MEASUREMENT MANIPULATION OF THE MATRIX SENSING PROBLEM TO IMPROVE OPTIMIZATION LANDSCAPE

Anonymous authors
Paper under double-blind review

ABSTRACT

This work studies the matrix sensing (MS) problem through the lens of the Restricted Isometry Property (RIP). It has been shown in several recent papers that two different techniques of convex relaxations and local search methods for the MS problem both require the RIP constant to be less than 0.5 while most real-world problems have their RIPs close to 1. The existing literature guarantees a small RIP constant only for sensing operators having an i.i.d. Gaussian distribution, and it is well-known that the MS problem could have a complicated landscape when the RIP is greater than 0.5. In this work, we address this issue and improve the optimization landscape by developing two results. First, we show that any sensing operator with a model not too distant from i.i.d. Gaussian has a slightly higher RIP than i.i.d. Gaussian, and that its RIP constant can be reduced to match the RIP constant of an i.i.d. Gaussian via slightly increasing the number of measurements. Second, we show that if the sensing operator has an arbitrary distribution, it can be modified in such a way that the resulting operator will act as a perturbed Gaussian with a lower RIP constant. Our approach is a preconditioning technique that replaces each sensing matrix with a weighted sum of all sensing matrices. We numerically demonstrate that the RIP constants for different distributions can be reduced from almost 1 to less than 0.5 via the preconditioning of the sensing operator.

1 INTRODUCTION

In this paper, we focus on an important class of problems in non-convex optimization and machine learning, named matrix sensing. The goal of the matrix sensing problem is to recover a low-rank matrix from a set of limited linear measurements. To be more specific, given m sensing matrices $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$, we define the linear sensing operator $\mathcal{A}$ as $\mathcal{A}(M) = [\langle A_1, M \rangle, \ldots, \langle A_m, M \rangle]^T$ for all M. The matrix sensing problem is formulated as the following non-convex optimization problem:

$$\min_{M \in \mathbb{R}^{n \times n}} \frac{1}{2} \|\mathcal{A}(M) - b\|^2 \quad \text{subject to} \quad \text{rank}(M) = r. \tag{1}$$

where $b = \mathcal{A}(M^*)$ is the observed vector, M^* is the unknown ground truth matrix, and r denotes the rank of M^*. Since the matrix sensing problem for an arbitrary solution M^* (being a rectangular matrix or a square sign indefinite matrix) can be converted to an expanded matrix sensing problem whose solution is a symmetric and positive semidefinite matrix (Zhang et al., 2021), we assume that M^* is positive semidefinite and symmetric without loss of generality.

The matrix sensing problem has a wide range of real-world applications in signal processing and machine learning, such as the training of neural networks (Li et al., 2018), reconstruction of images and videos (Fowler et al., 2012; Baraniuk et al., 2017), wireless sensor network (Razzaque et al., 2013), and quantum computing (Shabani et al., 2011; Ayanzadeh et al., 2020). It has attracted significant attention in recent years as it sheds light on a board range of non-convex optimization problems, serving as a theoretical guarantee in deep learning theory (Li et al., 2018; Scarlett et al., 2022). The complexity of the matrix sensing problem lies in the low-rank structure that creates spurious solutions, which makes local search algorithms with a random initialization become stuck at a wrong second-order critical point rather than the ground truth (Chen et al., 2019).

To overcome the above-mentioned non-convexity, one line of research relaxes this problem into a convex semi-definite program (SDP) (Candès & Recht, 2012; Recht et al., 2010), by replacing the rank constraint with a nuclear norm constraint. However, solving the SDP relaxation requires a large amount of calculations. Another popular way to deal with the low-rank constraint is the Burer-Monteiro (BM) factorization (Burer & Monteiro, 2003), which explicitly factorizes the low-rank matrix M into the form $M = XX^\top$ where $X \in \mathbb{R}^{n\times r}$ (note that this factorization uses the fact that M^* is positive definite and symmetric). Hence, the matrix sensing problem can be formulated as

$$\min_{X\in\mathbb{R}^{n\times r}} \frac{1}{2}\|\mathcal{A}(XX^\top) - b\|^2 \tag{2}$$

With this natural reparametrization, the number of parameters reduces from $O(n^2)$ in M to $O(nr)$ in X, where r is usually close to 1. Problem (2) is unconstrained, and therefore simple first-order methods, such as Gradient Descent (GD), can be applied to solve the problem. However, the factorized problem (2) is highly non-convex and $\mathcal{NP}$-hard to solve. There have been extensive studies on the optimization landscape of the matrix sensing problem (Candes & Tao, 2010; Candès & Recht, 2012; Recht et al., 2010; Ge et al., 2017; Zhang et al., 2018), and it turns out that the success of both SDP relaxation and local search methods relies on a condition named Restricted Isometry Property (RIP), which is defined below.

Definition 1 (RIP (Candès & Recht, 2012))**.** *Given a natural number s, the linear map $\mathcal{A} : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^m$ is said to satisfy the Restricted Isometry Property (RIP) condition of rank s for a constant δ, denoted as $\delta_s \in [0, 1)$, if the inequality*

$$(1 - \delta_s)\|M\|_F^2 \le \|\mathcal{A}(M)\|^2 \le (1 + \delta_s)\|M\|_F^2 \tag{3}$$

holds for all matrices $M \in \mathbb{R}^{n\times n}$ satisfying $\operatorname{rank}(M) \le s$.

Intuitively, the RIP is a condition guaranteeing that linear measurements approximately preserve the Euclidean geometry of low-rank matrices. Specifically, a sensing operator satisfies the RIP if it acts nearly as an isometry on the set of low-rank matrices, ensuring that the distances between these matrices are preserved after measurement. When $\delta_s = 0$, solving the matrix sensing problem is trivial, while δ_s close to 1 implies a complicated landscape for the matrix sensing problem where the number of local minima could be exponential (Yalçın et al., 2023). Note that the RIP constant is not unique. If δ_s is an RIP constant, every number greater than δ_s is also an RIP constant.

The RIP condition is crucial for the success of various recovery algorithms, as it underpins their ability to reconstruct the original matrix accurately from compressed measurements. Started by the convex relaxation approach, Recht et al. (2010) and Candès & Recht (2012) demonstrated that when the RIP constant satisfies the inequality $\delta_{5r} \le 1/10$, the SDP relaxation is exact, allowing for the exact recovery of the ground truth M^*. Later, Bhojanapalli et al. (2016) examined the factorized problem (2) and showed that $\delta_{2r} \le 1/5$ suffices to guarantee that all second-order critical points for (2) correspond to the ground truth solution. Zhu et al. (2018) further established that $\delta_{4r} \le 1/5$ is sufficient for the global recovery of the ground truth via a local search method. The recent paper (Zhang et al., 2021) showed that $\delta_{2r} < 1/2$ is the tightest bound for guaranteeing such global properties.

