OPTIMAL REDUCED RANK MODELING FOR GENERAL NOISE USING NULLSPACE ESTIMATION

Mats Viberg
Department of Mathematics and Natural Sciences
Blekinge Institute of Technology
Karlskrona, Sweden
mats.viberg@bth.se

Tomas McKelvey
Department of Electrical Engineering
Chalmers University of Technology
Göteborg, Sweden
tomas.mckelvey@chalmers.se

***Abstract*—The problem of optimal reconstruction of a low-rank matrix subject to additive noise of arbitrary noise color is addressed. We propose a non-iterative method based on modeling the nullspace of the data. The proposed technique is shown to yield statistically efficient estimates at sufficiently high Signal-to-Noise Ratio. Yet, the computational complexity is significantly reduced compared to existing methods. The empirical efficiency is verified using simulated data. In more difficult scenarios, the proposed NullSpace Estimator (NSE) can be used to initialize an iterative approach, and in the studied cases just one iteration of Alternating Least-Squares (ALS) was found enough.**

I. Introduction

To approximate a given matrix with one of lower rank is a problem that has been studied for almost a century. The classical Eckart-Young theorem solves the problem in the Frobenius norm sense using the Singular Value Decomposition (SVD). Reduced-rank approximation and the related problem reduced-rank linear regression [1], [2] is of great importance in a diverse set of applications in different disciplines, such as Chemometrics, Econometrics and Microbiology. Engineering applications include 2D filter design [3], image compression and de-noising [4], and localization from pairwise distance measurements [5]. In many of these applications, it is relevant to apply a weighting to the different matrix elements. The weighted low-rank approximation problem has been shown to be NP-complete [6]. Yet, several practical algorithms have been proposed, such as [7]–[9].

The present contribution considers a more general case, where the measured matrix is subject to an additive noise of an arbitrary color. This problem was studied in [10] and [11], and both of these contributions involve the idea to search for the nullspace of the low-rank matrix rather than the matrix itself (image space). We leverage on this idea by proposing a non-iterative solution to the low-rank matrix approximation problem for a general noise color. The proposed NullSpace Estimator (NSE) requires only the solution of a weighted linear least-squares problem. We derive the optimal weighting matrix and show that the so-obtained method attains the Cramér-Rao lower Bound (CRB) for the problem at hand, and thus has the same performance as the previously known iterative methods. A more precise relation to [10], [11] is given in Section V.

II. Reduced Rank Modeling

In many situations of practical interest, the information-bearing part of a data matrix can be assumed to be of low rank. Assuming additive noise, the observed data matrix $\mathbf{Y} \in \mathcal{R}^{M\times N}$, where $M \geq N$, is modeled as

$$\mathbf{Y} = \mathbf{X} + \mathbf{E}\,, \tag{1}$$

where $\mathbf{X}$ has rank r, $r < N$. Defining $\mathbf{e} = \mathrm{vec}(\mathbf{E})$, we assume

$$\mathrm{E}[\mathbf{e}] = 0\ , \quad \mathrm{E}[\mathbf{e}\mathbf{e}^T] = \mathbf{C}\,,$$

where $\mathbf{C}$ is known. For most of the analysis, $\mathbf{e}$ can have an arbitrary distribution, but for the Maximum Likelihood (ML) estimator and the Cramér-Rao lower Bound (CRB), we assume that $\mathbf{e} \in \mathcal{N}(\mathbf{0},\mathbf{C})$. For later reference, we introduce the Cholesky factorization of the positive definite matrix $\mathbf{C}$ and its inverse as

$$\mathbf{C} = \mathbf{C}^{1/2}\mathbf{C}^{T/2}\ , \quad \mathbf{C}^{-1} = \mathbf{C}^{-T/2}\mathbf{C}^{-1/2}\,. \tag{2}$$

The goal is to recover the low-rank matrix $\mathbf{X}$, given data $\mathbf{Y}$ and knowledge of $\mathbf{C}$. The rank r can be estimated from data in case it is not known [1], but we assume here that r is known. If the noise is white, the optimal solution (in a Maximum-Likelihood sense) to the matrix approximation problem is obtained by truncating the SVD, according to the classical Eckart-Young theorem

$$\begin{gathered}\min_{\mathbf{X}} \|\mathbf{Y}-\mathbf{X}\|_F^2 \quad \text{s. t. rank}(\mathbf{X}) = r \\ \Rightarrow \\ \hat{\mathbf{X}} = \sum_{k=1}^{r} \mathbf{u}_k s_k \mathbf{v}_k^T.\end{gathered} \tag{3}$$

Here, $\mathbf{Y} = \sum_{k=1}^{N} \mathbf{u}_k s_k \mathbf{v}_k^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$ is the SVD of $\mathbf{Y}$, and the singular values s_k are ordered in a non-increasing fashion. The case of main interest here is where $\mathbf{C}$ has no particular structure to exploit. For the statistical analysis, we will assume that the noise is ”small”, i.e. $\mathbf{C} = \sigma^2\mathbf{C}_0$, where $\mathbf{C}_0$ is fixed and σ^2 is sufficiently small.

