Estimation of Sensor Array Signal Model Parameters Using Factor Analysis

Andreas I. Koutrouvelis*, Richard C. Hendriks*, Richard Heusdens* and Jesper Jensen†‡
e-mails: {a.koutrouvelis, r.c.hendriks, r.heusdens}@tudelft.nl, and {jesj}@oticon.com
*Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands
†Dept. Electronic Systems, Aalborg University, Denmark, ‡Oticon A/S, Denmark

***Abstract*—Factor analysis is a popular tool in multivariate statistics, applied in several areas of study such as psychology, economics, chemistry and signal processing. Given a set of observed random variables, factor analysis aims at explaining and analyzing the correlation between these random variables. This is done by finding a meaningful structural model representation for the correlation matrix of the observed random variables, and subsequently estimating the underlying model parameters. In this paper, we focus on factor analysis methods applied to a commonly used signal model for sensor arrays applications and use it to jointly estimate the underlying model parameters. In addition we discuss practical considerations of these methods.**

***Index Terms*—Factor analysis, sensor array, signal processing.**

I. Introduction

In sensor array signal processing applications [1]–[3], the covariance matrix of the observed random variables at the sensors is often described by a signal model, which depends on a number of model parameters. As these parameters are often unknown, estimation is often required. Examples of such model parameters could be the relative transfer functions (RTFs) between the sources and the sensor array, the power spectral densities (PSDs) of the sources, the number of the sources and the PSDs of the self-noise of the sensors. Several methods have been proposed in the literature to jointly estimate such model parameters [4]–[11].

In this paper, we focus on parameter estimation methods that are based on factor analysis theory (see e.g., [12] for an overview). Unlike the well-known principal component analysis [13], [14] which aims to explain the variance of the observed random variables, factor analysis aims to explain the covariance of the observed random variables [15]. Specifically, factor analysis finds a set of factors (sources) that explain the correlations between the observed random variables.

Factor analysis mainly consists of two branches: the exploratory factor analysis (EFA) [12], [15]–[21] and the confirmatory factor analysis (CFA) [12], [22], [23]. EFA methods decompose the covariance matrix of the observed random variables into the sum of a low-rank non-diagonal covariance matrix and a diagonal covariance matrix and based on this one can extract the number of the important factors (sources) which are responsible for the correlation of the observed random variables. Once the number of sources is known, CFA methods go one step further and deal with the estimation of the powers of specific sources and how exactly they are correlated with the observed random variables and with each other.

In this paper, we demonstrate the usefulness of both EFA and CFA methods in the context of parameter estimation in sensor arrays signal processing applications. In [7], several CFA-based methods were proposed to estimate signal model parameters in a microphone array context. The methods in [7] assume that the number of sources is known and constant in all time-frequency bins. Specifically, in [7] the number of sources was set to the maximum number of sources present in the acoustic scene. However, in many time-frequency bins the number of sources is much less and, thus, the number of parameters that are needed to be estimated are much less. In order to reduce the parameter space and reduce the computational complexity of the methods proposed in [7], in this paper, we adapt the number of sources per time-frequency bin such that we estimate only the essential parameters.

Determining the number of sources is a topic that has been addressed from different angles, e.g., [24]–[27]. An important class of methods among these, is based on information theoretic concepts like the minimum description length (MDL) [28] and Aikaike's information criterion (AIC) [29]. However, the methods based on information theoretic concepts as in [25] make rather strong statistical assumptions. When translated to our scenario, these assumptions cannot always be validated. Among these is the assumption that the variance of the sensor self noise should be identical for all sensors. Violating this assumption will result in an incorrect estimation of the number of sources. Another source counting method with less strong assumptions is the scree test proposed in [24].

In this paper, we use the scree test and we adjust it to the broadband-signal scenario such that it becomes more robust to outliers. Finally, we examine the performance of one of the CFA-based methods proposed in [7] using the adaptive number of sources in a source separation framework. The results show that the computational complexity is significantly reduced at the price of slight degradation of the segmental-signal-to-noise-ratio (SSNR) [30] and predicted intelligibility of the separated sources compared to the original CFA-based method in [7].

