BAYESIAN CRAMER-RAO BOUNDS FOR FACTORIZED MODEL BASED LOW RANK MATRIX RECONSTRUCTION

Martin Sundin, Saikat Chatterjee and Magnus Jansson

ACCESS Linnaeus Center, School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm, Sweden
masundi@kth.se, sach@kth.se, magnus.jansson@ee.kth.se

ABSTRACT

Low-rank matrix reconstruction (LRMR) considers estimation (or reconstruction) of an underlying low-rank matrix from linear measurements. A low-rank matrix can be represented using a factorized model. In this article, we derive Bayesian Cramér-Rao bounds for LRMR where a factorized model is used. We first show a general informative bound, and then derive Bayesian Cramér-Rao bounds for different scenarios. We consider a low-rank random matrix model with hyper-parameters that are - deterministic known, deterministic unknown and random. Finally we compare the bounds with existing estimation algorithms through numerical simulations.

***Index Terms*—** Low-rank matrix reconstruction, matrix completion, Bayesian estimation, Cramér-Rao bounds.

1. INTRODUCTION

In the low-rank matrix reconstruction (LRMR) problem, a low-rank matrix $\mathbf{X} \in \mathbb{R}^{p\times q}$ is measured as

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

where $\mathbf{y} \in \mathbb{R}^m$ is measurements, $\mathbf{n} \in \mathbb{R}^m$ is $\mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I}_m)$ measurement noise with precision $\beta > 0$, $\text{vec}(\cdot)$ is a standard vectorization operator and the linear sensing operator $\mathcal{A} : \mathbb{R}^{p\times q} \to \mathbb{R}^m$ and the matrix $\mathbf{A} \in \mathbb{R}^{m\times pq}$ are two equivalent representations of the sensing process. The problem is to reconstruct $\mathbf{X}$ from the measurements $\mathbf{y}$. LRMR has applications in e.g. system identification [1–3] and recommendation systems [1, 4–10]. In several applications, the LRMR problem setup (1) is under-determined, i.e. $m < pq$. Here we mention that low-rank matrix completion is an important special case of LRMR where the sensing operator $\mathbf{A}$ has a special structure.

There exists several reconstruction algorithms for LRMR [3–9, 11]. A natural question is how to benchmark the performance of algorithms against theoretical bounds. In this paper, we address the question for Bayesian algorithms where the low-rank matrix is modeled as the product

$$\mathbf{X} = \mathbf{L}\mathbf{R}^\top, \tag{2}$$

where $\mathbf{L} \in \mathbb{R}^{p\times r}$, $\mathbf{R} \in \mathbb{R}^{q\times r}$ and $r < \min(p, q)$ is a user-defined constant. Note that $\text{rank}(\mathbf{X}) \leq r$. We refer to the model (2) as the factorized low-rank matrix model. The model is the basis for several algorithms, e.g. [3, 5–9, 11].

Our contribution in this article is to derive Bayesian Cramér-Rao bounds (BCRB) for the LRMR problem that uses the factorized model (2). Through numerical simulations, we compare the performance of Bayesian LRMR algorithms against the BCRB bounds. At this point we mention that there exists bounds for the deterministic scenario of LRMR, such as Cramér-Rao bounds for unstructured [9, 11] and structured [3] low-rank matrices. In the following subsections, we explain notations used in the article and provide preliminaries of BCRB.

1.1. Notations

We use $\mathcal{E}_{\mathbf{q}}[\cdot]$ to denote the expectation value of a random variable with respect to variables $\mathbf{q}$ and use the symbol $\otimes$ for the Kronecker product. The ℓ_2-norm and Frobenius norm are denoted by $\|\cdot\|$ and the $k \times k$ identity matrix by $\mathbf{I}_k$. We denote the (i, j)'th component of a matrix $\mathbf{X}$ by $[\mathbf{X}]_{ij}$ and the i'th standard unit vector by $\mathbf{e}_i$, that is $[\mathbf{e}_i]_j = 1$ if $j = i$ and zero otherwise. The *commutation matrix* $\mathbf{K}_{p,q} \in \mathbb{R}^{pq\times pq}$ is the matrix representation of the transpose operation, i.e. $\mathbf{K}_{p,q}\text{vec}(\mathbf{Z}) = \text{vec}(\mathbf{Z}^\top)$ for all $\mathbf{Z} \in \mathbb{R}^{p\times q}$. We also introduce the linear operators $\mathcal{T}_1$ and $\mathcal{T}_2$ which operate on Kronecker products as $\mathcal{T}_1(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{C}^\top \otimes \mathbf{D})$ and $\mathcal{T}_2(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{C} \otimes \mathbf{D}^\top)$ where $\mathbf{C} \in \mathbb{R}^{q\times q}$ and $\mathbf{D} \in \mathbb{R}^{p\times p}$. For ease of notation, we often set $\mathbf{x} \triangleq \text{vec}(\mathbf{X})$. Henceforth we use the variables $\mathbf{x}$ and $\mathbf{X}$ interchangeably and their explicit use will be clear from the context.

