Forty-Eighth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 1, 2010

The Stability of Low-Rank Matrix Reconstruction: a Constrained Singular Value Perspective

Gongguo Tang and Arye Nehorai

Department of Electrical and Systems Engineering
Washington University in St. Louis
St. Louis, MO 63130-1127
Email: nehorai@ese.wustl.edu

***Abstract*—The stability of low-rank matrix reconstruction 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 tight, and their derivations are less complex. Several random measurement ensembles are shown to have ℓ_*-CMSVs bounded away from zero with high probability, as long as the number of measurements is relatively large.**

***Index Terms*—ℓ_*-constrained minimal singular value, matrix Basis Pursuit, matrix Dantzig selector, matrix LASSO estimator, matrix 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 data set [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 to recover 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 $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ 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 addition, as in the case of sparse signal reconstruction [16], in certain applications we have the freedom to design the measurement operator $\mathcal{A}$ by selecting the best one from a collection of operators, which requires a precise quantification of the goodness of any given operator. All these considerations suggest that the stability measure should be *computable*, an aspect usually overlooked in literature.

This work is in parallel with our previous work on the stability study of sparse signal reconstruction. In [16], we demonstrated that the ℓ_1-constrained minimal singular value (ℓ_1-CMSV) of a measurement matrix quantifies the stability of sparse signal reconstruction. Several important random measurement ensembles are shown to have ℓ_1-CMSVs bounded away from zero with reasonable number of measurements. More importantly, we designed several algorithms to compute the ℓ_1-CMSV of any given measurement matrix.

In the current work, we define the ℓ_*-constrained minimal singular value (ℓ_*-CMSV) of a linear operator to measure the stability of low-rank matrix reconstruction. A large class of random linear operators are also shown to have ℓ_*-CMSVs bounded away from zero. The analysis for random linear operators acting on matrix space is more challenging. We need to employ advanced tools from geometrical functional analysis and empirical processes. The computational aspect of ℓ_*-CMSV is left to future work.

Several works in 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 *et.al.* 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]. Our bounds are more concise and are expected

This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, and ONR Grant N000140810849.

to be tighter. In both works [3] and [5], several important random measurement ensembles are shown to have the matrix restricted isometry constant (mRIC) bounded away from zero for reasonably large m. Our procedures for establishing the parallel results for ℓ_*-CMSV are significantly different from those in [3] and [5]. By analogy to ℓ_1-CMSV, we expect that the ℓ_*-CMSV is computationally more amenable than the mRIC [16].

The paper is organized as follows. Section II introduces notations and present the measurement model. Section III defines ℓ_*-CMSV and related concepts. The ℓ_*-CMSVs for subgaussian measurement operators are also analyzed in Section III. Section IV is devoted to deriving error bounds in term of ℓ_*-CMSVs for three convex relaxation algorithms. The paper is concluded in Section V.

II. Notations and the Measurement Model

A. Notations

We use ℓ_p^m to denote the space $\mathbb{R}^m$ equipped with the ℓ_p^m norm $\|\cdot\|_p$ defined as

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

and

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

for $\boldsymbol{x} = [x_1, \ldots, x_m]^T \in \mathbb{R}^m$. The notation $\|\boldsymbol{x}\|_0$ counts the number of nonzero elements of $\boldsymbol{x}$.

Suppose $X = [\boldsymbol{x}_1 \; \boldsymbol{x}_2 \; \ldots \; \boldsymbol{x}_{n_2}] \in \mathbb{R}^{n_1\times n_2}$. Define the Frobenius norm of X as $\|X\|_{\mathrm{F}} = \sqrt{\sum_{i,j}|X_{ij}|^2} = \sqrt{\sum_i \sigma_i^2(X)}$, the nuclear norm as $\|X\|_* = \sum_i \sigma_i(X)$, and the operator norm as $\|X\|_2 = \max\{\sigma_i(X)\}$, where $\sigma_i(X)$ is the ith singular value of X. The rank of X is denoted by $\mathrm{rank}(X) = \#\{i : \sigma_i(X) \neq 0\}$.