II. Multivariate Signal Model

Let us assume that we have a sensor array of M elements in total. When the nature of signals, in the application at

978-9-0827-9703-9/19/$31.00

hand, is broadband, it is frequently chosen to process each time-frequency bin independently assuming that they are uncorrelated. This assumption is based on the fact that the DFT approximately decorrelates the time samples if the time-frame is sufficiently long. Consequently, the signal model, used in this paper, represents the information of a single time-frequency bin[1]. That is,

$$\mathbf{y} = \underbrace{\sum_{i=1}^{r} \mathbf{a}_i s_i}_{\mathbf{As}} + \mathbf{n}, \tag{1}$$

with $\mathbf{y} \in \mathbb{C}^{M\times 1}$ the observed random variables, $\mathbf{a}_i \in \mathbb{C}^{M\times 1}$ the i-th RTF vector (also known as factor loading vector), $s_i \in \mathbb{C}$ the DFT coefficient of the i-th source signal (also known as common factor), r the number of sources, and $\mathbf{n}$ all remaining contributions that are not point sources. The matrix $\mathbf{A} \in \mathbb{C}^{M\times r}$ has as its columns the RTF vectors, and the vector $\mathbf{s} \in \mathbb{C}^{r\times 1}$ has as its elements the DFT coefficients of the sources' signals. If $\mathbf{A}$ and some realizations of $\mathbf{y}$ are known, the model in (1) is called a *linear regression model* and the parameters are estimated using classic linear regression methods (see e.g., [31] for an overview). If only some realizations of $\mathbf{y}$ are known it is called a *factor model* [12] and the parameters are estimated using factor analysis methods (see e.g., [12] for an overview). The factor model typically assumes that $\mathbf{n}$ is uncorrelated among sensors. If this assumption is not true, the model is called *approximate factor model* [32], [33]. Although, an approximate factor model may be in some cases more accurate than the factor model, in this paper, for simplicity, we focus mainly on the factor model.

If $\mathbf{s}$ is uncorrelated with $\mathbf{n}$, the population covariance matrix of the observed random variables, $\mathbf{P_y} = \mathrm{E}[(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y})(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y})^H] \in \mathbb{C}^{M\times M}$ in the factor model in (1) is given by [12]

$$\mathbf{P_y} = \underbrace{\mathbf{APA}^H}_{\boldsymbol{\Phi}} + \mathbf{D}, \tag{2}$$

where $\mathbf{P} = \mathrm{E}[(\mathbf{s} - \boldsymbol{\mu}_\mathbf{s})(\mathbf{s} - \boldsymbol{\mu}_\mathbf{s})^H] \in \mathbb{C}^{r\times r}$, $\mathbf{D} = \mathrm{Diag}(d_1, d_2, \cdots, d_M) \in \mathbb{R}^{M\times M}$, $d_i = \mathrm{E}[|n_i - \mu_{n_i}|^2] \in \mathbb{R}$, and $\mathbf{A}$ are the unknown parameters. Unfortunately, there is not an one-to-one relationship between the parameters $\mathbf{A}, \mathbf{P}$ and the population covariance matrix $\mathbf{P_y}$. In fact, for any non-singular matrix $\mathbf{T} \in \mathbb{C}^{r\times r}$, the following relationship holds [22]

$$\mathbf{P_y} = \underbrace{\mathbf{AT}^{-1}}_{\bar{\mathbf{A}}} \underbrace{\mathbf{TPT}^H}_{\bar{\mathbf{P}}} \underbrace{\mathbf{T}^{-H}\mathbf{A}^H}_{\bar{\mathbf{A}}^H} + \mathbf{D}. \tag{3}$$

That is, there are infinite many pairs of $\bar{\mathbf{A}}, \bar{\mathbf{P}}$ that produce exactly the same $\mathbf{P_y}$. The ambiguity introduced by $\mathbf{T}$ in (3) together with the large number of unknowns make the estimation of all the parameters of the factor model very challenging compared to the linear regression type of problems in which the number of unknowns is much smaller. In order to establish an one-to-one relationship between the parameters and $\mathbf{P_y}$, the number of equations should be sufficiently larger than the number of unknowns and some of the elements in $\mathbf{A}, \mathbf{P}$ should be fixed to constant values [22], [23].

Typically, the problem of parameter estimation in factor models is split into two phases. The first phase has a more exploratory nature in which EFA techniques can be used (see Sec. III) in order to determine the number of sources r. In the second phase, CFA methods (see Sec. IV) estimate all the remaining parameters of the factor model using the estimated r from the first phase. The only prior knowledge that is used in both phases is the sample estimate $\hat{\mathbf{P}}_\mathbf{y}$, which is in general different from $\mathbf{P_y}$ due to estimation and model-mismatch errors.

