Alternating Least-Squares for Low-Rank Matrix Reconstruction

Dave Zachariah, Martin Sundin, Magnus Jansson and Saikat Chatterjee

***Abstract*—For reconstruction of low-rank matrices from undersampled measurements, we develop an iterative algorithm based on least-squares estimation. While the algorithm can be used for any low-rank matrix, it is also capable of exploiting a-priori knowledge of matrix structure. In particular, we consider linearly structured matrices, such as Hankel and Toeplitz, as well as positive semidefinite matrices. The performance of the algorithm, referred to as alternating least-squares (ALS), is evaluated by simulations and compared to the Cramér-Rao bounds.**

***Index Terms*—Low-rank matrix reconstruction, Cramér-Rao bound, least squares, structured matrices**

I. Introduction

Low-rank matrices appear in various areas of signal processing and system identification [1]. In recent times the problem of reconstructing such matrices from a vector of linear measurements embedded in noise has attracted a lot of attention [2]. In this scenario the dimension of the measurement vector is lower than the number of matrix elements, and hence the problem consists of an underdetermined set of linear equations. This problem has a myriad of applications, for example, spectral imaging [3], wireless sensor networks [4], video error concealment [5].

In the gamut of designing low-rank matrix reconstruction algorithms, most of the existing research work deals with a specific setup known as 'matrix completion' where the measurement vector consists of a subset of elements of the underlying matrix. The algorithms can be separated into three broad categories: Convex-relaxation based [6], [7], minimization on Grassmannian manifold of subspaces [8], [9], least-squares matrix fitting [10], [11]. While we note that a substantial effort is paid to the matrix completion problem, far less effort is devoted to a general setup consisting of an underdetermined system of linear equations. There are few attempts, such as [7], [12], [13] and [14].

For the general underdetermined setup, we first consider the aspect of designing an algorithm that can deal with any low-rank matrix. In addition, we consider the aspect of exploiting a-priori knowledge of underlying matrix structure for further performance improvement. The motivation for using structured low-rank matrices is due to their frequent occurrence in various signal processing and system identification problems. We develop a new algorithm that addresses both aspects. Extending the approach of [11], the new algorithm is developed in an iterative framework based on least-squares (LS) estimation. For investigating matrix structure, we consider linearly structured matrices, such as Hankel and Toeplitz, as well as positive semidefinite matrices. The performance of the algorithm is evaluated by simulations and then compared to the Cramér-Rao bounds (CRBs). The CRBs are presented for measurements in Gaussian noise. Specifically we derived the CRBs for Hankel and positive semidefinite matrices.

The authors are with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm. E-mail: {dave.zachariah,magnus.jansson}@ee.kth.se, masundi@kth.se and saikatchatt@gmail.com

Notation: The invertible vectorization and matrix construction mappings are denoted $\text{vec}(\cdot) : \mathbb{C}^{n\times p} \to \mathbb{C}^{np\times 1}$ and $\text{mat}_{n,p}(\cdot) : \mathbb{C}^{np\times 1} \to \mathbb{C}^{n\times p}$, respectively. $\mathcal{X}_r \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times p} : \text{rank}(\mathbf{X}) = r\}$ and $\mathcal{X}_+ \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times n} : \mathbf{X} \succeq \mathbf{0}\}$ denote the sets of rank r matrices and positive semidefinite (p.s.d.) matrices, respectively. Similarly, $\mathcal{X}_S \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times p} : \text{vec}(\mathbf{X}) = \mathbf{S}\boldsymbol{\theta}, \boldsymbol{\theta} \in \mathbb{C}^q\}$ denotes a subspace of linearly structured matrices specified by $\mathbf{S} \in \mathbb{C}^{np\times q}$, which includes Hankel and Toeplitz matrices. $\mathbf{A}\otimes\mathbf{B}$ is the Kronecker product and $\langle \mathbf{A}, \mathbf{B}\rangle \triangleq \text{tr}\,(\mathbf{B}^*\mathbf{A})$ is the inner product. $\mathbf{A}^\top$, $\mathbf{A}^*$ and $\mathbf{A}^\dagger$ denote the transpose, Hermitian transpose and Moore-Penrose pseudoinverse of $\mathbf{A} \in \mathbb{C}^{n\times p}$, respectively while $\|\mathbf{A}\|_F$ is its Frobenius norm. $\|\mathbf{x}\|_{\mathbf{W}} \triangleq \sqrt{\mathbf{x}^*\mathbf{W}\mathbf{x}}$ is the weighted norm.

