TOWARDS INDEPENDENT SUBSPACE ANALYSIS IN CONTROLLED DYNAMICAL SYSTEMS

Zoltán Szabó **András Lőrincz**
Department of Information Systems, Eötvös Loránd University,
Pázmány P. sétány 1/C, Budapest H-1117, Hungary
WWW home page: `http://nipg.inf.elte.hu`
`szzoli@cs.elte.hu, andras.lorincz@elte.hu`

ABSTRACT

In this paper we extend Independent Component Analysis (ICA) task to controlled dynamical systems. To our best knowledge this is the first work that considers the control task in this field, which may open the door for extended ICA applications. We treat Independent Subspace Analysis (ISA) task, the multidimensional generalization of ICA. In particular, we consider the identification problem of ARX models, i.e., hidden AutoRegressive dynamical systems subject to eXogenous control inputs. In our case, these ARX models are driven by independent multidimensional noise processes. The goal is the estimation of the hidden variables, that is, the parameters of the system and the driving noise. 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 'InfoMax method'. To this end, we decouple the problem into a fully observable one and an ISA task. We solve the two problems and join the results to estimate the hidden variables. Numerical examples illustrate the efficiency of our method.

Keywords: D-optimal design, ARX models, independent subspace analysis, separation principle

1 INTRODUCTION

Our goal is to connect independent component analysis and a novel result, the D-optimal identification of dynamical systems.

One can think of independent component/subspace/process analysis (ICA-Independent Component Analysis [3, 4], ISA-Independent Subspace Analysis [2], IPA-Independent Process Analysis [8]) by considering the cocktail-party problem, where we have M pieces of independent sources or source groups ($\mathbf{s} = [\mathbf{s}^1; \ldots; \mathbf{s}^M]$), but although we observe their mixture only, which in the simplest ICA/ISA case means that we observe $\mathbf{x} = \mathbf{A}\mathbf{s}$, we still would like to recover the original $\mathbf{s}$ sources. The ICA task and its extensions have been successfully applied on a number of fields, including: (i) denoising, (ii) analysis of biomedical- (EEG, fMRI, MEG, ECG) and financial data, (iii) recognition of face- and facial pose, (iv) gene analysis, (v) processing of radar/sonar data, (vi) optics, (vii) seismic exploration. The ICA problem family allows for hidden variables but makes no reference to control.

It has been shown in a recent work [6] 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/ISA/IPA problem family can model hidden independent variables, but can not account for control. We unify these directions: we treat D-optimal identification parameter and noise estimation of controlled hidden dynamical systems driven by independent sources. We solve the problem by decoupling it into a fully observable task and an ISA problem. Our method may offer important extension possibilities for ICA applications.

In Section 2 we review the D-optimal identification of fully observed ARX models. The ARX-IPA model of hidden processes is introduced in Section 3. Our solution technique for the ARX-IPA task is derived in Section 4. Illustrations are provided in Section 5.

2 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

$$\mathbf{s}_{t+1} = \sum_{i=0}^{I-1} \mathbf{F}_i \mathbf{s}_{t-i} + \sum_{j=0}^{J-1} \mathbf{B}_j \mathbf{u}_{t+1-j} + \mathbf{e}_{t+1}, \qquad (1)$$

where (i) $\mathbf{s} \in \mathbb{R}^{D_s}$, $\mathbf{e} \in \mathbb{R}^{D_e}$ ($D_s = D_e$) represent the state of the system and the noise, respectively, (ii) $\mathbf{u} \in \mathbb{R}^{D_u}$ represents the control variables, and (iii) polynomial matrix $\mathbf{F}_{\text{AR}}[z] := \mathbf{I} - \sum_{i=0}^{I-1} \mathbf{F}_i z^{i+1}$ (given by matrices $\mathbf{F}_i \in \mathbb{R}^{D_s \times D_s}$ and identity matrix $\mathbf{I}$) is stable, that is $\det(\mathbf{F}_{\mathrm{AR}}[z]) \neq 0$, for all $z \in \mathbb{C}, |z| \leq 1$. Our task is the efficient estimation of parameters $\mathbf{\Theta} = [\mathbf{F}_0, \ldots, \mathbf{F}_{I-1}, \mathbf{B}_0, \ldots, \mathbf{B}_{J-1}]$ that determine the dynamics and 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

