

Controlled Complete ARMA Independent Process Analysis

Zoltán Szabó and András Lőrincz

Abstract—In this paper we address the controlled complete AutoRegressive Moving Average Independent Process Analysis (ARMAX-IPA; X-exogenous input or control) problem, which is a generalization of the Blind SubSpace Deconvolution (BSSD) task. Compared to our previous work that dealt with the undercomplete situation, (i) here we extend the theory to complete systems, (ii) allow an autoregressive part to be present, (iii) and include exogenous control. We investigate the case when the observed signal is a linear mixture of independent multidimensional ARMA processes that can be controlled. Our objective is to estimate the ARMA processes, their driving noises as well as the mixing. We aim efficient estimation by choosing suitable control values. For the optimal choice of the control we adapt the D-optimality principle, also known as the ‘InfoMax method’. We solve the problem by reducing it to a fully observable D-optimal ARX task and Independent Subspace Analysis (ISA) that we can solve. Numerical examples illustrate the efficiency of the proposed method.

I. INTRODUCTION

Recently, research on Independent Component Analysis (ICA) [1], [2] and its extensions has gained much attention. One can think of ICA as a cocktail-party problem, where there are D microphones, D one-dimensional sound sources, and the task is to estimate the original sources from the observed mixed signals. For a recent review about ICA see [3], [4], [5].

Applications, where *only certain groups* of the sources are independent may be highly relevant in practice. For example, at the cocktail-party, groups of people or groups of musicians may form *independent source groups*. This task is called Independent Subspace Analysis (ISA¹) [6]. The large number of different ISA algorithms [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31] shows the importance of this field. Successful applications of ISA involve: (i) the processing of EEG-fMRI (ElectroEncephaloGraphy, functional Magnetic Resonance Imaging) data [7], (ii) gene analysis [32], [33], [34], (iii) face view recognition [35], [36], and (iv) ECG (ElectroCardioGraphy) analysis [12], [15].

Another extension of the original ICA task is the Blind Source Deconvolution (BSD) problem. Such a problem emerges, for example, at a cocktail-party being held in an *echoic* room. Several BSD algorithms have been developed over the last decades, for a review see [37]. BSD shows potentials in the following areas: (i) remote sensing applications; passive radar/sonar processing [38], [39], (ii) image-

deblurring, image restoration [40], (iii) speech enhancement using microphone arrays, acoustics [41], [42], [43], [44], (iv) multi-antenna wireless communications, sensor networks [45], [46], (v) biomedical signal—EEG, ECG, MEG (MagnetoEncephaloGraphy), fMRI—analysis [47], [48], [49], (vi) optics [50], and (vii) seismic exploration [51].

The simultaneous assumption of the two extensions, that is, *ISA combined with BSD*, seems to be a more realistic model than either of the two models alone and has recently been introduced in the literature under the name of Blind SubSpace Deconvolution (BSSD). For example, at the cocktail-party, groups of people or groups of musicians may form *independent source groups and echoes* could be present. It has been shown that the undercomplete case of the BSSD problem can be reduced to ISA by means of (i) temporal concatenation [20] or (ii) Linear Predictive Approximation (LPA) [52]. Using the first approach the associated ISA problem can easily become ‘high dimensional’, this dimensionality problem can be circumvented by applying the LPA based reduction scheme. The LPA method has been extended to Integrated ARMA (ARIMA-IPA) processes [53], but the framework dealt with the undercomplete case and the estimation of the driving noises only.

It has been shown in a recent work [54] that the parameters and the driving noise of controlled dynamical systems (ARX models: AutoRegressive process with eXogenous inputs) can be efficiently estimated by means of D-optimality principles. This theory, which allows control variables, has been formulated only for the *fully observable* case. By contrast, the ‘ICA’ problem family can model hidden independent variables, but can *not* account for *control*.

We unify and generalize these directions: we treat D-optimal estimation of controlled independent multidimensional hidden dynamical systems, ARMAX processes. Beyond this extension, we estimate the independent multidimensional driving noises and the mixing process, too. Our method may offer important extension possibilities for ICA applications. Such potential applications motivate our work. Here, we present the necessary basic theoretical steps. Future applications, where interaction (‘X’) with the environment is present, may include for example, human-computer interaction serving the user.

The paper is structured as follows: In Section II we review the D-optimal identification of fully observed ARX models. Section III defines the problem domain, the Controlled ARMA Independent Process Analysis (ARMAX-IPA) task. In Section IV we detail our method. Section V contains numerical illustrations and conclusions are drawn in Section VI.

Department of Information Systems, Eötvös Loránd University, Pázmány P. sétány 1/C, Budapest H-1117, Hungary (email: szzoli@cs.elte.hu, andras.lorincz@elte.hu). ©2009 IEEE. International Joint Conference on Neural Networks (IJCNN), pages 3038-3045, Atlanta, Georgia, USA, June 14-19, 2009. <http://dx.doi.org/10.1109/IJCNN.2009.5178797>.

¹ISA is also called Multidimensional Independent Component Analysis, subspace ICA and group ICA in the literature.

II. D-OPTIMAL IDENTIFICATION OF ARX MODELS

We sketch the basic thoughts that lead to D-optimal identification of ARX models. The dynamical system is fully observed and evolves according to the ARX equation

$$\mathbf{s}_t = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}_{t-i} + \mathbf{e}_t + \sum_{k=0}^{L_u-1} \mathbf{B}_k \mathbf{u}_{t-k}, \quad (1)$$

where

- $\mathbf{s} \in \mathbb{R}^{D_s}$, $\mathbf{e} \in \mathbb{R}^{D_e}$ ($D_s = D_e$) represent the state of the system and the noise, respectively
- $\mathbf{u} \in \mathbb{R}^{D_u}$ represents the control variables, and
- polynomial matrix

$$\mathbf{F}[z] = \mathbf{I} - \sum_{i=1}^{L_s} \mathbf{F}_i z^i \quad (2)$$

(given by matrices $\mathbf{F}_i \in \mathbb{R}^{D_s \times D_s}$ and identity matrix \mathbf{I}) is *invertible*, that is

$$\det(\mathbf{F}[z]) \neq 0, \quad (3)$$

for all $z \in \mathbb{C}$, $|z| \leq 1$.

