal{D}(\mathbf{X})||_1 = ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_1 &\leq c_2 ||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_1 \\ &\leq c_2 ||\mathbf{X}_o||_F + \sqrt{n}||\operatorname{diag}(\mathbf{X})||_2 \\ &\leq \sqrt{2}\max(c_2, \sqrt{n})||\mathbf{X}||_F. \end{aligned} \tag{69}$$

Combining the two bounds gives (15). □

B.4 Proof of Lemma 2.3

On the event in Lemma 2.2, the upper bound is

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq & c_2||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2 \\ \leq & C_*(||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2) \\ \leq & \sqrt{2}C_*||\mathbf{X}||_F = c_2'||\mathbf{X}||_F. \end{aligned} \tag{70}$$

where $C_* = \max\{c_2, 1\}$ and $c_2' = \sqrt{2}C_*$.

For the lower bound,

$$\|\mathcal{B}(\mathbf{X})\|_1 + \|\mathcal{C}(\mathbf{X})\|_2 \geq c_1\|\mathbf{X}_o\|_F + \|\operatorname{diag}(\mathbf{X})\|_2 \geq c_4\left(\|\mathbf{X}_o\|_F + \|\operatorname{diag}(\mathbf{X})\|_2\right) \geq c_1'\|\mathbf{X}\|_F. \tag{71}$$

where $c_1' = c_4 = \min(c_1, 1) > 0$.

Both bounds hold on the same event as in Lemma 2.2, whose probability is at least $1 - 2\exp(-c_3 m)$. Renaming the constants establishes Lemma 2.3. □

B.5 Proof of Theorem 2.1

We apply the covering argument of Candes and Plan [2011, Section 3.B] to the complex matrix setting, using Lemma 2.3.

Let $S_r = \{\mathbf{X} \in \mathbb{C}^{n\times n} : \operatorname{rank}(\mathbf{X}) \leq r, ||\mathbf{X}||_F = 1\}$. Applying the SVD covering construction of Candes and Plan [2011, Lemma 3.1] to complex matrices, there exists an ε-net $\bar{S}_r \subset S_r$ with respect to the Frobenius norm obeying

$$|\bar{S}_r| \leq (9/\varepsilon)^{(4n+1)r}. \tag{72}$$

The exponent accounts for two complex singular-vector factors and the real singular values. Set $\varepsilon = \delta/(4\sqrt{2})$, where $0 < \delta < c_1/c_2 < 1$. Then (72) gives

$$|\bar{S}_r| \leq (36\sqrt{2}/\delta)^{(4n+1)r}. \tag{73}$$

Then Lemma 2.3 and a union bound give

$$\begin{aligned} \mathbb{P}\{\exists\, \bar{\mathbf{X}} \in \bar{S}_r : ||\mathcal{B}(\bar{\mathbf{X}})||_1 + ||\mathcal{C}(\bar{\mathbf{X}})||_2 \notin [c_1, c_2]\} &\leq 2|\bar{S}_r|\exp(-c_3 m) \\ &\leq 2\exp[(4n+1)r\ln(36\sqrt{2}/\delta) - c_3 m]. \end{aligned} \tag{74}$$

If $m \geq c_4(4n+1)r$, choosing $c_4 \geq 2\ln(72\sqrt{2}/\delta)/c_3$ absorbs the factor 2 and yields a failure probability of at most $\exp(-dm)$, where $d = c_3/2 > 0$.

On this event, every net point satisfies

$$c_1 \leq ||\mathcal{B}(\bar{\mathbf{X}})||_1 + ||\mathcal{C}(\bar{\mathbf{X}})||_2 \leq c_2, \qquad \forall\, \bar{\mathbf{X}} \in \bar{S}_r. \tag{75}$$

For the upper mRIP bound, set

$$1 + \delta_r^{ub} = \sup_{\mathbf{X}\in S_r} \left(||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2\right). \tag{76}$$

For any $\mathbf{X} \in S_r$, choose $\bar{\mathbf{X}} \in \bar{S}_r$ with $||\mathbf{X} - \bar{\mathbf{X}}||_F \leq \delta/(4\sqrt{2})$. Then

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq & ||\mathcal{B}(\mathbf{X} - \bar{\mathbf{X}})||_1 + ||\mathcal{C}(\mathbf{X} - \bar{\mathbf{X}})||_2 + ||\mathcal{B}(\bar{\mathbf{X}})||_1 + ||\mathcal{C}(\bar{\mathbf{X}})||_2 \\ \leq & ||\mathcal{B}(\mathbf{X} - \bar{\mathbf{X}})||_1 + ||\mathcal{C}(\mathbf{X} - \bar{\mathbf{X}})||_2 + c_2. \end{aligned} \tag{77}$$