$$J_{par}(\mathbf{u}_{t+1}) := I(\mathbf{\Theta}, \mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{s}_{t-1}, \ldots, \mathbf{u}_{t+1}, \mathbf{u}_t, \ldots),$$
$$J_{noise}(\mathbf{u}_{t+1}) := I(\mathbf{e}_{t+1}, \mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{s}_{t-1}, \ldots, \mathbf{u}_{t+1}, \mathbf{u}_t, \ldots)$$

for $\mathbf{u}_{t+1} \in U$. 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 [6], that if (i) $\mathbf{\Theta}$ has matrix Gaussian, (ii) $\mathbf{e}$ has Gaussian, and 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 $\mathbf{\Theta}$ 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 in the ARX-IPA model of the next section.

3 THE ARX-IPA MODEL

In the ARX-IPA model we assume that state $\mathbf{s}$ of the system can not be observed directly, but its linear and unknown mixture ($\mathbf{x}$) is available for observation:

$$\mathbf{s}_{t+1} = \sum_{i=0}^{I-1} \mathbf{F}_i \mathbf{s}_{t-i} + \sum_{j=0}^{J-1} \mathbf{B}_j \mathbf{u}_{t+1-j} + \mathbf{e}_{t+1}, \quad (2)$$

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t. \quad (3)$$

We assume for the *components* of $\mathbf{e} = [\mathbf{e}^1; \ldots; \mathbf{e}^M] \in \mathbb{R}^{D_e}$ $\mathbf{e}^m \in \mathbb{R}^{d_m}$ that at most one of them may be Gaussian, their temporal distributions are i.i.d. (independent identically distributed), and $I(\mathbf{e}^1; \ldots; \mathbf{e}^M) = 0$, that is, they satisfy the conditions of the ISA problem. Our task is to estimate the unknown invertible mixing matrix $\mathbf{A} \in \mathbb{R}^{D_s \times D_s}$, parameters $[\mathbf{F}_0, \ldots, \mathbf{F}_{I-1}, \mathbf{B}_0, \ldots, \mathbf{B}_{J-1}]$, $\mathbf{s}$ and $\mathbf{e}$ by means of observations $\mathbf{x}$ only.

4 IDENTIFICATION METHOD FOR ARX-IPA

There is a so-called *ISA separation principle* for the ISA task conjectured by Cardoso [2]. For some distribution types the conjecture has gained rigorous proof [10]. According to this principle, the ISA problem can be decomposed: (i) one may set aside that there are subspaces in the background and invoke a classical ICA algorithm, then (ii) cluster the estimated ICA elements into statistically dependent groups. One can reduce the estimation of multidimensional hidden components of AR processes [7], moving average (MA), ARMA, ARIMA (Integrated ARMA) processes, for real and complex variables, and also for post nonlinear mixing to the solution of the ISA task [8]. All of these estimations, however, are concerned with the *control free* case – subject to the independence of the components. Below, we solve the ARX-IPA model of Section 3, i.e., we include the control variables. Alike to other process model cases, we transform the solution of the problem to two subproblems: to that of a fully observed model (Section 2) and an ISA task.

One can apply the basis transformation rule of ARX processes and use (2) and (3) repeatedly to get

$$\mathbf{x}_{t+1} = \sum_{i=0}^{I-1} (\mathbf{A}\mathbf{F}_i\mathbf{A}^{-1})\mathbf{x}_{t-i} + \sum_{j=0}^{J-1} (\mathbf{A}\mathbf{B}_j)\mathbf{u}_{t+1-j} + (\mathbf{A}\mathbf{e}_{t+1}) \quad (4)$$

According to the d-central limit theorem [5] $\mathbf{A}\mathbf{e}_{t+1}$ is approximately Gaussian and thus the parameters $([\mathbf{A}\mathbf{F}_0\mathbf{A}^{-1}, \ldots, \mathbf{A}\mathbf{F}_{I-1}\mathbf{A}^{-1}, \mathbf{A}\mathbf{B}_0, \ldots, \mathbf{A}\mathbf{B}_{J-1})])$ and the noise ($\mathbf{A}\mathbf{e}_{t+1}$) of process $\mathbf{x}$ can be estimated by means of the D-optimality principle that assumes a fully observed process. The estimated noise can be seen as the observation of an ISA problem because components of $\mathbf{e}$ are independent. ISA techniques can be used to identify $\mathbf{A}$ and then from the estimated parameters of process $\mathbf{x}$, the estimations of $\mathbf{F}_i$ and $\mathbf{B}_j$ follow.

