ITERATIVELY REWEIGHTED LEAST SQUARES FOR RECONSTRUCTION OF LOW-RANK MATRICES WITH LINEAR STRUCTURE

Dave Zachariah, Saikat Chatterjee, Magnus Jansson

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

ABSTRACT

This paper considers the problem of reconstructing low-rank matrices from undersampled measurements, when the matrix has a known linear structure. Based on the iterative reweighted least-squares approach, we develop an algorithm that exploits the linear structure in an efficient way that allows for reconstruction in highly undersampled scenarios. The method also enables inferring an appropriate regularization parameter value from the observations. The performance of the method is tested in a missing data recovery problem.

Index Terms— low-rank matrix reconstruction, missing data recovery, Cramér-Rao bound

1. INTRODUCTION

In recent times advances have been made in the problem of low-rank matrix reconstruction from a set of linear measurements in noise [1, 2]. Low-rank matrices appear in various areas of signal processing and system identification, and has several fields of applications, including magnetic resonance and spectral imaging [3], wireless sensor networks [4], etc. A variety of methods exist for solving the general underdetermined reconstruction problem, cf. [5, 6, 7, 2], several of which are based on convex relaxation using the nuclear norm [8]. Computationally efficient methods for approximating nuclear norm minimization were developed with performance guarantees in [9, 10] for the noiseless scenario, based on the iteratively reweighted least-squares approach (IRLS) [11, 12].

In this paper we consider reconstruction of low-rank matrices with linear structure. Such matrices arise through e.g. data from low-order linear systems, pairwise distance measurements, autocorrelation sequences of periodic signals, etc. An alternating least-squares method for solving the problem was given in [13] but assumed that the rank of the matrix is known. In this work we draw upon [9, 10] and formulate an IRLS method for low-rank matrices with linear structure and unknown rank. This enables reconstruction in highly underdetermined scenarios since the effective number of parameters is reduced by structure. Further, we employ the cross-validation approach for inferring an appropriate regularization parameter value from the observations. For illustration purposes the IRLS method for linearly structured matrices (IRLS-L) is applied to a missing data recovery problem, and compared with the Cramér-Rao bound and an existing missing data recovery algorithm.

Notation: The invertible vectorization and matrix construction mappings are denoted $\mathrm{vec}(\cdot) : \mathbb{C}^{n\times p} \to \mathbb{C}^{np\times 1}$ and $\mathrm{mat}_{n,p}(\cdot) : \mathbb{C}^{np\times 1} \to \mathbb{C}^{n\times p}$, respectively. $\mathbf{X}^*$ and $\mathbf{X}^{1/2}$ denote the Hermitian transpose and matrix square root of $\mathbf{X}$. Further, $\mathbf{X}^{*/2} = (\mathbf{X}^{1/2})^*$. The nuclear norm can be computed as $\|\mathbf{X}\|_* = \mathrm{tr}\{(\mathbf{X}\mathbf{X}^*)^{1/2}\} = \sum_i \sigma_i$, where σ_i denotes the ith singular value of $\mathbf{X}$. $\lceil\cdot\rceil$ denotes the ceiling function. $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$. $\langle \mathbf{A}, \mathbf{B}\rangle \triangleq \mathrm{tr}\,(\mathbf{B}^*\mathbf{A})$ is the inner product.

This work was partially supported by the Swedish Research Council under contract 621-2011-5847.

2. PROBLEM FORMULATION

A matrix $\mathbf{X} \in \mathbb{C}^{n\times p}$ is observed through a linear mapping $\mathcal{A} : \mathbb{C}^{n\times p} \to \mathbb{C}^{m\times 1}$ in zero-mean noise

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

where the mapping can be written equivalently in forms,

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

The matrix $\mathbf{A}$ is assumed to be known and the measurement noise $\mathbf{n}$ is assumed to be zero-mean, $\mathrm{E}[\mathbf{n}\mathbf{n}^*] = \sigma^2\mathbf{I}_m$ and σ^2 is unknown. In matrix completion, $\{\mathbf{A}_k\}$ is nothing but the set of element-selecting operators.