Set $\Delta\mathbf{X} = \mathbf{X} - \bar{\mathbf{X}}$, so $\operatorname{rank}(\Delta\mathbf{X}) \leq 2r$. Split its SVD as $\Delta\mathbf{X} = \Delta\mathbf{X}_1 + \Delta\mathbf{X}_2$, where $\langle\Delta\mathbf{X}_1, \Delta\mathbf{X}_2\rangle_F = 0$ and $\operatorname{rank}(\Delta\mathbf{X}_i) \leq r$. Each nonzero $\Delta\mathbf{X}_i/||\Delta\mathbf{X}_i||_F$ belongs to S_r. Thus

$$\begin{aligned} ||\mathcal{B}(\Delta\mathbf{X})||_1 + ||\mathcal{C}(\Delta\mathbf{X})||_2 \leq & ||\mathcal{B}(\Delta\mathbf{X}_1)||_1 + ||\mathcal{C}(\Delta\mathbf{X}_1)||_2 + ||\mathcal{B}(\Delta\mathbf{X}_2)||_1 + ||\mathcal{C}(\Delta\mathbf{X}_2)||_2 \\ \leq & (1 + \delta_r^{ub})(||\Delta\mathbf{X}_1||_F + ||\Delta\mathbf{X}_2||_F). \end{aligned} \tag{78}$$

Frobenius orthogonality gives $||\Delta\mathbf{X}_1||_F + ||\Delta\mathbf{X}_2||_F \leq \sqrt{2}||\Delta\mathbf{X}||_F$. Since $||\Delta\mathbf{X}||_F \leq \delta/(4\sqrt{2})$, the last display is at most $(1 + \delta_r^{ub})\delta/4$. Substituting this bound into (77) gives

$$||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq (1 + \delta_r^{ub})\frac{\delta}{4} + c_2. \tag{79}$$

Taking the supremum over S_r gives $(1+\delta_r^{ub}) \leq (1+\delta_r^{ub})\delta/4+c_2$, and hence $1+\delta_r^{ub} \leq c_2/(1-\frac{\delta}{4}) \leq \frac{4}{3}c_2$. The lower bound follows from

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \geq & ||\mathcal{B}(\bar{\mathbf{X}})||_1 + ||\mathcal{C}(\bar{\mathbf{X}})||_2 - (||\mathcal{B}(\Delta\mathbf{X})||_1 + ||\mathcal{C}(\Delta\mathbf{X})||_2) \\ \geq & c_1 - \frac{4}{3}c_2\frac{\sqrt{2}\delta}{4\sqrt{2}} \\ \geq & c_1 - c_2\frac{\delta}{3} \\ \geq & \frac{2c_1}{3}. \end{aligned} \tag{80}$$

Since this holds for all $\mathbf{X} \in S_r$, we have $(1-\delta_r^{lb}) = \inf_{\mathbf{X}\in S_r}\left(||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2\right) \geq \frac{2c_1}{3}$.

Combining the bounds gives

$$1-\delta_r^{lb} \geq \frac{2}{3}c_1, 1+\delta_r^{ub} \leq \frac{4}{3}c_2, \tag{81}$$

with probability exceeding $1-\exp(-dm)$ provided that $m > c_4 nr$. Renaming some constants, we obtain Theorem 2.1. □

B.6 Proof of Proposition 2.1

Since $\mathbf{u} = e^{\mathrm{i}\theta}\mathbf{s}$, we have

$$\begin{aligned} \mathbf{u}^H\mathbf{X}\mathbf{u} &= (e^{\mathrm{i}\theta}\mathbf{s})^H\mathbf{X}(e^{\mathrm{i}\theta}\mathbf{s}) \\ &= e^{-\mathrm{i}\theta}e^{\mathrm{i}\theta}\mathbf{s}^T\mathbf{X}\mathbf{s} \\ &= \mathbf{s}^T\mathbf{X}\mathbf{s}. \end{aligned} \tag{82}$$

The global phase therefore cancels, leaving $\mathbf{s}^T\mathbf{X}\mathbf{s}$.

Next, decompose $\mathbf{X}$ into its transpose-symmetric and transpose-skew-symmetric parts:

$$\mathbf{X} = \frac{\mathbf{X}+\mathbf{X}^T}{2} + \frac{\mathbf{X}-\mathbf{X}^T}{2}. \tag{83}$$

Let

$$\mathbf{K} = \frac{\mathbf{X}-\mathbf{X}^T}{2}. \tag{84}$$

Then $\mathbf{K}^T = -\mathbf{K}$. Since $\mathbf{s}^T\mathbf{K}\mathbf{s}$ is a scalar, it equals its transpose, and hence

$$\begin{aligned} \mathbf{s}^T\mathbf{K}\mathbf{s} = (\mathbf{s}^T\mathbf{K}\mathbf{s})^T &= \mathbf{s}^T\mathbf{K}^T\mathbf{s} \\ &= -\mathbf{s}^T\mathbf{K}\mathbf{s}, \end{aligned} \tag{85}$$

which yields

$$\mathbf{s}^T\mathbf{K}\mathbf{s} = 0. \tag{86}$$

It follows from (83) that

$$\mathbf{s}^T\mathbf{X}\mathbf{s} = \mathbf{s}^T\left(\frac{\mathbf{X}+\mathbf{X}^T}{2}\right)\mathbf{s}. \tag{87}$$

