Constrained Least Squares for Extended Complex Factor Analysis

Ahmad Mouri Sardarabadi[1], Alle-Jan van der Veen[2] and L.V.E. Koopmans[1]

Abstract—For subspace estimation with an unknown colored noise, Factor Analysis (FA) and its extensions, denoted as Extended FA (EFA), are good candidates for replacing the popular eigenvalue decomposition (EVD). Finding the unknowns in (E)FA can be done by solving a non-linear least square problem. For this type of optimization problems, the Gauss-Newton (GN) algorithm is a powerful and simple method. The most expensive part of the GN algorithm is finding the direction of descent by solving a system of equations at each iteration. In this paper we show that for (E)FA, the matrices involved in solving these systems of equations can be diagonalized in a closed form fashion and the solution can be found in a computationally efficient way. We show how the unknown parameters can be updated without actually constructing these matrices. The convergence performance of the algorithm is studied via numerical simulations.

Index Terms—Factor Analysis, Non-Linear Optimization, Covariance Matching

I. Introduction

The eigenvalue decomposition (EVD) of a data covariance matrix (a.k.a. principal component analysis) is a powerful tool for subspace based and low-rank approximation techniques. Without noise, the data covariance matrix is considered to be rank-deficient, and its column span is called the signal subspace. In the presence of additive noise an implicit assumption is that this noise is white. If this is not the case but the noise covariance matrix is known from calibration, whitening techniques can be used as a pre-processing step. However, in many array processing applications this knowledge is not available. A preferable approach is to replace the EVD by techniques that jointly estimate the signal subspace and the noise covariance matrix. Using FA as a substitute for EVD in these cases was suggested by [1] and [2]. The latter also includes several extensions of the classical FA model which were denoted as Extended Factor Analysis (EFA).

The FA model was introduced by Spearman [3] in 1904 and became an established multivariate technique mainly due to the work of Lawley, Anderson, Jöreskog and others between 1940 and 1970 [4], [5], [6], [7], [8]. The FA model assumes a covariance matrix $\mathbf{R}$ of the data under study (e.g., samples acquired from an array of sensors) can be modeled as

$$\mathbf{R} = \mathbf{A}\mathbf{A}^H + \mathbf{D}, \tag{1}$$

where ^{H} is the Hermitian transpose, $\mathbf{A}$ is a "tall" matrix ($\mathbf{A}\mathbf{A}^H$ has low rank), and $\mathbf{D}$ is a positive diagonal matrix. In terms of subspace-based techniques, $\mathbf{A}$ captures the signal

1 A. Mouri Sardarabadi (ammsa@astro.rug.nl) and L.V.E. Koopmans are affiliated with Kapteyn Astronomical Institute, University of Groningen, The Netherlands. 2 Alle-Jan van der Veen is affiliated with Delft University of Technology, Delft, The Netherlands

subspace while $\mathbf{D}$ can model the noise covariance matrix. Given a sample covariance matrix $\hat{\mathbf{R}}$, the objective of FA is to estimate $\mathbf{A}$ and $\mathbf{D}$.

In this paper we use the EFA which was introduced in [9], [2] and present a new algorithm based on the Gauss-Newton (GN) method. By showing that a closed-form Singular Value Decomposition (SVD) of the Jacobian matrix can be found, we develop an efficient algorithm that has competitive complexity and performance.

The structure of the paper is as follows: in Sec. II we discuss the EFA model, in Sec. III we introduce the least squares (LS) formulation of the problem and discuss our method and in Sec. IV we use numerical simulations to evaluate the performance of the algorithm.