If we use $\boldsymbol{\sigma}(X) = [\sigma_1(X) \; \sigma_2(X) \; \ldots \; \sigma_{n_1}(X)]^T$ to represent the singular value vector, then clearly we have the following relations:

$$\begin{aligned} \|X\|_{\mathrm{F}} &= \|\boldsymbol{\sigma}(X)\|_2, \\ \|X\|_* &= \|\boldsymbol{\sigma}(X)\|_1, \\ \|X\|_2 &= \|\boldsymbol{\sigma}(X)\|_\infty, \\ \mathrm{rank}(X) &= \|\boldsymbol{\sigma}(X)\|_0. \end{aligned} \tag{4}$$

Note that we use $\|\cdot\|_2$ to represent both the matrix operator norm and the ℓ_2 norm of a vector. The exact meaning can always be inferred from the context. These singular value vector representations for the matrix norms immediately lead to

$$\begin{aligned} \|X\|_2 &\leq \|X\|_{\mathrm{F}} \leq \|X\|_* \\ &\leq \sqrt{\mathrm{rank}(X)}\|X\|_{\mathrm{F}} \\ &\leq \mathrm{rank}(X)\|X\|_2. \end{aligned}$$

The vectorization operator $\mathrm{vec}(X) = [\boldsymbol{x}_1^T \; \boldsymbol{x}_2^T \; \ldots \; \boldsymbol{x}_{n_2}^T]^T$ stacks the columns of X into a long column vector. The inner product of two matrices $X_1, X_2 \in \mathbb{R}^{n_1\times n_2}$ is defined as

$$\begin{aligned} \langle X_1, X_2\rangle &= \mathrm{trace}(X_1^T X_2) \\ &= \mathrm{vec}(X_1)^T \mathrm{vec}(X_2) \\ &= \sum_{i,j} (X_1)_{ij}(X_2)_{ij}. \end{aligned} \tag{5}$$

The following Cauchy-Schwarz type inequalities are due to the fact that the dual norm of the Frobenius norm is the Frobenius norm, and the dual norm of the operator norm is the nuclear norm [3]:

$$\begin{aligned} \langle X_1, X_2\rangle &\leq \|X_1\|_{\mathrm{F}}\|X_2\|_{\mathrm{F}}, \\ \langle X_1, X_2\rangle &\leq \|X_1\|_*\|X_2\|_2. \end{aligned}$$

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

$$\langle \mathcal{A}(X), \boldsymbol{z}\rangle = \langle X, \mathcal{A}^*(\boldsymbol{z})\rangle, \quad \forall X \in \mathbb{R}^{n_1\times n_2}, \boldsymbol{z} \in \mathbb{R}^m. \tag{6}$$

A linear operator $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \mapsto \mathbb{R}^m$ can be represented by m matrices $\mathscr{A} = \{A_1, A_2, \ldots, A_m\} \subset \mathbb{R}^{n_1\times n_2}$ as follows

$$\mathcal{A}(X) = \begin{bmatrix} \langle A_1, X\rangle \\ \langle A_2, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}. \tag{7}$$

We will interchangeably use $\mathcal{A}$ and $\mathscr{A}$ to represent the same linear operator. The adjoint operation for this representation is given by

$$\mathcal{A}^*(\boldsymbol{z}) = \sum_{k=1}^m \boldsymbol{z}_k A_k \in \mathbb{R}^{n_1\times n_2}. \tag{8}$$

B. The Measurement Model

Throughout the paper we assume $n_1 \leq n_2$ such that $n_1 = \min\{n_1, n_2\}$ and $n_2 = \max\{n_1, n_2\}$. Suppose we have a matrix $X \in \mathbb{R}^{n_1\times n_2}$ with $\mathrm{rank}(X) = r \ll n_1$. We observe $\boldsymbol{y} \in \mathbb{R}^m$ through the following linear model:

$$\boldsymbol{y} = \mathcal{A}(X) + \boldsymbol{w}, \tag{9}$$

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 the noise/disturbance vector, either deterministic or random. In the deterministic setting we assume boundedness: $\|\boldsymbol{w}\|_2 \leq \varepsilon$, while in the stochastic setting we assume Gaussianity: $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2\mathrm{I}_m)$. The model (9) is generally underdetermined with $m \ll n_1 n_2$, and the rank of X, r, is very small.

A fundamental problem pertaining to model (9) 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 $H \stackrel{\mathrm{def}}{=} \hat{X} - X$. 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 strength ε or σ^2.

III. ℓ_*-Constrained Singular Values

A. Matrix Restricted Isometry Constant

We first introduce the concept of mRIC used literature for low-rank matrix reconstruction analysis [3], [5]:

Definition 1 *For each integer $r \in \{1, \ldots, n_1\}$, the matrix restricted isometry constant (mRIC) δ_r of a linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ is defined as the smallest $\delta > 0$ such that*

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

holds for arbitrary non-zero 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. Hence, it is no surprise that the mRIC is involved in the stability of recovering X from $\mathcal{A}(X)$ corrupted by noise when X is of rank at most r.

The Rayleigh quotient $\frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}$ in the definition of the mRIC motivates us to define the rank constrained singular values, which are closely related to the mRIC.