Through the lens of RIP, one can guarantee benign optimization landscape and convergence to global optimality, solving the matrix sensing problem either using convex relaxations such as SDP or using non-convex methods such as the BM factorization with a random initialization. Furthermore, when the RIP constant is small, local search has a linear convergence rate for the factorized problem (2) (Zheng & Lafferty, 2015; Lee & Stöger, 2023). Moreover, strict-saddle property holds if $\delta_{2r} < 1/2$, and this result was developed for general low-rank optimization problems beyond matrix sensing (Bi et al., 2022). While the bound $\delta_{2r} < 1/2$ is sharp , it is not satisfied for most real-world problems except in special cases such as a class of isometric distributions.

Definition 2 (Nearly isometrically distributed (Recht et al., 2010))**.** *Let $\mathcal{A}$ be a random variable that takes values in linear maps from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^m$. We say that $\mathcal{A}$ is nearly isometrically distributed if for all $X \in \mathbb{R}^{n\times n}$ it holds that*

$$\mathbf{E}\left[\|\mathcal{A}(X)\|^2\right] = \|X\|_F^2$$

and for all $0 < \epsilon < 1$ *we have*

$$\mathbf{P}\left(\left|\|\mathcal{A}(X)\|^2 - \|X\|_F^2\right| \geq \epsilon\|X\|_F^2\right) \leq 2\exp\left(-\frac{m}{2}\left(\epsilon^2/2 - \epsilon^3/3\right)\right)$$

and for all $t > 0$ *we have*

$$\mathbf{P}\left(\sup_{X\neq 0} \frac{\|\mathcal{A}(X)\|}{\|X\|_F} \geq 1 + \sqrt{\frac{n^2}{m}} + t\right) \leq \exp\left(-\gamma m t^2\right)$$

for some constant $\gamma > 0$.

Given $0 < \delta < 1$ and $1 \leq r \leq m$, it turns out that $\mathcal{A}$ is a nearly isometric random variable, with high probability, i.e., $\delta_r(\mathcal{A}) \leq \delta$ if $m = \Theta(rn/\delta^2)$ (Recht et al., 2010; Candès & Plan, 2011). Independent and identically distributed (i.i.d.) Gaussian entries with variance $1/m$ are nearly isometrically distributed, and the literature of matrix sensing has heavily relied on the i.i.d. and Gaussian assumptions to justify the use of RIP. However, in practice we often have no prior knowledge of the distribution of the sensing matrices, and in addition the independence assumption is hardly satisfied.

Motivated by this inconsistency and the importance of RIP, this paper offers theoretical insights into deviations from nearly isometric distributions. We first study the case when the problem is not Gaussian due to small perturbations and derive an upper bound on the change to the RIP constant in terms of the distance of the distribution of the given operator from a Gaussian distribution. Moreover, we study whether an operator with an arbitrary distribution can be modified so that it acts as a perturbed Gaussian for which the above result on its RIP constant can be applied. For the case where the true distribution deviates significantly from normal distributions, we introduce a preconditioning algorithm that replaces each sensing matrix with a weighted sum of all sensing matrices. We discuss how this technique makes the resulting operator behave similarly to perturbed Gaussian distributions, leading to a reduction in the RIP constant and improving the optimization landscape.

The paper is organized as follows. In Section 2, we demonstrate the robustness of the RIP constant to small perturbations to the sensing operator. We show that nearly-isometric measurements under a modest perturbation continue to satisfy the RIP, thereby ensuring the reliable recovery of low-rank matrices. This finding is significant as it broadens the applicability of matrix sensing techniques to real-world scenarios by relaxing the restrictive Gaussian assumption.

Furthermore, we investigate the role of orthogonalization in enhancing the optimization landscape of the matrix sensing problem. In Section 3, we show that the orthogonalization of the sensing matrices can improve the RIP constant, making the landscape more favorable for efficient recovery algorithms. To achieve this, we propose a novel preconditioning method that optimizes the mixing of the measurements to reduce the RIP constant. We provide a theoretical analysis for the proposed method, and empirically show that it is highly effective on various types of measurement distributions, including Poisson, uniform, and correlated Gaussian distributions. In particular, we demonstrate that the original RIP constants for these distributions could be close to 1 for which the SDP relaxation and local search methods would fail to work, while the preconditioning technique reduces the RIP to less than 0.5 so that both of these optimization methods can correctly solve the modified problem.

By addressing the above two aspects, our work contributes to a deeper understanding of the matrix sensing problem with non-Gaussian models. We propose practical solutions to enhance recovery performance, paving the way for more robust and efficient applications in matrix sensing and beyond.

Definitions and Notations The symbol $\|v\|$ denotes the Euclidean norm of a vector v. $\|X\|_F$ denotes the Frobenius norm of a matrix X. $\|X\|_M = \max_{i,j}|X_{ij}|$ denotes the largest absolute entry of a matrix X. $\|\mathcal{A}\|_M = \max_k \max_{i,j}|{A_k}^{ij}|$ denotes the largest absolute entry of a sensing operator $\mathcal{A}$, where A_k^{ij} denotes the (i,j) entry of the matrix A_k. $\sigma_i(X)$ denotes the i-th largest singular value of a matrix X. $\lambda_i(X)$ denotes the i-th largest eigenvalue of a symmetric matrix X. $\langle A, B\rangle$ is defined as the inner product $\operatorname{tr}\left(A^T B\right)$ for two matrices A and B of the same size, where tr stands for trace. $\mathbf{E}(x)$ denotes the expectation of a random variable x. $\mathbf{P}(E)$ denotes the probability of en event E. $f = \Theta(g)$ denotes that there exist constants $c_1, c_2 > 0$ such that $c_1 * g \leq f \leq c_2 * g$. $f = \mathcal{O}(g)$ denotes that there exists a constant $c > 0$ such that $f \leq c * g$. For a matrix X, $\operatorname{vec}(X)$ is the usual vectorization operation by stacking the columns of the matrix X into a vector and $\mathrm{mat}(\cdot)$ is the inverse operator. $\mathrm{VStack}(\cdot)$ denotes concatenating the rows of a matrix into a vector. $[n]$ denotes the integer set $\{1, \ldots, n\}$. $\delta_s(\mathcal{A})$ denotes the smallest value for δ_s that satisfies the RIP condition of rank s for the sensing operator $\mathcal{A}$. The matrix orthogonality and the orthonormal basis are defined under the standard inner product $\langle \cdot, \cdot \rangle$.

2 PERTURBED ISOMETRICAL DISTRIBUTION

Given an arbitrary sensing operator $\mathcal{A}$, we first prove that if $\mathcal{A}$ is perturbed via another operator that is bounded by ε, then its RIP constant will be increased by at most $\mathcal{O}(mn^2\varepsilon)$.