III. Non-Linear Least-Squares Estimation

Since $\mathbf{X}$ is assumed to be deterministic, the ML principle provides the optimal estimate in the sense that the estimation error achieves the CRB in the small error regime (e.g. [12]). Let $\mathbf{y} = \mathrm{vec}(\mathbf{Y}) = \mathbf{x} + \mathbf{e}$, where $\mathbf{e} \in \mathcal{N}(0,\mathbf{C})$, so that $\mathbf{y} \in$

$\mathcal{N}(\mathbf{x}, \mathbf{C})$. As is well-known, the ML estimate for this case is obtained by solving the Non-Linear Least-Squares (NLLS) problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|^2_{\mathbf{C}^{-1}} \quad \text{subject to } \operatorname{rank}(\mathbf{X}) = r\,.$$

In general, this is difficult to solve due to the constraint $\operatorname{rank}(\mathbf{X}) = r$. But for $\mathbf{C} = \sigma^2\mathbf{I}$, we have

$$\|\mathbf{y} - \mathbf{x}\|^2_{\mathbf{C}^{-1}} = \frac{1}{\sigma^2}\|\mathbf{Y} - \mathbf{X}\|^2_F,$$

which verifies that the ML estimator for white noise is given by the truncated SVD.

A. Parameterization

The set of matrices satisfying the constraint $\operatorname{rank}(\mathbf{X}) = r$ is not convex. We can avoid the problem by realizing that if $\mathbf{X}$ has rank r, it can be factorized as

$$\mathbf{X} = \mathbf{L}\mathbf{R}\,, \tag{4}$$

where $\mathbf{L} \in \mathcal{R}^{M\times r}$ and $\mathbf{R} \in \mathcal{R}^{r\times N}$. Clearly, this factorization is not unique, but provided the first r columns of $\mathbf{R}$ are linearly independent, we can obtain a unique factorization by choosing

$$\mathbf{R} = [\mathbf{I}_r, \mathbf{R}_2] \quad \text{and} \quad \mathbf{L} \text{ arbitrary}\,, \tag{5}$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix and $\mathbf{R}_2 \in \mathcal{R}^{r\times(N-r)}$. The total number of unique parameters in $\mathbf{L}$ and $\mathbf{R}_2$ is thus $(M + N - r)r$.

Remark 1 While the parameterization (5) is not always possible, it can be "enforced" by a pre-conditioning of $\mathbf{Y}$ (and hence $\mathbf{X}$) using the orthogonal matrix $\mathbf{V}$ ($N \times N$) of right singular vectors of $\mathbf{Y}$. Since the columns of $\mathbf{YV}$ are orthogonal, we expect the first r columns of $\mathbf{XV}$ to be linearly independent so that the parameterization (5) exists. Thus, $\mathbf{Y}$ can be replaced by $\mathbf{Y}_V = \mathbf{YV}$ when applying the estimator, and the so-obtained $\mathbf{R}_V$ is replaced by $\mathbf{R} = \mathbf{R}_V\mathbf{V}^T$ to "undo" the transformation. Note that the noise color needs to be modified due to the transformation. The pre-conditioning is generally applicable and it guarantees that (5) exists for small enough σ^2. □

B. Separable NLLS Formulation

Given a unique parameterization, we write $\mathbf{x}(\boldsymbol{\theta}) = \operatorname{vec}(\mathbf{X}(\boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} \operatorname{vec}(\mathbf{L}) \\ \operatorname{vec}(\mathbf{R}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\ell} \\ \mathbf{r}_2 \end{bmatrix}. \tag{6}$$

Using the well-known formula

$$\operatorname{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\operatorname{vec}(\mathbf{B})\,,$$

where $\otimes$ is the Kronecker product, we have

$$\mathbf{x}(\boldsymbol{\theta}) = \operatorname{vec}(\mathbf{X}(\boldsymbol{\theta})) = \operatorname{vec}(\mathbf{LR}) = (\mathbf{R}^T \otimes \mathbf{I})\operatorname{vec}(\mathbf{L}) = (\mathbf{R}^T \otimes \mathbf{I})\boldsymbol{\ell}$$

and therefore the ML/NLLS estimate is the minimizing argument of the criterion function

$$V_{\text{ML}}(\boldsymbol{\theta}) = (\mathbf{y} - (\mathbf{R}^T \otimes \mathbf{I})\boldsymbol{\ell})^T\mathbf{C}^{-1}(\mathbf{y} - (\mathbf{R}^T \otimes \mathbf{I})\boldsymbol{\ell})\,. \tag{7}$$

