The Stability of Low-Rank Matrix Reconstruction: A Constrained Singular Value View

Gongguo Tang, *Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—The stability of low-rank matrix reconstruction with respect to noise is investigated in this paper. The ℓ_*-constrained minimal singular value (ℓ_*-CMSV) of the measurement operator is shown to determine the recovery performance of nuclear norm minimization-based algorithms. Compared with the stability results using the matrix restricted isometry constant, the performance bounds established using ℓ_*-CMSV are more concise, and their derivations are less complex. Isotropic and subgaussian measurement operators are shown to have ℓ_*-CMSVs bounded away from zero with high probability, as long as the number of measurements is relatively large. The ℓ_*-CMSV for correlated Gaussian operators are also analyzed and used to illustrate the advantage of ℓ_*-CMSV compared with the matrix restricted isometry constant. We also provide a fixed point characterization of ℓ_*-CMSV that is potentially useful for its computation.**

***Index Terms*—ℓ_*-constrained minimal singular value (CMSV), correlated design, matrix basis pursuit (mBP), matrix Dantzig selector (mDS), matrix LASSO estimator (mLASSO), restricted isometry property.**

I. Introduction

THE last decade witnessed the burgeoning of exploiting low-dimensional structures in signal processing, most notably the sparseness for vectors [1], [2], low-rankness for matrices [3]–[5], and low-dimensional manifold structure for general nonlinear datasets [6], [7]. This paper focuses on the stability problem of low-rank matrix reconstruction. Suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is a matrix of rank $r \ll \min\{n_1, n_2\}$; the low-rank matrix reconstruction problem aims at recovering matrix X from a set of linear measurements $\boldsymbol{y}$ corrupted by noise $\boldsymbol{w}$

$$\boldsymbol{y} = \mathcal{A}(X) + \boldsymbol{w} \tag{1}$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear measurement operator. Since the matrix X lies in a low-dimensional sub-manifold of $\mathbb{R}^{n_1 \times n_2}$, we expect $m \ll n_1 n_2$ measurements would suffice to reconstruct X from $\boldsymbol{y}$ by exploiting the signal structure. Application areas of model (1) include factor analysis, linear system realization [8], [9], matrix completion [10], [11], quantum state tomography [12], face recognition [13], [14], Euclidean embedding [15], to name a few (See [3]–[5] for discussions and references therein).

Several considerations motivate the study of the stability of low-rank matrix reconstruction. First, in practical problems, the linear measurement operator $\mathcal{A}$ is usually used repeatedly to collect measurement vectors $\boldsymbol{y}$ for different matrices X. Therefore, before taking the measurements, it is desirable to know the goodness of the measurement operator $\mathcal{A}$ as far as reconstructing X is concerned. Second, a stability analysis would offer means to quantify the confidence on the reconstructed matrix X, especially when there is no other ways to justify the correctness of the reconstructed signal.

In this study, we define the ℓ_*-constrained minimal singular value (ℓ_*-CMSV) of a linear operator to measure the stability of low-rank matrix reconstruction. By employing advanced tools from geometrical functional analysis and empirical processes, we show that a large class of random linear operators have ℓ_*-CMSVs bounded away from zero. We also derive a fixed point characterization of the ℓ_*-CMSV.

Several works in the literature also address the problem of low-rank matrix reconstruction. Recht *et al*. study the recovery of X in model (1) in the noiseless setting [3]. The matrix restricted isometry property (mRIP) is shown to guarantee exact recovery of X subject to the measurement constraint $\mathcal{A}(X) = \boldsymbol{y}$. Candés and Plan consider the noisy problem and analyze the reconstruction performance of several convex relaxation algorithms [5]. The techniques used in this paper for deriving the error bounds in terms of ℓ_*-CMSV draw ideas from [5]. In both works [3], [5], several important random measurement ensembles are shown to have the matrix restricted isometry constant (mRIC) close to zero for reasonably large m. Our procedures for establishing the parallel results for the ℓ_*-CMSV are significantly different from those in [3] and [5]. In particular, for correlated Gaussian operators, we show that the mRIC might fail with high probability, while the ℓ_*-CMSV is still well controlled.

The ℓ_*-CMSV has several advantages over the mRIC in stability analysis of low-rank matrix analysis. First, the error bounds involving ℓ_*-CMSV have more transparent relationships with the signal-to-noise ratio. For example, consider the matrix basis pursuit (mBP) algorithm; if we multiply the measurement operator $\mathcal{A}$ by a positive constant, the ℓ_*-CMSV will scale by the same constant and the error bound for the mBP will scale inverse proportionally, while the mRIC and

Manuscript received June 20, 2010; revised October 01, 2011; accepted March 23, 2012. Date of publication June 13, 2012; date of current version August 14, 2012. This work was supported in part by the Office of Naval Research under Grant N000140810849 and in part by the National Science Foundation under Grants CCF-1014908 and CCF-0963742.

G. Tang was with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA. He is now with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA (e-mail: gt2@ese.wustl.edu).

A. Nehorai is with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: nehorai@ese.wustl.edu).

Communicated by J. Romberg, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2012.2204535

associated error bounds have more complex scaling properties. Second, the ℓ_*-CMSV shows clear relations of low-rank matrix recovery with certain geometric properties of the nuclear ball, such as its mean width, Gaussian width, and the diameter of its sections. In addition, the derivation of the ℓ_*-CMSV bounds is less complicated and the resulting bounds have more concise forms. Last but not least, as shown by our probabilistic analysis for correlated Gaussian operators, the mRIC might fail, while the ℓ_*-CMSV is still meaningful.

This paper is organized as follows. Section II introduces notation, the measurement model, three convex relaxation-based recovery algorithms, and the definition and properties of the mRIC. Section III is devoted to deriving error bounds in terms of the ℓ_*-CMSV for three convex relaxation algorithms. In Section IV, we analyze the ℓ_*-CMSV for isotropic and subgaussian measurement operator, as well as correlated Gaussian operators. In Section V, we provide a fixed point characterization of the ℓ_*-CMSV. The paper is concluded in Section VI.

II. Notation, Measurement Model, Reconstruction Algorithms, and mRIC

A. Notation

We use bold lower case letters such as $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ to denote vectors whose ith components are represented by corresponding lower case letters x_i, y_i, and z_i, respectively. Subscripted bold lower case letters, e.g., $\boldsymbol{x}_i$, are reserved for vectors with subscript i. Matrices are denoted by upper case letters such as A, X, Z.

The ℓ_p norm $\|\cdot\|_p$ of $\boldsymbol{x} = [x_1, \ldots, x_m]^T \in \mathbb{R}^m$ is defined as

$$\|\boldsymbol{x}\|_p = (\sum_{k \leq m} |x_k|^p)^{1/p} \quad \text{for} \quad 1 \leq p < \infty \tag{2}$$

and

$$\|\boldsymbol{x}\|_\infty = \max_{k \leq m} |x_k|. \tag{3}$$

Suppose $X = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \ldots \ \boldsymbol{x}_{n_2}] \in \mathbb{R}^{n_1 \times n_2}$ is a matrix. Define the Frobenius norm of X as $\|X\|_{\mathrm{F}} = (\sum_{i,j} |X_{ij}|^2)^{1/2} = (\sum_i \sigma_i^2(X))^{1/2}$, the nuclear norm as $\|X\|_* = \sum_i \sigma_i(X)$, and the operator norm as $\|X\| = \max\{\sigma_i(X)\}$, where $\sigma_i(X)$ is the ith singular value of X in descending order. The rank and trace of X are denoted by $\mathrm{rank}(X)$ and $\mathrm{trace}(X)$, respectively. The inner product of two matrices $X_1, X_2 \in \mathbb{R}^{n_1 \times n_2}$ is defined as $\langle X_1, X_2 \rangle = \mathrm{trace}(X_1^T X_2)$. We use $\otimes$ to denote the Kronecker product, and $\mathrm{vec}(\cdot)$ to denote the vectorization of a matrix. A useful identity is $(B^T \otimes A)\mathrm{vec}(X) = \mathrm{vec}(AXB)$.

For any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$, its adjoint operator $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{R}^{n_1 \times n_2}$ is defined by the relation

$$\langle \mathcal{A}(X), \boldsymbol{z} \rangle = \langle X, \mathcal{A}^*(\boldsymbol{z}) \rangle, \forall X \in \mathbb{R}^{n_1 \times n_2}, \boldsymbol{z} \in \mathbb{R}^m. \tag{4}$$

A linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ can be represented by m matrices $\{A^1, A^2, \ldots, A^m\} \subset \mathbb{R}^{n_1 \times n_2}$ such that $\mathcal{A}(X) = \left[\langle A^1, X \rangle, \ldots, \langle A^m, X \rangle\right]^T$, or by a big matrix $A \in \mathbb{R}^{m \times n_1 n_2}$ whose kth row is $\mathrm{vec}(A^k)$ such that $\mathcal{A}(X) = A\mathrm{vec}(X)$. The adjoint operator is given by

$$\mathcal{A}^*(\boldsymbol{z}) = \sum_{k=1}^m z_k A^k \in \mathbb{R}^{n_1 \times n_2}. \tag{5}$$

Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. This notation is also generalized to Gaussian random vectors and matrix variate Gaussian distributions [16].

B. Measurement Model

Throughout this paper, we will assume $n_1 \leq n_2$. Suppose we have a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ with $\mathrm{rank}(X) = r \ll n_1$. We observe X through the following linear model:

$$\boldsymbol{y} = \mathcal{A}(X) + \boldsymbol{w} \tag{6}$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ is a linear operator and $\boldsymbol{w} \in \mathbb{R}^m$ is noise. Here, m is much less than $n_1 n_2$.

A fundamental problem pertaining to model (6) is to reconstruct the low-rank matrix X from the measurement $\boldsymbol{y}$ by exploiting the low-rank property of X, and the stability of the reconstruction with respect to noise. For any reconstruction algorithm, we denote the estimate of X as $\hat{X}$, and the error matrix $H \overset{\text{def}}{=} \hat{X} - X$. In this paper, the stability problem aims to bound $\|H\|_{\mathrm{F}}$ in terms of m, n_1, n_2, r, the linear operator $\mathcal{A}$, and the noise level.

C. Reconstruction Algorithms

We consider three low-rank matrix recovery algorithms based on convex relaxation: the mBP, the matrix Dantzig selector (mDS), and the matrix LASSO estimator (mLASSO).

The mBP algorithm [3], [5] minimizes the nuclear norm subject to bounded noise constraint

$$\text{mBP} : \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \text{ s.t. } \|\boldsymbol{y} - \mathcal{A}(Z)\|_2 \leq \varepsilon. \tag{7}$$

The mDS [5] reconstructs a low-rank matrix when its linear measurements are corrupted by unbounded noise. Its estimate for X is the solution to the nuclear norm regularization problem

$$\text{mDS} : \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \text{ s.t. } \|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(Z))\| \leq \mu. \tag{8}$$

The mLASSO solves the following optimization problem [5], [17]:

$$\text{mLASSO:} \quad \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(Z)\|_2^2 + \mu\|Z\|_*. \tag{9}$$

All three optimization problems can be solved using semidefinite programs.

D. mRIC

The reconstruction performance of the mBP, the mDS, and the mLASSO depends on the incoherence of the linear operator $\mathcal{A}$. A popular measure of incoherence is mRIC defined as follows [3], [5].

Definition 1: For each integer $r \in \{1, \dots, n_1\}$, the mRIC δ_r of a linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is defined as the smallest $\delta > 0$ such that

$$1 - \delta \le \frac{\|\mathcal{A}(X)\|_2^2}{\|X\|_{\mathrm{F}}^2} \le 1 + \delta \tag{10}$$

holds for arbitrary nonzero matrix X of rank at most r.

A linear operator $\mathcal{A}$ with a small δ_r roughly means that $\mathcal{A}$ is nearly an isometry when restricted onto all matrices with rank at most r. We cite stability results on the mBP, the mDS, and the mLASSO, which are expressed in terms of the mRIC. Assume X is of rank r and $\hat{X}$ is its estimate given by any of the three algorithms; then, we have the following.