Definition 2 *For any integer $1 \leq r \leq n_1$ and linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$, define the r-rank constrained minimal singular value $\nu_r^{\min}$ and $r-$rank constrained maximal singular value $\nu_r^{\max}$ of $\mathcal{A}$ via*

$$\nu_r^{\min}(\mathcal{A}) \stackrel{\text{def}}{=} \inf_{X \neq 0:\ \mathrm{rank}(X) \leq r} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}, \quad \textit{and} \tag{11}$$

$$\nu_r^{\max}(\mathcal{A}) \stackrel{\text{def}}{=} \sup_{X \neq 0:\ \mathrm{rank}(X) \leq r} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}, \tag{12}$$

respectively.

The mRIC δ_r for a linear operator $\mathcal{A}$ is related to the $r-$rank constrained minimal and maximal singular values by

$$\delta_r = \max\{|1 - (\nu_r^{\min})^2|, |(\nu_r^{\max})^2 - 1|\}. \tag{13}$$

Although the mRIC provides a measure quantifying the goodness of a linear operator, its computation poses great challenges. In the literature, the computation issue is circumvented by resorting to a random argument. We cite one general result below [5]:

- Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$ be a random linear operator satisfying the concentration inequality for any $X \in \mathbb{R}^{n_1 \times n_2}$ and $0 < \epsilon < 1$:

  $$\mathbb{P}\left(\left|\|\mathcal{A}(X)\|_2^2 - \|X\|_{\mathrm{F}}^2\right| \geq \epsilon \|X\|_{\mathrm{F}}^2\right) \leq Ce^{-mc_0(\epsilon)}. \tag{14}$$

  for fixed constant $C > 0$. Then, for any given $\delta \in (0, 1)$, there exist constants $c_1, c_2 > 0$ depending only on δ such that $\delta_r \leq \delta$, with probability not less than $1 - Ce^{-c_1 m}$, as long as

  $$m \geq c_2 n r. \tag{15}$$

B. ℓ_*-Constrained Singular Values for Random Operators

We define a quantity that continuously extends the concept of rank for a given matrix X. It is also an extension of the ℓ_1-sparsity level from vectors to matrices [16].

Definition 3 *The ℓ_*-rank of a non-zero matrix $X \in \mathbb{R}^{n_1 \times n_2}$ is defined as*

$$\tau(X) = \frac{\|X\|_*^2}{\|X\|_{\mathrm{F}}^2} = \frac{\|\boldsymbol{\sigma}(X)\|_1^2}{\|\boldsymbol{\sigma}(X)\|_2^2}. \tag{16}$$

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

$$\tau(X) \leq r, \tag{17}$$

and we have equality if and only if all non-zero singular values of X are equal. Therefore, the more non-zero 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 non-zero singular values but their magnitudes are spread in a wide range, then its ℓ_*-rank might be very small.

Definition 4 *For any $\tau \in [1, n_1]$ and any linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \mapsto \mathbb{R}^m$, define the ℓ_*-constrained minimal singular value (abbreviated as ℓ_*-CMSV) and the ℓ_*-constrained maximal singular value of $\mathcal{A}$ by*

$$\rho_\tau^{\min}(\mathcal{A}) \stackrel{\text{def}}{=} \inf_{X \neq 0,\ \tau(X) \leq \tau} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}, \quad \textit{and} \tag{18}$$

$$\rho_\tau^{\max}(\mathcal{A}) \stackrel{\text{def}}{=} \sup_{X \neq 0,\ \tau(X) \leq \tau} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_{\mathrm{F}}}, \tag{19}$$

respectively. Because we mainly use $\rho_\tau^{\min}$ in this paper, for notational simplicity, we sometimes use ρ_τ to denote $\rho_\tau^{\min}$ when it causes no confusion.

For an operator $\mathcal{A}$, a non-zero $\rho_\tau(\mathcal{A}) = \rho_\tau^{\min}(\mathcal{A})$ roughly means that $\mathcal{A}$ is invertible when restricted onto the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \tau(X) \leq \tau\}$, or equivalently, the intersection of the null space of $\mathcal{A}$ and $\{X \in \mathbb{R}^{n_1 \times n_2} : \tau(X) \leq \tau\}$ contains only the null vector of $\mathbb{R}^{n_1 \times n_2}$. The value of $\rho_\tau(\mathcal{A})$ measures the invertibility of $\mathcal{A}$ restricted onto $\{\tau(X) \leq \tau\}$. As we will see in Section IV-B, the error matrices for convex relaxation algorithms have small ℓ_*-ranks. Therefore, the error matrix is distinguishable from the zero matrix given the image of the error matrix under $\mathcal{A}$. Put it another way, given noise corrupted $\mathcal{A}(X)$, a signal matrix X is distinguishable from $X + H$, as long as the noise works in a way such that the error matrix H has a small ℓ_*-rank. This explains roughly why $\rho_\tau(\mathcal{A})$ determines the performance of convex relaxation algorithms.