II. Data Model

We assume to have access to N samples from the outputs of P receivers such that a sample covariance matrix can be constructed using

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}[n]\mathbf{y}[n]^H \tag{2}$$

where the $P \times 1$ vector $\mathbf{y}[n]$ represents the zero mean sampled output of the receivers. We also assume that the Extended Factor Analysis [2] model is valid

$$\mathbf{R} = \mathcal{E}\{\hat{\mathbf{R}}\} = \mathbf{A}\mathbf{A}^H + \mathbf{\Psi}, \tag{3}$$

where $\mathcal{E}\{.\}$ represents the expectation of the argument, $\mathbf{A}$ is a $P \times Q$ matrix with rank$(\mathbf{A}) = Q$ and $\mathbf{\Psi}$ is a structured matrix such that $\mathbf{\Psi} = \mathbf{M} \odot \mathbf{\Psi}$ for some mask matrix $\mathbf{M}$ consisting of zeros. We use $\odot$ to denote the element-wise or Hadamard product. We assume to know the structure of $\mathbf{\Psi}$ (and hence $\mathbf{M}$) in advance, and define a selection matrix, $\mathbf{S}$, for the non-zero elements of $\mathbf{M}$. Using this selection matrix we can stack the non-zero elements of $\mathbf{\Psi}$ into a column vector as $\psi = \mathbf{S}^T \text{vect}(\mathbf{\Psi})$ and similarly define the vectorized matrix as $\text{vect}(\mathbf{\Psi}) = \mathbf{S}\psi$. Here we used the operator vect(.), which stacks the columns of the argument matrix into a single column vector. For the classical FA, $\mathbf{M} = \mathbf{I}$ and $\mathbf{\Psi} = \mathbf{D}$ is diagonal. Given a sample covariance matrix we are interested in finding $\mathbf{A}$ and $\mathbf{\Psi}$ or only one of them, depending on the application.

This parametrization of the covariance model is not unique and for any unitary matrix $\mathbf{Q}$, $\mathbf{A}\mathbf{Q}$ results in exactly the same model. As a result, in the complex case, the problem needs additional Q^2 constraints for a unique parametrization [1], [2]. There are several ways to constrain $\mathbf{A}$ such as restricting it to be lower triangular with real (positive) diagonal elements,

or forcing its columns to be orthogonal such that $\mathbf{A}^H\mathbf{A} = \mathbf{\Gamma}$ is a diagonal matrix. Regardless of the chosen constraints, the total number of unique parameters for the problem is

$$\hat{n} = 2PQ + \|\mathbf{M}\|_1 - Q^2. \qquad (4)$$

We show that using the constraints of the form $\mathbf{A}^H\mathbf{A} = \mathbf{\Gamma}$ can be advantageous when solving the EFA model using the least squares cost function. In the next section we present our approach.

III. Least Squares for EFA

A. Gauss-Newton Method

The aim is to find the FA model for a sample covariance $\hat{\mathbf{R}}$ matrix by minimizing the following non-linear LS cost function:

$$\min_{\mathbf{A},\mathbf{\Psi}} \|\hat{\mathbf{R}} - \mathbf{A}\mathbf{A}^H - \mathbf{\Psi}\|_F^2.$$

Given the non-linearity of the cost function we use the Gauss-Newton method to find the solution similar to [10], [2]. The GN algorithm for solving this problem consists of the following updates

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \mu_k\boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is the direction of descent and the solution to the following system of linear equations

$$\mathbf{J}^H\mathbf{J}\boldsymbol{\delta} = \mathbf{J}^H(\hat{\mathbf{r}} - \mathbf{r}), \qquad (5)$$

where $\hat{\mathbf{r}} = \text{vect}(\hat{\mathbf{R}})$, $\mathbf{r} = \text{vect}(\mathbf{A}\mathbf{A}^H + \mathbf{D})$, and $\mathbf{J} = \partial\mathbf{r}/\partial\boldsymbol{\theta}^T$ is the Jacobian matrix. Using the relation between the Kronecker product and vectorization, we find the Jacobian matrix to be

$$\mathbf{J} = \frac{\partial\mathbf{r}}{\partial\boldsymbol{\theta}^T} = \begin{bmatrix} \mathbf{A}^* \otimes \mathbf{I}_P & (\mathbf{I}_P \otimes \mathbf{A})\mathbf{K}^{P,Q} & \mathbf{S} \end{bmatrix},$$

where

$$\boldsymbol{\theta} = \begin{bmatrix} \text{vect}(\mathbf{A})^T & \text{vect}(\mathbf{A})^H & \boldsymbol{\psi}^T \end{bmatrix}^T,$$