Our task is the efficient estimation of

- 1) the parameters $\Theta = [\Theta_{\text{dynamics}}, \Theta_{\text{noise}}]$, that is $\Theta_{\text{dynamics}} = [\mathbf{F}_1, \dots, \mathbf{F}_{L_s}, \mathbf{B}_0, \dots, \mathbf{B}_{L_u-1}]$ that determine the dynamics and noise parameters Θ_{noise} ,
- 2) the noise \mathbf{e} that drives the process

by the ‘optimal choice’ of control values \mathbf{u} . Formally, D-optimality aims to maximize one of the two objectives

$$J_{\text{par}}(\mathbf{u}_t) := I(\Theta, \mathbf{s}_t | \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots, \mathbf{u}_t, \mathbf{u}_{t-1}, \dots),$$

$$J_{\text{noise}}(\mathbf{u}_t) := I(\mathbf{e}_t, \mathbf{s}_t | \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots, \mathbf{u}_t, \mathbf{u}_{t-1}, \dots)$$

for $\mathbf{u}_t \in U \subseteq \mathbb{R}^{D_u}$. In other words, we choose control value \mathbf{u} from the achievable domain U such that it maximizes the mutual information between the next observation and the parameters (or the driving noise) of the system. It can be shown [54], that if

- Θ has matrix Gaussian,
- \mathbf{e} has Gaussian, and
- the covariance matrix of \mathbf{e} has inverted Wishart distributions,

then in the Bayesian setting

- maximization of the J objectives can be reduced to the solution of a quadratic programming task,
- priors of Θ and \mathbf{e} remain in their supposed distribution family and undergo simple updating.

The considerations allow for control, but assume full observability about the state variables. Now, we extend the method to hidden variables, to ARMA processes in the ARMAX-IPA model of the next section.

III. THE COMPLETE ARMAX-IPA MODEL

Here, we define the ARMAX-IPA (AutoRegressive Moving Average Independent Process Analysis with eXogenous input) task. Assume that we have M multidimensional independent noise processes that drive multidimensional ARMA processes that we can influence (control). Suppose also that only their

$$\mathbf{s}_t = \sum_{i=1}^{L_s} \mathbf{F}_i \mathbf{s}_{t-i} + \mathbf{e}_t + \sum_{j=1}^{L_e} \mathbf{H}_j \mathbf{e}_{t-j} + \sum_{k=0}^{L_u-1} \mathbf{B}_k \mathbf{u}_{t-k}, \quad (4)$$

$$\mathbf{x}_t = \mathbf{A} \mathbf{s}_t \quad (5)$$

mixture is available for observation², where

- $\mathbf{x}_t \in \mathbb{R}^{D_x}$, $\mathbf{s}_t = [\mathbf{s}_t^1; \dots; \mathbf{s}_t^M] \in \mathbb{R}^{D_s}$, $\mathbf{e}_t = [\mathbf{e}_t^1; \dots; \mathbf{e}_t^M] \in \mathbb{R}^{D_e}$ ($D_s = D_e$) represent the observation, the state of the system and the noise, respectively,
- $\mathbf{u}_t \in \mathbb{R}^{D_u}$ stands for the control variables at the t^{th} time instant,
- $\mathbf{A} \in \mathbb{R}^{D_x \times D_s}$ is the mixing matrix,
- $\mathbf{s}_t^m, \mathbf{e}_t^m \in \mathbb{R}^{d_m}$ ($m = 1, \dots, M$).

Denoting the time-shift operation by z , one may write (4)-(5) compactly as

$$\mathbf{F}[z] \mathbf{s} = \mathbf{H}[z] \mathbf{e} + \mathbf{B}[z] \mathbf{u}, \quad (6)$$

$$\mathbf{x} = \mathbf{A} \mathbf{s}, \quad (7)$$

using polynomial matrices

$$\mathbf{F}[z] = \mathbf{I} - \sum_{i=1}^{L_s} \mathbf{F}_i z^i \in \mathbb{R}[z]^{D_s \times D_s}, \quad (8)$$

$$\mathbf{H}[z] = \mathbf{I} + \sum_{j=1}^{L_e} \mathbf{H}_j z^j \in \mathbb{R}[z]^{D_s \times D_e}, \quad (9)$$

$$\mathbf{B}[z] = \sum_{k=0}^{L_u-1} \mathbf{B}_k z^k \in \mathbb{R}[z]^{D_s \times D_u}. \quad (10)$$

Our assumptions are the following:

- The problem is complete: $D := D_x = D_s = D_e$.
- Polynomial matrices $\mathbf{F}[z]$, $\mathbf{H}[z]$, and \mathbf{A} are invertible.
- Driving noises \mathbf{e}^m of processes \mathbf{s}^m are independent like in the ISA task and fulfill the ISA assumptions. In other words, \mathbf{e}^m s are
 - independent: $I(\mathbf{e}^1, \dots, \mathbf{e}^M) = 0$, where I denotes the mutual information,
 - i.i.d. (independent identically distributed) in t ,
 - and there is at most one Gaussian among the \mathbf{e}^m s.

Note: there is no block-diagonal restriction on polynomial matrices $\mathbf{F}[z]$ and $\mathbf{H}[z]$.

Our task is to estimate the unknown mixing matrix \mathbf{A} , hidden processes \mathbf{s}^m and their driving noises \mathbf{e}^m by means of observations \mathbf{x} only.

²Here, L_s, L_e, L_u denote the number of $\mathbf{F}_i, \mathbf{H}_j, \mathbf{B}_k$ matrices in the respective sums.

IV. METHOD

Below, we present our solution for the ARMAX-IPA task. According to our assumptions $\mathbf{H}[z]$ can be inverted, so we multiply (6) by $(\mathbf{H}[z])^{-1}$ from the left, use the invertibility of matrix \mathbf{A} and substitute relation

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \quad (11)$$

that follows from (7) and get

$$(\mathbf{H}[z])^{-1}\mathbf{F}[z]\mathbf{A}^{-1}\mathbf{x} = \mathbf{e} + (\mathbf{H}[z])^{-1}\mathbf{B}[z]\mathbf{u}. \quad (12)$$

Now, multiplying this equation by \mathbf{A} from the left we have

$$\mathbf{A}(\mathbf{H}[z])^{-1}\mathbf{F}[z]\mathbf{A}^{-1}\mathbf{x} = \mathbf{A}\mathbf{e} + \mathbf{A}(\mathbf{H}[z])^{-1}\mathbf{B}[z]\mathbf{u}. \quad (13)$$

In (13), the main coefficient of the polynomial matrix on the left hand side is \mathbf{I} giving rise to an AR(∞) form. Because $\mathbf{A}\mathbf{e}$ can be considered as an approximately Gaussian variable according to the d-dependent central limit theorem [55], we can apply the D-optimal ARX approximation for (13) in order to estimate noise $\mathbf{A}\mathbf{e}$. The result can be seen as the observation of an ISA problem because the $\mathbf{e}^m \in \mathbb{R}^{d_m}$ components of \mathbf{e} are independent. ISA techniques can be used to identify \mathbf{A} and \mathbf{e}^m . Our estimation for \mathbf{s} is $\hat{\mathbf{s}} = \hat{\mathbf{A}}^{-1}\mathbf{x}$.