Theorem 1. *Consider an arbitrary operator $\mathcal{A}$ with the RIP constant $\delta_s \in [0, 1)$. Let ε be a nonnegative constant such that $\varepsilon < \frac{1-\delta_s}{2mn^2\|\mathcal{A}\|_\infty}$. For every bounded perturbation operator $\mathcal{N}$ with $\|\mathcal{N}\|_\infty \leq \varepsilon$, the perturbed sensing operator $\mathcal{A} + \mathcal{N}$ satisfies the RIP condition of rank s with the constant $\delta_s + (4mn^2\|\mathcal{A}\|_\infty\varepsilon + mn^2\varepsilon^2(1-\delta))/(2 + mn^2\varepsilon^2)$.*

Proof. Let $N_1, \ldots, N_m$ denote the components of $\mathcal{N}$, i.e., $\mathcal{N}(X) = [\langle N_1, X \rangle, \ldots, \langle N_m, X \rangle]$. For every matrix $X \in \mathbb{R}^{n\times n}$ satisfying $\mathrm{rank}(X) \leqslant s$, it holds that

$$\begin{aligned}\|(\mathcal{A}+\mathcal{N})(X)\|^2 &= \sum_{i=1}^m \langle A_i + N_i, X \rangle^2 \\ &= \sum_{i=1}^m \langle A_i, X \rangle^2 + \sum_{i=1}^m \langle N_i, X \rangle^2 + 2\sum_{i=1}^m \langle A_i, X \rangle \langle N_i, X \rangle\end{aligned}$$

Since $\mathcal{A}$ satisfies the RIP condition with the constant δ_s, we have

$$(1-\delta_s)\|X\|_F^2 \leqslant \sum_{i=1}^m \langle A_i, X \rangle^2 \leqslant (1+\delta_s)\|X\|_F^2$$

Due to the Cauchy-Schwarz inequality, one can write

$$0 \leqslant \sum_{i=1}^m \langle N_i, X \rangle^2 \leqslant \sum_{i=1}^m \|N_i\|_F^2 \|X\|_F^2 \leqslant mn^2\varepsilon^2\|X\|_F^2,$$

and

$$\left|\sum_{i=1}^m \langle A_i, X \rangle \langle N_i, X \rangle\right| = \left|\sum_{i=1}^m \langle A_i, N_i \rangle\right| \|X\|_F^2 \leqslant mn^2\varepsilon\|\mathcal{A}\|_\infty\|X\|_F^2$$

Hence,

$$0 < \left(1-\delta_s - 2mn^2\|\mathcal{A}\|_\infty\varepsilon\right)\|X\|_F^2 \leqslant \|(\mathcal{A}+\mathcal{N})(X)\|^2 \leqslant \left(1+\delta_s + mn^2\varepsilon^2 + 2mn^2\|\mathcal{A}\|_\infty\varepsilon\right)\|X\|_F^2$$

indicating that $\mathcal{A}+\mathcal{N}$ satisfies the RIP condition with the constant $\delta_s + mn^2\varepsilon \cdot \frac{4\|\mathcal{A}\|_\infty + \varepsilon(1-\delta_s)}{2+mn^2\varepsilon^2}$. $\square$

If $\mathcal{N}$ is chosen as $-\mathcal{A}$, then the RIP condition is not satisfied. Similarly, if $N_1, \ldots, N_m$ are chosen in a way that the (i, j) entries of all matrices $A_1 + N_1, \ldots, A_m + N_m$ are zero for some indices i and j, then the RIP condition again no loner holds. For these reasons, the existence of an upper bound on ε in Theorem 1 is necessary.

Remark 1. *With series expansion at $\varepsilon = 0$, the RIP constant derived in Theorem 1 can be approximated by $\delta_s + mn^2\left(2\|\mathcal{A}\|_\infty\varepsilon + \frac{1}{2}(1-\delta_s)\varepsilon^2 - \|\mathcal{A}\|_\infty mn^2\varepsilon^3 + \mathcal{O}(\varepsilon^4)\right)$. On the other hand, since $\mathcal{A}$ satisfies the RIP condition with the constant δ_s, the term $\|\mathcal{A}\|_\infty$ can be bounded by choosing a matrix X whose entry at the position of the largest element of $\mathcal{A}$ is 1 and whose remaining entries are 0. Hence, $\|X\|_F^2 = 1$ and $\|\mathcal{A}\|_\infty^2 \leq \sum_{i=1}^m \langle A_i, X \rangle^2 \leqslant (1+\delta_s)\|X\|_F^2$, indicating that $\|\mathcal{A}\|_\infty \leq \sqrt{1+\delta_s}$. Thus, the RIP condition for $\mathcal{A}+\mathcal{N}$ can be upper bounded by $\delta_s + mn^2\varepsilon[2(1+\delta_s)^{1/2} + \frac{1}{2}(1-\delta_s)\varepsilon]$ up to the first-order approximation.*

Theorem 1 studies bounded perturbation operators $\mathcal{N}$ in the worst case. We will improve the results by relaxing the boundedness of the perturbation.

Corollary 1. *Consider an arbitrary operator $\mathcal{A}$ with the RIP constant $\delta_s \in [0, 1)$. Consider also a perturbation operator $\mathcal{N}$ such that $\|\mathcal{N}\|_\infty$ is sub-Gaussian with mean 0 and variance proxy σ^2/m. For every $c > 0$ and $\sigma < \frac{1-\delta_s}{2c\sqrt{m}n^2\|\mathcal{A}\|_\infty}$, with probability at least $1 - 2\exp(-c^2)$, the operator $\mathcal{A}+\mathcal{N}$ satisfies the RIP condition with the constant $\delta_s + c\sqrt{m}n^2\sigma[2(1+\delta_s)^{1/2} + \frac{c}{2\sqrt{m}}(1-\delta_s)\sigma]$.*

Proof. Since $\|\mathcal{N}\|_\infty$ is sub-Gaussian bounded, we have $\mathbf{P}(\|\mathcal{N}\|_\infty \geq \varepsilon) \leq 2\exp\left(-\frac{m\varepsilon^2}{\sigma^2}\right)$. This implies that $\mathbf{P}(\|\mathcal{N}\|_\infty \leq \varepsilon)$ with probability at least $1 - 2\exp\left(-m\varepsilon^2/\sigma^2\right)$. Combining Theorem 1 and $\varepsilon = c\sigma/\sqrt{m}$, it can be concluded that with probability at least $1 - 2\exp(-c^2)$, $\mathcal{A}+\mathcal{N}$ satisfies the RIP condition with the constant

$$\delta_s + c\sqrt{m}n^2\sigma[2(1+\delta_s)^{1/2} + \frac{c}{2\sqrt{m}}(1-\delta_s)\sigma].$$

This completes the proof. $\square$

Building on Corollary 1, we refine the RIP bound for a nearly isometrically distributed operator $\mathcal{A}$.