1.2. The factorized model

The factorized model promotes low rank by incurring column-wise block-sparsity in $\mathbf{L}$ and $\mathbf{R}$. For achieving column-wise

This work was partially supported by the Swedish Research Council under contract 2015-05484.

Table 1. BCRB for different cases

	BCRB-I	BCRB-II	BCRB-III
Variable $\mathbf{x} \triangleq \mathbf{X} = \text{vec}(\mathbf{L}\mathbf{R}^\top)$	Random	Random	Random
hyper-parameters $\boldsymbol{\gamma}, \beta$	Deterministic known	Deterministic unknown	Random
Relevant performance measures	$\mathcal{E}_{\mathbf{y},\mathbf{w}}\left[\lVert\mathbf{x}-\hat{\mathbf{x}}\rVert^2\right]$	$\mathcal{E}_{\mathbf{y},\mathbf{w}}\left[\lVert\mathbf{x}-\hat{\mathbf{x}}\rVert^2\right]$	$\mathcal{E}_{\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\beta}\left[\lVert\mathbf{x}-\hat{\mathbf{x}}\rVert^2\right]$
	-	$\mathcal{E}_{\mathbf{y},\mathbf{w}}\left[\lVert\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}\rVert^2\right]$	$\mathcal{E}_{\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\beta}\left[\lVert\boldsymbol{\gamma}-\hat{\boldsymbol{\gamma}}\rVert^2\right]$
	-	$\mathcal{E}_{\mathbf{y},\mathbf{w}}[(\hat{\beta}-\beta)^2]$	$\mathcal{E}_{\mathbf{y},\mathbf{w},\boldsymbol{\gamma},\beta}[(\hat{\beta}-\beta)^2]$

block-sparsity, one approach is to use the priors [6–8]

$$\begin{aligned} p(\mathbf{L}|\boldsymbol{\gamma}) &= \frac{|\boldsymbol{\Gamma}|^{p/2}}{(2\pi)^{pr/2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top)\right), \\ p(\mathbf{R}|\boldsymbol{\gamma}) &= \frac{|\boldsymbol{\Gamma}|^{q/2}}{(2\pi)^{qr/2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{R}\boldsymbol{\Gamma}\mathbf{R}^\top)\right), \end{aligned} \tag{3}$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_r]^\top$ and $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$. Here $\gamma_i > 0$ is the precision of ith column vector of $\mathbf{L}$ and $\mathbf{R}$. The precisions $\boldsymbol{\gamma}$ and β are typically assigned Gamma distributions

$$p(\gamma_i) = \text{Gamma}(\gamma_i|a, b) = \frac{b^a \gamma_i^{a-1} e^{-b\gamma_i}}{\Gamma(a)}, \tag{4}$$

$$p(\beta) = \text{Gamma}(\beta|c, d) = \frac{d^c \beta^{c-1} e^{-d\beta}}{\Gamma(c)}, \tag{5}$$

where $\Gamma(\cdot)$ denotes the Gamma function, when they are random. For ease of notation, we let $\mathbf{l}_i$ and $\mathbf{r}_i$ denote the i'th column vector of $\mathbf{L}$ and $\mathbf{R}$, respectively, and set $\mathbf{w} = [\text{vec}(\mathbf{L})^\top \; \text{vec}(\mathbf{R})^\top]^\top$. The joint distribution of the random variables is

$$p(\mathbf{y}, \mathbf{w}, \boldsymbol{\gamma}, \beta) = p(\mathbf{y}|\mathbf{w}, \beta)p(\mathbf{w}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\beta).$$