1) mBP [5]: Suppose that $\delta_{4r} < \sqrt{2} - 1$ and $\|\boldsymbol{w}\|_2 \le \varepsilon$. The solution to the mBP (7) satisfies

$$\|\hat{X} - X\|_{\mathrm{F}} \le \frac{4\sqrt{1 + \delta_{4r}}}{1 - (1 + \sqrt{2})\delta_{4r}} \varepsilon. \tag{11}$$

2) mDS [5]: If $\delta_{4r} < \sqrt{2} - 1$ and $\|\mathcal{A}^*(\boldsymbol{w})\| \le \mu$, then

$$\|\hat{X} - X\|_{\mathrm{F}} \le \frac{16\sqrt{r}}{1 - (\sqrt{2} + 1)\delta_{4r}} \mu. \tag{12}$$

3) mLASSO [5]: If $\delta_{4r} < (3\sqrt{2} - 1)/17$ and $\|\mathcal{A}^*(\boldsymbol{w})\| \le \mu/2$, then the solution to the mLASSO (9) satisfies

$$\|\hat{X} - X\|_{\mathrm{F}} \le C\delta_{4r}\sqrt{r}\mu \tag{13}$$

for some numerical constant C.

III. Recovery Error Bounds

In this section, we derive bounds on the recovery errors of the mBP, the mDS, and mLASSO. We first characterize the recovery errors by showing that the effective ranks of the error matrices are small.

A. Error Characteristics

We introduce a quantity that continuously extends the concept of rank for a given matrix X.

Definition 2: The ℓ_*-rank of a nonzero matrix $X \in \mathbb{R}^{n_1 \times n_2}$ is defined as

$$\tau(X) = \frac{\|X\|_*^2}{\|X\|_{\mathrm{F}}^2}. \tag{14}$$

The function $\tau(X)$ is indeed a measure of the effective rank. To see this, suppose $\mathrm{rank}(X) = r$; then, Cauchy–Schwarz inequality implies that

$$\tau(X) \le r \tag{15}$$

and we have equality if and only if all nonzero singular values of X are equal. Therefore, the more nonzero singular values X has and the more evenly the magnitudes of these non-zero singular values are distributed, the larger $\tau(X)$. In particular, if X is of rank 1, then $\tau(X) = 1$; if X is of full rank n_1 with all singular values having the same magnitudes, then $\tau(X) = n_1$. However, if X has n_1 nonzero singular values but their magnitudes are spread in a wide range, then its ℓ_*-rank might be very small.

The following proposition, whose proof is given in Appendix A, shows that the error matrices have small ℓ_*-rank.

Proposition 1: Suppose X in (6) is of rank r and the noise $\boldsymbol{w}$ satisfies $\|\boldsymbol{w}\|_2 \le \varepsilon$, $\|\mathcal{A}^*(\boldsymbol{w})\| \le \mu$, and $\|\mathcal{A}^*(\boldsymbol{w})\| \le \kappa\mu, \kappa \in (0, 1)$, for the mBP, the mDS, and the mLASSO, respectively. Then, the error matrix $H = \hat{X} - X$ for any of the three recovery algorithms (7) – (9) satisfies

$$\tau(H) \le cr \tag{16}$$

where $c = 8$ for the mBP and the mDS, and $c = 8/(1 - \kappa)^2$ for the mLASSO.

B. ℓ_*-CMSV and Error Bounds

The reconstruction performance of the recovery algorithms should depend on the invertibility of the linear operator $\mathcal{A}$. Proposition 1 indicates that we could restrict ourselves to the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \tau(X) \le cr\}$ when quantifying the invertibility of $\mathcal{A}$.

Definition 3: For any $\tau \in [1, n_1]$ and any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$, define the ℓ_*-constrained minimal singular value (abbreviated as ℓ_*-CMSV) of $\mathcal{A}$ by

$$\rho_\tau(\mathcal{A}) := \inf_{X \ne 0,\, \tau(X) \le \tau} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}. \tag{17}$$

As pointed out by one reviewer, one difference between the ℓ_*-CMSV and the mRIC is that the ℓ_*-CMSV does not require upper bounds on the restricted eigenvalues of the operator $\mathcal{A}$. This is known to be true in the vector case (i.e., sparsity recovery) and is established using the notions such as *restricted eigenvalues* [18] and m-sparse minimal eigenvalues [19]. This paper shares the common observation with previous work on the vector case that the reconstruction performance of the recovery algorithms should depend on the invertibility of the measurement matrix or operator when restricted to the error set, which is usually much smaller than the signal's ambient space. We use the ℓ_*-rank to differentiate the error set and to define the invertibility of $\mathcal{A}$. Probability analysis in Section IV shows that at least for isotropic and subgaussian operators, the ℓ_*-rank characterization is as good as the null space property characterization [20]. We also establish that, for correlated Gaussian operators, the mRIC might not be valid with high probability, even when the ℓ_*-CMSV is still bounded away from zero.

Now, we present bounds on the error matrices for the mBP, the mDS, and the mLASSO. The proof is given in Appendix B.

Theorem 1: Under the assumption of Proposition 1, we have

$$\|\hat{X} - X\|_{\mathrm{F}} \le \frac{2\varepsilon}{\rho_{8r}(\mathcal{A})} \tag{18}$$

for the mBP

$$\|\hat{X} - X\|_{\mathrm{F}} \le \frac{4\sqrt{2r}}{\rho_{8r}^2(\mathcal{A})} \mu \tag{19}$$

for the mDS, and

$$\|\hat{X}-X\|_{\mathrm{F}} \leq \frac{1+\kappa}{1-\kappa}\frac{2\sqrt{2r}}{\rho^2_{\frac{8r}{(1-\kappa)^2}}(\mathcal{A})}\mu \tag{20}$$

for the mLASSO.

Compared with the error bounds (11) – (13), the bounds given in Theorem 1 are simpler and their derivations are easier. When the noise levels are zero, namely, $\varepsilon = 0$ and $\mu = 0$, roughly speaking all three nuclear norm minimization algorithms reduce to

$$\min_{Z\in\mathbb{R}^{n_1\times n_2}} \|Z\|_* \text{ subject to } \boldsymbol{y} = \mathcal{A}(Z). \tag{21}$$

According to Theorem 1, if $\rho_{8r}(\mathcal{A}) > 0$, then we get exact recovery in the noise-free case. Therefore, $\rho_{8r} > 0$ is a sufficient condition for exact low-rank matrix recovery using nuclear norm minimization.

We observe that $\rho_{8r}(\mathcal{A}) > 0$ is equivalent to

$$r < \frac{1}{8}\min\{\tau(Z) : \mathcal{A}(Z) = 0\} \text{ or}$$
$$\frac{1}{2\sqrt{2r}} > \max\{\|Z\|_{\mathrm{F}} : \mathcal{A}(Z) = 0, \|Z\|_* \leq 1\}. \tag{22}$$

We note that the right-hand side of (22) is the diameter of the set $B_*^{n_1\times n_2} \bigcap \mathrm{null}(\mathcal{A})$, where $B_*^{n_1\times n_2}$ is the unit nuclear ball and $\mathrm{null}(\mathcal{A})$ is the null space of the operator $\mathcal{A}$. If the null space $\mathrm{null}(\mathcal{A})$ is chosen uniformly according to the Haar measure on the Grassmannian $\mathcal{G}_{n_1n_2, n_1n_2-m}$ of (n_1n_2-m)-dimensional subspaces of $\mathbb{R}^{n_1\times n_2}$ (e.g., when the entries of $\{A_i\}_{i=1}^m$ follow *i.i.d.* Gaussian with zero mean and unit variance), then the low M^* estimate [21] implies that

$$\begin{aligned}
&\mathrm{diam}(B_*^{n_1\times n_2} \textstyle\bigcap \mathrm{null}(\mathcal{A}))\\
&:= \max_{Z:Z\in B_*^{n_1\times n_2}\bigcap \mathrm{null}(\mathcal{A})} \|Z\|_{\mathrm{F}}\\
&\leq c\sqrt{\frac{n_1n_2}{m}} M^*(B_*^{n_1\times n_2})\\
&:= c\sqrt{\frac{n_1n_2}{m}}\int_{\mathbb{S}^{n_1n_2-1}} \|Z\|\mathrm{d}\sigma(Z)
\end{aligned} \tag{23}$$

with probability at least $1-e^{-m}$. Here, $\mathbb{S}^{n_1n_2-1}$ is the unit Euclidean sphere in $\mathbb{R}^{n_1\times n_2}$, $\sigma(\cdot)$ is the Haar measure on $\mathbb{S}^{n_1n_2-1}$, and c is a numerical constant.

According to Poincaré's lemma [22], the uniform measures on n-dimensional spheres with radius $\sqrt{n}$ approximate Gaussian measures. As a consequence, we have

$$\begin{aligned}
&M^*(B_*^{n_1\times n_2})\\
&= \frac{1}{\sqrt{n_1n_2}}\int_{\sqrt{n_1n_2}\mathbb{S}^{n_1n_2-1}} \|Z\|\mathrm{d}\sqrt{n_1n_2}\sigma(Z)\\
&\sim \frac{1}{\sqrt{n_1n_2}}\mathbb{E}\|Z\| \leq 2\sqrt{\frac{n_2}{n_1n_2}}.
\end{aligned} \tag{24}$$

Here, $\mathbb{E}$ is taken with respect to the canonical Gaussian measure in $\mathbb{R}^{n_1n_2}$, and the last inequality, which gives an upper bound on the expected largest singular value of a rectangular Gaussian matrix, is due to Slepian's lemma [23], [24]. Therefore, we obtain

$$\mathrm{diam}(B_*^{n_1\times n_2} \textstyle\bigcap \mathrm{null}(\mathcal{A})) \leq c\sqrt{\frac{n_2}{m}} \tag{25}$$

with high probability. Combining with the sufficient condition (22), we obtain the following corollary.

Corollary 1: If the null space of $\mathcal{A}$ follows uniform distribution on all subspaces of dimension n_1n_2-m and

$$m \geq cn_2r \tag{26}$$

then with probability greater than $1-e^{-m}$, we can recover any matrix X of rank less than r.

In the next section, we will directly analyze the probabilistic behavior of $\rho_\tau(\mathcal{A})$ and obtain Corollary 1 as a consequence.

IV. Probabilistic Analysis

This section is devoted to analyzing the properties of the ℓ_*-CMSVs for several important random operator ensembles. Although the bounds in Theorem 1 have concise forms, they are useless if the quantity involved, ρ_τ, is zero or approaches zero for most matrices as n_1, n_2, m, r vary in a reasonable manner. We show that for a large class of random linear operators, including both isotropic and subgaussian operators and correlated Gaussian operators, the ℓ_*-CMSVs are bounded away from zero with high probability.

A. Isotropic and Subgaussian Operators

We begin by defining the isotropic and subgaussian ensemble, after introducing some notations. For a scalar random variable a, the Orlicz ψ_2 norm [25, Sec. 4.1, p. 92] is defined as

$$\|a\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E}\exp\left(\frac{|a|^2}{t^2}\right) \leq 2\right\}. \tag{27}$$

Markov's inequality immediately gives that a with finite $\|a\|_{\psi_2}$ has a subgaussian tail

$$\mathbb{P}(|a| \geq t) \leq 2\exp(-ct^2/\|a\|_{\psi_2}). \tag{28}$$

The converse is also true, i.e., if a has subgaussian tail $\exp(-t^2/K^2)$, then $\|a\|_{\psi_2} \leq cK$. A random vector $\boldsymbol{a} \in \mathbb{R}^n$ is called isotropic and subgaussian with constant L if $\mathbb{E}|\langle \boldsymbol{a}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\|\langle \boldsymbol{a}, \boldsymbol{u}\rangle\|_{\psi_2} \leq L\|\boldsymbol{u}\|_2$ hold for any $\boldsymbol{u} \in \mathbb{R}^n$.

Recall that a linear operator $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^m$ can be represented by a collection of matrices $\{A^1, \ldots, A^m\}$. Based on this representation of $\mathcal{A}$, we have the following definition of isotropic and subgaussian operators.

Definition 4: Suppose $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^m$ is a linear operator with corresponding matrix representation $\{A^i\}_{i=1}^m$. We say $\mathcal{A}$ is from the isotropic and subgaussian ensemble if $\{A^i\}_{i=1}^m$ are independent isotropic and subgaussian vector with constant L, where L is a numerical constant independent of n_1, n_2.

