Real and Complex Independent Subspace Analysis by Generalized Variance

Zoltán Szabó, András Lőrincz

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

ICA Research Network
2006

- Cocktail party problem: groups of people / music bands
- ISA nicknames: MICA, group ICA, IVA
- $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$

$\mathbb{K}$-ISA Equations

Observation $\mathbf{z}$ is mixture of independent *components*:

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

where

- $\mathbf{s}^m(t) \in \mathbb{K}^{d_m}$ are i.i.d. sampled random variables in time,
- $\mathbf{s}^i$ is independent of $\mathbf{s}^j$, if $i \neq j$,
- mixing matrix $\mathbf{A} \in \mathbb{R}^{D \times D}$ is invertible, with $D := dim(\mathbf{s})$.

Goal: $\hat{\mathbf{s}}$. Specially for $\forall d_m = 1$: $\mathbb{R}$-ICA, $\mathbb{C}$-ICA.

Ambiguities of the $\mathbb{K}$-ISA Model

Hidden components can (only) be determined up to

- permutation, and
- invertible linear transformation within subspaces.

Whitening assumption:

$$\underbrace{\begin{array}{c} E[\mathbf{s}] = \mathbf{0}, cov[\mathbf{s}] = \mathbf{I}_D, \\ E[\mathbf{z}] = \mathbf{0}, cov[\mathbf{z}] = \mathbf{I}_D. \end{array}}$$

$$\Downarrow$$

Lessened ambiguities: invertible $\rightarrow$ orthogonal ($\mathbb{R}$) / unitary($\mathbb{C}$).

f-uncorrelatedness

- Whitening $\Leftrightarrow$ second order uncorrelatedness
- Our approach: independence is approximated as uncorrelatedness for a „lot of functions” (for $\forall \mathbf{f} \in \mathcal{F}$)
- Formally,
  1. Estimation of the hidden source in feedforward architecture:
  $$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t).$$
  2. **f**-covariance matrix is estimated empirically, after applying a function $\mathbf{f}(\in \mathcal{F})$:
  $$\mathbf{C}(\mathbf{f}, T) = \widehat{cov}[\mathbf{f}(\mathbf{y}), \mathbf{f}(\mathbf{y})],$$
  3. Cost function for $\mathbf{W}$ to make the blocks zero in $\mathbf{C}(\mathbf{f}, T)$ out of the block-diagonal, for *all* $\mathbf{f} \in \mathcal{F}$.

Cost for ISA Based on Joint f-decorrelation

- Cost function for ISA:

$$J(\mathcal{F}, T, \mathbf{W}) := \sum_{\mathbf{f}\in\mathcal{F}} \|\mathbf{M} \circ \mathbf{C}(\mathbf{f}, T, \mathbf{W})\|^2 \to \min_{\mathbf{W}:\text{orthogonal}},$$

  where **M** picks out the **f**-covariance of different subspaces.
- J should be minimized to solve the ISA problem.
- Too difficult: optimization on the Stiefel / flag manifold [Nishimori et al., 2006]!

$$\Downarrow$$

Reduce the task further.

The $\mathbb{K}$-ISA Separation Theorem

Essence:

$$\boxed{\mathbb{K}\text{-ISA} = \mathbb{K}\text{-ICA} + \text{permutation search (with } \mathbb{K}\text{-ISA cost).}}$$

Formally,

Theorem

Let H denote Shannon's differential entropy. Let us suppose that the $\mathbf{u} := \mathbf{s}^m$ components in $\mathbb{K}$-ISA satisfy

$$H\left(\sum_{i=1}^{d} w_i u_i\right) \geq \sum_{i=1}^{d} |w_i|^2 H(u_i), \quad \forall \|\mathbf{w}\|_{\mathbb{K}} = 1$$

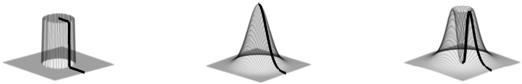
then $\mathbf{W}_{ISA} = \mathbf{P}\mathbf{W}_{ICA}$ with a $\mathbf{P} \in \mathbb{R}^{D \times D}$ permutation matrix.

References: [Cardoso, 1998], [Szabó et al., 2006a].