Finally, replacing $\mathbf{X}$ by any $\mathbf{K}$ with $\mathbf{K}^T = -\mathbf{K}$ in (82) gives

$$\mathbf{u}^H\mathbf{K}\mathbf{u} = \mathbf{s}^T\mathbf{K}\mathbf{s} = 0. \tag{88}$$

Thus, every transpose-skew-symmetric matrix lies in the kernel of $\mathbf{X} \mapsto \mathbf{u}^H\mathbf{X}\mathbf{u}$, and $\mathbf{X}$ and $\mathbf{X}+\mathbf{K}$ give the same observations. For $n \geq 2$, nonzero such $\mathbf{K}$ exist, so a general $\mathbf{X} \in \mathbb{C}^{n\times n}$ is not uniquely identifiable . □

B.7 Proof of Theorem 2.2

We follow the proof of Theorem 2.1 on the transpose-symmetric subspace. We first establish moment and concentration bounds, then apply the covering argument.

First, we compute the second moment of $[\mathcal{B}(\mathbf{X})]_j$.

Lemma B.4. *Let $\mathbf{X} \in \mathbb{S}_C^n$, and let $\mathcal{A}$ follow the unit-modulus model in* (3) *with $c_u = 1$. Then, for every $1 \le j \le \lfloor m/2 \rfloor$,*

$$\begin{aligned} &\mathbb{E}\{[\mathcal{B}(\mathbf{X})]_j\} = 0 \\ &\mathbb{E}\left\{\left|\left[\mathcal{B}(\mathbf{X})\right]_j\right|^2\right\} = \frac{2}{m^2}\|\mathbf{X}_o\|_F^2. \end{aligned} \tag{89}$$

Proof. (44) gives

$$\mathbb{E}\left\{[\mathcal{B}(\mathbf{X})]_j\right\} = 0. \tag{90}$$

Since $\mathbf{X}^T = \mathbf{X}$ and $c_u = 1$, (45) gives

$$\begin{aligned} \mathbb{E}\left\{\left|\left[\mathcal{B}(\mathbf{X})\right]_j\right|^2\right\} &= \frac{1}{m^2}\|\mathbf{X}_o\|_F^2 + \frac{c_u^2}{m^2}\sum_{p\neq q} X_{q,p}X_{p,q}^* \\ &= \frac{1}{m^2}\|\mathbf{X}_o\|_F^2 + \frac{1}{m^2}\times\sum_{p\neq q}|X_{p,q}|^2 \\ &= \frac{2}{m^2}\|\mathbf{X}_o\|_F^2. \end{aligned} \tag{91}$$

□

The next lemma bounds $\mathbb{E}\|\mathcal{B}(\mathbf{X})\|_1$.

Lemma B.5. *Let $m \ge 2$, let $\mathcal{A}(\cdot)$ be sampled from the unit-modulus model with $c_u = 1$ in* (3) *, and let $\mathcal{B}(\cdot)$ be given in* (10)*. For any matrix $\mathbf{X} \in \mathbb{S}_C^n$, there exists a universal constant $c_1 > 0$ such that*

$$c_1||\mathbf{X}_o||_F \le \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 \le \sqrt{2}||\mathbf{X}_o||_F. \tag{92}$$

Proof. By the preceding second-moment identity and Cauchy–Schwarz,

$$\begin{aligned} \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 &= \lfloor m/2\rfloor \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j| \\ &\le \lfloor m/2\rfloor\sqrt{\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^2} = \frac{\sqrt{2}\lfloor m/2\rfloor}{m}||\mathbf{X}_o||_F \le \sqrt{2}||\mathbf{X}_o||_F. \end{aligned} \tag{93}$$

For the lower bound, (58) gives

$$\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^4 \le \frac{c_2}{m^4}||\mathbf{X}_o||_F^4. \tag{94}$$

For $\mathbf{X}_o \neq 0$, the moment inequality used in Lemma B.1 gives

$$\begin{aligned} \mathbb{E}||\mathcal{B}(\mathbf{X})||_1 &\ge \lfloor m/2\rfloor\sqrt{\frac{(\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^2)^3}{\mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|^4}} \\ &\ge \frac{2\sqrt{2}\lfloor m/2\rfloor}{\sqrt{c_2}m}||\mathbf{X}_o||_F \ge c_1||\mathbf{X}_o||_F, \end{aligned} \tag{95}$$

Here $\lfloor m/2\rfloor/m \ge 1/3$ for $m \ge 2$. If $\mathbf{X}_o = 0$, both sides vanish. □

Next, we obtain a fixed-matrix concentration bound.

