Autoregressive Independent Process Analysis with Missing Observations

Zoltán Szabó

Department of Software Technology and Methodology,
Eötvös Loránd University, Budapest, Hungary
szzoli@cs.elte.hu

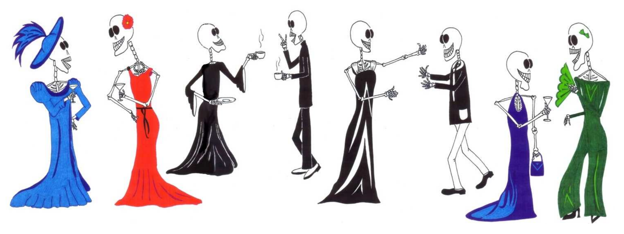
Abstract

The goal of this paper is to search for independent multidimensional 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 is independent and identically distributed (i.i.d.) in time. 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): 'cocktail party',

- we have: D speakers (sources), D microphones (sensors),
- observation: mixtures of the *independent* sources,
- task: recover the original sources from the mixed observations,



Independent Subspace Analysis (ISA):

- generalization of ICA to multidimensional source components: 'speakers can form *groups*' [1].
- one of the most fundamental hypotheses of the ICA research: ISA Separation Theorem ([1]–conjecture, [2]–proof for certain distribution types):
  $\boxed{\text{ISA = ICA + clustering}}$
  - forms the basis of the state-of-the-art ISA algorithms,
  - makes it possible to efficiently address the case of unknown source component dimensions.
- promising applications:
  - analysis of EEG, fMRI, ECG signals and gene data,
  - pattern and face direction recognition.

Goal:

- 'missing observation' case: has been addressed only for the simplest ICA model [3, 4].
- extension to:
  - multidimensional source components (ISA),
  - non-i.i.d. sources (sources have dynamics).

2. The AR-IPA Model with Missing Observations

mAR-IPA equations: observations ($\mathbf{y}$) are linear mixtures of independent AR sources ($\mathbf{x}$), available only at certain coordinates/time instants

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

where

- the driving noises $\mathbf{e}^m \in \mathbb{R}^{d_m}$ satisfy the ISA assumptions (and not $\mathbf{s}^m$),
- the unknown *mixing matrix* $\mathbf{A}$: invertible,
- the AR dynamics $\mathbf{F}[z] = \mathbf{I} - \sum_{l=0}^{L-1} \mathbf{F}_l z^{l+1}$ is *stable*: $\det(\mathbf{F}[z]) \neq 0$ for all $z \in \mathbb{C}, |z| \leq 1$,
- the $\mathcal{M}_t$ 'mask mappings': represent the coordinates and the time indices of the non-missing observations.

Goal: estimate hidden sources ($\mathbf{s}^m$) from observations $\mathbf{y}_t$.

Special cases: '$\mathcal{M}_t =$ identity and $L = 0$' = ISA, if '$\forall d_m = 1$ also holds' = ICA.

3. Method

The mAR-IPA task can be accomplished as follows:

- $\mathbf{x}$ is invertible linear transformation of the AR $\mathbf{s} \Rightarrow \mathbf{x}$ is AR with innovation $\mathbf{A}\mathbf{e}$:

$$\mathbf{x}_{t+1} = \sum_{l=0}^{L-1} \mathbf{A}\mathbf{F}_l\mathbf{A}^{-1}\mathbf{x}_{t-l} + \mathbf{A}\mathbf{e}_{t+1}. \tag{2}$$

- $\mathbf{A}\mathbf{e}$: approximately Gaussian ($\Leftarrow$ d-dependent CLT [5]).
- Estimation (separation technique):
  1. identify the partially observed AR process $\mathbf{y}_t$,
  2. estimate the independent components $\mathbf{e}^m$ from the estimated innovation by means of ISA ($\mathbf{W}_{\text{ISA}}$).

4. Illustrations

Databases:

- *ABC* dataset: $\mathbf{e}^m$s were uniform distributions on the images of the English alphabet ($d_m = 2$, $M = 3$, $D = 6$), see Fig. 1(a),
- *3D-geom* dataset: $\mathbf{e}^m$s were uniformly distributed on 3-dimensional geometric forms ($d_m = 3$, $M = 2$, $D = 6$), see Fig. 1(b),
- *Beatles* dataset: $\mathbf{s}^m$s were 8 kHz sampled portions of two stereo Beatles songs (A Hard Day's Night, Can't Buy Me Love; ($d_m = 2$, $M = 2$, $D = 4$), see http://rock.mididb.com/beatles/.

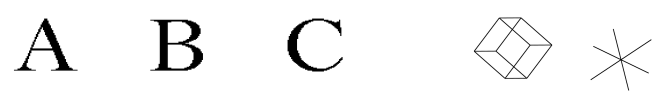


Figure 1: *(a):* ABC, *(b):* 3D-geom *database.*

Performance measure, the Amari-index: ISA ambiguities [6] $\Rightarrow$ components of the hidden sources ($\mathbf{s}^m$) can only be recovered up to

- permutation and
- invertible transformation within the subspaces.