We consider matrices subject to linear constraints on the elements, $\mathbf{X} = \sum_{i=1}^d \mathbf{S}_i\theta_i$, or equivalently, $\mathbf{X} \in \mathcal{X}_S$ where $\mathcal{X}_S \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times p} : \mathrm{vec}(\mathbf{X}) = \mathbf{S}\boldsymbol{\theta}, \boldsymbol{\theta} \in \mathbb{C}^d\}$ is a d-dimensional linear subspace parameterized by $\mathbf{S} \in \mathbb{C}^{np\times d}$. This includes Hankel, Toeplitz, symmetric, triangular and diagonal matrices [14]. Of interest here is the set of rank r matrices, $\mathcal{X}_r \triangleq \{\mathbf{X} \in \mathbb{C}^{n\times p} : \mathrm{rank}(\mathbf{X}) = r\}$, where $r \ll \min(n, p)$.

The goal is to estimate $\mathbf{X} \in \mathcal{X}_S \cap \mathcal{X}_r$ from $\mathbf{y}$ when the rank r is unknown. Note that $m < d$ leads to undersampling of the effective parameters.

3. ITERATIVELY REWEIGHTED LEAST SQUARES

One approach to estimate $\mathbf{X}$ with unknown r is to find a matrix $\widehat{\mathbf{X}}$ with minimum rank, subject to some constraint on the residual $\|\mathbf{y} - \mathcal{A}(\widehat{\mathbf{X}})\|_2^2$. As rank minimization is intractable in general, methods have been developed that are based on the convex relaxation of the problem using the nuclear norm [8]. In adopting this approach, reconstruction of $\mathbf{X}$ can be formulated as a regularized least-squares problem

$$\widehat{\mathbf{X}} = \underset{\mathbf{X}\in\mathcal{X}_S}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathrm{vec}(\mathbf{X})\|_2^2 + \lambda\|\mathbf{X}\|_*, \tag{3}$$

where λ is a regularization parameter which controls the cost of the nuclear norm $\|\mathbf{X}\|_*$ [2].

Suppose that $\mathbf{X}$ has full row rank such that $\mathbf{X}\mathbf{X}^*$ is invertible. Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be the singular value decomposition. Then $\mathrm{tr}\{(\mathbf{X}\mathbf{X}^*)^{1/2}\} = \mathrm{tr}\{(\mathbf{X}\mathbf{X}^*)^{-1/2}\mathbf{X}\mathbf{X}^*\} = \mathrm{tr}\{\mathbf{X}^*\mathbf{W}\mathbf{X}\}$ defining $\mathbf{W} \triangleq (\mathbf{X}\mathbf{X}^*)^{-1/2} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^*$ [10], so that

$$\begin{aligned}\|\mathbf{X}\|_* &= \mathrm{tr}\left\{(\mathbf{X}\mathbf{X}^*)^{1/2}\right\} \\ &= \mathrm{tr}\left\{\mathbf{X}^*\mathbf{W}^{*/2}\mathbf{W}^{1/2}\mathbf{X}\right\} \\ &= \|\mathbf{W}^{1/2}\mathbf{X}\|_F^2.\end{aligned}$$

Further, exploiting $\mathbf{X} \in \mathcal{X}_S$ yields

$$\begin{aligned}\|\mathbf{W}^{1/2}\mathbf{X}\|_F^2 &= \|\mathrm{vec}(\mathbf{W}^{1/2}\mathbf{X})\|_2^2 \\ &= \|(\mathbf{I}_p \otimes \mathbf{W}^{1/2})\mathrm{vec}(\mathbf{X})\|_2^2 \\ &= \|(\mathbf{I}_p \otimes \mathbf{W}^{1/2})\mathbf{S}\boldsymbol{\theta}\|_2^2 \\ &= \|\mathbf{G}\boldsymbol{\theta}\|_2^2,\end{aligned}$$

where $\mathbf{G}(\mathbf{W}) = (\mathbf{I}_p \otimes \mathbf{W}^{1/2})\mathbf{S}$. Thus the least-squares problem (3) can be recast as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\mathbb{C}^d}{\arg\min}\, V(\boldsymbol{\theta}), \tag{4}$$

where

$$V(\boldsymbol{\theta}) = \|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|_2^2 + \lambda\|\mathbf{G}\boldsymbol{\theta}\|_2^2$$

and $\mathbf{H} \triangleq \mathbf{A}\mathbf{S}$. When holding $\mathbf{W}$ fixed, the solution to (4) equals

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^*\mathbf{H} + \lambda\mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W})\mathbf{S}\right)^{-1}\mathbf{H}^*\mathbf{y},$$

where we used the fact that

$$\begin{aligned}\mathbf{G}^*\mathbf{G} &= \mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W}^{1/2})^*(\mathbf{I}_p \otimes \mathbf{W}^{1/2})\mathbf{S} \\ &= \mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W})\mathbf{S}.\end{aligned}$$

Thus starting from an initial weight matrix $\mathbf{W}$, (3) is amenable to an iteratively reweighed least-squares formulation, updating $\hat{\boldsymbol{\theta}}$ and $\mathbf{W}$ in an alternating fashion. By exploiting the linear structure, IRLS needs only to estimate d rather than np parameters which can be a considerable reduction.

