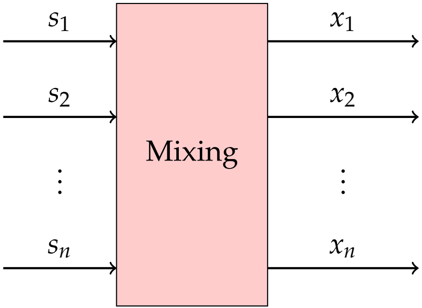
EE Department, Sharif University of Technology

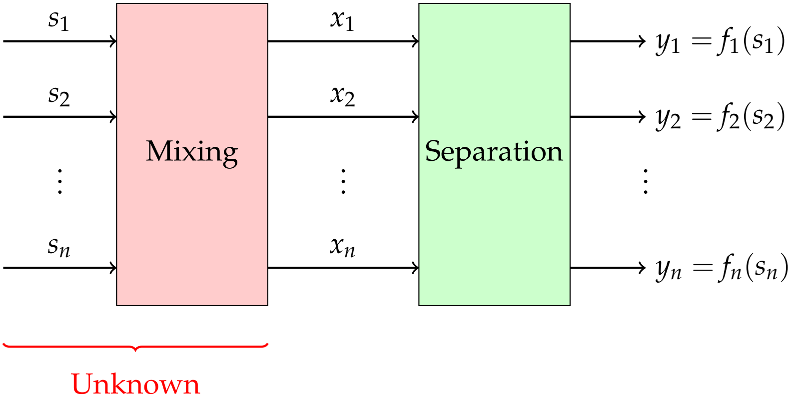
Blind Separation of Nonlinear Mixtures of Stochastic Processes

Behrad Moniri
bemoniri@ee.sharif.edu

Advisor:
Prof. Massoud Babaie-Zadeh

Outline


s_1
s_2
$\vdots$
s_n
Mixing
x_1
x_2
$\vdots$
x_n


s_1
s_2
s_n
Mixing
x_1
x_2
x_n
Separation
$y_1 = f_1(s_1)$
$y_2 = f_2(s_2)$
$y_n = f_n(s_n)$
Unknown

Darmois-Skitovic Theorem [Darmois-Skitovich 1950]

In the linear setting, the model is identifiable if the sources are **non-Gaussian** random variables.

Darmois-Skitovic Theorem [Darmois-Skitovich 1950]

In the linear setting, the model is identifiable if the sources are **non-Gaussian** random variables.

Non-Linear Mixtures

Non-linear mixtures are harder!

$$\begin{cases} S_1 = \text{Rayleigh}(\sigma) \\ S_2 = \text{Uniform}[0, 2\pi] \end{cases} \implies X_1 = S_1 \cos(S_2) \perp\!\!\!\perp X_2 = S_1 \sin(S_2)$$

Permutation Contrastive Separation [Hyvarinen et al., 2017]

The mixture $\mathbf{x}(t) = \mathbf{f}\big(\mathbf{s}(t)\big)$ is separable if:

- $\mathbf{f}$ is invertible and smooth!
- $s_i(t)$: **stationary**, **ergodic** and ***uniformly dependent***.
- $s_i(t)$ are not ***quasi-Gaussian***.

Time Contrastive Learning [Hyvarinen et al., 2016]

The mixture $\mathbf{x}(t) = \mathbf{f}\big(\mathbf{s}(t)\big)$ is separable if:

- $\log p_\tau(s_i) = \lambda_i(\tau) q(s_i) - \log Z(\lambda_i(\tau))$
- $[\mathbf{L}]_{\tau,i} = \lambda_{i,1}(\tau) - \lambda_{i,1}(1)$ has full column rank.

Gaussianity Preserving Functions

A fundamental question:
How can we separate Gaussian sources?

Gaussianity Preserving Functions

A fundamental question:
How can we separate Gaussian sources?

- Can nonlinear functions preserve Normality of random variables?

Gaussianity Preserving Functions

A fundamental question:
How can we separate Gaussian sources?