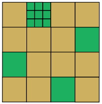
Simulations

- Test databases (can be scaled in d)
  - d-geom ($M = 4$):

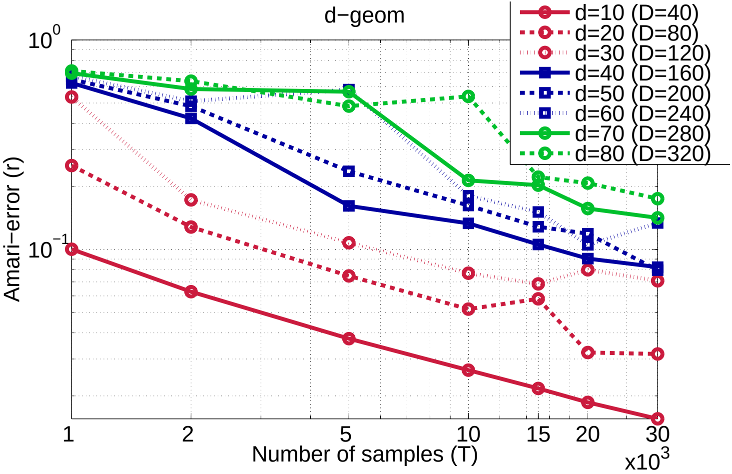
  - d-spherical ($M = 3$):



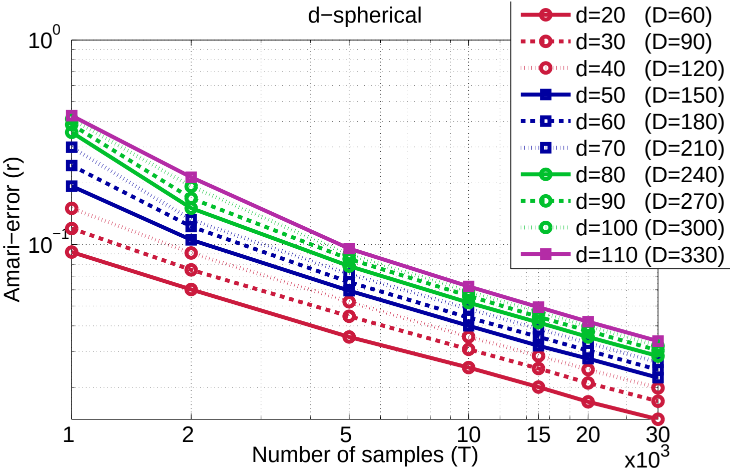
- Performance measure: *normalized* Amari-error $r \in [0, 1]$. Measures block permutation property of **WA**.



Results-1: d-geom



Results-2: *d*-spherical



Connection to Other Techniques ($d = 1$)

- Alike to KC (kernel covariance) we use a function set $\mathcal{F}$ [Gretton et al., 2003].
- For $\mathcal{F} = \{\mathbf{f}\}$:
  - $\mathbf{f}$-decorrelation is equivalent to minimization the cost

$$0 \leq Q_W(\mathbf{f}, T) := -\frac{1}{2} \log \left\{ \frac{\det[\mathbf{C}(\mathbf{f}, T)]}{\prod_{m=1}^{M} \det[\mathbf{C}^{m,m}(\mathbf{f}, T)]} \right\}.$$

    This is right the cost function of KGV (kernel generalized variance) [Bach and Jordan, 2002].
  - With RNN architecture (instead of feedforward), $\mathbb{K} = \mathbb{R}$ gives rise to self-organization ($\Leftarrow$ gradient) [Meyer-Bäse et al., 2006].

References-1

- Nishimori, Y., Akaho, S., and Plumbley, M. D. (2006).
  Riemannian optimization method on the flag manifold for independent subspace analysis.
  In *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume 3889 of *LNCS*, pages 295–302. Springer.

- Bach, F. R. and Jordan, M. I. (2002).
  Kernel independent component analysis.
  JMLR, 3:1–48.

- Gretton, A., Herbrich, R., and Smola, A. (2003).
  The kernel mutual information.
  In *IEEE ICASSP*, volume 4, pages 880–883.

- Meyer-Bäse, A., Gruber, P., Theis, F., and Foo, S. (2006).
  Blind source separation based on self-organizing neural network.
  Engineering Applications of AI, 19:305–311.

References-2

Cardoso, J. (1998).
Multidimensional independent component analysis.
In *Proceedings of ICASSP'98, Seattle, WA.*

Szabó, Z., Póczos, B., and Lőrincz, A. (2006a).
Cross-entropy optimization for independent process analysis.
In *Independent Component Analysis and Blind Signal Separation (ICA 2006)*, LNCS 3889, pages 909–916. Springer.

Szabó, Z., Póczos, B., and Lőrincz, A. (2006b).
Separation theorem for $\mathbb{K}$-independent subspace analysis with sufficient conditions.
Technical report, Eötvös Loránd University, Budapest.
http://arxiv.org/abs/math.ST/0608100.

Summary

- Presented $\mathbb{K}$-ISA method: joint decorrelation on function set $\mathcal{F}$:
  - with feedforward architecture:
    - Reduction using the $\mathbb{K}$-ISA Separation Theorem

      $\mathbb{K}$-ISA = $\mathbb{K}$-ICA + permutation search (with $\mathbb{K}$-ISA cost)
    - $\Leftrightarrow$ KGV for $\mathcal{F} = \{\mathbf{f}\}$.
  - with recurrent architecture: self-organization.
- First step toward large scale problems:
  - few hundred dimensions,
  - „power law" decrease of estimation error.

Thank you for the paying attention!