Isotropic and subgaussian operators include operators with i.i.d. centered subgaussian entries of unit variance (Gaussian and Bernoulli entries in particular) as well as operators whose matrices A_i ($\mathrm{vec}(A_i)$, more precisely) are independent copies of random vectors distributed according to the normalized volume measure of unit balls of $(\mathbb{R}^{n_1n_2}, \|\cdot\|_p)$ for $2 \leq p \leq \infty$.

An important concept in studying empirical processes of isotropic and subgaussian random vectors is the Gaussian width defined as follows.

Definition 5: Let $\mathcal{H} \subset \mathbb{R}^n$ and the components of $\boldsymbol{g}$ follow i.i.d. Gaussian with mean zero and variance one, i.e., $\boldsymbol{g} \sim \mathcal{N}(0, \mathrm{I}_n)$. Denote by $w(\mathcal{H}) = \mathbb{E}\ \sup_{\boldsymbol{u}\in\mathcal{H}} \langle \boldsymbol{g}, \boldsymbol{u}\rangle$.

With these preparations, we combine [26, Th. D] and the discussion following it to present.

Theorem 2 [26, Th. D]: Let $\{\boldsymbol{a}, \boldsymbol{a}_i, i = 1, \ldots, m\} \subset \mathbb{R}^n$ be i.i.d. isotropic and subgaussian random vectors, $\mathcal{H}$ be a subset of the unit sphere of $\mathbb{R}^n$, and $\mathcal{F} = \{f_{\boldsymbol{u}}(\cdot) = \langle \boldsymbol{u}, \cdot\rangle : \boldsymbol{u} \in \mathcal{H}\}$. Suppose $\mathrm{diam}(\mathcal{F}, \|\cdot\|_{\psi_2}) = \max_{f,g\in\mathcal{F}} \|f - g\|_{\psi_2} = \alpha$. Then, there exist absolute constants c_1, c_2, c_3 such that for any $\epsilon > 0$ and $m \geq 1$ satisfying

$$m \geq c_1 \frac{\alpha^2 w^2(\mathcal{H})}{\epsilon^2}, \tag{29}$$

with probability at least $1 - \exp(-c_2\epsilon^2 m/\alpha^4)$, we have

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{m}\sum_{k=1}^{m} f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right| \leq \epsilon. \tag{30}$$

Furthermore, if $\mathcal{F}$ is symmetric, we have

$$\begin{aligned} &\mathbb{E}\sup_{f\in\mathcal{F}} \left| \frac{1}{m}\sum_{k=1}^{m} f^2(\boldsymbol{a}_k) - \mathbb{E}f^2(\boldsymbol{a})\right| \\ &\leq c_3 \max\left\{\alpha \frac{w(\mathcal{H})}{\sqrt{m}}, \frac{w^2(\mathcal{H})}{m}\right\}. \end{aligned} \tag{31}$$

Using Theorem 2, we show that for any isotropic and subgaussian operator $\sqrt{m}\mathcal{A}$, the typical value of $\rho_\tau(\mathcal{A})$ concentrates around 1 for relatively large m (but $\ll n_1n_2$). More precisely, we have the following theorem.

Theorem 3: Let $\sqrt{m}\mathcal{A}$ be an isotropic and subgaussian operator with some numerical constant L. Then, there exist absolute constants c_1, c_2 depending on L only,such that for any $\epsilon > 0$ and $m \geq 1$ satisfying

$$m \geq c_1 \frac{\tau n_2}{\epsilon^2}, \tag{32}$$

we have

$$\mathbb{E}|\rho_\tau^2(\mathcal{A}) - 1| \leq \epsilon \tag{33}$$

and

$$\mathbb{P}\left\{|\rho_\tau^2(\mathcal{A}) - 1| \leq \epsilon\right\} \geq 1 - \exp(-c_2\epsilon^2 m). \tag{34}$$

Proof of Theorem 3: Since the linear operator $\mathcal{A}$ is generated in a way such that $\mathbb{E}\left\langle \sqrt{m}A^k, X\right\rangle^2 = \|X\|_{\mathrm{F}}^2$ for any $X \in \mathbb{R}^{n_1\times n_2}$, we have $|\rho_\tau^2(\mathcal{A}) - 1| < 1 - \epsilon$ is a consequence of

$$\begin{aligned} &\sup_{X\in\mathcal{H}_\tau} \left|\mathcal{A}(X)^T\mathcal{A}(X) - 1\right| \\ &= \sup_{X\in\mathcal{H}_\tau} \left|\frac{1}{m}\sum_{k=1}^{m}\left\langle \sqrt{m}A^k, X\right\rangle^2 - 1\right| \leq \epsilon. \end{aligned} \tag{35}$$

Here, the set

$$\mathcal{H}_\tau = \{X \in \mathbb{R}^{n_1\times n_2} : \|X\|_{\mathrm{F}} = 1,\ \|X\|_*^2 \leq \tau\}. \tag{36}$$

As usual, the operator $\mathcal{A}$ is represented by a collection of matrices $\{A^1, \ldots, A^m\}$. We define a class of functions parameterized by X as $\mathcal{F}_\tau := \{f_X(\cdot) = \langle X, \cdot\rangle : X \in \mathcal{H}_\tau\}$.

It remains to compute $w(\mathcal{H}_\tau)$ as follows:

$$\begin{aligned} w(\mathcal{H}_\tau) &= \mathbb{E}\sup_{X\in\mathcal{H}_\tau} \langle G, X\rangle \\ &\leq c\,\|X\|_*\ \mathbb{E}\,\|G\|_2 \\ &\leq c\,\sqrt{\tau}\sqrt{n_2} \end{aligned} \tag{37}$$

where G is a Gaussian matrix with i.i.d. entries from $\mathcal{N}(0, 1)$. Again, we have used an upper bound on the expected largest singular value of a rectangular Gaussian matrix due to Slepian's lemma [25, Ch. 3.1]. As a consequence, the conclusions of Theorem 3 hold. ■

If we take $\epsilon = 1/2$ and $\tau = 8r$ in Theorem 3, we get Corollary 1 as a consequence. The bound $m = \Omega(rn_2)$ is the same as the one obtained for the mRIC. Thus, the ℓ_*-CMSV is as good as the mRIC for isotropic and subgaussian operators.

B. Correlated Gaussian Operators

In this section, we consider Gaussian measurement operators with a correlation structure. Correlated sensing matrices are considered in [27] and [28] in the context of compressive sensing. For low-rank matrix recovery, correlated measurement operators are potentially useful for multivariate regression and vector autoregressive processes [29].

Suppose that the entries of $G^k \in \mathbb{R}^{n_1\times n_2}, k = 1, \ldots, m$ follow i.i.d. Gaussian distribution $\mathcal{N}\left(0, \frac{1}{m}\right)$, and $\Sigma_1 \in \mathbb{R}^{n_1\times n_1}$ and $\Sigma_2 \in \mathbb{R}^{n_2\times n_2}$ are positive semidefinite matrices. Then, $\Sigma_1^{1/2}G^k\Sigma_2^{1/2}, k = 1, \ldots, m$ follow i.i.d. matrix variate Gaussian distribution $\mathcal{N}\left(0, \frac{1}{m}\Sigma_2 \otimes \Sigma_1\right)$, namely, $\mathrm{vec}\left(\Sigma_1^{1/2}G^k\Sigma_2^{1/2}\right) \sim \mathcal{N}\left(0, \frac{1}{m}\Sigma_2 \otimes \Sigma_1\right)$ [16]. Here, $\Sigma^{1/2}$ denotes the matrix square root of a positive semidefinite matrix Σ. We call the linear operator $\mathcal{A}_{\Sigma_1,\Sigma_2}$ represented by $\left\{\Sigma_1^{1/2}G^k\Sigma_2^{1/2}, k = 1, \ldots, m\right\}$ a correlated Gaussian measurement operator.

The following theorem shows that $\rho_\tau\left(\mathcal{A}_{\Sigma_1,\Sigma_2}\right)$ is controlled by $\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)$ with high probability.

Theorem 4: Suppose $\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right) > 0$. Then, there exist universal positive constants c, c_1, c_2 such that if

$$m \geq c\,\frac{\mathrm{trace}\left(\Sigma_1\right) + \mathrm{trace}\left(\Sigma_2\right)}{\rho_\tau^2\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)}\tau \tag{38}$$

then the linear operator $\mathcal{A}_{\Sigma_1,\Sigma_2}$ has ℓ_*-CMSV

$$\rho_\tau\left(\mathcal{A}_{\Sigma_1,\Sigma_2}\right) \geq \frac{1}{8}\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right) \tag{39}$$

with probability at least $1 - c_1\exp\left(-c_2 m\right)$.

Remark 1: When $\Sigma_1 = I_{n_1}$ and $\Sigma_2 = I_{n_2}$, clearly we have $\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right) = \rho_\tau\left(I_{n_1n_2}\right) = 1$, and $\mathrm{trace}\left(\Sigma_1\right) +$

$\operatorname{trace}(\Sigma_2) = n_1 + n_2$. So, Theorem 4 is consistent with Theorem 3 applied to Gaussian measurement operators.

Remark 2: In this remark, we provide an example where the ℓ_*-CMSV is well controlled, while the mRIP fails with high probability. Refer to [27] for more examples constructed in the similar compressive sensing setting. For simplicity, we set $n_1 = n_2 = n$. Consider $\Sigma_1 = \Sigma_2 = \Sigma = (1-a) I_n + a\mathbf{1}\mathbf{1}^T$ for some fixed $a \in (0,1)$. Here, $\mathbf{1} \in \mathbb{R}^n$ is the vector with all ones. Clearly, we have $\operatorname{trace}(\Sigma) = n$. The inequality $\rho_\tau^2\left(\Sigma^{1/2} \otimes \Sigma^{1/2}\right) \geq \lambda_{\min}(\Sigma \otimes \Sigma) = \lambda_{\min}^2(\Sigma) = (1-a)^2$ together with Theorem 4 imply that $\rho_\tau\left(\mathcal{A}_{\Sigma_1,\Sigma_2}\right) \geq \frac{1}{8}(1-a)$ with high probability as long as $m \geq c\frac{n\tau}{(1-a)^2}$.

However, $\Sigma^{1/2} \otimes \Sigma^{1/2}$ does not satisfy the mRIC when n is large. To see this, assume that the eigendecomposition of Σ is $\sum_{j=1}^n \sigma_j \boldsymbol{u}_j \boldsymbol{u}_j^T$ and observe that when $X = \sum_{i=1}^r \lambda_i \boldsymbol{u}_{j_i} \boldsymbol{u}_{j_i}^T$, we have

$$\begin{aligned} &\frac{\left\|\left(\Sigma^{1/2} \otimes \Sigma^{1/2}\right) \operatorname{vec}(X)\right\|_2^2}{\|X\|_{\mathrm{F}}^2} \\ &= \frac{\left\|\Sigma^{1/2} X \Sigma^{1/2}\right\|_{\mathrm{F}}^2}{\|X\|_{\mathrm{F}}^2} = \frac{\sum_{i=1}^r \sigma_{j_i}^2 \lambda_i^2}{\sum_{i=1}^r \lambda_i^2}. \end{aligned} \tag{40}$$

Taking supremum and infimum leads to

$$\begin{aligned} &\sup_{X:\operatorname{rank}(X)\leq r} \frac{\left\|\left(\Sigma^{1/2} \otimes \Sigma^{1/2}\right) \operatorname{vec}(X)\right\|_2^2}{\|X\|_{\mathrm{F}}^2} \\ &\geq \sup_{\lambda_i} \frac{\sum_{i=1}^r \sigma_{j_i}^2 \lambda_i^2}{\sum_{i=1}^r \lambda_i^2} = \max\left\{\sigma_i^2\right\} \\ &= ((1-a)+na)^2 \end{aligned} \tag{41}$$

$$\begin{aligned} &\inf_{X:\operatorname{rank}(X)\leq r} \frac{\left\|\left(\Sigma^{1/2} \otimes \Sigma^{1/2}\right) \operatorname{vec}(X)\right\|_2^2}{\|X\|_{\mathrm{F}}^2} \\ &\leq \inf_{\lambda_i} \frac{\sum_{i=1}^r \sigma_{j_i}^2 \lambda_i^2}{\sum_{i=1}^r \lambda_i^2} = \min\left\{\sigma_i^2\right\} \\ &= (1-a)^2. \end{aligned} \tag{42}$$

Therefore, independent of the rank parameter, $\Sigma^{1/2} \otimes \Sigma^{1/2}$ does not satisfy the mRIC when n is large. A large deviation argument similar to the one given in [27] shows that the same statement is true with high probability for $\mathcal{A}_{\Sigma_1,\Sigma_2}$. Hence, the mRIP might be violated with high probability for correlated Gaussian operators, while the ℓ_*-CMSV is still well controlled. Roughly speaking, the problem with mRIC is that it involves the maximal eigenvalue while only the minimal eigenvalue is essential for low-rank matrix recovery.

