Lower Bounds on the Mean-Squared Error of Low-Rank Matrix Reconstruction

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

***Abstract*—We investigate the behavior of the mean-square error (MSE) of low-rank matrix reconstruction and its special case, matrix completion. We first derive the constrained Cramér–Rao bound (CRB) on the MSE matrix of any locally unbiased estimator, and then analyze the behavior of the constrained CRB when a subset of entries of the underlying matrix is randomly observed. We design an alternating minimization procedure to compute the maximum likelihood estimator (MLE) for the low-rank matrix, and demonstrate through numerical simulations that the performance of the MLE approaches the constrained CRB when the signal-to-noise ratio is high. Applying a Chapman–Robbins type Barankin bound allows us to derive lower bounds on the worst-case scalar MSE. We demonstrate that the worst-case scalar MSE is infinite even if the model is identifiable. However, the infinite scalar MSE is achieved only on a set of low-rank matrices with measure zero. We discuss the implications of these bounds and compare them with the empirical performance of the matrix LASSO estimator and the existing bounds in the literature.**

***Index Terms*—Barankin bound, Chapman–Robbins bound, constrained Cramér–Rao bound, low-rank matrix reconstruction, matrix completion, maximum likelihood estimator, mean-square error.**

I. Introduction

RECONSTRUCTION of a low-rank matrix from noisy linear measurements, especially from a subset of its entries corrupted by noise, appears in many signal processing branches, such as factor analysis, linear system realization [1], [2], matrix completion [3], [4], quantum state tomography [5], face recognition [6], [7], and Euclidean embedding [8], to name a few (see [9]–[11] for discussions and references therein). Suppose $X \in \mathbb{R}^{n\times p}$ is a low-rank matrix with rank $r \ll \min(n,p)$, then the goal of low-rank matrix reconstruction is to determine X from the linear measurements:

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

where $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement vector, $\mathcal{A} : \mathbb{R}^{n\times p} \to \mathbb{R}^m$ is the sensing operator, and $\boldsymbol{w} \in \mathbb{R}^m$ is the noise vector. In particular, when the operator $\mathcal{A}$ observes a subset of entries of the matrix X, the resulting problem is called matrix completion. Since the size of the measurement vector m is usually less than the size of the matrix np, the measurement model (1) is underdetermined.

In this paper, we investigate the behavior of the MSE in estimating X under the unbiasedness condition. For a fixed matrix X, we derive a constrained Cramér–Rao bound (CRB) on the MSE matrix that applies to any *locally* unbiased estimator. The bound depends on the sensing operator $\mathcal{A}$, the row and column spaces of the underlying matrix X, and the noise level of $\boldsymbol{w}$. We approximate the typical behavior of the constrained CRB using a concentration of measure argument. We design an alternating algorithm to compute the maximum likelihood estimator (MLE) of the low-rank matrix. Numerical simulations show that when the signal-to-noise ratio (SNR) is relatively high, the scalar MSE of the MLE, which is equal to the trace of the MSE matrix, approaches the trace of the constrained CRB. The constrained CRB is helpful for system design as it provides insight into which properties of the sensing operator $\mathcal{A}$ are important for low-rank matrix recovery. Under a *globally* unbiased condition, we show that the worst-case scalar MSE is infinite for any estimator. Actually this infinite MSE is achieved by any matrix that is not strictly of rank r.

We review approaches to solve X from its measurements. In a noiseless setting, solutions to the model (1) are not unique. A natural strategy to obtain the true X is to find the solution with lowest rank that is consistent with the measurement, i.e.,

$$\min \operatorname{rank}(X) \text{ s.t. } \boldsymbol{y} = \mathcal{A}(X), \tag{2}$$

where $\operatorname{rank}(\cdot)$ denotes the rank of a matrix. Unfortunately, the optimization problem (2) is NP-hard. A variety of computationally affordable methods, which work well for the noisy case, have been proposed to estimate X by exploiting its low-rankness. We are particular interested in the behavior of the matrix LASSO estimator, which solves the following regularized nuclear norm minimization problem:

$$\min_{X\in\mathbb{R}^{n\times p}} \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(X)\|_2^2 + \mu\|X\|_* \tag{3}$$

where $\|\cdot\|_*$ is the nuclear norm of the matrix. We use the fixed point continuation with approximate SVD (FPCA) algorithm [12] to efficiently solve (3). For the matrix completion problem, we also design an alternating minimization procedure to compute the MLE, assuming the knowledge of rank r for X. More specifically, we write $X = LR$ for $L \in \mathbb{R}^{n\times r}, R \in \mathbb{R}^{r\times p}$, and alternatingly minimize

$$\ell(L,R) = \|\boldsymbol{y} - \mathcal{A}(LR)\|_2^2 \tag{4}$$

with respect to L and R while fixing the other. We compare the performance of the FPCA, the MLE, and the derived constrained Cramér–Rao bound. Numerical simulations show that when the

Manuscript received October 19, 2010; revised March 28, 2011; accepted June 17, 2011. Date of publication July 12, 2011; date of current version September 14, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean Pierre Delmas. This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, ONR Grant N000140810849, and NSF Grants CCF-1014908 and CCF-0963742.

The authors are with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: gt2@ese.wustl.edu; nehorai@ese.wustl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2161471

SNR is high, the biased matrix LASSO estimator is suboptimal, and the constrained CRB is achieved by the MLE.

Universal lower bounds on the MSE matrix (or error covariance matrix) of any unbiased estimator, most notably the CRB, have long been used as a benchmark for system performance in the signal processing field [13]. However, the application of the CRB requires a regular parameter space (an open set in $\mathbb{R}^N$, for example) which is not satisfied by the low-rank matrix reconstruction problem. The parameter space $\mathcal{X}_r$ for low-rank matrices can not even be represented by $f(\mathbf{x}) = 0, g(\mathbf{x}) \leq 0$ for continually differentiable f and g, a common form of constraints in the theory of constrained CRB [14]–[16]. In [17]–[19], the constrained CRB is applied to study unbiased estimators for sparse vectors. In this paper, we analyze the low-rank matrix reconstruction problem by employing the Chapman–Robbins form of the Barankin bound [20]–[22], as well as a multiparameter Cramér–Rao type lower bound with parameter space constrained to nonopen subset of $\mathbb{R}^N$ [14]–[16]. The most significant challenge presented by applying the Chapman–Robbins type Barankin bound is to optimize the lower bound over all possible test points. We address the challenge by establishing a technical lemma on the behavior of a matrix function. To address the issue of representing the constraints in applying the constrained CRB, we directly derive the constrained Cramér–Rao lower bound from the Chapman–Robbins bound with additional regularity conditions [14].

The paper is organized as follows. In Section II, we introduce model assumptions for (1), the Chapman–Robbins type Barankin bound, and the constrained CRB. In Section III, we derive the constrained CRB on the MSE matrix for any locally unbiased estimator. Section IV shows that, by applying the Barankin bound, the worst-case scalar MSE for a globally unbiased low-rank matrix estimator is infinite. In Section V, the constrained CRB is compared with the empirical performance of the matrix LASSO estimator and the MLE. Section VI is a concluding summary.

II. Model Assumptions, the Chapman–Robbins Bound, and the Constrained Cramér–Rao Bound

In this section, we introduce model assumptions, and review the Chapman–Robbins type Barankin bound and constrained CRB. Suppose we have a low-rank matrix $X \in \mathcal{X}_r$, where the parameter space

$$\mathcal{X}_r = \{X \in \mathbb{R}^{n\times p} : \operatorname{rank}(X) \leq r\}. \tag{5}$$

For any matrix $X \in \mathbb{R}^{n\times p}$, we use $\mathbf{x} = \operatorname{vec}(X)$ to denote the vector obtained by stacking the columns of X into a single column vector. Similarly, for a vector $\mathbf{x} \in \mathbb{R}^{np}$, we use $\operatorname{mat}_{n,p}(\cdot)$ to denote the operation of reshaping $\mathbf{x}$ into an $n \times p$ matrix such that $X = \operatorname{mat}_{n,p}(\operatorname{vec}(X))$. Without introducing any ambiguity, we identify $\mathcal{X}_r$ with $\{\mathbf{x} \in \mathbb{R}^{np} : \operatorname{rank}(\operatorname{mat}_{n,p}(\mathbf{x})) \leq r\}$.

We observe X through the linear measurement mechanism

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

where the noise vector $\boldsymbol{w} \sim \mathcal{N}(0, \Sigma)$. It is convenient to rewrite (6) in the following matrix-vector form:

$$\boldsymbol{y} = \boldsymbol{A}\mathbf{x} + \boldsymbol{w} \tag{7}$$

where $\mathbf{x} = \operatorname{vec}(X)$, and $\boldsymbol{A} \in \mathbb{R}^{m\times np}$ is the matrix corresponding to the operator $\mathcal{A}$, namely, $\mathcal{A}(X) = \boldsymbol{A}(\operatorname{vec}(X))$. Therefore, the measurement vector $\boldsymbol{y}$ follows $\mathcal{N}(\boldsymbol{A}\mathbf{x}, \Sigma)$ with a probability density function (pdf)

$$f_{\mathbf{x}}(\boldsymbol{y}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{(\boldsymbol{y}-\boldsymbol{A}\mathbf{x})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{A}\mathbf{x})}{2}\right]. \tag{8}$$

Our goal is to derive lower bounds on the MSE matrix for any unbiased estimator $\hat{\mathbf{x}}(\boldsymbol{y})$ that infers the deterministic parameter $\mathbf{x} \in \mathcal{X}_r$ from $\boldsymbol{y}$.

We consider two types of unbiasedness requirements: global unbiasedness and local unbiasedness. Global unbiasedness requires that an estimator $\hat{\mathbf{x}}(\boldsymbol{y})$ is unbiased at any parameter point $\mathbf{x} \in \mathcal{X}_r$; that is,

$$\mathbb{E}\{\hat{\mathbf{x}}(\boldsymbol{y})\} = \mathbf{x}, \quad \forall \mathbf{x} \in \mathcal{X}_r. \tag{9}$$

Here, the expectation $\mathbb{E}\{\cdot\}$ is taken with respect to the noise. The local unbiased condition only imposes the unbiased constraint on parameters in the neighborhood of a single point. More precisely, we require

$$\mathbb{E}\{\hat{\mathbf{x}}(\boldsymbol{y})\} = \mathbf{z}, \quad \forall \mathbf{z} \in B_\varepsilon(\mathbf{x}) \cap \mathcal{X}_r \tag{10}$$

where $B_\varepsilon(\mathbf{x}) = \{\mathbf{z} \in \mathbb{R}^{np} : \|\mathbf{z}-\mathbf{x}\|_2 \leq \varepsilon\}$. Refer to [17] for more discussion on implications of unbiased conditions in the similar sparse estimator scenario.

It is good to distinguish two kinds of stability results for low-rank matrix reconstruction. For the first kind, we seek a condition on the operator $\mathcal{A}$ (e.g., the matrix restricted isometry property) such that we could reconstruct *all* matrices $X \in \mathcal{X}_r$ stably from the measurement $\boldsymbol{y}$. For the second kind, we study the stability of reconstructing a *specific* $X \in \mathcal{X}_r$ and find the set of sensing operators $\mathcal{A}$ that work well for this *particular* X. We usually need to identify "good" low-rank matrices X that can be reconstructed stably by all or most of the sensing operators. We see that, for the first kind of problem, it is suitable to consider the worst MSE among all matrices X with rank less than a specified value and global unbiasedness, while for the second kind it is more appropriate to focus on locally unbiased estimators and a fixed low-rank matrix X. Therefore, for the first kind of problems, we will apply the Chapman–Robbins type Barankin bound in a worst-case framework, and for the second kind we will apply the constrained CRB.

