Towards Independent Subspace Analysis in Controlled Dynamical Systems

Zoltán Szabó, András Lőrincz

Neural Information Processing Group,
Department of Information Systems,
Eötvös Loránd University,
Budapest, Hungary

ICA Research Network 2008

Acknowledgements:

NKTH
National Office for Research and Technology

Tools to Integrate-1 (Independent Subspace Analysis)

- Cocktail party problem
- Generalization of ICA:
  - *multidimensional* components,
  - groups of ’people/music bands’
- Hidden, independent, multidimensional processes – NO CONTROL.

Tools to Integrate-2 (D-optimal Identification of Dynamical Systems)

- Problem: estimate the parameters of a fully observable controlled dynamical system by the ‘optimal’ choice of the control.
  - ’Parameters’: dynamics, noise.
  - ’Optimal’: in information theoretical sense $\rightarrow$ D-optimality.
- Synonyms: active learning, optimal experimental design.
- For ARX models: QP in Bayesian framework.
- Controlled dynamical system – FULLY OBSERVABLE.

Motivation

- Goal: integrate the former methodologies
  - hidden multidimensional sources,
  - optimal design in controlled systems.

Motivation

- Goal: integrate the former methodologies
  - hidden multidimensional sources,
  - optimal design in controlled systems.
- EEG data analysis: recognition + prediction, e.g., epileptic

Motivation

- Goal: integrate the former methodologies
  - hidden multidimensional sources,
  - optimal design in controlled systems.
- EEG data analysis: recognition + prediction, e.g., epileptic

- EU-FP7: interested in partnership

Contents

Independent Subspace Analysis (ISA/MICA)

- ISA equations: Observation $\mathbf{x}$ is linear mixture of independent multidimensional *components*:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),$$
$$\mathbf{s}(t) = [\mathbf{s}^1(t); \ldots ; \mathbf{s}^M(t)],$$

  where
  - $\mathbf{s}^m(t) \in \mathbb{R}^{d_m}$ are i.i.d. sampled random variables in time,
  - $I(\mathbf{s}^1, \ldots, \mathbf{s}^M) = 0$,
  - mixing matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ is invertible, with $D := dim(\mathbf{s})$.
- Goal: $\hat{\mathbf{s}}$. Specially for $\forall d_m = 1$: ICA.
- Ambiguities: permutation, linear (/orth.) transformation.

D-optimal ARX Identification

- Observation equation (ARX model; **u**: control, **e**: noise):

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{B}\mathbf{u}(t+1) + \mathbf{e}(t+1).$$

- Task: 'efficient' estimation of
  - system parameters: $\boldsymbol{\Theta} = [\mathbf{F}, \mathbf{B}, parameters(\mathbf{e})]$, or
  - noise: **e**

  by the 'optimal' choice of control **u**.
- Optimality (D-optimal/'InfoMax'):

$$J_{par}(\mathbf{u}_{t+1}) := I(\boldsymbol{\Theta}, \mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{s}_{t-1}, \ldots, \mathbf{u}_{t+1}, \mathbf{u}_t, \ldots) \to \max_{\mathbf{u}_{t+1} \in U}, \text{ or}$$

$$J_{noise}(\mathbf{u}_{t+1}) := I(\mathbf{e}_{t+1}, \mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{s}_{t-1}, \ldots, \mathbf{u}_{t+1}, \mathbf{u}_t, \ldots) \to \max_{\mathbf{u}_{t+1} \in U}.$$

- Result (Póczos & Lőrincz, 2008): In the Bayesian setting, optimization of J can be reduced to QP.

D-optimal Hidden ARX Identification

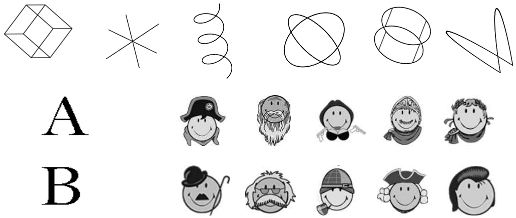
- State (**s**) + observation (**x**) equation:

$$\mathbf{s}(t+1) = \mathbf{F}\mathbf{s}(t) + \mathbf{B}\mathbf{u}(t+1) + \mathbf{e}(t+1), \qquad (1)$$
$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \qquad (2)$$

- Assumptions: $I(\mathbf{e}^1, \ldots, \mathbf{e}^M) = 0$ – hidden non-Gaussian independent multiD components
- Trick: reduce the problem to the fully observable case (d-dep. CLT) + ISA:
  - $\mathbf{x}(t+1) = [\mathbf{AFA}^{-1}]\mathbf{x}(t) + [\mathbf{AB}]\mathbf{u}(t+1) + [\mathbf{Ae}(t+1)]$,
  - **x** – 'fully observable tool' $\rightarrow$ $[\mathbf{AFA}^{-1}]$, $[\mathbf{AB}]$, $\mathbf{Ae}$,
  - **Ae** – ISA$\rightarrow$ $\mathbf{A}\}$ $\Rightarrow$ $\mathbf{F}, \mathbf{B}, \mathbf{e}$.
- Note: for higher order ARX systems the same idea holds.

