NONPARAMETRIC INDEPENDENT PROCESS ANALYSIS

Zoltán Szabó[1] *and Barnabás Póczos*[2]

[1] Faculty of Informatics, Eötvös Loránd University,
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
phone: +36-70-548-3369, fax: +36-1-381-2140,
email: szzoli@cs.elte.hu
web: http://nipg.inf.elte.hu

[2] School of Computer Science, Carnegie Mellon University
5000 Forbes Ave, 15213, Pittsburgh, PA, USA
phone: + 1-412-251-9209,
email: bapoczos@cs.cmu.edu,
web: http://www.autonlab.org

ABSTRACT

Linear dynamical systems are widely used tools to model stochastic time processes, but they have severe limitations; they assume linear dynamics with Gaussian driving noise. Independent component analysis (ICA) aims to weaken these limitations by allowing independent, non-Gaussian sources in the model. Independent subspace analysis (ISA), an important generalization of ICA, has proven to be successful in many source separation applications. Still, the general ISA problem of separating sources with nonparametric dynamics has been hardly touched in the literature yet. The goal of this paper is to extend ISA to the case of (i) nonparametric, asymptotically stationary source dynamics and (ii) unknown source component dimensions. We make use of functional autoregressive (fAR) processes to model the temporal evolution of the hidden sources. An extension of the well-known ISA separation principle is derived for the solution of the introduced fAR independent process analysis (fAR-IPA) task. By applying fAR identification we reduce the problem to ISA. The Nadaraya-Watson kernel regression technique is adapted to obtain strongly consistent fAR estimation. We illustrate the efficiency of the fAR-IPA approach by numerical examples and demonstrate that in this framework our method is superior to standard linear dynamical system based estimators.

1. INTRODUCTION

Linear dynamical systems (LDS) play a central role in the field of stochastic processes. There exist efficient algorithms for the identification (parameter estimation) and filtering (state estimation) of these systems in the literature. One of the most popular identification algorithms is the expectation maximization (EM) based method [12, 23]. This is a maximum likelihood estimation method that also incorporates the Kalman filter/smoother (KF/KS) problems to estimate the values of the hidden states [18]. These linear dynamical systems, however, have several limitations. They assume that (i) the driving noise in their hidden layer is Gaussian, and (ii) the evolution of the hidden states can be described by simple linear dynamics. These limitations inspired considerable research efforts to derive efficient techniques for the estimation of the hidden state in nonlinear systems, i.e., to solve nonlinear filtering problems. They include the extended KF [16], and the sigma-point KF family [33] with important special cases such as the unscented KF and the central difference KF. These methods can also be applied to the dual estimation problem, where our goal is to estimate both the system parameters and the hidden states. Although these methods are nonlinear by nature, they still assume special *parametric* dynamical models. Estimation in the nonparametric case, especially when the driving noise is non-Gaussian, is even more challenging, and only very little work has been published on this problem so far. In our paper we investigate this problem. Particularly, we consider the dual estimation of state space models with unknown, nonparametric dynamics in the hidden layer. We derive a simple and efficient estimation method for the special case where the driving noise of the hidden layer consists of *independent*, multidimensional, non-Gaussian components.

Independent component analysis (ICA), the problem of separating mixed non-Gaussian independent sources, has received considerable attention in signal processing [17, 9]. The original form of ICA considers *one-dimensional* independent sources only. One may think of this task as a cocktail party problem: we have D independent speakers (sources) and D microphones (sensors) which measure the mixed signals emitted by the sources. The task is to recover the original sources from the mixed observations only. For a recent review about ICA, see [8]. The model is more realistic if one assumes that not all, but only some *groups* of the hidden sources are independent ('speakers are talking in groups'). This is the independent subspace analysis (ISA) generalization of the ICA problem [6, 10]. The ISA model has already had some exciting applications including (i) the processing of EEG-fMRI data [1, 20, 22] and natural images [14, 27], (ii) gene expression analysis [19], (iii) learning of face view-subspaces [21], (iv) ECG analysis [10, 6, 28, 31, 26], (v) motion segmentation [11], (vi) single-channel source separation [7], (vii) texture classification [25]. For a recent review of ISA techniques, see [29].