We now present the Chapman–Robbins version of the Barankin bound [14], [20], [21] on the MSE matrix (or error covariance matrix), defined as follows:

$$\Sigma_{\mathbf{x}} \overset{\text{def}}{=} \mathbb{E}\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^T\} \tag{11}$$

for any unbiased estimator $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\boldsymbol{y})$. For any integer $N \geq 1$, and arbitrary vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{X}_r$ that are not equal to $\mathbf{x}$, we define the finite differences, $\delta_i\mathbf{x}$ and $\delta_i f_{\mathbf{x}}$, as

$$\delta_i\mathbf{x} \overset{\text{def}}{=} \mathbf{x}_i - \mathbf{x} \text{ and } \delta_i f_{\mathbf{x}} \overset{\text{def}}{=} f_{\mathbf{x}_i} - f_{\mathbf{x}}. \tag{12}$$

If $\hat{\mathbf{x}}$ is unbiased at $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{X}_r$, the Chapman–Robbins bound states that the MSE matrix $\Sigma_{\mathbf{x}}$ satisfies the matrix inequality

$$\Sigma_{\mathbf{x}} \geq B_{\mathbf{x}} \stackrel{\text{def}}{=} [\delta \mathbf{x}] \left(\mathbb{E} \left[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}} \right]^T \left[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}} \right] \right)^{\dagger} [\delta \mathbf{x}]^T \tag{13}$$

where $\delta f_{\mathbf{x}} \stackrel{\text{def}}{=} [\delta_1 f_{\mathbf{x}}, \ldots, \delta_N f_{\mathbf{x}}]$, $\delta \mathbf{x} \stackrel{\text{def}}{=} [\delta_1 \mathbf{x}, \ldots, \delta_N \mathbf{x}]$ and the $\dagger$ denotes the pseudo-inverse. We use $A \geq B$ in the sense that $A - B$ is positive semidefinite. Taking the trace of both sides of (13) yields a bound on the scalar MSE $\mathrm{MSE}_{\mathbf{x}}$:

$$\mathrm{MSE}_{\mathbf{x}} \stackrel{\text{def}}{=} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 = \mathrm{tr}(\Sigma_{\mathbf{x}}) \geq \mathrm{tr}(B_{\mathbf{x}}). \tag{14}$$

While the MSE matrix $\Sigma_{\mathbf{x}}$ is a more accurate and complete measure of system performance, the scalar MSE $\mathrm{MSE}_{\mathbf{x}}$ is sometimes more amenable to analysis. We will use the Chapman–Robbins bound to derive a lower bound on the worst-case scalar MSE as follows:

$$\mathrm{MSE}_{\mathrm{w}} \stackrel{\text{def}}{=} \sup_{\mathbf{x}} \mathrm{MSE}_{\mathbf{x}} \geq B_{\mathrm{w}} \stackrel{\text{def}}{=} \sup_{N, \mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}}). \tag{15}$$

We include a proof for (15) in Appendix A.

The worst-case scalar MSE has been used by many researchers as an estimation performance criterion [23]–[26]. In many cases, it is desirable to minimize the scalar MSE to obtain a good estimator. However, the scalar MSE depends explicitly on the unknown parameters when the parameters are deterministic, and hence can not be optimized directly [26]. To circumvent this difficulty, the authors of [23]–[26] resort to a minimax framework to find estimators that minimize the worst-case scalar MSE. When an analytical expression of the scalar MSE is not available, it makes sense for system design to identify a lower bound on the worst-case MSE [27] and minimize it [28].

A constrained Cramér–Rao bound for locally convex parameter space under local unbiasedness condition is obtained by taking $\mathbf{x}_1, \ldots, \mathbf{x}_N$ arbitrarily close to $\mathbf{x}$. Suppose that $\mathbf{x}$ and N test points $\{\mathbf{x}_i = \mathbf{x} + \Delta_i \boldsymbol{\nu}_i, i = 1, \ldots, N\}$ are contained in $\mathcal{X}_r$ for sufficiently small $\Delta_i > 0, i = 1, \ldots, N$. Then under certain regularity conditions, the MSE matrix for any locally unbiased estimator $\hat{\mathbf{x}}$ satisfies [14, Lemma 2]

$$\Sigma_{\mathbf{x}} \geq C_{\mathbf{x}} \stackrel{\text{def}}{=} P[P^T J_{\mathbf{x}} P]^{\dagger} P^T \tag{16}$$

where P is any $np \times N$ matrix whose column space equals $\mathrm{span}\{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N\}$, and where the $np \times np$ Fisher information matrix

$$\begin{aligned} J_{\mathbf{x}} &= \mathbb{E} \left\{ \left[\frac{\nabla f_{\mathbf{x}}}{f_{\mathbf{x}}} \right] \left[\frac{\nabla f_{\mathbf{x}}}{f_{\mathbf{x}}} \right]^T \right\} \\ &= \mathbb{E} \{ [\nabla \ln f_{\mathbf{x}}] [\nabla \ln f_{\mathbf{x}}]^T \} = -\mathbb{E} \{ \nabla^2 \ln f_{\mathbf{x}} \}. \end{aligned} \tag{17}$$

Note that the positive definiteness of $J_{\mathbf{x}}$ in Lemma 2 of [14] can be relaxed to $\mathrm{rank}(P^T J_{\mathbf{x}} P) = \mathrm{rank}(P)$.

III. The Constrained Cramér–Rao Bound for Any Locally Unbiased Estimator

In this section, we apply (16) to derive the constrained CRB on the MSE matrix for any locally unbiased estimator. We are particularly interested in the matrix completion problem, for which we study the typical behavior of the derived constrained CRB in a probabilistic framework. We also propose an alternating minimization algorithm which computes the MLE for matrix completion.

A. Constrained Cramér–Rao Lower Bound

For any $\mathbf{x} \in \mathcal{X}_r$, in order to employ (16) and let N test points $\{\mathbf{x}_i = \mathbf{x} + \Delta_i \boldsymbol{\nu}_i\}_{i=1}^N$ lie in $\mathcal{X}_r$, we need to carefully select the N direction vectors $\{\boldsymbol{\nu}_i\}_{i=1}^N$. Denote $X = \mathrm{mat}_{n,p}(\mathbf{x})$ and $X_i = \mathrm{mat}_{n,p}(\mathbf{x}_i)$ for $i = 1, \ldots, N$. Suppose that $X = U_0 \Lambda_0 V_0^T$ is the singular value decomposition of X with $U_0 = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r] \in \mathbb{R}^{n \times r}, \Lambda_0 = \mathrm{diag}([\lambda_1, \ldots, \lambda_r]) \in \mathbb{R}^{r \times r}$, and $V_0 = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r] \in \mathbb{R}^{p \times r}$. If we define

$$\boldsymbol{\nu}_i = \mathrm{vec}(U_0 R_i), \quad i = 1, \ldots, N \tag{18}$$

with $R_i \in \mathbb{R}^{r \times p}$, then $\mathbf{x}_i = \mathbf{x} + \Delta_i \boldsymbol{\nu}_i \in \mathcal{X}_r$. If additionally $\{R_i\}_{i=1}^N$ are linearly independent when viewed as vectors, then $\{U_0 R_i\}_{i=1}^n$ (hence $\{\mathrm{vec}(U_0 R_i)\}_{i=1}^N$) are also linearly independent. To see this, note that multiplying both sides of

$$\sum_{i=1}^N \alpha_i U_0 R_i = U_0 \left(\sum_{i=1}^N \alpha_i R_i \right) = 0 \tag{19}$$

with U_0^T yields $\sum_{i=1}^N \alpha_i R_i = 0$. Therefore, we can find at most $r \times p$ linearly independent directions in this manner. Similarly, if we take $\boldsymbol{\nu}_i = \mathrm{vec}(L_i V_0^T)$, we find another $n \times r$ linearly independent directions. However, the union of the two sets of directions $\{\mathrm{vec}(U_0 R_i)\}_{i=1}^{pr} \cup \{\mathrm{vec}(L_j V_0)\}_{j=1}^{nr}$ is linearly dependent. As a matter of fact, we have only $(n+p)r - r^2$ linearly independent directions, as explicitly constructed in the proof of Theorem 1.

We first need the following lemma, whose proof is given in Appendix B:

Lemma 1: Suppose $P_1 \in \mathbb{R}^{np \times N_1}$ and $P_2 \in \mathbb{R}^{np \times N_2}$ are two full rank matrices with $\mathrm{span}\{P_1\} \subseteq \mathrm{span}\{P_2\}$. If $P_2^T J_{\mathbf{x}} P_2$ is positive definite, then $P_1 [P_1^T J_{\mathbf{x}} P_1]^{\dagger} P_1^T \leq P_2 [P_2^T J_{\mathbf{x}} P_2]^{\dagger} P_2^T$.

We have the following theorem:

Theorem 1: Suppose $\mathbf{x} \in \mathcal{X}_r$ has the full-size singular value decomposition $X = \mathrm{mat}_{n,p}(\mathbf{x}) = U \Lambda V^T$ with $U = [U_0 \;\; U_1] = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \;\; \boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{n \times n}, \Lambda = \mathrm{diag}([\lambda_1, \ldots, \lambda_r]) \in \mathbb{R}^{n \times p}$, and $V = [V_0 \;\; V_1] = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \;\; \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_p] \in \mathbb{R}^{p \times p}$. The MSE matrix at $\mathbf{x}$ for any unbiased estimator $\hat{\mathbf{x}}$ satisfies

$$\Sigma_{\mathbf{x}} \geq P[P^T (\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A}) P]^{-1} P^T \tag{20}$$

with

$$P = [V_1 \otimes U_0 \; V_0 \otimes U_0 \; V_0 \otimes U_1] \in \mathbb{R}^{np \times (n+p)r - r^2} \tag{21}$$

as long as $\mathrm{rank}(\boldsymbol{A} P) = (n+p)r - r^2$.

Proof: It is easy to compute that the Fisher information matrix for (6) is $J_{\mathbf{x}} = \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A}$. Lemma 1 tells us that we should find linearly independent directions $\{\boldsymbol{\nu}_i\}_{i=1}^N$ spanning as large a subspace as possible. If two sets of directions span the same subspace, then they are equivalent in maximizing the

lower bound. Therefore, without loss of generality, we take the following directions:

$$\begin{aligned} V_1 \otimes U_0 &= \left[\mathrm{vec}\left(\boldsymbol{u}_i \boldsymbol{v}_j^T\right)\right] \\ &= [\boldsymbol{v}_j \otimes \boldsymbol{u}_i], \quad 1 \le i \le r, r+1 \le j \le p \\ V_0 \otimes U_0 &= \left[\mathrm{vec}\left(\boldsymbol{u}_i \boldsymbol{v}_j^T\right)\right] \\ &= [\boldsymbol{v}_j \otimes \boldsymbol{u}_i], \quad 1 \le i \le r, 1 \le j \le r \\ V_0 \otimes U_1 &= \left[\mathrm{vec}\left(\boldsymbol{u}_i \boldsymbol{v}_j^T\right)\right] \\ &= [\boldsymbol{v}_j \otimes \boldsymbol{u}_i], \quad r+1 \le i \le n, 1 \le j \le r. \end{aligned} \tag{22}$$

We denote the index set of valid (i,j)s in the above directions as $\mathcal{I}$. Defining $P = [V_1 \otimes U_0 \; V_0 \otimes U_0 \; V_0 \otimes U_1] \in \mathbb{R}^{np \times ((n+p)r - r^2)}$, we have

$$\begin{aligned} P^T P &= \begin{bmatrix} V_1^T \otimes U_0^T \\ V_0^T \otimes U_0^T \\ V_0^T \otimes U_1^T \end{bmatrix} [V_1 \otimes U_0 \quad V_0 \otimes U_0 \quad V_0 \otimes U_1] \\ &= \mathrm{I}_{(n+p)r - r^2} \end{aligned} \tag{23}$$

which implies that $\mathrm{rank}(P) = (n+p)r - r^2$. Therefore, we obtain

$$\begin{aligned} \mathrm{rank}(P^T J_{\mathbf{x}} P) &= \mathrm{rank}((\boldsymbol{A}P)^T \Sigma^{-1} (\boldsymbol{A}P)) \\ &= \mathrm{rank}(\boldsymbol{A}P) = (n+p)r - r^2. \end{aligned} \tag{24}$$

The constrained CRB (16) then implies that

$$\Sigma_{\mathbf{x}} \ge P[P^T(\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})P]^{-1} P^T, \tag{25}$$

with P defined in (21). ∎

An immediate corollary is the following lower bound on the scalar MSE.