Lemma B.6. *Let $\mathcal{A}(\cdot)$ be sampled from the unit-modulus model with $c_u = 1$ in* (3) *, and let $\mathcal{B}(\cdot)$ be given by* (10)*. For any fixed $\mathbf{X} \in \mathbb{S}_C^n$, there exist universal constants $c_1, c_2, c_3 > 0$ and $m_0 \ge 2$ such that, for $m \ge m_0$, with probability at least $1 - \exp(-c_3 m)$, we have*

$$c_1||\mathbf{X}_o||_F \le ||\mathcal{B}(\mathbf{X})||_1 \le c_2||\mathbf{X}_o||_F. \tag{96}$$

Proof. Let $Z_j = |[\mathcal{B}(\mathbf{X})]_j| - \mathbb{E}|[\mathcal{B}(\mathbf{X})]_j|$ for $1 \le j \le \lfloor m/2 \rfloor$. As in the proof of Lemma 2.2, these variables are independent and satisfy

$$\mathbb{E}Z_j = 0, \qquad ||Z_j||_{\psi_1} \le \frac{c_4}{m}||\mathbf{X}_o||_F. \tag{97}$$

Bernstein's inequality gives, for every $\epsilon > 0$,

$$|\,||\mathcal{B}(\mathbf{X})||_1 - \mathbb{E}||\mathcal{B}(\mathbf{X})||_1\,| \le \epsilon||\mathbf{X}_o||_F \tag{98}$$

with probability at least $1 - 2\exp[-c_5 m \min\{\epsilon^2, \epsilon\}]$. Combining this with the expectation bounds gives

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 &\le (\sqrt{2} + \epsilon)||\mathbf{X}_o||_F, \\ ||\mathcal{B}(\mathbf{X})||_1 &\ge (c_1 - \epsilon)||\mathbf{X}_o||_F. \end{aligned} \tag{99}$$

Set $\epsilon = c_1/2$ and rename the constants to obtain the result for sufficiently large m. □

Finally, we add the diagonal term.

Lemma B.7. *Let $\mathcal{A}(\cdot)$ follow the unit-modulus model with $c_u = 1$ in* (3)*, and let $\mathcal{C}(\cdot)$ be the diagonal sampling model in* (8)*. For any fixed $\mathbf{X} \in \mathbb{S}_C^n$, there exist universal constants $c_1, c_2, c_3 > 0$ and $m_0 \ge 2$ such that, for $m \ge m_0$, with probability at least $1 - \exp(-c_3 m)$, we have*

$$c_1||\mathbf{X}||_F \le ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \le c_2||\mathbf{X}||_F.$$

Proof. By the preceding concentration bound and $||\mathcal{C}(\mathbf{X})||_2 = ||\operatorname{diag}(\mathbf{X})||_2$, the upper bound is

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 &\le c_2||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2 \\ &\le \max(c_2, 1)(||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2) \\ &\le \sqrt{2}\max(c_2, 1)||\mathbf{X}||_F. \end{aligned} \tag{100}$$

For the lower bound,

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 &\ge c_1||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2 \\ &\ge \min(c_1, 1)(||\mathbf{X}_o||_F + ||\operatorname{diag}(\mathbf{X})||_2) \\ &\ge \min(c_1, 1)||\mathbf{X}||_F. \end{aligned} \tag{101}$$

Renaming the constants proves the lemma. □

Proof of Theorem 2.2. We follow the covering argument in Appendix B.5 . Let

$$S_r = \{\mathbf{X} \in \mathbb{S}_C^n : \operatorname{rank}(\mathbf{X}) \le r,\ ||\mathbf{X}||_F = 1\}. \tag{102}$$

An $\varepsilon/2$-net for the ambient complex rank-r set gives an ε-net $\bar{S}_r \subset S_r$: retain the balls intersecting S_r and replace each center by a point of the intersection. For any matrix in S_r, its distance to the new center is at most ε by the triangle inequality. The number of centers does not increase, so

$$|\bar{S}_r| \le (18/\varepsilon)^{(4n+1)r}. \tag{103}$$

Take $\varepsilon = \delta/(4\sqrt{2})$, where $0 < \delta < c_1/c_2 < 1$ and c_1, c_2 are from Lemma B.7. By a union bound,

$$\begin{aligned} \mathbb{P}\big\{\exists\, \bar{\mathbf{X}} \in \bar{S}_r : ||\mathcal{B}(\bar{\mathbf{X}})||_1 + ||\mathcal{C}(\bar{\mathbf{X}})||_2 \notin [c_1, c_2]\big\} &\le |\bar{S}_r|\exp(-c_3 m) \\ &\le \exp\big[(4n+1)r\ln(72\sqrt{2}/\delta) - c_3 m\big]. \end{aligned} \tag{104}$$

Thus, for $m \ge c_4(4n+1)r$ with $c_4 \ge 2\ln(72\sqrt{2}/\delta)/c_3$ also large enough to ensure $m \ge m_0$, all net points satisfy the bounds with probability at least $1 - \exp(-c_3 m/2)$.