III. Source Counting Using EFA Methods

The estimation of the number of sources in the factor model has been investigated by many researchers. The method in [25] estimates the number of sources by minimizing the MDL and AIC information theoretic criteria (introduced in [28], [29], [34]) under the assumption that the elements on the diagonal of $\mathbf{D}$ are all equal. When this assumption is violated, this method can lead to large overestimation [35], which becomes even worse when the number of samples used to estimate $\mathbf{P_y}$ increases [35], [36]. In this paper, we do not constrain the diagonal elements of $\mathbf{D}$ to be equal and, thus, this method is not applicable here. Preliminary tests using the method in [25] with non-identical values on the diagonal of $\mathbf{D}$ demonstrated that indeed the overestimation error is quite high.

Another popular (but heuristically motivated) approach, which does not make the aforementioned limiting assumption, is the scree test [12], [24], which first sorts the eigenvalues of $\hat{\mathbf{P}}_\mathbf{y}$, or $\hat{\boldsymbol{\Phi}}$ in descending order and finds at which point the last smooth decrease of eigenvalues starts. This point is called scree point and the index of the eigenvalue before this point is the estimated number of sources, i.e., the scree point is $\hat{r}+1$. After the scree point the eigenvalues descend gradually forming a smoothly changing segment until the last (smallest) eigenvalue. Trying to determine the scree point based on the eigenvalues of $\hat{\mathbf{P}}_\mathbf{y}$ may be very challenging, especially when the diagonal elements of $\mathbf{D}$ are not all equal. We therefore first obtain an estimate of $\boldsymbol{\Phi}$ and then apply the scree test on this estimate.

To estimate $\boldsymbol{\Phi}$, one can use an EFA-based method. Several EFA methods estimate the matrix $\boldsymbol{\Phi}$ and the diagonal matrix $\mathbf{D}$ in (2), e.g., [19]–[21]. EFA assumes that there are only a few sources that explain the covariance of the observed random variables, i.e., that the matrix $\boldsymbol{\Phi}$ is low-rank (i.e., $M \gg r$). One popular EFA method is the minimum-rank factor analysis [19], [20] which is formulated as

$$\begin{aligned} \hat{\boldsymbol{\Phi}}, \hat{\mathbf{D}} = \underset{\boldsymbol{\Phi},\mathbf{D}}{\arg\min}\ & \mathrm{rank}\,(\boldsymbol{\Phi}) \\ \text{s.t. } & \hat{\mathbf{P}}_\mathbf{y} = \boldsymbol{\Phi} + \mathbf{D} \\ & \boldsymbol{\Phi} \succeq 0 \\ & \mathbf{D} = \mathrm{Diag}(d_1, d_2, \cdots, d_M) \succeq 0. \end{aligned} \tag{4}$$

The problem in (4) is a non-convex optimization problem that is hard to solve. A convex relaxation of the non-convex rank

[1]The same signal model can be used for narrowband signals as well.

operation is to replace it with the trace norm operation, leading to the suboptimal minimum trace factor analysis (MTFA) problem [19], [21]. The MRFA and MTFA formulations result in the same solution under certain conditions. One of these conditions is that the number of sensors should be sufficiently larger than the number of sources [21]. In this paper, we use the MTFA problem formulation to obtain $\hat{\mathbf{\Phi}}$. Using an estimate $\hat{\mathbf{\Phi}}$ obtained with an EFA method, one may trivially compute the number of sources as $\hat{r} = \text{rank}(\hat{\mathbf{\Phi}})$. Since $\hat{r} = \text{rank}(\hat{\mathbf{\Phi}})$ is based on the estimated $\hat{\mathbf{\Phi}}$, we will have estimation errors on $\hat{r}$ due to estimation errors in $\hat{\mathbf{\Phi}}$. To obtain a more robust estimate of $\hat{r}$, we propose in Sec. V a modified version of the scree test originating from [24]. In Sec. V we use the scree test applied to $\hat{\mathbf{\Phi}}$ with some adjustments to the sensor array context, to estimate the number of sources r.