Corollary 1: Under the notations and assumptions of Theorem 1, we have

$$\mathrm{MSE}_{\mathbf{x}} \ge \mathrm{tr}([P^T(\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})P]^{-1}). \tag{26}$$

Proof: Taking the trace of both sides of (20) yields

$$\begin{aligned} \mathrm{MSE}_{\mathbf{x}} &\ge \mathrm{tr}(P[P^T(\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})P]^{-1} P^T) \\ &= \mathrm{tr}([P^T(\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})P]^{-1} P^T P) \\ &= \mathrm{tr}([P^T(\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})P]^{-1}) \end{aligned} \tag{27}$$

which is (26). ∎

The condition $\mathrm{rank}(\boldsymbol{A}P) = (n+p)r - r^2$ is satisfied if $\mathcal{A}(X) \neq 0$ for any nonzero matrix with a rank of at most $2r$. To see this, first note that

$$\mathrm{rank}(\boldsymbol{A}P) = \mathrm{rank}\left([\,\mathcal{A}(\boldsymbol{u}_i \boldsymbol{v}_j)\,]_{(i,j) \in \mathcal{I}}\right) \tag{28}$$

due to the definition of $\boldsymbol{A}$ and P. Thus, we only need to show

$$\sum_{(i,j) \in \mathcal{I}} \alpha_{ij} \mathcal{A}\left(\boldsymbol{u}_i \boldsymbol{v}_j^T\right) = \mathcal{A}\left(\sum_{(i,j) \in \mathcal{I}} \alpha_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^T\right) \neq 0 \tag{29}$$

for any $\boldsymbol{\alpha} = [\alpha_{ij}]_{(i,j) \in \mathcal{I}} \neq 0$, which is actually a consequence of the fact that

$$\begin{aligned} &\sum_{(i,j) \in \mathcal{I}} \alpha_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^T \\ &= \sum_{\substack{1 \le i \le r \\ 1 \le j \le p}} \alpha_{ij} \left(U_0 \mathbf{e}_i^r\right) \boldsymbol{v}_j^T + \sum_{\substack{r+1 \le i \le n \\ 1 \le j \le r}} \alpha_{ij} \boldsymbol{u}_i \left(V_0 \mathbf{e}_j^r\right)^T \\ &= \underbrace{U_0 \left(\sum_{\substack{1 \le i \le r \\ 1 \le j \le p}} \alpha_{ij} \mathbf{e}_i^r \boldsymbol{v}_j^T\right)}_{\text{rank at most } r} + \underbrace{\left(\sum_{\substack{r+1 \le i \le n \\ 1 \le j \le r}} \alpha_{ij} \boldsymbol{u}_i \mathbf{e}_j^{rT}\right) V_0^T}_{\text{rank at most } r} \end{aligned} \tag{30}$$

is nonzero and has a rank of at most $2r$. Here, $\mathbf{e}_i^n \in \mathbb{R}^n$ denotes the ith canonical basis of $\mathbb{R}^n$, i.e., the vector with the ith component one and the rest of the components zeros.

The more restrictive condition $\mathcal{A}(X) \neq 0$ for any nonzero matrix with rank at most $2r$ is of a global nature and is always not satisfied by the matrix completion problem. To see this, suppose $\mathcal{A}$ selects entries in the index set Ω. Then for any $(i,j) \notin \Omega$, we have $\mathcal{A}(\mathbf{e}_i^n \mathbf{e}_j^{pT}) = 0$ with $\mathrm{rank}(\mathbf{e}_i^n \mathbf{e}_j^{pT}) = 1$. However, the condition $\mathrm{rank}(\boldsymbol{A}P) = (n+p)r - r^2$ is met for matrix completion if the operation $\boldsymbol{A}P$ selects $(n+p)r - r^2$ linearly independent rows of P. This will happen with high probability for random selection operators unless the singular vectors of X are very spiky.

We note that U_0 and V_0 in Theorem 1 are determined by the underlying matrix X, while the semi-orthogonal matrices U_1 and V_1 are arbitrary as long as they span the spaces orthogonal to the column spaces of U_0 and V_0, respectively. Suppose we have another choice of $\tilde{U}_1$ and $\tilde{V}_1$. Since U_1 (V_1 respectively) spans the same space as $\tilde{U}_1$ ($\tilde{V}_1$ respectively), we have

$$\tilde{U}_1 = U_1 F \text{ and } \tilde{V}_1 = V_1 G \tag{31}$$

for the invertible matrices $F \in \mathbb{R}^{(n-r) \times (n-r)}$ and $G \in \mathbb{R}^{(p-r) \times (p-r)}$. Then, it is easy to see that

$$\begin{aligned} \tilde{P} &= [\tilde{V}_1 \otimes U_0 \; V_0 \otimes U_0 \; V_0 \otimes \tilde{U}_1] \\ &= [V_1 G \otimes U_0 \; V_0 \otimes U_0 \; V_0 \otimes U_1 F] \\ &= [V_1 \otimes U_0 \; V_0 \otimes U_0 \; V_0 \otimes U_1] \begin{bmatrix} G \otimes \mathrm{I}_r & & \\ & \mathrm{I}_r \otimes \mathrm{I}_r & \\ & & \mathrm{I}_r \otimes F \end{bmatrix} \end{aligned} \tag{32}$$

which together with Lemma 1 implies that the choice of U_1 and V_1 does not affect the bound (20).

We now present a simplified bound:

Corollary 2: Under the conditions of Theorem 1, we have the following simplified but slightly looser bound:

$$\begin{aligned} \mathrm{MSE}_{\mathbf{x}} \ge \max\big\{ &\mathrm{tr}([(\mathrm{I}_p \otimes U)^T (\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})(\mathrm{I}_p \otimes U)]^{-1}), \\ &\mathrm{tr}([(V \otimes \mathrm{I}_n)^T (\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})(V \otimes \mathrm{I}_n)]^{-1}) \big\}. \end{aligned} \tag{33}$$

In particular, for the matrix completion problem with white noise $\Sigma = \sigma^2 \mathrm{I}_m$, the above bound further simplifies to

$$\mathrm{MSE}_{\mathbf{x}} \geq \sigma^2 \max\left\{ \sum_{j=1}^{p} \mathrm{tr}([U(\boldsymbol{s}_j,:)^T U(\boldsymbol{s}_j,:)]^{-1}), \sum_{i=1}^{n} \mathrm{tr}([V(\boldsymbol{t}_i,:)^T V(\boldsymbol{t}_i,:)]^{-1}) \right\}. \quad (34)$$

Here, $\{\boldsymbol{s}_j, 1 \leq j \leq p\}$ is the index set for observed entries in the jth column of matrix X; $\{\boldsymbol{t}_i, 1 \leq i \leq n\}$ is similarly the observation index set for the ith row of X; $U(\boldsymbol{s}_j,:)$ is the submatrix of U with rows indicated by $\boldsymbol{s}_j$; and $V(\boldsymbol{t}_i,:)$ is defined similarly.

Proof: Note that, in Theorem 1, we could also equivalently take P to be

$$P_1 = [\mathrm{I}_p \otimes U_0 \; V_0 \otimes U_1] = P \begin{bmatrix} V^T \otimes \mathrm{I}_r & 0 \\ 0 & \mathrm{I}_{r(n-r)} \end{bmatrix} \quad (35)$$

or

$$P_2 = [V_1 \otimes U_0 \; V_0 \otimes \mathrm{I}_n] = P \begin{bmatrix} \mathrm{I}_{r(p-r)} & 0 \\ 0 & \mathrm{I}_r \otimes U^T \end{bmatrix}. \quad (36)$$

If we take $\tilde{P}_1 = \mathrm{I}_p \otimes U_0$ and $\tilde{P}_2 = V_0 \otimes \mathrm{I}_n$, respectively, then according to Lemma 1 it is easy to get two lower bounds on $\mathrm{tr}([P^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})P]^{-1})$:

$$\mathrm{tr}([P^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})P]^{-1}) \geq \mathrm{tr}([(\mathrm{I}_p \otimes U_0)^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})(\mathrm{I}_p \otimes U_0)]^{-1}) \quad (37)$$

and

$$\mathrm{tr}([P^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})P]^{-1}) \geq \mathrm{tr}([(V_0 \otimes \mathrm{I}_n)^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})(V_0 \otimes \mathrm{I}_n)]^{-1}). \quad (38)$$

When $\Sigma = \sigma^2 \mathrm{I}_m$ and we observe a subset of entries of X, i.e., the matrix completion problem with white noise, each row of the observation matrix $\boldsymbol{A}$ has a single 1 and all other elements are zeros. If we exclude the possibility of repeatedly observing entries, then $\boldsymbol{A}^T\boldsymbol{A}$ is a diagonal matrix with m ones and $np-m$ zeros in the diagonal, where the ones correspond to the observed locations in $\mathbf{x}$. Then algebraic manipulations of (37) and (38) yield the desired (34). Thus, the conclusion of the corollary holds. ∎

The simplified bound given in (34) is not as tight as the one given in Theorem 1, as shown in Fig. 3. However, it is much easier to compute (34) when n and p are large.

B. Probability Analysis of the Constrained Cramér-Rao Bound

We analyze the behavior of the bound $\mathrm{tr}([P^T(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})P]^{-1})$ for the matrix completion problem when $\Sigma = \sigma^2 \mathrm{I}_m$. Suppose we randomly and uniformly observe m entries of matrix X, the corresponding index set of which is denoted by Ω. Based on this measurement model, we rewrite

$$P^T(\boldsymbol{A}^T\boldsymbol{A})P = \frac{1}{np}\sum_{i\in\Omega} \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T \quad (39)$$

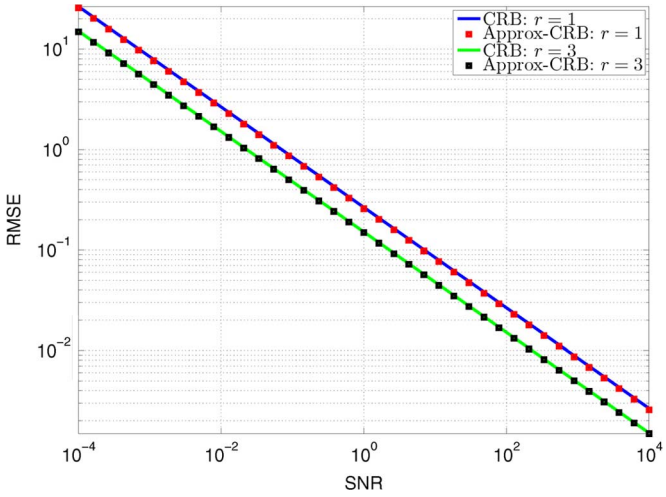


Fig. 1. Normalized constrained Cramér–Rao bound and its approximation with $n = p = 300, \rho = [m/np] = 0.3$.

where $\boldsymbol{\xi}_i^T/\sqrt{np}$ is the ith row of P. Note that

$$\begin{aligned} \frac{np}{m}\mathbb{E}\{P^T(\boldsymbol{A}^T\boldsymbol{A})P\} &= \frac{1}{m}\mathbb{E}\sum_{i\in\Omega}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T \\ &= \frac{1}{m}\frac{1}{\binom{np}{m}}\sum_{i=1}^{np}\binom{np-1}{m-1}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T, \\ &= \sum_{i=1}^{np}\frac{\boldsymbol{\xi}_i}{\sqrt{np}}\frac{\boldsymbol{\xi}_i^T}{\sqrt{np}} = \mathrm{I}_{(n+p)r-r^2}. \end{aligned} \quad (40)$$

Since $\mathrm{tr}(A^{-1})$ is a convex function for positive semidefinite matrices [29, p. 283, Proposition 8.5.15, xviii], Jensen's inequality implies that

$$\begin{aligned} &\mathbb{E}\{\mathrm{tr}([P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\}\sigma^2 \\ &\geq \mathrm{tr}([\mathbb{E}P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\sigma^2 \\ &= \mathrm{tr}\left(\left(\frac{m}{np}\mathrm{I}_{(n+p)r-r^2}\right)^{-1}\right)\sigma^2 \\ &= \frac{np[(n+p)r-r^2]}{m}\sigma^2. \end{aligned} \quad (41)$$

Our result of Theorem 3, specifically (50), strengthens the above result and says the bound $\mathrm{tr}([P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\sigma^2$ actually concentrates around $\frac{np[(n+p)r-r^2]}{m}\sigma^2$ with high probability. As a matter of fact, when n and p are relatively large, the bound is very close to $\frac{np[(n+p)r-r^2]}{m}\sigma^2$, as illustrated in Fig. 1.

