ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Autoregressive Independent Process Analysis with Missing Observations

Zoltán Szabó*

Eötvös Loránd University - Department of Software Technology and Methodology
Pázmány P. sétány 1/C, Budapest, H-1117 - Hungary

Abstract. The goal of this paper is to search for independent multi-dimensional processes subject to missing and mixed observations. The corresponding cocktail-party problem has a number of successful applications, however, the case of missing observations has been worked out only for the simplest Independent Component Analysis (ICA) task, where the hidden processes (i) are one-dimensional, and (ii) signal generation in time is independent and identically distributed (i.i.d.). Here, the missing observation situation is extended to processes with (i) autoregressive (AR) dynamics and (ii) multidimensional driving sources. Performance of the solution method is illustrated by numerical examples.

1 Introduction

Independent Component Analysis (ICA) [1] has received considerable attention in signal processing. One may consider ICA as a cocktail party problem: we have D 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. One assumes in this estimation that the sources are independent. ICA model has gained successful applications, e.g., in (i) feature extraction, (ii) denoising, (iii) analysis of financial and neurobiological data, and (iv) face processing. For a recent review about ICA, see [2]. The model is more realistic if we assume 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 [3]. Applications of the ISA model include (i) the analysis of EEG, fMRI, ECG signals and gene data, (ii) pattern and face direction recognition. For a recent review of ISA techniques, see [4].

In spite of the large number of successful applications, the case of missing observation is considered only for the simplest ICA model in the literature [5, 6]. We extend the solution to (i) multidimensional sources (ISA) and (ii) ease the i.i.d. constraint; we consider AR Independent Process Analysis (AR-IPA problem).

The paper is structured as follows: Section 2 formulates the problem domain. Section 3 treats the solution technique. Section 4 contains the numerical illustrations. Conclusions are drawn in Section 5.

*This work has been supported by the National Office for Research and Technology.

2 The AR-IPA Model with Missing Observations

We define the AR-IPA model for missing observations (mAR-IPA). Let us assume that we can only partially (at certain coordinates/time instants) observe ($\mathbf{y}$) the mixture ($\mathbf{x}$) of independent AR sources, that is

$$\mathbf{s}_{t+1} = \sum_{l=0}^{L-1} \mathbf{F}_l \mathbf{s}_{t-l} + \mathbf{e}_{t+1}, \qquad \mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \qquad \mathbf{y}_t = \mathcal{M}_t(\mathbf{x}_t), \tag{1}$$

where (i) the driving noises, or the innovation process $\mathbf{e}^m$ of the components $\mathbf{s}^m \in \mathbb{R}^{d_m}$ of the hidden source $\mathbf{s} = [\mathbf{s}^1; \ldots; \mathbf{s}^M]$ are independent, non-Gaussian, and i.i.d. in time, (ii) the unknown *mixing matrix* $\mathbf{A} \in \mathbb{R}^{D \times D}$ is invertible, $(D = \sum_{m=1}^M d_m)$, (iii) the AR dynamics $\mathbf{F}[z] = \mathbf{I} - \sum_{l=0}^{L-1} \mathbf{F}_l z^{l+1}$, where $\mathbf{I}$ is the identity matrix and z is the time step operator, is *stable*, that is $\det(\mathbf{F}[z]) \neq 0$, for all $z \in \mathbb{C}, |z| \leq 1$, (iv) the $\mathcal{M}_t$ 'mask mappings' represent the coordinates and the time indices of the non-missing observations. Our task is the estimation of the hidden source $\mathbf{s}$ and the mixing matrix $\mathbf{A}$ (or its inverse $\mathbf{W}$) from observation $\mathbf{y}$. The case $\mathcal{M}_t$ = identity and $L = 0$ corresponds to the ISA task, and if $\forall d_m = 1$ also holds then the ICA task is recovered.

3 Method

The mAR-IPA task can be accomplished as follows. The $\mathbf{x}$ process corresponds to an invertible linear transformation of the hidden AR process $\mathbf{s}$ and thus the $\mathbf{x}$ process is also an AR process with innovation $\mathbf{A}\mathbf{e}_{t+1}$: $\mathbf{x}_{t+1} = \sum_{l=0}^{L-1} \mathbf{A}\mathbf{F}_l\mathbf{A}^{-1}\mathbf{x}_{t-l} + \mathbf{A}\mathbf{e}_{t+1}$. According to the d-dependent central limit theorem [7], the distribution of variable $\mathbf{A}\mathbf{e}$ is approximately Gaussian, so one carry out the estimation by

1. identifying the partially observed AR process $\mathbf{y}_t$, and then by
2. estimating the independent components $\mathbf{e}^m$ from the estimated innovation by means of ISA.