Theorem 2. *Assume that $\mathcal{A}$ is nearly isometrically distributed and $\|\mathcal{N}\|_\infty$ is sub-Gaussian with mean 0 and variance proxy σ^2/m. There exist positive constants c_1 and c_2, independent of the parameters of $\mathcal{N}$ (such as σ) such that for every $c > 0$ and $\sigma < \frac{1-\delta_s}{2c\sqrt{m}n^2\|\mathcal{A}\|_\infty}$, with probability at least $1 - 2\exp(-c^2) - \exp(-c_1 m)$, the operator $\mathcal{A}+\mathcal{N}$ satisfies the RIP condition with the constant $c_2\sqrt{ns\log n/m} + c\sqrt{m}n^2\sigma[2(1+\delta_s)^{1/2} + \frac{c}{2\sqrt{m}}(1-\delta_s)\sigma]$.*

Proof. It has been proved in Recht et al. (2010) that if $\mathcal{A}$ is nearly isometrically distributed, then there exist positive constants c_1 and c_2 with c_1 depending on the RIP constant of $\mathcal{A}$ such that, with probability at least $1 - \exp(-c_1 m)$, we have $\delta_s(\mathcal{A}) \leq c_2\sqrt{ns\log n/m}$. Now, it follows from Corollary 1 that with probability at least $1 - 2\exp(-c^2) - \exp(-c_1 m)$, it holds that $\mathcal{A}+\mathcal{N}$ satisfies the RIP condition with the constant

$$c_2\sqrt{ns\log n/m} + c\sqrt{m}n^2\sigma[2(1+\delta_s)^{1/2} + \frac{c}{2\sqrt{m}}(1-\delta_s)\sigma].$$

This completes the proof. $\square$

Remark 2. *Due to Theorem 2, the RIP constant of the perturbed operator $\mathcal{A}+\mathcal{N}$ compared to the RIP of $\mathcal{A}$ has increased from $\mathcal{O}(1/\sqrt{m})$ to $\mathcal{O}(1/\sqrt{m}) + \mathcal{O}(\sqrt{m}\sigma) + \mathcal{O}(\sigma^2)$. Thus, when the perturbation σ is small, one can compensate for the influence of the perturbation on the RIP constant by slightly increasing the number of measurements m, which will reduce the RIP constant of the perturbed operator to the RIP constant of the unperturbed operator $\mathcal{A}$. This formula shows how many additional measurements are needed to nullify the effect of deviation from a Gaussian distribution.*

To summarize the results of this section, Theorem 1 provides an RIP bound for a fixed sensing operator $\mathcal{A}$ and a bounded perturbation $\mathcal{N}$, and Corollary 1 extends this result to a random perturbation $\mathcal{N}$. In Theorem 2, we further derive a high probability bound for any nearly isometric random distributed sensing operator $\mathcal{A}$, and prove that the impact of a small perturbation on RIP is small and that increasing the number of measurements m on a small scale can compensate for the increase in RIP.

3 Preconditioning of Matrix Sensing

In the previous section, we proved that small deviations from nearly isometrically distributed sensing matrices will slightly increase the RIP constant. However, real-world sensing matrices often have unknown probability distributions that cannot be approximated by Gaussian models, for which several empirical results have shown that the RIP constant is often close to 1. To address this issue, we consider a sensing operator $\mathcal{A}$ coming from an arbitrary probability distribution and develop a preconditioning algorithm to improve its RIP constant and make it act as a perturbed Gaussian.

It has been proved in Ma et al. (2023; 2024) that the RIP constant can be reduced if the optimization complexity of the matrix sensing problem is increased, e.g., via a tensor-based lifting technique. However, this incurs a high computational cost and is not applicable to large-scale matrix sensing problems. To avoid this computational complexity, we propose a simple and scalable linear preconditioning method, which replaces every sensing matrix with a linear combination of all the original sensing matrices. More precisely, consider a weight matrix $P \in \mathbb{R}^{m\times m}$ with its (i, j) entry denoted as P_{ij}. We construct a preconditioned operator $\tilde{\mathcal{A}}$ with the components $\tilde{A}_1, ..., \tilde{A}_m$ as follows:

$$\tilde{A}_i = \sum_{j=1}^{m} P_{ij} A_j, \quad \forall i \in \{1, ..., m\}$$

Therefore,

$$\langle \tilde{A}_i, X\rangle = \sum_{j=1}^{m} P_{ij}\langle A_j, X\rangle = \sum_{j=1}^{m} P_{ij} b_j, \quad \forall i \in \{1, \ldots, m\}, \forall X$$

Hence, $\tilde{\mathcal{A}}(X) = Pb$. The preconditioning is independent of the optimization method (such as local search or convex relaxation) to be used to solve the matrix sensing problem, and the goal of preconditioning is to create a better structure for the sensing operator and thus a better RIP constant. In what follows, we will develop a simple method for designing P and study its impact on the RIP constant.

3.1 Orthonormal Bases as Sensing Matrices

The following lemma for Haar distribution is the basis of our method.

Lemma 1 (Frankl & Maehara (1988)). *Let $\{x_j\}_{j=1}^{n} \subseteq \mathbb{R}^d$, and let P be a $k \times d$ random matrix, consisting of the first k rows of a Haar-distributed random matrix in the orthogonal group $\mathbb{O}(d)$. Given $\epsilon > 0$ and $k = \frac{a\log(n)}{\epsilon^2}$, there are absolute constants c and C such that with probability at least $1 - Cn^{2-\frac{ac}{4}}$ the inequalities*

$$(1-\epsilon)\left\|x_i - x_j\right\|^2 \le \binom{d}{k}\left\|Px_i - Px_j\right\|^2 \le (1+\epsilon)\left\|x_i - x_j\right\|^2$$

hold for all $i, j \in \{1, \ldots, n\}$.

The orthonormal vectors from the unitary matrix in QR decomposition of i.i.d. Gaussian matrices follow a Haar distribution (Mezzadri, 2007). Given that those orthonormal bases maintain the distances during projection, we are inspired to transform our original sensing operator $\mathcal{A}$ into a preconditioned operator $\tilde{\mathcal{A}}$ with orthonormal bases as vectorized sensing matrices. To be more specific, we first write the sensing operator $\mathcal{A}$ into the vectorized form

$$\mathbf{A} = [\text{vec}(A_1), \text{vec}(A_2), \ldots, \text{vec}(A_m)]^T \in \mathbb{R}^{m\times n^2}.$$

Then, since the inner product of two matrices can be defined as a vector product, it holds that

$$\mathbf{A}\,\text{vec}(X) = \mathcal{A}(X), \quad \forall X \in \mathbb{R}^{n\times n}.$$

By pre-multiplying the above equation with a weight matrix $P \in \mathbb{R}^{m\times m}$, we ideally intend to make the rows of $P\mathbf{A}$ normalized and orthogonal to each other. Since the individual entries of a random orthogonal matrix are approximately Gaussian for large matrices (Meckes, 2019), as m increases, these preconditioned operators are likely to act as i.i.d. Gaussian.