- Can nonlinear functions preserve Normality of random variables? YES!

Example

Define the function h as follows

$$h(x) = \begin{cases} -x & a \leq |x| < b \\ x & \text{otherwise} \end{cases}$$

If X is a Normal Random variable, then $h(X)$ is also a Normal random variable.

There are also many other examples!

Gaussianity Preserving Functions

A fundamental question:
How can we separate Gaussian sources?

- Can nonlinear functions preserve Normality of random variables? YES!

Example

Define the function h as follows

$$h(x) = \begin{cases} -x & a \leq |x| < b \\ x & \text{otherwise} \end{cases}$$

If X is a Normal Random variable, then $h(X)$ is also a Normal random variable.

There are also many other examples!

- How about specific classes of functions?

High Dimensional Polynomial Mappings

- We have proved that a high dimensional polynomial function preserves normality, if and only if it is linear.

Polynomial Mixing Theorem

Let $\mathbf{s} = (s_1, s_2, \ldots, s_n)^\top \sim \mathcal{N}(\mu, \Sigma_s)$ be a vector of jointly normal random variables and $\mathbf{p} : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible polynomial mapping.

$$\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top \triangleq \mathbf{p}(\mathbf{s}) \tag{1}$$

The random variables $y_1, \ldots, y_n$ are jointly normally distributed if and only if the polynomial $\mathbf{p}$ satisfies

$$\mathbf{y} = \mathbf{p}(\mathbf{s}) = \mathbf{A}\mathbf{s} + \mathbf{b}, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$.

Based on the previous theorem, we can prove the following corollary:

Corollary

Let $\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t))$*, where*

- $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ *is an unknown invertible polynomial.*
- *For all* $i \in [n]$, $s_i(t)$ *are mean zero Gaussian processes.*

If there exists polynomial $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ *such that* $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t))$ *are Gaussian Processes, then* $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ *is linear.*

A parametric model for polynomials:

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta_1}^\top \\ \boldsymbol{\theta_2}^\top \\ \vdots \\ \boldsymbol{\theta_n}^\top \end{bmatrix} \mathbf{k}(\mathbf{x}) = \boldsymbol{\Theta}\mathbf{k}(\mathbf{x})$$

Calculate Neg-Entropy as a measure of Gaussanity:

$$\boldsymbol{\mathcal{J}}(y_i) = \boldsymbol{H}(\tilde{y}_i) - \boldsymbol{H}(y_i)$$

Thus the algorithm should solve the following problem:

$$\min_{\boldsymbol{\Theta}} \|\boldsymbol{\mathcal{J}}(\boldsymbol{\Theta}\mathbf{k}(\mathbf{x}))\|_2^2$$

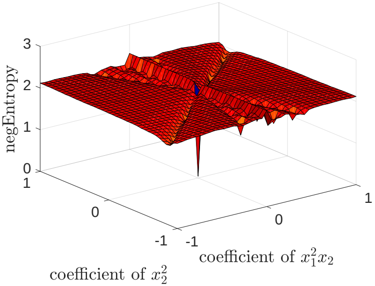
Simulations

Let $s_1, s_2 \sim \mathcal{N}(0,1)$ and $s_1 \perp\!\!\!\perp s_2$.

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} s_1 + (s_1 + s_2)^3 \\ s_2 - (s_1 + s_2)^3 \end{bmatrix}$$

This function can be exactly inverted as

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} x_1 - (x_1 + x_2)^3 \\ x_2 + (x_1 + x_2)^3 \end{bmatrix} \leftarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$



Algebraic Functions and Rotations

Example

If $(X_1, X_2)^\top \sim \mathcal{N}\big(\mathbf{0}, \Sigma_X\big)$, then $Y_1 = \frac{X_1^2 - X_2^2}{\sqrt{X_1^2 + X_2^2}}$ and $Y_2 = \frac{2X_1X_2}{\sqrt{X_1^2 + X_2^2}}$ is also a Gaussian random vector.

