Cross-Entropy Optimization for Independent Process Analysis

Zoltán Szabó, Barnabás Póczos, and András Lőrincz

Department of Information Systems

Eötvös Loránd University, Budapest, Hungary

szzoli@cs.elte.hu, pbarn@cs.elte.hu, lorincz@inf.elte.hu

Abstract

We treat the problem of searching for hidden multi-dimensional independent auto-regressive processes. First, we transform the problem to Independent Subspace Analysis (ISA). Our main contribution concerns ISA. We show that under certain conditions, ISA is equivalent to a combinatorial optimization problem. For the solution of this optimization we apply the cross-entropy method. Numerical simulations indicate that the cross-entropy method can provide considerable improvements over other state-of-the-art methods.

1. The IPA Model

1.1 The IPA Equations

THE IPA (Independent Process Analysis) model is

$$\mathbf{s}^m(t+1) = \mathbf{F}^m\mathbf{s}^m(t) + \mathbf{e}^m(t), \quad m = 1, \ldots, M \tag{1}$$
$$\mathbf{z}(t) = \mathbf{A}\mathbf{s}(t). \tag{2}$$

Here: the unknown *mixing matrix* $\mathbf{A} \in \mathbb{R}^{D\times D}$, the hidden *components* $\mathbf{s}^m \in \mathbb{R}^d$, and $\mathbf{s}(t) := [\mathbf{s}^1(t); \ldots; \mathbf{s}^M(t)] \in \mathbb{R}^D$.
Goal of IPA: estimate $\mathbf{s}(t)$ and $\mathbf{A}$ (or $\mathbf{W} := \mathbf{A}^{-1}$: *separation matrix*) by using observations $\mathbf{z}(t)$ only. Specially: (i) ISA ($\forall \mathbf{F}^m = \mathbf{0}$), (ii) Independent Component Analysis (ICA), when $\forall \mathbf{F}^m = \mathbf{0}$ and $d = 1$.

1.2 Assumptions

- $\mathbf{e}^m(t)$ is i.i.d. in t, $\mathbf{e}^i(t)$ is independent from $\mathbf{e}^j(t)$, if $i \neq j$
- $\mathbf{F}^m$s correspond to stable AR processes
- $\mathbf{A}$: invertible
- *whitened* noise process $\mathbf{e}(t)$ and orthogonal $\mathbf{A}$ [without loss of generality (invertible $\mathbf{A}$, innovation trick)], that is

$$E[\mathbf{e}(t)] = \mathbf{0}, E\left[\mathbf{e}(t)\mathbf{e}(t)^T\right] = \mathbf{I}_D, \quad \forall t, \tag{3}$$
$$\mathbf{I}_D = \mathbf{A}\mathbf{A}^T. \tag{4}$$

1.3 Uncertainties of the IPA Model

- IPA identification ambiguities, alike to ICA and ISA
- IPA $\xrightarrow{\text{innovation trick [1, 2, 3]}}$ ISA, where the *innovation* of a stochastic process $\mathbf{u}(t)$ is

$$\tilde{\mathbf{u}}(t) := \mathbf{u}(t) - E[\mathbf{u}(t)|\mathbf{u}(t-1), \mathbf{u}(t-2), \ldots]. \tag{5}$$

For an AR process, the innovation is identical to the noise that drives the process $\Rightarrow$ IPA model [$\mathbf{F} := blockdiag(\mathbf{F}^1, \ldots, \mathbf{F}^M)$]:

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{e}(t), \tag{6}$$
$$\mathbf{z}(t) = \mathbf{A}\mathbf{F}\mathbf{A}^{-1}\mathbf{z}(t-1) + \mathbf{A}\mathbf{e}(t-1), \tag{7}$$
$$\tilde{\mathbf{z}}(t) = \mathbf{A}\mathbf{e}(t-1) = \mathbf{A}\tilde{\mathbf{s}}(t). \tag{8}$$

- Concerning the ISA task, if $\mathbf{s}$ and $\mathbf{z}$ are white, then
  - lessened ISA ambiguities: (i) permutation of the components, (i) orthogonal transformation within subspaces,
  - $\mathbf{W}$ is orthogonal.

Identification ambiguities of the ISA task are detailed in [4].