Define the s-sparse set $\text{span}_s(\mathcal{A})$ as the set of all matrices X that can be written as $X = \sum_{i=1}^{m} \alpha_i A_i$ for some coefficients $\alpha_1, ..., \alpha_m$ such that at most s coefficients are nonzero. We say that $A_1, ..., A_m$ are orthonormal if $\langle A_i, A_j\rangle = 0, \forall i \neq j$ and $\langle A_i, A_i\rangle = 1$ otherwise.

Theorem 3. *Assume that $A_1, ..., A_m$ are orthogonal. It holds that*

$$\frac{||\mathcal{A}(X)||^2}{||X||_F^2} = 1, \quad \forall X \in \text{span}_s(\mathcal{A})$$

Proof. We expand the orthonormal matrices $A_1, ..., A_m$ into a basis for $\mathbb{R}^{n\times n}$. More precisely, consider orthonormal bases $V_1, \ldots, V_{n^2} \in \mathbb{R}^{n\times n}$ such that $V_i = A_i$ for $i = 1, \ldots, m$. Given a matrix $X \in \text{span}_s(\mathcal{A})$, we can write it as $\sum_{i=1}^m \alpha_i A_i$ with at most s nonzero α_i's. Without loss of generality, we assume that $\|X\|_F^2 = \sum_{i=1}^m \alpha_i^2 = 1$. Now, one can write:

$$\frac{\|\mathcal{A}(X)\|^2}{\|X\|_F^2} = \sum_{i=1}^m \sum_{j=1}^m \langle A_i, \alpha_j V_j \rangle^2 = \sum_{i=1}^m \alpha_i^2 = 1.$$

This completes the proof. $\square$

The set $\text{span}_s(\mathcal{A})$ includes matrices that can be written as the sum of at most s matrices from the set $\{A_1, ..., A_m\}$. As m increases, if this set continues to include orthonormal matrices, the set $\text{span}_s(\mathcal{A})$ grows until it completely covers the low-rank set $\{X \mid \text{rank}(X) \leq s\}$. Thus, it follows from Theorem 3 that as m grows, the RIP constant for orthonormal matrices approaches zero (note that RIP is about taking the minimum and maximum of the ratio $\|\mathcal{A}(X)\|^2/\|X\|_F$ over matrices of rank at most s). Hence, Theorem 3 justifies the conversion of arbitrary sensing matrices into orthogonal matrices.

3.2 Preconditioning Algorithm

Based on the idea of using orthonormal bases as sensing matrices, we propose Algorithm 1, which applies the singular value decomposition (SVD) to extract unitary sensing matrices from the given sensing operator.

Algorithm 1 Preconditioned Matrix Sensing

1: **for** $iteration = 1, 2, \ldots, m$ **do**
2: $\quad a_i \leftarrow \text{vec}(A_i)$
3: **end for**
4: $U, S, V^\top \leftarrow \text{SVD}(\text{VStack}(a_1, a_2, \ldots, a_m))$
5: **for** $iteration = 1, 2, \ldots, m$ **do**
6: $\quad \tilde{A}_i \leftarrow \text{mat}(V_i^\top)$
7: **end for**
8: **Return** $\tilde{\mathcal{A}} = [\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m]$

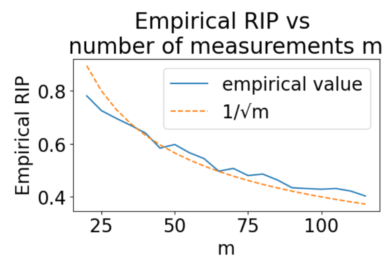

Figure 1: Empirical RIP curve

Remark 3. *The singular value decomposition of* $\mathbf{A}$ *written as* $U[S, \mathbf{0_{m\times(n^2-m)}}]V^\top$ *will obtain a unitary matrix* $V^\top$ *whose rows are the eigenvectors of* $\mathbf{A}^\top\mathbf{A}$. *The new sensing matrices* $\tilde{A}_i$ *obtained by reshaping the rows of* $V^\top$ *into matrices are perpendicular to each other. For* $S = \text{diag}([\sigma_1(\mathbf{A}), \ldots, \sigma_m(\mathbf{A})]) \in \mathbb{R}^{m\times m}$, *we could assume* $\mathbf{A}$ *to be full rank in practice, and since* $\sigma_m(\mathbf{A}) > 0$, S *becomes invertible. Since the extraction step can be considered as a linear transformation, we can easily calculate the corresponding vector* $b' = U^\top S^{-1} b$, *and the weight matrix is* $P = U^\top S^{-1}$.

In the following, we will derive an upper bound on the RIP constant of the preconditioned sensing operator. Furthermore, we will prove that the RIP constant of the preconditioned $\tilde{\mathcal{A}}$ is smaller that the RIP constant of the original sensing operator $\mathcal{A}$ with high probability.

Theorem 4. *Consider an arbitrary operator* A *with the RIP constant* $\delta_s \in [0, 1)$. *Then, the conditioned operator* $\tilde{\mathcal{A}}$ *also satisfies the RIP condition with the constant* $1 - \frac{1-\delta_s}{\sigma_1^2(\mathbf{A})}$.

Proof. Since $\mathcal{A}$ satisfied the RIP condition, the following inequality holds for every matrix M such that $\text{rank}(M) \leq s$:

$$(1-\delta_s)\|M\|_F^2 \leqslant \|\mathcal{A}(M)\|_2^2 = \|\mathbf{A}\,\text{vec}\,(M)\|_2^2 \leqslant (1+\delta_s)\|M\|_F^2.$$

As $\tilde{\mathbf{A}} = P\mathbf{A}$, we introduce the operator norm of P and write

$$\sup_{M:\mathbf{A}\,\text{vec}\,(M)\neq 0} \frac{\|P\mathbf{A}\,\text{vec}\,(M)\|_2^2}{\|\mathbf{A}\,\text{vec}\,(M)\|_2^2} = \lambda_1\left(P^\top P\right), \qquad \inf_{M:\mathbf{A}\,\text{vec}\,(M)\neq 0} \frac{\|P\mathbf{A}\,\text{vec}\,(M)\|_2^2}{\|\mathbf{A}\,\text{vec}\,(M)\|_2^2} = \lambda_m\left(P^\top P\right).$$