One can relax the traditional independent identically distributed (i.i.d.) assumption of ISA and model the temporal evolution of the sources, for example, by autoregressive (AR) processes [24]. However, the general case of sources with unknown, nonparametric dynamics is more challenging, and very little is known about it [31, 3]. [3] focused on the separation of stationary and ergodic source components of known and equal dimensions in case of constrained mixing matrices. [31] was dealing with wide sense stationary sources that (i) are supposed to be block-decorrelated for all time-shifts and (ii) have equal and known dimensional source components.

The contributions of our paper are as follows:

- we address the problem of ISA with nonparametric, asymptotically stationary dynamics,
- beyond this extension we also treat the case of unknown and possibly different source component dimensions,
- we allow the temporal evolution of the sources to be coupled; it is sufficient that their driving noises are indepen-

dent.

- we propose a simple estimation scheme by reducing the problem to kernel regression and ISA.

The paper is structured as follows. Section 2 formulates the problem set-up. In Section 3 we describe our identification method. Section 4 contains our numerical experiments to illustrate the efficiency of our algorithm. Here we also show that in this problem setting our method performs better than standard EM-LDS based estimators. Conclusions are drawn in Section 5.

2. THE FUNCTIONAL AUTOREGRESSIVE INDEPENDENT PROCESS ANALYSIS MODEL

In this section we formally define the problem set-up. In our framework we use functional autoregressive (fAR) processes to model nonparametric stochastic time series. The goal of this paper is to develop dual estimation methods for the functional autoregressive independent process analysis (fAR-IPA) model, which is defined as follows. Assume that the observation ($\mathbf{x}$) is a linear mixture ($\mathbf{A}$) of the hidden source ($\mathbf{s}$), which evolves according to an unknown fAR dynamics ($\mathbf{f}$) with independent driving noises ($\mathbf{e}$). Formally,

$$\mathbf{s}_t = \mathbf{f}(\mathbf{s}_{t-1}, \ldots, \mathbf{s}_{t-p}) + \mathbf{e}_t, \tag{1}$$

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \tag{2}$$

where the unknown mixing matrix $\mathbf{A} \in \mathbb{R}^{D_x \times D_s}$ is of full column rank, p is the order of the process and the $\mathbf{e}^m \in \mathbb{R}^{d_m}$ components of $\mathbf{e} = \left[\mathbf{e}^1; \ldots; \mathbf{e}^M\right] \in \mathbb{R}^{D_s}$ ($D_s = \sum_{m=1}^{M} d_m$) are (i) non-Gaussian, (ii) i.i.d. in time t and (iii) independent, $I(\mathbf{e}^1, \ldots, \mathbf{e}^M) = 0$, where I denotes mutual information. The goal of the fAR-IPA problem is to estimate (i) the mixing matrix $\mathbf{A}$ (or its left inverse $\mathbf{W} = \mathbf{A}^{-1}$) and (ii) the original source $\mathbf{s}_t$ by using observations $\mathbf{x}_t$ only.

We list a few interesting special cases:

- If we knew the parametric form of $\mathbf{f}$, and if it were linear, then the problem would be the AR-IPA (autoregressive IPA) task [24].
- If we assume that the dynamics of the hidden layer is zero-order AR, then the problem reduces to the original ISA problem [6].
- If we deal with the ISA problem, and the independent subspaces are one-dimensional, ($d_m = 1, \forall m$), then our problem is the traditional ICA problem.