Note, however, that $\mathbf{W}$ is predicated on $\mathbf{X}(\boldsymbol{\theta})\mathbf{X}(\boldsymbol{\theta})^*$ being invertible, but as the goal is low rank the computation of $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^*$ becomes ill-conditioned. Following [10], stability can be ensured by truncating the smallest singular values to some ε. The weight matrix is then defined as $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{U}^*$, where $\mathbf{\Sigma}_\varepsilon = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\overline{r}-1}, \bar{\sigma}_{\overline{r}}, \ldots, \bar{\sigma}_K)$ and $K = \min(n, p)$. For all $k \geq \overline{r}$, the singular values are truncated $\bar{\sigma}_k \equiv \varepsilon$, where $\overline{r} = \lceil K/\kappa \rceil$ and κ is user-defined. The threshold value is adaptively lowered by $\varepsilon := \min(\varepsilon, \sigma_{\overline{r}}/(\kappa\overline{r}))$. For large matrices, the weight matrix can alternatively be computed using $\overline{r}$-truncated singular value decompositions [9, 10].

The resulting iteratively reweighted least-squares estimator (IRLS-L) is given in Algorithm 1, starting with some initial weight matrix $\mathbf{W}_0$ and threshold ε_0. It is set to terminate when the difference between iterates is smaller than some threshold, $\|\hat{\boldsymbol{\theta}}^\ell - \hat{\boldsymbol{\theta}}^{\ell-1}\|_2^2 \leq \epsilon_\theta$.

Note that when $\mathcal{X}_S$ is the subspace of diagonal matrices, IRLS-L includes recovery of sparse vectors as a special case.

Algorithm 1 IRLS-L

Input: $\mathbf{y}$, $\mathbf{A}$, $\mathbf{S}$, $\mathbf{W}_0$, ε_0, κ and λ
Set $\mathbf{H} = \mathbf{A}\mathbf{S}$ and $\overline{r} = \lceil \min(n, p)/\kappa \rceil$
$\mathbf{W} := \mathbf{W}_0$ and $\varepsilon := \varepsilon_0$
repeat
  $\hat{\boldsymbol{\theta}} := \left(\mathbf{H}^*\mathbf{H} + \lambda\mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W})\mathbf{S}\right)^{-1}\mathbf{H}^*\mathbf{y}$
  $[\mathbf{U}, \mathbf{\Sigma}] = \mathrm{svd}\Big(\mathrm{mat}_{n,p}(\mathbf{S}\hat{\boldsymbol{\theta}})\Big)$
  Set $\varepsilon := \min\{\varepsilon, \sigma_{\overline{r}}/(\kappa\overline{r})\}$ and form $\mathbf{\Sigma}_\varepsilon$
  $\mathbf{W} := \mathbf{U}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{U}^*$
until convergence
Output: $\widehat{\mathbf{X}} = \mathrm{mat}_{n,p}(\mathbf{S}\hat{\boldsymbol{\theta}})$

4. SELECTING REGULARIZATION PARAMETER

Lacking any prior knowledge, an appropriate value of the regularization parameter λ must be inferred from the data $\mathbf{y}$ alone. A cross-validation method aims to predict the ith component y_i using the remaining observations, denoted $\mathbf{y}^{(i)} \in \mathbb{C}^{m-1}$, and opts for the λ that minimizes the resulting prediction errors [15]. Let $\hat{\boldsymbol{\theta}}^{(i)}(\lambda)$ denote the estimate of $\boldsymbol{\theta}$ using $\mathbf{y}^{(i)}$ for a fixed λ and $\mathbf{h}_i^*$ be the ith row of $\mathbf{H}$, then the weighted sum of squared prediction errors is

$$C(\lambda) = \frac{1}{m}\sum_{i=1}^{m} g_i(\lambda)\left|y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}^{(i)}(\lambda)\right|^2. \tag{5}$$