Now, we aim to bound $\|\tilde{\mathcal{A}}(M)\|_2^2$ by the eigenvalues of $P^\top P$. Since $P = U^\top S^{-1}$, U is a unitary matrix, and S is diagonal, we have $P^\top P = S^{-2}$, $\lambda_1\left(P^\top P\right) = \sigma_m^{-2}(\mathbf{A})$, and $\lambda_m\left(P^\top P\right) = \sigma_1^{-2}(\mathbf{A})$. Hence,

$$\sigma_1^{-2}(\mathbf{A})\|\mathbf{A}\,\text{vec}\,(M)\|_2^2 \leq \|P\mathbf{A}\,\text{vec}\,(M)\|_2^2 \leq \sigma_m^{-2}(\mathbf{A})\|\mathbf{A}\,\text{vec}\,(M)\|_2^2.$$

As a result, we obtain the lower bound

$$\|P\mathbf{A}\,\mathrm{vec}\,(M)\|_2^2 \geq \frac{1}{\sigma_1^2(\mathbf{A})}\|\mathbf{A}\,\mathrm{vec}\,(M)\|_2^2 \geq \frac{1-\delta_s}{\sigma_1^2(\mathbf{A})}\|\,\mathrm{vec}\,(M)\|_2^2.$$

On the other hand, since V is a unitary matrix, one can write

$$\|P\mathbf{A}\,\mathrm{vec}\,(M)\|_2^2 = \|S^{-1}U^\top\mathbf{A}\,\mathrm{vec}\,(M)\|_2^2 = \|[\mathbf{I_m}, \mathbf{0_{m\times(n^2-m)}}]V^\top\,\mathrm{vec}\,(M)\|_2^2 \leq \|\,\mathrm{vec}\,(M)\|_2^2$$

By combining the above two inequalities, we obtain the desired result for the RIP constant of $\tilde{\mathcal{A}}$. □

Remark 4. *$1-\frac{1-\delta_s}{\sigma_1^2(\mathbf{A})}$ in Theorem 4 can be smaller or larger than δ_s, depending on whether $\delta_s+1 > \sigma_1^2(\mathbf{A})$.*

Assumption 1. *Assume that singular values of the matrix $\mathbf{A}\in\mathbb{R}^{m\times n^2}$ with $m<n^2$ satisfy*

$$\Pr\left\{\sqrt{n^2/m}(1-\epsilon)-1\leq\sigma_i(\mathbf{A})\leq 1+\sqrt{n^2/m}(1+\epsilon), i\in[m]\right\}\geq 1-2\exp\left(-n^2\epsilon^2/2\right),\quad\forall\epsilon>0.$$

Assumption 2. *Consider two constants ϵ and δ such that*

$$0<\epsilon<1-\sqrt{\frac{m}{n^2}},\quad \frac{[1+\sqrt{\frac{m}{n^2}}(1+\epsilon)]^2-1}{2[1+\sqrt{\frac{m}{n^2}}(1+\epsilon)]^2-1}<\delta<\frac{1}{2}.$$

Theorem 5. *Let $\mathcal{A}$ be a nearly isometrically distributed operator. Under Assumption 1 and Assumption 2, there exist positive constants c_0 and c_1 depending only on δ_s such that, with probability at least $1-\exp(-c_1 m)-2\exp(-n^2\epsilon^2/2)$, as long as $m\geq c_2 sn\log(n)$, the original sensing operator satisfies $\delta_s(\mathcal{A})\leq 2\delta$ and the conditioned sensing operator satisfies $\delta_s(\tilde{\mathcal{A}})\leq 1-(1-\delta)/[1+\sqrt{\frac{n^2}{m}}(1+\epsilon)]^2$.*

Proof. Inspired by the proof of Theorem 1 in Chen & Lin (2021), define the following events

$$E_1 \doteq \{\mathcal{A}\text{ satisfies the RIP of rank } s \text{ with the constant } 2\delta\},$$
$$E_2 \doteq \left\{\tilde{\mathcal{A}}\text{ satisfies the RIP of rank } s \text{ with the constant } 1-(1-\delta)/[1+\sqrt{\tfrac{n^2}{m}}(1+\epsilon)]^2\right\},$$
$$F_1 \doteq \{\mathcal{A}\text{ satisfies the RIP of rank } s \text{ with the constant } \delta\},$$
$$F_2 \doteq \left\{\sqrt{\tfrac{n^2}{m}}(1-\epsilon)-1\leq\sigma_i(\mathbf{A})\leq 1+\sqrt{\tfrac{n^2}{m}}(1+\epsilon), i\in[m]\right\}.$$

We will show that $\Pr(E_1E_2)\geq\Pr(F_1F_2)$. Consider the singular value decomposition of $\mathbf{A}$ as $\mathbf{A}=U[S,\mathbf{0_{m\times(n^2-m)}}]V^\top$, where $U\in\mathbb{R}^{m\times m}, S=\mathrm{diag}([\sigma_1(\mathbf{A}),\ldots,\sigma_m(\mathbf{A})])\in\mathbb{R}^{m\times m}, V\in\mathbb{R}^{n^2\times n^2}$. Under Assumption 2, we have $\sqrt{\frac{n^2}{m}}(1-\epsilon)-1>0$, and therefore S is nonsingular. Hence, the preconditioning matrix defined as $P=S^{-1}U^\top$ is valid. If $\mathbf{A}\in F_1F_2$, in light of Theorem 4, F_1 implies that the conditioned operator $\tilde{\mathcal{A}}$ satisfies the RIP of rank s with the constant $1-\frac{1-\delta}{\sigma_1^2(\mathbf{A})}$. With F_2 implying an upper bound on $\sigma_1^2(\mathbf{A})$, obtain that $\tilde{\mathcal{A}}$ satisfies the RIP inequality (3) with the constant $1-(1-\delta)/[1+\sqrt{\frac{n^2}{m}}(1+\epsilon)]^2$. We could also have $\mathbf{A}\in E_2$. Since $F_1\subset E_1$, we have $\Pr(E_1E_2)\geq\Pr(F_1F_2)$. With the union bound $\Pr(F_1F_2)\geq\Pr(F_1)+\Pr(F_2)-1$, we estimate the probabilities $\Pr(F_1)$ and $\Pr(F_2)$ using Theorem 4.2 in Recht et al. (2010) and Assumption 1 to arrive at

$$\begin{aligned}\Pr(E_1E_2)&\geq\Pr(F_1F_2)\\&\geq\Pr(F_1)+\Pr(F_2)-1\\&\geq 1-\exp(-c_1m)-2\exp(-n^2\epsilon^2/2)\end{aligned}$$

This completes the proof. □

To shed light on the two assumptions used in Theorem 5, note that Gaussian random matrices satisfy Assumption 1 as an example. Regarding Assumption 2, when $\epsilon\to 0, \delta\to\frac{1}{2}$, and the lower bound of δ is always smaller than $\frac{1}{2}$. As a result, such pair (ϵ,δ) satisfying Assumption 2 always exists. We immediately obtain the following corollary due to the fact that Gaussian random matrices satisfy the concentration inequality of the singular values naturally.