3. METHOD

We consider the dual estimation of the system described in (1)–(2). In the important special case when $\mathbf{f}$ has known, linear form with order $p = 0$, then one can use the *ISA separation principle* of Jean-François Cardoso [6], who conjectured that the solution of the ISA problem can be separated into two steps: (i) applying traditional ICA and then (ii) clustering the ICA elements into statistically independent groups. This principle forms the basis of the state-of-the-art ISA solvers. While the extent of this conjecture is still on open issue, *sufficient* conditions have recently been given for the principle [30]. In what follows, we will propose a similar reduction scheme with which we can reduce the fAR-IPA estimation problem ((1)–(2)) to a functional AR process identification and an ISA problem. To obtain strongly consistent fAR estimation, the Nadaraya-Watson kernel regression technique is invoked.

More formally, the estimation of the fAR-IPA problem (1)-(2) can be accomplished as follows. The observation process $\mathbf{x}$ is left invertible linear transformation of the hidden fAR source process $\mathbf{s}_t$ and thus it is also fAR process with innovation $\mathbf{A}\mathbf{e}_t$

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t = \mathbf{A}\mathbf{f}(\mathbf{s}_{t-1}, \ldots, \mathbf{s}_{t-p}) + \mathbf{A}\mathbf{e}_t \tag{3}$$

$$= \mathbf{A}\mathbf{f}(\mathbf{A}^{-1}\mathbf{x}_{t-1}, \ldots, \mathbf{A}^{-1}\mathbf{x}_{t-p}) + \mathbf{A}\mathbf{e}_t \tag{4}$$

$$= \mathbf{g}(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-p}) + \mathbf{n}_t, \tag{5}$$

where function

$$\mathbf{g}(\mathbf{u}_1, \ldots, \mathbf{u}_p) = \mathbf{A}\mathbf{f}(\mathbf{A}^{-1}\mathbf{u}_1, \ldots, \mathbf{A}^{-1}\mathbf{u}_p) \tag{6}$$

describes the temporal evolution of $\mathbf{x}_t$, and

$$\mathbf{n}_t = \mathbf{A}\mathbf{e}_t \tag{7}$$

stands for the driving noise of the observation. Making use of this form, the fAR-IPA estimation can be carried out by fAR fit to observation $\mathbf{x}_t$ followed by ISA on $\hat{\mathbf{n}}_t$, the estimated innovation of $\mathbf{x}_t$.

Note that Eq. (5) can be considered as a nonparametric regression problem; we have $\mathbf{u}_t = [\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-p}]$, $\mathbf{v}_t = \mathbf{x}_t$ ($t = 1, \ldots, T$) samples from the unknown relation

$$\mathbf{v}_t = \mathbf{g}(\mathbf{u}_t) + \mathbf{n}_t, \tag{8}$$

where $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{n}$ are the explanatory-, response variables and noise, respectively, and $\mathbf{g}$ is the unknown conditional mean or regression function. Nonparametric techniques can be applied to estimate the unknown mean function

$$\mathbf{g}(\mathbf{U}) = \mathbb{E}(\mathbf{V}|\mathbf{U}), \tag{9}$$

e.g., by carrying out kernel density estimation for random variables ($\mathbf{u}$,$\mathbf{v}$) and $\mathbf{u}$, where $\mathbb{E}$ stands for expectation. The resulting Nadaraya-Watson estimator (i) takes the simple form

$$\hat{\mathbf{g}}_0(\mathbf{u}) = \frac{\sum_{t=1}^{T} \mathbf{v}_t K\left(\frac{\mathbf{u}-\mathbf{u}_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{\mathbf{u}-\mathbf{u}_t}{h}\right)}, \tag{10}$$