The proof of Theorem 4 is based on the following proposition.

Proposition 2: If each matrix A^k associated with $\mathcal{A}_{\Sigma_1,\Sigma_2}$ are i.i.d. random matrices following $\mathcal{N}\left(0, \frac{1}{m}\Sigma_2 \otimes \Sigma_1\right)$, then there exist universal constants c, c_1 such that for all $X \in \mathbb{R}^{n_1 \times n_2}$

$$\begin{aligned} \left\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\right\|_2 \geq{}& \frac{1}{4}\left\|\Sigma_1^{1/2} X \Sigma_2^{1/2}\right\|_{\mathrm{F}} \\ &- 3\frac{\sqrt{\operatorname{trace}(\Sigma_1)} + \sqrt{\operatorname{trace}(\Sigma_2)}}{\sqrt{m}}\|X\|_* \end{aligned} \tag{43}$$

with probability at least $1 - c_1 \exp(-cm)$.

Proof of Theorem 4: By the definition of ℓ_*-CMSV, for any X with $\tau(X) \leq \tau$, we have

$$\begin{aligned} \left\|\Sigma_1^{1/2} X \Sigma_2^{1/2}\right\|_{\mathrm{F}} &= \left\|\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right) \operatorname{vec}(X)\right\|_2 \\ &\geq \rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)\|X\|_{\mathrm{F}}. \end{aligned} \tag{44}$$

As a consequence of Proposition 2, we obtain

$$\begin{aligned} &\left\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\right\|_2 \\ &\geq \frac{1}{4}\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)\|X\|_{\mathrm{F}} \\ &\qquad - 3\frac{\sqrt{\operatorname{trace}(\Sigma_1)} + \sqrt{\operatorname{trace}(\Sigma_2)}}{\sqrt{m}}\|X\|_* \\ &\geq \frac{1}{4}\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)\|X\|_{\mathrm{F}} \\ &\qquad - 3\frac{\sqrt{\operatorname{trace}(\Sigma_1)} + \sqrt{\operatorname{trace}(\Sigma_2)}}{\sqrt{m}}\sqrt{\tau}\|X\|_{\mathrm{F}}. \end{aligned} \tag{45}$$

The sample size condition

$$m \geq c\frac{\operatorname{trace}(\Sigma_1) + \operatorname{trace}(\Sigma_2)}{\rho_\tau^2\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)}\tau \tag{46}$$

with $c = 2 \times 24^2$ then leads to

$$\left\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\right\|_2 \geq \frac{1}{8}\rho_\tau\left(\Sigma_2^{1/2} \otimes \Sigma_1^{1/2}\right)\|X\|_{\mathrm{F}} \tag{47}$$

for all X such that $\tau(X) \leq \tau$, yielding the desired result. ■

The proof of Proposition 2 uses similar techniques developed in [27] for correlated Gaussian design in the compressive sensing setting. More specifically, we first show that (43) is true on the set $V(r) = \left\{X \in \mathbb{R}^{n_1 \times n_2} : \left\|\Sigma_1^{1/2} X \Sigma_2^{1/2}\right\|_{\mathrm{F}} = 1, \|X\|_* \leq r\right\}$ for fixed $r > 0$ with high probability. A peeling argument is then used to extend the result to all $X \in \mathbb{R}^{n_1 \times n_2}$. Refer to Appendix D for more details.

V. Fixed Point Characterization

In this section, we derive a fixed point characterization of $\rho_\tau(\mathcal{A})$. Recall that the optimization problem defining ρ_τ is as follows:

$$\rho_\tau(\mathcal{A}) = \min_Z \frac{\|\mathcal{A}(Z)\|_2}{\|Z\|_{\mathrm{F}}} \text{ s.t. } \frac{\|Z\|_*}{\|Z\|_{\mathrm{F}}} \leq \sqrt{\tau} \tag{48}$$

or equivalently

$$\frac{1}{\rho_\tau(\mathcal{A})} = \max_Z\{\|Z\|_{\mathrm{F}} : \|\mathcal{A}(Z)\|_2 \leq 1, \frac{\|Z\|_*}{\|Z\|_{\mathrm{F}}} \leq \sqrt{\tau}\}. \tag{49}$$

Denote by $\tau^* = \min_{Z:\mathcal{A}(Z)=0} \tau(Z)$. For any $\tau \in (1, \tau^*)$, we define functions over $[0, \infty)$ parameterized by τ

$$f_\tau(\eta; Y) = \max_Z\left\{\langle Y, Z\rangle : \|\mathcal{A}(Z)\|_2 \leq 1, \|Z\|_* \leq \sqrt{\tau}\eta\right\} \tag{50}$$

for $Y \in \mathbb{S}^{n_1 n_2 - 1}$ and

$$\begin{aligned} f_\tau(\eta) &= \max_{Z} \left\{ \|Z\|_{\mathrm{F}} : \|\mathcal{A}(Z)\|_2 \le 1, \|Z\|_* \le \sqrt{\tau}\eta \right\} \\ &= \max_{Y,Z} \{ \langle Y, Z\rangle : \|\mathcal{A}(Z)\|_2 \le 1, \|Z\|_* \le \sqrt{\tau}\eta \\ &\qquad \|Y\|_{\mathrm{F}} \le 1 \} \\ &= \sup_{Y \in \mathbb{S}^{n_1 n_2 - 1}} f_\tau(\eta; Y). \end{aligned} \tag{51}$$

Here, $\mathbb{S}^{n_1 n_2 - 1}$ is the unit sphere in $(\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_{\mathrm{F}})$. The continuity of $f_\tau(\eta; Y)$ with respect to Y, as established in Theorem 5, implies that the supremum in (51) can be replaced by maximum. In the definition of $f_\tau(\eta)$, we basically replaced the $\|Z\|_{\mathrm{F}}$ in the denominator of the fractional constraint in (49) with η.

For $\eta > 0$, it is easy to show that strong duality holds for the optimization problem defining $f_\tau(\eta; Y)$. As a consequence, we have the dual form of $f_\tau(\eta; Y)$

$$f_\tau(\eta; Y) = \min_{\lambda} \sqrt{\tau}\eta \|Y - \mathcal{A}^*(\lambda)\| + \|\lambda\|_2. \tag{52}$$

It turns out that the unique positive fixed point of $f_\tau(\eta)$ is exactly $1/\rho_\tau(\mathcal{A})$, as shown by the following theorem. See Appendix C for the proof.

Theorem 5: The functions $f_\tau(\eta; Y)$ and $f_\tau(\eta)$ have the following properties:

1) $f_\tau(\eta; Y)$ and $f_\tau(\eta)$ are jointly continuous in τ, η, and Y.
2) $f_\tau(\eta; Y)$ and $f_\tau(\eta)$ are strictly increasing in η.
3) $f_\tau(\eta; Y)$ is concave for each $Y \in \mathbb{S}^{n_1 n_2 - 1}$.
4) $f_\tau(0) = 0$, $f_\tau(\eta) \ge s\eta > \eta$ for sufficiently small $\eta > 0$, and there exists $\rho < 1$ such that $f_\tau(\eta) < \rho\eta$ for sufficiently large η; the same holds for $f_\tau(\eta; Y)$ if $Y = \boldsymbol{u}\boldsymbol{v}^T$ with $\|\boldsymbol{u}\|_2 = 1$ and $\|\boldsymbol{v}\|_2 = 1$, and the existence of ρ holds for all $f_\tau(\eta; Y)$.
5) $f_\tau(\eta; Y)$ has unique positive fixed point for $Y = \boldsymbol{u}\boldsymbol{v}^T$ with $\|\boldsymbol{u}\|_2 = 1$ and $\|\boldsymbol{v}\|_2 = 1$; $f_\tau(\eta)$ has at least one positive fixed points
$$\eta^* = \max \bigcup_{Y} \{\eta : \eta = f_\tau(\eta; Y)\}. \tag{53}$$
6) The positive fixed point η^* of $f_\tau(\eta)$ is unique and satisfies
$$\eta^* = \frac{1}{\rho_\tau(\mathcal{A})}. \tag{54}$$
7) For $\eta \in (0, \eta^*)$, we have $f_\tau(\eta) > \eta$; and for $\eta \in (\eta^*, \infty)$, we have $f_\tau(\eta) < \eta$; the same statement holds also for $f_\tau(\eta; Y)$ if $Y = \boldsymbol{u}\boldsymbol{v}^T$ with $\|\boldsymbol{u}\|_2 = 1$ and $\|\boldsymbol{v}\|_2 = 1$.
8) For any $\epsilon > 0$, there exists $\rho_1(\epsilon) > 1$ such that $f_\tau(\eta) > \rho_1(\epsilon)\eta$ as long as $0 < \eta \le (1-\epsilon)\eta^*$, and there exists $\rho_2(\epsilon) < 1$ such that $f_\tau(\eta) < \rho_2(\epsilon)\eta$ as long as $\eta > (1+\epsilon)\eta^*$.

For sparsity recovery and block-sparsity recovery, the fixed point characterizations yield efficient algorithms to compute certain incoherence measures [30], [31]. We develop the fixed point characterization in this paper in the hope that it might also lead to a way to compute $\rho_\tau(\mathcal{A})$. However, at this point, it is not clear how to compute or approximate $f_\tau(\eta)$ efficiently at a particular η.

VI. Conclusion

In this paper, the ℓ_*-CMSV of a measurement operator, which measures the invertibility of the measurement operator restricted to matrices with small ℓ_*-rank, is proposed to quantify the stability of low-rank matrix reconstruction. The reconstruction errors of the mBP, the mDS, and the mLASSO are concisely bounded using the ℓ_*-CMSV. We demonstrate that the ℓ_*-CMSV is bounded away from zero with high probability for isotropic and subgaussian measurement operators, as long as the number of measurements is relatively large. We also show that for correlated Gaussian operator, the ℓ_*-CMSV is lower bounded by that of its covariance matrix. Finally, we derive a fixed point characterization that is potentially useful for computing ℓ_*-CMSV.

In the future work, we will design algorithms to efficiently compute or approximate the ℓ_*-CMSV using the fixed point characterization. We also plan to extend the result for correlated Gaussian operator to subgaussian operators with correlation structure.

Appendix A
Proof of Proposition 1

We need two lemmas about the properties of nuclear norms derived in [3]

Lemma 1 [3, Lemma 2.3]: Let A and B be matrices of the same dimensions. If $AB^T = 0$ and $A^T B = 0$, then $\|A+B\|_* = \|A\|_* + \|B\|_*$.

Lemma 2 [3, Lemma 3.4]: Let A and B be matrices of the same dimensions. Then, there exist matrices B_1 and B_2 such that:

1) $B = B_1 + B_2$;
2) $\mathrm{rank}(B_1) \le 2\mathrm{rank}(A)$;
3) $AB_2^T = 0$ and $A^T B_2 = 0$;
4) $\langle B_1, B_2\rangle = 0$.

Proof of Proposition 1: We first deal with the mBP and the mDS. We decompose the error matrix $B = H$ according to Lemma 2 with $A = X$; more explicitly, we have the following:

1) $H = H_0 + H_c$;
2) $\mathrm{rank}(H_0) \le 2\mathrm{rank}(X) = 2r$;
3) $XH_c^T = 0$ and $X^T H_c = 0$;
4) $\langle H_0, H_c\rangle = 0$.