For further details, see Section 4.2.

4 Illustrations

Here, we illustrate the efficiency of the proposed mAR-IPA estimation technique. Test cases are introduced in Section 4.1. Numerical results are presented in Section 4.2.

4.1 Databases

We define three databases to study our identification algorithm. In the *ABC* database, hidden sources $\mathbf{e}^m$ were uniform distributions defined by 2-dimensional images ($d_m = 2$) of the English alphabet (see Fig. 1(a); $M = 3$, $D = 6$). In the

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

(a) (b)

Fig. 1: Illustration of the (a): *ABC* and (b): *3D-geom* databases.

3D-geom test $\mathbf{e}^m$s were random variables uniformly distributed on 3-dimensional geometric forms ($d_m = 3$, $M = 2$, $D = 6$; see Fig. 1(b)). In the *Beatles* test hidden sources ($\mathbf{s}^m$) are 8 kHz sampled portions of two stereo Beatles songs (A Hard Day's Night, Can't Buy Me Love; $d_m = 2$, $M = 2$, $D = 4$).[1]

4.2 Simulations

Results on databases *ABC*, *3D-geom* and *Beatles* are provided here. Because of the ambiguities of the ISA problem, components of the hidden sources can only be recovered up to permutation and invertible transformation within the subspaces [8]. For this reason, and because we demonstrate the case of d-dimensional sources, the optimal estimation provides matrix $\mathbf{G} := \hat{\mathbf{W}}_{\text{ISA}}\mathbf{A}$, which is a block-permutation matrix made of $d \times d$ sized blocks. This block-permutation property can be measured by the *Amari-index*

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

where $g_{ij} = \sum_{k,l=1}^{d} |G_{kl}^{ij}|$ with the $d \times d$ sized block decomposition of $\mathbf{G} = \left[\mathbf{G}^{ij} \in \mathbb{R}^{d\times d}\right]_{i,j=1,\ldots,M} \in \mathbb{R}^{D\times D}$.[2] The Amari-index takes values in $[0,1]$: for an ideal block-permutation matrix $\mathbf{G}$ it takes 0; for the worst case it is 1. This measure was used to evaluate the performance of the proposed mAR-IPA method. For each individual parameter, 10 random runs ($\mathbf{A}$, $\mathbf{F}[z]$, $\mathbf{e}$) were averaged. Our parameters are: T, the sample number of observations $\mathbf{y}_t$, L, the order of the AR process, p, the probability of missing observation in $\mathcal{M}_t$ ($x_{t,i}$s, the coordinates of process $\mathbf{x}$, were not observed with probability p, independently), and λ, the (contraction) parameter of the stable polynomial matrix $\mathbf{F}[z]$. It is expected that if the roots of $\mathbf{F}[z]$ are close to the unit circle then our estimation will deteriorate. We investigated this by generating the polynomial matrix $\mathbf{F}[z]$ as $\mathbf{F}[z] = \prod_{l=1}^{L}(\mathbf{I} - \lambda\mathbf{O}_i z) \quad (|\lambda| < 1, \lambda \in \mathbb{R})$, where matrices $\mathbf{O}_i \in \mathbb{R}^{D\times D}$ were random orthogonal and the $\lambda \to 1$ limit was studied. Mixing matrix $\mathbf{A}$ was a random orthogonal matrix. AR fit subject to missing observations was accomplished by means of (i) the maximum likelihood (ML) principle [9], (ii) the

[1] See http://rock.mididb.com/beatles/.

[2] One can apply a similar construction in case of different d_m source dimensions [4].

subspace technique [10], and (iii) in a Bayesian framework using normal-inverted Wishart (shortly *NIW*) conjugate prior and filling in the next missing data using the maximum-a-posteriori estimation of the parameters [11]. In the ISA subtask, we used the ISA separation theorem [12] and grouped the computed ICA components. FastICA [13] was used for the ICA estimation and the dependency of the elements was estimated by means of the kernel canonical correlation method [14]. 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.

