Independent Subspace Analysis and ISA on Innovation Using Entropy Estimation

Barnabás Póczos
Zoltán Szabó
András Lőrincz

Eötvös Loránd University
Budapest, Hungary
Independent Subspace Analysis

Sources

\[ s^1 \in \mathbb{R}^m \]

\[ s^2 \in \mathbb{R}^m \]

\[ \vdots \]

\[ s^d \in \mathbb{R}^m \]

Observation

\[ x^1 \in \mathbb{R}^m \]

\[ x^2 \in \mathbb{R}^m \]

\[ \vdots \]

\[ x^d \in \mathbb{R}^m \]

Estimation

\[ y^1 \in \mathbb{R}^m \]

\[ y^2 \in \mathbb{R}^m \]

\[ \vdots \]

\[ y^d \in \mathbb{R}^m \]

\[ A \in \mathbb{R}^{md \times md} \]

\[ W \in \mathbb{R}^{md \times md} \]

\[ s = [(s^1)^T, \ldots, (s^d)^T]^T \in \mathbb{R}^{dm} \]: hidden, independent subspaces

\[ x = [(x^1)^T, \ldots, (x^d)^T]^T \in \mathbb{R}^{dm}, x = As \]: observation

\[ y = [(y^1)^T, \ldots, (y^d)^T]^T \in \mathbb{R}^{dm}, y = Wx \]: estimated sources
The Cost Function

Mutual Information:

\[ I(y^1, \ldots, y^d) = \int \log \frac{p(y)}{p(y^1) \cdots p(y^d)} dy \]

Shannon-entropy:

\[ H(y) = -\int p(y) \log p(y) dy \]

Assume \( y = Wx \). Then \( H(Wx) = H(x) + \log |W| \)

\[ I(y^1, \ldots, y^d) = -H(x) + \log |W| + \sum_{i=1}^{d} H(y^i) \]

The cost function:

\[ J(W) = H(y^1) + \ldots + H(y^d) \]
Multi-dimensional Entropy Estimations Using Nearest Neighbours

Let us apply Rényi’s-entropy for estimating the Shannon-entropy:

\[
H_\alpha = \frac{1}{1-\alpha} \log \int f_\alpha(z) \, dz
\]

\[
\lim_{\alpha \to 1} H_\alpha = -\int f(z) \log f(z) \, dz
\]

Let \( \{z(1), \ldots, z(n)\} \) denote \( n \) i.i.d. samples drawn from the distribution of \( z \in \mathbb{R}^m \). Let \( \mathcal{N}_{k,j} \) be the \( k \) nearest neighbours of \( z(j) \) in the sample set. Let \( \gamma = m - m\alpha \), then

\[
\frac{1}{1-\alpha} \log \left( \frac{1}{n^\alpha} \sum_{j=1}^{n} \sum_{v \in \mathcal{N}_{k,j}} \|v - z(j)\|^{\gamma} \right) \to H_\alpha(z) + c,
\]

as \( n \to \infty \)
Multi-dimensional Entropy Estimations Using Geodesic Spanning Forests

- Build first an **Euclidean neighborhood graph**
  - use the edges of the k nearest nodes to each node

- Find **geodesic spanning forests** on this graph
  - (minimal spanning forests of the Euclidean neighborhood graph)
Numerical Simulations
2D Letters (i.i.d.)

Sources

Observation
Estimated sources
Performance matrix
Numerical Simulations

3D Curves (i.i.d.)

Sources

Observation

Estimated sources

Performance matrix
Numerical Simulations

Facial images (i.i.d.)

Sources

Observation

Estimated sources

Performance matrix
Numerical Simulations
Movies
These methods for entropy estimation need i.i.d. processes.

What can we do with $\tau$-order AR sources?

Then the innovations are i.i.d. processes:

$$s^i(t) = F_1s^i(t-1) + \ldots + F_\tau s^i(t-\tau) + \mu(t)$$

and the mixing matrix is the same for the innovations, so we can use ISA on innovations.

$$A\hat{s}(t) = x(t) - E(x(t)|x(t-1), x(t-2)\ldots) = \hat{x}(t)$$
Results Using Innovations

Original ‘AR’ sources

Estimated sources by plain ISA

Estimated sources using ISA on innovations

Mixed sources

Performance of plain ISA

Performance using innovations
References

Independent Subspace Analysis on Innovations
B. Póczos, B. Takács and A. Lőrincz

Independent subspace analysis using geodesic spanning trees
B. Póczos and A. Lőrincz

Independent subspace analysis using k-nearest neighborhood distances
B. Póczos and A. Lőrincz

Cross-Entropy Optimization for Independent Process Analysis
Z. Szabó, B. Póczos and A. Lőrincz
submitted