II. System model

A. General underdetermined setup

We consider a matrix $\mathbf{X} \in \mathcal{X}_r$, where $r \ll \min(n, p)$ is assumed to be known. It is observed by an undersampled linear measurement process,

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) + \mathbf{n} \in \mathbb{C}^{m\times 1}, \tag{1}$$

where $m < np$ and the linear sensing operator $\mathcal{A} : \mathbb{C}^{n\times p} \to \mathbb{C}^{m\times 1}$ can be written equivalently in forms,

$$\mathcal{A}(\mathbf{X}) = \begin{bmatrix} \langle \mathbf{X}, \mathbf{A}_1\rangle \\ \vdots \\ \langle \mathbf{X}, \mathbf{A}_m\rangle \end{bmatrix} = \mathbf{A}\text{vec}(\mathbf{X}). \tag{2}$$

The matrix $\mathbf{A}$ is assumed to be known and the measurement noise $\mathbf{n}$ is assumed to be zero-mean, with $\text{E}[\mathbf{n}\mathbf{n}^*] = \mathbf{C} \in \mathbb{C}^{m\times m}$ given. Note that in matrix completion, $\{\mathbf{A}_k\}$ is nothing but the set of element-selecting operators.

B. Applications in signal processing and system identification

Applications of the general underdetermined setup are illustrated in the following examples:

1) Recovery of data matrices $\mathbf{X} \in \mathcal{X}_r$, compressed by some randomly chosen linear operation $\mathcal{A}$ into $\mathbf{y}$.
2) Reconstruction of covariance matrices $\text{Cov}(\mathbf{z}) = \mathbf{R} \in \mathcal{X}_r \cap \mathcal{X}_+$ from a subset of second-order moments $r_{ij} =$

$\mathrm{E}[z_i z_j^*]$, estimated with zero-mean error ε_{ij}, $(i,j) \in \Omega$. Then (1) can be applied, $\mathbf{y} = \mathcal{A}(\mathbf{R}) + \boldsymbol{\varepsilon}$. In certain applications $\mathbf{R} \in \mathcal{X}_r \cap \mathcal{X}_+ \cap \mathcal{X}_S$, e.g. Toeplitz or Persymmetric structure.

3) Recovery of distance matrices $\mathbf{D} \in \mathcal{X}_r$, where $d_{ji} = d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2_{\mathbf{W}}$ and $\mathbf{W} \succ \mathbf{0}$. A subset of distance measurements are observed in noise.
4) Identification of a system matrix $\mathbf{D} \in \mathcal{X}_r$. The system $\mathbf{z}_t = \mathbf{D}\mathbf{u}_t \in \mathbb{C}^{n\times 1}$ is sampled by a varying linear operator $\mathbf{y}_t = \mathbf{A}_t\mathbf{z}_t + \mathbf{n}_t$, where $\mathbf{A}_t \in \mathbb{C}^{s\times p}$. A special case is $\mathbf{A}_t = \mathbf{e}^\top_{\ell(t)}$, where the index $\ell(t)$ varies the sampling of $\mathbf{z}_t$ at each t. Given a set of input and output data, $\{\mathbf{y}_t, \mathbf{u}_t\}_{t=1}^T$, where $sT < np$, the observations can be stacked by vectorization,

$$\mathbf{y} = \begin{bmatrix} \mathbf{u}_1^\top \otimes \mathbf{A}_1 \\ \vdots \\ \mathbf{u}_T^\top \otimes \mathbf{A}_T \end{bmatrix} \mathrm{vec}(\mathbf{D}) + \mathbf{n}.$$

In certain scenarios $\mathbf{D}$ may also have linear structure.