We need to establish a result stronger than (40) that the eigenvalues of $\frac{np}{m}P^T(\boldsymbol{A}^T\boldsymbol{A})P$ concentrate around one with high probability, which implies that the lower bound $\mathrm{tr}([P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\sigma^2$ concentrates around $\frac{np[(n+p)r-r^2]}{m}\sigma^2$. For this purpose, we need to study the following quantity:

$$\left\| \mathrm{I}_{(n+p)r-r^2} - \frac{1}{m}\sum_{i\in\Omega}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T \right\|_2. \quad (42)$$

We first present a lemma about the behavior of a Rademacher average [30], whose proof is given in Appendix C.

Lemma 2: Let $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m \in \mathbb{R}^{np}$ have uniformly bounded entries, i.e., $\|\boldsymbol{\xi}_i\|_{\ell_\infty} \leq \mu_B$ for all i. Then

$$\mathbb{E}\left\|\sum_{i=1}^{m} \varepsilon_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T\right\|_2 \leq C\sqrt{(n+p)r - r^2} \times \log^{3/2}((n+p)r - r^2)\sqrt{\log m}\left\|\sum_{i=1}^{m} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T\right\|_2^{1/2} \quad (43)$$

where $\{\varepsilon_i\}$ are independent symmetric $\{-1, 1\}$ valued random variables (the Rademacher sequence), and C is a constant that depends on μ_B.

We establish the following theorem, whose proof is given in Appendix D.

Theorem 2: Suppose

$$\|\mathrm{vec}(U)\|_{\ell_\infty} \leq \sqrt{\mu_B/n}, \quad \|\mathrm{vec}(V)\|_{\ell_\infty} \leq \sqrt{\mu_B/p} \quad (44)$$

for some $\mu_B = O(1)$. Then if

$$m \geq C\frac{[(n+p)r - r^2]\log^3((n+p)r - r^2)}{\varepsilon^2} \times \log\left(\frac{(n+p)r - r^2}{\varepsilon^2}\right), \quad (45)$$

we have

$$\mathbb{E}\left\|\mathrm{I}_{(n+p)r-r^2} - \frac{1}{m}\sum_{i\in\Omega} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T\right\|_2 \leq \varepsilon. \quad (46)$$

The proof essentially follows [31]. Note that a standard symmetrization technique implies that the left-hand side of (46) is bounded by

$$2\mathbb{E}\left\|\frac{1}{m}\sum_{i\in\Omega} \varepsilon_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T\right\|_2 \quad (47)$$

which is a scaled version of the Rademacher average considered in Lemma 2.

The assumption (44) means that the singular vectors spread across all coordinates, i.e., they are not very spiky. Assumptions similar to this are used in many theoretical results for matrix completion [4], [32]. Recall that for $U = [U_0\ U_1]$ and $V = [V_0\ V_1]$, only U_0 and U_1 are determined by the underlying matrix X, while U_1 and V_1 are arbitrary aside from forming orthogonal unit bases for the spaces orthogonal to the column spaces of U_0 and V_0, respectively. Furthermore, the constrained CRB $\mathrm{tr}([P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\sigma^2$ does not depend on the choice of U_1 and V_1. Hence, it is not very natural to impose conditions such as (44) on U_1 and V_1. We conjecture that if the rank r is not extremely large, it is always possible to construct U_1 and V_1 such that the assumption (44) holds for U_1 and V_1 if it holds for U_0 and V_0. If this conjecture can be shown, our assumption is the same as the weak incoherence property imposed in [4] and [32]. However, currently, we are not able to prove this conjecture. Note that our Theorem 2 and Theorem 3 are based on the assumption (44) and not on the conjecture we raise here.

Because the elements of $P = [V_1 \otimes U_0\ V_0 \otimes U_0\ V_0 \otimes U_1]$ are of the form $U_{ij}V_{kl}$, we conclude from (44) that

$$\|\mathrm{vec}(P)\|_{\ell_\infty} \leq \frac{\mu_B}{\sqrt{np}}, \quad (48)$$

or equivalently

$$\|\boldsymbol{\xi}_i\|_{\ell_\infty} \leq \mu_B \quad \text{for } i = 1, \ldots, np. \quad (49)$$

Hence, the uniform boundedness condition in Lemma 2 is satisfied.

We proceed to use a concentration inequality to show the following high probability result.

Theorem 3: Under the assumption (44), we have

$$\frac{2}{3}\frac{np[(n+p)r - r^2]}{m}\sigma^2 \leq \mathrm{tr}([P^T(\boldsymbol{A}^T\boldsymbol{A})P]^{-1})\sigma^2 \leq 2\frac{np[(n+p)r - r^2]}{m}\sigma^2 \quad (50)$$

with probability greater than $1 - 5e^{-\frac{c}{\varepsilon^2}}$ for some constant $c > 0$, as long as m satisfies (45).

Theorem 3 follows from the following concentration of measure result:

$$\mathbb{P}\left\{\left\|\mathrm{I}_{(n+p)r-r^2} - \frac{1}{m}\sum_{i\in\Omega} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T\right\|_2 \geq \frac{1}{2}\right\} \leq 5e^{-\frac{c}{\varepsilon^2}} \quad (51)$$

whose proof follows [31] with minor modifications. Hence, we omit the proof in this paper. The implication of (51) is that the eigenvalues of $\frac{np}{m}P^T(\boldsymbol{A}^T\boldsymbol{A})P$ are between 1/2 and 3/2 with high probability.

We compare our approximated bound $\frac{np[(n+p)r-r^2]}{m}\sigma^2$ with existing results for matrix completion. Our result is an approximation of the universal lower-bound. In some sense, our result is more of a necessary condition. The results of [4], [32] are for particular algorithms, and they are sufficient conditions to guarantee the stability of these algorithms. Consider the following optimization problem:

$$\text{minimize} \quad \|X\|_* \quad (52)$$

$$\text{subject to} \quad \|\mathcal{A}(X) - \boldsymbol{y}\|_{\ell_2} \leq c\sqrt{m}\sigma. \quad (53)$$

For this problem, a typical result of ([4], Equation III.3) states that when $n = p$ the solution $\hat{X}$ obeys with high probability

$$\frac{\|\hat{X} - X\|_{\mathrm{F}}}{n} \leq C\sqrt{n}\sigma. \quad (54)$$

Our probability analysis says that any locally unbiased estimator approximately satisfies

$$\frac{\|\hat{X} - X\|_{\mathrm{F}}}{n} \geq C\sqrt{\frac{nr}{m}}\sigma. \quad (55)$$

Since m must be greater than $2nr$, the right-hand side of (55) is essentially σ, which means the per-element error is proportional to the noise level. However, the existing bound in (54) is proportional to $\sqrt{n}\sigma$. Considering that the MLE approaches the constrained CRB (see Fig. 3), we see that the bound (54) is not optimal.

TABLE I
MAXIMUM LIKELIHOOD ESTIMATOR ALGORITHM

$\texttt{MLE}(n, p, r, Y, \Omega)$
Input: Matrix sizes n and p, matrix rank r, observation matrix Y, index set Ω.
Output: A rank r approximation of the target matrix.

1. **initialize:** $L^{(0)} = U_0$, $R^{(0)} = \Lambda_0 V_0^T$ where $Y = U_0 \Lambda_0 V_0^T$ is the economical SVD of Y;
2. **while** not converged **do**
3. compute $L_k^{(t)} = \left(\sum_{l:(k,l)\in\Omega} R_l^{(t-1)} R_l^{(t-1)^T}\right)^{-1} \left(\sum_{l:(k,l)\in\Omega} Y_{kl} R_l^{(t-1)}\right), k = 1, \cdots, n;$
4. perform the QR decomposition $[Q, P] = \texttt{qr}(L^{(t)})$ and let $L^{(t)} = Q$;
5. compute $R_l^{(t)} = \left(\sum_{k:(k,l)\in\Omega} L_k^{(t)} L_k^{(t)^T}\right)^{-1} \left(\sum_{k:(k,l)\in\Omega} Y_{kl} L_k^{(t)}\right), l = 1, \cdots, p;$
6. **end while**
7. **output:** $X = LR$.

C. Maximum Likelihood Estimation for Matrix Completion

In this subsection, assuming knowledge of the matrix rank r, we present a simple alternating minimization algorithm to heuristically compute the maximum likelihood estimator of a low-rank matrix based on a few noise corrupted entries. When the SNR is relatively high, the algorithm performs very well. Suppose $X \in \mathbb{R}^{n\times p}$ is of rank r, then we can write X as

$$X = LR \quad \text{where} \quad L \in \mathbb{R}^{n\times r}, R \in \mathbb{R}^{r\times p}. \tag{56}$$

Assume that we observe a few entries of X that are corrupted by noise:

$$Y_{kl} = X_{kl} + W_{kl} = L_k^T R_l + W_{kl}, (k,l) \in \Omega \tag{57}$$

where L_k^T is the kth row of L, R_l is the lth column of R, W_{kl} is independent identically distributed (i.i.d.) Gaussian noise with variance σ^2, and Ω is the index set of all observed entries. The MLE of R and L are obtained by minimizing

$$\ell(L, R) = \sum_{(k,l)\in\Omega} \left(Y_{kl} - L_k^T R_l\right)^2. \tag{58}$$

We adopt an alternating minimization procedure. First, for fixed R, setting the derivative of $\ell(L, R)$ with respect to L_k to zero gives

$$L_k = \left(\sum_{l:(k,l)\in\Omega} R_l R_l^T\right)^{-1} \left(\sum_{l:(k,l)\in\Omega} Y_{kl} R_l\right) \tag{59}$$

for $k = 1, \ldots, n$. Similarly, when L is fixed, we get R_l

$$R_l = \left(\sum_{k:(k,l)\in\Omega} L_k L_k^T\right)^{-1} \left(\sum_{k:(k,l)\in\Omega} Y_{kl} L_k\right) \tag{60}$$

for $l = 1, \ldots, p$. Suppose $Y = U_0 \Lambda_0 V_0^T$ with $U_0 \in \mathbb{R}^{n\times r}, \Lambda_0 \in \mathbb{R}^{r\times r}$, and $V_0 \in \mathbb{R}^{p\times r}$ is the SVD of X. Setting $L^{(0)} = U_0$ and $R^{(0)} = \Lambda_0 V_0^T$, we then alternate between (59) and (60). To increase stability, we do a QR decomposition for $L^{(t)}$ and set the obtained orthogonal matrix as $L^{(t)}$. Hence, $L^{(t)}$ is always a matrix with orthogonal columns. The overall algorithm goes as shown in Table I.