Equation (17) implies that $\{X \neq 0 : \mathrm{rank}(X) \leq r\} \subseteq \{X \neq 0 : \tau(X) \leq r\}$. As a consequence, the rank constrained singular values satisfy the following inequality

$$\rho_r^{\min} \leq \nu_r^{\min} \leq \nu_r^{\max} \leq \rho_r^{\max}, \tag{20}$$

which combined with (13) yields the following relationship between mRIC and ℓ_*-constrained singular values:

$$\delta_r \leq \max\left\{|1-(\rho_r^{\min})^2|, |(\rho_r^{\max})^2 - 1|\right\}. \tag{21}$$

We begin to define a class of important random operator ensembles: the isotropic and subgaussian ensemble. For a scalar random variable x, the Orlicz ψ_2 norm is defined as

$$\|x\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E}\exp\left(\frac{|x|^2}{t^2}\right) \leq 2\right\}. \tag{22}$$

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

$$\mathbb{P}(|x| \geq t) \leq 2\exp(-ct^2/\|x\|_{\psi_2}). \tag{23}$$

The converse is also true, *i.e.,* if x has subgaussian tail $\exp(-t^2/K^2)$, then $\|x\|_{\psi_2} \leq cK$.

Definition 5 *A random vector $\boldsymbol{x} \in \mathbb{R}^n$ is called* isotropic and subgaussian *with constant L if $\mathbb{E}|\langle \boldsymbol{x}, \boldsymbol{u}\rangle|^2 = \|\boldsymbol{u}\|_2^2$ and $\|\langle \boldsymbol{x}, \boldsymbol{u}\rangle\|_{\psi_2} \leq L\|\boldsymbol{u}\|_2$ hold for any $\boldsymbol{u} \in \mathbb{R}^n$.*

A random vector $\boldsymbol{x}$ with independent subgaussian entries $x_1, \ldots, x_n$ is a subgaussian vector in the sense of Definition 5 because [17]

$$\begin{aligned}\|\langle \boldsymbol{x}, \boldsymbol{u}\rangle\|_{\psi_2} &\leq c\left(\sum_{i=1}^n u_i^2\|x_i\|_{\psi_2}^2\right)^{1/2} \\ &\leq c\|\boldsymbol{u}\|_2 \max_{1\leq i\leq n}\|x_i\|_{\psi_2}. \end{aligned}\tag{24}$$

Clearly, if in addition $\{x_i\}_{i\leq n}$ are centered and has unit variance, then $\boldsymbol{x}$ is also isotropic. In particular, the standard Gaussian vector on $\mathbb{R}^n$ and the sign vector with *i.i.d.* $1/2$ Bernoulli entries are isotropic and subgaussian. Isotropic and subgaussian random vectors also include the vectors with the normalized volume measure on various convex symmetric bodies , for example, the unit balls of ℓ_p^n for $2 \leq p \leq \infty$ [18].

Clearly, any linear operator $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^m$ can be represented by a collection of matrices $\mathscr{A} = \{A_1, \ldots, A_m\}$ (Refer to Section II-A for more details). This representation allows us to define isotropic and subgaussian operators:

Definition 6 *Suppose $\mathcal{A} : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^m$ is a linear operator with corresponding matrix representation $\mathscr{A}$. We say $\mathcal{A}$ is from the isotropic and subgaussian ensemble if for each $A_i \in \mathscr{A}$, $\mathrm{vec}(A_i)$ is an independent isotropic and subgaussian vector with constant L, and 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 $\ell_p^{n_1 n_2}$ for $2 \leq p \leq \infty$. For any isotropic and subgaussian operator $\mathcal{A}$, the typical values of $\rho_\tau^{\min}(\mathcal{A}/\sqrt{m})$ and $\rho_\tau^{\max}(\mathcal{A}/\sqrt{m})$ concentrate around 1 for relatively large m (but $\ll n_1 n_2$). More precisely, we have the following theorem:

Theorem 1 *Let $\mathcal{A}$ be an isotropic and subgaussian operator with some numerical constant L. Then there exists constants c_1, c_2, c_3 depending on L only such that for any $\epsilon > 0$ and $m \geq 1$ satisfying*

$$m \geq c_1\frac{\tau n_2}{\epsilon^2}, \tag{25}$$

we have

$$\begin{aligned}1 - c_2\epsilon \leq \mathbb{E}\rho_\tau^{\min}&\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \\ &\leq \mathbb{E}\rho_\tau^{\max}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq 1 + c_2\epsilon\end{aligned} \tag{26}$$

and

$$\begin{aligned}\mathbb{P}\left[1 - \epsilon \leq \rho_\tau^{\min}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq \rho_\tau^{\max}\left(\frac{\mathcal{A}}{\sqrt{m}}\right) \leq 1 + \epsilon\right] \\ \geq 1 - \exp(-c_3\epsilon^2 m).\end{aligned} \tag{27}$$

Using the relation between ℓ_*-constrained singular values and the mRIC, we have the following immediate corollary which was also established in [5] using different approaches :

Corollary 1 *Under the conditions of Theorem 1, there exists numerical constants c_1, c_2 such that for any $\epsilon > 0$ the mRIC constant $\delta_r(\mathcal{A}/\sqrt{m})$ satisfies*

$$\mathbb{P}\left[\delta_r(\mathcal{A}/\sqrt{m}) > \epsilon\right] \leq \exp(-c_1\epsilon^2 m), \tag{28}$$

as long as

$$m \geq c_2\frac{rn_2}{\epsilon^2}. \tag{29}$$

Corollary 1 was established in [5] using an $\epsilon-$net argument. The same procedure can not be generalized trivially to prove Theorem 1. The idea behind an $\epsilon-$net argument is that any point in the set under consideration can be approximated using a point in its $\epsilon-$cover with an error at most ϵ. One key ingredient in the proof given by [5] is that the approximation error matrix also has low rank. This is not satisfied by the error matrix in our case, because the difference between two matrices with small ℓ_*-ranks does not necessarily have a small ℓ_*-rank. This difficulty might be circumvented by resorting to the Dudley's inequality [19]. However, a good estimate of the covering number of the set $\{X \in \mathbb{R}^{n_1\times n_2} : \|X\|_{\mathrm{F}} = 1, \|X\|_*^2 \leq \tau\}$ that makes the Dudley's inequality tight enough is not readily available.

It is relatively easy to establish Theorem 1 for the Gaussian ensemble. The comparison theorems for Gaussian processes, the Gordon's inequality and the Slepian's inequality in particular, can be used to show that the expected values of ℓ_*-constrained singular values fall within the neighborhood of one. Then the concentration of measure phenomenon in Gauss space immediately yields the desired results of Theorem 1 for Gaussian ensemble as both $\rho_\tau^{\min}(\cdot)$ and $\rho_\tau^{\max}(\cdot)$ are Lipschitz functions. The Gordon's and Slepian's inequalities rely heavily on Gaussian processes' highly symmetric property, which is not satisfied by general subgaussian processes.

The corresponding results for subgaussian operators are more difficult to establish. In [20], the problem is formulated as one of empirical processes. We then employ a recent general result of empirical processes established in [21]. The challenges of deriving Theorem 1 using more direct approaches such as Dudley's inequality and/or the general generic chaining bound [22] is also discussed in [20].

One common ground of the proofs for the Gaussian case and the more general subgaussian case reveals the reason that $\mathcal{A}$ is invertible on $\{\tau(X) \leq \tau\}$ while it is far from invertible on $\mathbb{R}^{n_1 \times n_2}$. Both proofs rely on the fact that the canonical Gaussian process indexed by the intersection of $\{\tau(X) \leq \tau\}$ and the unit sphere of $(\mathbb{R}^{n_1 \times n_2}, \|\cdot\|_\mathrm{F})$ can not go too far from zero in its life. This essentially means that the set $\{\tau(X) \leq \tau\} \bigcap \{\|X\|_\mathrm{F} = 1\}$ with a small τ is significantly smaller than the unit sphere itself, on which the canonical Gaussian process would drift far away from zero. Refer to [20] for more precise meanings of these discussions.

IV. Stability of Convex Relaxation Algorithms

A. Reconstruction Algorithms

We briefly review three low-rank matrix recovery algorithms based on convex relaxation: the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator. A common theme of these algorithms is enforcing the low-rankness of solutions by penalizing large nuclear norms, or equivalently, the ℓ_1 norms of the singular value vectors. As a relaxation of the matrix rank, the nuclear norm remains a measure of low-rankness while being a convex function. In fact, the nuclear norm $\|\cdot\|_*$ is the convex envelop of $\mathrm{rank}(\cdot)$ on the set $\{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_2 \leq 1\}$ [3, Theorem 2.2]. Most computational advantages of the aforementioned three algorithms result from the convexity of the nuclear norm. As demonstrated in Section IV-B, the nuclear norm enforcement guarantees a small ℓ_*-rank of the error matrix.