For notational convenience, let $\hat{\boldsymbol{\theta}}(\lambda) = \mathbf{B}^{-1}(\lambda)\mathbf{H}^*\mathbf{y}$ denote the full IRLS-L estimate, where $\mathbf{B}(\lambda) = (\mathbf{H}^*\mathbf{H} + \lambda\mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W}_\lambda)\mathbf{S})$ and $\mathbf{W}_\lambda$ is the converged weight matrix. Then the Hermitian matrix $\mathbf{\Gamma}(\lambda) \triangleq \mathbf{I}_m - \mathbf{H}\mathbf{B}^{-1}(\lambda)\mathbf{H}^*$ produces the residuals $\mathbf{e} = \mathbf{y} - \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{\Gamma}(\lambda)\mathbf{y}$. Selecting the weights as $g_i(\lambda) \propto [\mathbf{\Gamma}(\lambda)]_{ii}^2 \geq 0$ emphasizes the components y_i with large residual variances and leads to the generalized cross-validation method [16].

The computation of the cost function $C(\lambda)$ can further be simplified. For a fixed $\mathbf{W}$ and defining $\tilde{\mathbf{y}}^{(i)} = \mathbf{H}^*\mathbf{y} - \mathbf{h}_i y_i$,

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{(i)} &= \left(\mathbf{H}_i^*\mathbf{H}_i + \lambda\mathbf{S}^*(\mathbf{I}_p \otimes \mathbf{W})\mathbf{S}\right)^{-1}\mathbf{H}_i^*\mathbf{y}^{(i)} \\ &= \left(\mathbf{B} - \mathbf{h}_i\mathbf{h}_i^*\right)^{-1}\left(\mathbf{H}^*\mathbf{y} - \mathbf{h}_i y_i\right) \\ &= \mathbf{B}^{-1}\left(\mathbf{H}^*\mathbf{y} - \mathbf{h}_i y_i\right) + \frac{\mathbf{B}^{-1}\mathbf{h}_i\mathbf{h}_i^*\mathbf{B}^{-1}\left(\mathbf{H}^*\mathbf{y} - \mathbf{h}_i y_i\right)}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i} \\ &= \hat{\boldsymbol{\theta}} - \frac{\mathbf{B}^{-1}\mathbf{h}_i y_i(1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i) - \mathbf{B}^{-1}\mathbf{h}_i\mathbf{h}_i^*\mathbf{B}^{-1}\tilde{\mathbf{y}}^{(i)}}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i} \\ &= \hat{\boldsymbol{\theta}} - \frac{\mathbf{B}^{-1}\mathbf{h}_i y_i - \mathbf{B}^{-1}\mathbf{h}_i\mathbf{h}_i^*\hat{\boldsymbol{\theta}}}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i} \\ &= \hat{\boldsymbol{\theta}} - \frac{\mathbf{B}^{-1}\mathbf{h}_i}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i}(y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}),\end{aligned}$$

where $\mathbf{H}_i \in \mathbb{C}^{(m-1)\times d}$ is the observation matrix $\mathbf{H}$ after removing the ith row. Then

$$\begin{aligned}y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}^{(i)} &= y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}} + \frac{\mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i}(y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}) \\ &= \frac{1}{1 - \mathbf{h}_i^*\mathbf{B}^{-1}\mathbf{h}_i}(y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}) \\ &= \frac{1}{[\mathbf{\Gamma}(\lambda)]_{ii}}(y_i - \mathbf{h}_i^*\hat{\boldsymbol{\theta}}).\end{aligned}$$

When the weights are normalized as $g_i = [\mathbf{\Gamma}(\lambda)]_{ii}^2/(\frac{1}{m}\mathrm{tr}\{\mathbf{\Gamma}(\lambda)\})^2$ [16], (5) can be written as

$$C(\lambda) = \frac{\|\mathbf{\Gamma}(\lambda)\mathbf{y}\|^2}{\eta(\lambda)}, \tag{6}$$

where $\eta = \mathrm{tr}\{\mathbf{\Gamma}(\lambda)\}^2/m$. The regularization parameter λ is chosen to minimize $C(\lambda)$.