2. The ISA Separation Theorem

ISA task $\Leftrightarrow$ minimization of mutual information between the components $\Leftrightarrow$

$$J(\mathbf{W}) := \sum_{m=1}^{M} H(\mathbf{y}^m) \longrightarrow \min_{\mathbf{W}\in\mathbb{R}^{D\times D}:\text{ orthogonal}} \tag{9}$$

Here, (i) $\mathbf{y} = \mathbf{W}\mathbf{z} = \left[\mathbf{y}^1; \ldots; \mathbf{y}^M\right]$, $\mathbf{y}^m$ are the estimated components and (ii) H is Shannon's (multi-dimensional) differential entropy. Our main result:

Theorem 1 (Separation theorem for ISA) *Let us suppose, that all the* $\mathbf{u} = [u_1; \ldots; u_d] = \mathbf{s}^m$ *components of source* $\mathbf{s}$ *in the ISA task satisfy*

$$H\left(\sum_{i=1}^{d} w_i u_i\right) \geq \sum_{i=1}^{d} w_i^2 H(u_i), \forall \mathbf{w} : \sum_{i=1}^{d} w_i^2 = 1. \tag{10}$$

Assuming that $\mathbf{W}_{\text{ICA}}(\mathbf{z})$ *is unique (up to permutation and sign of the components), then it is* $\mathbf{W}_{\text{ISA}}(\mathbf{z})$ *(up to permutation and sign of the components). In other words*

$$\mathbf{W}_{\text{ISA}} = \mathbf{P}\mathbf{W}_{\text{ICA}}, \tag{11}$$

where $\mathbf{P}\left(\in \mathbb{R}^{D\times D}\right)$ *is a permutation matrix to be determined. (Proof in [5], e.g., for elliptically symmetric sources)*

$\Rightarrow$ IPA estimation steps:

1. observe $\mathbf{z}(t)$ and estimate the AR model,
2. whiten the innovation of the AR process and perform ICA on it,
3. solve the combinatorial problem: search for the permutation of the ICA sources that minimizes the cost J.

Thus IPA needs only two (more) steps: (i) $\hat{H}$, and (ii) optimization of J in S_D (permutations of length D).

3. Assistants

3.1 Multi-dimensional Entropy Estimation by the k-nearest Neighbor Method

Entropy estimation (similar to [3]) based on k-nearest neighbors [6, 7]: asymptotically unbiased and strongly consistent [6]. Basic idea:

$$\hat{H}(\{\mathbf{u}_1, \ldots, \mathbf{u}_T\}, k, \gamma) \xrightarrow{T\to\infty} H_\alpha(\mathbf{u}) + c, \tag{12}$$
$$H_\alpha(\mathbf{u}) \xrightarrow{\alpha\to 1 \quad (\gamma\to 0)} H(\mathbf{u}), \tag{13}$$

where (i) $\mathbf{u}(1), \ldots, \mathbf{u}(T)$ is an i.i.d. sample from the distribution of $\mathbf{u} \in \mathbb{R}^d$, (ii) H_α denotes Rényi's α-entropy and (iii) $\alpha := \frac{d-\gamma}{d}$. [3]: (i) only IPA algorithm at present (to our best knowledge), (ii) Jacobi rotations for pairs, after ICA preprocessing (ICA-Jacobi).

3.2 Cross-Entropy Method for Combinatorial Optimization

For permutation search ($\mathbf{P}$) CE [8] technique, cost function

$$J : \mathbf{x} \in S_D \rightarrow J(\mathbf{P}_\mathbf{x}\mathbf{W}_{ICA}), \tag{14}$$

where $\mathbf{P}_\mathbf{x}$ is the permutation matrix associated to $\mathbf{x}$.
Our method is similar to the Travelling Salesman Problem (TSP) solved by CE: travel cost $\leftrightarrow J(\mathbf{x})\} \Rightarrow$ *ICA-TSP*.

4. Numerical Studies

4.1 Databases

Four databases (as the innovation of the hidden processes), three in Fig. 1, the fourth:

- uniform $u_i(t)$ coordinates ($i = 1, \ldots, k$) on $\{0, \ldots, k-1\}$,
- $u_{k+1} := mod(u_1 + \ldots + u_k, k)$.

$\Rightarrow$ every k-element subset of $\{u_1, \ldots, u_{k+1}\}$ is made of independent variables; *all-k-independent* problem [9], in our simulations $M = 5$ and $d = k + 1 = 4$.