The individual factor matrices $\mathbf{L}$ and $\mathbf{R}$ are not identifiable since $(\mathbf{L}\mathbf{Q}^\top)(\mathbf{R}\mathbf{Q}^{-1})^\top = \mathbf{L}\mathbf{R}^\top$ for any invertible matrix $\mathbf{Q} \in \mathbb{R}^{r\times r}$. The precisions $\{\gamma_i\}$ are also not identifiable since they can be interchanged without changing the model. We can therefore only estimate invariant quantities such as

$$\boldsymbol{\eta} = \mathbf{g}(\mathbf{z}) = \begin{bmatrix} \text{vec}(\mathbf{L}\mathbf{R}^\top) \\ s(\boldsymbol{\gamma}) \\ \beta \end{bmatrix}, \tag{6}$$

where $\mathbf{z} = [\mathbf{w}, \boldsymbol{\gamma}, \beta]$ and $s(\boldsymbol{\gamma})$ is a symmetric function of $\boldsymbol{\gamma}$.

1.3. The Bayesian Cramér-Rao bound

The Bayesian Cramér-Rao bound (BCRB), also known as the van-Trees inequality [12, 13] and the Borovkov-Sakhanenko inequality [13, 14], provides a lower bound on the variance of unbiased estimators. To derive the BCRB, we need to compute the Fisher information matrix $\mathbf{F}$ of $\mathbf{z}$, given by

$$\mathbf{F} = \mathcal{E}_{\mathbf{y}}\left[\frac{\partial \log p(\mathbf{y}, \mathbf{z})}{\partial \mathbf{z}} \frac{\partial \log p(\mathbf{y}, \mathbf{z})}{\partial \mathbf{z}^\top}\right].$$

We denote the covariance matrix of the estimation error $\boldsymbol{\epsilon} \triangleq \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}$ as

$$\mathbf{C}_{\boldsymbol{\epsilon}} \triangleq \mathcal{E}_{\mathbf{y},\mathbf{z}}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top\right] = \mathcal{E}_{\mathbf{y},\mathbf{z}}\left[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^\top\right].$$

Proposition 1. *For any unbiased estimator $\hat{\boldsymbol{\eta}}$, the covariance $\mathbf{C}_{\boldsymbol{\epsilon}}$ of estimation error $\boldsymbol{\epsilon}$ is bounded as*

$$\mathbf{C}_{\boldsymbol{\epsilon}} \succeq \mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right] \left(\mathcal{E}_{\mathbf{z}}[\mathbf{F}]\right)^{-1} \mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right]^\top. \tag{7}$$

It also holds that

$$\mathbf{C}_{\boldsymbol{\epsilon}} \succeq \mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top\right] \left(\mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\mathbf{F}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top\right]\right)^{-1} \mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top\right]. \tag{8}$$

■

The proof of Proposition 1 is given in [13].We obtain a lower bound on MSE by taking the trace of the inequalities. At this point, we mention that (7) can be non-informative, for example when $\mathcal{E}_{\mathbf{z}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\right] = \mathbf{0}$. Therefore we derived (8) as a relevant informative BCRB. Table 1 shows a nomenclature of various BCRB of associated variables.

2. BCRB FOR THE FACTORIZED MODEL

The following proposition gives the Fisher information matrix of the factorized model.

Proposition 2 (Fisher information matrix)**.** *For the factorized model, the Fisher information matrix is given by*

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\mathbf{w}\mathbf{w}} & \mathbf{F}_{\mathbf{w}\boldsymbol{\gamma}} & \mathbf{F}_{\mathbf{w}\beta} \\ \mathbf{F}_{\boldsymbol{\gamma}\mathbf{w}} & \mathbf{F}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} & \mathbf{F}_{\boldsymbol{\gamma}\beta} \\ \mathbf{F}_{\beta\mathbf{w}} & \mathbf{F}_{\beta\boldsymbol{\gamma}} & \mathbf{F}_{\beta\beta} \end{bmatrix}, \tag{9}$$

where $\mathbf{F}_{\mathbf{w}\mathbf{w}}$ is given in (10) *on the next page, $\mathbf{F}_{\beta\beta} = \frac{m}{2\beta^2} + \left(d - \frac{c-1}{\beta}\right)^2$ and $\mathbf{F}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} = \mathbf{h}\mathbf{h}^\top$ for $\mathbf{h} = [h_1, h_2, \ldots, h_r]^\top$ with*

$$h_i = \frac{p + q + 2(a-1)}{2\gamma_i} - \frac{||\mathbf{l}_i||^2 + ||\mathbf{r}_i||^2}{2} - b.$$