III. Alternating least-squares estimator

A. General low-rank matrix reconstruction

We begin by considering the case of reconstructing $\mathbf{X} \in \mathcal{X}_r$. Using a weighted least-squares criterion, the estimator is

$$\hat{\mathbf{X}} = \underset{\mathbf{X}\in\mathcal{X}_r}{\arg\min} \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|^2_{\mathbf{C}^{-1}}. \tag{3}$$

When the measurement noise $\mathbf{n}$ is Gaussian, this estimator coincides with the maximum likelihood estimator. For brevity we assume spatially uncorrelated noise, $\mathbf{C} = \sigma^2\mathbf{I}_m$, without loss of generality. Then minimizing the ℓ_2-norm is equivalent to (3). For general $\mathbf{C}$ the observation model is pre-whitened by forming $\bar{\mathbf{y}} = \mathbf{C}^{-1/2}\mathbf{y}$ and $\bar{\mathbf{A}} = \mathbf{C}^{-1/2}\mathbf{A}$.

Since $\mathbf{X} \in \mathcal{X}_r$, we express $\mathbf{X} = \mathbf{L}\mathbf{R}$ where $\mathbf{L} \in \mathbb{C}^{n\times r}$ and $\mathbf{R} \in \mathbb{C}^{r\times p}$. Then the square of the measurement residual can be written as

$$\begin{aligned} J(\mathbf{L},\mathbf{R}) &\triangleq \|\mathbf{y} - \mathcal{A}(\mathbf{L}\mathbf{R})\|_2^2 \\ &= \|\mathbf{y} - \mathbf{A}(\mathbf{I}_p \otimes \mathbf{L})\mathrm{vec}(\mathbf{R})\|_2^2 \\ &= \|\mathbf{y} - \mathbf{A}(\mathbf{R}^\top \otimes \mathbf{I}_n)\mathrm{vec}(\mathbf{L})\|_2^2. \end{aligned} \tag{4}$$

The cost $J(\mathbf{L},\mathbf{R})$ is minimized in an alternating fashion by the following steps:

- minimizing $\mathbf{R}$ with a fixed $\mathbf{L}$,
- minimizing $\mathbf{L}$ with a fixed $\mathbf{R}$.

In the new algorithm, the alternating minimization is performed through iterations. Starting with an initial $\mathbf{L}$, the iterations continue as long as the decreasing trend of $J(\mathbf{L},\mathbf{R})$ is observed. Given $\mathbf{L}$, the minimizer of $\mathbf{R}$ is computed by $\mathrm{vec}(\hat{\mathbf{R}}) = [\mathbf{A}(\mathbf{I}_p \otimes \mathbf{L})]^\dagger\mathbf{y}$ and similarly, given $\mathbf{R}$, the minimizer of $\mathbf{L}$ is computed by $\mathrm{vec}(\hat{\mathbf{L}}) = [\mathbf{A}(\mathbf{R}^\top \otimes \mathbf{I}_n)]^\dagger\mathbf{y}$. Since the Kronecker products are sparse, $\mathbf{A}(\mathbf{I}_p \otimes \mathbf{L}) \in \mathbb{C}^{m\times rp}$ and $\mathbf{A}(\mathbf{R}^\top \otimes \mathbf{I}_n) \in \mathbb{C}^{m\times nr}$ can be computed efficiently. Henceforth, we refer to the algorithm as alternating LS (ALS).

B. Structured low-rank matrix reconstruction

Next, we consider a structured low-rank matrix $\mathbf{X} \in \mathcal{X}_r$, and develop an ALS for a known matrix structure in Algorithm 1. In the algorithm, for each iteration, we approach the LS problem by first relaxing the structural constraint, and compute $\mathbf{R}$ with a fixed $\mathbf{L}$. Then, to impose the structural constraint on $\mathbf{R}$, the low-rank matrix estimate is projected onto the set of structured matrices by $\bar{\mathbf{X}} \triangleq \mathcal{P}(\mathbf{L}\mathbf{R})$, similar to 'lift and project' [15]. $\mathbf{R}$ is subsequently modified as the least-squares solution of $\bar{\mathbf{X}}$,

$$\min_{\mathbf{R}} \|\mathbf{L}\mathbf{R} - \bar{\mathbf{X}}\|_F^2.$$

$\mathbf{L}$ is updated in the same fashion. Here we mention that the algorithm is initialized by $\mathbf{L} := \mathbf{U}_0\boldsymbol{\Sigma}_0$ where $[\mathbf{U}_0, \boldsymbol{\Sigma}_0, \mathbf{V}_0]$ = $\mathtt{svdtrunc}(\mathrm{mat}_{n,p}(\mathbf{A}^*\mathbf{y}), r)$ and $\mathtt{svdtrunc}(\mathbf{Z}, r)$ denotes the singular value decomposition of $\mathbf{Z} \in \mathbb{C}^{n\times p}$ truncated to the rth singular value.