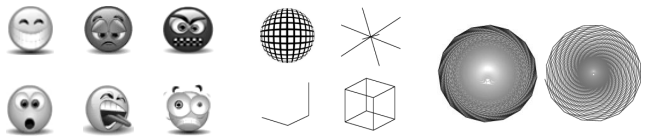
where K and $h > 0$ denotes the applied kernel (a non-negative real-valued function that integrates to one) and bandwith, respectively. It can be used to provide a strongly consistent estimation of the regression function $\mathbf{g}$ for stationary $\mathbf{x}_t$ processes [5]. It has been shown recently [13] that for first order and only *asymptotically stationary* fAR processes, under mild regularity conditions, one can get strongly constistent estimation for the innovation $\mathbf{n}_t$ by applying the recursive version of the Nadaraya-Watson estimator

$$\hat{\mathbf{g}}(\mathbf{u}) = \frac{\sum_{t=1}^{T} t^{\beta D_x} \mathbf{v}_t K(t^{\beta}(\mathbf{u} - \mathbf{u}_t))}{\sum_{t=1}^{T} t^{\beta D_x} K(t^{\beta}(\mathbf{u} - \mathbf{u}_t))}, \tag{11}$$

where the bandwith is parameterized by $\beta \in (0, 1/D_x)$.

4. ILLUSTRATIONS

Now we illustrate the efficiency of the algorithm presented in Section 3. Test databases are described in Section 4.1. To evaluate the solutions, we use the performance measure given in Section 4.2. The numerical results are summarized in Section 4.3.



(a) (b) (c)

Figure 1: Illustration of the (a) *smiley*, (b) *d-geom* and (c) *ikeda* datasets.

4.1 Databases

We define three databases to study the performance of our identification scheme. The *smiley* dataset has 2-dimensional source components ($d_m = 2$) generated from images of the 6 basic facial expressions[1], see Fig. 1(a). Sources $\mathbf{e}^m$ were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities. In other words, the 2-dimensional images were considered as density functions. In the *d-geom* dataset $\mathbf{e}^m$s were random variables uniformly distributed on d_m-dimensional geometric forms. Geometrical forms were chosen as follows. We used: (i) the surface of the unit ball, (ii) the straight lines that connect the opposing corners of the unit cube, (iii) the broken line between $d_m + 1$ points $\mathbf{0} \to \mathbf{e}_1 \to \mathbf{e}_1 + \mathbf{e}_2 \to \ldots \to \mathbf{e}_1 + \ldots + \mathbf{e}_{d_m}$ (where $\mathbf{e}_i$ is the i canonical basis vector in $\mathbb{R}^{d_m}$, i.e., all of its coordinates are zero except the i^{th}, which is 1), and (iv) the skeleton of the unit square. Thus, the number of components M was equal to 4, and the dimension of the components (d_m) can be different. For illustration, see Fig. 1(b). In the *ikeda* test, the hidden $\mathbf{s}_t^m = [s_{t,1}^m, s_{t,2}^m] \in \mathbb{R}^2$ sources realized the ikeda map

$$s_{t+1,1}^m = 1 + \lambda_m [s_{t,1}^m \cos(w_t^m) - s_{t,2}^m \sin(w_t^m)], \quad (12)$$

$$s_{t+1,2}^m = \lambda_m [s_{t,1}^m \sin(w_t^m) + s_{t,2}^m \cos(w_t^m)], \quad (13)$$

where λ_m is a parameter of the dynamical system and

$$w_t^m = 0.4 - \frac{6}{1 + (s_{t,1}^m)^2 + (s_{t,2}^m)^2}. \quad (14)$$

$M = 2$ was chosen with initial points $\mathbf{s}_1^1 = [20; 20]$, $\mathbf{s}_1^2 = [-100; 30]$ and parameters $\lambda_1 = 0.9994$, $\lambda_2 = 0.998$, see Fig. 1(c) for illustration.

4.2 Performance Measure, the Amari-index

The identification of the fAR-IPA model is ambiguous; the hidden $\mathbf{s}^m$ sources can be estimated up to the ISA ambiguities. Nonetheless, these ambiguities are simple [32]: the components of equal dimension can be recovered up to permutation and invertible transformation within the subspaces. Thus, in the ideal case, the product of the ISA demixing matrix $\mathbf{W}_{\text{ISA}}$ and the ISA mixing matrix $\mathbf{A}$,