$\mathbf{K}^{P,Q}$ is a permutation such that $\text{vect}(\mathbf{X}^T) = \mathbf{K}^{P,Q}\text{vect}(\mathbf{X})$ for any $P \times Q$ matrix $\mathbf{X}$, $\otimes$ represents the Kronecker product and * represent the complex conjugate. We also have $\mathbf{K}^{P,Q}\mathbf{K}^{Q,P} = \mathbf{I}$.

The most expensive operation during the GN updates is the solution of (5). We also know that the non-uniqueness of $\mathbf{A}$ causes the Jacobian to be rank deficient by Q^2, which makes the system singular. A closed form solution of this system was first introduced in [2] using a block LDU decomposition. While that method is more general and also works for weighted least squares (WLS), it does not produce a minimum length solution and also does not clearly show how (5) is reduced to a non-singular system. In the next section we show that a unique minimum length solution can be found when the constraint $\mathbf{A}^H\mathbf{A}$ is enforced at each iteration.

B. Model Reduction

In this section we show how the constraints $\mathbf{A}^H\mathbf{A} = \mathbf{\Gamma}$ can be used to transform the problem into a reduced one which does not suffer from singularities. We also show how this transformation can be used to solve the Gauss-Newton system of equations in an efficient way. One way to find

$\mathbf{A}$ that satisfies the constraints is to use the singular value decomposition. Let the SVD of $\mathbf{A}$ be

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_0 & | & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{0} \end{bmatrix}^{1/2} \mathbf{Q}^H. \qquad (6)$$

If $\mathbf{A}$ already satisfies the constraints then $\mathbf{Q} = \mathbf{I}_Q$. Otherwise, we can always multiply $\mathbf{A}$ from right by $\mathbf{Q}$ to make sure that the constraints are satisfied. For simplicity we also define

$$\mathbf{J_A} = \begin{bmatrix} \mathbf{A}^* \otimes \mathbf{I}_P & (\mathbf{I}_P \otimes \mathbf{A})\mathbf{K}^{P,Q} \end{bmatrix},$$

which is the submatrix of $\mathbf{J}$ corresponding to derivatives with respect to $\mathbf{A}$ and $\mathbf{A}^*$.

Theorem 1. *Given the SVD of $\mathbf{A}$ as defined by (6), the eigenvalue decomposition of $\mathbf{J_A}^H\mathbf{J_A}$ is given by*

$$\mathbf{J_A}^H\mathbf{J_A} = \tilde{\mathbf{V}}\tilde{\mathbf{\Gamma}}\tilde{\mathbf{V}}^H, \qquad (7)$$

where

$$\tilde{\mathbf{\Gamma}} = \left[\begin{array}{ccc|c} \mathbf{\Gamma} \otimes \mathbf{I} & & & \\ & \mathbf{I} \otimes \mathbf{\Gamma} & & \\ & & \mathbf{\Gamma} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{\Gamma} & \\ \hline & & & \mathbf{0} \end{array}\right],$$

is a diagonal matrix containing the eigenvalues and $\tilde{\mathbf{V}}$ is a unitary matrix containing the eigenvectors. The matrix $\tilde{\mathbf{V}}$ can be defined as $\tilde{\mathbf{V}} = [\mathbf{V}|\mathbf{Z}]$ where

$$\mathbf{V} = \begin{bmatrix} \mathbf{I}_{PQ} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}^{Q,P} \end{bmatrix} \begin{bmatrix} \mathbf{I}_Q \otimes \mathbf{U}_n & \mathbf{0} & (\mathbf{\Gamma}^{1/2} \otimes \mathbf{U}_0)\mathbf{G}^{1/2} \\ \mathbf{0} & (\mathbf{U}_n^* \otimes \mathbf{I}_Q) & (\mathbf{U}_0^* \otimes \mathbf{\Gamma}^{1/2})\mathbf{G}^{1/2} \end{bmatrix},$$