As observed by Recht *et al.* in [3] (see also [5], [32], and [33]), the fact that $\|\hat{X}\|_* = \|X + H\|_*$ is the minimum among all Zs satisfying the constraint in (7) implies that $\|H_c\|_*$ cannot be very large. To see this, we observe that

$$\begin{aligned} \|X\|_* &\ge \|X + H\|_* \\ &= \|X + H_c + H_0\|_* \\ &\ge \|X + H_c\|_* - \|H_0\|_* \\ &= \|X\|_* + \|H_c\|_* - \|H_0\|_*. \end{aligned} \tag{55}$$

Here, for the last equality, we used Lemma 1 and $XH_c^T = 0, X^T H_c = 0$. Therefore, we obtain

$$\|H_c\|_* \leq \|H_0\|_* \tag{56}$$

which leads to

$$\begin{aligned}\|H\|_* &\leq \|H_0\|_* + \|H_c\|_* \\ &\leq 2\|H_0\|_* \\ &\leq 2\sqrt{\mathrm{rank}(H_0)}\|H_0\|_\mathrm{F} \\ &= 2\sqrt{2r}\|H\|_\mathrm{F}\end{aligned} \tag{57}$$

where for the next to the last inequality we used the fact that $\|H\|_* \leq \sqrt{\mathrm{rank}(H)}\|H\|_\mathrm{F}$, and for the last inequality we used the Pythagorean theorem $\|H\|_\mathrm{F}^2 = \|H_0\|_\mathrm{F}^2 + \|H_c\|_\mathrm{F}^2 \geq \|H_0\|_\mathrm{F}^2$ because $\langle H_0, H_c\rangle = 0$. Inequality (57) is equivalent to

$$\tau(H) \leq 8\ \mathrm{rank}(X) = 8r. \tag{58}$$

We now turn to the LASSO estimator (9) . Suppose the noise $\boldsymbol{w}$ satisfies $\|\mathcal{A}^*(\boldsymbol{w})\| \leq \kappa\mu$ for some small $\kappa > 0$. Because $\hat{X}$ is a solution to (9), we have

$$\begin{aligned}&\frac{1}{2}\|\mathcal{A}(\hat{X}) - \boldsymbol{y}\|_2^2 + \mu\|\hat{X}\|_* \\ &\leq \frac{1}{2}\|\mathcal{A}(X) - \boldsymbol{y}\|_2^2 + \mu\|X\|_*.\end{aligned} \tag{59}$$

Consequently, substituting $\boldsymbol{y} = \mathcal{A}(X) + \boldsymbol{w}$ yields

$$\begin{aligned}\mu\|\hat{X}\|_* &\leq \left\langle \mathcal{A}(\hat{X} - X), \boldsymbol{w}\right\rangle + \mu\|X\|_* \\ &= \left\langle \hat{X} - X, \mathcal{A}^*(\boldsymbol{w})\right\rangle + \mu\|X\|_*.\end{aligned} \tag{60}$$

Using the Cauchy–Schwarz type inequality, we get

$$\begin{aligned}\mu\|\hat{X}\|_* &\leq \|\hat{X} - X\|_*\|\mathcal{A}^*(\boldsymbol{w})\| + \mu\|X\|_* \\ &= \kappa\mu\|H\|_* + \mu\|X\|_*\end{aligned} \tag{61}$$

which leads to

$$\|\hat{X}\|_* \leq \kappa\|H\|_* + \|X\|_*. \tag{62}$$

Therefore, similar to the argument in (55), we have

$$\|X\|_* \geq \|X\|_* + (1-\kappa)\|H_c\|_* - (1+\kappa)\|H_0\|_*. \tag{63}$$

Consequently, we have

$$\|H_c\|_* \leq \frac{1+\kappa}{1-\kappa}\|H_0\|_* \tag{64}$$

an inequality slightly worse than (56) for small κ. Therefore, an argument similar to the one leading to (57) yields

$$\|H\|_* \leq \frac{2}{1-\kappa}\sqrt{2r}\|H\|_\mathrm{F} \tag{65}$$

or equivalently

$$\tau(H) \leq \frac{8r}{(1-\kappa)^2}. \tag{66}$$

■

Appendix B
Proof of Theorem 1

Proof: To prove Theorem 1, we only need to obtain upper bounds on $\|\mathcal{A}(H)\|_2$ and then invoke the definition of the ℓ_*-CMSV. For mBP (7), this is trivial as both X and $\hat{X}$ satisfy constraint $\|\boldsymbol{y} - \mathcal{A}(Z)\| \leq \varepsilon$ in (7) . Therefore, the triangle inequality yields

$$\begin{aligned}\|\mathcal{A}(H)\|_2 &= \|\mathcal{A}(\hat{X} - X)\|_2 \\ &\leq \|\mathcal{A}(\hat{X}) - \boldsymbol{y}\|_2 + \|\boldsymbol{y} - \mathcal{A}(X)\|_2 \\ &\leq 2\varepsilon.\end{aligned} \tag{67}$$

It then follows from Definition 3 that

$$\rho_{8r}\|H\|_\mathrm{F} \leq \|\mathcal{A}(H)\|_2 \leq 2\varepsilon. \tag{68}$$

Hence, we get

$$\|\hat{X} - X\|_\mathrm{F} \leq \frac{2\varepsilon}{\rho_{8r}}. \tag{69}$$

For the mDS (8), the condition $\|\mathcal{A}^*(\boldsymbol{w})\| \leq \mu$ and the constraint in (8) yield

$$\|\mathcal{A}^*(\mathcal{A}(H))\| \leq 2\mu \tag{70}$$

because

$$\begin{aligned}\mathcal{A}^*(\boldsymbol{w} - \hat{\boldsymbol{r}}) &= \mathcal{A}^*\left((\boldsymbol{y} - \mathcal{A}(X)) - (\boldsymbol{y} - \mathcal{A}(\hat{X}))\right) \\ &= \mathcal{A}^*\left(\mathcal{A}(\hat{X}) - \mathcal{A}(X)\right) \\ &= \mathcal{A}^*(\mathcal{A}(H))\end{aligned} \tag{71}$$

where $\hat{\boldsymbol{r}} = \boldsymbol{y} - \mathcal{A}(\hat{X})$ is the residual corresponding to the mDS solution $\hat{X}$. Therefore, we obtain an upper bound on $\|\mathcal{A}(H)\|_2^2$ as follows:

$$\begin{aligned}\langle \mathcal{A}(H), \mathcal{A}(H)\rangle &= \langle H, \mathcal{A}^*(\mathcal{A}(H))\rangle \\ &\leq \|H\|_*\|\mathcal{A}^*(\mathcal{A}(H))\| \\ &\leq 2\mu\|H\|_*.\end{aligned} \tag{72}$$

Equation (72), the definition of ρ_{8r}, and $\tau(H) \leq 8r$ together yield

$$\begin{aligned}\rho_{8r}^2\|H\|_\mathrm{F}^2 &\leq \langle \mathcal{A}(H), \mathcal{A}(H)\rangle \\ &\leq 2\mu\|H\|_* \\ &\leq 2\mu\sqrt{8r}\|H\|_\mathrm{F}.\end{aligned} \tag{73}$$

We conclude that

$$\|H\|_\mathrm{F} \leq \frac{4\sqrt{2r}}{\rho_{8r}^2}\mu. \tag{74}$$

Now, we establish an upper bound on $\|\mathcal{A}(H)\|_2^2$ for the mLASSO (9) using a procedure similar to the one used for the mDS given previously. First note that

$$\begin{aligned}&\|\mathcal{A}^*(\mathcal{A}(H))\| \\ &\leq \|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(X))\| + \|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X}))\| \\ &\leq \|\mathcal{A}^*(\boldsymbol{w})\| + \|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X}))\| \\ &= \kappa\mu + \|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X}))\|.\end{aligned} \tag{75}$$

We follow the procedure in [5] [see also [18]] to estimate $\|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X}))\|$. Since $\hat{X}$ is the solution to (9), the optimality condition yields that

$$\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X})) \in \mu\partial\|\hat{X}\|_* \tag{76}$$

where $\partial\|\hat{X}\|_*$ is the family of subgradient of $\|\cdot\|_*$ evaluated at $\hat{X}$. According to [10], if the singular value decomposition of $\hat{X}$ is $U\Sigma V^T$, then we have

$$\partial\|\hat{X}\|_* = \{UV^T + W : \|W\| \leq 1 \quad U^T W = 0, WV = 0\}. \tag{77}$$

As a consequence, we obtain $\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X})) = \mu(UV^T + W)$ and

$$\begin{aligned}\|\mathcal{A}^*(\boldsymbol{y} - \mathcal{A}(\hat{X}))\| &\leq \|\mu(UV^T + W)\| \\ &= \mu.\end{aligned} \tag{78}$$

We used $\|UV^T + W\| = 1$ because

$$\begin{aligned}&\max_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \|(UV^T + W)\boldsymbol{x}\| \\ &= \max_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \|(UV^T + W)V\boldsymbol{y}\| \leq 1.\end{aligned} \tag{79}$$

Following the same lines in (72), we get

$$\|\mathcal{A}(H)\|_2^2 \leq (\kappa + 1)\mu\|H\|_*. \tag{80}$$

Then, (65), (75), and (78)

$$\begin{aligned}\rho^2_{\frac{8r}{(1-\kappa)^2}}\|H\|_{\mathrm{F}}^2 &\leq \|\mathcal{A}(H)\|_2^2 \\ &\leq (\kappa+1)\mu\frac{\sqrt{8r}}{1-\kappa}\|H\|_{\mathrm{F}}.\end{aligned} \tag{81}$$

As a consequence, the error bound (20) holds. ■

APPENDIX C
PROOF OF THEOREM 5

Proof:

1) Since in the optimization problem defining $f_\tau(\eta; Y)$, the objective function $\langle Y, Z\rangle$ is jointly continuous in τ, η, and Y, and the constraint correspondence

$$\begin{aligned}&C(\tau, \eta) : [0, \infty) \twoheadrightarrow \mathbb{R}^{n_1 \times n_2} \\ &\eta \mapsto \left\{Z : \|\mathcal{A}(Z)\|_2 \leq 1, \|Z\|_* \leq \sqrt{\tau}\eta\right\}\end{aligned} \tag{82}$$

is compact-valued and continuous (both upper and lower hemicontinuous), according to Berge's maximum theorem [34], the optimal value function $f_\tau(\eta; Y)$ is jointly continuous in τ, η, and Y. The continuity of $f_\tau(\eta)$ can be proved in a similar manner.

2) To show the strict increasing property, suppose $0 < \eta_1 < \eta_2$ and the dual variable λ_2^* achieves $f_\tau(\eta_2; Y)$ in (52) . Then, we have

$$\begin{aligned}f_\tau(\eta_1; Y) &\leq \sqrt{\tau}\eta_1\|Y - \mathcal{A}(\lambda_2^*)\| + \|\lambda_2^*\|_2 \\ &< \sqrt{\tau}\eta_2\|Y - \mathcal{A}(\lambda_2^*)\| + \|\lambda_2^*\|_2 \\ &= f_\tau(\eta_2; Y).\end{aligned} \tag{83}$$

The case for $\eta_1 = 0$ is proved by continuity, and the strict increasing of $f_\tau(\eta)$ follows immediately.

3) The concavity of $f_\tau(\eta; Y)$ follows from the dual representation (52) and the fact that $f_\tau(\eta; Y)$ is the minimization of a function of variables η and λ, and when λ, the variable to be minimized, is fixed, the function is linear in η.

4) Next, we show that when $\eta > 0$ is sufficiently small, $f_\tau(\eta; Y) \geq \sqrt{\tau}\eta$ if $Y = \boldsymbol{u}\boldsymbol{v}^T$ with $\|\boldsymbol{u}\|_2 = 1$ and $\|\boldsymbol{v}\|_2 = 1$. Taking $Z = \sqrt{\tau}\eta\boldsymbol{u}\boldsymbol{v}^T$, we have $\|Z\|_* = \sqrt{\tau}\eta$ and $\langle Y, Z\rangle = \sqrt{\tau}\eta > \eta$ (recall $\tau \in (1, \tau^*)$). In addition, when $0 < \eta \leq 1/(\sqrt{\tau}\|\mathcal{A}(\boldsymbol{u}\boldsymbol{v}^T)\|_2)$, we also have $\|\mathcal{A}(Z)\|_2 \leq 1$. Therefore, for sufficiently small η, we have $f_\tau(\eta; Y) \geq \sqrt{\tau}\eta > \eta$. Clearly, $f_\tau(\eta) = \max_Y f_\tau(\eta; Y) \geq \sqrt{\tau}\eta > \eta$ for such η.