Algorithm 1 : ALS with known matrix structure

1: Input: $\mathbf{y}$, $\mathbf{A}$ and r
2: $[\mathbf{U}_0, \boldsymbol{\Sigma}_0, \mathbf{V}_0]$ = $\mathtt{svdtrunc}(\mathrm{mat}_{n,p}(\mathbf{A}^*\mathbf{y}), r)$
3: $\mathbf{L} := \mathbf{U}_0\boldsymbol{\Sigma}_0$
4: **while** $J(\mathbf{L},\mathbf{R})$ decreases **do**
5: $\quad \mathbf{R} := \mathrm{mat}_{r,p}\left([\mathbf{A}(\mathbf{I}_p \otimes \mathbf{L})]^\dagger\mathbf{y}\right)$
6: $\quad \bar{\mathbf{X}} := \mathcal{P}(\mathbf{L}\mathbf{R})$
7: $\quad \mathbf{R} := \mathbf{L}^\dagger\bar{\mathbf{X}}$
8: $\quad \mathbf{L} := \mathrm{mat}_{n,r}\left([\mathbf{A}(\mathbf{R}^\top \otimes \mathbf{I}_n)]^\dagger\mathbf{y}\right)$
9: $\quad \bar{\mathbf{X}} := \mathcal{P}(\mathbf{L}\mathbf{R})$
10: $\quad \mathbf{L} := \bar{\mathbf{X}}\mathbf{R}^\dagger$
11: **end while**
12: Output: $\hat{\mathbf{X}} = \mathbf{L}\mathbf{R}$

For linearly structured matrices the projection, $\bar{\mathbf{X}} = \mathcal{P}_{\mathcal{X}_S}(\mathbf{L}\mathbf{R})$, is computed by obtaining $\bar{\boldsymbol{\theta}} = \mathbf{S}^\dagger\mathrm{vec}(\mathbf{L}\mathbf{R})$ and setting $\bar{\mathbf{X}} = \mathrm{mat}_{n,p}(\mathbf{S}\bar{\boldsymbol{\theta}})$. For p.s.d. matrices, $\bar{\mathbf{X}} = \mathcal{P}_{\mathcal{X}_+}(\mathbf{L}\mathbf{R})$ is computed by first performing an eigenvalue decomposition of a symmetrized matrix, which projects the matrix onto the set of Hermitian matrices, $\frac{1}{2}(\mathbf{L}\mathbf{R} + (\mathbf{L}\mathbf{R})^*) = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^*$. The decomposition is performed to the rth largest eigenvalue, then setting $\bar{\mathbf{X}} = \mathbf{V}_r\widetilde{\boldsymbol{\Lambda}}_r\mathbf{V}_r^*$, where positive eigenvalues have been retained, projects the matrix onto $\mathcal{X}_+$.

IV. Cramér-Rao bounds

For relevant comparison, we use CRBs. In this section, we describe the CRB expressions.

A. Unstructured matrix

For real-valued $\mathbf{X}$ and Gaussian measurement noise $\mathbf{n}$, $\mathbf{y}$ is distributed as $\mathcal{N}(\mathbf{A}\mathrm{vec}(\mathbf{X}), \mathbf{C})$. The Cramér-Rao bound for unbiased low-rank matrix estimators without structure was derived in [11] (see also [1]): $\mathrm{E}_{y|X}[\|\mathbf{X} - \hat{\mathbf{X}}(\mathbf{y})\|_F^2] \geq \mathrm{CRB}(\mathbf{X})$, where

$$\mathrm{CRB}(\mathbf{X}) = \mathrm{tr}\left((\mathbf{P}^\top\mathbf{A}^\top\mathbf{C}^{-1}\mathbf{A}\mathbf{P})^{-1}\right). \tag{5}$$