Algebraic Functions and Rotations

Example

If $(X_1, X_2)^\top \sim \mathcal{N}(\mathbf{0}, \Sigma_X)$, then $Y_1 = \frac{X_1^2 - X_2^2}{\sqrt{X_1^2 + X_2^2}}$ and $Y_2 = \frac{2X_1 X_2}{\sqrt{X_1^2 + X_2^2}}$ is also a Gaussian random vector.

General Rotation Conjecture [Eidlin, Linnik, Kagan]

Let $\sigma > 0$ and Consider a random vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ with every $x_j \sim \mathcal{N}(0, \sigma^2)$ and an algebraic transformation $\boldsymbol{\mathcal{A}}$. If $\mathbf{y} = \boldsymbol{\mathcal{A}}(\mathbf{x})$ is normally distributed, then $\|\mathbf{y}\|_2 = \|\mathbf{x}\|_2$.

Ongoing Works

- Gradual non-convexity methods for optimization.

Ongoing Works

- Gradual non-convexity methods for optimization.
- Other measures of non-Gaussanity, e.g. 4th cumulants.

Ongoing Works

- Gradual non-convexity methods for optimization.
- Other measures of non-Gaussanity, e.g. 4th cumulants.
- Can the Theorem 1 be extended to "polynomials" with positive and negative powers?
  I believe the answer is positive and I think that the proof is not very difficult. But I am still working on this.

Ongoing Works

- Gradual non-convexity methods for optimization.
- Other measures of non-Gaussanity, e.g. 4th cumulants.
- Can the Theorem 1 be extended to "polynomials" with positive and negative powers?
  I believe the answer is positive and I think that the proof is not very difficult. But I am still working on this.
- Finding two "good" sets of functions $\mathcal{F}$ and $\mathcal{G}$ such that $\forall f \in \mathcal{F} \;\; \forall g \in \mathcal{G}$, the function $f \circ g$ is a polynomial. This will result in a new corollary and a new separation algorithm.

Selected Papers I

[1] **M. Babaie-Zadeh and Ch. Jutten.**
A general approach for mutual information minimization and its application to blind source separation.
Signal Processing, 85(5):975–995, 2005.

[2] **Bahram Ehsandoust, Massoud Babaie-Zadeh, Bertrand Rivet, and Christian Jutten.**
Blind source separation in nonlinear mixtures: separability and a basic algorithm.
IEEE Transactions on Signal Processing, 65(16):4339–4352, 2017.

[3] **Shahram Hosseini and Christian Jutten.**
On the separability of nonlinear mixtures of temporally correlated sources.
IEEE signal processing letters, 10(2):43–46, 2003.

[4] **Aapo Hyvarinen and Hiroshi Morioka.**
Unsupervised feature extraction by time-contrastive learning and nonlinear ica.
Advances in Neural Information Processing Systems, pages 3765–3773, 2016.

[5] **Aapo Hyvarinen and Hiroshi Morioka.**
Nonlinear ICA of temporally dependent stationary sources.
20th International Conference on Artificial Intelligence and Statistics (AISTAT), 2017.

Selected Papers II

[6] **Aapo Hyvärinen and Erkki Oja.**
Independent component analysis: algorithms and applications.
Neural networks, 13(4-5):411–430, 2000.

[7] **Aapo Hyvarinen, Hiroaki Sasaki, and Richard E Turner.**
Nonlinear ica using auxiliary variables and generalized contrastive learning.
arXiv preprint arXiv:1805.08651, 2018.

[8] **Ch. Jutten, M. Babaie-Zadeh, and J. Karhunen.**
Nonlinear Mixtures in Handbook of Blind Source Separation: Independent component analysis and applications.
Academic press, 2010.

[9] **Abram M Kagan, Calyampudi Radhakrishna Rao, and Yurij Vladimirovich Linnik.**
Characterization problems in mathematical statistics.
1973.

[10] **Yu V Linnik and VL Eidlin.**
Remark on analytic transformations of normal vectors.
Theory of Probability & Its Applications, 13(4):707–710, 1968.

Thank You!