IV. APPLICATION OF THE BARANKIN BOUND FOR LOW-RANK MATRIX ESTIMATION

In this section, we apply the Chapman–Robbins type Barankin bound to the low-rank matrix reconstruction problem (6) with Gaussian noise.

A. Chapman–Robbins Type Barankin Bound With One Test Point

We first consider the Chapman–Robbins bound with $N = 1$ test point $\mathbf{x}_1 \neq \mathbf{x}$. Suppose that $\hat{\mathbf{x}}$ is any globally unbiased estimator for $\mathbf{x} \in \mathcal{X}_r$, we derive lower bounds on the worst-case scalar MSE for $\hat{\mathbf{x}}$. The intent of this subsection is mainly to demonstrate the application of the Chapman–Robbins bound to low-rank matrix estimation without going into complicated matrix manipulations as in the next subsection. According to (13), (14), and (15), we have

$$\mathrm{MSE}_{\mathrm{w}} \geq \sup_{\mathbf{x}\neq\mathbf{x}_1} \frac{\|\delta\mathbf{x}\|_2^2}{\mathbb{E}\left(\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}}\right)^2} = \sup_{\mathbf{x}\neq\mathbf{x}_1} \frac{\|\mathbf{x}_1 - \mathbf{x}\|_2^2}{\int \frac{f_{\mathbf{x}_1}^2(\boldsymbol{y})}{f_{\mathbf{x}}(\boldsymbol{y})} \mathrm{d}\boldsymbol{y} - 1}. \tag{61}$$

Using the pdf for the Gaussian distribution, we calculate the integral in (61) as

$$\begin{aligned} \int \frac{f_{\mathbf{x}_1}^2(\boldsymbol{y})}{f_{\mathbf{x}}(\boldsymbol{y})} \mathrm{d}\boldsymbol{y} &= \exp[(\mathbf{x}_1 - \mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A} (\mathbf{x}_1 - \mathbf{x})] \\ &= \exp[(\delta\mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A} (\delta\mathbf{x})]. \end{aligned} \tag{62}$$

Since maximization with respect to $\mathbf{x}_1$ and $\mathbf{x}$ over $\mathcal{X}_r \backslash \{\mathbf{x}_1 = \mathbf{x}\}$ is equivalent to maximization with respect to $\delta\mathbf{x}$ over $\mathcal{X}_{2r} \backslash \{\mathbf{0}\}$, the lower bound in (61) becomes

$$\mathrm{MSE}_{\mathrm{w}} \geq \sup_{\delta\mathbf{x}\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}} \frac{\|\delta\mathbf{x}\|_2^2}{e^{(\delta\mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A} (\delta\mathbf{x})} - 1}. \tag{63}$$

In order to perform the maximization in (63), we establish the following lemma, whose proof is fairly easy (See also the proof of Lemma 5 in Appendix F).

Lemma 3: For any $\delta\mathbf{x} \in \mathcal{X}_{2r} \backslash \{\mathbf{0}\}$ and $t > 0$, define a function $g(t)$ as

$$g(t) \stackrel{\mathrm{def}}{=} \frac{t\|\delta\mathbf{x}\|_2^2}{e^{t(\delta\mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A} (\delta\mathbf{x})} - 1}, \quad t > 0. \tag{64}$$

Then, we have the following:

1) $g(t)$ is decreasing in t, and is strictly decreasing if $(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x}) > 0$;
2) $\lim_{t\to 0} g(t) = \frac{\|\delta\mathbf{x}\|_2^2}{(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})}$.

Since $\mathcal{X}_{2r}\backslash\{\mathbf{0}\} = \bigcup_{t>0}\{t(\delta\mathbf{x}) : \delta\mathbf{x} \in \mathcal{X}_{2r}\backslash\{\mathbf{0}\}, \|\delta\mathbf{x}\|_2 = 1\}$, Lemma 3 allows us to rewrite the maximization in (63) as

$$\begin{aligned}\mathrm{MSE_w} &\geq \max_{\substack{\delta\mathbf{x}\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta\mathbf{x}\|_2=1}} \sup_{t>0} \frac{t\|\delta\mathbf{x}\|_2^2}{e^{t(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})} - 1} \\ &= \max_{\substack{\delta\mathbf{x}\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta\mathbf{x}\|_2=1}} \frac{1}{(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})} \\ &= \max_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \frac{1}{\mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X)}. \end{aligned} \tag{65}$$

Note that when the supremum is achievable, we switch to the maximization notation.

Therefore, we have the following proposition:

Proposition 1: The worst-case scalar MSE for any globally unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathrm{MSE_w} \geq \frac{1}{\min_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X)}. \tag{66}$$

The result of Proposition 1 merits discussion. The quantity $\min_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X)$ is closely related to the matrix restricted isometry property (RIP) constant [33]:

$$\begin{aligned}\delta_{2r} = \max\Bigg\{ &1 - \min_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X), \\ &\max_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X) - 1\Bigg\} \end{aligned} \tag{67}$$

which is believed to guarantee stable low-rank matrix recovery. From Proposition 1 or from (63), it is easy to see that the worst-case MSE is infinite if $\min_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X) = 0$. This is because the model (6) is not identifiable under this condition. The requirement of $\min_{\substack{\delta X\in\mathcal{X}_{2r}\backslash\{\mathbf{0}\}\\ \|\delta X\|_{\mathrm{F}}=1}} \mathcal{A}(\delta X)^T\Sigma^{-1}\mathcal{A}(\delta X) > 0$ guarantees that there are not two matrices of rank r that give rise to the same measurement in the noiseless setting; i.e., $\mathcal{A}(X_1) \neq \mathcal{A}(X_2)$ for $X_1, X_2 \in \mathcal{X}_r$ and $X_1 \neq X_2$. This condition was discussed following Theorem 1; recall that it is not satisfied by the matrix completion problem.

B. The Worst-Case Scalar MSE Is Infinite

In this subsection, by optimizing the lower bound in (13) for multiple test points, we demonstrate that the worst-case scalar MSE is infinite even if the model is identifiable. Denote $H_{\mathbf{x}} \overset{\mathrm{def}}{=} \mathbb{E}\{[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}}]^T[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}}]\} \in \mathbb{R}^{N\times N}$. Then, we compute the (i,j)th element of $H_{\mathbf{x}}$ as

$$\begin{aligned}&\mathbb{E}_{\mathbf{x}}\left\{\frac{\delta_i f_{\mathbf{x}}}{f_{\mathbf{x}}}\frac{\delta_j f_{\mathbf{x}}}{f_{\mathbf{x}}}\right\} \\ &\quad= \mathbb{E}_{\mathbf{x}}\left\{\left(\frac{f_{\mathbf{x}_i}}{f_{\mathbf{x}}} - 1\right)\left(\frac{f_{\mathbf{x}_j}}{f_{\mathbf{x}}} - 1\right)\right\} \\ &\quad= \mathbb{E}_{\mathbf{x}}\left\{\frac{f_{\mathbf{x}_i} f_{\mathbf{x}_j}}{f_{\mathbf{x}}^2}\right\} - 1 = \int \frac{f_{\mathbf{x}_i}(\boldsymbol{y}) f_{\mathbf{x}_j}(\boldsymbol{y})}{f_{\mathbf{x}}(\boldsymbol{y})}\mathrm{d}\boldsymbol{y} - 1 \\ &\quad= \exp[(\delta_i\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta_j\mathbf{x})] - 1 \end{aligned} \tag{68}$$

where for the second equality we used $\mathbb{E}\frac{f_{\mathbf{x}_i}}{f_{\mathbf{x}}} = \mathbb{E}\frac{f_{\mathbf{x}_j}}{f_{\mathbf{x}}} = 1$. Note that (68) coincides with (62) when $i = j$. Therefore, we obtain $H_{\mathbf{x}} = \exp\odot[(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})] - \mathbf{1}\mathbf{1}^T$ where $\exp\odot(\cdot)$ is the elementwise exponential function of a matrix and $\mathbf{1}$ is the column vector with all ones.

Although the lower bound (13) uses a pseudoinverse for generality, the following lemma shows that $H_{\mathbf{x}}$ is always invertible. The proof, which is given in Appendix E, relies on the fact that any N Gaussian pdfs with a common covariance matrix but distinct means are linearly independent.

Lemma 4: If $\mathcal{A}(X) \neq 0$ for any matrix X with rank of at most $2r$, the covariance matrix $H_{\mathbf{x}}$ is positive definite.

The lower bound in Proposition 1 is not very strong since we have only one test point. In order to consider multiple test points, we extend Lemma 3 to the matrix case in the following lemma. For generality, we consider both $m \geq np$ and $m < np$. The proof is given in Appendix F.

Lemma 5: Suppose $\boldsymbol{A} \in \mathbb{R}^{m\times np}$ and $\delta\mathbf{x} \in \mathbb{R}^{np\times N}$ are matrices of full rank and $\mathrm{rank}(\boldsymbol{A}(\delta\mathbf{x})) = \min(m, np, N)$. For $t > 0$, define a matrix valued function $G(t)$ as

$$\begin{aligned}G(t) &\overset{\mathrm{def}}{=} t(\delta\mathbf{x})\{H(t)\}^\dagger(\delta\mathbf{x})^T \\ &\overset{\mathrm{def}}{=} t(\delta\mathbf{x})\{\exp\odot[t(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})] - \mathbf{1}\mathbf{1}^T\}^\dagger(\delta\mathbf{x})^T \end{aligned} \tag{69}$$

