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
[2]Dept. of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250
[3]Dept. of Mathematics and Statistics, American University, Washington, DC 20016

2018

Introduction

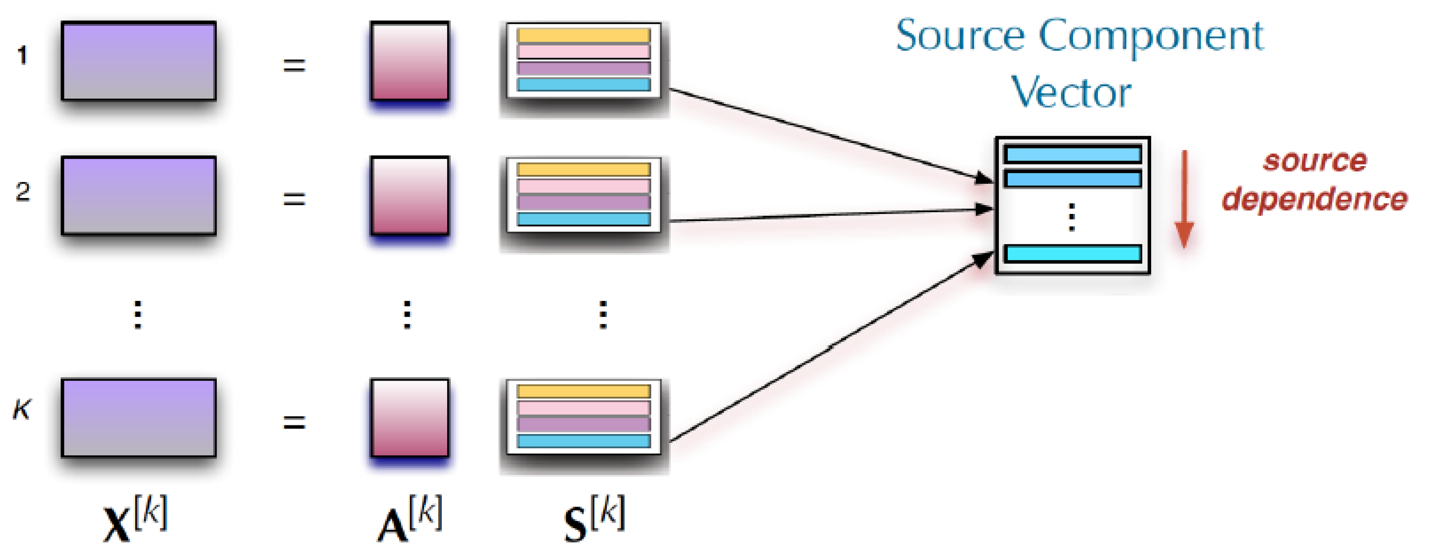
- Independent vector analysis (IVA) is a recent generalization of independent component analysis (ICA) that enables the joint factorization of multiple datasets.
- The success of IVA is tied to proper characterization of the multivariate probability density function (PDF) of the latent sources; information that is generally *unknown*.

Contribution

- We propose a new flexible and efficient multivariate PDF estimation technique based on entropy maximization with kernels, (M-EMK), which jointly uses global and local multidimensional measuring functions to provide flexible PDFs while keeping the complexity low by integrating a multidimensional Monte-Carlo (MC) integration technique.
- We use M-EMK to derive an efficient IVA algorithm, IVA by multivariate entropy maximization with kernels (IVA-M-EMK) that accurately separates sources from a wide range of distributions.
- We verify the effectiveness of the new estimation technique and further demonstrate the superior performance of the new IVA algorithm numerically using simulated data.

Independent vector analysis model

- Generative model: $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}$, where $\mathbf{x}^{[k]}$ are the observations and $\mathbf{s}$ are the latent sources linearly mixed by the *mixing matrix* $\mathbf{A}^{[k]}$



- IVA can separate mixed sources from multiple subjects by assuming source *dependence* across datasets.

[1] Z. Boukouvalas, Y. Levin-Schwartz, R. Mowakeaa, G. Fu, and T. Adalı,"Independent Component Analysis Using Semi-Parametric Density Estimation Via Entropy Maximization,"*IEEE Statistical Signal Processing Workshop (SSP)*, pp. 403-407, 2018.
[2] S.Y. Kung, "Kernel Method and Machine Learning,"*Cambridge University Press*, 2014.
[3] G. McLachlan and D. Peel," Finite Mixture Models,"*ser. Wiley Series in Probability and Statistics*, 2004.
[4] http://mlsp.umbc.edu/jointBSS_introduction.html

Mutual information is a natural measure of dependence

The goal in IVA is to estimate K *demixing matrix*, $\mathbf{W}^{[k]}$, 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. In order to estimate $\mathbf{W}^{[k]}$, we use a *decoupling procedure*[1] and *minimize* the mutual information with respect to each row vector $\mathbf{w}_n^{[k]}$ given by

$$J_{\text{IVA}} = \sum_{n=0}^{N} \underbrace{H(\mathbf{y}_n)}_{-E\{\log p(\mathbf{y}_n)\}} - \sum_{k=0}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - \underbrace{C}_{H(\mathbf{x}^{[1]},\dots,\mathbf{x}^{[K]})},$$

and its gradient by

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

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

Maximum entropy principle

Maximum entropy principle enables a semi-parametric PDF estimation scheme

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

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

where where $\mathbf{y} \in \mathbb{R}^K$ represents the dataset, $r_i(\mathbf{y})$ are the constraints. The solution is given by

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

λ_i are estimated by Newton iterations.