The remaining terms are zero when $\boldsymbol{\gamma}$ and β are deterministic (corresponding to $a = c = 1$ and $b = d = 0$) and zero mean when $\boldsymbol{\gamma}$ and β are random. They are therefore omitted. ■

$$\mathbf{F}_{\mathbf{ww}} = \begin{bmatrix} \beta(\mathbf{R}^\top \otimes \mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{R}\otimes\mathbf{I}_p) & \beta\mathbf{K}_{q,r}(\mathbf{I}_q\otimes\mathbf{L}^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{R}\otimes\mathbf{I}_p) \\ \beta(\mathbf{R}^\top \otimes \mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q\otimes\mathbf{L})\mathbf{K}_{r,q} & \beta\mathbf{K}_{q,r}(\mathbf{I}_q\otimes\mathbf{L}^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q\otimes\mathbf{L})\mathbf{K}_{r,q} \end{bmatrix} + \begin{bmatrix} \mathrm{vec}(\mathbf{L}\boldsymbol{\Gamma}) \\ \mathrm{vec}(\mathbf{R}\boldsymbol{\Gamma}) \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\mathbf{L}\boldsymbol{\Gamma}) \\ \mathrm{vec}(\mathbf{R}\boldsymbol{\Gamma}) \end{bmatrix}^\top, \quad (10)$$

The proof of Proposition 2 will be shown later in an extended manuscript. Next, we evaluate

$$\frac{\partial \mathbf{g}}{\partial \mathbf{z}} = \begin{bmatrix} [(\mathbf{R}\otimes\mathbf{I}_p),\ (\mathbf{I}_q\otimes\mathbf{L})\mathbf{K}_{r,q}] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \nabla_{\boldsymbol{\gamma}} s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \quad (11)$$

where we used (6) and the algebraic relation

$$\mathrm{vec}(\mathbf{L}\mathbf{R}^\top) = (\mathbf{R}\otimes\mathbf{I}_p)\mathrm{vec}(\mathbf{L}) = (\mathbf{I}_q\otimes\mathbf{L})\mathbf{K}_{r,q}\mathrm{vec}(\mathbf{R}).$$

Noting that $\mathcal{E}_{\mathbf{w}}\left[[(\mathbf{R}\otimes\mathbf{I}_p),\ (\mathbf{I}_q\otimes\mathbf{L})\mathbf{K}_{r,q}]\right] = \mathbf{0}$ as $\mathbf{L}$ and $\mathbf{R}$ are zero-mean, we find that the BCRB (7) is non-informative. We therefore compute the BCRB (8). We first evaluate

$$\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top = \begin{bmatrix} (\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p) + (\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & ||\nabla_{\boldsymbol{\gamma}} s||^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix},$$

where we used (11) and the standard relation $\mathbf{K}_{r,q}\mathbf{K}_{r,q}^\top = \mathbf{I}_{rq}$. We find that

$$\mathcal{E}_{\mathbf{w}}\left[\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top\right] = \begin{bmatrix} 2\sum_{i=1}^r \gamma_i^{-1}\mathbf{I}_{pq} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{E}_{\mathbf{z}}[||\nabla_{\boldsymbol{\gamma}} s||^2] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \quad (12)$$

where we used that $\mathcal{E}_{\mathbf{w}}[(\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p) + (\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top)] = 2\sum_{i=1}^r \gamma_i^{-1}\mathbf{I}_{pq}$. Using (9) and (11) we find that