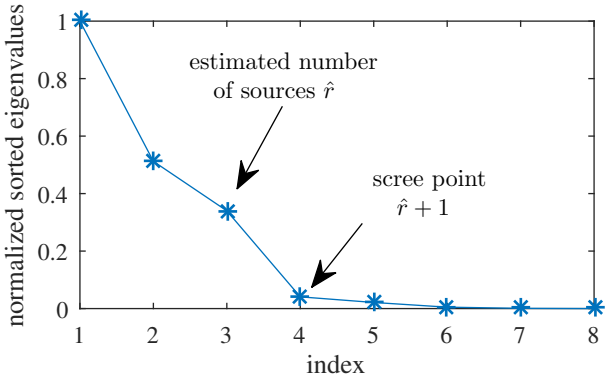

Fig. 1. Scree point and estimated number of sources based on scree-test.

IV. Confirmatory factor analysis

Unlike EFA, CFA [22], [23] estimates all parameters of the signal model in (2) (except for r), including $\mathbf{A}$ and $\mathbf{P}$. This results in an increased number of unknowns compared to EFA. In order to guarantee that the number of equations is greater than the number of unknowns and to avoid identifiability problems (see Sec. II), we need to fix some of the variables using linear equality constraints and share some of the variables over multiple time-frames [7]. Specifically, a common assumption used in sensor array applications is to assume that the source signals are uncorrelated to each other so that the matrix $\mathbf{P}$ is real and diagonal [5], [7]. Moreover, typically, $\mathbf{A}$ and $\mathbf{P}$ are estimated relative to a reference sensor which can be translated as constraining the row of $\mathbf{A}$ corresponding to the reference sensor to be equal to the all-ones vector [7]. The matrices $\mathbf{A}$ and/or $\mathbf{D}$ can also be shared over multiple time-frames within a *time-segment* [5], [7]. The CFA problem formulation used in this paper is given by [7]

$$\begin{aligned}\hat{\mathbf{A}}, \{\hat{\mathbf{P}}(t) : t \in \mathcal{T}\}, \hat{\mathbf{D}} = & \underset{\substack{\mathbf{A},\mathbf{D},\\ \{\mathbf{P}(t):t\in\mathcal{T}\}}}{\arg\min} \sum_{\forall \tau \in \mathcal{T}} F(\hat{\mathbf{P}}_{\mathbf{y}}(\tau), \mathbf{P}_{\mathbf{y}}(\tau)) \\ \text{s.t. } & \mathbf{P}_{\mathbf{y}}(t) = \mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{D}, \ \forall t \in \mathcal{T} \\ & \mathbf{P}(t) = \text{Diag}(p_1, p_2, \cdots, p_M) \succeq 0 \\ & \mathbf{D} = \text{Diag}(d_1, d_2, \cdots, d_M) \succeq 0, \\ & a_{\rho j} = 1, \ \forall j, \\ & -b_a \leq \Re\left(a_{ij}\right), \Im\left(a_{ij}\right) \leq b_a, \ \forall i, j, \\ & d_i \leq \eta_i \min_{\forall t \in \mathcal{T}} \left(\hat{p}_{\mathbf{y},ii}(t)\right), \ \forall i, \end{aligned} \tag{5}$$

with $\mathcal{T}$ the set of the time-frames in which the matrices $\mathbf{A}, \mathbf{D}$ are shared, $\hat{p}_{\mathbf{y},ii}$ the i-th diagonal element of $\hat{\mathbf{P}}_{\mathbf{y}}$, a_{ij} the (i,j)-th element of $\mathbf{A}$, $b_a = (Nc)/(f_s\lambda) + 1$, N the time-frame length, c the speed of sound, f_s the sampling frequency, λ the minimum possible distance between any source-sensor pair, ρ the index of the reference sensor, η_i a bias compensation parameter for d_i, and $F(\cdot)$ a discrepancy function measuring the error between the sample estimate $\hat{\mathbf{P}}_{\mathbf{y}}$ and the signal model in (2). In the current paper, we use the maximum likelihood discrepancy function [22], [23]. Note that the last constraint in (5) is less tight than the corresponding constraint proposed in [7] in which it was assumed that $d_i = d_j, \forall i, j$. Moreover, unlike the corresponding constraint proposed in [7], in this paper we also use the scaling η_i which avoids biased estimates of d_i. η_i can be estimated using the technique in [37]. The CFA problem is applied to each time-segment independently.

For $|\mathcal{T}| \geq 2$, the CFA problem in (5) can always provide uniquely identifiable solutions up to a permutation ambiguity in $\mathbf{A}$ if the ratio between the number of sensors and the number of sources is large enough [7]. Several approaches have been proposed to resolve the permutation problem (see e.g., [8], [38]). In this paper we do not focus on this problem and we assume that we know the perfect permutation matrix.