For fixed $\mathbf{R}$, $V_{ML}(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\ell}$, implying that this is a *separable* NLLS problem, see e.g. [12]. The weighted LS solution w.r.t. $\boldsymbol{\ell}$ is given by

$$\hat{\boldsymbol{\ell}}(\mathbf{R}) = \left\{(\mathbf{R} \otimes \mathbf{I})\mathbf{C}^{-1}(\mathbf{R}^T \otimes \mathbf{I})\right\}^{-1}(\mathbf{R} \otimes \mathbf{I})\mathbf{C}^{-1}\mathbf{y}\,. \tag{8}$$

Similarly, one could fix $\mathbf{L}$ and solve for $\mathbf{R}$. The procedure that solves for one parameter vector at the time in an iterative fashion is termed Alternating Least Squares (ALS), and it does not really require a unique parameterization as in (5). An alternative is to substitute (8) back into (7), which results in a minimization over $\mathbf{r}_2$ only:

$$V_{\text{sep}}(\mathbf{r}_2) = \left\|\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{R}^T\otimes\mathbf{I})}\mathbf{C}^{-1/2}\mathbf{y}\right\|^2. \tag{9}$$

Here, we have used the Cholesky factorization of $\mathbf{C}^{-1}$ introduced in (2) and $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is an orthogonal projection onto the span of $\mathbf{A}$, whereas $\mathbf{P}^{\perp}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$ projects onto its orthogonal complement. Minimizing (9) requires a non-linear search over the parameter $\mathbf{r}_2$, which can be done using e.g. a Newton-type method [10], [11]. This can be successful if a sufficiently good initial estimate is available.

C. The Cramér-Rao lower Bound

The Cramér-Rao lower Bound (CRB) is the inverse of the Fisher Information Matrix (FIM). For our case it is given by (see e.g. [12])

$$\mathbf{FIM}_{\boldsymbol{\theta}} = \frac{\partial \mathbf{x}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\mathbf{C}^{-1}\frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\,, \tag{10}$$

and the CRB inequality is

$$\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \geq \mathbf{FIM}^{-1}_{\boldsymbol{\theta}}, \tag{11}$$

where $\hat{\boldsymbol{\theta}}$ is any unbiased estimate of $\boldsymbol{\theta}$. With the parameterization (5), the Jacobian is obtained as

$$\frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \begin{bmatrix}\mathbf{R}^T \otimes \mathbf{I} & \tilde{\mathbf{I}}^T \otimes \mathbf{L}\end{bmatrix} \triangleq \mathbf{B}\,, \tag{12}$$

where

$$\tilde{\mathbf{I}} = \begin{bmatrix}\mathbf{0}_{(N-r)\times r} & \mathbf{I}_{(N-r)}\end{bmatrix}.$$

It is well-known that the ML/NLLS estimator asymptotically achieves (11) with equality as $\sigma^2 \to 0$.

IV. The Nullspace Estimator

The idea of the proposed NullSpace Estimator (NSE) is to parameterize the nullspace of $\mathbf{X}$ rather than the matrix itself. Since $\mathbf{X}$ has rank r, there exists a full-rank matrix $\mathbf{N} \in \mathcal{R}^{N\times(N-r)}$ such that $\mathbf{XN} = \mathbf{0}$. Clearly, $\mathbf{N}$ is not unique, but provided its $N - r$ last rows are linearly independent we can uniquely parameterize $\mathbf{N}$ as

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 \\ \mathbf{I}_{N-r} \end{bmatrix}, \tag{13}$$

where $\mathbf{N}_1 \in \mathcal{R}^{r\times(N-r)}$. The benefit of this parameterization is that it only requires one unknown matrix $\mathbf{N}$, in contrast with modeling $\mathbf{X}$ itself as in (4), which is a bi-linear model.

Note that (13) is related to (5), since $\mathbf{XN} = \mathbf{0}$ implies $\mathbf{RN} = \mathbf{0}$, i.e.

$$[\mathbf{I}_r, \mathbf{R}_2] \left[\begin{array}{c} \mathbf{N}_1 \\ \mathbf{I}_{N-r} \end{array} \right] = \mathbf{0} \;\Rightarrow\; \mathbf{N}_1 = -\mathbf{R}_2 \,. \quad (14)$$

From Remark 1, we note that $\mathbf{Y}$ can be pre-conditioned to make (13) possible if needed.