Recall that

$$\tau^* = \min_Z \frac{\|Z\|_*^2}{\|Z\|_{\mathrm{F}}^2} \text{ subject to } \mathcal{A}(Z) = 0 \tag{84}$$

or equivalently $\frac{1}{\sqrt{\tau^*}}$ is

$$\max_{Z, Y:\|Y\|_{\mathrm{F}}=1} \{\langle Y, Z\rangle : \mathcal{A}(Z) = 0, \|Z\|_* \leq 1\}. \tag{85}$$

Since the dual program of

$$\max_Z \langle Y, Z\rangle \text{ s.t. } \mathcal{A}(Z) = 0, \|Z\|_* \leq 1 \tag{86}$$

is

$$\max_\lambda \|Y - \mathcal{A}^*(\lambda)\| \tag{87}$$

we have

$$\frac{1}{\sqrt{\tau^*}} = \max_{Y:\|Y\|_{\mathrm{F}}=1} \min_\lambda \|Y - \mathcal{A}^*(\lambda)\|. \tag{88}$$

Suppose λ_Y^* is the optimal solution for each $\min_\lambda \|Y - \mathcal{A}^*(\lambda)\|$. For each Y, we then have

$$\frac{1}{\sqrt{\tau^*}} \geq \|Y - \mathcal{A}^*(\lambda_Y^*)\| \tag{89}$$

which implies

$$\begin{aligned}f_\tau(\eta; Y) &= \min_\lambda \sqrt{\tau}\eta\|Y - \mathcal{A}^*(\lambda)\| + \|\lambda\|_2 \\ &\leq \sqrt{\tau}\eta\|Y - \mathcal{A}^*(\lambda_Y^*)\| + \|\lambda_Y^*\|_2 \\ &\leq \sqrt{\frac{\tau}{\tau^*}}\eta + \|\lambda_Y^*\|_2.\end{aligned} \tag{90}$$

As a consequence, we obtain

$$\begin{aligned}f_\tau(\eta) &= \max_Y f_\tau(\eta; Y) \\ &\leq \sqrt{\frac{\tau}{\tau_*}}\eta + \sup_Y \|\lambda_Y^*\|_2.\end{aligned} \tag{91}$$

Viewing $\mathcal{A}^* : (\mathbb{R}^m, \|\cdot\|_2) \to (\mathbb{R}^{n_1 \times n_2}, \|\cdot\|)$ as an operator, we obtain that $\theta_{\mathcal{A}} \stackrel{\text{def}}{=} \inf_{\lambda \neq 0} \|\mathcal{A}^*(\lambda)\|/\|\lambda\|_2 > 0$ when

$\{A^k\}_{k=1}^m$ are linearly independent, because $\mathcal{A}^*(\lambda) \neq 0$ for $\lambda \neq 0$. Triangle inequality and (89) imply

$$\begin{aligned}\frac{1}{\sqrt{\tau^*}} + 1 &\geq \frac{1}{\sqrt{\tau^*}} + \|Y\| \\ &\geq \|\mathcal{A}^*(\lambda_Y)\| \geq \theta_{\mathcal{A}}\|\lambda_Y\|_2. \end{aligned} \tag{92}$$

Therefore, the quantify $\sup \|\lambda_Y^*\|_2$ is finite. Pick $\rho \in (\sqrt{\tau/\tau_*}, 1)$. Then, we have the following when $\eta > \sup_Y \|\lambda_Y^*\|_2/(\rho - \sqrt{\tau/\tau_*})$:

$$\begin{aligned} f_\tau(\eta; Y) &\leq \rho\eta, \ \forall Y \in \mathbb{S}^{n_1 n_2 - 1} \text{ and} \\ f_\tau(\eta) &\leq \rho\eta. \end{aligned} \tag{93}$$

5) Properties 1) and 4) imply that $f_\tau(\eta; Y)$ has at least one positive fixed point for $Y = \boldsymbol{u}\boldsymbol{v}^T$. (Interestingly, 2) and 4) also imply the existence of a positive fixed point, see [35].) The positive fixed point for such $f_\tau(\eta; Y)$ is also unique. Suppose there are two fixed points $0 < \eta_1^* < \eta_2^*$. Pick η_0 small enough such that $f_\tau(\eta_0; Y) > \eta_0 > 0$ and $\eta_0 < \eta_1^*$. Then, $\eta_1^* = \lambda\eta_0 + (1-\lambda)\eta_2^*$ for some $\lambda \in (0,1)$, which implies that $f_\tau(\eta_1^*;Y) \geq \lambda f_\tau(\eta_0;Y) + (1-\lambda) f_\tau(\eta_2^*;Y) > \lambda\eta_0 + (1-\lambda)\eta_2^* = \eta_1^*$ due to the concavity, contradicting $\eta_1^* = f_\tau(\eta_1^*; Y)$.
The set of positive fixed point for $f_\tau(\eta)$, $\{\eta \in (0,\infty) : \eta = f_\tau(\eta) = \max_Y f_\tau(\eta; Y)\}$, is a subset of $\{\eta \in (0,\infty) : \eta = f_\tau(\eta; Y) \text{ for some } Y\}$, which is nonempty due to the existence of fixed points for $f_\tau(\eta; Y)$ with $Y = \boldsymbol{u}\boldsymbol{v}^T$. We argue that η^* is the supremum of

$$\{\eta \in (0,\infty) : \eta = f_\tau(\eta; Y) \text{ for some } Y\} \tag{94}$$

which is the unique positive fixed point for $f_\tau(\eta)$.
First of all, η^* must be finite as all the fixed points for $f_\tau(\eta; Y)$ are less than $\sup_Y \|\lambda_Y^*\|_2/(\rho - \sqrt{\tau/\tau_*})$ according to the proof of property 4). Second, η^* is a fixed point for some $f_\tau(\eta; Y^*)$, namely, the supremum is achievable and can be replaced by maximum. To see this, we construct $\{\eta_k\}_{k=1}^\infty$ converging to η^* using the definition of η^*, and corresponding $\{Y_k\}_{k=1}^\infty$ (i.e., η_k is a fixed point of $f_\tau(\eta; Y_k)$) converging to some Y^* using the compactness of $\mathbb{S}^{n_1 n_2 - 1}$. The joint continuity of $f_\tau(\eta; Y)$ in both η and Y implies

$$\begin{aligned}\eta^* &= \lim_{k\to\infty} \eta_k = \lim_{k\to\infty} f_\tau(\eta_k; Y_k) \\ &= f_\tau(\eta^*; Y^*). \end{aligned} \tag{95}$$

We proceed to show that η^* is a fixed point of $f_\tau(\eta)$. It suffices to show that $\max_Y f_\tau(\eta^*; Y) = f_\tau(\eta^*; Y^*)$. If this is not the case, there exists $Y_1 \neq Y^*$ such that $f_\tau(\eta^*; Y_1) > f_\tau(\eta^*; Y^*) = \eta^*$. The continuity of $f_\tau(\eta; Y_1)$ and the property 4) imply that there exists $\eta > \eta^*$ with $f_\tau(\eta; Y_1) = \eta$, contradicting the definition of η^*.

6) Next, we show $\eta^* = \gamma^* \stackrel{\text{def}}{=} 1/\rho_\tau(\mathcal{A})$ for any positive fixed point of $f_\tau(\eta)$, hence the uniqueness. We first prove $\gamma^* \geq \eta^*$ for any fixed point $\eta^* = f_\tau(\eta^*)$. Suppose Z^* achieves the optimization problem defining $f_\tau(\eta^*)$; then, we have

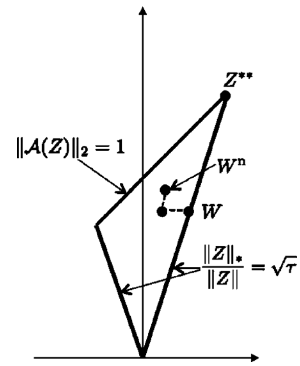

Fig. 1. Illustration of the proof for $\rho > 1$.

$$\eta^* = f_\tau(\eta^*) = \|Z^*\|_{\mathrm{F}}, \|\mathcal{A}(Z^*)\|_2 \leq 1 \tag{96}$$
$$\text{and } \|Z^*\|_* \leq \sqrt{\tau}\eta^*. \tag{97}$$

Since $\|Z^*\|_*/\|Z^*\|_{\mathrm{F}} \leq \sqrt{\tau}\eta^*/\eta^* \leq \sqrt{\tau}$, we have

$$\gamma^* \geq \frac{\|Z^*\|_{\mathrm{F}}}{\|\mathcal{A}(Z^*)\|_2} \geq \eta^*. \tag{98}$$

If $\eta^* < \gamma^*$, we define $\eta_0 = (\eta^* + \gamma^*)/2$ and

$$\begin{aligned} Z^{\mathrm{c}} &= \operatorname{argmax}_Z \frac{\sqrt{\tau}\|Z\|_{\mathrm{F}}}{\|Z\|_*} \\ &\quad \text{s.t. } \|\mathcal{A}(Z)\|_2 \leq 1, \|Z\|_{\mathrm{F}} \geq \eta_0 \end{aligned} \tag{99}$$
$$\rho = \frac{\sqrt{\tau}\|Z^{\mathrm{c}}\|_{\mathrm{F}}}{\|Z^{\mathrm{c}}\|_*}. \tag{100}$$

Suppose Z^{**} with $\|\mathcal{A}(Z^{**})\|_2 = 1$ achieves the optimum of the optimization (48) defining $\gamma^* = 1/\rho_\tau(\mathcal{A})$. Clearly, $\|Z^{**}\|_{\mathrm{F}} = \gamma^* > \eta_0$, which implies Z^{**} is a feasible point of the optimization problem (99) defining Z^{c} and ρ. As a consequence, we have

$$\rho \geq \frac{\sqrt{\tau}\|Z^{**}\|_{\mathrm{F}}}{\|Z^{**}\|_*} \geq 1. \tag{101}$$

Actually, we will show that $\rho > 1$. If $\|Z^{**}\|_* < \sqrt{\tau}\|Z^{**}\|_{\mathrm{F}}$, we are done. If not (i.e., $\|Z^{**}\|_* = \sqrt{\tau}\|Z^{**}\|_{\mathrm{F}}$), as illustrated in Fig. 1, we consider $W = \frac{\eta_0}{\gamma^*} Z^{**}$, which satisfies

$$\|\mathcal{A}(W)\|_2 = \frac{\eta_0}{\gamma^*} < 1 \tag{102}$$
$$\|W\|_{\mathrm{F}} = \eta_0, \text{ and} \tag{103}$$
$$\|W\|_* = \sqrt{\tau}\eta_0. \tag{104}$$

Suppose σ is the singular value vector of W. To get W^n as shown in Fig. 1, pick the smallest nonzero singular value, and scale it by a small positive constant κ less than 1. Because $\tau > 1$, σ has more than one nonzero component, elementary mathematics then show that this first scaling

will decrease the ratio $\|\sigma\|_1/\|\sigma\|_2$. We then scale the entire vector σ so that its ℓ_2 norm restores to its original value. This latter process of course does not change the ratio $\|\sigma\|_1/\|\sigma\|_2$.
If the scaling constant κ is close enough to 1, $\|\mathcal{A}(W^n)\|_2$ will remain less than 1 due to continuity. But the good news is that the ratio $\|\sigma\|_1/\|\sigma\|_2 = \|W^n\|_*/\|W^n\|_{\mathrm{F}}$ decreases, and hence $\rho \geq \frac{\sqrt{\tau}\|W^n\|_{\mathrm{F}}}{\|W^n\|_*}$ becomes greater than 1.
Now, we proceed to obtain a contradiction that $f_\tau(\eta^*) > \eta^*$. If $\|Z^{\mathrm{c}}\|_* \leq \sqrt{\tau}\cdot\eta^*$, then it is a feasible point of

$$\max_Z \|Z\|_{\mathrm{F}} \quad \text{s.t. } \|\mathcal{A}(Z)\|_2 \leq 1, \|Z\|_* \leq \sqrt{\tau}\cdot\eta^*. \tag{105}$$

As a consequence, $f_\tau(\eta^*) \geq \|Z^{\mathrm{c}}\|_{\mathrm{F}} \geq \eta_0 > \eta^*$, contradicting η^* is a fixed point and we are done. If this is not the case, i.e., $\|Z^{\mathrm{c}}\|_* > \sqrt{\tau}\cdot\eta^*$, we define a new point

$$Z^n = \tau Z^{\mathrm{c}} \tag{106}$$

with

$$\tau = \frac{\sqrt{\tau}\cdot\eta^*}{\|Z^{\mathrm{c}}\|_*} < 1. \tag{107}$$

Note that Z^n is a feasible point of the optimization problem defining $f_\tau(\eta^*)$ since

$$\|\mathcal{A}(Z^n)\|_2 = \tau\|\mathcal{A}(Z^{\mathrm{c}})\|_2 < 1 \tag{108}$$

$$\|Z^n\|_* = \tau\|Z^{\mathrm{c}}\|_* = \sqrt{\tau}\cdot\eta^*. \tag{109}$$

Furthermore, we have

$$\|Z^n\|_{\mathrm{F}} = \tau\|Z^{\mathrm{c}}\|_{\mathrm{F}} = \rho\eta^*. \tag{110}$$

As a consequence, we obtain a contradiction

$$f_\tau(\eta^*) \geq \rho\eta^* > \eta^*. \tag{111}$$

Therefore, for the fixed point η^*, we have $\eta^* = \gamma^* = 1/\rho_\tau(\mathcal{A})$.