5 ILLUSTRATIONS

The D-optimal ARX-IPA identification algorithm of Section 4 is illustrated below. Test cases are introduced in Section 5.1. The quality of the solutions will be measured by the Amari-index (Section 5.2). Numerical results are provided in Section 5.3.

5.1 Databases

Three databases were defined to study our algorithm. The databases are illustrated in Fig. 1. In the *3D-geom* test $\mathbf{e}^m$s were random variables uniformly distributed on 3-dimensional geometric forms ($d_m = 3$). We chose 6 different components ($M = 6$) and, as a result, the dimension of the hidden source $\mathbf{e}$ is $D_e = 18$. In the *ABC* database, hidden sources $\mathbf{e}^m$ were uniform distributions defined by 2-dimensional images ($d = 2$) of the English alphabet. M was 10 (A-J). The *celebrities* test has 2-dimensional source components generated from cartoons of celebrities ($d_m = 2$, $M = 10$).[1] Sources $\mathbf{e}^m$ were generated by sampling 2-dimensional coordinates proportional to the corresponding pixel intensities. In other words, 2-dimensional images of celebrities were considered as density functions. $M = 10$ was chosen.

5.2 Performance measure, the Amari-index

Recovery of components $\mathbf{e}^m$ are subject to the ambiguities of ISA task. Namely, components of equal dimension can be recovered up to the permutation (of equal dimension) and invertible transformation within the subspaces [11]. Let us suppose, that all components are d-dimensional ($d = d_m$, $\forall m$). Then, in the ideal case, the product of the estimated ISA demixing matrix

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

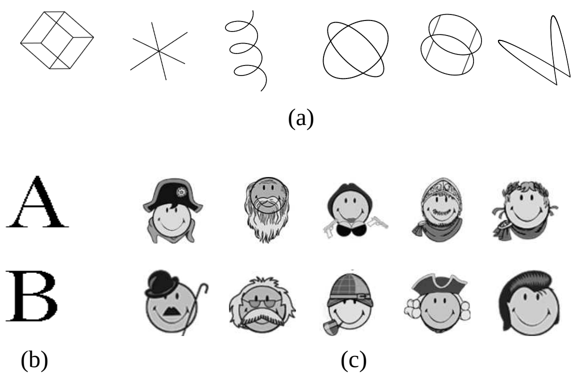


Figure 1: Illustration of the test datasets. (a): *3D-geom*, (b): *ABC* (first two components), (c): *celebrities* set.

$\hat{\mathbf{W}}_{\text{ISA}}$ and mixing matrix $\mathbf{A}$ is a block-permutation matrix $\mathbf{G} := \hat{\mathbf{W}}_{\text{ISA}}\mathbf{A} \in \mathbb{R}^{D_e \times D_e}$ made of $d \times d$ sized blocks. This block-permutation property can be measured by the *Amari-index*. Namely, let matrix $\mathbf{G}$ be decomposed into $d \times d$ blocks: $\mathbf{G} = \left[\mathbf{G}^{ij}\right]_{i,j=1,\ldots,M}$. Let g^{ij} denote the sum of the absolute values of the elements of matrix $\mathbf{G}^{ij} \in \mathbb{R}^{d \times d}$. The Amari-error [1] can be adapted to the ISA task [12] that we normalize to interval [0,1] to get the *Amari-index*:

$$r(\mathbf{G}) := \frac{1}{2M(M-1)} \left[\sum_{i=1}^{M} \left(\frac{\sum_{j=1}^{M} g^{ij}}{\max_j g^{ij}} - 1 \right) + \sum_{j=1}^{M} \left(\frac{\sum_{i=1}^{M} g^{ij}}{\max_i g^{ij}} - 1 \right) \right]. \quad (5)$$

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. $r(\mathbf{G}) = 1$ is in the worst case, i.e, when all the elements of $\mathbf{G}$ are equal in absolute value.