and

$$\mathbf{Z} = \begin{bmatrix} (\mathbf{I}_Q \otimes \mathbf{A}) \\ -\mathbf{K}^{Q,P}(\mathbf{A}^* \otimes \mathbf{I}_Q) \end{bmatrix} \mathbf{G}^{1/2}.$$

with $\mathbf{G} = (\mathbf{\Gamma} \otimes \mathbf{I}_Q + \mathbf{I}_Q \otimes \mathbf{\Gamma})^{-1}$.

Proof. Carrying out the multiplications on both sides of (7) verifies the equality, but it is important to show that $\tilde{\mathbf{V}}$ is a unitary matrix. It is trivial to show that $\mathbf{V}^H\mathbf{V} = \mathbf{I}$ using the fact that $\mathbf{U}_n^H\mathbf{U}_n = \mathbf{I}_{P-Q}$, $\mathbf{U}_0^H\mathbf{U}_0 = \mathbf{I}_Q$, $\mathbf{U}_n^H\mathbf{U}_0 = \mathbf{0}$ and the property $(\mathbf{A} \otimes \mathbf{B})^H(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}^H\mathbf{C} \otimes \mathbf{B}^H\mathbf{D})$. What is left to show is that both $\mathbf{J_A}\mathbf{Z}$ and $\mathbf{V}^H\mathbf{Z}$ are zero.

We can easily verify that

$$\mathbf{J_A}\mathbf{Z} = [(\mathbf{A}^* \otimes \mathbf{A}) - (\mathbf{A}^* \otimes \mathbf{A})]\mathbf{G}^{1/2} = \mathbf{0},$$

This shows that $\mathbf{Z}$ is in the null-space of $\mathbf{J_A}$ (and $\mathbf{J}$). We also have

$$\mathbf{Z}^H\mathbf{Z} = \mathbf{G}^{1/2}[(\mathbf{A}^T\mathbf{A}^* \otimes \mathbf{I}_Q) + (\mathbf{I}_Q \otimes \mathbf{A}^H\mathbf{A})]\mathbf{G}^{1/2}$$
$$= \mathbf{G}^{1/2}[(\mathbf{\Gamma} \otimes \mathbf{I}_Q) + (\mathbf{I}_Q \otimes \mathbf{\Gamma})]\mathbf{G}^{1/2} = \mathbf{I}_{Q^2}$$

which confirms that $\text{rank}(\mathbf{Z}) = Q^2$. Now that we have shown that $\mathbf{Z}$ is a basis for the null-space of $\mathbf{J}$, it is sufficient to show that $\mathbf{V}^H\mathbf{Z} = \mathbf{0}$. Again, using $\mathbf{U}_n^H\mathbf{A} = \mathbf{0}$ and $\mathbf{U}_0^H\mathbf{A} = \mathbf{\Gamma}^{1/2}$ we can easily verify that $\mathbf{V}^H\mathbf{Z} = \mathbf{0}$. $\square$

Using this result we can define a new basis

$$\mathcal{V} = \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

which allows us to remove the null-space of $\mathbf{J}$ and reformulate the system of linear equations for solving the direction of

decent $\boldsymbol{\delta}$, (5), as

$$\begin{aligned}
\mathbf{J}^H\mathbf{J}\boldsymbol{\delta} &= \mathbf{J}^H(\hat{\mathbf{r}} - \mathbf{r}) \\
\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^H\mathbf{J}^H\mathbf{J}\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^H\boldsymbol{\delta} &= \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^H\mathbf{J}^H(\hat{\mathbf{r}} - \mathbf{r}) \\
(\boldsymbol{\mathcal{V}}^H\mathbf{J}^H\mathbf{J}\boldsymbol{\mathcal{V}})\tilde{\boldsymbol{\delta}} &= \boldsymbol{\mathcal{V}}^H\mathbf{J}^H(\hat{\mathbf{r}} - \mathbf{r}),
\end{aligned}$$

where in the last equation we changed the variable $\tilde{\boldsymbol{\delta}} \equiv \boldsymbol{\mathcal{V}}^H\boldsymbol{\delta}$. Because the systems are consistent we have $\boldsymbol{\delta} = \boldsymbol{\mathcal{V}}\tilde{\boldsymbol{\delta}}$. This transformation removes the Q^2 redundant dimensions due to the unitary ambiguity.