The matrix Basis Pursuit algorithm [3], [5] tries to minimize the nuclear norm of solutions subject to the measurement constraint. It is applicable to both noiseless settings and bounded noise settings with a known noise bound ε. The matrix Basis Pursuit algorithm was originally developed for the noise-free case in [3], *i.e.*, $\varepsilon = 0$ in (30). In this paper, we refer to both cases as matrix Basis Pursuit. Mathematically, the matrix Basis Pursuit solves:

$$\mathrm{mBP}: \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \text{ subject to } \|\boldsymbol{y} - \mathcal{A}(Z)\|_2 \leq \epsilon. \quad (30)$$

The matrix Dantzig selector [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:

$$\mathrm{mDS}: \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \|Z\|_* \text{ subject to } \|\mathcal{A}^*(\boldsymbol{r})\|_2 \leq \lambda, \quad (31)$$

where $\boldsymbol{r} = \boldsymbol{y} - \mathcal{A}^*(\boldsymbol{z})$ is the residual vector, σ the noise standard deviation, and λ_n a control parameter.

The matrix LASSO estimator solves the following optimization problem [5], [23]:

$$\text{mLASSO:} \quad \min_{Z \in \mathbb{R}^{n_1 \times n_2}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(Z)\|_2^2 + \lambda_n \|Z\|_*. \quad (32)$$

We emphasize that all three optimization problems can be solved using convex programs.

Now we cite stability results on the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator, which are expressed in terms of the mRIC. The aim of stability analysis is to derive error bounds of the solutions of these algorithms. 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) matrix Basis Pursuit [5]: Suppose that $\delta_{4r} < \sqrt{2} - 1$ and $\|\boldsymbol{w}\|_2 \leq \varepsilon$. The solution to the matrix Basis Pursuit (30) satisfies
$$\|\hat{X} - X\|_\mathrm{F} \leq \frac{4\sqrt{1 + \delta_{4r}}}{1 - (1 + \sqrt{2})\delta_{4r}} \cdot \varepsilon. \quad (33)$$
2) matrix Dantzig selector [5]: If $\delta_{4r} < \sqrt{2} - 1$ and $\|\mathcal{A}^*(\boldsymbol{w})\|_2 \leq \lambda$, then
$$\|\hat{X} - X\|_\mathrm{F} \leq \frac{16}{1 - (\sqrt{2} + 1)\delta_{4r}} \cdot \sqrt{r} \cdot \lambda. \quad (34)$$
3) matrix LASSO estimator [5]: If $\delta_{4r} < (3\sqrt{2} - 1)/17$ and $\|\mathcal{A}^*(\boldsymbol{w})\| \leq \mu/2$, then the solution to the matrix LASSO (32) satisfies
$$\|\hat{X} - X\|_\mathrm{F} \leq C(\delta_{4r})\sqrt{r} \cdot \mu, \quad (35)$$
for some numerical constant C.

B. Stability of Convex Relaxation Algorithms using ℓ_*-CMSV

In this section, we present the stability results in terms of the ℓ_*-CMSV for three convex relaxation algorithms: the matrix Basis Pursuit, the matrix Dantzig Selector, and the matrix LASSO estimator. The procedure of establishing these theorems has two steps:

1) Show that the error matrix $H = \hat{X} - X$ has a small ℓ_*-rank: $\tau(H) \leq \tau$ for some suitably defined τ, which automatically leads to a lower bound $\|\mathcal{A}(H)\|_2 \geq \rho_\tau \|H\|_\mathrm{F}$. Here X is the true matrix and $\hat{X}$ is its estimate given by convex relaxation algorithms.
2) Obtain an upper bound on $\|\mathcal{A}(H)\|_2$.

As shown later, these are all relatively easy to show for the matrix Basis Pursuit algorithm. We have the following stability result:

Theorem 2 *If matrix X has rank r and the noise $\boldsymbol{w}$ is bounded; that is, $\|\boldsymbol{w}\|_2 \le \epsilon$, then the solution $\hat{X}$ to the matrix Basis Pursuit* (30) *obeys*

$$\|\hat{X} - X\|_{\mathrm{F}} \quad \le \quad \frac{2\epsilon}{\rho_{8r}}. \tag{36}$$

The corresponding bound (33) using mRIC is expressed as $\frac{4\sqrt{1+\delta_{4r}}}{1-(1+\sqrt{2})\delta_{4r}} \cdot \varepsilon$ under the condition $\delta_{4r} \le \sqrt{2} - 1$. Here δ_r is the mRIC defined in Definition 1. We note the ℓ_*-CMSV bound (36) is more concise and only requires $\rho_{8r} > 0$. Of course, we pay a price by replacing the subscript $4r$ with $8r$. A similar phenomena is also observed in the sparse signal reconstructions case [16]. As suggested by the numerical simulation in [16], by analogy we expect that it is easier to get $\rho_{8r} > 0$ than $\delta_{4r} \le \sqrt{2} - 1$. However, we did not run simulations in this paper because it is not clear how to compute δ_{4r} within reasonable time even for small scale problems.