$$\frac{\partial \mathbf{g}}{\partial \mathbf{z}}\mathbf{F}\frac{\partial \mathbf{g}}{\partial \mathbf{z}}^\top = \begin{bmatrix} \mathbf{G}_{\mathbf{ww}} & \mathbf{G}_{\mathbf{w}\boldsymbol{\gamma}} & \mathbf{G}_{\mathbf{w}\beta} \\ \mathbf{G}_{\boldsymbol{\gamma}\mathbf{w}} & \mathbf{G}_{\boldsymbol{\gamma\gamma}} & \mathbf{G}_{\boldsymbol{\gamma}\beta} \\ \mathbf{G}_{\beta\mathbf{w}} & \mathbf{G}_{\beta\boldsymbol{\gamma}} & G_{\beta\beta} \end{bmatrix}, \quad (13)$$

where

$$\begin{aligned} \mathbf{G}_{\mathbf{ww}} &= \beta(\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p) \\ &\quad +\beta(\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top) \\ &\quad +\beta(\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top) \\ &\quad +\beta(\mathbf{I}_q\otimes\mathbf{L}\mathbf{L}^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{R}\mathbf{R}^\top\otimes\mathbf{I}_p), \\ &\quad +(\mathbf{R}\boldsymbol{\Gamma}\mathbf{R}^\top\otimes\mathbf{I}_p) + (\mathbf{I}_q\otimes\mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top), \\ \mathbf{G}_{\boldsymbol{\gamma\gamma}} &= (\nabla_{\boldsymbol{\gamma}} s)\mathbf{F}_{\boldsymbol{\gamma\gamma}}(\nabla_{\boldsymbol{\gamma}} s)^\top = ((\nabla_{\boldsymbol{\gamma}} s)^\top\mathbf{h})^2, \\ G_{\beta\beta} &= F_{\beta\beta}. \end{aligned} \quad (14)$$

The remaining terms are zero when $\boldsymbol{\gamma}$ and β are deterministic ($a = c = 1$, $b = d = 0$) and zero mean when $\boldsymbol{\gamma}$ and β are random. They are therefore omitted.

2.1. BCRB-I and II

We compute the BCRB (8) by taking expectation values with respect to $\mathbf{w}$. We state the bound in the following proposition.

Proposition 3. *The BCRB-II of the factorized model is given by*

$$\mathbf{C}_{\boldsymbol{\epsilon}} \succeq \begin{bmatrix} \left(2\sum_{i=1}^r \gamma_i^{-1}\right)^2 (\mathcal{E}_{\mathbf{w}}[\mathbf{G}_{\mathbf{w}}])^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\boldsymbol{\gamma}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_\beta^{-1} \end{bmatrix}$$

where $\mathbf{G}_{\boldsymbol{\gamma}}$ *and* G_β *are given in* (14) *and*

$$\begin{aligned} \mathcal{E}_{\mathbf{w}}[\mathbf{G}_{\mathbf{w}}] &= 2\beta\left(\sum_{n=1}^r \gamma_n^{-1}\right)^2 \mathbf{A}^\top\mathbf{A} + 2r\mathbf{I}_{pq} \\ &+ \beta\left(\sum_{n=1}^r \gamma_n^{-2}\right)\left(\mathcal{T}_1(\mathbf{A}^\top\mathbf{A}) + \mathcal{T}_2(\mathbf{A}^\top\mathbf{A})\right) \\ &+ \beta\left(\sum_{n=1}^r \gamma_n^{-2}\right)\left(\mathbf{I}_q\otimes\left(\sum_{m=1}^q (\mathbf{e}_m^\top\otimes\mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{e}_m\otimes\mathbf{I}_p)\right)\right) \\ &+ \beta\left(\sum_{n=1}^r \gamma_n^{-2}\right)\left(\left(\sum_{m=1}^p (\mathbf{I}_q\otimes\mathbf{e}_m^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q\otimes\mathbf{e}_m)\right)\otimes\mathbf{I}_p\right). \end{aligned}$$

The linear operators $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *are defined in Section 1.1. Since* (12) *and* (13) *are block diagonal, the BCRB-I of* $\mathbf{x}$ *is*

$$\mathcal{E}_{\mathbf{y},\mathbf{w}}\left[(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^\top\right] \succeq \left(2\sum_{i=1}^r \gamma_i^{-1}\right)^2 (\mathcal{E}_{\mathbf{w}}[\mathbf{G}_{\mathbf{w}}])^{-1}.$$

■

The proof of the above proposition will be given later in an extended manuscript.