Separation Theorem for Independent Subspace Analysis
Z. Szabó, B. Póczos and A. Lőrincz
ELU, Budapest, Technical Report
Multi-dimensional Entropy Estimations Using Euclidean Spanning Trees

Let us define

- Euclidean graph

\[ E = \{ e : e(p, q) = z(p) - z(q) \in \mathbb{R}^m, p \neq q \} \]

- Weight of minimal (\( \gamma \)-weighted) Euclidean spanning tree:

\[ L_\gamma(z) = \min_{T \in \mathcal{T}} \sum_{e \in T} \| e \|_\gamma \]

Where \( \mathcal{T} \) is the set of all \( \gamma \)-weighted Euclidean spanning trees

Let \( \gamma = m - m\alpha \), then

\[ \frac{m}{\gamma} \log \frac{L_\gamma(z)}{n^\alpha} \rightarrow H_\alpha(z) + c, \text{ as } n \rightarrow \infty \]
Independent Subspace Analysis
The Ambiguity

- Sources can be recovered only up to:
  - arbitrary permutation
  - arbitrary invertible transformation

- Pairwise independence of the subspaces ≠ joint independence of the subspaces

  Proof:
  Let \( \{s_1^1, s_1^2, s_1^3\}, \{s_2^1, s_2^2, s_2^3\}, \{s_3^1, s_3^2, s_3^3\} \) 3 of 3 dimensional independent sources, where the elements of each subspace are pairwise independent.

  Than \( \{s_1^1, s_2^1, s_3^1\}, \{s_1^2, s_2^2, s_3^2\}, \{s_1^3, s_2^3, s_3^3\} \) is a wrong ISA solution.
Estimation of the Shannon-entropy

The graph shows the entropy as a function of the rotation angle. The lines represent the true and estimated entropy, with parameters $k=20$ and $\gamma=0.01$. The graph indicates a periodic behavior of the entropy with peaks and troughs.
Pairwise Independent Sources
Optimization
Using Jacobi-rotations

Let us define the Jacobi-rotation matrix:

\[
J(W) = H(y^1) + \ldots + H(y^d) \\
= \sum_{j=1}^{d} \sum_{i=1}^{m} H(y^j_i) - \sum_{j=1}^{d} I(y^j_1, \ldots, y^j_m)
\]

It’s worth starting with ICA

Let us define the Jacobi-rotation matrix:

\[
G(p, q, \theta) = \begin{pmatrix}
1 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \ldots & \cos(\theta) & \ldots & -\sin(\theta) & \ldots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \ldots & \sin(\theta) & \ldots & \cos(\theta) & \ldots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & \ldots & 0 & \ldots & 1
\end{pmatrix} \in \mathbb{R}^{md \times md}
\]

Examine \(H(\hat{y}^1(\theta)) + \ldots + H(\hat{y}^d(\theta))\), where \([\hat{y}^1(\theta)^T, \ldots, \hat{y}^d(\theta)^T]^T = G(p, q, \theta)y\)
The Pseudo-code

INITIALIZATION

$x(1), \ldots, x(n) \in \mathbb{R}^{md}$: measured signals
$y \in \mathbb{R}^{md}$: ICA estimations from $x$

ITERATE UNTIL CONVERGENCE

for $p = 1 : md - m$, $q = m \lfloor \frac{p - 1}{m} \rfloor + m + 1 : dm$

Choose $\theta^* = \arg \max_{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]} J_{p,q}(\theta)$

where

$J_{p,q}(\theta) = H(\hat{y}^1(\theta)) + \ldots + H(\hat{y}^d(\theta))$,

$[\hat{y}^1(\theta)^T, \ldots, \hat{y}^d(\theta)^T]^T = G(p, q, \theta)y$,

Set $y := G(p, q, \theta^*)y$

endfor

OUTPUT: Estimated sources $y^* := y$
One-dimensional Entropy Estimation

- Estimating first the density function
  - Edgeworth expansion
    - Hermite polinoms and higher order cumulants
  - Mixture of Gaussians estimation
  - B-spline estimation (Pham)
  - Parzen windows, kernel methods

- Direct entropy estimation
  - Spacing methods (Wasicek, Learned-Miller)
Making More Robust Estimation

- Delete the longest $k$ edges from the spanning tree and consider the remaining sum of the edge lengths (Banks et al, 1992)

- Replace the set of points $\{z(1), \ldots, z(n)\}$ by the $k$-element subset which produces the smallest sum of edge lengths.
  - NP complete
  - Greedy approximation exists (Hero & Michel, 1998)

- Utilize geodesic distances (Costa & Hero, 2004)
Geodesic Distances

- Build first an *Euclidean neighborhood graph*
  - use the edges of the k nearest nodes to each node $z(p)$
  - use the edges of the nodes within radius $\varepsilon$ to each node $z(p)$