The L order of the AR process was 1 and 2 for the *ABC* and the *3D-geom* tasks, respectively, contraction parameter λ was varied between values 0.1 and 0.99, the probability of missing observations took different values ($p = 0.01$, 0.1, 0.15, 0.2), and sample number T was set to $1,000$, $2,000$, and $5,000$. According to our experiences, the methods are efficient on both tasks. The most precise method is *ML* followed by the *subspace* method and the *NIW* technique (see Fig. 2(a)). Running time of the algorithms is the opposite and the *ML* technique is computation time demanding (see Fig. 2(b)). Considering the ratio of missing observations – in the parameter range we studied – the *ML*, the *subspace* and the *NIW* method can handle parameter p up to $0.2 - 0.3$ (see Fig. 2(c)-(d)), $p = 0.15 - 0.2$, and $p = 0.1 - 0.15$, respectively. Figure 2(c)-(d) demonstrate that the *ML* method works robustly for the contraction parameter λ and provides reasonable estimations for values around 1. Figure 2(e)-(j) illustrate the ML component estimations for different p values.

Because of the high computation demands of the *ML* technique, the performances of the *subspace* and *NIW* methods were studied on the *Beatles* test. According to the Schwarz's Bayesian Criterion we used the crude $L = 10$ AR estimation. Results for sample number $T = 30,000$ are summarized in Fig. 3. According to the figure, the methods give reasonable estimations up to $p = 0.1 - 0.15$. In accord with our previous experiences, the *subspace* method is more precise, but it is somewhat slower.

5 Conclusions

We addressed the problem of independent process analysis subject to missing observations. Previous works assumed 1D, i.i.d. ICA sources that we extended to multidimensional Independent Subspace Analysis (ISA) and we allowed an autoregressive dynamics, too. During the solution, we used the separation principle that breaks the problem into two subproblems, namely to the identification of the AR process subject to missing observations and to ISA. Numerical examples were used to illustrate the performance of different methods.

References

[1] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

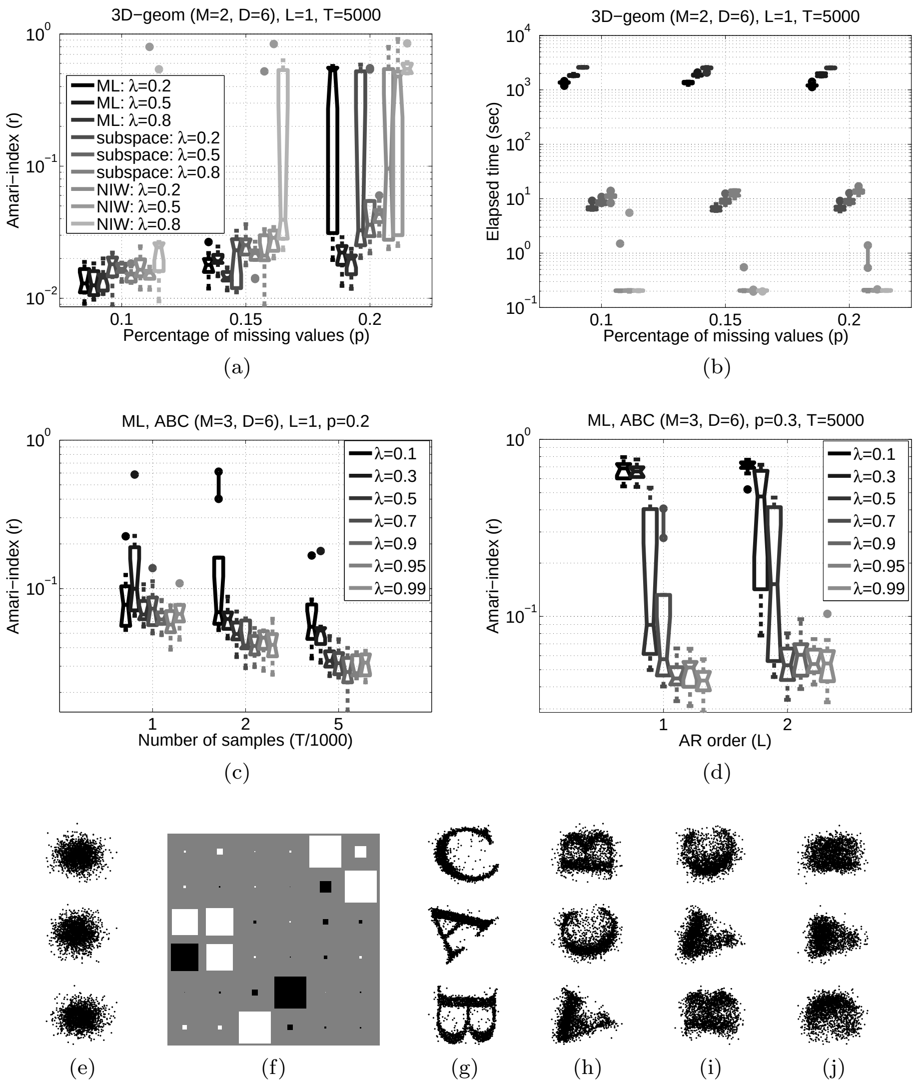
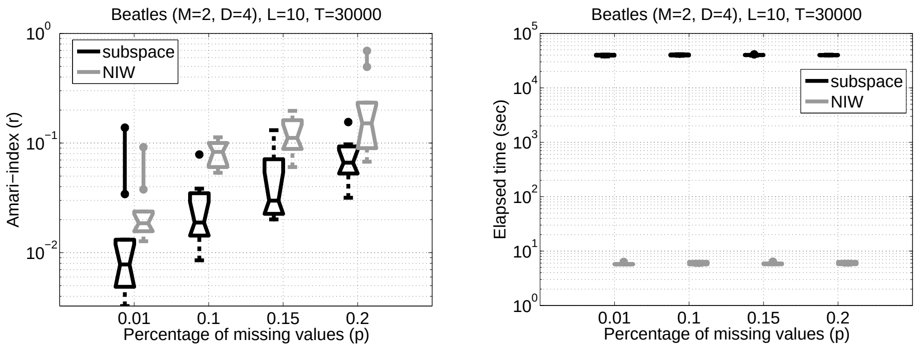