On this event, set

$$1 + \delta_r^{ub} = \sup_{\mathbf{X} \in S_r} \big(||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2\big). \tag{105}$$

For any $\mathbf{X} \in S_r$, choose $\bar{\mathbf{X}} \in \bar{S}_r$ with $||\mathbf{X} - \bar{\mathbf{X}}||_F \le \delta/(4\sqrt{2})$. Put $\Delta\mathbf{X} = \mathbf{X} - \bar{\mathbf{X}}$. Since $\Delta\mathbf{X}$ is transpose-symmetric and has rank at most $2r$, splitting its Takagi factorization yields

$$\Delta\mathbf{X} = \Delta\mathbf{X}_1 + \Delta\mathbf{X}_2, \quad \langle\Delta\mathbf{X}_1, \Delta\mathbf{X}_2\rangle = 0, \quad \operatorname{rank}(\Delta\mathbf{X}_i) \le r, \tag{106}$$

where each $\Delta\mathbf{X}_i \in \mathbb{S}_C^n$. Consequently,

$$\begin{aligned} ||\mathcal{B}(\Delta\mathbf{X})||_1 + ||\mathcal{C}(\Delta\mathbf{X})||_2 &\leq (1+\delta_r^{ub})(||\Delta\mathbf{X}_1||_F + ||\Delta\mathbf{X}_2||_F) \\ &\leq \sqrt{2}(1+\delta_r^{ub})||\Delta\mathbf{X}||_F \\ &\leq (1+\delta_r^{ub})\delta/4. \end{aligned} \tag{107}$$

For the upper bound, the triangle inequality gives

$$||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 \leq c_2 + (1+\delta_r^{ub})\delta/4. \tag{108}$$

Taking the supremum over S_r yields $1+\delta_r^{ub} \leq c_2/(1-\delta/4) \leq 4c_2/3$. For the lower bound,

$$\begin{aligned} ||\mathcal{B}(\mathbf{X})||_1 + ||\mathcal{C}(\mathbf{X})||_2 &\geq c_1 - (1+\delta_r^{ub})\delta/4 \\ &\geq c_1 - c_2\delta/3 \geq 2c_1/3. \end{aligned} \tag{109}$$

Therefore,

$$1-\delta_r^{lb} \geq \frac{2}{3}c_1, \qquad 1+\delta_r^{ub} \leq \frac{4}{3}c_2. \tag{110}$$

Renaming the sampling and probability constants proves Theorem 2.2. □

B.8 Proof of Lemma 3.1

We adapt the nuclear-norm decomposition of Chen et al. [2015] to the two measurement constraints. Write $\mathbf{X} = \mathbf{X}_r + \mathbf{X}_c$ and $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{H}$. Let $\mathbf{X}_r = \mathbf{U}\Lambda\mathbf{V}^H$ be a truncated SVD, where $\mathbf{U}$ and $\mathbf{V}$ have r orthonormal columns, and define

$$T = \{\mathbf{U}\mathbf{M}_1 + \mathbf{M}_2\mathbf{V}^H : \mathbf{M}_1 \in \mathbb{C}^{r\times n},\ \mathbf{M}_2 \in \mathbb{C}^{n\times r}\}.$$

Set $\mathbf{H}_T = \mathcal{P}_T\mathbf{H}$ and $\mathbf{H}_{T^\perp} = (I - \mathbf{U}\mathbf{U}^H)\mathbf{H}(I - \mathbf{V}\mathbf{V}^H)$, so $\operatorname{rank}(\mathbf{H}_T) \leq 2r$. Partition the singular values of $\mathbf{H}_{T^\perp}$, in decreasing order, into consecutive blocks of at most K_1 values. This gives mutually Frobenius-orthogonal matrices $\mathbf{H}_1, \ldots, \mathbf{H}_M$, each of rank at most K_1 and in $T^\perp$. Take empty blocks to be zero.

Lemma B.8. *With the above notation,*

$$\|\mathbf{H}_{T^\perp}\|_* \leq \|\mathbf{H}_T\|_* + 2\|\mathbf{X}_c\|_*, \tag{111}$$

and

$$\sum_{j\geq 2}\|\mathbf{H}_j\|_F \leq \frac{\|\mathbf{H}_{T^\perp}\|_*}{\sqrt{K_1}} \leq \sqrt{\frac{2r}{K_1}}\|\mathbf{H}_T\|_F + \frac{2\|\mathbf{X}_c\|_*}{\sqrt{K_1}}. \tag{112}$$