With the parameterization (13), the NSE is formulated as

$$\hat{\mathbf{N}}_1 = \arg\min_{\mathbf{N}_1} \left\| \mathrm{vec}\left([\mathbf{Y}_1 \; \mathbf{Y}_2] \left[\begin{array}{c} \mathbf{N}_1 \\ \mathbf{I}_{N-r} \end{array} \right] \right) \right\|_{\mathbf{W}}^2 \quad (15)$$

where $\mathbf{Y}_1 \in \mathcal{R}^{M\times r}$, $\mathbf{Y}_2 \in \mathcal{R}^{M\times(N-r)}$ and $\mathbf{W} \in \mathcal{R}^{M(N-r)\times M(N-r)}$ is a positive definite weighting matrix as yet to be defined. In the unweighted case, $\mathbf{W} = \mathbf{I}$, we get the standard pseudo-inverse solution

$$\hat{\mathbf{N}}_1 = -(\mathbf{Y}_1^T\mathbf{Y}_1)^{-1}\mathbf{Y}_1^T\mathbf{Y}_2 = -\mathbf{Y}_1^{+}\mathbf{Y}_2 \,. \quad (16)$$

This can be used as an initial estimate, since, as we shall see, the optimal weighting depends on the unknown $\mathbf{N}_1$. Note that (16) is a consistent estimator in the sense that gives a correct basis for the nullspace in the noiseless case.

For a general weighting, we vectorize the matrices as $\mathbf{y}_k = \mathrm{vec}(\mathbf{Y}_k),\ k = 1, 2$ and $\boldsymbol{\eta} = \mathrm{vec}(\mathbf{N}_1)$. Applying the Kronecker product formula again, the criterion (15) is expressed as

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \|(\mathbf{I}_{N-r} \otimes \mathbf{Y}_1)\boldsymbol{\eta} + \mathbf{y}_2\|_{\mathbf{W}}^2 \,. \quad (17)$$

Defining the Cholesky factorization of the weighting matrix as $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{T/2}$, the solution is obtained as

$$\hat{\boldsymbol{\eta}} = -\left\{\mathbf{W}^{T/2}(\mathbf{I}_{N-r} \otimes \mathbf{Y}_1)\right\}^{+} \mathbf{W}^{T/2}\mathbf{y}_2 \,. \quad (18)$$

Re-arranging the vector $\hat{\boldsymbol{\eta}}$ into the matrix $\hat{\mathbf{N}}_1$, we now have an estimate of the nullspace matrix $\mathbf{N}$ from (13). In addition, we can estimate $\mathbf{R}$ using (5) with $\hat{\mathbf{R}}_2 = -\hat{\mathbf{N}}_1$. Inserting this into (8) gives $\hat{\mathbf{L}}$, which finally yields the reconstructed low-rank matrix $\hat{\mathbf{X}} = \hat{\mathbf{L}}\hat{\mathbf{R}}$.

A. Performance Analysis and Choice of Weighting Matrix

First, note that as $\sigma^2 \to 0$, we have

$$\hat{\boldsymbol{\eta}} \to -\left\{\mathbf{W}^{T/2}(\mathbf{I}_{N-r} \otimes \mathbf{X}_1)\right\}^{+} \mathbf{W}^{T/2}\mathbf{X}_2 = \boldsymbol{\eta}_0 \,, \quad (19)$$

where $\boldsymbol{\eta}_0$ denotes the true value. Thus, for small σ^2, we can perform a first-order analysis of the estimation error, $\Delta\boldsymbol{\eta} = \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0$. Let $\mathbf{Y} = \mathbf{X} + \mathbf{E}$ and partition $\mathbf{X} = [\mathbf{X}_1 \; \mathbf{X}_2]$ and $\mathbf{E} = [\mathbf{E}_1 \; \mathbf{E}_2]$ in the same way as $\mathbf{Y}$. Note that $\mathbf{X}_1 = \mathbf{L}\mathbf{R}_1 = \mathbf{L}$, since $\mathbf{R}_1 = \mathbf{I}$. The estimate using noisy data is then expressed as

$$\hat{\boldsymbol{\eta}} = -\left\{(\mathbf{I} \otimes [\mathbf{L} + \mathbf{E}_1])^T \mathbf{W} (\mathbf{I} \otimes [\mathbf{L} + \mathbf{E}_1])\right\}^{-1} \times (\mathbf{I} \otimes [\mathbf{L} + \mathbf{E}_1])^T \mathbf{W}(\mathbf{x}_2 + \mathbf{e}_2) \,. \quad (20)$$

Using that $\mathbf{x}_2 = -(\mathbf{I} \otimes \mathbf{X}_1)\boldsymbol{\eta} = -(\mathbf{I} \otimes \mathbf{L})\boldsymbol{\eta}$ and keeping only first-order terms we arrive at

$$\Delta\boldsymbol{\eta} \simeq -\left\{(\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W}(\mathbf{I} \otimes \mathbf{L})\right\}^{-1} (\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W} \times \left\{(\mathbf{I} \otimes \mathbf{E}_1)\boldsymbol{\eta} + \mathbf{e}_2\right\} \,, \quad (21)$$

where the notation $\simeq$ means that terms that tend to zero faster than $\Delta\boldsymbol{\eta}$ in probability as $\sigma^2 \to 0$ have been neglected. Rearranging the rightmost Kronecker product, the above becomes

$$\Delta\boldsymbol{\eta} \simeq -\mathbf{H}^{-1}(\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W} \left[\mathbf{N}_1^T \otimes \mathbf{I}_M \quad \mathbf{I}_{M(N-r)}\right] \mathbf{e} \,, \quad (22)$$

where we introduced the matrix

$$\mathbf{H} = (\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W}(\mathbf{I} \otimes \mathbf{L}) \,. \quad (23)$$