The CRB holds when $\mathrm{rank}(\mathbf{A}\mathbf{P}) = r(n+p) - r^2$ where $\mathbf{P} \triangleq \begin{bmatrix}\mathbf{V}_1 \otimes \mathbf{U}_0 & \mathbf{U}_0 \otimes \mathbf{V}_0 & \mathbf{V}_0 \otimes \mathbf{U}_1\end{bmatrix}$. The submatrices are obtained from the left-singular vectors, $\mathbf{U}_0 = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$ and $\mathbf{U}_1 = [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n]$, and right-singular vectors, $\mathbf{V}_0 = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ and $\mathbf{V}_1 = [\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n]$, of $\mathbf{X} \in \mathcal{X}_r$.

B. Structured matrices

1) Hankel and Toeplitz matrices: For certain linearly structured matrices, such as Hankel or Toeplitz, low-rank parametrizations $\mathbf{S}\boldsymbol{\theta} = \mathbf{S}\mathbf{g}(\boldsymbol{\alpha})$ exist that enable the derivation of CRBs. E.g. when $\mathcal{X}_S$ is Hankel, $\mathbf{X}$ can be parameterized in the *canonical controllable form* [16] as $[\mathbf{X}]_{ij} = \mathbf{b}^\top \boldsymbol{\Phi}^{i+j-2}\mathbf{e}_1$, where $\mathbf{b} \in \mathbb{R}^r$,

$$\boldsymbol{\Phi} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_{r-1} & -a_r \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{r\times r},$$

and $\mathbf{e}_1$ is the first standard basis vector in $\mathbb{R}^r$. Here the parameters are $\boldsymbol{\alpha} = \begin{bmatrix}\mathbf{a}^\top & \mathbf{b}^\top\end{bmatrix}^\top$, where $\mathbf{a} = \begin{bmatrix}a_1 & \dots a_r\end{bmatrix}^\top$. A similar parametrization can be found for Toeplitz matrices.

The Cramér-Rao bound is then given by

$$\mathrm{CRB}_S(\mathbf{X}) = \mathrm{tr}\left(\boldsymbol{\Delta}_{\boldsymbol{\alpha}}^\top \mathbf{J}^{-1}(\boldsymbol{\alpha})\boldsymbol{\Delta}_{\boldsymbol{\alpha}}\right) \tag{6}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\alpha}} = \frac{\partial \mathbf{g}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^\top \mathbf{S}^\top$ and $\mathbf{J}(\boldsymbol{\alpha}) \triangleq -\mathrm{E}\left[\nabla^2_{\boldsymbol{\alpha}} p(\mathbf{y};\alpha)\right]$ is the Fisher information matrix of $\boldsymbol{\alpha}$ [17],

$$\mathbf{J}(\boldsymbol{\alpha}) = \boldsymbol{\Delta}_{\boldsymbol{\alpha}}\mathbf{A}^\top\mathbf{C}^{-1}\mathbf{A}\boldsymbol{\Delta}_{\boldsymbol{\alpha}}^\top.$$

2) Positive semidefinite matrix: Positive semidefinite matrices can be parameterized as $\mathbf{X} = \mathbf{M}\mathbf{M}^\top$, where $\mathbf{M} \in \mathbb{R}^{n\times r}$. Let $\boldsymbol{\alpha} = \mathrm{vec}(\mathbf{M})$, so that $\mathbf{X} = \mathbf{g}(\boldsymbol{\alpha})$ then $\mathbf{J}(\boldsymbol{\alpha})$ has the same form as above, with $\boldsymbol{\Delta}_{\boldsymbol{\alpha}} = \frac{\partial \mathbf{g}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^\top$. The gradient can be written compactly as $\frac{\partial \mathbf{g}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = [(\mathbf{I}_n \otimes \mathbf{M})\mathbf{T} + (\mathbf{M} \otimes \mathbf{I}_n)]$ where $\mathbf{M} = \mathrm{mat}_{n,r}(\boldsymbol{\alpha})$ and $\mathbf{T}$ is the matrix form of the transpose operator, $\mathbf{T}\mathrm{vec}(\mathbf{M}) = \mathrm{vec}(\mathbf{M}^\top)$. Note that parametrization by $\mathbf{M}$ is unique only up to an orthonormal transformation, hence $\mathbf{J}(\boldsymbol{\alpha})$ is not invertible in general. The Cramér-Rao bound is then given by