5.3 Simulations

Results on databases *3D-geom*, *ABC* and *celebrities* are provided here. The Amari-index was used to measure the performance (the precision of estimation of $\hat{\mathbf{e}}$) of the ARX-IPA method of Section 4.[2] For each individual parameter, 20 random runs were averaged. Our parameters are: T, the sample number of observations $\mathbf{x}(t)$, I, the order of dynamics of the AR part, J, the temporal memory of the effect of the control applied, δ_u, the upper limit of the magnitude of the control ($U := \{\mathbf{u} : max_i |u_i| \leq \delta_u\}$), and λ, parameter of the stable $\mathbf{F}_{\text{AR}}[z]$. 'Random run' means random choice of quantities $\mathbf{F}_{\text{AR}}[z]$, $\mathbf{H}_j$s, $\mathbf{A}$ and $\mathbf{e}$. In each simulation $\mathbf{A}$ was a random orthogonal matrix, sample number T varied between $1,000$ and $100,000$, we optimized J_{pars} and J_{noise} on intervals $[1, T/2]$ and $[T/2+1, T]$, respectively (see footnote 2), the dimension of the control was equal to the the dimension of $\mathbf{s}$ ($D_u = D_s$), the ISA task was solved by using the JFD method [9], the elements of matrices $\mathbf{H}_j$ were generated independently from standard Gaussian distributions, and stable $\mathbf{F}_{\text{AR}}[z]$ was generated as follows

$$\mathbf{F}_{\text{AR}}[z] = \prod_{i=0}^{I-1} (\mathbf{I} - \lambda \mathbf{O}_i z) \quad (|\lambda| < 1, \lambda \in \mathbb{R}), \quad (6)$$

where matrices $\mathbf{O}_i \in \mathbb{R}^{D_s \times D_s}$ were random orthogonal.

We sum up our experiences about the ARX-IPA method here:

1. Dependence on δ_u: We studied the effect of the magnitude of control (δ_u) on the precision of the estimation for 'small' I, J ($I, J \leq 3$) values and for $\lambda = 0.95$. We found that for a range of not too large control values δ_u the estimation is precise (Fig. 2(a)) and the error follows a power law in the number of samples: $r(T) \propto T^{-c}$ ($c > 0$) is a straight line on log-log scale. Similar results were found for all three databases in all experiments (Fig. 2(b)). Figure 3 illustrates the results of the estimations. In the rest of the studies we fixed the maximum of the control magnitude to $\delta_u = 0.2$ and show the results of the *3D-geom* database.

2. Dependence on J: Increasing the temporal memory of the effect of the control applied ($J = 3, 5, 10, 20, 50$) precise estimation was found even for $J = 50$. The estimation error is shown in Fig. 4(a).

3. Dependencies on I and λ: We found that the order of the dynamics of the AR process (I) can be increased provided that λ in Eq. (6) is decreased: For $J = 1$ and for $I = 5$, 10, 20, 50, the estimation is precise up to values approximately equal to $\lambda = 0.85 - 0.9$, $0.65 - 0.7$, $0.45 - 0.5$, $0.25 - 0.3$, respectively. Results are depicted in Fig. 4(b).

ACKNOWLEDGEMENTS

This work has been supported by the NKTH.

References

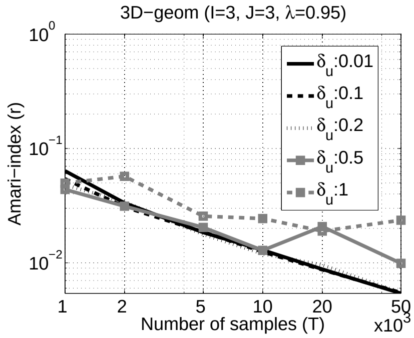
[1] S. Amari, A. Cichocki, and H.H. Yang. A new learning algorithm for blind signal separation. *NIPS'96*, 8:757–763.

[2] J.-F. Cardoso. Multidimensional independent component analysis. In *ICASSP'98*, volume 4, pages 1941–1944.