Corollary 2. *Let $A_1, \dots, A_m$ be i.i.d. Gaussian random matrices of mean zero and variance $\frac{1}{m}$. Under Assumption 2, there exist positive constants c_0 and c_1 such that, with probability at least $1-\exp(-c_1 m) - 2\exp(-n^2\epsilon^2/2)$, as long as $m \geq c_0 s(m + n^2 \log(mn^2))$, , it holds that $\delta_s(\mathcal{A}) \leq 2\delta$ and $\delta_s(\tilde{\mathcal{A}}) \leq 1 - (1-\delta)/[1 + \sqrt{\frac{n^2}{m}}(1+\epsilon)]^2$.*

Remark 5. *As explained before, the RIP constant is not unique and if δ_s satisfies the inequality 3, then any number between δ_s and 1 also satisfies the inequality. Based on this argument, it follows from Corollary 2 is that the RIP constants of the original operator $\mathcal{A}$ and preconditioned operator $\tilde{\mathcal{A}}$ are 2δ and $1-(1-\delta)/[1+\sqrt{\frac{n^2}{m}}(1+\epsilon)]^2$. The preconditioning improves the RIP if $2\delta > 1-(1-\delta)/[1+\sqrt{\frac{n^2}{m}}(1+\epsilon)]^2$, which holds true under Assumption 2 for our choice of the parameters. Hence, the preconditioning algorithm can improve the RIP constant for nearly isometrically distributed random matrices with high probability. Under Assumption 2, although 2δ can be close to 1, the preconditioning technique has the ability to reduce it to less than 1/2.*

3.3 Simulation Experiments

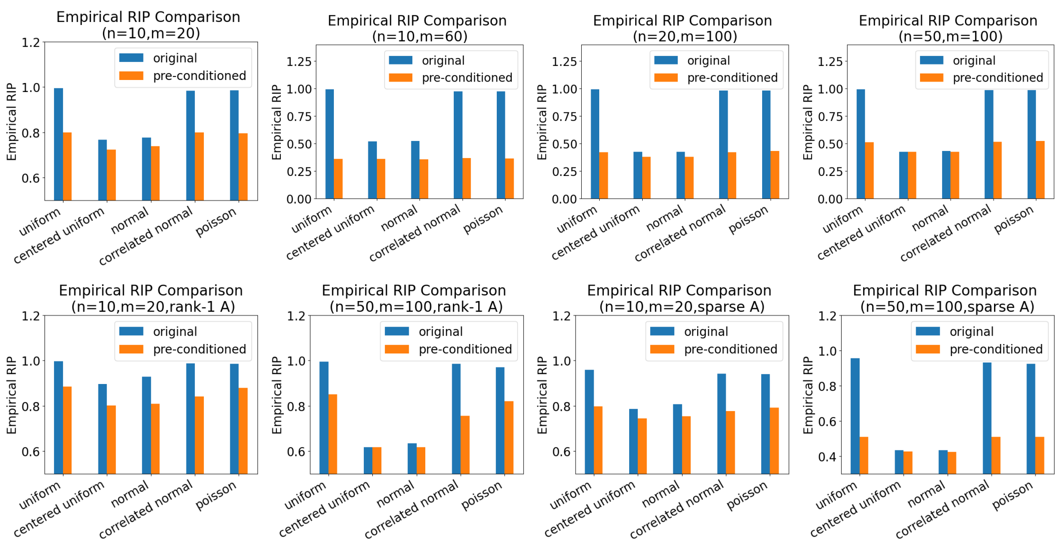

Figure 2: Empirical RIP comparison before and after preconditioning; the horizontal axis shows that the sensing matrices are sampled from uniform distribution $[0, 1]$, centered uniform distribution $[-1, 1]$, standard normal distribution, multivariate correlated normal distribution with $\rho = 0.5$, and poisson distribution separately. The first row is for general sensing matrices, and the second row is for matrices with special structures.

In this subsection, we will demonstrate the performance of the preconditioning Algorithm 1 for $s = 2r$ since δ_{2r} determines whether or not SDP relaxations or local search methods succeed to solve the matrix sensing problem. However, measuring the true RIP value δ_{2r} for any given sensing operator $\mathcal{A}$ requires checking the inequalities (3) for all low-rank matrices X of rank at most $2r$ and determining the maximum and minimum possible values of $\|\mathcal{A}(X)\|_2^2/\|X\|_F^2$ over all rank-$2r$ matrices. This is equivalent to solving a non-convex optimization problem, which is known to be $\mathcal{NP}-$hard. Hence, we will instead measure the empirical RIP constant in our experiments.

By randomly selecting 1000 Gaussian distributed matrices $M \in \mathbb{R}^{n\times 2r}$ (we simply choose $r = 1$ in the following experiments), we generate 1000 rank-$2r$ matrices $X = M^\top M \in \mathbb{R}^{n\times n}$ to be rank-$2r$ matrices. Afterwards, we calculate $\|\mathcal{A}(M)\|_2^2/\|X\|_F^2$ for all those X matrices and compute the smallest and the largest values, denoted as α and β correspondingly. Hence, we obtain the following inequalities over the generated samples of rank-$2r$ matrices:

$$\alpha\|X\|_F^2 \leq \|\mathcal{A}(X)\|^2 \leq \beta\|X\|_F^2. \tag{4}$$

Since rescaling (multiplying the sensing operator $\mathcal{A}$ by a constant c) will not affect the landscape of the matrix sensing problem, we multiply all of the above inequalities by $\frac{2}{\alpha+\beta}$ and calculate the empirical RIP constant for $\frac{2}{\alpha+\beta}\mathcal{A}$, which is $\frac{\beta+\alpha}{\beta-\alpha}$. Given that the set of simulated X is a subset of all low-rank matrices, the simulated RIP is a lower bound for true RIP value. We can see in Figure 1 that for Gaussian distributed sensing matrices, the empirical RIP value decreases as the number of measurements m increases. The empirical RIP curve matches the $m^{-1/2}$ curve, which is the result of the true RIP bound in Recht et al. (2010). Hence, we could treat the empirical RIP value as an accurate measure of the true RIP constant.

We randomly generate m sensing matrices under different distributions, including nearly isometric distributions such as Gaussian and non-isometric distributions such as Poisson. Besides, we also generate $\mathcal{A}$ with special structures, including low-rank structures and sparse structures. We numerically calculate the empirical RIP value before and after the preconditioning step. We run experiments under different scenarios from $n = 10$ to $n = 50$, and run 100 trials for each scenario to obtain the average empirical RIP value.