$$\mathbf{G} = \mathbf{W}_{\text{ISA}} \mathbf{A}, \quad (15)$$

is a block-permutation matrix. This property can be measured by a simple extension of the Amari-index [2] as follows. (i) Assume without loss of generality that the component dimensions and their estimations are ordered in increasing order ($d_1 \leq \ldots \leq d_M$, $\hat{d}_1 \leq \ldots \leq \hat{d}_M$), (ii) decompose $\mathbf{G}$ into $d_i \times d_j$ blocks ($\mathbf{G} = \left[\mathbf{G}^{ij}\right]_{i,j=1,\ldots,M}$), and (iii) define g^{ij} as the sum of the absolute values of the elements of the matrix $\mathbf{G}^{ij} \in \mathbb{R}^{d_i \times d_j}$. Then the Amari-index adapted to the ISA task of different component dimensions is defined as

$$r(\mathbf{G}) := \kappa \left[\sum_{i=1}^{M} \left(\frac{\sum_{j=1}^{M} g^{ij}}{\max_j g^{ij}} - 1 \right) + \sum_{j=1}^{M} \left(\frac{\sum_{i=1}^{M} g^{ij}}{\max_i g^{ij}} - 1 \right) \right], \quad (16)$$

where $\kappa = 1/(2M(M-1))$. 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 block-permutation matrix with $d_i \times d_j$ sized blocks. $r(\mathbf{G}) = 1$ is in the worst case, i.e, when all the g^{ij} elements are equal in absolute value.

4.3 Simulations

We provide empirical results on the *smiley*, *d-geom*, and *ikeda* datasets. For illustration purposes, we chose fAR order $p = 1$ and used the recursive Nadaraya-Watson (11) for functional AR estimation with the Gaussian kernel. The ISA subproblem was solved with the application of the "ISA separation theorem" [6, 30]: FastICA [15] was used as a preprocessing step for the ICA estimation, and then the estimated ICA elements were clustered. The kernel canonical correlation technique [4] was applied to estimate the dependence of the ICA elements. The clustering was carried out by greedy optimization for tasks when the component dimensions were known (*smiley*, *ikeda* datasets). We also studied the case when these component dimensions were unknown (*d-geom* dataset); in this case we used the NCut [34] spectral technique to cluster the estimated ICA components into ISA subspaces. Mixing matrix $\mathbf{A}$ was random orthogonal ($D = D_x = D_s$). For dataset *smiley* and *d-geom*, $\mathbf{f}$ was the composition of a random $\mathbf{F}$ matrix with entries distributed uniformly on interval $[0, 1]$ and the noninvertible sine function, $\mathbf{f}(\mathbf{u}) = \sin(\mathbf{F}\mathbf{u})$. The Amari-index (Section 4.2) was used to evaluate the performance of the proposed fAR-IPA method. For each individual parameter, 10 random runs were averaged. Our parameters included T, the sample number of observations $\mathbf{x}_t$, and bandwith $\beta \in (0, 1/D)$ to study the robustness of the kernel regression approach. β was reparameterized as $\beta = \frac{\beta_c}{D}$ and β_c was chosen from the set $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}\}$. The performance of the method is summarized by notched boxed plots, which show the quartiles (Q_1, Q_2, Q_3), depict the outliers, i.e., those that fall outside of interval $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ by circles, and whiskers represent the largest and smallest non-outlier data points.

For the *smiley* dataset, Fig. 2 demonstrates that the algorithm was able to estimate the hidden components with high precision. Fig. 2(a) shows the Amari-index as a function of the sample number, for $M = 2$ ($D = 4$). The estimation error is plotted on log scale for different bandwith parameters. Fig. 2(c-d) indicate that the problem with $M = 6$ components ($D = 12$) is still amenable to our method when the sample size is large enough ($T = 100,000$). Fig. 2(c) shows the estimated subspaces, and Fig. 2(d) presents the Hinton-diagram. It is approximately a block-permutation matrix with 2×2 blocks indicating that the algorithm could successfully estimate the hidden subspaces.