5. EXPERIMENTAL RESULTS

For illustration we apply IRLS-L to a missing data recovery problem, where $d \ll np$. Consider a set of d samples from a sum of K sinusoids

$$\theta(t) = \sum_{i=1}^{K} \alpha_i \sin(\omega_i t + \phi_i), \quad t = 0, \ldots, d-1,$$

where the model structure is unknown. The only assumption is that $\theta(t)$ can be modeled as the output of a low-order linear system. A subset of samples are observed in noise

$$y(t) = \theta(t) + n(t), \quad t \in \mathcal{T} \subset \{0, \ldots, d-1\}, \tag{7}$$

where $|\mathcal{T}| = m$ and $n(t) \sim \mathcal{N}(0, \sigma^2)$. The goal is to estimate all samples $\boldsymbol{\theta} = [\theta(0) \cdots \theta(d-1)]^\top \in \mathbb{R}^d$ from the noisy subset $\mathbf{y} \in \mathbb{R}^m$. By arranging $\boldsymbol{\theta}$ in a Hankel matrix $\mathbf{X}(\boldsymbol{\theta}) \in \mathcal{X}_S$, with unknown rank $r = 2K$ [17], we can write (7) as $\mathbf{y} = \mathbf{A}\mathrm{vec}(\mathbf{X}(\boldsymbol{\theta})) + \mathbf{n}$, where, and $\mathbf{A}$ is a sampling matrix selecting the observed elements. Thus (7) equivalent to the linear observation model (1) and the estimation problem can be posed as (3).

The performance of IRLS-L is evaluated with respect to the normalized mean square error, NMSE $\triangleq \mathrm{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2]/\,\mathrm{E}[\|\boldsymbol{\theta}\|^2]$. When the model order r is known, a low-rank parameterization $\boldsymbol{\alpha} \in \mathbb{R}^{2r}$ exists such that $\boldsymbol{\theta} = \mathbf{g}(\boldsymbol{\alpha})$, cf. [13]. The Cramér-Rao bound (CRB) on the mean square error of unbiased estimators $\hat{\boldsymbol{\theta}}(\mathbf{y}, r)$ is given by

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

where $\mathbf{\Delta}_\alpha = \partial_\alpha \mathbf{g}(\boldsymbol{\alpha})^\top$ and $\mathbf{J}(\boldsymbol{\alpha})$ is the Fisher information matrix. As r is unknown, CRB is an oracle bound in this problem.

Further, we compare with the state of the art iterative adaptive approach for missing temporal data recovery (MIAA-T) [18]. In this approach the data is modeled as $\theta(t) = \sum_{k=1}^{K_g} \alpha(\omega_k) e^{j\omega_k t}$ over a grid of K_g frequencies $\{\omega_k\}_{k=1}^{K_g}$. First, the complex amplitudes $\alpha(\omega_k)$ and covariance matrix $\mathbf{R}$ of the observed data are estimated in an alternating manner based on the model. This results in the estimates $\{\hat{\alpha}_{\text{iaa}}(\omega_k)\}$ and $\widehat{\mathbf{R}}_{\text{iaa}}$. Then, the MIAA-T estimator of the $d-m$ missing samples is

$$\hat{\boldsymbol{\theta}}_0 = \sum_{k=1}^{K_g} |\hat{\alpha}_{\text{iaa}}(\omega_k)|^2 \mathbf{a}_0(\omega_k)\mathbf{a}_1^*(\omega_k)\widehat{\mathbf{R}}_{\text{iaa}}^{-1}\mathbf{y} \in \mathbb{C}^{d-m},$$

where $\mathbf{a}_0(\omega_k)$ and $\mathbf{a}_1(\omega_k)$ denote Fourier vectors corresponding to the missing and observed samples, respectively [18]. For the remaining samples, we use the MSE optimal unbiased estimate, $\hat{\boldsymbol{\theta}}_1 = \mathbf{y} \in \mathbb{C}^m$.

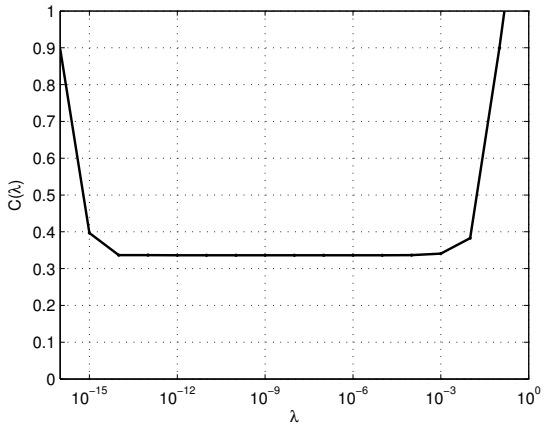

Fig. 1. Realization of weighted sum of prediction errors $C(\lambda)$ for SNR $= 25$ dB and $\rho = 0.30$.