Next, note that $\left[\mathbf{N}_1^T \otimes \mathbf{I}_M \quad \mathbf{I}_{M(N-r)}\right] = \mathbf{N}^T \otimes \mathbf{I}_M$, so

$$\Delta\boldsymbol{\eta} \simeq -\mathbf{H}^{-1}(\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W}(\mathbf{N}^T \otimes \mathbf{I}_M)\,\mathbf{e} \,. \quad (24)$$

Hence, we see that if $\|\mathbf{e}\|^2$ is "small", the estimation error $\Delta\boldsymbol{\eta}$ has zero mean and covariance matrix

$$\mathrm{E}[\Delta\boldsymbol{\eta}\Delta\boldsymbol{\eta}^T] = \mathbf{H}^{-1}(\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W}(\mathbf{N}^T \otimes \mathbf{I}_M) \times \mathbf{C}(\mathbf{N} \otimes \mathbf{I}_M)\mathbf{W}(\mathbf{I} \otimes \mathbf{L})\mathbf{H}^{-1} \,. \quad (25)$$

Using standard arguments in weighted Least-Squares we can now show that

$$\mathrm{E}[\Delta\boldsymbol{\eta}\Delta\boldsymbol{\eta}^T] \geq \mathbf{H}^{-1} \,, \quad (26)$$

with equality if the weighting matrix is chosen in an optimal way as

$$\mathbf{W}_{\text{opt}} = \left\{(\mathbf{N}^T \otimes \mathbf{I}_M)\mathbf{C}(\mathbf{N} \otimes \mathbf{I}_M)\right\}^{-1} \,. \quad (27)$$

The optimal weighting matrix depends on the unknown matrix $\mathbf{N}_1$, but replacing $\mathbf{W}$ by any consistent estimate $\hat{\mathbf{W}}$, for example using (16), will not alter the first-order approximation.

Hence, we have proved the following key result:

Theorem 1: Let the estimate $\hat{\boldsymbol{\eta}}$ be obtained from (17), and assume that $\sigma^2 \ll 1$. Let the weighting matrix be chosen as (27). Then, to first order in σ^2, the estimation error $\Delta\boldsymbol{\eta}$ is zero mean with covariance matrix

$$\mathrm{E}[\Delta\boldsymbol{\eta}\Delta\boldsymbol{\eta}^T] = \left\{(\mathbf{I} \otimes \mathbf{L}^T)\mathbf{W}_{\text{opt}}(\mathbf{I} \otimes \mathbf{L})\right\}^{-1} \,. \quad (28)$$

B. Asymptotic Optimality

The proposed NSE is non-iterative and requires only the solution of linear equations. Next, we show that this does not imply a performance degradation asymptotically. Notice first that the parameter vector $\mathbf{r}_2$ in (6) is identical to $-\boldsymbol{\eta}$. Therefore, it is possible to directly compare the covariance matrix of $\hat{\boldsymbol{\eta}}$ with the CRB for $\mathbf{r}_2$.

Theorem 2: The asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$ as given by (28), is identical to the CRB for the corresponding parameter $\mathbf{r}_2$, obtained from (10)–(12). Thus, the optimally weighted null-space estimator is statistically efficient for high SNR.

Proof: See the appendix.

C. Computational Cost

Being non-iterative, the proposed NSE is a computationally attractive alternative to the ALS or Newton-type techniques. The computational cost for reconstructing $\mathbf{X}$ is dominated by solving (8). This can be done by computing the QR-factorization of the matrix $(\mathbf{R} \otimes \mathbf{I}_M)\mathbf{C}^{-T/2}$, which requires

$\mathcal{O}(M^3Nr^2)$ flops, not taking the structure of the matrices into account.

For the white noise case, the proposed NSE is in fact a competitive alternative to the SVD. It is easy to verify that the optimally weighted NSE reduces to the unweighted solution (16) when $\mathbf{C} = \sigma^2\mathbf{I}$. Thus, a basis for the nullspace of $\mathbf{X}$ is obtained using only $\mathcal{O}((M+N-r)r^2)$ flops, as opposed to the SVD, which requires $\mathcal{O}(MN^2)$ flops. We note that there are other alternatives to the SVD for the white noise case, such as the approach of [13], which is based on parallel subspace computation of submatrices of smaller size.