Our experiences concerning the *d-geom* dataset are summarized in Fig. 3. In contrast to the previous experiment,

[1] See http://www.smileyworld.com.

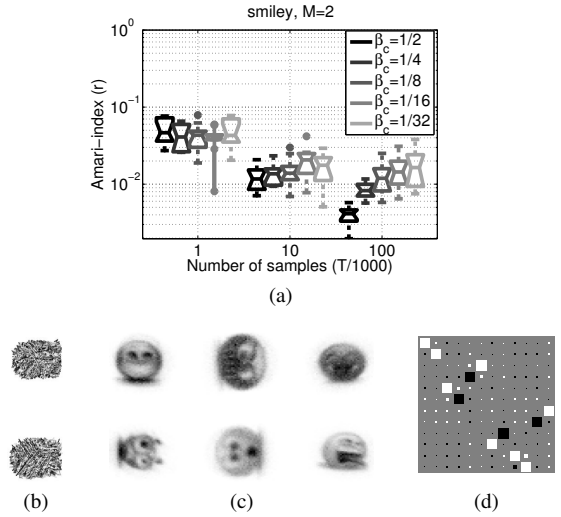


Figure 2: Illustration of the estimations on the *smiley* dataset. (a): Amari-index as a function of the sample number, for $M = 2$. (b): observed signal $\mathbf{x}_t$, the first two 2-dimensional projections when $M = 6$. (c): estimated components ($\hat{\mathbf{e}}^m$) with average (closest to the median) Amari-index for $M = 6$, $\beta_c = \frac{1}{32}$, $T = 100,000$. (d): Hinton-diagram of matrix $\mathbf{G}$.

here the dimensions of the hidden components were different and unknown to the algorithm: $d_1 = 2$, $d_2 = d_3 = 3$, $d_4 = 4$ ($D = 12$). As it can be seen from Fig. 3(a), our method provides precise estimations on this dataset for sample size $T = 100,000 - 150,000$. The Hinton-diagram of matrix $\mathbf{G}$ with average (closest to the median) Amari-index is depicted in Fig. 3(b). Again, this is a block-permutation matrix indicating that the proposed method was able to estimate the hidden subspaces.

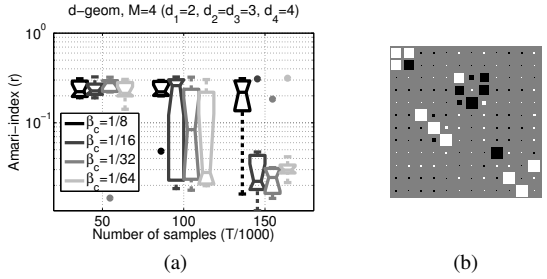


Figure 3: Illustration of the estimations on the *d-geom* dataset. (a) Amari-index on log scale as a function of the sample number for different bandwith parameters on the *d-geom* dataset (with component dimensions: $d_1 = 2$, $d_2 = d_3 = 3$, $d_4 = 4$). (b): Hinton-diagram of $\mathbf{G}$ with average (closest to the median) Amari-index for dataset *d-geom*, $\beta_c = \frac{1}{32}$, $T = 150,000$–it is approximately a block-permutation matrix with one 2×2, two 3×3 and one 4×4 sized block.

.

We ran experiments on the *ikeda* dataset too. Fig. 4(a) illustrates that is we simply use a standard autoregressive approximation method (AR-IPA) [24], then we can not find the proper subspaces. Nevertheless, the Amari-index values of Fig. 4(a) show that the functional AR-IPA approach was able to estimate the hidden subspaces for sample number $T \geq 10,000$. The figure also shows that the estimation is precise for a wide range of bandwith parameters. The Hinton-diagram of matrix $\mathbf{G}$ with average (closest to the median) Amari-index is depicted in Fig. 4(c). This is a block diagonal matrix, which demonstrates that our method was able to separate the mixed subspaces. The estimated hidden sources (with average Amari-index) are illustrated in Fig. 4(d).