7) This property simply follows from the continuity, the uniqueness, and property 4).
8) We use contradiction to show the existence of $\rho_1(\epsilon)$ in 8). In view of 4), we need only to show the existence of such a $\rho_1(\epsilon)$ that works for $\eta_L \leq \eta \leq (1-\epsilon)\eta^*$ where $\eta_L = \sup\{\eta : f_\tau(\xi) \geq \sqrt{\tau}\xi, \forall 0 < \xi \leq \eta\}$. Suppose otherwise, we then construct sequences $\{\eta^{(k)}\}_{k=1}^\infty \subset [\eta_L, (1-\epsilon)\eta^*]$ and $\{\rho_1^{(k)}\}_{k=1}^\infty \subset (1,\infty)$ with

$$\lim_{k\to\infty} \rho_1^{(k)} = 1$$
$$f_\tau(\eta^{(k)}) \leq \rho^{(k)}\eta^{(k)}. \tag{112}$$

Due to the compactness of $[\eta_L, (1-\epsilon)\eta^*]$, there must exist a subsequence $\{\eta^{(k_l)}\}_{l=1}^\infty$ of $\{\eta^{(k)}\}$ such that $\lim_{l\to\infty}\eta^{(k_l)} = \eta_{\lim}$ for some $\eta_{\lim} \in [\eta_L, (1-\epsilon)\eta^*]$. As a consequence of the continuity of $f_\tau(\eta)$, we have

$$f_\tau(\eta_{\lim}) = \lim_{l\to\infty} f_\tau(\eta^{(k_l)}) \leq \lim_{l\to\infty} \rho_1^{(k_l)}\eta^{(k_l)} = \eta_{\lim}. \tag{113}$$

Again, due to the continuity of $f_\tau(\eta)$ and the fact that $f_\tau(\eta) > \eta$ for $\eta < \eta_L$, there exists $\eta_c \in [\eta_L, \eta_{\lim}]$ such that

$$f_\tau(\eta_c) = \eta_c \tag{114}$$

contradicting the uniqueness of the fixed point for $f_\tau(\eta)$. The existence of $\rho_2(\epsilon)$ can be proved in a similar manner. ■

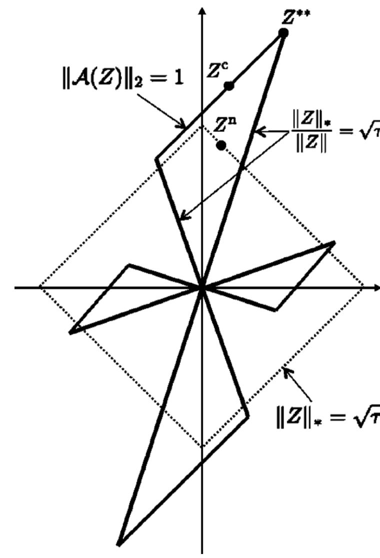

Fig. 2. Illustration of the proof for $f_\tau(\eta^*) \geq \rho\eta^*$.

APPENDIX D
PROOF OF PROPOSITION 2

In this appendix, we provide a proof of Proposition 2. For any r such that the set $V(r) = \left\{X \in \mathbb{R}^{n_1\times n_2} : \left\|\Sigma_1^{1/2}X\Sigma_2^{1/2}\right\|_{\mathrm{F}} = 1, \|X\|_* \leq r\right\}$ is nonempty, define a random variable

$$M(r, \mathcal{A}_{\Sigma_1,\Sigma_2}) := 1 - \inf_{X\in V(r)} \|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\|_2 = \sup_{X\in V(r)} \left\{1 - \|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\|_2\right\}. \tag{115}$$

We first show that

$$\mathbb{E}M(r, \mathcal{A}_{\Sigma_1,\Sigma_2}) \leq \frac{1}{4} + \frac{\sqrt{\mathrm{trace}(\Sigma_1)} + \sqrt{\mathrm{trace}(\Sigma_2)}}{\sqrt{m}}r. \tag{116}$$

To this end, we define two Gaussian processes indexed by $\boldsymbol{u} \in S^{m-1}, X \in V(r)$

$$Y_{\boldsymbol{u},X} = \langle \boldsymbol{u}, \mathcal{A}_{\Sigma_1,\Sigma_2}(X)\rangle = \sum_{k,i,j} u_k G_{ij}^k \left(\Sigma_1^{1/2} X \Sigma_2^{1/2}\right)_{ij} \quad (117)$$

$$Z_{\boldsymbol{u},X} = \langle \boldsymbol{u}, \boldsymbol{g}\rangle + \left\langle \Sigma_1^{1/2} X \Sigma_2^{1/2}, G\right\rangle \quad (118)$$

where S^{m-1} is the unit sphere in $\mathbb{R}^m$, and g_k, G_{ij}, G_{ij}^k are i.i.d. $\mathcal{N}(0, 1/m)$ random variables. Observe that

$$\sup_{X\in V(r)} \inf_{\boldsymbol{u}\in S^{m-1}} Y_{\boldsymbol{u},X} = -\inf_{X\in V(r)} \sup_{\boldsymbol{u}\in S^{m-1}} Y_{\boldsymbol{u},X} = -\inf_{X\in V(r)} \left\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\right\|_2. \quad (119)$$

We apply Gordon's comparison theorem [25, Ch. 3.1] which states that if

$$\mathbb{E}\left(Y_{\boldsymbol{u},X} - Y_{\boldsymbol{u}',X'}\right)^2 \le \mathbb{E}\left(Z_{\boldsymbol{u},X} - Z_{\boldsymbol{u}',X'}\right)^2 \quad (120)$$

for all $(\boldsymbol{u}, X), (\boldsymbol{u}', X')$ in $S^{m-1}\times V(r)$ and the inequality becomes equality when $X = X'$, then

$$\mathbb{E} \sup_{X\in V(r)} \inf_{\boldsymbol{u}\in S^{m-1}} Y_{\boldsymbol{u},X} \le \mathbb{E} \sup_{X\in V(r)} \inf_{\boldsymbol{u}\in S^{m-1}} Z_{\boldsymbol{u},X}. \quad (121)$$

Elementary calculations show that

$$\begin{aligned} &\mathbb{E}\left(Y_{\boldsymbol{u},X} - Y_{\boldsymbol{u}',X'}\right)^2 \\ &= \frac{1}{m}\sum_{k,i,j}\left(u_k\left(\Sigma_1^{1/2} X \Sigma_2^{1/2}\right)_{ij} - u_k'\left(\Sigma_1^{1/2} X' \Sigma_2^{1/2}\right)_{ij}\right)^2 \\ &\le \frac{1}{m}\sum_k\left(u_k - u_k'\right)^2 \\ &\quad + \frac{1}{m}\sum_{i,j}\left(\left(\Sigma_1^{1/2} X \Sigma_2^{1/2}\right)_{ij} - \left(\Sigma_1^{1/2} X' \Sigma_2^{1/2}\right)_{ij}\right)^2 \\ &= \mathbb{E}\left(Z_{\boldsymbol{u},X} - Z_{\boldsymbol{u}',X'}\right)^2 \end{aligned} \quad (122)$$

and

$$\mathbb{E}\left(Y_{\boldsymbol{u},X} - Y_{\boldsymbol{u}',X}\right)^2 = \mathbb{E}\left(Z_{\boldsymbol{u},X} - Z_{\boldsymbol{u}',X}\right)^2 = \frac{1}{m}\sum_k\left(u_k - u_k'\right)^2. \quad (123)$$

As a consequence, we upper bound the expectation of $M(r, \mathcal{A}_{\Sigma_1,\Sigma_2})$ as follows:

$$\begin{aligned} &\mathbb{E}M\left(r, \mathcal{A}_{\Sigma_1,\Sigma_2}\right) \\ &= 1 + \mathbb{E}\sup_{X\in V(r)} \inf_{\boldsymbol{u}\in S^{m-1}} Y_{\boldsymbol{u},X} \\ &\le 1 + \mathbb{E}\sup_{X\in V(r)} \inf_{\boldsymbol{u}\in S^{m-1}} Z_{\boldsymbol{u},X} \\ &\le 1 - \mathbb{E}\left\|\boldsymbol{g}\right\|_2 + \mathbb{E}\sup_{X\in V(r)}\left|\left\langle \Sigma_1^{1/2} X \Sigma_2^{1/2}, G\right\rangle\right|. \end{aligned} \quad (124)$$

For $\mathbb{E}\left\|\boldsymbol{g}\right\|_2$, we use a simple lower bound $\mathbb{E}\left\|\boldsymbol{g}\right\|_2 \ge \frac{3}{4}\sqrt{\frac{m}{m}} = \frac{3}{4}$ [27]. For the last term in (124), by the definition of $V(r)$, we have

$$\begin{aligned} &\mathbb{E}\sup_{X\in V(r)}\left|\left\langle \Sigma_1^{1/2} X \Sigma_2^{1/2}, G\right\rangle\right| \\ &= \mathbb{E}\sup_{X\in V(r)}\left|\left\langle X, \Sigma_1^{1/2} G \Sigma_2^{1/2}\right\rangle\right| \\ &\le \mathbb{E}\left\|X\right\|_* \left\|\Sigma_1^{1/2} G \Sigma_2^{1/2}\right\| \\ &\le r\mathbb{E}\left\|\Sigma_1^{1/2} G \Sigma_2^{1/2}\right\|. \end{aligned} \quad (125)$$

The problem then boils down to estimating $\mathbb{E}\left\|\Sigma_1^{1/2} G \Sigma_2^{1/2}\right\|$, which is achieved by applying Slepian's comparison theorem [25, Ch. 3.1] to the following two Gaussian processes indexed by $\boldsymbol{u} \in S^{n_1-1}, \boldsymbol{v} \in S^{n_2-1}$:

$$Y_{\boldsymbol{u},\boldsymbol{v}} = \boldsymbol{u}^T \Sigma_1^{1/2} G \Sigma_2^{1/2} \boldsymbol{v} \quad (126)$$

$$Z_{\boldsymbol{u},\boldsymbol{v}} = \boldsymbol{u}^T \Sigma_1^{1/2} \boldsymbol{g} + \boldsymbol{v}^T \Sigma_2^{1/2} \boldsymbol{h} \quad (127)$$

where $\boldsymbol{g} \sim \mathcal{N}\left(0, \frac{1}{m} I_{n_1}\right)$ and $\boldsymbol{h} \sim \mathcal{N}\left(0, \frac{1}{m} I_{n_2}\right)$. It is easy to verify that the variance condition for Slepian's comparison theorem holds

$$\mathbb{E}\left(Y_{\boldsymbol{u},\boldsymbol{v}} - Y_{\boldsymbol{u}',\boldsymbol{v}'}\right)^2 \le \mathbb{E}\left(Z_{\boldsymbol{u},\boldsymbol{v}} - Z_{\boldsymbol{u}',\boldsymbol{v}'}\right)^2. \quad (128)$$