V. Relation to Previous Work

The proposed NSE estimator is close in spirit to work published in [10], [11]. Therein, the authors exploit a re-parameterization of the ML/NLLS criterion (9) in terms of the nullspace matrix $\mathbf{N}$ as follows: $\mathbf{XN} = \mathbf{0}$ means that $\mathbf{RN} = \mathbf{0}$, which implies that the matrix $\mathbf{C}^{T/2}(\mathbf{N}\otimes\mathbf{I})$ spans the orthogonal complement of span$\{\mathbf{C}^{-1/2}(\mathbf{R}^T\otimes\mathbf{I})$, appearing in (9). This means that the projections onto these spaces coincide:

$$\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{R}^T\otimes\mathbf{I})} = \mathbf{P}_{\mathbf{C}^{T/2}(\mathbf{N}\otimes\mathbf{I})}\,. \quad (29)$$

Therefore, the concentrated criterion (9) can be expressed as

$$V_{\text{sep}}(\mathbf{N}) = \left\|\mathbf{P}_{\mathbf{C}^{T/2}(\mathbf{N}\otimes\mathbf{I})}\mathbf{C}^{-1/2}\mathbf{y}\right\|^2 = \mathbf{y}^T(\mathbf{N}\otimes\mathbf{I})\mathbf{W}_{\text{opt}}(\mathbf{N}^T\otimes\mathbf{I})\mathbf{y}\,, \quad (30)$$

which we recognize as the NSE criterion (15). The differences to the proposed approach are twofold. In the referenced work, $\mathbf{W}_{\text{opt}} = \mathbf{W}(\mathbf{N})$ is a non-linear function of $\mathbf{N}$. In contrast, we show that any choice of $\mathbf{W} > \mathbf{0}$ yields consistent estimates of $\mathbf{N}$. Such an estimate can be used to calculate a fixed $\hat{\mathbf{W}}_{\text{opt}}$, which in turn gives an asymptotically optimal estimate of $\mathbf{N}$. The second difference lies in the parameterization of $\mathbf{N}$. In [11], $\mathbf{N}$ is filled with parameters, resulting in an over-parameterized (singular) optimization problem. In [10], $\mathbf{N}$ is instead constrained to be an orthogonal matrix, resulting in an optimization over the Grassman manifold. In contrast, we propose to use a linear parameterization that allows a computationally efficient solution by linear least-squares. Finally, we remark that the idea to fix the normalizing "middle matrix" in the projection matrix has been used before in a different context, see e.g. [14], [15].

VI. Numerical Examples

In this section, we present results from computer simulations using randomly generated low-rank matrices and noise.

Example 1 Missing data
In the first example, we consider the case where some data entries are missing and the measurement noise is white. The missing data is modeled by adding a large number to the diagonal of $\mathbf{C}$ at locations corresponding to the missing entries. The data dimensions are $M = 40$, $N = 20$, $r = 4$, the noise is $\mathbf{e} \in \mathcal{N}(0, \sigma^2\mathbf{I})$, and 10% data at random locations are missing. The low-rank matrix $\mathbf{X}$ is generated from (4), where the entries of $\mathbf{L}$ and $\mathbf{R}$ are i.i.d. $\mathcal{N}(0, 1)$. Figure 1 shows the total MSE of the reconstructed $\hat{\mathbf{X}}$ versus the SNR, defined as

$$\text{SNR} = \frac{\text{E}[|x_{i,j}|^2]}{\sigma^2}.$$

The following methods are compared: SVD, the proposed NSE, NSE-V, ALS with random initialization and ALS initialized by NSE-V. Here, NSE-V refers to the NSE applied to transformed data according to Remark 1. Only one iteration of ALS is applied, and the MSE:s as well as the CRB are all averaged over 1000 Monte-Carlo trials.

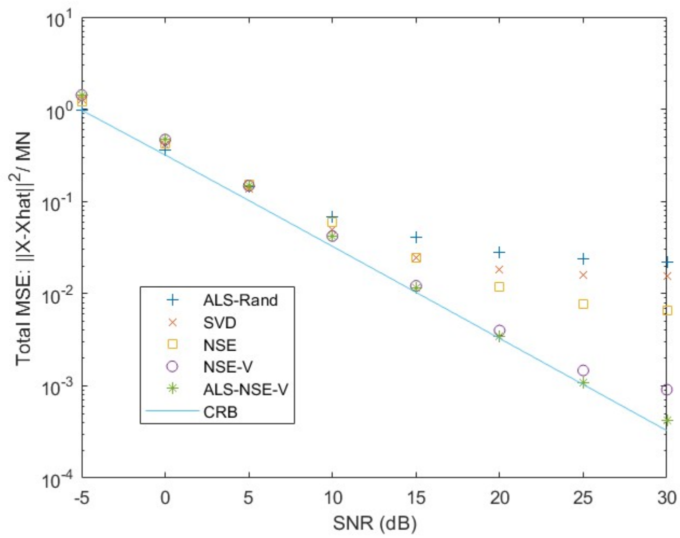

Fig. 1. Total MSE of $\hat{\mathbf{X}}$ vs SNR in dB. Missing data scenario.