Proof. Since $\mathbf{X}$ is feasible, nuclear-norm optimality gives $\|\mathbf{X}+\mathbf{H}\|_* \leq \|\mathbf{X}\|_*$. The row and column spaces of $\mathbf{X}_r$ and $\mathbf{H}_{T^\perp}$ are orthogonal. Thus,

$$\begin{aligned} \|\mathbf{X}_r\|_* + \|\mathbf{X}_c\|_* = \|\mathbf{X}\|_* &\geq \|\mathbf{X}+\mathbf{H}\|_* \\ &\geq \|\mathbf{X}_r + \mathbf{H}_{T^\perp}\|_* - \|\mathbf{X}_c + \mathbf{H}_T\|_* \\ &\geq \|\mathbf{X}_r\|_* + \|\mathbf{H}_{T^\perp}\|_* - \|\mathbf{X}_c\|_* - \|\mathbf{H}_T\|_*. \end{aligned}$$

Rearranging proves (111). For $j \geq 2$, every singular value in block j is at most the average singular value in the preceding full block. Hence $\|\mathbf{H}_j\|_F \leq \|\mathbf{H}_{j-1}\|_*/\sqrt{K_1}$. Sum this inequality and use (111) and $\|\mathbf{H}_T\|_* \leq \sqrt{2r}\|\mathbf{H}_T\|_F$ to obtain (112) . □

Proof of Lemma 3.1. Feasibility of $\mathbf{X}$ and $\hat{\mathbf{X}}$ gives

$$\|\mathcal{A}(\mathbf{H})\|_1 \leq 2\varepsilon_1, \qquad \|\mathcal{C}(\mathbf{H})\|_2 \leq 2\varepsilon_2.$$

By (10),

$$\|\mathcal{B}(\mathbf{H})\|_1 \leq \frac{1}{\sqrt{2}m}\|\mathcal{A}(\mathbf{H})\|_1 \leq \frac{\sqrt{2}\varepsilon_1}{m}.$$

By the triangle inequality and mRIP,

$$\begin{aligned}\frac{\sqrt{2}\varepsilon_1}{m} + 2\varepsilon_2 &\geq \|\mathcal{B}(\mathbf{H})\|_1 + \|\mathcal{C}(\mathbf{H})\|_2 \\ &\geq (1-\delta^{lb}_{2r+K_1})\|\mathbf{H}_T + \mathbf{H}_1\|_F - (1+\delta^{ub}_{K_1})\sum_{j\geq 2}\|\mathbf{H}_j\|_F \\ &\geq \frac{1-\delta^{lb}_{2r+K_1}}{\sqrt{2}}(\|\mathbf{H}_T\|_F + \|\mathbf{H}_1\|_F) \\ &\quad - (1+\delta^{ub}_{K_1})\left(\sqrt{\frac{2r}{K_1}}\|\mathbf{H}_T\|_F + \frac{2\|\mathbf{X}_c\|_*}{\sqrt{K_1}}\right).\end{aligned} \tag{113}$$

Here $\operatorname{rank}(\mathbf{H}_T + \mathbf{H}_1) \leq 2r + K_1$, and Frobenius orthogonality gives $\|\mathbf{H}_T + \mathbf{H}_1\|_F \geq (\|\mathbf{H}_T\|_F + \|\mathbf{H}_1\|_F)/\sqrt{2}$. Hence (27) gives

$$\begin{aligned}\|\mathbf{H}_T\|_F + \|\mathbf{H}_1\|_F \leq \frac{2(1+\delta^{ub}_{K_1})}{\beta_1}\frac{\|\mathbf{X}_c\|_*}{\sqrt{K_1}} \\ + \frac{2}{\beta_1}\left(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2\right).\end{aligned} \tag{114}$$

Finally, combining (112) and (114) with $K_1 > 2r$ gives

$$\begin{aligned}\|\mathbf{H}\|_F &\leq \|\mathbf{H}_T\|_F + \|\mathbf{H}_1\|_F + \sum_{j\geq 2}\|\mathbf{H}_j\|_F \\ &\leq 2(\|\mathbf{H}_T\|_F + \|\mathbf{H}_1\|_F) + \frac{2\|\mathbf{X}_c\|_*}{\sqrt{K_1}} \\ &\leq \left(\frac{4(1+\delta^{ub}_{K_1})}{\beta_1} + 2\right)\frac{\|\mathbf{X}_c\|_*}{\sqrt{K_1}} \\ &\quad + \frac{4}{\beta_1}\left(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2\right).\end{aligned} \tag{115}$$

This proves the lemma with $C_1 = 4(1+\delta^{ub}_{K_1})$, $C_2 = 4$, and $C_3 = 2$. □