2.2. BCRB-III

Computation of BCRB-III requires computing the expectation with respect to $\mathbf{y}$ and $\mathbf{z} = [\mathbf{w}^\top,\ \boldsymbol{\gamma}^\top,\ \beta]^\top$. We state the bound in the following proposition.

Proposition 4. *Assume that* $a > 2$ *in* (4) *and* $c > 2$ *in* (5). *The BCRB-III of the factorized model is given by*

$$\mathbf{C}_{\boldsymbol{\epsilon}} \succeq \begin{bmatrix} \left(\frac{2rb}{a-1}\right)^2 (\mathcal{E}_{\mathbf{z}}[\mathbf{G}_{\mathbf{w}}])^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathcal{E}_{\mathbf{z}}[\mathbf{G}_{\boldsymbol{\gamma}}])^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathcal{E}_{\mathbf{z}}[G_\beta])^{-1} \end{bmatrix}$$

where now

$$\mathcal{E}_{\mathbf{z}}\left[\mathbf{G}_{\mathbf{w}}\right] = 2\frac{c}{d}\left(\frac{r^2b^2}{(a-1)^2} + \frac{rb^2}{(a-1)^2(a-2)}\right)\mathbf{A}^\top\mathbf{A} + 2r\mathbf{I}_{pq}$$
$$+ \frac{c}{d}\frac{rb^2}{(a-1)(a-2)}\left(\mathcal{T}_1(\mathbf{A}^\top\mathbf{A}) + \mathcal{T}_2(\mathbf{A}^\top\mathbf{A})\right)$$
$$+ \frac{c}{d}\frac{rb^2}{(a-1)(a-2)}\left(\mathbf{I}_q \otimes \left(\sum_{m=1}^{q}(\mathbf{e}_m^\top \otimes \mathbf{I}_p)\mathbf{A}^\top\mathbf{A}(\mathbf{e}_m \otimes \mathbf{I}_p)\right)\right)$$
$$+ \frac{c}{d}\frac{rb^2}{(a-1)(a-2)}\left(\left(\sum_{m=1}^{p}(\mathbf{I}_q \otimes \mathbf{e}_m^\top)\mathbf{A}^\top\mathbf{A}(\mathbf{I}_q \otimes \mathbf{e}_m)\right) \otimes \mathbf{I}_p\right),$$
$$\mathcal{E}_{\mathbf{z}}[\mathbf{G}_{\boldsymbol{\gamma}}] = \frac{p+q+2(a-1)}{2}\mathcal{E}_{\boldsymbol{\gamma}}\left[(\nabla_{\boldsymbol{\gamma}}s)\mathbf{\Gamma}^{-2}(\nabla_{\boldsymbol{\gamma}}s)^\top\right],$$
$$\mathcal{E}_{\mathbf{z}}[G_\beta] = \frac{(m+2(c-1))d^2}{2(c-1)(c-2)}.$$

■

3. NUMERICAL EVALUATION

Here we numerically evaluate the BCRB bounds and compare them to the performance of two low-rank matrix estimation methods - (1) variational Bayesian (VB) estimator of [8] and - (2) nuclear norm (NN) minimization based estimator (convex optimization based) [1]. The VB estimator uses the factorized model with (3), (4) and (5) while NN is a convex optimization based estimator that does not use the factorized model. We provide the performance of NN estimator due to its widespread use. For simulations we only considered matrix completion where the sensing matrix $\mathbf{A}$ has one component in each row set to one and all other components zero, with the constraint that $\mathbf{A}$ has full row-rank. We measure the performance in terms of normalized mean square error

$$\text{NMSE} = \frac{\mathcal{E}_{\mathbf{z}}[||\hat{\mathbf{x}} - \mathbf{x}||^2]}{\mathcal{E}_{\mathbf{z}}[||\mathbf{x}||^2]}$$

which we evaluate empirically for different values of the model parameters and averaged out for many realizations. In the simulations we fixed the signal-to-noise-ratio

$$\text{SNR} = \frac{\mathcal{E}_{\mathbf{z}}[||\mathcal{A}(\mathbf{L}\mathbf{R}^\top)||^2]}{\mathcal{E}_{\mathbf{z}}[||\mathbf{n}||^2]} = \frac{r(c-1)b^2}{(a-1)(a-2)d}.$$

In the simulations we first fixed the parameter values of $\{p, q, r, m, a, b, c, \text{SNR}\}$ and then randomly generated the measurement matrix $\mathbf{A}$ and the precisions $\boldsymbol{\gamma}$, the noise precision β. The factor matrices were drawn from (3) and the measurements from (1).