As seen in the plot, the ALS with random initialization performs the worst (more iterations are needed), and NSE without pre-conditioning suffers in the high SNR region due to some realizations of $\mathbf{X}$ where the parameterization (5) is nearly invalid. Applying the transformation (NSE-V) leads to performance close to the CRB, and applying one iteration of ALS can only make a minor improvement, except at very high SNR, where also NSE-V suffers somewhat.

Example 2 Clutter-like Noise
In the second example, the noise has a strong (20 dB above noise floor) low-rank component, like a radar clutter, plus white "receiver noise":

$$\mathbf{C} = \sigma^2\left(\mathbf{C}_c + \mathbf{I}\right)$$

where $\mathbf{C}_c$ is a randomly generated rank-2 matrix with $\text{Tr}(\mathbf{C}_c)/(MN) = 20$ dB.

The rest of the parameters are as in Example 1, and the same methods are compared in Figure 2. In this plot we omitted the 5% smallest and 5% largest values when computing the MSE to reduce the influence of outliers (affected mainly NSE).

Figure 2 shows a similar general behavior as Figure 1, but in this more stressful scenario the performance differences are

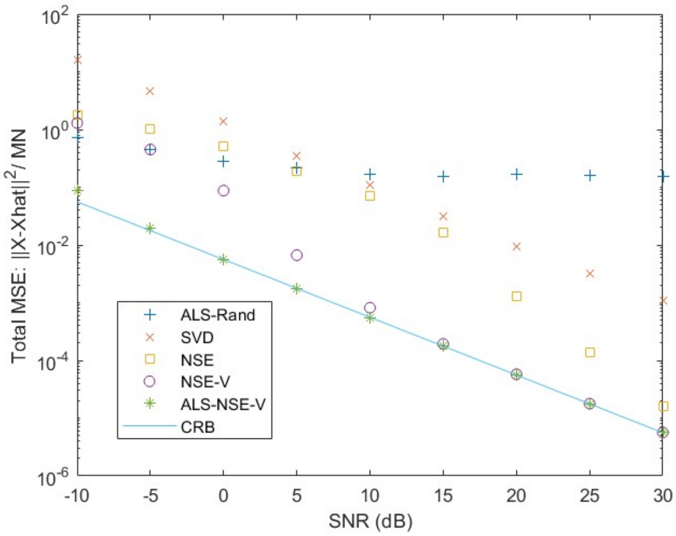

Fig. 2. Total MSE of $\hat{\mathbf{X}}$ vs SNR in dB. Clutter-like noise scenario.

more notable. Again, NSE-V achieves the CRB for high SNR, whereas one iteration of ALS initialized by NSE-V suffices to yield efficient estimates also for low SNR.

VII. Conclusions

We have proposed a non-iterative method for reconstructing a low-rank matrix from data corrupted by additive noise of arbitrary color. The proposed Null-Space Estimator (NSE) is based on modeling the nullspace rather than the image space of the data, and it is shown to yield statistically efficient estimates in the high SNR regime. For more stressful scenarios, it can be used to initialize an iterative method, like Alternating Least-Squares or a Newton-type optimization. In the computer simulations, just one iteration using ALS was found sufficient. The proposed approach can potentially reduce the computational complexity of low-rank matrix approximation substantially, thus allowing larger problems to be solved. Possibilities for future work include statistical analysis for large M (as opposed to high SNR), subspace tracking as M is incremented, and extension to tensor-valued data.

Appendix

Express the FIM using (10) and (12) as the 2×2 block matrix

$$\mathbf{B}^T\mathbf{C}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{R}\otimes\mathbf{I} \\ \tilde{\mathbf{I}}\otimes\mathbf{L}^T \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} \mathbf{R}^T\otimes\mathbf{I} & \tilde{\mathbf{I}}^T\otimes\mathbf{L} \end{bmatrix} \quad (31)$$

$$= \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix}, \quad (32)$$

where the blocks are given by

$$\mathbf{J}_{11} =(\mathbf{R}\otimes\mathbf{I})\mathbf{C}^{-1}(\mathbf{R}^T\otimes\mathbf{I}) \quad (33)$$

$$\mathbf{J}_{12} =(\mathbf{R}\otimes\mathbf{I})\mathbf{C}^{-1}(\tilde{\mathbf{I}}^T\otimes\mathbf{L}) \quad (34)$$

$$\mathbf{J}_{21} =(\tilde{\mathbf{I}}\otimes\mathbf{L}^T)\mathbf{C}^{-1}(\mathbf{R}^T\otimes\mathbf{I}) \quad (35)$$

$$\mathbf{J}_{22} =(\tilde{\mathbf{I}}\otimes\mathbf{L}^T)\mathbf{C}^{-1}(\tilde{\mathbf{I}}^T\otimes\mathbf{L})\,. \quad (36)$$