It can be shown [56] that AR estimation of order

$$p = o(T^{\frac{1}{3}}) \xrightarrow{T \rightarrow \infty} \infty \quad (14)$$

for the inverse of polynomial matrix $\mathbf{H}[z]$ gives rise to an asymptotically consistent estimation. Thus, taking into account (13), we shall apply ARX estimations with the following orders

$$L'_s = p + L_s, \quad (15)$$

$$L'_u = p + L_u. \quad (16)$$

For the particular choice of p , see Section V-C.

Note:

- 1) In the above described complete ARMAX-IPA technique, the D-optimal ARX procedure is an online estimation for the innovation $\boldsymbol{\varepsilon} = \mathbf{A}\mathbf{e}$, the input of the ISA method, for fixed p . Online ISA method on $\boldsymbol{\varepsilon}$ would enable online estimation of the inverse of matrix \mathbf{A} and then the estimation of \mathbf{s} and \mathbf{e} . In the absence of an efficient online ISA estimation, the ISA step was executed in batch mode. For one-dimensional hidden sources ($d_m = 1, \forall m$) efficient online ICA methods, e.g., [57], could be used here.
- 2) The pseudocode of our method can be found in Table I.
- 3) In the absence of control ($L_u = -1$) one can apply a simple AR fit instead of the ARX estimation. Efficient AR fit methods can be found in [58], [59]. These methods use least squares estimations for AR fit, so recursive online estimations are feasible here.

Table I: Pseudocode of the ARMAX-IPA algorithm.

<p>Input of the algorithm AR order: L'_s control order: L'_u observation: $\{\mathbf{x}_t\}_{t=1,\dots,T}$</p> <p>Optimization D-optimal ARX estimation: for $t = 1, \dots, T$ Using \mathbf{x}_t update D-optimally the distribution of $\hat{\boldsymbol{\Theta}}_t = [\hat{\mathbf{F}}_{1,t}, \dots, \hat{\mathbf{F}}_{L'_s,t}, \hat{\mathbf{B}}_{0,t}, \dots, \hat{\mathbf{B}}_{L'_u-1,t}]$ Estimate the innovation process $\boldsymbol{\varepsilon}_t = \mathbf{A}\mathbf{e}_t$: $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{x}_t - (\sum_{i=1}^{L'_s} \hat{\mathbf{F}}_{i,t}\mathbf{x}_{t-i} + \sum_{k=0}^{L'_u-1} \hat{\mathbf{B}}_{k,t}\mathbf{u}_{t-k})$ end</p> <p>ISA estimation: on $\{\hat{\boldsymbol{\varepsilon}}_t\}_{t=1,\dots,T} \Rightarrow$ demixing matrix: $\hat{\mathbf{W}}_{\text{ISA}}$</p> <p>Estimation Estimated driving noise: $\hat{\mathbf{e}} = \hat{\mathbf{W}}_{\text{ISA}}\hat{\boldsymbol{\varepsilon}}$ Estimated source: $\hat{\mathbf{s}} = \hat{\mathbf{W}}_{\text{ISA}}\hat{\mathbf{x}}$</p>
--

V. ILLUSTRATIONS

Here, we illustrate the efficiency of the proposed complete ARMAX-IPA estimation technique. Test cases are introduced in Section V-A. To evaluate the solutions we use a performance measure given in Section V-B. Numerical results are presented in Section V-C.

A. Databases

We define three databases (e) to study our identification algorithm. The databases are depicted in Fig. 1.

1) *ABC database*: In the *ABC* database, hidden sources \mathbf{e}^m were uniform distributions on 2-dimensional images ($d_m = 2$) of the English alphabet. The number of components was $M = 4$, and thus the dimension of the source was $D = 8$. For illustration, see Fig. 1(a).

2) *Tale database*: The *tale* test has 2-dimensional source components generated from drawings of fairy tale characters ($d_m = 2$).³ Sources \mathbf{e}^m were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities. In other words, 2-dimensional images of tale characters were considered as density functions. $M = 4$ was chosen, thus the dimension of the hidden source was $D = 8$. For illustration, see Fig. 1(b).

3) *3D-geom database*: In the *3D-geom* test \mathbf{e}^m s were random variables uniformly distributed on 3-dimensional geometric forms ($d_m = 3$). We chose 3 different components ($M = 3$) and, as a result, the dimension of the hidden source is $D = 9$. For illustration, see Fig. 1(c).

B. Performance Measure, the Amari-index

Recovery of source components \mathbf{s}^m (and \mathbf{e}^m) are subject to the ambiguities of the ISA task. Namely, components of equal dimension can be recovered up to permutation and invertible transformation within the subspaces [60]. Let us suppose that the hidden components are d-dimensional ($d = d_m$). Then, in the ideal case, the product of the estimated ISA demixing matrix $\hat{\mathbf{W}}_{\text{ISA}}$ and mixing matrix \mathbf{A} , that is $\mathbf{G} := \hat{\mathbf{W}}_{\text{ISA}}\mathbf{A} \in \mathbb{R}^{D \times D}$ is a *block-permutation matrix* with $d \times d$ sized blocks. This block-permutation

³See <http://www.smileyworld.com>.

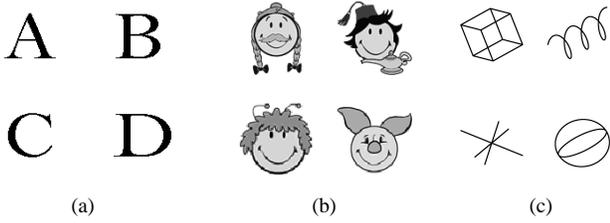


Figure 1: Illustration of the *ABC* (a), *tale* (b), and *3D-geom* (c) datasets.

structure can be measured by the Amari-index. Namely, let matrix $\mathbf{G} \in \mathbb{R}^{D \times D}$ be decomposed into $d \times d$ blocks: $\mathbf{G} = [\mathbf{G}^{ij}]_{i,j=1,\dots,M}$. Let g^{ij} denote the sum of the absolute values of the elements of matrix $\mathbf{G}^{ij} \in \mathbb{R}^{d \times d}$. We normalized the ISA adapted version [15], [16] of the Amari-error [61] into interval $[0, 1]$ [62]:

$$r(\mathbf{G}) := \frac{1}{2M(M-1)} \left[\sum_{i=1}^M \left(\frac{\sum_{j=1}^M g^{ij}}{\max_j g^{ij}} - 1 \right) + \sum_{j=1}^M \left(\frac{\sum_{i=1}^M g^{ij}}{\max_i g^{ij}} - 1 \right) \right]. \quad (17)$$

We refer to the normalized Amari-error as the *Amari-index*. One can see that $0 \leq r(\mathbf{G}) \leq 1$ for any matrix \mathbf{G} , and $r(\mathbf{G}) = 0$ if and only if \mathbf{G} is a block-permutation matrix with $d \times d$ sized blocks.