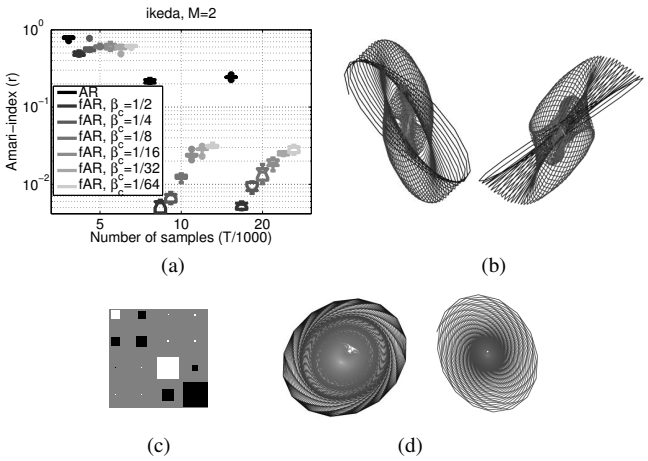


Figure 4: Illustration of the estimations on *ikeda* dataset. (a): Amari-index as a function of the sample number for different bandwith parameters, for AR-IPA and the proposed fAR-IPA approach. (b): Observation, $\mathbf{x}_t$. (c): Hinton-diagram of $\mathbf{G}$ with average (closest to the median) Amari-index. (d): Estimated subspaces using the fAR-IPA method ($\beta_c = \frac{1}{2}$, $T = 20,000$).

Our model (Eq. (1)-(2)) belongs to the family of state space models. Though the dynamics of the hidden variables $\mathbf{s}_t$ is nonlinear, one might wonder whether with a standard linear dynamical system (LDS) based identification method we could identify the parameter $\mathbf{A}$ and the driving noise $\mathbf{e}_t$. The following experiment demonstrates that this is not the case; while our method is efficiently able to cope with this problem, the LDS based identification leads to very poor results. For this purpose we treated the observations $\mathbf{x}_t$ as if they had been generated by an LDS with unknown parameters. We estimated its parameters with the EM method [23, 12], and then using these estimated parameters we applied a Kalman smoother to estimate the hidden dynamical layer $\mathbf{s}_t$ and the driving noise $\mathbf{e}_t$. After this estimation we post-processed the estimated noise $\hat{\mathbf{e}}_t$ with ISA. We performed these estimations on the *smiley* and *d-geom* datasets. Using 10 independent experiments, the EM-LDS based estimators led to $r = 0.56$ and $r = 0.48$ Amari-indices (minima of the Q_2 medians), respectively. These results are very poor; the EM-LDS based method was not able to identify the noise components. On the contrary, the proposed fAR-IPA method successfully estimated the noise components and provided $r = 0.0041$ and $r = 0.0055$ Amari-indices (Fig. 2, Fig. 3).

5. CONCLUSIONS

In this paper we (i) extended independent subspace analysis (ISA) to asymptotically stationary sources, (ii) relaxed the constraint of decoupled (block-decorrelated) dynamics, and (iii) simultaneously addressed the case of unknown

source component dimensions. The temporal evolution of the sources was captured by nonparametric, functional autoregressive (fAR) processes. We generalized the ISA separation technique to the derived fAR setting, and reduced the solution of the problem to fAR identification and ISA. The fAR estimation was carried out by the Nadaraya-Watson kernel regression method with strong consistency guarantee. We extended the Amari-index to different dimensional components and illustrated our technique by numerical experiments. According to our experiences, the fAR-IPA identification can be accomplished robustly and can be advantageous compared to parametric approaches.

Acknowledgments. The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1./B-09/1/KMR-2010-0003.

REFERENCES

[1] S. Akaho, Y. Kiuchi, and S. Umeyama. MICA: Multimodal independent component analysis. In *IJCNN '99*, pages 927–932.

[2] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. *NIPS '96*, pages 757–763.