Unlike the CFA method proposed in [7], in which a constant (a priori known) $r = r_{\text{max}}$ is used for all time-frequency bins, in this paper, to reduce the computational complexity, we estimate the number of sources using the method that we will present in Sec. V over each frequency bin of an entire time-segment. Each time segment consists of multiple time-frames, which share the same $\mathbf{A}$ and $\mathbf{D}$. Note that we did not estimate r for every time-frame within the same time-segment, since CFA shares the same $\mathbf{A}$ and $\mathbf{D}$ and assumes a common signal model for all time-frames within a time-segment. Hence, in this paper we adapt r per time-segment and frequency bin.

V. Implementation Details on Source Counting

Having an estimate of $\hat{\mathbf{\Phi}}$ using the EFA method in Sec. III, the scree test [24] sorts the eigenvalues of $\hat{\mathbf{\Phi}}$ in descending order $(\lambda_1, \lambda_2, \cdots, \lambda_M$, where $\lambda_i \geq \lambda_j \geq 0, i < j)$ and finds the scree point which is equal to $\hat{r}+1$. The scree point is the first point of the last segment (which can be well approximated by a line) which is formed after the last jump of eigenvalues downwards (see the example in Fig. 1).

As the second derivative of an affine function is zero, we first find where the second derivative becomes approximately zero and then we determine the scree point. Specifically, we obtain the scree point using the following steps. We first normalize all eigenvalues with the maximum eigenvalue λ_1 leading to the normalized eigenvalues, $\tilde{\lambda}_1, \tilde{\lambda}_2, \cdots, \tilde{\lambda}_M$, where $\tilde{\lambda}_i = \lambda_i/\lambda_1$. After that, we compute the set of all differences

(which can be seen as an approximation of the first derivative) between consecutive normalized eigenvalues, i.e.,

$$\mathcal{D} = \{\bar{d}_i : \ \bar{d}_i = \tilde{\lambda}_i - \tilde{\lambda}_{i+1}, \ 1 \leq i \leq M-1\}. \quad (6)$$

Then we compute the set of all absolute-valued differences between the consecutive differences in the set $\mathcal{D}$, i.e.,

$$\mathcal{G} = \{\Delta\bar{d}_i : \ \Delta\bar{d}_i = |\bar{d}_i - \bar{d}_{i+1}|, \ 1 \leq i \leq M-2\}. \quad (7)$$

Note that $\mathcal{G}$ basically approximates the (absolute value of the) second derivative. Selecting all indices i where $\Delta\bar{d}_i \geq \epsilon$ (with ϵ a small positive number), we find all eigenvalues that are not part of a line segment. That is, we calculate the set $\mathcal{K}$,

$$\mathcal{K} = \{i : \ \Delta\bar{d}_i \in \mathcal{G}, \ \Delta\bar{d}_i \geq \epsilon\}. \quad (8)$$

The estimate $\hat{r}$ is then given by

$$\hat{r} = \max(\mathcal{K}). \quad (9)$$

Some remarks are in place here. For sensor arrays with a small aperture, the low frequency components of the signals will not have very distinguishable phase differences among the sensors and the prominent eigenvalues can be much less than r leading to underestimation of r using the scree test. When it is difficult to have a good estimate of r, we believe that is better to have an overestimate rather than an underestimate for the CFA methods (see Sec. IV). When we have an underestimate, the signal model is not accurate, while when we have an overestimate it still is, however, at the expense of additional (unnecessary) computational complexity. Hence, for the $\phi\%$ lowest frequencies we therefore set $\hat{r} = \hat{r}_{\max}$, where $\hat{r}_{\max}$ is an estimate of the maximum number of sources among all frequency bins. The value of ϕ depends on the size of the array aperture. Finally, to reduce outliers we set $\hat{r} = \hat{r}_{\max}$ in those frequency bins where $\hat{r} > \hat{r}_{\max}$.

A. Proposed Estimation of r_{max}

To obtain a robust estimate $\hat{r}_{\max}$, we create a histogram of the percentages $(f_{r=1}, f_{r=2}, \cdots, f_{r=M-2})$, where $f_{r=i}$ is the percentage of the frequency bins that have i sources. We then find a robust max number of sources as follows

$$\hat{r}_{\max} = \max_i \{i : f_{r=i} > \lambda\}, \quad (10)$$

where λ is a small percentage threshold (e.g., 10%). This avoids to find an overestimate of $\hat{r}_{\max}$.