The results are plotted in Figure 2. We can see from the figure that for uniform, correlated normal and poisson distribution, the original sensing operator has a RIP constant close to 1, which means that with the i.i.d. Gaussian assumption violated, these measuring operators are no longer nearly isometric and thus cannot guarantee a benign optimization landscape for the matrix sensing problem. However, after preconditioning, we observe a clear decrease in the corresponding empirical RIP value. The preconditioned sensing matrices have the same level of RIP constant compared to the standard normal distribution with the same m, n values. On the other hand, for centered uniform and standard normal distribution, we can decrease the RIP constant by increasing m, and the preconditioning step can still slightly help to decrease the RIP value. This improvement becomes more obvious for the cases with a large m/n^2.

In addition to unstructured operators $\mathcal{A}$, we also study sensing matrices with special structures. For the low-rank structure, we generate $a_i \in \mathbb{R}^{n\times 1}$ and define $A_i = a_i a_i^\top \in \mathbb{R}^{n\times n}$. For the sparse structure, we generate a binomial distributed mask with $p = 0.3$, and only 30% elements of $\mathcal{A}$ are likely to be non-zero. The results are similar to the case of unstructured operators (see Figure 1), and the preconditioning effectively decreases the empirical RIP value in both low-rank and sparse cases. Even centered uniform and normal distribution will be affected by these special structures and show high empirical RIP values. One can observe that our preconditioning algorithm has a universal impressive performance in a wide range of situations.

Moreover, we can see that for the same level of (m, n), whatever the original distribution is, the empirical RIP value after preconditioning for different types of distributions are almost the same, which means that in practice we may not need to make additional assumptions on the distribution of sensing matrices; the landscape after preconditioning as well as the RIP constant will mainly depend on the value of r, m, n. As is shown by simulation experiments, Algorithm 1 makes best use of the current information provided by the original sensing operator and remains stable under different scenarios. The computational cost is also not high, only requiring $\mathcal{O}(m^2n^2)$ for a singular value decomposition of a matrix of dimension $m \times n^2$.

4 Conclusion

The results presented in this paper highlight several critical insights into the behavior of sensing operators and their impact on the Restricted Isometry Property (RIP) constant. When dealing with a nearly isometric operator perturbed by a sub-Gaussian term, the impact of deviation from the nearly isometric case can be effectively mitigated by increasing the number of measurements. Specifically, the RIP constant ensures that a benign optimization landscape can be preserved even in the presence of perturbations to the sensing operator. Thus, even in the presence of perturbations, careful adjustment of the number of measurements provides a practical approach to deal with non-Gaussian distributions. Our findings also demonstrate both theoretically and empirically that the proposed preconditioning algorithm significantly improves the RIP constant for various distributions. A notable observation is that, after preconditioning, the RIP constant is nearly independent of the original distribution. This finding simplifies practical implementations, as it eliminates the need for distribution-specific assumptions about the sensing matrices. Practitioners can rely on the preconditioned sensing matrices to provide consistent RIP performance, primarily governed by the values of r, m, and n.

References

Ramin Ayanzadeh, Milton Halem, and Tim Finin. An ensemble approach for compressive sensing with quantum annealers. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3517–3520, 2020.

Richard G Baraniuk, Thomas Goldstein, Aswin C Sankaranarayanan, Christoph Studer, Ashok Veeraraghavan, and Michael B Wakin. Compressive video sensing: Algorithms, architectures, and applications. *IEEE Signal Processing Magazine*, 34(1):52–66, 2017.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 3880–3888, 2016.

Yingjie Bi, Haixiang Zhang, and Javad Lavaei. Local and global linear convergence of general low-rank matrix recovery problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36:10129–10137, 2022.

Samuel Burer and Renato D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

Emmanuel Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Commun. ACM*, 55(6):111–119, 2012.

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

Emmanuel J. Candès 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.

Yingtong Chen and Shoujin Lin. Preconditioning with rip improvement for compressed sensing systems in total noise environment. *Signal Processing*, 179:107765, 2021.

Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, 176:5–37, 2019.

James E. Fowler, Sungkwang Mun, and Eric W. Tramel. Block-based compressed sensing of images and video. *Found. Trends Signal Process.*, 4:297–416, 2012.

P Frankl and H Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44(3):355–362, 1988.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1233–1242. PMLR, 2017.

Kiryung Lee and Dominik Stöger. Randomly initialized alternating least squares: Fast convergence for matrix sensing. *SIAM Journal on Mathematics of Data Science*, 5(3):774–799, 2023.

Yuanzhi Li, Tengyu Ma, and Hongyang Zhang. Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In *Proceedings of the 31st Conference On Learning Theory*, volume 75, pp. 2–47. PMLR, 2018.

Ziye Ma, Javad Lavaei, and Somayeh Sojoudi. Algorithmic regularization in tensor optimization: Towards a lifted approach in matrix sensing. In *Thirty-seventh Conference on Neural Information Processing Systems*, volume 36, 2023.

Ziye Ma, Ying Chen, Javad Lavaei, and Somayeh Sojoudi. Absence of spurious solutions far from ground truth: A low-rank analysis with high-order losses. In *International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 1603–1611. PMLR, 2024.

Elizabeth S. Meckes. *The Random Matrix Theory of the Classical Compact Groups*. Cambridge University Press, 2019.

Francesco Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society*, 54(5):592 – 604, 2007.

M. A. Razzaque, Chris Bleakley, and Simon Dobson. Compression in wireless sensor networks: A survey and comparative evaluation. *ACM Trans. Sen. Netw.*, 10(1), 2013.

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.

Jonathan Scarlett, Reinhard Heckel, Miguel R. D. Rodrigues, Paul Hand, and Yonina C. Eldar. Theoretical perspectives on deep learning methods in inverse problems. *IEEE Journal on Selected Areas in Information Theory*, 3(3):433–453, 2022.

A Shabani, RL Kosut, M Mohseni, H Rabitz, MA Broome, MP Almeida, A Fedrizzi, and AG White. Efficient measurement of quantum dynamics via compressive sensing. *Physical review letters*, 106(10):100401, 2011.

Baturalp Yalçın, Ziye Ma, Javad Lavaei, and Somayeh Sojoudi. Semidefinite programming versus burer-monteiro factorization for matrix sensing. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 2023.

Haixiang Zhang, Yingjie Bi, and Javad Lavaei. General low-rank matrix optimization: Geometric analysis and sharper bounds. In *Advances in Neural Information Processing Systems*, volume 34, pp. 27369–27380, 2021.

Richard Zhang, Cédric Josz, Somayeh Sojoudi, and Javad Lavaei. How much restricted isometry is needed in nonconvex matrix recovery? *Advances in Neural Information Processing Systems*, 31, 2018.

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

Zhihui Zhu, Qiuwei Li, Gongguo Tang, and Michael B. Wakin. Global optimality in low-rank matrix optimization. *IEEE Transactions on Signal Processing*, 66(13):3614–3628, 2018.