Thus, $\mathbf{G} = \mathbf{W}_{\text{ISA}}\mathbf{A}$ is ideally a block-permutation matrix made of $d \times d$ sized blocks. This property can be measured by the Amari-index [7, 2]: $r(\mathbf{G}) \in [0, 1]$, $r(\mathbf{G}) = 0 \leftrightarrow$ perfect estimation, $r(\mathbf{G}) = 1 \leftrightarrow$ worst possible.

Simulation parameters:

- performance measure: Amari-index over 10 random runs ($\mathbf{A}$, $\mathbf{F}[z]$, $\mathbf{e}$),
- 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 $\mathcal{M}_t$),
  - $\lambda \to 1$: the (contraction) parameter of the stable polynomial matrix $\mathbf{F}[z]$ ($\mathbf{O}_i$: random orthogonal)

$$\mathbf{F}[z] = \prod_{l=1}^{L}(\mathbf{I} - \lambda\mathbf{O}_i z) \quad (|\lambda| < 1, \lambda \in \mathbb{R}), \tag{3}$$

- mixing matrix $\mathbf{A}$: random orthogonal,
- mAR fit:
  - maximum likelihood (ML) principle [8], or
  - the subspace technique [9], or
  - in a Bayesian framework using normal-inverted Wishart (*NIW*) conjugate prior [10].
- ISA subtask: using the ISA separation theorem [2];
  - ICA: FastICA [11],
  - dependence: kernel canonical correlation [12],
  - clustering: greedy, with given $d = d_m$ dimensions.

Simulations:

- performance is summarized with notched boxplots: quartiles (Q_1, Q_2, Q_3), outliers $\notin [Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$, whiskers–largest/smallest non-outliers.
- Dataset *ABC*, *3D-geom*: $L \in \{1, 2\}$, $0.1 \leq \lambda \leq 0.99$, $p \in \{0.01, 0.1, 0.15, 0.2, 0.3\}$, $T/1,000 \in \{1, 2, 5\}$.
  - Precision: *ML > subspace > NIW*, see Fig. 3(a),
  - Running times: the opposite, see Fig. 3(b),
  - Ratio of missing observations (p): *ML* ($p \leq 0.2 - 0.3$), *subspace* ($p \leq 0.15 - 0.2$), *NIW* ($p \leq 0.1 - 0.15$), see Fig. 3(a), (c)-(d),
  - *ML* is robust with respect to contraction parameter λ, see Fig. 3(c)-(j).
- Dataset *Beatles*: *subspace* and *NIW* methods, crude AR estimation of $L = 10$ ($\Leftarrow$ Schwarz's Bayesian Criterion). Results for $T = 30,000$ in Fig. 2:
  - reasonable estimations up to $p = 0.1 - 0.15$,
  - *subspace* method: more precise, but somewhat slower.

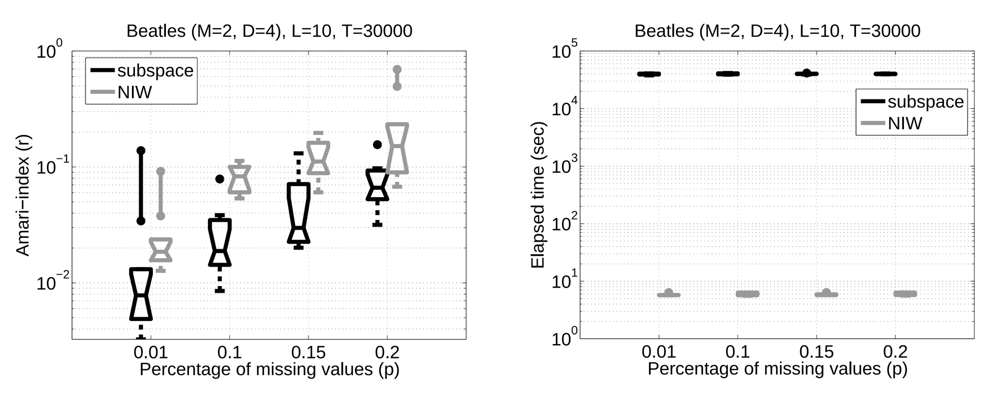


Figure 2: *Illustration on the* Beatles *test (methods:* subspace, NIW; $T = 30,000$, $L = 10$*). Left: Amari-index as a function of the rate of missing observations* p*, right: elapsed time; log scales.*