$$\mathrm{CRB}_+(\mathbf{X}) = \mathrm{tr}\left(\boldsymbol{\Delta}_{\boldsymbol{\alpha}}^\top \mathbf{J}^\dagger(\boldsymbol{\alpha})\boldsymbol{\Delta}_{\boldsymbol{\alpha}}\right) \tag{7}$$

provided that

$$\boldsymbol{\Delta}_{\boldsymbol{\alpha}}^\top = \boldsymbol{\Delta}_{\boldsymbol{\alpha}}^\top \mathbf{J}(\boldsymbol{\alpha})\mathbf{J}^\dagger(\boldsymbol{\alpha})$$

holds, or equivalently that $\mathbf{P}_{\mathbf{J}}^\perp \boldsymbol{\Delta}_{\boldsymbol{\alpha}} = \mathbf{0}$, where $\mathbf{P}_{\mathbf{J}}^\perp$ is the orthogonal projection onto the nullspace of $\mathbf{J}(\boldsymbol{\alpha})$ [1].

For sake of brevity the CRB derivations for Hankel and p.s.d. matrices are given in a supplementary note [18].

V. Experiments and Results

A. Experiment setup and performance measure

In the following we consider real-valued matrices of dimension $n = p \equiv 100$, with Hankel and p.s.d. structure respectively. For Hankel structure we generate $\mathbf{X}$ randomly by first creating a matrix with elements from an i.i.d. $\mathcal{N}(0,1)$ and fitting $\boldsymbol{\alpha}$ using Prony's method [19]. Then set $\mathbf{X} = \mathrm{mat}_{n,p}(\mathbf{S}\mathbf{g}(\boldsymbol{\alpha}))$. For p.s.d. structure, $\mathbf{X}$ is generated by $\mathbf{X} = \mathbf{M}\mathbf{M}^\top$, where the elements of $\mathbf{M}$ are generated by i.i.d. $\mathcal{N}(0,1)$. We let parameter $\lambda \triangleq r/\min(n,p) \in (0,1]$ determine the rank, which controls the degrees of freedom.

The linear sensing operator, modeled by $\mathbf{A}$ in (2), is selected randomly by $a_{ij} \sim \mathcal{N}(0, \frac{1}{m})$ [13]. The measurement noise is generated as $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$. In the experiments, two signal parameters are varied:

1) Signal to measurement noise ratio,

$$\mathrm{SMNR} \triangleq \frac{\mathrm{E}\left[\|\mathbf{X}\|_F^2\right]}{\mathrm{E}\left[\|\mathbf{n}\|_F^2\right]}. \tag{8}$$

2) Measurement fraction, $\rho \in (0,1]$, so that $m \equiv \lceil \rho np \rceil$.

The signal-to-reconstruction error ratio, or inverse of the normalized mean square error (NMSE),

$$\mathrm{SRER} \triangleq \frac{\mathrm{E}\left[\|\mathbf{X}\|_F^2\right]}{\mathrm{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2\right]} = \frac{1}{\mathrm{NMSE}}, \tag{9}$$

is used as the performance measure as it increases with SMNR. For an unbiased estimator, $\mathrm{SRER} \leq \mathrm{E}[\|\mathbf{X}\|_F^2]/\,\mathrm{E}[\mathrm{CRB}(\mathbf{X})]$. When SRER = 0 dB there is no reconstruction gain over the zero solution $\hat{\mathbf{X}} = \mathbf{0}$.

B. Results

The experiments were repeated for 500 Monte Carlo simulations. For each run a new realization of $\mathbf{X}$, $\mathbf{y}$ and $\mathbf{A}$ was generated and the Cramér-Rao bounds were computed correspondingly. The algorithm was set to terminate when the measurement residual $J(\mathbf{L},\mathbf{R})$ stops decreasing.