C. Simulations

Results on databases *ABC*, *tale* and *3D-geom* are provided here. We focused on the following questions:

- 1) The error of the source estimation as a function of the sample size.
- 2) It is expected that if the roots of $\mathbf{F}[z]$ and $\mathbf{H}[z]$ are close to the unit circle then our estimation will deteriorate. We investigated this by generating the polynomial matrix $\mathbf{F}[z]$ and $\mathbf{H}[z]$ as follows:

$$\mathbf{F}[z] = \prod_{i=0}^{L_s} (\mathbf{I} - \lambda_s \mathbf{O}_i z) \quad (|\lambda_s| < 1), \quad (18)$$

$$\mathbf{H}[z] = \prod_{j=0}^{L_e} (\mathbf{I} - \lambda_e \mathbf{U}_j z) \quad (|\lambda_e| < 1), \quad (19)$$

where matrices \mathbf{O}_i and \mathbf{U}_j were chosen uniformly (according to the Haar-measure) from the orthogonal group, $\lambda_s, \lambda_e \in \mathbb{R}$ and the $\lambda_s \rightarrow 1, \lambda_e \rightarrow 1$ limits were studied.

The Amari-index was used to measure the performance of the proposed complete ARMAX-IPA method. For each individual parameter (T, λ_s, λ_e), 20 random runs were averaged. ‘Random run’ means random choice of quantities $\mathbf{F}[z], \mathbf{H}[z], \mathbf{B}[z], \mathbf{A}$ and \mathbf{e} . In our simulations:

- mixing matrix \mathbf{A} was uniformly distributed on the orthogonal group,

- distributions of coordinates of polynomial matrix $\mathbf{B}[z]$ were independent and normal,
- control \mathbf{u} was limited to a hypercube

$$U := \{\mathbf{u} \in \mathbb{R}^{D_u} : \max_{i \in \{1, \dots, D_u\}} |u_i| \leq \delta_u\}, \quad (20)$$

with upper limit δ_u equal to 0.1,

- sample number T varied between 1,000 and 20,000,
- dimension of the control was equal to the the dimension of \mathbf{s} ($D_u = D$),
- invertibility parameters λ_s and λ_e of polynomial matrices $\mathbf{F}[z]$ and $\mathbf{H}[z]$, respectively were chosen independently from the set $\{0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95\}$,
- order p of the AR approximation (see Section IV) was⁴

$$p = \lfloor T^{\frac{1}{3} - \frac{1}{1000}} \rfloor, \quad (21)$$

thus the orders in the D-optimal ARX estimation were (see (15)-(16))

$$L'_s = \lfloor T^{\frac{1}{3} - \frac{1}{1000}} \rfloor + L_s, \quad (22)$$

$$L'_u = \lfloor T^{\frac{1}{3} - \frac{1}{1000}} \rfloor + L_u, \quad (23)$$

- the ISA subtask on the estimated innovation, that is on the estimation of $\mathbf{A}\mathbf{e}$ was carried out by the joint f-decorrelation method [18].

We present our results for invertibility parameters λ_s and λ_e for matrices $\mathbf{F}[z]$ and $\mathbf{H}[z]$, respectively for maximal sample number $T = 20,000$. The average quality of the estimations is shown in Fig. 2, Fig. 3, and Fig. 4 for the *ABC*, the *tale*, and the *3D-geom* databases, respectively. The average quality is depicted by filled, 30 level contour plots for the studied parameter region between 0.4 – 0.95. One may conclude from these figures that our method

- provides reliable estimates for λ_s and λ_e even if they are close to 1,
- is more sensitive for parameter λ_s , and
- is robust up to 0.9 in both parameters.

Estimation curves for the full $1,000 \leq T \leq 20,000$ sample interval and for $(\lambda_s, \lambda_e) = (0.4, 0.4), (0.7, 0.7), (0.85, 0.85), (0.9, 0.85), (0.9, 0.9), (0.9, 0.95)$ are provided in Fig. 8, Fig. 9, and Fig. 10 for the *ABC*, the *tale*, and the *3D-geom* databases, respectively. Estimation errors for $\lambda_s, \lambda_e \leq 0.85$ approximate a power law $r(T) \propto T^{-c}$ ($c > 0$) –manifested by straight lines on log-log scale,– and this characteristics can be observed for $\lambda_s = 0.9$, too. Precise values (mean \pm standard deviation) of the Amari-index are provided for sample number $T = 20,000$ in Table II, Table III and Table IV, for the *ABC*, the *tale*, and the *3D-geom* databases, respectively. These tables demonstrate that

- estimations for $\lambda_s, \lambda_e \leq 0.85$ are highly precise (1–3%) with small standard deviations,
- estimation errors start to increase around $\lambda_s = 0.9$: for $\lambda_e = 0.85, 0.9, 0.95$ values estimation errors are about

⁴We found this choice of p reliable in our numerical experiments within the studied parameter domain.

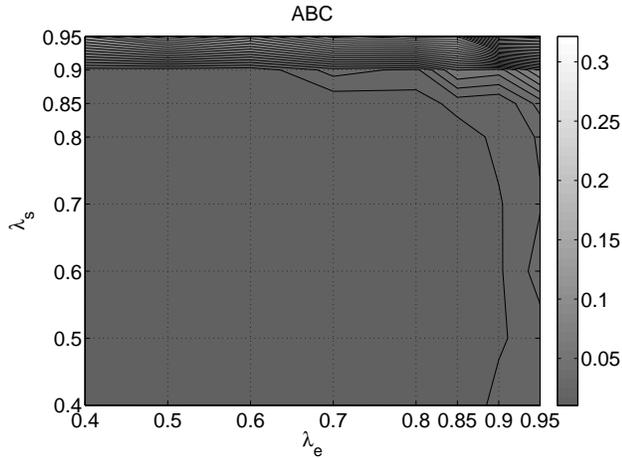


Figure 2: Contour plot of the Amari-index as a function of the λ_s, λ_e invertibility parameters on the *ABC* database. Number of samples: $T = 20,000$. For error curves and numerical values, see Fig. 8 and Table II, respectively.