Fig. 2: Illustration of the estimations on the *3D-geom* and *ABC* datasets. (a), (b): Amari-index and elapsed time, respectively as a function of the probability of missing observation (p) for the *3D-geom* dataset on log-log scale and for AR order $L = 1$ and sample number $T = 5000$. (c)-(d): Amari-index for the *ML* method for $p = 0.2$ and for $p = 0.3$ as a function of the AR order for the *ABC* test. (e)-(j): illustration of the estimation for the *ML* method: $L = 1$, $T = 5,000$, $\lambda = 0.9$; (e) observation before mapping $\mathcal{M}_t$ $(\mathbf{x})$. (g): estimated components ($\hat{\mathbf{e}}^m$) with average Amari-index for $p = 0.01$. (f): Hinton-diagram of matrix $\mathbf{G}$ for (g)–it is approximately a block-permutation matrix with 2×2 blocks. (h)-(j): like (g), but for $p = 0.1$, $p = 0.2$, and $p = 0.3$, respectively.



(a) (b)

Fig. 3: Illustration of the *subspace* and the *NIW* methods for the *Beatles* dataset for sample number $T = 30,000$ and AR order $L = 10$. (a): Amari-index as a function of the rate of missing observations p on log-log scale, (b): elapsed time.

[2] Andrzej Cichocki and Shun-ichi Amari. *Adaptive blind signal and image processing.* John Wiley & Sons, 2002.

[3] Jean-François Cardoso. Multidimensional independent component analysis. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, volume 4, pages 1941–1944, 1998.

[4] Zoltán Szabó. *Separation Principles in Independent Process Analysis.* PhD thesis, Eötvös Loránd University, Budapest, 2009.

[5] Kwokleung Chan, Te-Won Lee, and Terrence J. Sejnowski. Variational Bayesian learning of ICA with missing data. *Neural Computation*, 15(8):1991 – 2011, 2003.

[6] A. Taylan Cemgil, Cédric Févotte, and Simon J. Godsill. Variational and stochastic inference for Bayesian source separation. *Digital Signal Processing*, 17:891–913, 2007.

[7] V.V. Petrov. Central limit theorem for m-dependent variables. In *Proceedings of the All-Union Conference on Probability Theory and Mathematical Statistics*, pages 38–44, 1958.

[8] Fabian J. Theis. Uniqueness of complex and multidimensional independent component analysis. *Signal Processing*, 84(5):951–956, 2004.

[9] Jaime Terceiro Lomba. *Estimation of Dynamic Econometric Models with Errors in Variables*, volume 339 of *Lecture notes in economics and mathematical systems.* Berlin; New York:Springer-Verlag, 1990.

[10] Alfredo García-Hiernaux, José Casals, and Miguel Jerez. Fast estimation methods for time series models in state-space form. *Journal of Statistical Computation and Simulation*, 79(2):121–134, February 2009. DOI: 10.1080/00949650701617249.

[11] K. Rao Kadiyala and Sune Karlsson. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12:99–132, 1997.

[12] Zoltán Szabó, Barnabás Póczos, and András Lőrincz. Undercomplete blind subspace deconvolution. *Journal of Machine Learning Research*, 8:1063–1095, 2007.

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

[14] Francis R. Bach and Michael I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.