The block matrix inversion lemma now shows that

$$\begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} X & X \\ X & \left(\mathbf{J}_{22}-\mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12}\right)^{-1} \end{bmatrix}, \quad (37)$$

where X denotes blocks that are not of current interest and the lower right block is the CRB for $\mathbf{r}_2$. We get

$$\begin{aligned}\mathbf{J} &=\mathbf{J}_{22}-\mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12} = \cdots \\ &=(\tilde{\mathbf{I}}\otimes\mathbf{L}^T)\mathbf{C}^{-T/2}\,\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{R}^T\otimes\mathbf{I})}\,\mathbf{C}^{-1/2}(\tilde{\mathbf{I}}^T\otimes\mathbf{L})\,, \quad (38)\end{aligned}$$

where $\mathbf{P}^{\perp}_{\mathbf{C}^{-1/2}(\mathbf{R}^T\otimes\mathbf{I})}$ was introduced in (9). Next, we apply the re-parameterization (29), which inserted into (38) yields

$$\mathbf{J} = (\tilde{\mathbf{I}}\otimes\mathbf{L}^T)(\mathbf{N}\otimes\mathbf{I})\mathbf{W}_{\text{opt}}(\mathbf{N}^T\otimes\mathbf{I})(\tilde{\mathbf{I}}^T\otimes\mathbf{L})\,. \quad (39)$$

The last step is to observe that $\tilde{\mathbf{I}}\mathbf{N} = [\mathbf{0}\ \mathbf{I}][\mathbf{N}_1\ \mathbf{I}]^T = \mathbf{I}$, so that $(\tilde{\mathbf{I}}\otimes\mathbf{L}^T)(\mathbf{N}\otimes\mathbf{I}) = \mathbf{I}\otimes\mathbf{L}^T$, and therefore

$$\mathbf{J} = \mathbf{J}_{22}-\mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12} = (\mathbf{I}\otimes\mathbf{L}^T)\mathbf{W}_{\text{opt}}(\mathbf{I}\otimes\mathbf{L})\,, \quad (40)$$

which is precisely the matrix appearing in (28).

References

[1] K. Werner and M. Jansson, “Reduced rank linear regression and weighted low rank approximations,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2063–2075, 2006.

[2] G. Reinsel, R. Velu, and K. Chen, *Multivariate Reduced-Rank Regression*. New York, NY: Springer, 2022.

[3] W.-S. Lu, S.-C. Pei, and P.-H. Wang, “Weighted low-rank approximation of general complex matrices and its application in the design of 2-D digital filters,” *IEEE Trans Circ Sys I*, vol. 44, pp. 650–655, July 1997.

[4] M. Nejati, S. Samavi, H. Derksen, and K. Najarian, “Denoising by low-rank and sparse representations,” *J Vis Comm and Image Repr*, vol. 36, pp. 28–39, 2016.

[5] S. Lichtenberg and A. Tasissa, “Localization from structured distance matrices via low-rank matrix recovery,” *IEEE Trans IT*, 12 2024.

[6] N. Gillis and F. Glineur, “Low-Rank Matrix Approximation With Weights or Missing Data is NP-Hard,” *Siam J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1149–1165, 2011.

[7] N. Srebro and T. Jaakkola, “Weighted Low Rank Approximations,” in *Proc. ICML’03*, (Washington, DC), Aug. 2003.

[8] Y. Li, Y. Liang, and A. Risteski, “Recovery guarantee of weighted low-rank approximation via alternating minimization,” in *Proc. ICML’16*, (New York City, NY), Aug. 2016.

[9] Z. Song, M. Ye, J. Yin, and L. Zhang, “Efficient Alternating Minimization with Applications to Weighted Low Rank Approximation,” in *Proc. ICLR 2025*, (Singapore), Apr. 2025.

[10] J. Manton, R. Mahony, and Y. Hua, “The geometry of weighted low-rank approximations,” *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 500–514, 2003.

[11] K. Werner and M. Jansson, “Weighted low rank approximation and reduced rank linear regression,” in *ICASSP 2004*, pp. 501–504, 2004.

[12] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[13] V.-D. Nguyen, K. Abed-Meraim, N. Linh-Trung, and R. Weber, “Generalized minimum noise subspace for array processing,” *IEEE Trans on SP*, vol. 65, no. 14, pp. 3789–3802, 2017.

[14] Y. Bresler and A. Macovski, “Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise,” *IEEE Trans. on ASSP*, vol. ASSP-34, pp. 1081–1089, Oct. 1986.

[15] P. Stoica and K. Sharman, “Maximum Likelihood Methods for Direction-of-Arrival Estimation,” *IEEE Trans. on ASSP*, vol. ASSP-38, pp. 1132–1143, July 1990.