In the next section we show how to find the solution to this system of equations without actually constructing the matrices $\mathbf{J}$ and $\boldsymbol{\mathcal{V}}$.

C. Improved Linear Solver

In [2] it was shown that we can split the direction of descent, $\boldsymbol{\delta}$ as

$$\boldsymbol{\delta} = [\boldsymbol{\delta}_\mathbf{A}^T, \boldsymbol{\delta}_\Psi]^T,$$

and find each sub-vector sequentially. This can be done because $\mathbf{J}_\mathbf{A}$ and $\mathbf{S}$ must be linearly independent for the (E)FA problem to be identifiable [2]. First, using the generalized matrix inversion-lemma [11] on the Gramian of $\mathbf{J}$ in (5), we find $\boldsymbol{\delta}_\Psi$ by solving

$$(\mathbf{I} - \mathbf{T})\boldsymbol{\delta}_\Psi = \mathbf{S}^H\text{vect}[\mathbf{P}^\perp(\hat{\mathbf{R}} - \Psi)\mathbf{P}^\perp], \quad (8)$$

where $\mathbf{T} = \mathbf{S}^H(\mathbf{P}^T \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{P} - \mathbf{P}^T \otimes \mathbf{P})\mathbf{S}$, $\mathbf{P} = \mathbf{U}_0\mathbf{U}_0^H$ and $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$. The matrices $\mathbf{I} - \mathbf{T}$ and $\mathbf{T}$ are positive definite and positive semi-definite respectively. This means that $\mathbf{T}$ is convergent, i.e. its spectral radius satisfies $\rho(\mathbf{T}) < 1$. This allows us to use simple iterative solvers to find the solution in an efficient way. Once we have $\boldsymbol{\delta}_\Psi$, we use substitution and find $\boldsymbol{\delta}_\mathbf{A}$ by solving

$$(\mathbf{V}^H\mathbf{J}_\mathbf{A}^H\mathbf{J}_\mathbf{A}\mathbf{V})\tilde{\boldsymbol{\delta}}_\mathbf{A} = \mathbf{V}^H\mathbf{J}_\mathbf{A}^H(\hat{\mathbf{r}} - \mathbf{r} - \mathbf{S}\boldsymbol{\delta}_\Psi),$$

followed by $\boldsymbol{\delta}_\mathbf{A} = \mathbf{V}\tilde{\boldsymbol{\delta}}_\mathbf{A}$. Using the results from Theorem 1 we have

$$(\mathbf{V}^H\mathbf{J}_\mathbf{A}^H\mathbf{J}_\mathbf{A}\mathbf{V})^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{-1} \otimes \mathbf{I} & & \\ & \mathbf{I} \otimes \boldsymbol{\Gamma}^{-1} & \\ & & \mathbf{G} \end{bmatrix},$$

which makes finding $\tilde{\boldsymbol{\delta}}_\mathbf{A}$ trivial. Similar to the the definition of $\boldsymbol{\theta}$, for $\boldsymbol{\delta}_\mathbf{A}$ we have

$$\boldsymbol{\delta}_\mathbf{A} = \begin{bmatrix} \text{vect}(\boldsymbol{\Delta}_\mathbf{A}) \\ \text{vect}(\boldsymbol{\Delta}_\mathbf{A}^*) \end{bmatrix}.$$