- Find *geodesic spanning forests* on this graph
  - (minimal spanning forests of the Euclidean neighborhood graph)

Other applications:
- Manifold learning, ISOMAP algorithm (Tenenbaum et al, 2000)
- Estimating intrinsic dimensions (Costa & Hero, 2004)
Geodesic Spanning Trees
The goal of ICA is to recover linearly, or nonlinearly mixed independent, possibly noisy sources.

\[ y = Wx \]

Sources \[ \rightarrow \text{Mixing} \rightarrow \text{Observation} \rightarrow \text{Estimation} \]

\[ x = As \]

\[ y = Wx \]
Some ICA Applications

- Blind source separation
- Image denoising
- Medical signal processing – fMRI, ECG, EEG
- Modeling of the hippocampus
- Modeling of the visual cortex
- Feature extraction
- Face recognition
- Compression, redundancy reduction
- Clustering
- Time series analysis
- Financial applications
Independent Subspace Analysis
The Ambiguity

Let
\[ C = \begin{pmatrix}
  0 & 0 & C_3 & 0 \\
  C_1 & 0 & 0 & 0 \\
  0 & 0 & 0 & C_4 \\
  0 & C_2 & 0 & 0 \\
\end{pmatrix} \in \mathbb{R}^{md \times md} \]

where
\[ C_i \in \mathbb{R}^{m \times m}, \text{ invertible, } i = 1 \ldots d \]

Then
\[ x = (AC^{-1})(Cs) \text{ and } C_is^i \text{ is independent of } C_js^j \]

In the ISA model:

matrices \( A \) and \( AC^{-1} \)
sources \( s^i \) and \( C_is^i \) are indistinguishable.
Multi-dimensional Entropy Estimations

Let us apply Rényi’s-entropy for estimating the Shannon-entropy:

\[ H_\alpha = \frac{1}{1-\alpha} \log \int f^\alpha(z) dz \]

\[ \lim_{\alpha \to 1} H_\alpha = -\int f(z) \log f(z) dz \]

Let \( \{z(1), \ldots, z(n)\} \) denote \( n \) i.i.d. samples drawn from distribution \( z \in \mathbb{R}^m \).
Beadword - Halton - Hammersley Theorem

Let $N_{k,j}$ be the $k$ nearest neighbours of $z(j)$ in the sample set. Let $\gamma = m - m_\alpha$, then

$$\frac{1}{1-\alpha} \log \left( \frac{1}{n^{\alpha}} \sum_{j=1}^{n} \sum_{v \in N_{k,j}} \|v - z^i(j)\|^{\gamma} \right) \rightarrow H_\alpha(z) + c,$$

as $n \rightarrow \infty$

This estimation is
- asymptotically unbiased and strongly consistent
(Yukich, 1998)
Let \{z(1), \ldots, z(n)\} denote \(n\) i.i.d. samples drawn from distribution \(z \in \mathbb{R}^m\)

Let us define

- Euclidean graph

\[
E = \{e : e(p, q) = z(p) - z(q) \in \mathbb{R}^m, p \neq q\}
\]

- Weight of minimal (\(\gamma\)-weighted) Euclidean spanning tree:

\[
L_\gamma(z) = \min_{T \in \mathcal{T}} \sum_{e \in T} \|e\|^{\gamma}
\]

Where \(\mathcal{T}\) is the set of all \(\gamma\)-weighted Euclidean spanning trees.
Beadword - Halton - Hammersley
Theorem

Let $\gamma = m - m\alpha$, then

$$\frac{m}{\gamma} \log \frac{L_\gamma(y^i)}{n^\alpha} \to H_\alpha(y^i) + c, \text{ as } n \to \infty$$

This estimation is

- asymptotically unbiased and strongly consistent (Yukich, 1998)

- sensitive to outliers if the spanning tree has long edges
Making More Robust Estimation

- Delete the longest k edges from the spanning tree and consider the remaining sum of the edge lengths (Banks et al, 1992)

- Replace the set of points \( \{y^i(1), \ldots, y^i(n)\} \) by the k-element subset which produces the smallest sum of edge lengths.
  - NP complete
  - Greedy approximation exists (Hero & Michel, 1998)

- Utilize geodesic distances
Geodesic Distances

- Build first an *Euclidean neighborhood graph*
  - use the edges of the k nearest nodes to each node $y^i(p)$
  - use the edges of the nodes within radius $\varepsilon$ to each node $y^i(p)$

- Find **geodesic spanning forests** on this graph
  - (minimal spanning forests of the Euclidean neighborhood graph)

Other applications:
- Manifold learning, ISOMAP algorithm (Tenenbaum et al, 2000)
- Estimating intrinsic dimensions (Costa & Hero, 2004)