VI. Experiments

In this section, we examine the performance of the factor analysis methods discussed in Secs. III and IV in the context of source separation in an anechoic acoustic environment. Similar to [7], the source separation was implemented by using a separate multi-channel Wiener filter for each source using the appropriate estimated parameters from the factor analysis methods presented in this paper. The parameters of the proposed method in Sec. V are selected as $\epsilon = 0.07$, $\phi = 25\%$ and $\lambda = 10\%$. Moreover, the bias compensation parameter η_i was selected to be always 1 for simplicity.

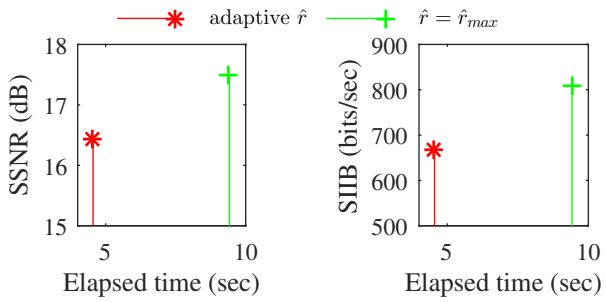

Fig. 2. Evaluation of the source separation performance in terms of SSNR gain (in dB) and SIIB gain (in bits/sec), using a constant $\hat{r} = \hat{r}_{\max}$, and an adaptive $\hat{r}$ per time-frequency bin.

The microphone array is uniform circular with radius 5 cm consisting of $M = 8$ microphones. The self-noise of each microphone was simulated as white Gaussian noise with a different power randomly chosen for each microphone. This simulates microphones with different characteristics. Specifically, the SSNRs of the mixture of the recordings at each microphone with respect to the corresponding microphone self noise was $[26.7, 24.6, 17.6, 27.6, 26, 29.7, 15.1, 18.1]$ dB. There are three acoustic sources ($r = 3$), a female talker at 0^{o} and two male talkers at 70^{o} and -110^{o}. The sources are 2 meters from the center of the microphone array. The sampling frequency is 16 kHz, the time-frame length is 0.125 sec and the time-segment consists of $|\mathcal{T}| = 16$ time-frames and, thus, its length is 2 sec. The total duration of the recorded mixture of signals is 28 sec.

Note that r will not always be equal to 3 at each time-frequency bin (because sources are not always simultaneously active at all time-frequency bins), but its true maximum value will be $r_{\max} = 3$. We examine the difference in performance of the method in Sec. IV using a constant $\hat{r} = \hat{r}_{\max}$ per time-segment (where $\hat{r}_{\max}$ is estimated as described in V-A) as proposed in [7], and with the adaptive r as proposed in this paper. The reconstruction accuracy and predicted intelligibility are measured in terms of the SSNR gain [30] and speech intelligibility in bits (SIIB) gain [39]. We also measure the elapsed time for our MATLAB implementations. Note that the elapsed time was measured as the average time in seconds that it takes to estimate all parameters including $r, r_{\max}$ per frequency-bin and time-segment. Fig. 2 shows the trade-off between SSNR gain and elapsed time and SIIB gain and elapsed time. We also computed the mean absolute error of $\hat{r}$ compared to the true r over all time-segments and frequency bins, i.e.,

$$\text{Error} = \frac{1}{NB} \sum_{i=1}^{B} \sum_{j=1}^{N} |\hat{r}_{ij} - r_{ij}|, \quad (11)$$

where B is the total number of time-segments and N is the number of frequency bins. The true r was obtained by checking the activity of each recorded source separately by checking if the source's PSD is above a certain small threshold. If the outcome of this check is positive, this source is active,

otherwise not. The error was found to be 0.5399. The estimate $\hat{r}_{\max}$ was correct (i.e., $\hat{r}_{\max} = 3$) in 71.43% of the time-segments, while $\hat{r}_{\max} = 4$ in 28.57% of the time-segments.

It is clear from Fig. 2 that the separation of the sources is slightly better with the constant $\hat{r} = \hat{r}_{\max}$ compared to the adaptive $\hat{r}$, because as explained in Sec. V, overestimation errors do not negatively influence the signal model compared to the underestimation errors. This slight advantage in SSNR and predicted intelligibility for the constant $\hat{r} = \hat{r}_{\max}$ method comes with the cost of a much larger computational load compared to the adaptive $\hat{r}$ method.