[3] J. Anemüller. Second-order separation of multidimensional sources with constrained mixing system. In *ICA 2006*, pages 16–23.

[4] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *J. Mach. Learn. Res.*, 3:1–48, 2002.

[5] D. Bosq. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, 1998.

[6] J. Cardoso. Multidimensional independent component analysis. In *ICASSP '98*, pages 1941–1944.

[7] M. A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In *ICMC 2000*, pages 154–161.

[8] A. Cichocki and S. Amari. *Adaptive blind signal and image processing*. Wiley, 2002.

[9] P. Comon. Independent component analysis, a new concept? *Signal Process.*, 36(3):287–314, 1994.

[10] L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *IEEE SP/Athos Workshop on HOS '95*, pages 134–138.

[11] Z. Fan, J. Zhou, and Y. Wu. Motion segmentation based on independent subspace analysis. In *ACCV 2004*.

[12] Z. Ghahramani and G. E. Hinton. Parameter estimation for linear dynamical systems. Technical report, 1996.

[13] N. Hilgert and B. Portier. Strong uniform consistency and asymptotic normality of a kernel based error density estimator in functional autoregressive models. Technical report, 2009.

[14] A. Hyvärinen and P. O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.*, 12:1705–1720, 2000.

[15] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput.*, 9(7):1483–1492, 1997.

[16] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[17] C. Jutten and J. Hérault. Blind separation of sources: An adaptive algorithm based on neuromimetic architecture. *Signal Process.*, 24:1–10, 1991.

[18] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans ASME J. Basic Eng.*, 82:35–45, 1960.

[19] J. K. Kim and S. Choi. Tree-dependent components of gene expression data for clustering. In *ICANN 2006*, pages 837–846.

[20] F. Kohl, G. Wübbeler, D. Kolossa, C. Elster, M. Bär, and R. Orglmeister. Non-independent BSS: A model for evoked MEG signals with controllable dependencies. In *ICA 2009*, pages 443–450.

[21] S. Z. Li, X. Lv, and H. Zhang. View-subspace analysis of multi-view face patterns. In *RATFG-RTS-2001*, pages 125–132.

[22] S. Ma, X.-L. Li, N. Correa, T. Adali, and V. Calhoun. Independent subspace analysis with prior information for fMRI data. In *ICASSP-2010*, pages 1922–1925.

[23] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience, 2008.

[24] B. Póczos, B. Takács, and A. Lőrincz. Independent subspace analysis on innovations. In *ECML 2005*, pages 698–706.

[25] C. S. Santos, J. E. Kögler, and E. D. M. Hernandez. Using independent subspace analysis for selecting filters used in texture processing. In *ICIP 2005*, pages 465–468.

[26] A. Sharma and K. K. Paliwal. Subspace independent component analysis using vector kurtosis. *Pattern Recognit.*, 39:2227–2232, 2006.

[27] F. Sinz, E. Simoncelli, and M. Bethge. Hierarchical modeling of local image features through L_p-nested symmetric distributions. In *NIPS 2009*, pages 1696–1704.

[28] H. Stögbauer, A. Kraskov, S. A. Astakhov, and P. Grassberger. Least dependent component analysis based on mutual information. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, 70(6):066123, 2004.

[29] Z. Szabó. *Separation Principles in Independent Process Analysis*. PhD thesis, 2009.

[30] Z. Szabó, B. Póczos, and A. Lőrincz. Undercomplete blind subspace deconvolution. *J. Mach. Learn. Res.*, 8:1063–1095, 2007.

[31] F. J. Theis. Blind signal separation into groups of dependent signals using joint block diagonalization. In *ISCAS 2005*, pages 5878–5881.

[32] F. J. Theis. Towards a general independent subspace analysis. In *NIPS 2007*, volume 19, pages 1361–1368.

[33] R. van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*. PhD thesis, 2004.

[34] S. X. Yu and J. Shi. Multiclass spectral clustering. In *ICCV 2003*, volume 1, pages 313–319.