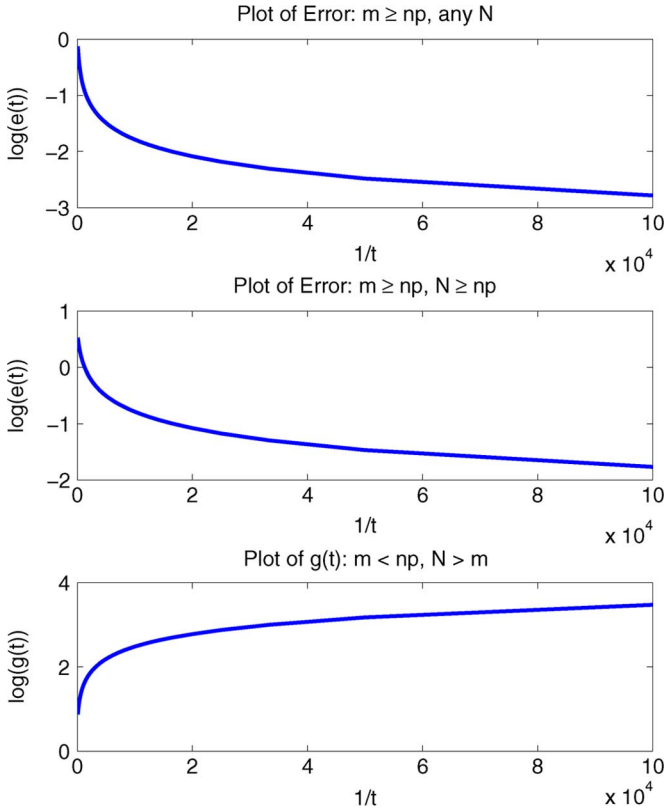
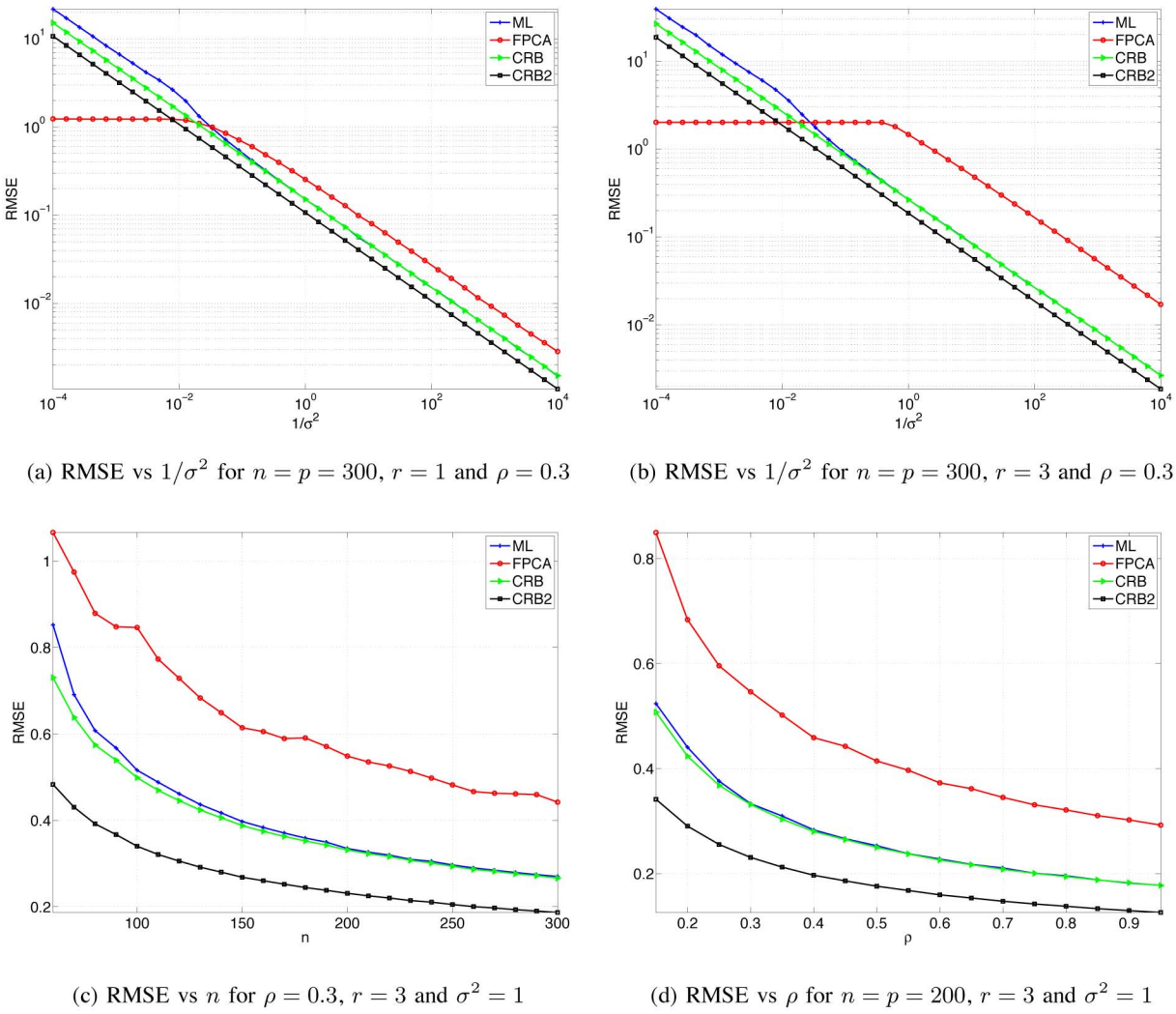


Fig. 2. The limiting behavior of $G(t)$ and $g(t)$. Note the horizontal axis is for $1/t$ instead of t. In the first and second plots, the error functions $e(t)$ are defined as $\log_{10}(\|G(t) - (\delta\mathbf{x})[(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})]^\dagger(\delta\mathbf{x})^T\|_2)$ and $\log_{10}(\|G(t) - (\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A})^{-1}\|_2)$, respectively. In the third plot, the function $g(t) = \mathrm{tr}(G(t))$.



(a) RMSE vs $1/\sigma^2$ for $n = p = 300$, $r = 1$ and $\rho = 0.3$

(b) RMSE vs $1/\sigma^2$ for $n = p = 300$, $r = 3$ and $\rho = 0.3$

(c) RMSE vs n for $\rho = 0.3$, $r = 3$ and $\sigma^2 = 1$

(d) RMSE vs ρ for $n = p = 200$, $r = 3$ and $\sigma^2 = 1$

Fig. 3. Performance of the MLE and the FPCA compared with two Cramér–Rao bounds.

and define the trace of $G(t)$ as

$$g(t) \stackrel{\text{def}}{=} \text{tr}(G(t)). \tag{70}$$

Then, the following hold.

1) $G(t)$ is strictly decreasing in the Löwner partial order if $\mathcal{A}(X) \neq 0$ for any matrix X with rank of at most $2r$.
2) $\lim_{t\to 0} G(t) = (\delta\mathbf{x})[(\delta\mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A}(\delta\mathbf{x})]^\dagger (\delta\mathbf{x})^T$ if $N \leq \min(m, np)$. In particular, if $m \geq np = N$, we have $\lim_{t\to 0} G(t) = (\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})^{-1}$.
3) $\lim_{t\to 0} g(t) = \infty$ if $m < np = N$.

Numerical simulations show that under the condition of Lemma 5 more general results hold.

1) If $m \geq np$, we always have $\lim_{t\to 0} G(t) = (\delta\mathbf{x})[(\delta\mathbf{x})^T \boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A}(\delta\mathbf{x})]^\dagger (\delta\mathbf{x})^T$. Furthermore, if $N \geq np$, we have $\lim_{t\to 0} G(t) = (\boldsymbol{A}^T \Sigma^{-1} \boldsymbol{A})^{-1}$.
2) If $m < np$, $N > m$ is sufficient for $\lim_{t\to 0} g(t) = \infty$.

We illustrate these results in Fig. 2. For example, the second subfigure shows that if $m \geq np$ and $N \geq np$, the matrix function $G(t)$ converges to the the classic Cramér–Rao bound. Although we have not yet been successful in showing the above general results analytically, the special cases covered by Lemma 5 suffice for our purpose.

Lemma 5 leads to the following theorem.

Theorem 4: Suppose that $m < np$, $\text{rank}(\boldsymbol{A}) = m$, and $\mathcal{A}(X) \neq 0$ for any matrix with rank of at most $2r$. Then, the worst-case scalar MSE for any unbiased estimator $\hat{\mathbf{x}}$ is $+\infty$. In addition, the infinite scalar MSE is achieved at any $\mathbf{x}$ such that $\text{rank}(\text{mat}_{n,p}(\mathbf{x})) < r$.

Proof: Consider any $\mathbf{x} \in \mathcal{X}_r$ such that $\text{rank}(\text{mat}_{n,p}(\mathbf{x})) < r$. For any $t > 0$, define $\mathbf{x}_i = \mathbf{x} + \sqrt{t}\boldsymbol{\eta}_i$ where $\boldsymbol{\eta}_i \in \mathbb{R}^{np}$ is the ith column of the np dimensional identity matrix I_{np}. Due to the rank inequality

$$\begin{aligned}&\text{rank}(\text{mat}_{n,p}(\mathbf{x}_i)) \\ &\quad \leq \text{rank}(\text{mat}_{n,p}(\mathbf{x})) + \text{rank}(\text{mat}_{n,p}(\sqrt{t}\boldsymbol{\eta}_i)) \leq r,\end{aligned} \tag{71}$$

we get $\mathbf{x}_i \in \mathcal{X}_r$. Thus, we obtain that $\delta\mathbf{x} = \sqrt{t}\text{I}_{np}$. According to Lemma 5, the trace of $B_{\mathbf{x}}$ in (13) can be made arbitrarily large by letting t approach 0. Hence, the MSE for $\mathbf{x}$ is $+\infty$ and all conclusions of the theorem hold. ■

Theorem 4 essentially says there is no globally unbiased estimator that has finite worst-case scalar MSE, even if the model is identifiable. Actually there is no locally unbiased estimator with finite worst-case scalar MSE for matrix X with rank less than r. Fortunately these sets of matrices form a measure zero subset of $\mathcal{X}_r$. The applicability of the key lemma (Lemma 5) to the worst-case scalar MSE also hinges on the fact that the signal/parameter vector can be arbitrarily small, which enables

TABLE II
PARAMETER CONFIGURATION FOR SIMULATIONS

Experiment	n	rank r	ρ	s^2	$\log_{10}\sigma^2$	T
RMSE vs σ^2 (Fig. 3a)	300	1	0.3	$\frac{20}{\sqrt{n}}$	$-4:0.1:4$	10
RMSE vs σ^2 (Fig. 3b)	300	3	0.3	$\frac{20}{\sqrt{n}}$	$-4:0.1:4$	10
RMSE vs n (Fig. 3c)	$60:10:300$	3	0.3	1	0	40
RMSE vs ρ (Fig. 3d)	200	3	$0.15:0.05:0.95$	1	0	20

us to drive the parameter t in $G(t)$ to zero. We might obtain a bounded MSE if we impose additional restrictions on the parameter space, for example, a positive threshold for the absolute values of nonzero singular values.

V. NUMERICAL SIMULATIONS

In this section, we show several numerical examples to demonstrate the performances of the FPCA that solves the matrix LASSO estimator (3) and the MLE algorithm given in Table I, and compare them with the derived constrained CRB (26).

We describe the experiment setup for the constrained CRB first. In four experiments, whose results are shown in Fig. 3, we generated a rank r matrix $X = s^2LR \in \mathbb{R}^{n\times p}$. We took $n = p$ in all experiments. Here $L \in \mathbb{R}^{n\times r}$ and $R \in \mathbb{R}^{r\times p}$ were matrices whose entries followed i.i.d. Gaussian distribution with mean zero and variance one, and s^2 was the signal level. We randomly observed $m = \lceil \rho np \rceil$ entries of X corrupted by white Gaussian noise of variance σ^2. We then ran the FPCA and the MLE algorithms T times for different realizations of the noise and recorded the averaged MSE. The parameter configurations of the four experiments are summarized in Table II. The parameter μ in the matrix LASSO estimator (3) solved by the FPCA was set to be $\mu = \max(4(\sqrt{n}+\sqrt{p})\sqrt{\rho}\sigma, 10^{-8})$, or $\mu = \max(4(\sqrt{n}+\sqrt{p})\sqrt{\rho}\sigma, 10^{-4})$ depending on whether the optimization problem was "hard" or "easy" [12]. We varied different parameters, e.g., the noise level σ^2, the matrix size n and p, and the fraction of observed entries ρ, and plotted the root mean-square error (RMSE) as a function of these varying variables. Note that the RMSE is related to the scalar MSE by $\mathrm{RMSE} = \sqrt{\mathrm{MSE}/(np)}$, i.e., it is the square root of the per-element error. In addition to the empirical RMSE, we also plotted the constrained CRB (26) and the relaxed bound (34) as a function of the varying variables. For comparison with the RMSE, we also divided them by np and took the square root. In Fig. 3, the bound given by (26) is labeled CRB and that of (34) is labeled CRB2.