Figure 1 shows the performance of ALS when varying SMNR for Hankel and p.s.d. matrices, respectively. The measurement fraction was fixed to $\rho = 0.3$ and the relative rank was $\lambda = 0.03$. The algorithm is tested both with and without prior knowledge of matrix structure. Without such information it quickly approaches the bound for unstructured matrices (5). The reconstruction gain is significantly raised when exploiting the matrix structure. ALS remains at a SRER gap from the bounds (6) and (7) because it relaxes the problem by alternating projections. The gaps are 2.75 and 0.77 dB for Hankel and p.s.d., respectively, at SMNR = 10 dB. The estimator inserts a proportionally larger bias into the MSE compared to the estimator without prior information.

In Figure 2 the experiment is repeated for varying ρ, while $\lambda = 0.03$ and SMNR is fixed to 10 dB. As the measurement fraction increases the SRER of the algorithm without prior matrix structure approaches the corresponding CRB. At very low ρ, below 0.1, the bound (5) tends to break down as the rank constraint is violated. Given prior structural information the algorithm performs similar to the SMNR case.

Reproducible research: The MATLAB code to reproduce the figures in this paper are available at http://sites.google.com/site/saikatchatt/, along with the figures for varying λ that have been omitted here due to lack of space.

VI. Conclusions

We have developed an algorithm based on least-squares estimation for reconstruction of low-rank matrices in a general underdetermined setup. Furthermore it is capable of exploiting structures, in particular linearly structured matrices and positive semidefinite matrices, leading to better performance. Through simulations, we found that the algorithm provides good performance compared to the Cramér-Rao bound.

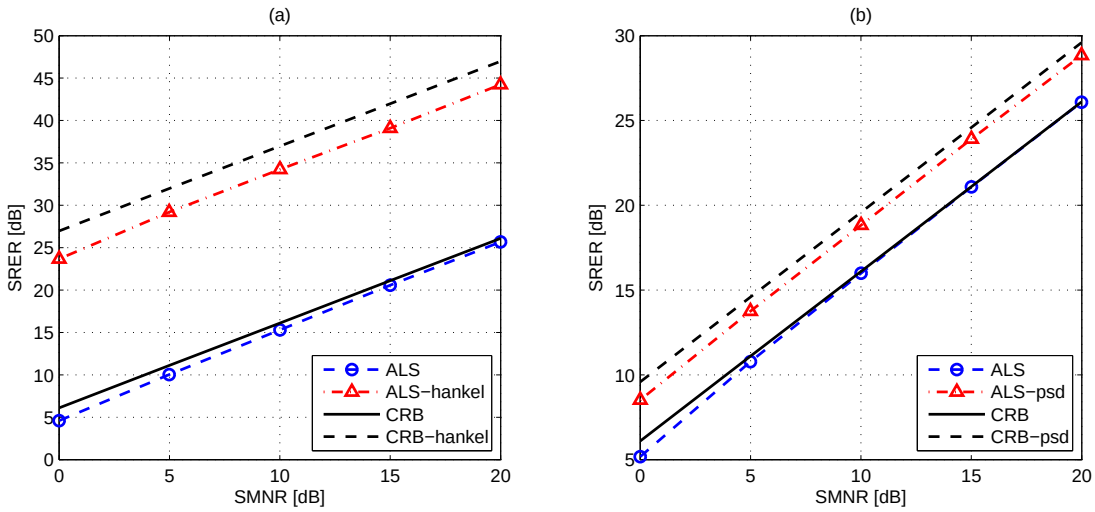

Fig. 1. SMNR versus SRER for $n = p = 100$, $\lambda = 0.03$ and $\rho = 0.3$. (a) Hankel structure, (b) Positive semidefinite structure.