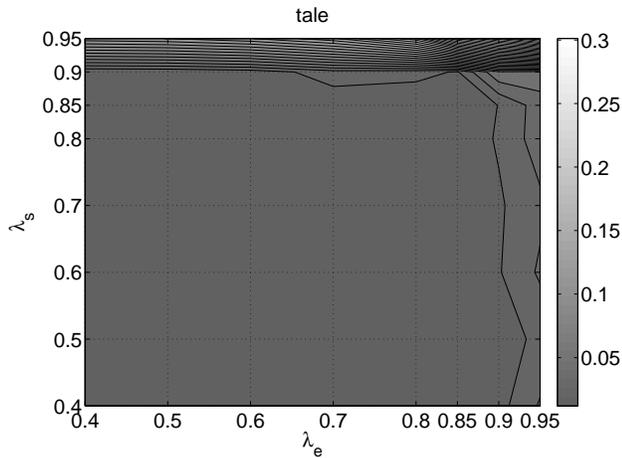


Figure 3: Contour plot of the Amari-index as a function of the λ_s, λ_e invertibility parameters on the *tale* database. Number of samples: $T = 20,000$. For error curves and numerical values, see Fig. 9 and Table III, respectively.

6 – 10% (*ABC*), 2 – 5% (*tale*), 5 – 11% (*3D-geom* database) with standard deviations being about the same order of magnitude as the respective means.

Estimations with average Amari-indices are shown in Fig. 5, Fig. 6, and Fig. 7. According to these figures, our ARMAX-IPA method can provide acceptable estimations up to about $(\lambda_s, \lambda_e) = (0.9, 0.9) - (0.9, 0.95)$ values.

VI. CONCLUSIONS

In this paper we addressed the controlled complete Autoregressive Moving Average Independent Process Analysis (ARMAX-IPA) problem. We treated the model of hidden multidimensional ARMA processes (i) driven by hidden independent multidimensional noise processes, (ii) observed through their linear mixtures, and (iii) subject to exogenous

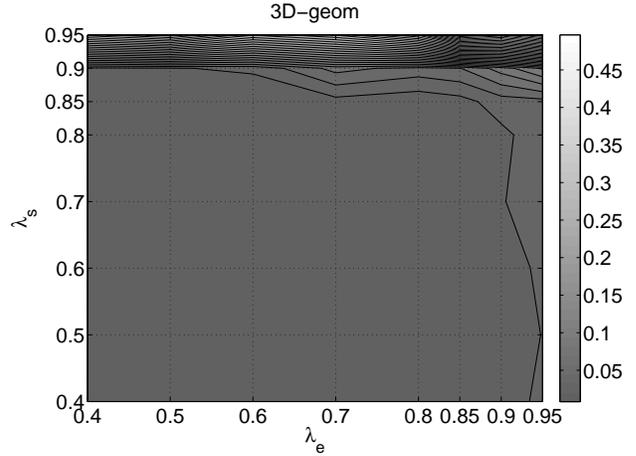


Figure 4: Contour plot of the Amari-index as a function of the λ_s, λ_e invertibility parameters on the *3D-geom* database. Number of samples: $T = 20,000$. For error curves and numerical values, see Fig. 10 and Table IV, respectively.

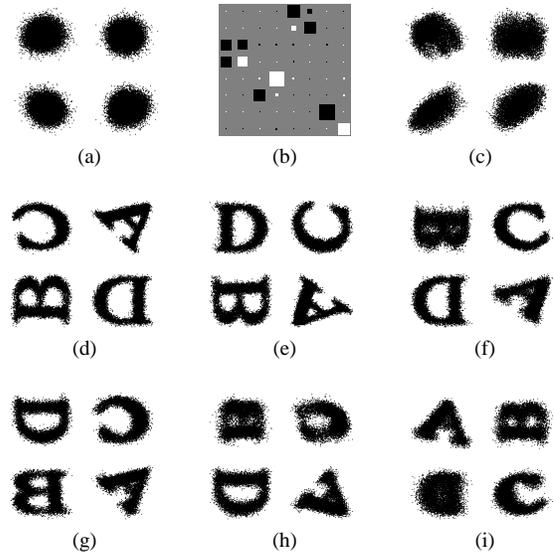


Figure 5: Illustration of the estimations on the *ABC* dataset. Number of samples: $T = 20,000$. In (a)-(d): $(\lambda_s, \lambda_e) = (0.4, 0.4)$. (a): observed signal $\mathbf{x}(t)$. (c): estimation of the \mathbf{Ae} innovation, input of the ISA procedure. (d): estimated components $\hat{\mathbf{e}}^m$, recovered up to the ISA ambiguities. (b): Hinton-diagram of \mathbf{G} , ideally a block-permutation matrix with 2×2 blocks. (e)-(i): the same as (d), but for $(\lambda_s, \lambda_e) = (0.7, 0.7), (0.85, 0.85), (0.9, 0.85), (0.9, 0.9), (0.9, 0.95)$, respectively. All the plotted estimations have average Amari-indices, see Table II.

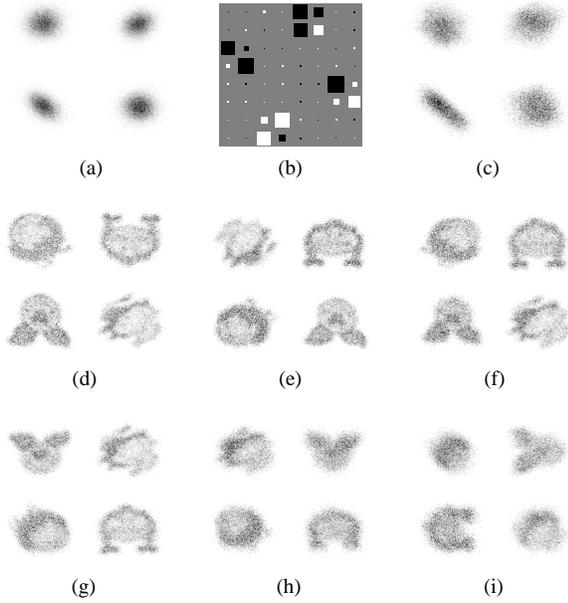


Figure 6: Illustration of the estimations on the *tale* dataset. Number of samples: $T = 20,000$. In (a)-(d): $(\lambda_s, \lambda_e) = (0.4, 0.4)$. (a): observed signal $\mathbf{x}(t)$. (c): estimation of the \mathbf{Ae} innovation, input of the ISA procedure. (d): estimated components $\hat{\mathbf{e}}^m$, recovered up to the ISA ambiguities. (b): Hinton-diagram of \mathbf{G} , ideally a block-permutation matrix with 2×2 blocks. (e)-(i): the same as (d), but for $(\lambda_s, \lambda_e) = (0.7, 0.7)$, $(0.85, 0.85)$, $(0.9, 0.85)$, $(0.9, 0.9)$, $(0.9, 0.95)$, respectively. All the plotted estimations have average Amari-indices, see Table III.

Table II: Amari-index in percentages on the *ABC* dataset for different (λ_s, λ_e) invertibility parameter pairs: mean \pm standard deviation. Number of samples: $T = 20,000$. For other (i) (λ_s, λ_e) pairs, (ii) sample numbers between $1,000 \leq T < 20,000$, see Fig. 2 and Fig. 8, respectively. For the illustration of the estimations, see Fig. 5.

