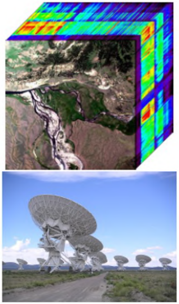
Independent Vector Analysis using Semi-Parametric Density Estimation via Multivariate Entropy Maximization

Lucas P. Damasceno[1], Charles C. Cavalcante[1], Tülay Adalı[2], and Zois Boukouvalas[3]

[1]Dept. of Teleinformatics Eng., Federal University of Ceará, Fortaleza - CE, 60455-760, Brazil
[2]Dept. of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
[3]Dept. of Mathematics and Statistics, American University, Washington, DC 20016, USA

June, 2021

Joint blind source separation is an active area of research due to its numerous applications



The goal is to process the data

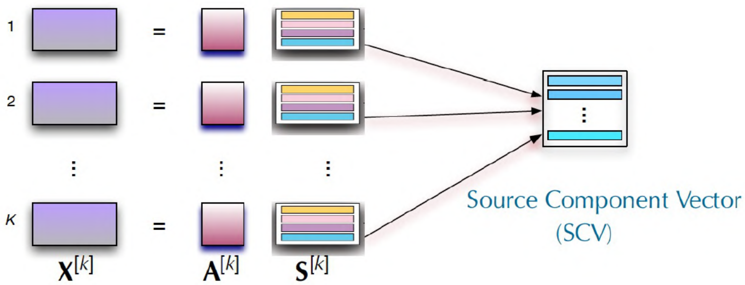
- to explore the underlying structure of multi-modal datasets, i.e., information extraction by integrating and modeling multiple modalities
- to extract meaningful information about the underlying sources collected multiple subjects

IVA is an effective solution for joint blind source separation

IVA is an extension of ICA to **multiple datasets**

ICA: $\mathbf{x} = \mathbf{A}\mathbf{s}$ $\mathbf{y} = \mathbf{W}\mathbf{x}$, $\mathbf{W} \in \mathbb{R}^{N \times N}$

IVA: $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}$ $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, $\mathbf{W}^{[k]} \in \mathbb{R}^{N \times N}$, $\mathrm{k} = 1,\ldots,\mathrm{K}$



IVA and mutual information (MI)

The goal in IVA is to estimate K **demixing matrices**, $\mathbf{W}^{[k]}$, to yield maximally independent source estimates

$$\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$$

such that each SCV is maximally independent of all other SCVs.

MI cost function

$$J_{\text{IVA}}(\mathbf{W}^{[k]}) = \sum_{n=0}^{N} \underbrace{H(\mathbf{y}_n)}_{-E\{\log p(\mathbf{y}_n)\}} - \sum_{k=0}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C$$

where $\mathbf{y}_n$ is the nth estimated SCV and $H(\mathbf{y}_n)$ denotes its differential entropy.

- Instead of minimizing the MI cost function with respect to $\mathbf{W}^{[k]}$, we use a **decoupling procedure** to minimize with respect to each row of $\mathbf{W}^{[k]}$.

Decoupling MI cost function provides optimization benefits

Decoupling procedure

$$J_{\text{IVA}}(\mathbf{w}_n^{[k]}) = H[\mathbf{y}_n] - \log \left| \left(\mathbf{h}_n^{[k]}\right)^{\top} \mathbf{w}_n^{[k]} \right| - \mathbf{C}_n^{[k]}$$

for $n = 1, ..., N$, where $\mathbf{h}_n^{[k]}$ is the unit length vector is perpendicular to $\mathbf{w}_j^{[k]}, \forall j \neq n$.

Gradient

$$\frac{\partial J_{\text{IVA}}(\mathbf{w}_n^{[k]})}{\partial \mathbf{w}_n^{[k]}} = -E\left\{ \frac{\partial \log\left(p_n\left(\mathbf{y}_n\right)\right)}{\partial y_n^{[k]}} \mathbf{x}^{[k]} \right\} - \frac{\mathbf{h}_n^{[k]}}{(\mathbf{h}_n^{[k]})^{\top} \mathbf{w}_n^{[k]}}$$

where $p(\mathbf{y}_n)$ denotes its **probability density function (PDF).**

Estimation of PDF for each SCV $\mathbf{y}_n$ plays an important role in the development of IVA algorithms.