B.9 Proof of Theorem 3.1

Let c_1, c_2 be the constants in Theorem 2.1, and choose the integer

$$K_1 = \left\lceil 32\frac{c_2^2}{c_1^2}r \right\rceil > 2r.$$

By Definition 2.1, rank bounds above n are equivalent to rank n. Applying Theorem 2.1 at ranks $\min\{K_1, n\}$ and $\min\{2r + K_1, n\}$ gives

$$1 - \delta^{lb}_{2r+K_1} \geq \frac{2c_1}{3}, \qquad 1 + \delta^{ub}_{K_1} \leq \frac{4c_2}{3}$$

with probability at least $1 - 2\exp(-c_3 m)$ whenever $m > c_4 n(2r + K_1)$. By the choice of K_1 ,

$$\frac{1-\delta^{lb}_{2r+K_1}}{\sqrt{2}} - (1+\delta^{ub}_{K_1})\sqrt{\frac{2r}{K_1}} \geq \frac{2c_1}{3}\left(\frac{1}{\sqrt{2}} - \frac{1}{2}\right) =: \beta_1 > 0.$$

Thus Lemma 3.1 applies. Since K_1 is bounded above and below by constant multiples of r, we obtain

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_F \leq C_1\frac{\|\mathbf{X}_c\|_*}{\sqrt{r}} + C_2\left(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2\right).$$

Absorbing the fixed ratio K_1/r into the sampling constant yields $m > c_1 nr$; renaming c_3 as c_0 proves Theorem 3.1. □

B.10 Proof of Theorem 3.2

We use the error decomposition in Lemma 3.1. The Takagi decomposition keeps every error block in $\mathbb{S}_C^n$.

Lemma B.9. *Assume $||\mathbf{z}||_1 \le \varepsilon_1$ and $||\mathbf{z}_d||_2 \le \varepsilon_2$. For a given integer $1 \le r \le n$, consider any matrix $\mathbf{X} = \mathbf{X}_r + \mathbf{X}_c \in \mathbb{S}_C^n$, where $\mathbf{X}_r$ is the best rank-r approximation of $\mathbf{X}$, and $\delta_{2r+K_1}^{lb}, \delta_{K_1}^{ub}$ are the mRIP constants for $\mathcal{D}(\cdot) : \mathbb{S}_C^n \to \mathbb{C}^{\lfloor \frac{m}{2} \rfloor + n}$ in* (11)*. If there exists an integer $K_1 > 2r$ such that*

$$\frac{1-\delta_{2r+K_1}^{lb}}{\sqrt{2}} - (1+\delta_{K_1}^{ub})\sqrt{\frac{2r}{K_1}} \ge \beta_1 > 0, \tag{116}$$

holds for some numerical value β_1, the minimizer $\hat{\mathbf{X}}$ to (30) *obeys*

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \le (\frac{C_1}{\beta_1} + C_3)\frac{||\mathbf{X}_c||_*}{\sqrt{K_1}} + \frac{C_2}{\beta_1}(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2), \tag{117}$$

where $C_1, C_2, C_3 > 0$ are universal constants depending only on $\delta_{2r+K_1}^{lb}, \delta_{K_1}^{ub}$.

Proof. Write $\mathbf{X} = \mathbf{X}_r + \mathbf{X}_c$ and $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{H}$. By the Takagi factorization,

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{U}^T, \qquad \mathbf{X}_r = \mathbf{U}_r\Sigma_r\mathbf{U}_r^T. \tag{118}$$

The error components are

$$\mathbf{H}_{T^\perp} = (I - \mathbf{U}_r\mathbf{U}_r^H)\mathbf{H}(I - \mathbf{U}_r\mathbf{U}_r^H)^T, \qquad \mathbf{H}_T = \mathbf{H} - \mathbf{H}_{T^\perp}. \tag{119}$$

Both components are transpose-symmetric. Split a Takagi factorization of $\mathbf{H}_{T^\perp}$ into consecutive blocks of at most K_1 singular values to obtain transpose-symmetric matrices $\mathbf{H}_j$. Thus all matrices in the mRIP inequalities belong to $\mathbb{S}_C^n$, and (111) and (112) apply. Feasibility gives

$$||\mathcal{B}(\mathbf{H})||_1 + ||\mathcal{C}(\mathbf{H})||_2 \le \frac{\sqrt{2}\varepsilon_1}{m} + 2\varepsilon_2. \tag{120}$$

Applying the mRIP and block estimates in (114) gives

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \le \left(\frac{C_1}{\beta_1} + C_3\right)\frac{||\mathbf{X}_c||_*}{\sqrt{K_1}} + \frac{C_2}{\beta_1}\left(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2\right). \tag{121}$$

This proves the lemma. □

Proof of Theorem 3.2. As in Appendix B.9, let c_1, c_2 be the constants in Theorem 2.2 and choose

$$K_1 = \left\lceil 32\frac{c_2^2}{c_1^2}r \right\rceil. \tag{122}$$

By the convention in Definition 2.1, we apply Theorem 2.2 at ranks $\min\{K_1, n\}$ and $\min\{2r+K_1, n\}$. If $m > c_4 n(2r + K_1)$, the following bounds hold with probability at least $1 - 2\exp(-c_3 m)$:

$$1 - \delta_{2r+K_1}^{lb} \ge \frac{2c_1}{3}, \qquad 1 + \delta_{K_1}^{ub} \le \frac{4c_2}{3}. \tag{123}$$