3.1. Numerical rank

The random matrix $\mathbf{X} = \mathbf{L}\mathbf{R}^\top$ has rank r with probability one. However, the *effective rank* is smaller when some precisions are large. Here we introduce the *numerical rank* that serves as a lower bound on the actual rank as

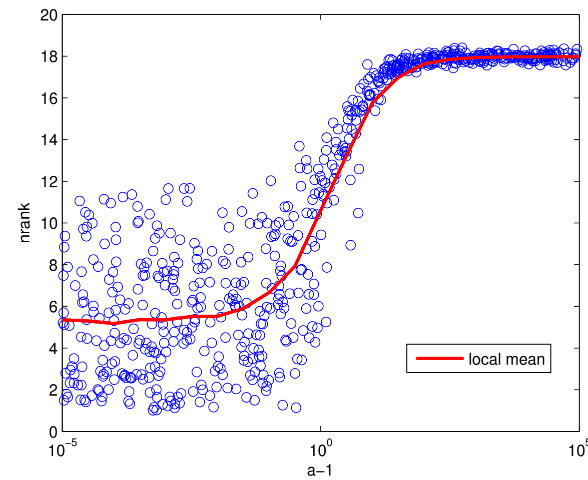

Fig. 1. Realizations and mean of the numerical rank for different values of a for $p = q = 100$, $r = 20$ and $b = 1$.

$$\text{nrank}(\mathbf{X}) \triangleq \frac{||\mathbf{X}||_*^2}{||\mathbf{X}||^2} \leq \text{rank}(\mathbf{X}),$$

where $||\mathbf{X}||_* = \sum_{i=1}^{k}\sigma_i(\mathbf{X})$ and $k = \min(p, q)$. Since the parameter b in (4) affects the magnitude of $\mathbf{X}$, only the parameter a affects the numerical rank. To investigate how the numerical rank varies with a, we empirically evaluate the mean numerical rank $\mathcal{E}_{\mathbf{w},\boldsymbol{\gamma}}[\text{nrank}(\mathbf{X})]$. The results for $p = q = 100$, $r = 20$ and $b = 1$ are given in Figure 1. We find that the mean numerical rank is about 18 for $a - 1 > 100$ and about 5.3 for $a - 1 < 0.03$. However, the variance is large for smaller values of $a - 1$ and BCRB-III does not exists for $a \leq 2$. Small values of $a - 1$ thus correspond to low numerical rank.

3.2. Performance of practical algorithms and BCRB

In the second experiment we evaluated BCRB-II and BCRB-III for $\mathbf{X}$ for matrix completion by setting $a = c = 2 + 10^{-3}$, $b = d = 1$, $p = q = 10$, $r = 5$ and SNR $= 20$ dB. We generated 25 measurement matrices, 25 values of $\boldsymbol{\gamma}$ and β and 10 matrices $\mathbf{X}$ and measurements $\mathbf{y}$ for each value of m. The results are given in Figure 2 where the NMSE is plotted against the sub-sampling factor $\frac{m}{pq}$. Note that NN estimator provides better performance than VB. The main reason is that NN estimator knows the noise parameter exclusively, whereas VB infers all necessary parameters including noise parameter using a variational Bayes learning technique. For the choice of parameters, it turns out that the BCRB-II is consistently lower than the BCRB-III. However, the VB algorithm has a large gap in performance from both bounds and hence the problem of designing Bayesian algorithms for matrix completion and deriving new bounds remains valuable. We also measured the numerical rank of the realizations of $\mathbf{X}$. The distribution of