Databases, Performance Measure, Questions

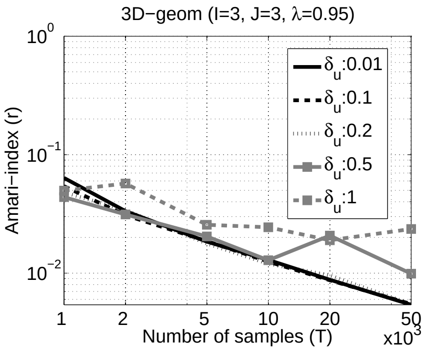
- Databases (*3D-geom*, *ABC*, *celebrities*):



- Performance measure: Amari-index (r) $\in [0, 1]$, 0–perfect.
- Questions:
  1. Dependence on $\delta_u = |U_{\text{control}}|$,
  2. Dependence on $J = deg(\mathbf{B}_{\text{control}}[z])$,
  3. Dependencies on $I = deg(\mathbf{F}_{\text{AR}}[z])$ and λ:

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

Decline of the estimation error: power-law $[r(T) \propto T^{-c}\ (c > 0)]$



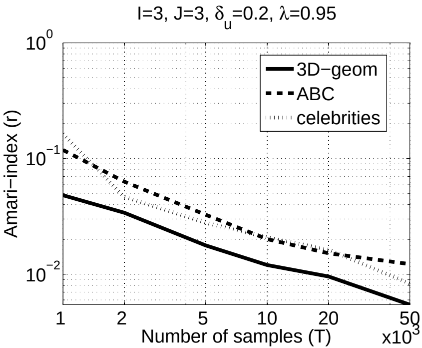


Illustration: *3D-geom*

- observation:

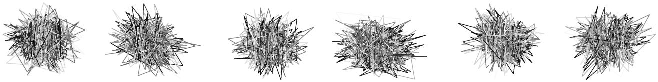
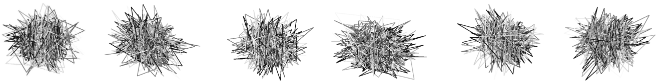


Illustration: *3D-geom*

- observation:



- estimated innovation (input of ISA):

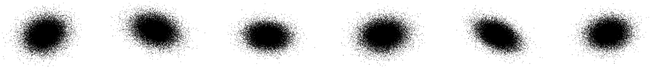
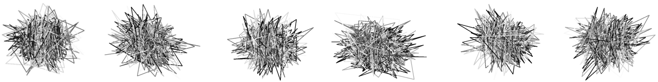


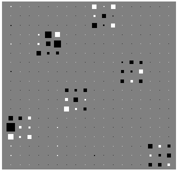
Illustration: *3D-geom*

- observation:



- estimated innovation (input of ISA):

- Hinton-diagram, estimated components:



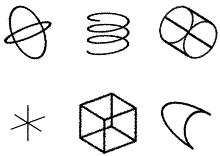
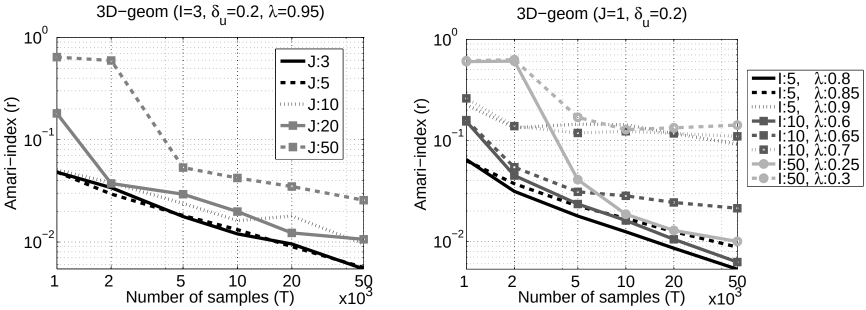


Illustration: Dependencies on $J = deg(\mathbf{B}_{\text{control}}[z])$, $I = deg(\mathbf{F}_{\text{AR}}[z])$

Precise even for $J = 50$; $I = 50$ $(\nearrow, \lambda \searrow)$



Summary

- Integration of two methodologies:
  - hidden independent multidimensional sources,
  - optimal design in controlled dynamical systems.
- Numerical experiences:
  - Decline of the estimation error follows power-law.
  - Robust against:
    - the order of the AR process,
    - temporal memory of the control.
- Possibility to apply ICA, ISA, ... in controlled systems.

TYFYA!

References

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`).

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.

Z. Szabó.
Separation Principles in Independent Process Analysis.
PhD thesis, Eötvös Loránd University, Budapest, 2008.
(submitted; available at `nipg.inf.elte.hu`: Download).