The matrix form of the solution, $\boldsymbol{\Delta}_\mathbf{A}$, can be found by introducing some intermediate results:

$$\begin{aligned}
\mathbf{E} &= \hat{\mathbf{R}} - \mathbf{A}\mathbf{A}^H - \Psi - \boldsymbol{\Delta}_\Psi \\
\boldsymbol{\Delta}_1 &= \mathbf{U}_0[\tilde{\mathbf{G}} \odot (\boldsymbol{\Gamma}\mathbf{U}_0^H\mathbf{E}\mathbf{U}_0 + \mathbf{U}_0^H\mathbf{E}\mathbf{U}_0\boldsymbol{\Gamma})\boldsymbol{\Gamma}^{1/2}] \\
\boldsymbol{\Delta}_2 &= (\mathbf{I} - \mathbf{P})\mathbf{E}\mathbf{U}_0\boldsymbol{\Gamma}^{-1/2} \\
\boldsymbol{\Delta}_\mathbf{A} &= \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2, \quad (9)
\end{aligned}$$

where $\text{vect}(\boldsymbol{\Delta}_\Psi) \equiv \mathbf{S}\boldsymbol{\delta}_\Psi$ is the matrix form of the direction of descent for Ψ and $\tilde{\mathbf{G}} = \text{unvect}(\text{vectdiag}(\mathbf{G}^2))$ is a $Q \times Q$ matrix[1] constructed by unvectorizing the diagonal elements of $\mathbf{G}^2$. By closer inspection we see that $\boldsymbol{\Delta}_1$ is the update in the

---

[1] Let $\boldsymbol{\gamma} = \text{vectdiag}(\boldsymbol{\Gamma})$ then $\tilde{\mathbf{G}} = (\boldsymbol{\gamma}\mathbf{1}_Q^T + \mathbf{1}_Q\boldsymbol{\gamma}^T)^{\odot -2}$ where $\odot^{-1}$ is the Hadamard or element-wise inversion of a matrix (i.e. for $\mathbf{X} = [x_{ij}]$, $\mathbf{X}^{\odot -n} = [1/x_{ij}^n]$).

current subspace of $\mathbf{A}$ and $\boldsymbol{\Delta}_2$ is the update in its null-space. It is also clear that we do not need to explicitly find $\mathbf{U}_n$ which allows us to use the economic-size SVD rather than a full one (or use EVD of $\mathbf{A}^H\mathbf{A}$).

To summarize, we use the following updates during the GN iterations:

$$\begin{aligned}
\Psi^{(k+1)} &= \Psi^{(k)} + \mu_k\boldsymbol{\Delta}_\Psi \\
\mathbf{A}^{(k+1)} &= \mathbf{A}^{(k)} + \mu_k\boldsymbol{\Delta}_\mathbf{A}
\end{aligned}$$

where $\boldsymbol{\Delta}_\Psi$ and $\boldsymbol{\Delta}_\mathbf{A}$ are given by (8) and (9) respectively. Once the direction of descent is found we can find the optimal μ_k using the procedure described in [2] which requires solving for the roots of a third order polynomial with real coefficients for which closed form solutions exist.

IV. SIMULATION

We evaluate the convergence speed of various algorithms for the classical factor analysis model. An array with $P = 100$ elements is simulated. The matrix $\mathbf{A}$ is chosen randomly with a standard complex Gaussian distribution (i.e. each element is distributed as $\mathcal{CN}(0, 1)$) and the diagonal entries of $\mathbf{D}$ are chosen randomly with a uniform distribution between 1 and 5.

For $P = 100$, the maximum number of sources is $Q_{\max} = 89$. We show simulation results for $Q = 20$, representative for low-rank cases, and for $Q = 80$ for high-rank cases. Sources and noise are generated using standard unit power complex Gaussian distributions.

We compare our algorithm (denoted as "Reduced method") with the classical ad-hoc alternating LS (see [12], [8], [2] for details), the Block LDU based solvers proposed in [2], Jöreskog [13], and the more recent CM method by [14]. Because the method presented here is based on LS we set the weighting matrix for the Block LDU based algorithms equal to identity. Fig. 1 shows the convergence behavior based on the gradient.