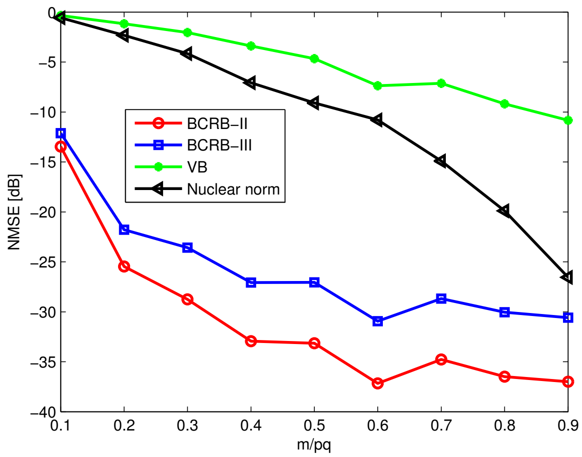

Fig. 2. NMSE vs. m for matrix completion with $p = q = 10$, $r = 5$, SNR = 20 dB, $b = d = 1$ and $a = c = 1 + 10^{-3}$.

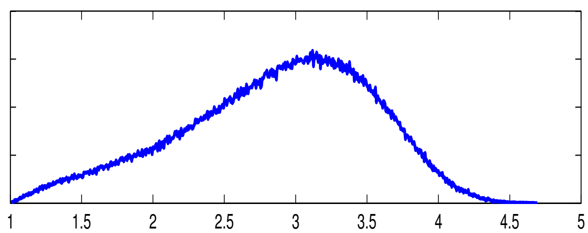

Fig. 3. Distribution of the numerical rank for $p = q = 10$, $r = 5$, $b = d = 1$ and $a = 1 + 10^{-3}$.

numerical rank is given in Figure 3. We see that the numerical rank is consistently lower than the actual rank $r = 5$.

4. CONCLUSION

In this article we computed Bayesian Cramér-Rao bounds for matrix reconstruction from linear measurements. For the factorized model, we found that while the standard BCRB (7) is non-informative, the BCRB (8) provides an informative lower bound. By evaluating the numerical rank we found that the model promotes low numerical rank when $a < 1$. However, the bound BCRB-III only exists for $a > 1$. We found that both the nuclear norm estimator and the variational Bayes estimator showed a considerable gap from the BCRB bounds. This shows that there still exists room for improvement of Bayesian low-rank matrix reconstruction algorithms and bounds.

REFERENCES

[1] Emmanuel J Candes and Yaniv Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[2] Maryam Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, PhD thesis, Stanford University, 2002.

[3] Dave Zachariah, Martin Sundin, Magnus Jansson, and Saikat Chatterjee, "Alternating least-squares for low-rank matrix reconstruction," *Signal Processing Letters, IEEE*, vol. 19, no. 4, pp. 231–234, 2012.

[4] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[5] Yehuda Koren, Robert Bell, and Chris Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, , no. 8, pp. 30–37, 2009.

[6] Yew Jin Lim and Yee Whye Teh, "Variational Bayesian approach to movie rating prediction," in *Proceedings of KDD Cup and Workshop*. Citeseer, 2007, vol. 7, pp. 15–21.

[7] Andriy Mnih and Ruslan Salakhutdinov, "Probabilistic matrix factorization," in *Advances in neural information processing systems*, 2007, pp. 1257–1264.

[8] S Derin Babacan, Martin Luessi, Rafael Molina, and Aggelos K Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 3964–3977, 2012.

[9] Gongguo Tang and Arye Nehorai, "Lower bounds on the mean-squared error of low-rank matrix reconstruction," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4559–4571, 2011.

[10] Emmanuel J Candès and Terence Tao, "The power of convex relaxation: Near-optimal matrix completion," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2053–2080, 2010.

[11] Karl Werner and Magnus Jansson, "Reduced rank linear regression and weighted low rank approximations," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2063–2075, 2006.

[12] Harry L Van Trees, *Detection, estimation, and modulation theory*, John Wiley & Sons, 2004.

[13] Richard D Gill and Boris Y Levit, "Applications of the van Trees inequality: a Bayesian Cramér-Rao bound," *Bernoulli*, pp. 59–79, 1995.

[14] AA Borovkov and AU Sakhanienko, "On estimates of the expected quadratic risk," *Probab. Math. Statist*, vol. 1, pp. 185–195, 1980.