$(\lambda_s, \lambda_e) = (0.4, 0.4)$	$(\lambda_s, \lambda_e) = (0.7, 0.7)$	$(\lambda_s, \lambda_e) = (0.85, 0.85)$
1.11% (± 0.27)	1.14% (± 0.17)	2.40% (± 2.30)
$(\lambda_s, \lambda_e) = (0.9, 0.85)$	$(\lambda_s, \lambda_e) = (0.9, 0.9)$	$(\lambda_s, \lambda_e) = (0.9, 0.95)$
6.24% (± 8.76)	5.73% (± 6.54)	9.40% (± 11.76)

Table III: Amari-index in percentages on the *tale* dataset for different (λ_s, λ_e) invertibility parameter pairs: mean \pm standard deviation. Number of samples: $T = 20,000$. For other (i) (λ_s, λ_e) pairs, (ii) sample numbers between $1,000 \leq T < 20,000$, see Fig. 3 and Fig. 9, respectively. For the illustration of the estimations, see Fig. 6.

$(\lambda_s, \lambda_e) = (0.4, 0.4)$	$(\lambda_s, \lambda_e) = (0.7, 0.7)$	$(\lambda_s, \lambda_e) = (0.85, 0.85)$
1.22% (± 0.12)	1.23% (± 0.15)	1.50% (± 0.46)
$(\lambda_s, \lambda_e) = (0.9, 0.85)$	$(\lambda_s, \lambda_e) = (0.9, 0.9)$	$(\lambda_s, \lambda_e) = (0.9, 0.95)$
2.07% (± 1.37)	4.90% (± 4.64)	4.75% (± 4.01)

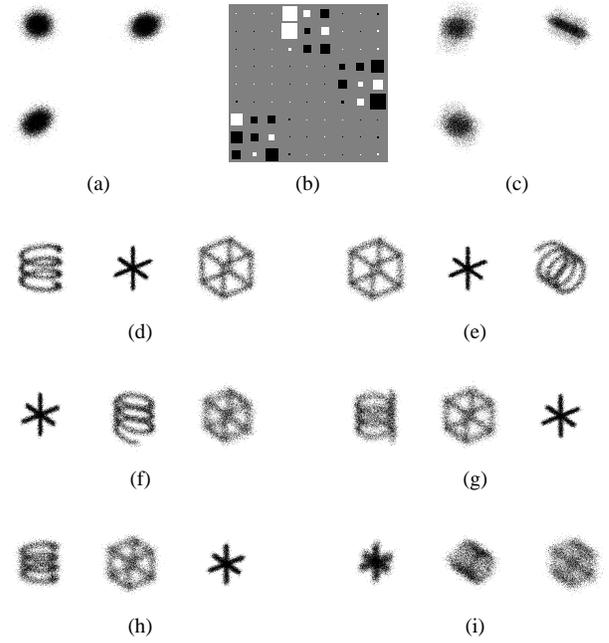


Figure 7: Illustration of the estimations on the *3D-geom* dataset. Number of samples: $T = 20,000$. In (a)-(d): $(\lambda_s, \lambda_e) = (0.4, 0.4)$. (a): observed signal $\mathbf{x}(t)$. (c): estimation of the \mathbf{Ae} innovation, input of the ISA procedure. (d): estimated components $\hat{\mathbf{e}}^m$, recovered up to the ISA ambiguities. (b): Hinton-diagram of \mathbf{G} , ideally a block-permutation matrix with 3×3 blocks. (e)-(i): the same as (d), but for $(\lambda_s, \lambda_e) = (0.7, 0.7)$, $(0.85, 0.85)$, $(0.9, 0.85)$, $(0.9, 0.9)$, $(0.9, 0.95)$, respectively. All the plotted estimations have average Amari-indices, see Table IV.

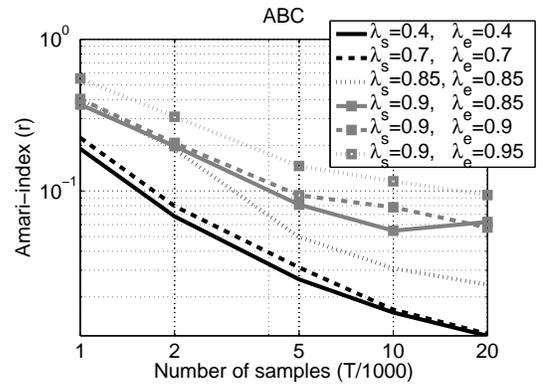


Figure 8: Amari-index as a function of the sample number on log-log scale for different λ_s, λ_e invertibility parameters on the *ABC* database. For different (λ_s, λ_e) pairs (contour plot) and numerical values, see Fig. 2 and Table II, respectively.

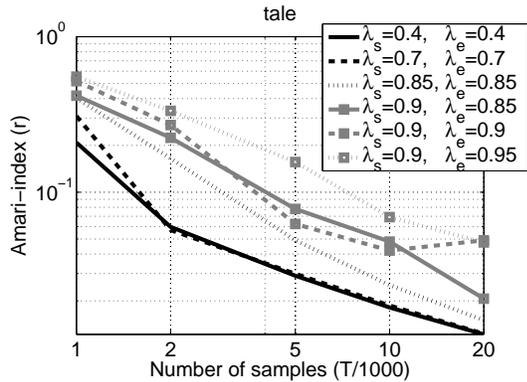


Figure 9: Amari-index as a function of the sample number on log-log scale for different λ_s , λ_e invertibility parameters on the *tale* database. For different (λ_s, λ_e) pairs (contour plot) and numerical values, see Fig. 3 and Table III, respectively.

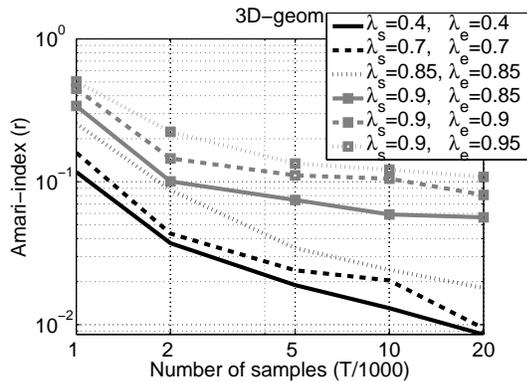


Figure 10: Amari-index as a function of the sample number on log-log scale for different λ_s , λ_e invertibility parameters on the *3D-geom* database. For different (λ_s, λ_e) pairs (contour plot) and numerical values, see Fig. 4 and Table IV, respectively.

Table IV: Amari-index in percentages on the *3D-geom* dataset for different (λ_s, λ_e) invertibility parameter pairs: mean \pm standard deviation. Number of samples: $T = 20,000$. For other (i) (λ_s, λ_e) pairs, (ii) sample numbers between $1,000 \leq T < 20,000$, see Fig. 4 and Fig. 10, respectively. For the illustration of the estimations, see Fig. 7.