Slepian's inequality $\mathbb{E}\sup_{\boldsymbol{u},\boldsymbol{v}} Y_{\boldsymbol{u},\boldsymbol{v}} \le \mathbb{E}\sup_{\boldsymbol{u},\boldsymbol{v}} Z_{\boldsymbol{u},\boldsymbol{v}}$ and Jensen's inequality then imply

$$\begin{aligned} &\mathbb{E}\left\|\Sigma_1^{1/2} G \Sigma_2^{1/2}\right\| \\ &\le \mathbb{E}\left\|\Sigma_1^{1/2}\boldsymbol{g}\right\|_2 + \mathbb{E}\left\|\Sigma_2^{1/2}\boldsymbol{h}\right\|_2 \\ &\le \sqrt{\mathbb{E}\left\|\Sigma_1^{1/2}\boldsymbol{g}\right\|_2^2} + \sqrt{\mathbb{E}\left\|\Sigma_2^{1/2}\boldsymbol{h}\right\|_2^2} \\ &= \frac{1}{\sqrt{m}}\left(\sqrt{\operatorname{trace}(\Sigma_1)} + \sqrt{\operatorname{trace}(\Sigma_2)}\right). \end{aligned} \quad (129)$$

Plugging back into (124) yields

$$\mathbb{E}M\left(r, \mathcal{A}_{\Sigma_1,\Sigma_2}\right) \le \frac{1}{4} + \frac{\sqrt{\operatorname{trace}(\Sigma_1)} + \sqrt{\operatorname{trace}(\Sigma_2)}}{\sqrt{m}} r. \quad (130)$$

We next obtain a high probability result using concentration of measure in Gauss space for Lipschitz functions. Denote $\boldsymbol{e}_k$ as the kth canonical basis vector. The following manipulation of

$$\begin{aligned} &h\left(G^1, \ldots, G^m\right) \\ &:= \sup_{X\in V(r)}\left(1 - \left\|\sum_k\left\langle G^k, \Sigma_1^{1/2} X \Sigma_2^{1/2}\right\rangle \boldsymbol{e}_k\right\|_2\right) \\ &= M\left(r, \mathcal{A}_{\Sigma_1,\Sigma_2}\right) \end{aligned} \quad (131)$$

as a function of $\left\{G^1, \ldots, G^m\right\}$

$$h\left(G^1, \ldots, G^m\right)$$

$$\begin{aligned}&\leq 1+\sup_{X\in V(r)}-\left\|\sum_k\left\langle G'^k,\Sigma_1^{1/2}X\Sigma_2^{1/2}\right\rangle \boldsymbol{e}_k\right\|_2\\&\quad+\sup_{X\in V(r)}\left\|\sum_k\left\langle G^k-G'^k,\Sigma_1^{1/2}X\Sigma_2^{1/2}\right\rangle \boldsymbol{e}_k\right\|_2\\&\leq h\left(G'^1,\ldots,G'^m\right)+\sqrt{\sum_k\|G^k-G'^k\|_{\mathrm{F}}^2}\end{aligned}\tag{132}$$

shows that the Lipschitz constant of $h\left(G^1,\ldots,G^m\right)$ is 1. Here, we have used the triangle inequality for the first inequality, Cauchy–Schwarz inequality and $\Sigma_1^{1/2}X\Sigma_2=1$ on $V(r)$ for the last inequality. By concentration of measure in Gauss space [36], we obtain

$$\begin{aligned}&\mathbb{P}\left\{\left|M\left(r,\mathcal{A}_{\Sigma_1,\Sigma_2}\right)-\mathbb{E}M\left(r,\mathcal{A}_{\Sigma_1,\Sigma_2}\right)\right|\geq t(r)/2\right\}\\&\leq 2\exp\left(-mt^2(r)/8\right)\end{aligned}\tag{133}$$

where $t(r):=\frac{1}{4}+\frac{\sqrt{\mathrm{trace}(\Sigma_1)}+\sqrt{\mathrm{trace}(\Sigma_2)}}{\sqrt{m}}r$, implying

$$\begin{aligned}&\mathbb{P}\left\{M\left(r,\mathcal{A}_{\Sigma_1,\Sigma_2}\right)\geq\frac{3t(r)}{2}\right\}\\&\leq 2\exp\left(-mt^2(r)/8\right).\end{aligned}\tag{134}$$

To complete the proof, we need the following lemma whose proof is based on a peeling argument.

Lemma 3 ([27, Lemma 3]): Consider a random objective function $f(\boldsymbol{v};\boldsymbol{u})$ with $\boldsymbol{u}$ the underlying random variable and $\boldsymbol{v}\in\mathbb{R}^p$ the variable to be optimized with. Suppose that $g:\mathbb{R}^p\to\mathbb{R}_+$ specifies an increasing constraint function, namely, $\{\boldsymbol{v}:g(\boldsymbol{v})\leq r\}\subseteq\{\boldsymbol{v}:g(\boldsymbol{v})\leq r'\}$ for $r\leq r'$, Ω is a nonempty constraint set, and

$$E=\{\boldsymbol{u}:\exists\boldsymbol{v}\in\Omega\text{ such that }f(\boldsymbol{v};\boldsymbol{u})\geq 2q(g(\boldsymbol{v}))\}$$

where $q:\mathbb{R}\to\mathbb{R}$ is nonnegative and strictly increasing with $q(r)\geq\mu$ for all $r\geq 0$. Further assume that there exists constant $d>0$ (which may depend on some other parameters) such that

$$\begin{aligned}&\mathbb{P}\left\{\boldsymbol{u}:\sup_{\boldsymbol{v}\in\Omega,g(\boldsymbol{v})\leq r}f(\boldsymbol{v};\boldsymbol{u})\geq q(r)\right\}\\&\leq 2\exp\left(-dq^2(r)\right).\end{aligned}\tag{135}$$

Then, we have

$$\mathbb{P}\{E\}\leq\frac{2\exp\left(-4d\mu^2\right)}{1-\exp\left(-4d\mu^2\right)}.\tag{136}$$

We apply Lemma 3 to the objective function $f\left(X;\mathcal{A}_{\Sigma_1,\Sigma_2}\right)=1-\left\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\right\|_2$, the constraint function $g(X)=\|X\|_*$, the constraint set $\Omega=\left\{X:\left\|\Sigma_1^{1/2}X\Sigma_2^{1/2}\right\|_{\mathrm{F}}=1\right\}$, $q(r)=\frac{3}{2}t(r)=\frac{3}{8}+\frac{3}{2}\frac{\sqrt{\mathrm{trace}(\Sigma_1)}+\sqrt{\mathrm{trace}(\Sigma_2)}}{\sqrt{m}}r$, and $\mu=3/8$. Since

$$\begin{aligned}&\mathbb{P}\left\{\sup_{g(X)\leq r,X\in\Omega}f\left(X;\mathcal{A}_{\Sigma_1,\Sigma_2}\right)\geq q(r)\right\}\\&\leq 2\exp\left(-cmq^2(r)\right)\end{aligned}\tag{137}$$

Lemma 3 implies that for properly chosen constant $c>0$, the event

$$\begin{aligned}E=\{&\exists X\in\Omega\text{ such that }1-\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\|_2\\&\geq\frac{3}{4}+3\frac{\sqrt{\mathrm{trace}(\Sigma_1)}+\sqrt{\mathrm{trace}(\Sigma_2)}}{\sqrt{m}}\|X\|_*\}\end{aligned}\tag{138}$$

has probability

$$\mathbb{P}\{E\}\leq 2\exp(-cm).\tag{139}$$

So, with probability at least $1-2\exp(-cm)$, we have

$$\begin{aligned}&\|\mathcal{A}_{\Sigma_1,\Sigma_2}(X)\|_2\\&\geq\frac{1}{4}\left\|\Sigma_1^{1/2}X\Sigma_2^{1/2}\right\|_{\mathrm{F}}\\&\quad-3\frac{\sqrt{\mathrm{trace}(\Sigma_1)}+\sqrt{\mathrm{trace}(\Sigma_2)}}{\sqrt{m}}\|X\|_*.\end{aligned}\tag{140}$$

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their helpful suggestions in improving the paper.

REFERENCES

[1] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[4] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2002.

[5] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.

[6] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[7] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, Dec. 2000.

[8] L. El Ghaoui and P. Gahinet, "Rank minimization under LMI constraints: A framework for output feedback problems," presented at the presented at the Eur. Control Conf., Groningen, The Netherlands, Jun. 1993.

[9] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Control Conf.*, 2001, vol. 6, pp. 4734–4739.

[10] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[11] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[12] D. Gross, Y. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.*, vol. 105, no. 15, pp. 150401–150404, Oct. 2010.

[13] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[14] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, Jun. 2011.

[15] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, pp. 215–245, 1995.

[16] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Boca Raton, FL: CRC, 1999.

[17] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, no. 2, pp. 321–353, 2011.

[18] P. Bickel, Y. Ritov, and A. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Statist.*, vol. 37, no. 4, pp. 1705–1732, 2009.

[19] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," *Ann. Statist.*, vol. 37, pp. 246–270, 2009.

[20] B. Recht, W. Xu, and B. Hassibi, "Null space conditions and thresholds for rank minimization," *Math. Program.*, vol. 127, no. 1, pp. 175–202, 2011.

[21] K. J. Böröczky and R. Schneider, "Mean width of circumscribed random polytopes," *Can. Math. Bull.*, vol. 53, pp. 614–628, 2010.

[22] D. Stroock, *Probability Theory: An Analytic View*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[23] K. Davidson and S. Szarek, "Local operator theory, random matrices and Banach spaces," in *Handbook of the Geometry of Banach Spaces*, W. B. Johnson and J. Lindenstrauss, Eds. New York: North Holland, 2001, vol. 1, pp. 317–366.

[24] R. Vershynin, *Lecture notes on non-asymptotic random matrix theory*, 2007.

[25] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*. New York: Springer-Verlag, 1991.

[26] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geom. Funct. Anal.*, vol. 14, no. 7, pp. 1248–1282, Nov. 2007.

[27] G. Raskutti, M. J. Wainwright, and B. Yu, "Restricted eigenvalue properties for correlated Gaussian designs," *J. Mach. Learn. Res.*, vol. 11, pp. 2241–2259, 2010.

[28] M. Rudelson and S. Zhou, "Reconstruction from anisotropic random measurements," Jun. 2011, arXiv:1106.1151.

[29] S. Negahban and M. J. Wainwright, "Estimation of (near) low-rank matrices with noise and high-dimensional scaling," Dec. 2009, arXiv:0912.5100v1.

[30] G. Tang and A. Nehorai, "Verifiable and computable performance analysis of sparsity recovery," Feb. 2011, arXiv:1102.4868v2.

[31] G. Tang and A. Nehorai, "Fixed point theory and semidefinite programming for computable performance analysis of block-sparsity recovery," Oct. 2011, arXiv:1110.1078v1.

[32] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589–592, 2008.

[33] G. Tang and A. Nehorai, "Performance analysis of sparse recovery based on constrained minimal singular values," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5734–5745, Dec. 2011.

[34] C. Berge, *Topological Spaces*, Reprinted ed. Mineola, NY: Dover, 1997, Paperback.

[35] A. Tarski, "A lattice-theoretical fix point theorem and its applications," *Pacific J. Math.*, vol. 5, pp. 285–309, 1955.

[36] P. Massart, *Concentration Inequalities and Model Selection*. New York: Springer, 2003.

Gongguo Tang (S'09–M'11) received the B.Sc. degree in mathematics from the Shandong University, China, in 2003, and the M.Sc. degree in systems science from Chinese Academy of Sciences, in 2006, and the Ph.D. degree in electrical and systems engineering from Washington University in St. Louis, St. Louis, MO, in 2011. He is currently a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Wisconsin-Madison. His research interests are in the area of signal processing, convex optimization, statistics, and their applications.

Arye Nehorai (S'80–M'83–SM'90–F'94) is the Eugene and Martha Lohman Professor and Chair of the Preston M. Green Department of Electrical and Systems Engineering (ESE) at Washington University in St. Louis (WUSTL). He serves as the Director of the Center for Sensor Signal and Information Processing at WUSTL. Earlier, he was a faculty member at Yale University and the University of Illinois at Chicago. He received the B.Sc. and M.Sc. degrees from the Technion, Israel and the Ph.D. from Stanford University, California. Dr. Nehorai served as Editor-in-Chief of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002. From 2003 to 2005 he was the Vice President of the IEEE Signal Processing Society (SPS), the Chair of the Publications Board, and a member of the Executive Committee of this Society. He was the founding editor of the special columns on Leadership Reflections in *IEEE Signal Processing Magazine* from 2003 to 2006. He has been a Fellow of the IEEE since 1994 and of the Royal Statistical Society since 1996.