In Fig. 3(a) and (b), we see that the FPCA performs better than both the MLE and the predictions of the constrained CRB for high levels of noise. This is achieved by introducing large bias toward the zero matrix. However, for relatively high SNR, the performance of the MLE is better than that of the FPCA, especially when the matrix rank is high. This finding confirms that biased estimators are suboptimal when the signal is strong and implies that there is room to improve the performance of current matrix completion techniques in the relatively high SNR region. In addition, the constrained CRB (26) predicts the behavior of the MLE very well. It also serves as a lower bound on the performance of the matrix LASSO estimator for low levels of noise. However, the constrained CRB fails to capture the threshold phenomenon, in which the performance of the MLE suddenly improves at certain SNR level. This threshold phenomenon is observed in many signal processing problems and can be captured by considering bounds tighter than the CRB. We also notice a gap between the constrained CRB (26) and its relaxed version (34) labeled CRB2 in the figure. In addition, from Fig. 3(c) and (d) we see that the performance of both the FPCA and MLE improve as n and ρ increase, which is correctly predicted by the constrained CRB.

VI. CONCLUSION

We analyzed the behavior of the MSE matrix and the scalar MSE of locally and globally unbiased estimators for low-rank matrix reconstruction. Compared with the performance analysis of low-rank matrix recovery for specific algorithms, these lower bounds apply to any unbiased estimator. The global and local unbiasedness requirements are related to the two kinds of stability problems raised in low-rank matrix reconstruction: stability results applying to all low-rank matrices and those applying to a specific low-rank matrix. We derived a constrained CRB for any locally unbiased estimator and showed that the predicted performance bound is approached by the MLE. Due to the good performance of the MLE, our ongoing work involves designing more efficient implementations of the basic MLE algorithm presented in this paper, demonstrating its convergence, and incorporating procedures to automatically estimate the rank r. We also demonstrated that the worst-case MSE for any globally unbiased estimator is infinite, which is achieved by matrices of rank strictly less than r.

APPENDIX A
PROOF OF (15)

Proof: Note that in the lower bound on the scalar MSE [see (14)]

$$\mathrm{MSE}_{\mathbf{x}} \geq \mathrm{tr}(B_{\mathbf{x}}), \tag{72}$$

the right-hand side depends on the integer N and the test points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{X}_r$, while the left-hand side does not depend on these quantities. The tightest bound on the scalar MSE for any globally unbiased estimator of a particular $\mathbf{x}$ is obtained by maximizing $\mathrm{tr}(B_{\mathbf{x}})$ over all integers $N \geq 1$ and all possible test points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{X}_r$:

$$\mathrm{MSE}_{\mathbf{x}} \geq \sup_{N,\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}}). \tag{73}$$

Intuitively, the tightest bound on the worst-case scalar MSE is obtained by taking an additional maximization over all possible $\mathbf{x}$. We develop this intuition more rigorously in the following. For any $\epsilon > 0$, suppose there exists $\mathbf{x}^* \in \mathcal{X}_r$ such that

$$\begin{aligned}\sup_{N,\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}^*}) &\geq \sup_{\mathbf{x}} \sup_{N,\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}}) - \epsilon \\ &= \sup_{N,\mathbf{x},\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}}) - \epsilon. \end{aligned} \tag{74}$$

Then, we have

$$\begin{aligned}\mathrm{MSE}_{\mathrm{w}} &= \sup_{\mathbf{x}} \mathrm{MSE}_{\mathbf{x}} \geq \mathrm{MSE}_{\mathbf{x}^*} \\ &\geq \sup_{N,\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}^*}) \\ &\geq \sup_{N,\mathbf{x},\mathbf{x}_1,\ldots,\mathbf{x}_N} \mathrm{tr}(B_{\mathbf{x}}) - \epsilon. \end{aligned} \tag{75}$$

Due to the arbitrariness of ϵ, we have proved (15). ∎

Appendix B
Proof of Lemma 1

Proof: The conditions of the lemma imply that there exists a full rank matrix $Q_1 \in \mathbb{R}^{N_2 \times N_1}$ such that $P_1 = P_2 Q_1$. Since $P_2^T J_{\mathbf{x}} P_2$ is positive definite, we construct a matrix $Q_2 \in \mathbb{R}^{N_2 \times (N_2 - N_1)}$ with $Q_1^T (P_2^T J_{\mathbf{x}} P_2) Q_2 = O$, the $N_1 \times (N_2 - N_1)$ zero matrix, through the Gram-Schmidt orthogonalization process with inner product defined by $P_2^T J_{\mathbf{x}} P_2$. Define the $N_2 \times N_2$ nonsingular matrix $Q \overset{\mathrm{def}}{=} [\, Q_1 \quad Q_2 \,]$. Then, we have

$$\begin{aligned} & P_2 \left[P_2^T J_{\mathbf{x}} P_2\right]^{\dagger} P_2^T \\ &= P_2 Q \left[Q^T P_2^T J_{\mathbf{x}} P_2 Q\right]^{\dagger} Q^T P_2^T \\ &= P_2 Q \left[\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} P_2^T J_{\mathbf{x}} P_2 [\, Q_1 \quad Q_2 \,]\right]^{\dagger} Q^T P_2^T \\ &= P_2 Q \begin{bmatrix} Q_1^T P_2^T J_{\mathbf{x}} P_2 Q_1 & O \\ O & Q_2^T P_2^T J_{\mathbf{x}} P_2 Q_2 \end{bmatrix}^{\dagger} Q^T P_2^T \\ &= [\, P_1 \quad P_2 Q_2 \,] \\ &\quad \times \begin{bmatrix} \left[P_1^T J_{\mathbf{x}} P_1\right]^{\dagger} & O \\ O & \left[Q_2^T P_2^T J_{\mathbf{x}} P_2 Q_2\right]^{\dagger} \end{bmatrix} \begin{bmatrix} P_1^T \\ Q_2^T P_2^T \end{bmatrix} \\ &= P_1 \left[P_1^T J_{\mathbf{x}} P_1\right]^{\dagger} P_1^T + P_2 Q_2 \left[Q_2^T P_2^T J_{\mathbf{x}} P_2 Q_2\right]^{\dagger} Q_2^T P_2^T \\ &\geq P_1 \left[P_1^T J_{\mathbf{x}} P_1\right]^{\dagger} P_1^T \end{aligned} \tag{76}$$

where for the first equality we relied on the fact that $[Q^T J Q]^{\dagger} = Q^{-1} J^{\dagger} Q^{-T}$ if Q is invertible. ∎

Appendix C
Proof of Lemma 2

Proof: The proof essentially follows [34]. Denote $M = (n+p)r - r^2$. Using the comparison principle, we replace the Rademacher sequence $\{\varepsilon_i\}$ in (47) by the standard Gaussian sequence $\{g_i\}$ and then apply Dudley's inequality [35]:

$$\begin{aligned} \mathbb{E} \left\| \sum_{i=1}^m g_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \right\|_2 &= \mathbb{E} \sup_{\mathbf{z} \in B_2} \left| \sum_{i=1}^m g_i \langle \boldsymbol{\xi}_i, \mathbf{z} \rangle^2 \right| \\ &\leq C \int_0^{\infty} \log^{1/2} N(B_2, d, u)\, \mathrm{d}u \end{aligned} \tag{77}$$

where B_2 is the unit ball in $(\mathbb{R}^M, \|\cdot\|_2)$, and $N(B_2, d, u)$ is the minimal number of balls with radius u under metric d that covers the set B_2. The metric d is defined by the Gaussian process as follows:

$$\begin{aligned} d(\mathbf{z}_1, \mathbf{z}_2) &= \left[\sum_{i=1}^m (\langle \boldsymbol{\xi}_i, \mathbf{z}_1 \rangle^2 - \langle \boldsymbol{\xi}_i, \mathbf{z}_2 \rangle^2)^2 \right]^{1/2} \\ &\leq \left[\sum_{i=1}^m \left(\langle \boldsymbol{\xi}_i, \mathbf{z}_2 \rangle + \langle \boldsymbol{\xi}_i, \mathbf{z}_2 \rangle \right)^2 \right]^{1/2} \\ &\quad \times \max_{i \leq m} |\langle \boldsymbol{\xi}_i, \mathbf{z}_1 - \mathbf{z}_2 \rangle| \\ &\leq 2 \max_{\mathbf{z} \in B_2} \left[\sum_{i=1}^m \langle \boldsymbol{\xi}_i, \mathbf{z} \rangle^2 \right]^{1/2} \max_{i \leq m} |\langle \boldsymbol{\xi}_i, \mathbf{z}_1 - \mathbf{z}_2 \rangle| \\ &= 2 \left\| \sum_{i=1}^m \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \right\|_2^{1/2} \max_{i \leq m} |\langle \boldsymbol{\xi}_i, \mathbf{z}_1 - \mathbf{z}_2 \rangle|. \end{aligned} \tag{78}$$

As a consequence, we have

$$\begin{aligned} \int_0^{\infty} \log^{1/2} N(B_2, d, u) \mathrm{d}u \leq 2\sqrt{M} \left\| \sum_{i=1}^m \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \right\|_2^{1/2} \\ \times \int_0^{\infty} \log^{1/2} N\left(\frac{1}{\sqrt{M}} B_2, \|\cdot\|_{\boldsymbol{\xi}}, u \right) \mathrm{d}u \end{aligned} \tag{79}$$

where $\|\cdot\|_{\boldsymbol{\xi}} = \max_{i \leq m} |\langle \boldsymbol{\xi}_i, \cdot \rangle|$. Now note the following containments:

$$\frac{1}{\sqrt{M}} B_2 \subseteq B_1 \subseteq \mu_B B_{\boldsymbol{\xi}} \tag{80}$$

where $B_{\boldsymbol{\xi}}$ is the unit ball under $\|\cdot\|_{\boldsymbol{\xi}}$. We use the following two estimates on the covering numbers:

$$\log N(B_1, \|\cdot\|_{\boldsymbol{\xi}}, u) = C \mu_B^2 \log(2M) \log M / u^2 \tag{81}$$

$$\log N(B_{\boldsymbol{\xi}}, \|\cdot\|_{\boldsymbol{\xi}}, u) = M \log(1 + 2\mu_B / u). \tag{82}$$

We compute the Dudley integral as follows:

$$\begin{aligned} & \int_0^{\infty} \log^{1/2} N\left(\frac{1}{\sqrt{M}} B_2, \|\cdot\|_{\boldsymbol{\xi}}, u \right) \mathrm{d}u \\ &\leq \int_0^{A} \log^{1/2} N(B_{\boldsymbol{\xi}}, \|\cdot\|_{\boldsymbol{\xi}}, u) \mathrm{d}u \\ &\quad + \int_A^{\mu_B} \log^{1/2} N(B_1, \|\cdot\|_{\boldsymbol{\xi}}, u) \mathrm{d}u \\ &\leq CA\sqrt{M} \log(1 + 2/A) \\ &\quad + C \log(1/A) \sqrt{\log m} \sqrt{\log M}. \end{aligned} \tag{83}$$