$(\lambda_s, \lambda_e) = (0.4, 0.4)$	$(\lambda_s, \lambda_e) = (0.7, 0.7)$	$(\lambda_s, \lambda_e) = (0.85, 0.85)$
0.85% (± 0.15)	0.94% (± 0.23)	1.81% (± 0.75)
$(\lambda_s, \lambda_e) = (0.9, 0.85)$	$(\lambda_s, \lambda_e) = (0.9, 0.9)$	$(\lambda_s, \lambda_e) = (0.9, 0.95)$
5.64% (± 6.91)	8.08% (± 11.00)	10.82% (± 13.27)

control. For the estimation, we adapted the D-optimality principle. We divided the solution of the problem into two parts, the estimation of a fully observable ARX problem and the Independent Subspace Analysis (ISA) task that we can solve. We also demonstrated the efficiency of the algorithm on different datasets. Our simulations revealed that (i) the error of the estimation of the hidden sources decreases approximately in a power law fashion as the sample size increases and (ii) estimation is robust against values of the invertibility parameter. The problem family that we treated may gain applications among others in human-computer interaction serving the user.

REFERENCES

- [1] C. Jutten and J. Herault, "Blind separation of sources: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.
- [2] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [3] A. Cichocki and S. Amari, *Adaptive blind signal and image processing*. John Wiley & Sons, 2002.
- [4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
- [5] S. Choi, A. Cichocki, H. Park, and S. Lee, "Blind source separation and independent component analysis," *Neural Information Processing - Letters and Reviews*, vol. 6, pp. 1–57, 2005.
- [6] J. Cardoso, "Multidimensional independent component analysis," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 4, 1998, pp. 1941–1944.
- [7] S. Akaho, Y. Kiuchi, and S. Umeyama, "MICA: Multimodal independent component analysis," in *International Joint Conference on Neural Networks (IJCNN '99)*, vol. 2, 1999, pp. 927–932.
- [8] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, pp. 1705–1720, 2000.
- [9] A. Hyvärinen and U. Köster, "FastISA: A fast fixed-point algorithm for independent subspace analysis," in *European Symposium on Artificial Neural Networks (ESANN 2006)*. Evere, Belgium: d-side, 2006, pp. 371–376.
- [10] R. Vollgraf and K. Obermayer, "Multi-dimensional ICA to separate correlated sources," in *Neural Information Processing Systems (NIPS 2001)*, vol. 14. Cambridge, MA: MIT Press, 2001, pp. 993–1000.
- [11] F. R. Bach and M. I. Jordan, "Beyond independent components: Trees and clusters," *Journal of Machine Learning Research*, vol. 4, pp. 1205–1233, 2003.
- [12] H. Stögbauer, A. Kraskov, S. A. Astakhov, and P. Grassberger, "Least dependent component analysis based on mutual information," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 066123, Dec. 2004.
- [13] B. Póczos and A. Lőrincz, "Independent subspace analysis using k-nearest neighborhood distances," *Artificial Neural Networks: Formal Models and their Applications (ICANN 2005)*, vol. 3697, pp. 163–168, 2005.
- [14] —, "Independent subspace analysis using geodesic spanning trees," in *International Conference on Machine Learning (ICML 2005)*, vol. 119. New York, NY, USA: ACM Press, 2005, pp. 673–680.
- [15] F. J. Theis, "Blind signal separation into groups of dependent signals using joint block diagonalization," in *International Society for Computer Aided Surgery (ISCAS 2005)*, vol. 6, 2005, pp. 5878–5881.
- [16] —, "Multidimensional independent component analysis using characteristic functions," in *European Signal Processing Conference (EU-SIPCO 2005)*, 2005.
- [17] —, "Towards a general independent subspace analysis," in *Neural Information Processing Systems (NIPS 2006)*, vol. 19, 2006.
- [18] Z. Szabó and A. Lőrincz, "Real and complex independent subspace analysis by generalized variance," in *ICA Research Network International Workshop (ICARN'06)*, 2006, pp. 85–88, <http://arxiv.org/abs/math.ST/0610438>.
- [19] G. Nolte, F. C. Meinecke, A. Ziehe, and K.-R. Müller, "Identifying interactions in mixed and noisy complex systems," *Physical Review E*, vol. 73, no. 051913, 2006.