5.1. Setup

We consider $K = 3$ sinusoids with $d = 100$ samples. The samples are nominally arranged in a 50×51 Hankel matrix $\mathbf{X}$. The amplitudes are equal $\alpha_i \equiv \alpha = 1$ and the phases are drawn independently as $\phi_i \sim \mathcal{U}(0, 2\pi)$. The frequencies are $\omega_1 = 0.80\pi$, $\omega_2 = 0.10\pi$ and $\omega_3 = 0.06\pi$. Two ratios are varied: sampling factor $\rho \in (0, 1]$, such that $m = \lceil \rho d \rceil$, and the signal to noise ratio SNR $\triangleq \alpha^2/\sigma^2$.

We set $\kappa = 5$, which effectively assumes an upper bound on the rank to $\overline{r} = 50/5 = 10$. The regularization parameter λ is set to minimize $C(\lambda)$ in (6). Fig. 1 shows a typical realization of $C(\lambda)$ when sweeping over $\lambda \in [10^{-16}, 10^0]$ in steps of decades. Over a large span of ρ and SNR-levels, it is found that $C(\lambda)$ reaches its minimum and is virtually constant for λ in the interval $[10^{-14}, 10^{-3}]$. As higher parameter value penalizes rank, we opt for a value that promotes the 'sparsest' solution, i.e., $\lambda = 10^{-3}$. The termination criterion was set to $\epsilon_\theta = 5 \times 10^{-4}$. We initialize the algorithm by setting the missing samples to $\mathbf{0}$ and the remaining samples to $\mathbf{y}$. This defines $\mathbf{W}_0$ and we select ε_0 as the largest singular value.

For MIAA-T, we follow [18] and set $K_g = 10^3$ over a uniform grid and terminate after 15 iterations.

5.2. Results

Fig. 2 shows a realization of $\theta(t)$ and a reconstruction from a given realization $\mathbf{y}$, when SNR=25 dB and 70% of the samples are randomly discarded. As can be seen IRLS-L is capable of recovering the missing samples assuming only that they can be modeled as the output from an unknown low-order linear system.

When computing the NMSE we perform 10^3 Monte Carlo runs. Fig. 3 shows the NMSE versus the sampling factor ρ at SNR=25 dB. The estimation errors of MIAA-T and IRLS-L rapidly decline as ρ rises to about 0.5. Initially, MIAA-T declines faster but saturates around $\rho = 0.3$ and is unable to improve the estimate as more samples are observed. This is consistent with the results presented in [18]. By contrast, IRLS-L keeps reducing the NMSE when more than half of the samples are observed by exploiting the underlying low-rank structure but remains at a certain gap from the oracle CRB. In this scenario, the gain of the low-rank method over MIAA-T approaches 4 dB as ρ rises. Fig. 4 shows that an advantage remains as

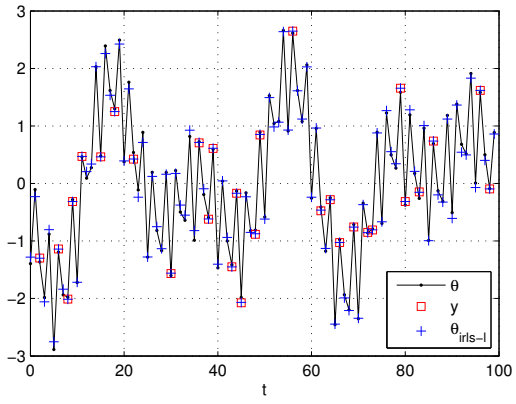

Fig. 2. Example of realization of $\theta(t)$ and observation $y(t)$ for $\rho = 0.30$ and SNR = 25 dB.