In the following, we establish a proof of Theorem 2. For matrix Basis Pursuit (30), the second step is trivial as both X and $\hat{X}$ satisfy constraint $\|\boldsymbol{y} - \mathcal{A}(Z)\| \le \epsilon$ in (30). Therefore, the triangle inequality yields

$$\begin{aligned} \|\mathcal{A}(H)\|_2 &= \|\mathcal{A}(\hat{X} - X)\|_2 \\ &\le \|\mathcal{A}(\hat{X}) - \boldsymbol{y}\|_2 + \|\boldsymbol{y} - \mathcal{A}(X)\|_2 \\ &\le 2\epsilon. \end{aligned} \tag{37}$$

In order to establish that the error matrix has a small ℓ_*-rank in the first step, we present two lemmas on the properties of nuclear norms derived in [3]:

Lemma 1 [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] *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) $\langle B_1, B_2 \rangle = 0$.

Now we give a proof of Theorem 2:

Proof of Theorem 2: We decompose the error matrix $B = H$ according to Lemma 2 with $A = X$, more explicitly, we have:

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], the fact that $\|\hat{X}\|_* = \|X + H\|_1$ is the minimum among all Zs satisfying the constraint in (30) implies that $\|H_c\|_*$ cannot be very large. To see this, we observe that

$$\begin{aligned} \|X\|_* &\ge \|X + H\|_* \\ &= \|X + H_c + H_0\|_* \\ &\ge \|X + H_c\|_* - \|H_0\|_* \\ &= \|X\|_* + \|H_c\|_* - \|H_0\|_*. \end{aligned} \tag{38}$$

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

$$\|H_c\|_* \le \|H_0\|_*, \tag{39}$$

which leads to

$$\begin{aligned} \|H\|_* &\le \|H_0\|_* + \|H_c\|_* \\ &\le 2\|H_0\|_* \\ &\le 2\sqrt{\mathrm{rank}(H_0)}\|H_0\|_{\mathrm{F}} \\ &= 2\sqrt{2r}\|H\|_{\mathrm{F}}, \end{aligned} \tag{40}$$

where for the next to the last inequality we used the fact that $\|A\|_* \le \sqrt{\mathrm{rank}(A)}\|A\|_F$, and for the last inequality we used $\|H\|_{\mathrm{F}}^2 = \|H_0\|_{\mathrm{F}}^2 + \|H_c\|_{\mathrm{F}}^2 \ge \|H_0\|_{\mathrm{F}}^2$ because $\langle H_0, H_c \rangle = 0$. Inequality (40) is equivalent to

$$\tau(H) \le 8 \ \mathrm{rank}(X) = 8r. \tag{41}$$

It follows from (37) and Definition 4 that

$$\rho_{8r}\|H\|_{\mathrm{F}} \quad \le \quad \|\mathcal{A}(H)\|_2 \le 2\epsilon. \tag{42}$$

Hence, we get the conclusion of Theorem 2

$$\|\hat{X} - X\|_{\mathrm{F}} \quad \le \quad \frac{2\epsilon}{\rho_{8r}}. \tag{43}$$

■

Before stating the results for the matrix Dantzig Selector and the matrix Lasso estimator, we cite a lemma of [5]:

Lemma 3 [5] *Suppose $\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_m)$. If $C \ge 4\sqrt{(1 + \rho_1^{\max}(\mathcal{A}))\log 12}$, then there exists a numerical constant $c > 0$ such that with probability greater than $1 - 2\exp(-cn_2)$ that*

$$\|\mathcal{A}^*(\boldsymbol{w})\| \le C\sqrt{n_2}\sigma, \tag{44}$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$.

Lemma 3 allows to transform statements under the condition of $\|\mathcal{A}^*(\boldsymbol{w})\|_2 \le \lambda$, *e.g.* Theorem 3 and 4, into ones that hold with large probability. We now present the error bounds for the matrix Dantzig Selector and the matrix LASSO estimator, whose proofs can be found in [20].