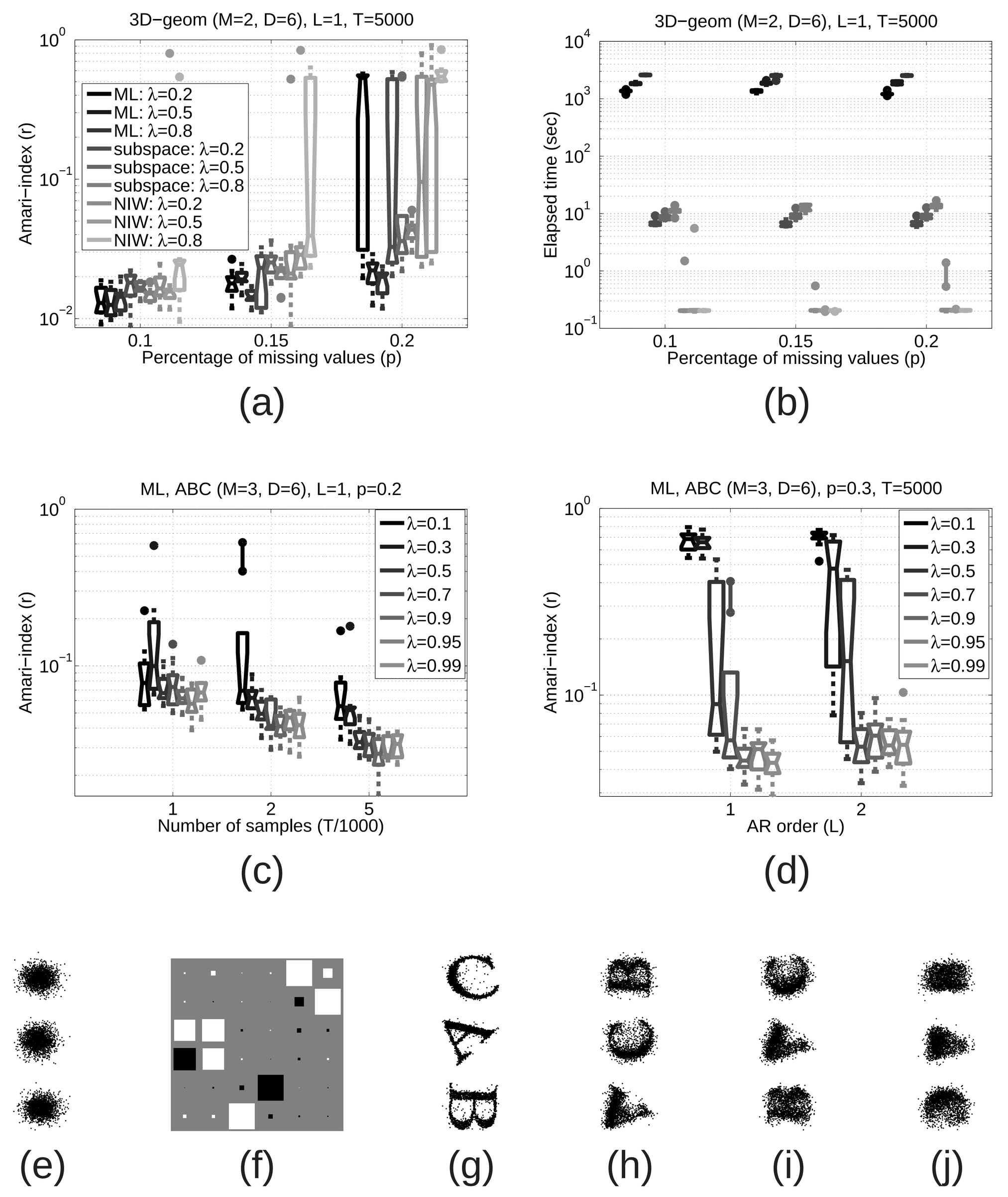


Figure 3: *Illustration on the* 3D-geom *and* ABC *datasets. (a), (b): Amari-index and elapsed time vs. rate of missing observations* p*,* 3D-geom *dataset, log scale,* $L = 1$*,* $T = 5,000$*. (c)-(d): Amari-index,* ML *method,* ABC *test,* $p = 0.2$ *and* $p = 0.3$ *vs. sample number and AR order,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

Acknowledgments.

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

References

[1] Cardoso, J.: Multidimensional independent component analysis. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98). Volume 4. (1998) 1941–1944

[2] Szabó, Z., Póczos, B., Lőrincz, A.: Undercomplete blind subspace deconvolution. Journal of Machine Learning Research **8** (2007) 1063–1095

[3] Chan, K., Lee, T.W., Sejnowski, T.J.: Variational Bayesian learning of ICA with missing data. Neural Computation **15** (2003) 1991 – 2011

[4] Cemgil, A.T., Févotte, C., Godsill, S.J.: Variational and stochastic inference for Bayesian source separation. Digital Signal Processing **17** (2007) 891–913

[5] Petrov, V.: Central limit theorem for m-dependent variables. In: Proc. of the All-Union Conference on Probability Theory and Mathematical Statistics. (1958) 38–44

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

[7] Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. Advances in Neural Information Processing Systems **8** (1996) 757–763

[8] Lomba, J.T.: 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)

[9] García-Hiernaux, A., Casals, J., Jerez, M.: Fast estimation methods for time series models in state-space form. Journal of Statistical Computation and Simulation **79** (2009) 121–134

[10] Kadiyala, K.R., Karlsson, S.: Numerical methods for estimation and inference in Bayesian VAR-models. Journal of Applied Econometrics **12** (1997) 99–132

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

[12] Bach, F.R., Jordan, M.I.: Kernel independent component analysis. Journal of Machine Learning Research **3** (2002) 1–48

ESANN 2010, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, April 28-30 2010, Bruges, Belgium

NKTH
National Office for Research and Technology