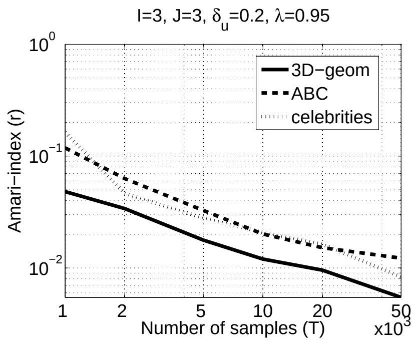
[3] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

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

[2] We note that the InfoMax objectives J_{par} and J_{noise} look forward only by one-step, so the method is greedy. The objective could be extended to include long-term cumulated contributions, but the solution is not yet known for this task. According to experiences, estimation of noise $\mathbf{e}$ can proceed by using J_{par} first for a some iterations and then use J_{noise} to compute the control values [6].

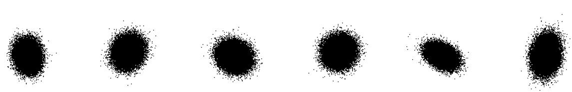


(a)

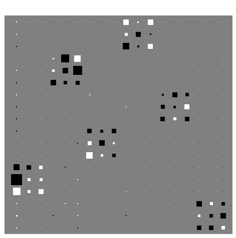


(b)

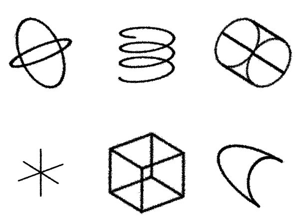
Figure 2: Estimation error (Amari-index) as a function of sample number on log-log scale for different control magnitudes (a), and databases (b).



(a)

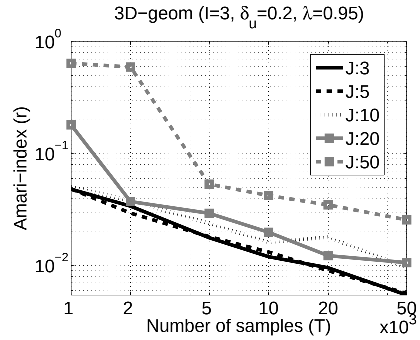


(b)

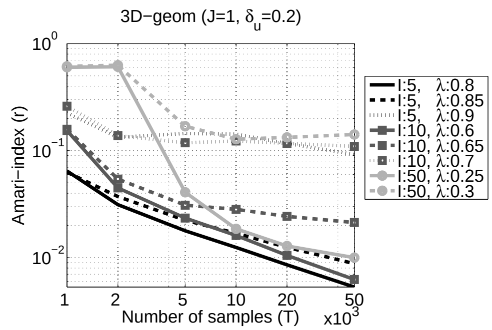


(c)

Figure 3: Illustration for the *3D-geom* database ($I = J = 3$, $\delta_u = 0.2$, $\lambda = 0.95$, $T = 50,000$), for an estimation with average estimation error (100 · Amari-index = 0.55%). (a): observed signal $\mathbf{x}(t)$. (b) Hinton-diagram: the product of the estimated demixing matrix and the mixing matrix of the derived ISA task (= approximately block-permutation matrix with 3×3 blocks). (c): estimated components–recovered up to the ISA ambiguities.



(a)



(b)

Figure 4: Estimation error (Amari-index) as a function of (a) temporal memory of control J, and (b) order of the AR process I.

[5] 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.

[6] B. Póczos and A. Lőrincz. D-optimal Bayesian interrogation for parameter and noise identification of recurrent neural networks. 2008. (submitted; available at `http://arxiv.org/abs/0801.1883`).

[7] B. Póczos, B. Takács, and A. Lőrincz. Independent subspace analysis on innovations. In *ECML 2005*, volume 3720 of *LNAI*, pages 698–706.

[8] Z. Szabó. *Separation Principles in Independent Process Analysis*. PhD thesis, Eötvös Loránd University, Budapest, 2008. (submitted; available at `http://nipg.inf.elte.hu/index.php?option=com_remository&Itemid=27&func=fileinfo&id=153`).

[9] Z. Szabó and A. Lőrincz. Real and complex independent subspace analysis by generalized variance. In *ICARN 2006*, pages 85–88.

[10] Z. Szabó, B. Póczos, and A. Lőrincz. Undercomplete blind subspace deconvolution. *JMLR*, 8:1063–1095, 2007.

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

[12] F. J. Theis. Blind signal separation into groups of dependent signals using joint block diagonalization. In *ISCAS 2005*, volume 6, pages 5878–5881.