Post Nonlinear Independent Subspace Analysis

Zoltán Szabó, Barnabás Póczos,
Gábor Szirtes and András Lőrincz

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

ICANN 2007

Post Nonlinear Independent Subspace Analysis

Cocktail party problem:

- independent groups of people / music bands,
- nonlinear (post nonlinear) mixing.

PNL ISA Equations

- The PNL ISA model:

$$\mathbf{x}(t) = \mathbf{f}[\mathbf{A}\mathbf{s}(t)]. \tag{1}$$

- Assumptions ($\mathbf{s} = [\mathbf{s}_1; \ldots; \mathbf{s}_M] \in \mathbb{R}^{Md} = \mathbb{R}^D$):
  - source $\mathbf{s}$ is *d-independent*: $I(\mathbf{s}_1, \ldots, \mathbf{s}_M) = 0$,
  - $\mathbf{s}(t) \in \mathbb{R}^D$ is i.i.d. in time t,
  - $\mathbf{A} \in \mathbb{R}^{D \times D}$ invertible and 'mixing', that is: $\mathbf{A} = [\mathbf{A}_{ij} \in \mathbb{R}^{d \times d}]$, $\forall\, i \Rightarrow \exists (j,k) : \mathbf{A}_{ij}$ and $\mathbf{A}_{ik}$ are invertible.
  - $\mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$ invertible, acts component-wise.
- Goal: $\hat{\mathbf{s}}$.

Ambiguites of PNL ISA

- PNL mixing structure $\Rightarrow$ mirror demixing [$\hat{\mathbf{s}} = \mathbf{W}\mathbf{g}(\mathbf{x})$].
- Question: d-independence of $\hat{\mathbf{s}}$ $\Rightarrow$ true $\mathbf{s}$ has been found?

$$\Downarrow$$

- Yes: PNL ISA separability theorem.

Separability; PNL-ISA Ambiguities with Locally-Constant Nonzero C^2 Densities

Theorem

Supposing that:

- $\mathbf{A}, \mathbf{W}$: *invertible and 'mixing' matrices,*
- $\mathbf{s}$: *(i) existing covariance matrix, (ii) somewhere locally constant, C^2 density function,*
- $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ *is a component-wise bijection, with analytical coordinate functions.*

In this case, if $\mathbf{e} := [\mathbf{e}_1; \ldots; \mathbf{e}_M] = \mathbf{Wh}(\mathbf{As})$ *is d-independent with somewhere locally constant density function, then:*

- $\mathbf{e}$ *recovers the hidden source (up to ISA ambiguities + constant translation within subspaces).*

Separability $\Rightarrow$ PNL ISA algorithm

Sketch:

1. Estimate $\mathbf{g} = \hat{\mathbf{f}}^{-1}$:

   d-dependent Central Limit Theorem

   $\Downarrow$

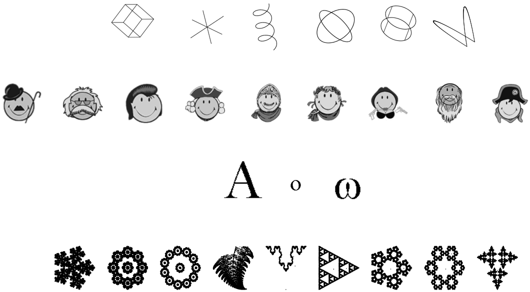
   As is asymptotically Gaussian ($D \to \infty$)

   $\Downarrow$

   g: 'gaussianization' transformation
2. Estimate **W**: linear ISA to $\mathbf{g}(\mathbf{x})$.

Test databases (s)

- I.i.d. tests:
  1. *3D-geom* ($d = 3, M = 6$),
  2. *celebrities* ($d = 2, M = 10$),
  3. *letters* ($d = 2, M \leq 50$).
- Non-i.i.d. test: *IFS* (self-similar structures; $d = 2, M = 9$).