Hence

$$\frac{1-\delta_{2r+K_1}^{lb}}{\sqrt{2}} - (1+\delta_{K_1}^{ub})\sqrt{\frac{2r}{K_1}} \ge \frac{2c_1}{3}\left(\frac{1}{\sqrt{2}} - \frac{1}{2}\right) =: \beta_1 > 0. \tag{124}$$

Applying Lemma B.9 and using $K_1 = O(r)$ gives

$$||\hat{\mathbf{X}} - \mathbf{X}||_F \le C_1\frac{||\mathbf{X}_c||_*}{\sqrt{r}} + C_2\left(\frac{\varepsilon_1}{\sqrt{2}m} + \varepsilon_2\right). \tag{125}$$

Renaming the constants gives $m > c_1 nr$ and probability $1 - 2\exp(-c_0 m)$, as stated in Theorem 3.2. □

C Limitations

The current framework requires the diagonal entries of $\mathbf{X}$ to be known or directly measurable. Acquiring them may incur additional hardware or measurement costs and may not be feasible in all applications. Additionally, the nuclear norm minimization estimator in (26) and (30) can be computationally expensive for large matrices, limiting scalability .

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the three main contributions: (1) identification of the diagonal ambiguity of symmetric unit-modulus measurements, (2) establishment of the mRIP for the stacked operator $\mathcal{D}(\cdot)$, and (3) exact and stable recovery guarantees in both noiseless and noisy settings. These claims are fully supported by Theorems 2.1, 2.2, 3.1 and 3.2, and validated by numerical experiments in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are discussed in Appendix C, including the requirement that diagonal entries of $\mathbf{X}$ be known or directly measurable, and the computational cost of the nuclear norm minimization estimators in (26) and (30), which limits scalability to large-scale problems.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The theoretical results, including the supporting lemmas and Theorems 2.1, 2.2, 3.1 and 3.2, are stated with explicit assumptions and proved in Appendix B. The assumptions on the unit-modulus measurement model are clearly specified in (3), and the intermediate results used in the proofs are stated or cross-referenced.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 4 and Appendix A describe the experimental setup, including matrix size (30×30 or 50×50), rank r, measurement vector generation, noise model, number of repetitions, and the estimators in (26) and (30). The main experiments and Gaussian comparison use 20 trials per parameter setting, while the additional phase-transition and joint-noise experiments use 10. The convex programs are implemented in MATLAB using CVX and SDPT3. These details support independent reproduction of the reported recovery trends.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The experimental code is publicly available at `https://github.com/tongyuz941/Matrix-Recovery-Experiments`. The repository provides MATLAB scripts, experiment configurations, dependency requirements, and instructions for running the experiments and plotting the results. All data are generated synthetically by the provided code, so no external dataset is required. The documentation describes the scope of reproducibility and numerical solver considerations.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental details are provided in Section 4 and Appendix A, including matrix dimensions, rank settings, measurement vector distributions, noise levels and constraint radii, number of trials, and the estimators used. The convex programs are modeled in CVX in MATLAB and solved using SDPT3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The paper reports average recovery errors and empirical success probabilities over repeated trials: 20 trials per parameter setting for the main experiments and Gaussian comparison, and 10 for the additional phase-transition and joint-noise experiments. It does not report error bars, confidence intervals, or formal statistical significance tests. The observed recovery transitions and noise-dependent error trends are consistent with the theoretical guarantees. The main recovery guarantees are established analytically and do not depend on statistical significance tests; the experiments provide supporting numerical evidence.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The experiments involve small-scale synthetic matrices (at most 50×50) solved via standard convex optimization, and do not require significant computational resources. No GPU or large-scale cluster was needed.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: This paper studies a theoretical problem in matrix recovery and does not involve human subjects, sensitive data, or harmful applications. The research conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This is a foundational theoretical work on matrix recovery with unit-modulus measurements. The results provide theoretical insight into sensing with phase-only measurement vectors, motivated by hardware-constrained systems such as coded diffraction and RIS-aided communications. The paper does not deploy such systems or introduce new applications with direct, identified negative societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release any pretrained models, datasets scraped from the Internet, or other assets that pose risks of misuse.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use external datasets or pretrained models. All experiments are conducted on synthetically generated data using standard optimization software (MATLAB/CVX with SDPT3), rather than third-party research implementations.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We release the experimental code at `https://github.com/tongyuz941/Matrix-Recovery-Experiments`, together with documentation describing dependencies, experiment configurations, execution and plotting instructions, and reproducibility limitations. All experimental data are generated synthetically by the code.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human subjects and therefore does not require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not used as part of the core methodology: they are not components of the measurement model, recovery estimators, or experimental methods.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.