The maximum entropy principle

The best probability distribution to model a dataset is the one with largest entropy

$$\max_{p(\mathbf{y})} H(p(\mathbf{y})) = -\int_{\mathbb{R}^K} p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y}$$

$$\text{s.t.} \quad \int_{\mathbb{R}^K} r_i(\mathbf{y}) p(\mathbf{y}) d\mathbf{y} = \alpha_i, \ \text{for } i = 0, ..., M$$

where $\mathbf{y} \in \mathbb{R}^K$ represents the dataset and $r_i(\mathbf{y})$ are the constraints.

Lagrangian form

$$\mathcal{L}(p(\mathbf{y})) = -\int_{\mathbb{R}^K} p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} + \sum_{i=0}^{M} \lambda_i \int_{\mathbb{R}^K} (r_i(\mathbf{y}) - \alpha_i) p(\mathbf{y}) d\mathbf{y}$$

where λ_i are the Lagrangian multipliers.

Density function

$$p(\mathbf{y}) = e^{-1+\sum_{i=0}^{M} \lambda_i r_i(\mathbf{y})}$$

λ_i are estimated by Newton iterations.

Global and local constraints provide flexible model

Global constraints

$r_0(\mathbf{y}) = 1$ $\qquad r_2(\mathbf{y}) = y^2$

$r_1(\mathbf{y}) = y$ $\qquad r_3(\mathbf{y}) = \frac{y}{(1+y^2)}$

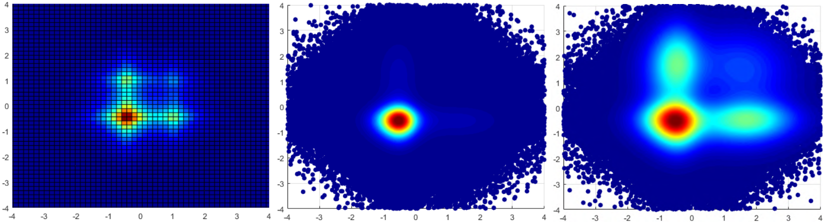
provide information about the PDF's overall statistics, such as the mean, variance, and certain higher order statistics.

Local constraint

Gaussian Kernel:

$$q(\mathbf{y}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|(2\pi)^K}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mu)\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mu)'\right)$$

provides localized information about the PDF.



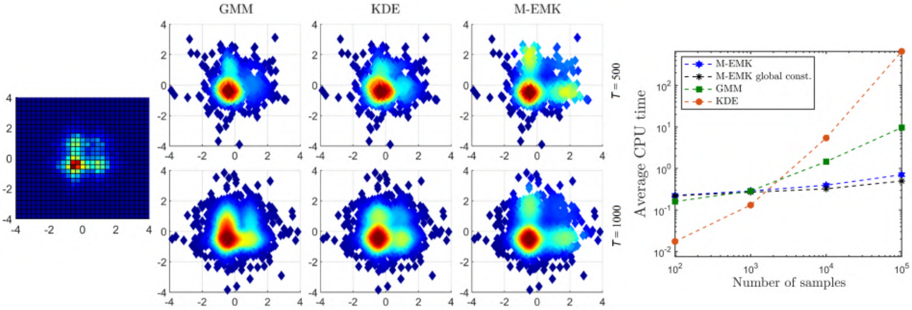
Estimation performance only using global constraints and using global and local constraints.

Multivariate entropy maximization with kernels (M-EMK)

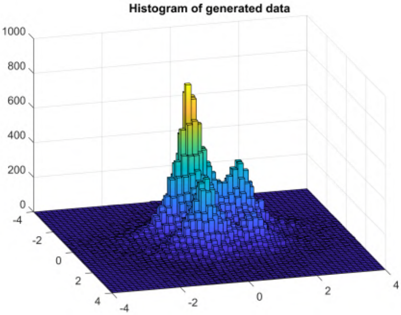
Quasi-Monte Carlo multidimensional integral approximation

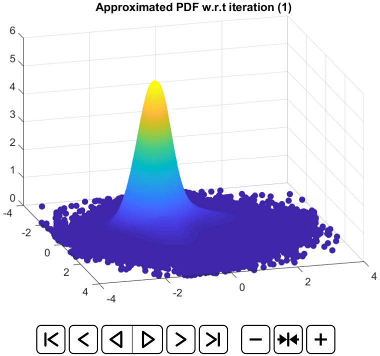
$$Q_{T,K}\left(p(\mathbf{y})\right) = \Omega\left(\frac{1}{T}\sum_{i=0}^{T-1} p(s)\right)$$

where Ω, T, p and s denote the dimensional measure of the integration region, sample size, density function and the generated quasi-random points sequence, respectively.