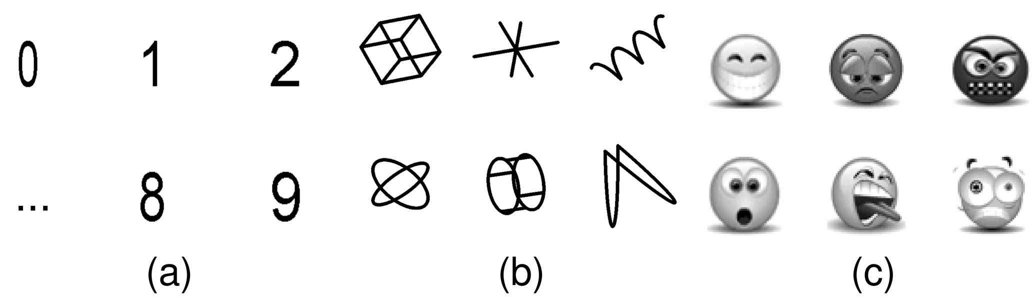

Figure 1: *3 test databases: densities of* $\mathbf{e}^m$. *Each object represents a probability density. Left: numbers:* $10 \times 2 = 20$*-dimensional problem, uniform distribution on the images of numbers. Middle: 3D-geom:* $6 \times 3 = 18$*-dimensional problem, uniform distribution on 3-dimensional geometric objects. Right: smiley: 6 basic facial expressions [10], non-uniform distribution defined in 2 dimensions,* $6 \times 2 = 12$*-dimensional problem.*

In the test examples:

- entropy estimation: $k = 3$, $\gamma = 0.01$
- dimensions: $D = 12, 18, 20$ and $d = 2, 3, 4$
- sample number: $T = 300, 400, \ldots, 1500$
- measure of goodness: *normalized* Amari-distance (r, average of 10 computer runs) $\rightarrow$ measure of block-permutation property.
  That is, for matrix $\mathbf{B} \in \mathbb{R}^{D\times D}$: (i) $0 \leq r(\mathbf{B}) \leq 1$, and (ii) $r(\mathbf{B}) = 0 \Leftrightarrow \mathbf{B}$ is a block-permutation matrix with $d \times d$ sized blocks ($\Leftrightarrow 0$ for optimal IPA estimation: $\mathbf{B} := \mathbf{W}\mathbf{A}$).

4.2 Results and Discussion

- ICA-Jacobi: exhaustive search for all Jacobi pairs with 50 angles in $[0, \pi/2]$ several times until convergence
- Still, ICA-TSP is superior in all of the studied examples.
- Quantitative results in Table 1, innovations estimated by the ICA-TSP method on facial expressions in Fig. 2.
- Greedy ICA-Jacobi method seems to be similar or sometimes inferior to the global ICA-TSP, in spite of the much smaller search space available for the latter.
- Simulations indicate that conditions of the 'Separation Theorem' may be too restrictive.

- Non-combinatorial IPA approach (based on the Separation Theorem) in [11].

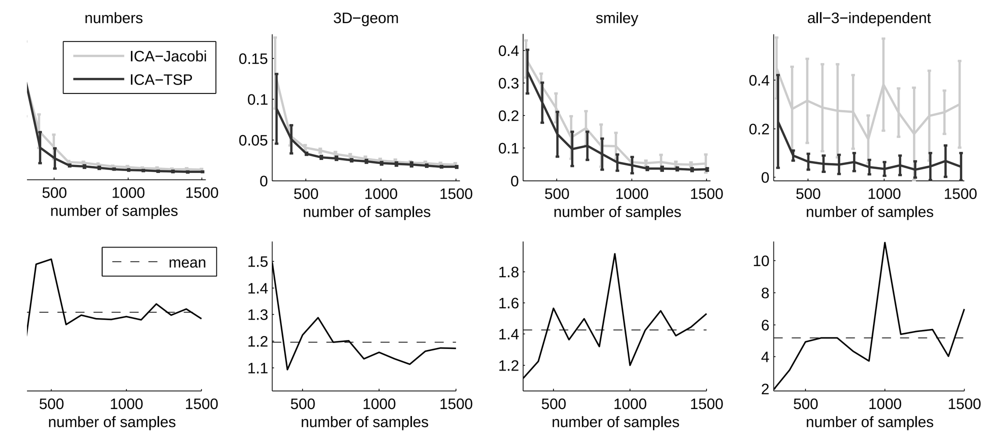

Figure 2: *Mean±standard deviation of* $r(T)$ *(upper row). Gray: ICA-Jacobi, black: ICA-TSP. In the lower row, black: relative precision of estimation, dashed: average over the different sample numbers. Columns from left to right correspond to databases 'numbers', '3D-geom', 'smiley', 'all-3-independent', respectively.*

Table 1: *Average normalized Amari-errors (in* $100 \cdot r\%\pm$ *standard deviation, for* $T = 1500$*) and precision of the ICA-TSP relative to that of ICA-Jacobi in sample domain* $300 - 1500$.