Theorem 3 *Suppose the noise vector in model* (1) *satisfies $\|\mathcal{A}^*(\boldsymbol{w})\|_2 \le \lambda$, and suppose $X \in \mathbb{R}^{n_1 \times n_2}$ is of rank r. Then, the solution $\hat{X}$ to the matrix Dantzig Selector* (31) *satisfies*

$$\|\hat{X} - X\|_{\mathrm{F}} \le \frac{4\sqrt{2}}{\rho_{8r}^2} \cdot \sqrt{r} \cdot \lambda. \tag{45}$$

Theorem 4 *Suppose the noise vector in model* (1) *satisfies* $\|\mathcal{A}^*(\boldsymbol{w})\|_2 \leq \kappa\mu$ *for some* $\kappa \in (0,1)$*, and suppose* $X \in \mathbb{R}^{n_1 \times n_2}$ *is of rank* r*. Then, the solution* $\hat{X}$ *to* (31) *satisfies*

$$\|\hat{X} - X\|_{\mathrm{F}} \leq \frac{1+\kappa}{1-\kappa} \cdot \frac{2\sqrt{2}}{\rho^2_{\frac{8r}{(1-\kappa)^2}}} \cdot \sqrt{r} \cdot \mu. \tag{46}$$

For example, if we take $\kappa = 1 - 2\sqrt{2}/3$, then the bound becomes

$$\|\hat{X} - X\|_{\mathrm{F}} \leq 6(1 - \frac{\sqrt{2}}{3}) \cdot \frac{1}{\rho_{9r}} \cdot \sqrt{r} \cdot \mu. \tag{47}$$

The readers are encouraged to compare the statements of Theorem 3 and 4 with those using mRIC as cited in Section III-A (Equations (34), (35) and the conditions for them to be valid).

V. Conclusions

In this paper, the ℓ_*-constrained minimal singular value of a measurement operator, which measures the invertibility of the measurement operator restricted to matrices with small ℓ_*-ranks, is proposed to quantify the stability of low-rank matrix reconstruction. The reconstruction errors of the matrix Basis Pursuit, the matrix Dantzig selector, and the matrix LASSO estimator are concisely bounded using the ℓ_*-CMSV. Using a generic chaining bound for empirical processes, we demonstrate that the ℓ_*-CMSV is bounded away from zero with high probability for the subgaussian measurement ensembles, as long as the number of measurements is relatively large.

In the future work, we will study the feasibility of using the ℓ_*-CMSV to bound the reconstruction error of iterative algorithms. More importantly, we will also design algorithms to efficiently compute the ℓ_*-CMSV and use the ℓ_*-CMSV as a basis for designing optimal measurement operators.

References

[1] E. J. Candés 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. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization," Submitted to *SIAM Review*, 2007.

[4] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, 2002.

[5] E. J. Candés and Y. Plan, "Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements," Submitted for publication, 2009.

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

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

[8] L. El Ghaoui and P. Gahinet, "Rank minimization under LMI constraints: A framework for output feedback problems," in *Proceedings of the European Control Conference*. 1993.

[9] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference*. 2001.

[10] E. J. Candés and B. Recht, "Exact matrix completion via convex optimization," *Found. of Comput. Math.*, vol. 9, no. 7, pp. 717–772, 2008.

[11] E. J. Candès and Y. Plan, "Matrix completion with noise," Submitted to *Proceedings of the IEEE*.

[12] D. Gross, Y. K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," 2009.

[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. Cand'es, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," Technical report, Stanford University, 2009.

[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] G. Tang and A. Nehorai, "The ℓ_1-constrained minimal singular value: a computable quantification of the stability of sparse signal reconstruction," Submitted to *IEEE Trans. Inf. Theory*.

[17] S. Mendelson and N. Tomczak-Jaegermann, "A subgaussian embedding theorem," *Israel Journal of Mathematics*, pp. 349–364, Mar. 2008.

[18] F. Barthe, O. Guédon, S. Mendelson, and A. Naor, "A probabilistic approach to the geometry of the ℓ_p^n-ball," *Ann. Probab.*, pp. 480–513, 2005.

[19] M. Ledoux and M. Talagrand, *Probability in Banach spaces: Isoperimetry and processes*, vol. 23. of *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 1991.

[20] G. Tang and A. Nehorai, "The stability of low-rank matrix reconstruction: a constrained singular value perspective," in preparation.

[21] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geometric And Functional Analysis*, pp. 1248–1282, Nov. 2007.

[22] M. Talagrand, *The generic chaining: upper and lower bounds of stochastic processes*, Springer, 2005.

[23] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, 2009.