Simulations

- Performance index: Amari-index ($r \in [0,1]$) to measure the *block-permutation matrix* property of the linear approximation of

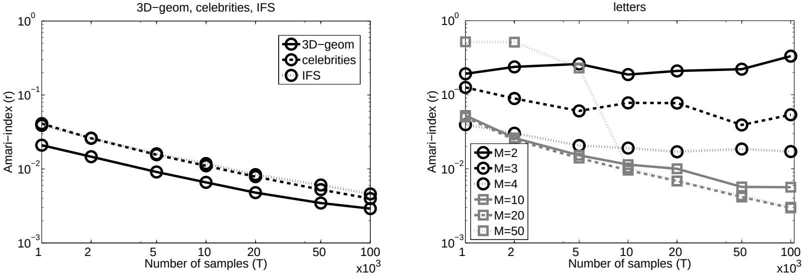
$$\mathbf{s} - E[\mathbf{s}] \in \mathbb{R}^D \mapsto \hat{\mathbf{s}} := \mathbf{Wg}[\mathbf{f}(\mathbf{As})] - E[\hat{\mathbf{s}}] \in \mathbb{R}^D.$$

- Components of the PNL ISA algorithm:
  - gaussianization based on ranks of samples,
  - ISA by joint f-decorrelation (JFD).
- Simulation parameters:
  - goodness: average of 50 random ($\mathbf{A}, \mathbf{s}, \mathbf{f}$) runs,
  - mixing matrix $\mathbf{A}$: random orthogonal,
  - coordinate-wise distortions: $f_i(z) = c_i[a_i z + \tanh(b_i z)] + d_i$.

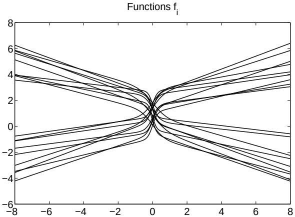
Illustrations-1: $r(T)$

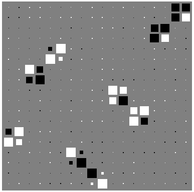
Amari-index as a function of

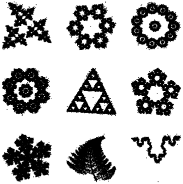
- the sample number (T): *3D-geom*, *celebrites*, *IFS*.
- dimensionality of the problem ($\leftrightarrow M$): *letters*.



Power-law decline: $r(T) \propto T^{-c}$ $(c > 0)$, as $D\nearrow$.

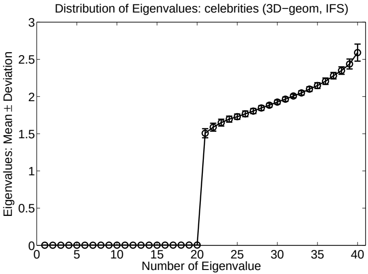

Functions f_i
8
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8

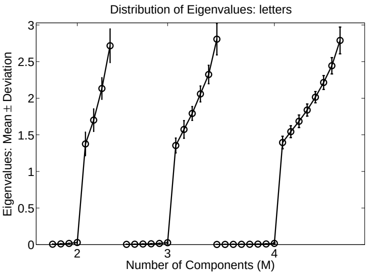




Illustrations-2: $\hat{D}$ $(T = 10,000)$

- Estimation of $D = dim(\mathbf{s})$ - in the background: $D_x = 2D$.
- Gaussianization $\rightarrow$ ordered eigenvalues of $cov[\mathbf{g}(\mathbf{x})]$.
- Results: average over 50 random runs $(\mathbf{A}, \mathbf{f})$.





Summary

- PNL ISA problem
- Separability of PNL ISA

$\Downarrow$ ($\leftarrow$ d-dependent Central Limit Theorem)

$$\boxed{\text{PNL ISA = gaussianization + ISA}}$$

- Simulations:
  - Estimation error vs. sample number:
    - power-law decline, as $D\nearrow$.
  - Possibility to estimate the dimension of the hidden source.
    - The dimensions of the hidden sources: can also be estimated using the ISA Separation Theorem [Szabó et al., JMLR 8 (2007), 1063-1095] . . .

Thank you for the attention!