The proposed algorithm has similar convergence properties as the Block LDU. This is expected because these algorithms also use the GN updates. CM and the ad-hoc method are more sensitive to the number of sources, Q, and converge very slowly for large Q. Also, for large Q, both Jöreskog and CM might fail to converge for the same model but different noise realization (see Fig. 1c). While it is possible to find models for which our algorithm also fails to converge as the problem is non-convex and non-linear it is our experience that in those scenarios the other methods also fail to converge with high probability. Also for the same model (i.e. the same $\mathbf{A}$ and $\mathbf{D}$) our method is less sensitive to the noise realization.

V. CONCLUSION

In this paper we have shown how constraining the columns of $\mathbf{A}$ to be orthogonal in a(n) (E)FA problem allows us to find a diagonalization of the Gauss-Newton approximation of the Hessian. Using this diagonalization we have developed a new non-linear least squares algorithm with competitive convergence speed. We have also shown that the system of
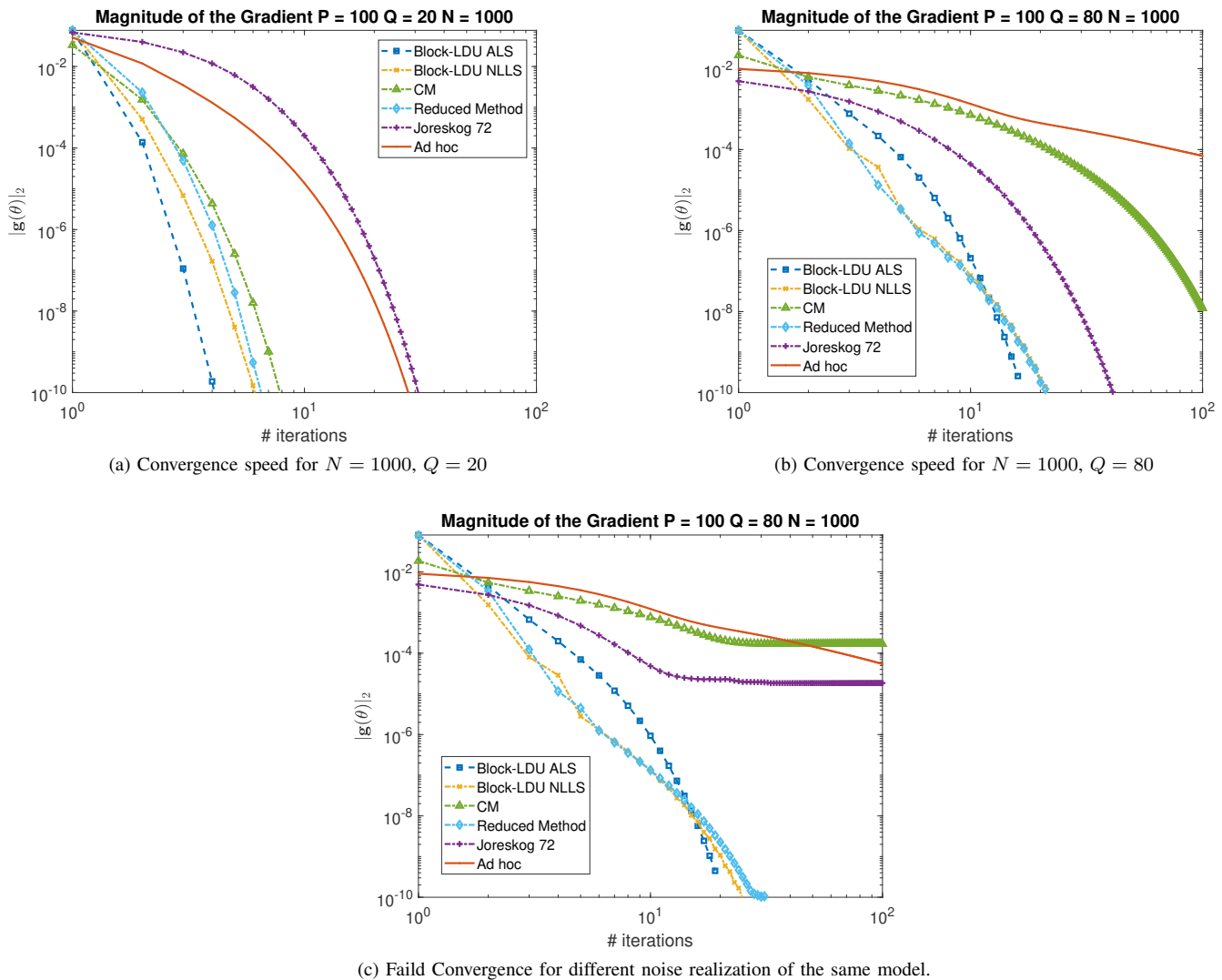
(a) Convergence speed for $N = 1000$, $Q = 20$