Density estimation performance of M-EMK





Approximation to $T = 10000$ samples generated from a mixture of multivariate generalized Gaussians.

Incorporation into the IVA model

M-EMK density function

$$p(\mathbf{y}) = \exp\left\{-1 + \sum_{i=0}^{M} \lambda_i r_i(\mathbf{y})\right\}$$

IVA-M-EMK

$$J_{\text{IVA}}(\mathbf{W}^{[k]}) = \sum_{n=0}^{N} -E\{\log p(\mathbf{y})\} - \sum_{k=0}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C$$

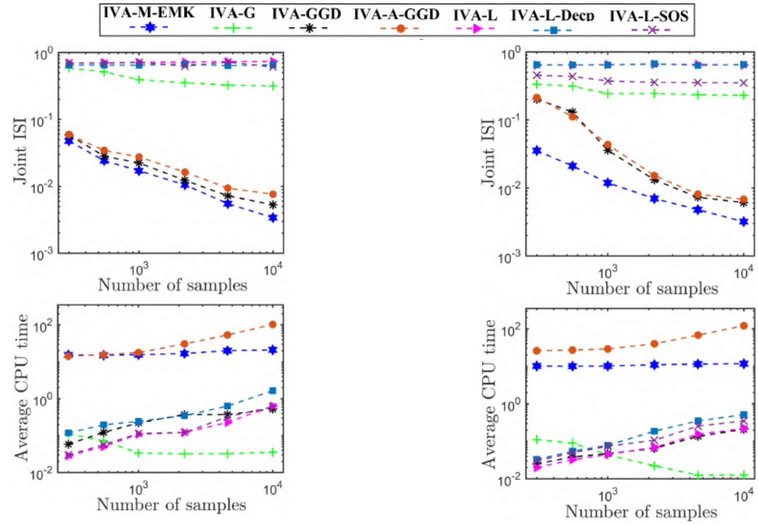
The gradient w.r.t. $\mathbf{w}_n^{[k]}$ is given by

$$\frac{\partial J_{\text{IVA}}(\mathbf{w}_n^{[k]})}{\partial \mathbf{w}_n^{[k]}} = -\sum_{i=0}^{M} \lambda_i \frac{\partial r_i(\mathbf{y}_n)}{\partial y_n^{[k]}} E\left\{\mathbf{x}^{[k]}\right\} - \frac{\mathbf{h}_n^{[k]}}{\left(\mathbf{h}_n^{[k]}\right)^{\top} \mathbf{w}_n^{[k]}}$$

λ_i are provided by M-EMK.

IVA-M-EMK takes advantage of the accurate estimation capability of M-EMK to improve source separation performance.

Experimental results using simulated sources



$K = 3$, one unimodal MGGD, and two multi-modal MGGD sources.

$K = 2$, three multi-modal MGGD sources.

Performance comparison in terms of Joint ISI and average CPU time for different number of sample size.

Summary of contributions

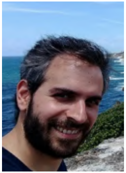
- Present a new multivariate probability density estimator based on the maximum entropy principle that
  - provides flexible multivariate PDFs while keeping computational complexity low;
  - provides superior performance over popular density estimation procedures;
- Derive an efficient IVA algorithm that
  - accurately separates sources from a wide range of multivariate PDFs outperforming widely used IVA algorithms;
  - enables the application of IVA to many practical applications where a flexible multivariate density modeling is needed;

Research team

Lucas P. Damasceno
Machine learning
Statistical signal processing

Dr. Charles C. Cavalcante
Information geometry
Statistical signal processing

Dr. Tülay Adali
Machine learning for
Signal processing



Dr. Zois Boukouvalas
Machine learning
Multi-modal learning

UNIVERSIDADE FEDERAL DO CEARÁ

UMBC

AMERICAN UNIVERSITY
WASHINGTON, DC