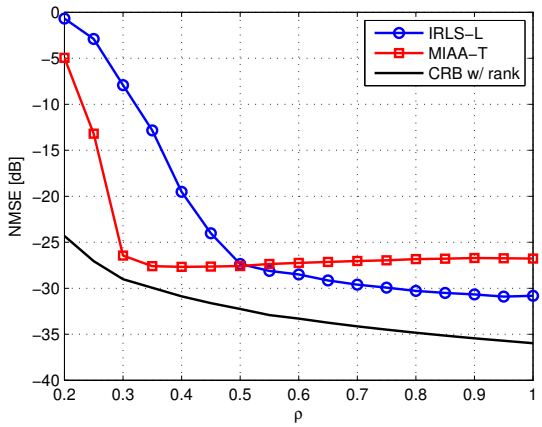

Fig. 3. NMSE versus sampling factor ρ for SNR = 25 dB.

the signal to noise ratio is varied over a wide range of values. When the algorithms perform approximately equal, at $\rho = 0.5$, the average computation times for the current implementations of IRLS-L and MIAA-T are 0.893 and 4.053 seconds, respectively.

6. CONCLUSIONS AND RELATION TO PRIOR WORK

This paper addressed the problem of reconstructing low-rank matrices with linear structure from undersampled measurements in noise. The proposed method draws upon the nuclear norm relaxation framework [8, 5, 7]. Unlike [13], which also deals with structured low-rank matrices, it does not assume the rank to be known. The method, denoted IRLS-L, extends the iterative reweighted least-squares methods developed in [9, 10] to exploit the linear structure, which can potentially reduce the parameter space significantly. This enables reconstruction in highly underdetermined scenarios. It further enables the formulation of a computationally efficient cross-validation function for inferring an appropriate value of the regularization parameter from the data alone.

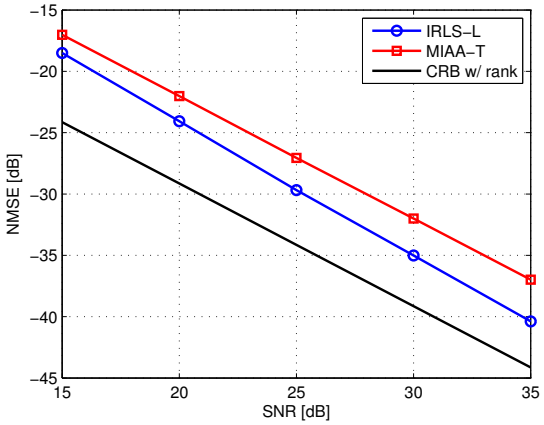

Fig. 4. NMSE versus SNR for $\rho = 0.7$.

Finally, IRLS-L was applied to a missing data recovery problem and compared with the Cramér-Rao bound and a state-of-the-art method [18].

7. REFERENCES

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

[2] G. Tang and A. Nehorai, "Lower bounds on the mean-squared error of low-rank matrix reconstruction," *IEEE Trans. Signal Processing*, vol. 59, no. 10, pp. 4559–4571, Oct. 2011.

[3] M. Signoretto, R. Van de Plas, B. De Moor, and J.A.K. Suykens, "Tensor versus matrix completion: A comparison with application to spectral data," *IEEE Signal Process Lett.*, vol. 18, no. 7, pp. 403–406, July 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 Telecom. Conf. (GLOBECOM) 2010*, Dec. 2010, pp. 1–5.

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

[6] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4402 –4416, Sept. 2010.

[7] D. Goldfarb and S. Ma, "Convergence of fixed-point continuation algorithms for matrix rank minimization," *Found. Comput. Math.*, vol. 11, no. 2, pp. 183–210, Apr. 2011.

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

[9] K. Mohan and M. Fazel, "Iterative reweighted least squares for matrix rank minimization," in *Annual Allerton Conf. Com., Control, and Comp.*, Oct. 2010, pp. 653–661.

[10] 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.

[11] Å. Björck, *Numerical Methods for Least Square Problems*, SIAM, 1996.

[12] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[13] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Alternating least-squares for low-rank matrix reconstruction," *IEEE Signal Process Lett.*, vol. 19, no. 4, pp. 231–234, Apr. 2012.

[14] J. R. Magnus, "L-structured matrices and linear matrix equations," *Linear and Multilinear Algebra*, vol. 14, no. 1, pp. 67–88, 1983.

[15] D.M. Allen, "The relationship between variable selection and data augmentation and a method for prediction," *Technometrics*, vol. 16, no. 1, pp. 125–127, 1974.

[16] G.H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

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

[18] P. Stoica, J. Li, and J. Ling, "Missing data recovery via a nonparametric iterative adaptive approach," *IEEE Signal Process Lett.*, vol. 16, no. 4, pp. 241–244, Apr. 2009.