Database	ICA-Jacobi	ICA-TSP	Improvement (min-mean-max)
numbers	3.06% (± 0.22)	2.40% (± 0.11)	1.03 - 1.30 - 1.54
3D-geom	1.99% (± 0.17)	1.69% (± 0.10)	1.09 - 1.20 - 1.50
smiley	5.26% (± 2.76)	3.44% (± 0.36)	1.16 - 1.43 - 1.92
all-3-indep.	30.05% (± 17.90)	4.31% (± 5.61)	1.96 - 5.18 - 11.12

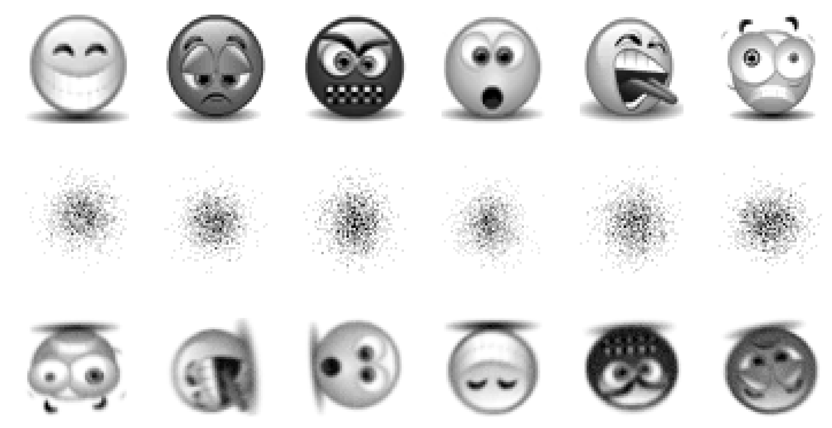

Figure 3: *Illustration of the ICA-TSP algorithm on the 'smiley' database. Upper row: density function of the sources (using* 10^6 *data points). Middle row: 1,500 samples of the observed mixed signals (*$\mathbf{z}(t)$*). The ICA-TSP algorithm works on these data. Lower row: Estimated separated sources (recovered up to permutation and orthogonal transformation).*

References

[1] Hyvärinen, A.: Independent Component Analysis for Time-dependent Stochastic Processes. In: ICANN 1998. (1998) 541–546

[2] Cheung, Y., Xu, L.: Dual Multivariate Auto-Regressive Modeling in State Space for Temporal Signal Separation. IEEE Tr. on Syst. Man Cyb. B **33** (2003) 386–398

[3] Póczos, B., Takács, B., Lőrincz, A.: Independent Subspace Analysis on Innovations. In: ECML, Porto, Portugal. (2005) 698–706

[4] Theis, F.J.: Uniqueness of Complex and Multidimensional Independent Component Analysis. Signal Proc. **84** (2004) 951–956

[5] Szabó, Z., Póczos, B., Lőrincz, A.: Separation Theorem for Independent Subspace Analysis. Technical report, Eötvös Loránd University, Budapest (2005) http://people.inf.elte.hu/lorincz/Files/TR-ELU-NIPG-31-10-2005.pdf.

[6] Yukich, J.E.: Probability Theory of Classical Euclidean Optimization Problems. Volume 1675 of Lecture Notes in Math. Springer-Verlag, Berlin (1998)

[7] Costa, J.A., Hero, A.O.: Manifold Learning Using k-Nearest Neighbor Graphs. In: ICASSP, Montreal, Canada. (2004)

[8] Rubinstein, R.Y., Kroese, D.P.: The Cross-Entropy Method. Springer (2004)

[9] Póczos, B., Lőrincz, A.: Independent Subspace Analysis Using Geodesic Spanning Trees. In: ICML. (2005) 673–680

[10] Ekman, P.: Emotion in the Human Face. Cambridge Univ. Press, New York (1982)

[11] Póczos, B., Lőrincz, A.: Non-combinatorial Estimation of Independent Autoregressive Sources. Neurocomputing Letters (2006) accepted.

ICA 2006, 6th International Conference on Independent Component Analysis and Blind Source Separation, March 5-8 2006, Charleston, SC, USA