VII. Conclusions

We reviewed factor analysis in the context of parameter estimation in a commonly used signal model for sensor arrays. We discussed how exploratory and confirmatory factor analysis can be used to estimate all the parameters of the signal model. That is, the number of sources, PSDs of the sources, relative transfer functions of the sources and PSDs of the sensor-self noises. We also proposed a robust source counting method using an EFA method and we used this in a CFA method to estimate all the remaining signal model parameters. As a result, the complexity was reduced significantly, while the separation of the sources was only slightly worse compared to the case where a constant number of sources is used.

References

[1] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, July 1996.

[2] M. Brandstein and D. Ward (Eds.), *Microphone arrays: signal processing techniques and applications*. Springer, 2001.

[3] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. John Wiley & Sons, 2004.

[4] H. Ye and D. DeGroat, "Maximum likelihood doa estimation and asymptotic cramér-rao bounds for additive unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 938–949, 1995.

[5] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Audio, Speech, Language Process.*, vol. 8, no. 3, pp. 320–327, 2000.

[6] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[7] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Robust joint estimation of multi-microphone signal model parameters," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 7, pp. 1136–1150, 2019.

[8] D. Nion, K. Mokios, N. D. Sidiropoulos, and A. Potamianos, "Batch and adaptive parafac-based blind separation of convolutive speech mixtures," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1193–1207, 2010.

[9] S. Markovich and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2015, pp. 544–548.

[10] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood psd estimation for speech enhancement in reverberation and noise," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1599–1612, 2016.

[11] R. Varzandeh, M. Taseska, and E. A. P. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *Int. Workshop Hands-Free Speech Commun.*, 2017, pp. 11–15.

[12] S. A. Mulaik, *Foundations of factor analysis*. CRC press, 2009.

[13] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[14] H. Hotelling, "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[15] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method*. London Butterworths, 1963.

[16] C. Spearman, ""general intelligence," objectively determined and measured," *The American Journal of Psychology*, vol. 15, no. 2, pp. 201–292, 1904.

[17] A. M. Sardarabadi and A. J. van der Veen, "Complex factor analysis and extensions," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 954–967.

[18] D. Bertsimas, M. S. Copenhaver, and R. Mazumder, "Certifiably optimal low rank factor analysis," *Journal of Machine Learning Research*, vol. 18, no. 29, pp. 1–53, 2017.

[19] W. Ledermann, "On a problem concerning matrices with variable diagonal elements," *Proc. Roy. Soc. Endinburgh*, vol. 60, pp. 1–17, 1940.

[20] J. M. Ten Berge and H. A. Kiers, "A numerical approach to the approximate and the exact minimum rank of a covariance matrix," *Psychometrika*, vol. 56, no. 2, pp. 309–315, 1991.

[21] G. Della Riccia and A. Shapiro, "Minimum rank and minimum trace of covariance matrices," *Psychometrika*, vol. 47, no. 4, pp. 443–448, 1982.

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

[23] ——, "Simultaneous factor analysis in several populations," *Psychometrika*, vol. 36, no. 4, pp. 409–426, 1971.

[24] R. B. Cattell, "The scree test for the number of factors," *Multivariate behavioral research*, vol. 1, no. 2, pp. 245–276, 1966.

[25] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.

[26] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 10, pp. 2193–2206, 2013.

[27] A. Alexandridis and A. Mouchtaris, "Multiple sound source location estimation and counting in a wireless acoustic sensor network," in *IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA)*. IEEE, 2015, pp. 1–5.

[28] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[29] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[30] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2007.

[31] S. M. Kay, "Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1)," *PTR Prentice-Hall, Englewood Cliffs*, 1993.

[32] G. Connor and R. A. Korajczyk, "A test for the number of factors in an approximate factor model," *the Journal of Finance*, vol. 48, no. 4, pp. 1263–1291, 1993.

[33] J. Bai and S. Ng, "Determining the number of factors in approximate factor models," *Econometrica*, vol. 70, no. 1, pp. 191–221, 2002.

[34] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[35] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, 2001.

[36] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, 1995.

[37] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

[38] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 832–844, 2005.

[39] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, "An instrumental intelligibility metric based on information theory," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 115–119, Jan. 2018.