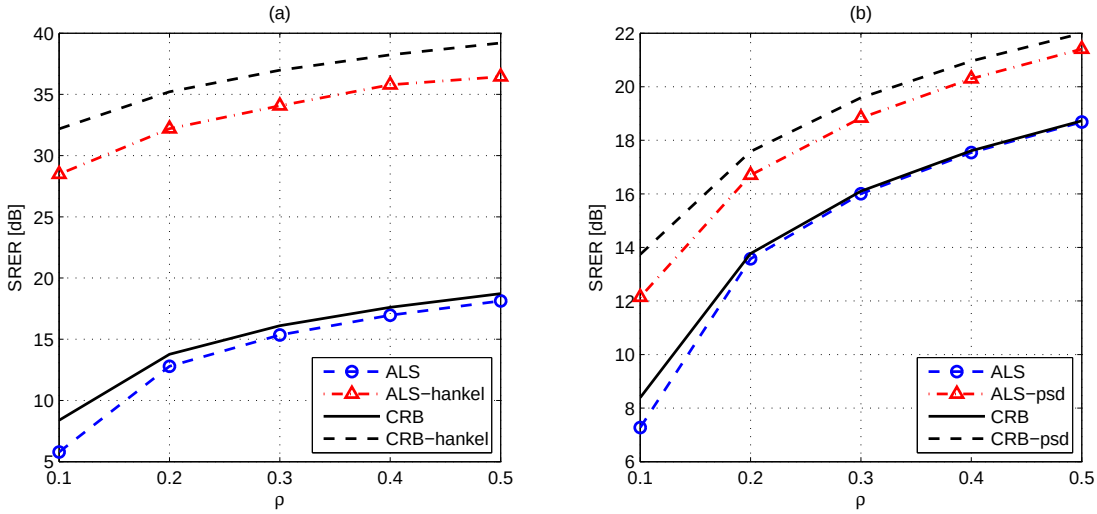

Fig. 2. ρ versus SRER for $n = p = 100$, $\lambda = 0.03$ and SMNR $= 10$ dB. (a) Hankel structure, (b) Positive semidefinite structure.

REFERENCES

[1] K. Werner and M. Jansson, "Reduced rank linear regression and weighted low rank approximations," *IEEE Trans. Signal Processing*, vol. 54, pp. 2063–2075, Jun. 2006.

[2] E. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, pp. 925–936, Jun. 2010.

[3] M. Signoretto, R. Van de Plas, B. De Moor, and J. Suykens, "Tensor versus matrix completion: A comparison with application to spectral data," *IEEE Signal Processing Letters*, vol. 18, pp. 403–406, Jul. 2011.

[4] J. Cheng, H. Jiang, X. Ma, L. Liu, L. Qian, C. Tian, and W. Liu, "Efficient data collection with sampling in wsns: Making use of matrix completion techniques," in *IEEE Global Telecommunications Conf. (GLOBECOM) 2010*, pp. 1–5, Dec. 2010.

[5] M. Dao, D. Nguyen, Y. Cao, and T. Tran, "Video concealment via matrix completion at high missing rates," in *Conf. on Signals, Systems and Computers (ASILOMAR)*, pp. 758–762, Nov. 2010.

[6] J.-F. Cai, E. Candes, and S. Shen, "A singular value thresholding algorithm for matrix completion," 2008. http://arxiv.org/abs/0810.3286.

[7] K. Toh and Y. Sangwoon, "An accelerated proximal gradient algorithm," *Pacific J. Optimization*, vol. 6, no. 3, pp. 615–640, 2010.

[8] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," 2010. http://arxiv.org/abs/1006.4046v2.

[9] W. Dai, O. Milenkovic, and E. Kerman, "Subspace evolution and transfer (set) for low-rank matrix completion," *IEEE Trans. Signal Processing*, vol. 59, pp. 3120 –3132, july 2011.

[10] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," tech. rep., Rice University, 2010. CAAM Technical Report TR10-07.

[11] G. Tang and A. Nehorai, "Lower bounds on the mean-squared error of low-rank matrix reconstruction," *IEEE Trans. Signal Processing*, vol. 59, pp. 4559–4571, Oct. 2011.

[12] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Trans. Information Theory*, vol. 56, pp. 4402–4416, Sep. 2010.

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

[14] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM J. on Optimization*, vol. 21, no. 4, pp. 1614–1640, 2011.

[15] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, "Structured low rank approximation," *Linear Algebra Appl.*, vol. 366, pp. 157–172, 2003.

[16] T. Kailath, *Linear Systems*. Prentice-Hall, 1980.

[17] S. Kay, *Fundamentals Of Statistical Signal Processing, Vol.1: Estimation Theory*. Prentice Hall, 1993.

[18] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Supplement: Alternating least-squares for low-rank matrix reconstruction," 2011. http://sites.google.com/site/saikatchatt/.

[19] M. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, 1996.