(b) Convergence speed for $N = 1000$, $Q = 80$

(c) Faild Convergence for different noise realization of the same model.

Fig. 1: Convergence for $P = 100$ sensors and varying number of samples N and sources Q.

equations needed for updating the noise covariance matrix, $\mathbf{\Psi}$, has a convergent matrix which allows fast iterative algorithms to be applied.

Using simulation we demonstrated the newly proposed algorithm has a competitive convergence properties and is robust to different noise realizations.

The proposed algorithm does not guarantee the positive (semi-)definiteness of $\mathbf{\Psi}$. This will be address in future works.

REFERENCES

[1] A.-J. van der Veen, A. Leshem, and A.-J. Boonstra, "Array signal processing for radio astronomy," in *The Square Kilometre Array: An Engineering Perspective*, P. Hall, Ed. Dordrecht: Springer, 2005, pp. 231–249, iSBN 1-4020-3797-x. Reprinted from Experimental Astronomy, 17(1-3),2004. [Online]. Available: http://ens.ewi.tudelft.nl/pubs/aj04ska.pdf

[2] A. M. Sardarabadi and A. J. van der Veen, "Complex factor analysis and extensions," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 954–967, Feb 2018.

[3] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, Jan 1904.

[4] D. N. Lawley, "The estimation of factor loadings by the method of maximum likelihood." *Proceedings of the Royal Society of Edinburgh*, vol. 60, no. 01, pp. 64–82, 1940.

[5] T. W. Anderson and H. Rubin, "Statistical inference in factor analysis," *In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5, pp. 111 – 150, 1956. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0024379573900232

[6] K. G. Jöreskog, "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika*, vol. 34, no. 2, pp. 183–202, 1969.

[7] D. N. Lawley and A. Maxwell, *Factor analysis as a statistical method*, ser. 2nd. ed. New York: Am. Elsevier Publ., 1971.

[8] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[9] A. Mouri Sardarabadi and A.-J. van der Veen, "Subspace estimation using factor analysis," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012, pp. 477 –480.

[10] S. Y. Lee, "The Gauss-Newton algorithm for the Weighted Least Squares factor analysis," *Journal of the Royal Statistical Society*, vol. 27, no. 2, June 1978.

[11] C. A. Rohde, "Generalized inverses of partitioned matrices," *Journal of*

the Society for Industrial and Applied Mathematics, vol. 13, no. 4, pp. 1033–1035, 1965.

[12] T. Anderson, *An Introduction to Multivariate Statistical Analysis*, third edition ed.   Wiley, 2003.

[13] K. G. Jöreskog and A. S. Goldberger, "Factor analysis by generalized least squares," *Psychometrika*, vol. 37, no. 3, pp. 243–260, Sep 1972. [Online]. Available: http://dx.doi.org/10.1007/BF02306782

[14] J.-H. Zhao, P. Yu, and Q. Jiang, "ML estimation for factor analysis: EM or non-EM?" *Statistics and Computing*, vol. 18, pp. 109–123, 2008, 10.1007/s11222-007-9042-y. [Online]. Available: http://dx.doi.org/10.1007/s11222-007-9042-y