- [20] Z. Szabó, B. Póczos, and A. Lőrincz, "Undercomplete blind subspace deconvolution," *Journal of Machine Learning Research*, vol. 8, pp. 1063–1095, 2007.
- [21] D. N. Levin, "Using state space differential geometry for nonlinear blind source separation," University of Chicago, Tech. Rep., 2006, <http://arxiv.org/abs/cs/0612096>.
- [22] A. Sharma and K. K. Paliwal, "Subspace independent component analysis using vector kurtosis," *Pattern Recognition*, vol. 39, pp. 2227–2232, 2006.
- [23] J.-F. Cardoso and M. Martin, "A flexible component model for precision ICA," in *Independent Component Analysis and Signal Separation (ICA 2007)*, ser. Lecture Notes in Computer Science, M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, Eds., vol. 4666. Heidelberg: Springer, 2007, pp. 1–8.
- [24] Y. Nishimori, S. Akaho, and M. D. Plumbley, "Riemannian optimization method on the flag manifold for independent subspace analysis," in *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, ser. Lecture Notes in Computer Science, J. Rosca, D. Erdogmus, J. C. Principe, and S. Haykin, Eds., vol. 3889. Springer, 5–8 Mar. 2006, pp. 295–302.
- [25] —, "Riemannian optimization method on the generalized flag manifold for complex and subspace ICA," in *MaxEnt 2006: International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, CNRS, Paris, France, 8–13 July 2006.
- [26] Y. Nishimori, S. Akaho, S. A. Abdallah, and M. D. Plumbley, "Flag manifolds for subspace ICA problems," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, vol. 4, Honolulu, Hawaii, USA, 15–20 Apr. 2007, pp. 1417–1420.
- [27] H. Shen, K. Hüper, and M. Kleinstueber, "Block-Jacobi-type methods for log-likelihood based linear independent subspace analysis," in *IEEE International Workshop on Machine Learning (IEEE MLSP07)*, Thessaloniki, Greece, 27–29 Aug. 2007, pp. 133–138.
- [28] H. Shen and K. Hüper, "Generalised FastICA for independent subspace analysis," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, vol. 4, Honolulu, Hawaii, USA, 15–20 Apr. 2007, pp. 1409–1412.
- [29] H. Choi and S. Choi, "Relative gradient learning for independent subspace analysis," in *International Joint Conference on Neural Networks (IJCNN 2006)*, Vancouver, Canada, 16–21 July 2006, pp. 3919–3924.
- [30] M. Castella and P. Comon, "Blind separation of instantaneous mixtures of dependent sources," in *Independent Component Analysis and Signal Separation (ICA 2007)*, ser. Lecture Notes in Computer Science, M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, Eds., vol. 4666. Heidelberg: Springer, 2007, pp. 9–16.
- [31] S. Kirschner and B. Póczos, "ICA and ISA using Schweizer-Wolff measure of dependence," in *International Conference on Machine Learning (ICML 2008)*, A. McCallum and S. Roweis, Eds. Omnipress, 2008, pp. 464–471.
- [32] H. Kim, S. Choi, and S. Bang, "Membership scoring via independent feature subspace analysis for grouping co-expressed genes," in *International Joint Conference on Neural Networks (IJCNN 2003)*, vol. 3, 20–24 July 2003, pp. 1690–1695.
- [33] H. Kim and S. Choi, "Independent subspaces of gene expression data," in *IASTED International Conference on Artificial Intelligence and Applications*, Innsbruck, Austria, 14–16 Feb. 2005.
- [34] J. K. Kim and S. Choi, "Tree-dependent components of gene expression data for clustering," in *International Conference on Artificial Neural Networks (ICANN 2006)*, ser. Lecture Notes in Computer Science, vol. 4132. Springer Berlin / Heidelberg, 2006, pp. 837–846.
- [35] S. Z. Li, X. Lv, and H. Zhang, "View-based clustering of object appearances based on independent subspace analysis," in *International Conference on Computer Vision (ICCV'01)*, vol. 2, Vancouver, BC, Canada, 2001, pp. 295–300.
- [36] —, "View-subspace analysis of multi-view face patterns," in *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, Vancouver, BC, Canada, 2001, pp. 125–132.
- [37] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," in *Springer Handbook of Speech Processing*. Springer Press, Nov. 2007. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?4924>
- [38] A. MacDonald and S. Cain, "Derivation and application of an anisoplanatic optical transfer function for blind deconvolution of laser radar imagery," *Unconventional Imaging*, vol. 5896, pp. 9–20, 2005.
- [39] J. B. Hedgepeth, V. F. Gallucci, F. O'Sullivan, and R. E. Thorne, "An expectation maximization and smoothing approach for indirect acoustic estimation of fish size and density," *ICES Journal of Marine Science*, vol. 56, no. 1, pp. 36–50, 1999.
- [40] C. Vural and W. A. Sethares, "Blind image deconvolution via dispersion minimization," *Digital Signal Processing*, vol. 16, pp. 137–148, 2006.
- [41] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multi-channel blind deconvolution and speech separation using causal FIR filters," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 92–104, Jan. 2005.
- [42] N. Mitianoudis and M. E. Davies, "Audio source separation of convolutive mixtures," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 489–497, Sept. 2003.
- [43] M. J. Roan, M. R. Gramann, J. G. Erling, and L. H. Sibul, "Blind deconvolution applied to acoustical systems identification with supporting experimental results," *The Journal of the Acoustical Society of America*, vol. 114, no. 4, pp. 1988–1996, Oct. 2003.
- [44] S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolved mixture of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 2, pp. 109–116, Mar. 2003.
- [45] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, Mar. 2002.
- [46] F. Deligianni, B. Lo, and G. Yang, "Source recovery for body sensor network," in *International Workshop on Wearable and Implantable Body Sensor Networks 2006 (BSN 2006)*, Apr. 2006, pp. 199–202.
- [47] T. Jung, S. Makeig, T. Lee, M. J. McKeown, G. Brown, A. J. Bell, and T. J. Sejnowski, "Independent component analysis of biomedical signals," in *International Workshop on Independent Component Analysis and Signal Separation (ICA 2000)*, June 2000, pp. 633–644.
- [48] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *NeuroImage*, vol. 9, pp. 416–429, 1999.
- [49] M. Dyrholm, S. Makeig, and L. K. Hansen, "Model selection for convolutive ICA with an application to spatio-temporal analysis of EEG," *Neural Computation*, Apr. 2007.
- [50] T. Kotzer, N. Cohen, and J. Shamir, "Generalized projection algorithms with applications to optics and signal restoration," *Optics Communications*, vol. 156, no. 1, pp. 77–91, 1998.
- [51] H. Karşlı, "Further improvement of temporal resolution of seismic data by autoregressive (AR) spectral extrapolation," *Journal of Applied Geophysics*, vol. 59, pp. 324–336, 2006.
- [52] Z. Szabó, B. Póczos, and A. Lőrincz, "Undercomplete blind subspace deconvolution via linear prediction," in *European Conference on Machine Learning (ECML'07)*, ser. Lecture Notes in Artificial Intelligence, J. N. K. et al., Ed., vol. 4701. Berlin Heidelberg: Springer-Verlag, 2007, pp. 740–747.
- [53] B. Póczos, Z. Szabó, M. Kiszlinger, and A. Lőrincz, "Independent process analysis without a priori dimensional information," in *Independent Component Analysis and Blind Signal Separation (ICA 2007)*, ser. Lecture Notes in Computer Science, vol. 4666. Berlin Heidelberg: Springer-Verlag, 2007, pp. 252–259.
- [54] B. Póczos and A. Lőrincz, "D-optimal Bayesian interrogation for parameter and noise identification of recurrent neural networks," 2008, (accepted for publication in the *Journal of Machine Learning Research*; preliminary version is available at <http://arxiv.org/abs/0801.1883>).
- [55] V. Petrov, "Central limit theorem for m-dependent variables," in *Proceedings of the All-Union Conference on Probability Theory and Mathematical Statistics*, 1958, pp. 38–44.
- [56] J. Galbraith, A. Ullah, and V. Zinde-Walsh, "Estimation of the vector moving average model by vector autoregression," *Econometric Reviews*, vol. 21, no. 2, pp. 205–219, 2002.
- [57] J. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [58] A. Neumaier and T. Schneider, "Estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Transactions on Mathematical Software*, vol. 27, no. 1, pp. 27–57, 2001.
- [59] T. Schneider and A. Neumaier, "Algorithm 808: ARfit - a Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Transactions on Mathematical Software*, vol. 27, no. 1, pp. 58–65, 2001.
- [60] F. J. Theis, "Uniqueness of complex and multidimensional independent

component analysis,” *Signal Processing*, vol. 84, no. 5, pp. 951–956, 2004.

- [61] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.
- [62] Z. Szabó, B. Póczos, and A. Lőrincz, “Cross-entropy optimization for independent process analysis,” in *Independent Component Analysis and Blind Signal Separation (ICA'06)*, ser. Lecture Notes in Computer Science, vol. 3889. Springer, 2006, pp. 909–916.