Choosing $A = 1/\sqrt{M}$ yields an upper bound of the form

$$\log^{3/2} M \sqrt{\log m}. \tag{84}$$

∎

APPENDIX D
PROOF OF THEOREM 2

Proof: Denote the left-hand side of (46) by E and denote $M = (n+p)r - r^2$. Conditioned on a choice of Ω and using Lemma 2, we get

$$E \le \frac{C\sqrt{M}\log^{3/2} M\sqrt{\log m}}{\sqrt{m}}\mathbb{E}\left\|\frac{1}{m}\sum_{i\in\Omega}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right\|_2^{1/2}$$
$$\le \frac{C\sqrt{M}\log^{3/2} M\sqrt{\log m}}{\sqrt{m}}(E+1)^{1/2} \quad (85)$$

which implies that

$$E \le \frac{2C\sqrt{M}\log^{3/2} M\sqrt{\log m}}{\sqrt{m}} \quad (86)$$

as long as $C\sqrt{M}\log^{3/2} M\sqrt{\log m}/\sqrt{m} \le 1/2$. Thus, if we take

$$m \ge C\frac{M\log^3 M}{\varepsilon^2}\log\left(\frac{M}{\varepsilon^2}\right), \quad (87)$$

then we have (46). ∎

APPENDIX E
PROOF OF LEMMA 4

Proof: Suppose $\boldsymbol{a} \in \mathbb{R}^N$ is a nonzero column vector. We have

$$\boldsymbol{a}^T H_{\mathbf{x}}\boldsymbol{a} = \mathbb{E}\left\{\left[\sum_{i=1}^N \boldsymbol{a}_i\left(\frac{f_{\mathbf{x}_i}}{f_{\mathbf{x}}} - 1\right)\right]^2\right\}$$
$$= \int \frac{\left(\sum_{i=1}^N \boldsymbol{a}_i f_{\mathbf{x}_i} - \left(\sum_{i=1}^N \boldsymbol{a}_i\right) f_{\mathbf{x}}\right)^2}{f_{\mathbf{x}}}\mathrm{d}\boldsymbol{y}. \quad (88)$$

Therefore, the quadratic form $\boldsymbol{a}^T H_{\mathbf{x}}\boldsymbol{a} = 0$ is equivalent to $\sum_{i=1}^N \boldsymbol{a}_i f_{\mathbf{x}_i} - (\sum_{i=1}^N \boldsymbol{a}_i) f_{\mathbf{x}} = 0$. If we can show that any $N+1$ Gaussian functions $\mathcal{F} \overset{\text{def}}{=} \{f_{\mathbf{x}_0}, f_{\mathbf{x}_1}, \ldots, f_{\mathbf{x}_N}\}$ with $\mathbf{x}_0 = \mathbf{x}$ are linearly independent, then $H_{\mathbf{x}}$ is positive definite under the lemma's conditions. To this end, we compute that the Gram matrix associated with $\mathcal{F}$ is

$$G = \exp\odot\left[-\frac{1}{2}\|\Sigma^{-1/2}A\mathbf{x}_i - \Sigma^{-1/2}A\mathbf{x}_j\|_2^2\right]. \quad (89)$$

We note that if $\mathcal{A}(X) \ne 0$ for any matrix X with rank of at most $2r$, the mean vectors $\{A\mathbf{x}_i\}_{i=0}^p$ for the Gaussian pdfs in $\mathcal{F}$ are distinct. According to [36, p. 14], the Gram matrix G is nonsingular, which implies that the functions in $\mathcal{F}$ are linearly independent. ∎

APPENDIX F
PROOF OF LEMMA 5

Proof:

1) If $\mathcal{A}(X) \ne 0$ for any matrix with rank of at most $2r$, according to Lemma 4, $H(t)$ is positive definite. Hence, the pseudoinverse in the definition of $G(t)$ is actually an inverse. Taking the derivative of $G(t)$ with respect to t yields

$$\frac{\mathrm{d}G(t)}{\mathrm{d}t}$$
$$= (\delta\mathbf{x})\{H(t)\}^{-1}(\delta\mathbf{x})^T + t(\delta\mathbf{x})\left\{\frac{\mathrm{d}[H(t)^{-1}]}{\mathrm{d}t}\right\}(\delta\mathbf{x})^T$$
$$= (\delta\mathbf{x})\{H(t)\}^{-1}(\delta\mathbf{x})^T - t(\delta\mathbf{x})\left\{[H(t)]^{-1}\frac{\mathrm{d}H(t)}{\mathrm{d}t}[H(t)]^{-1}\right\}(\delta\mathbf{x})^T$$
$$= -(\delta\mathbf{x})\left\{[H(t)]^{-1}\left(t\frac{\mathrm{d}H(t)}{\mathrm{d}t} - H(t)\right)[H(t)]^{-1}\right\}(\delta\mathbf{x})^T. \quad (90)$$

It suffices to show that $t\frac{\mathrm{d}H(t)}{\mathrm{d}t} - H(t)$ is positive semidefinite. To this end, suppose $\boldsymbol{a} = [a_i] \in \mathbb{R}^N$ is a nonzero column vector, and construct the function

$$h(t) = \boldsymbol{a}^T\left(t\frac{\mathrm{d}H(t)}{\mathrm{d}t} - H(t)\right)\boldsymbol{a}$$
$$= \sum_{i,j}[tb_{ij}\exp(tb_{ij}) + 1 - \exp(tb_{ij})]a_i a_j \quad (91)$$

where $b_{ij} = (\delta_i\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta_j\mathbf{x})$. Obviously, we have $h(0) = 0$. Therefore, it suffices to show that $h'(t) > 0$ for $t > 0$. Taking the derivative of $h(t)$ gives

$$h'(t) = t\sum_{i,j} b_{ij}^2\exp(tb_{ij})a_i a_j$$
$$= t\boldsymbol{a}^T(B\circ B\circ\exp\odot(tB))\boldsymbol{a} \quad (92)$$

where $B = [b_{ij}]_{N\times N}$. We note that $\exp\odot(tB) = H(t) + \mathbf{1}^T\mathbf{1}$ is a positive definite matrix and B is a positive semidefinite matrix with positive diagonal elements under the lemma's conditions. Therefore, according to the Shur product theorem [37, Theorem 7.5.3, p. 458] and [38, Theorem 8.17, p. 300], the Hadamard product $B\circ B\circ\exp\odot(tB)$ is positive definite. Therefore, the function $h(t)$ is increasing, which implies that $\mathrm{d}G(t)/\mathrm{d}t$ is negative definite. Thus, we conclude that $G(t)$ is strictly increasing in t in the Löwner partial order.

2) For the second claim, if $N \le \min(m, np)$, the full-rankness of $\boldsymbol{A}(\delta\mathbf{x})$ implies that $\mathrm{rank}(\boldsymbol{A}(\delta\mathbf{x})) = N$. Therefore, both $H(t)$ and $(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})$ are invertible. The conclusion follows from the continuity of the matrix inverse.

3) If $N = np$ and $\delta\mathbf{x}$ is of full rank, we get

$$G(t) = \left\{(\delta\mathbf{x})^{-T}\frac{H(t)}{t}(\delta\mathbf{x})^{-1}\right\}^{-1} \quad (93)$$

and

$$g(t) = \sum_{i=1}^N \frac{1}{\lambda_i\left(\left\{(\delta\mathbf{x})^{-T}\frac{H(t)}{t}(\delta\mathbf{x})^{-1}\right\}\right)} \quad (94)$$

where $\lambda_i(\cdot)$'s are the eigenvalues of a matrix in the decreasing order. Due to the continuity of matrix eigenvalues with respect to matrix entries, we have

$$\lim_{t\to 0}\lambda_i\left(\left\{(\delta\mathbf{x})^{-T}\frac{H(t)}{t}(\delta\mathbf{x})^{-1}\right\}\right)$$
$$= \lambda_i\left(\lim_{t\to 0}\left\{(\delta\mathbf{x})^{-T}\frac{H(t)}{t}(\delta\mathbf{x})^{-1}\right\}\right)$$
$$= \lambda_i\left(\left\{(\delta\mathbf{x})^{-T}(\delta\mathbf{x})^T\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}(\delta\mathbf{x})(\delta\mathbf{x})^{-1}\right\}\right)$$
$$= \lambda_i(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}). \quad (95)$$

Since $\mathrm{rank}(\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}) = m < np = N$, the last $np - m$ eigenvalues of $\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A}$ are zeros. Therefore, we obtain

$$\lim_{t\to 0} g(t) = \infty. \tag{96}$$

■

REFERENCES

[1] L. El Ghaoui and P. Gahinet, "Rank minimization under LMI constraints: A framework for output feedback problems," presented at the Eur. Control Conf., Groningen, The Netherlands, Jun. 28, 1993.

[2] 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.

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

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

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

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

[11] 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.

[12] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Programm.*, vol. 120, no. 2, pp. 1–33, 2009.

[13] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér–Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[14] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.

[15] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér–Rao bound," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2247–2249, Jun. 1993.

[16] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[17] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.

[18] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér-Rao bound with a singular fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 453–456, Jun. 2009.

[19] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, "Unbiased estimation of a sparse vector in white Gaussian noise," ArXiv E-Prints May 2010 [Online]. Available: http://arxiv.org/abs/1005.5697

[20] J. M. Hammersley, "On estimating restricted parameters," *J. Roy. Stat. Soc. Series B*, vol. 12, no. 2, pp. 192–240, 1950.

[21] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Ann. Math. Stat.*, vol. 22, no. 4, pp. 581–586, 1951.

[22] R. McAulay and E. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Trans. Inf. Theory*, vol. 17, no. 6, pp. 669–676, Nov. 1971.

[23] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Stat. Planning Inference*, vol. 13, pp. 297–318, 1986.

[24] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[25] Y. Guo and B. C. Levy, "Worst-case MSE precoder design for imperfectly known mimo communications channels," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2918–2930, Aug. 2005.

[26] Y. C. Eldar, "Minimax MSE estimation of deterministic parameters with noise covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 138–145, Jan. 2006.

[27] P. A. Parker, P. Mitran, D. W. Bliss, and V. Tarokh, "On bounds and algorithms for frequency synchronization for collaborative communication systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3742–3752, Aug. 2008.

[28] A. Funai and J. A. Fessler, "Cramér Rao bound analysis of joint b1/t1 mapping methods in MRI," in *IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, Apr. 2010, pp. 712–715.

[29] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas With Application to Linear Systems Theory*. Princeton, NJ: Princeton Univ. Press, 2005.

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

[31] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 40th Conf. Inf. Sci. Syst. (CISS 2006)*, Mar. 2006, pp. 207–212.

[32] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

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

[34] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, pp. 1025–1045, 2008.

[35] M. Talagrand, *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. New York: Springer, 2005.

[36] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1990.

[38] J. R. Schott, *Matrix Analysis for Statistics*, 2nd ed. New York: Wiley-Interscience, 2005, 1.

Gongguo Tang (S'09–M'11) received the B.Sc. degree in mathematics from the Shandong University, China, in 2003, the M.Sc. degree in systems science from the Chinese Academy of Sciences, China, in 2006, and the Ph.D. degree in electrical and systems engineering from Washington University in 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 sparse signal processing, matrix completion, mathematical programming, statistical signal processing, detection and estimation, and their applications.

Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, and the Ph.D. degree from Stanford University, Stanford, CA.

He was formerly a faculty member at Yale University and the University of Illinois at Chicago. He is currently the Eugene and Martha Lohman Professor and Chair of the Department of Electrical and Systems Engineering at Washington University in St. Louis (WUSTL) and serves as the Director of the Center for Sensor Signal and Information Processing at WUSTL.

Dr. Nehorai served as Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002. From 2003 to 2005, he was Vice-President (Publications) of the IEEE Signal Processing Society (SPS), Chair of the Publications Board, and member of the Executive Committee of this Society. He was the Founding Editor of the special columns on Leadership Reflections in the *IEEE Signal Processing Magazine* from 2003 to 2006. He received the 2006 IEEE SPS Technical Achievement Award and the 2010 IEEE SPS Meritorious Service Award. He was elected Distinguished Lecturer of the IEEE SPS for the term 2004 to 2005. He was corecipient of the IEEE SPS 1989 Senior Award for Best Paper coauthor of the 2003 Young Author Best Paper Award and corecipient of the 2004 Magazine Paper Award. In 2001, he was named University Scholar of the University of Illinois. He is the Principal Investigator of the Multidisciplinary University Research Initiative (MURI) project entitled Adaptive Waveform Diversity for Full Spectral Dominance. He has been a Fellow of the Royal Statistical Society since 1996.