Global and local measuring functions

Global measuring functions

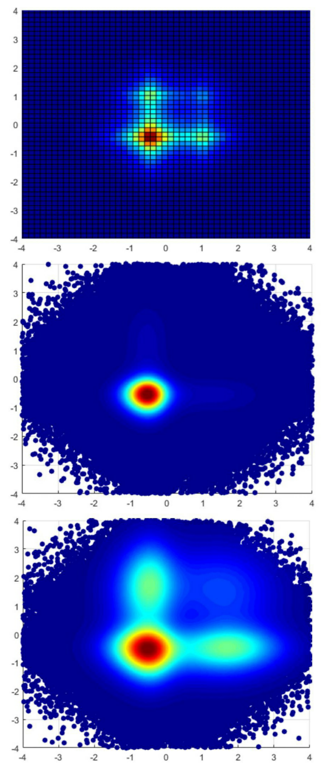
$$r_0(\mathbf{y}) = 1;\ r_1(\mathbf{y}) = y;\ r_2(\mathbf{y}) = y^2;\ r_3(\mathbf{y}) = \frac{y}{1+y^2},$$

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

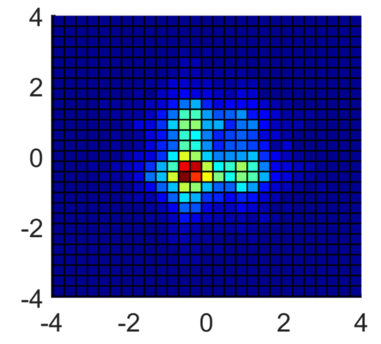
Local measuring functions

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

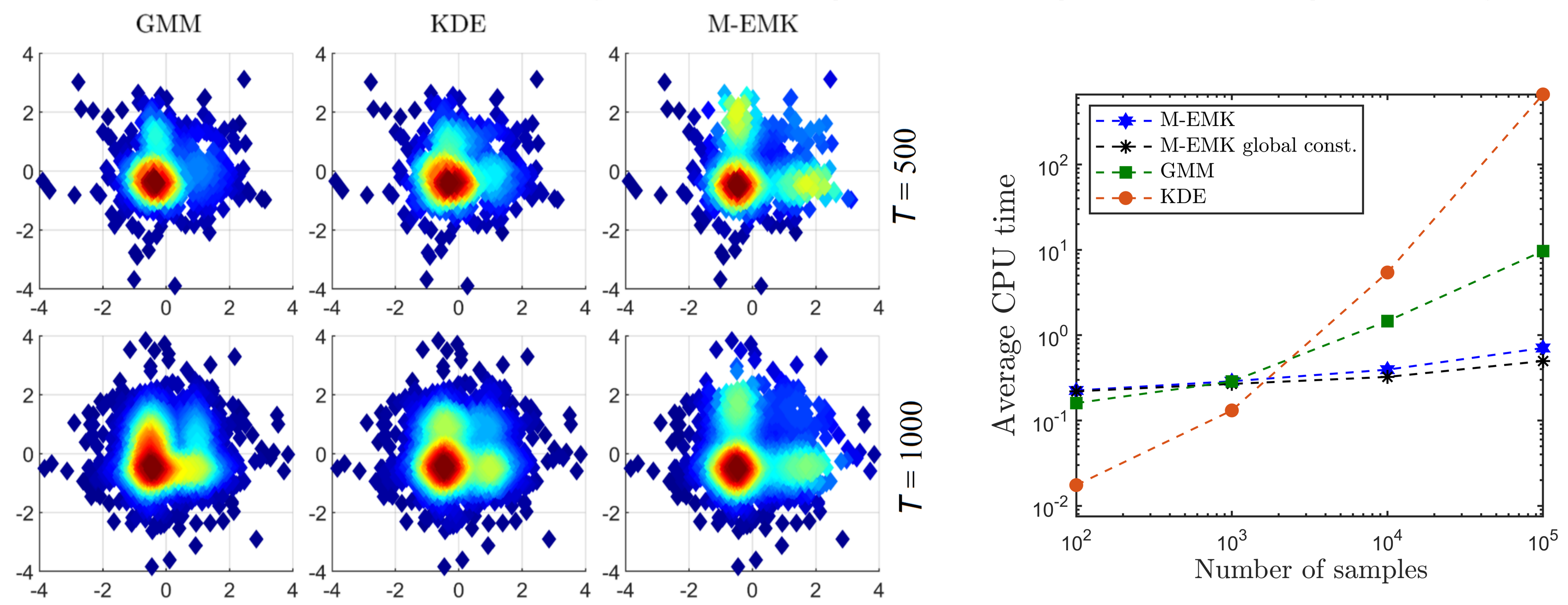
provide *localized information* about the PDF.



M-EMK density estimation performance

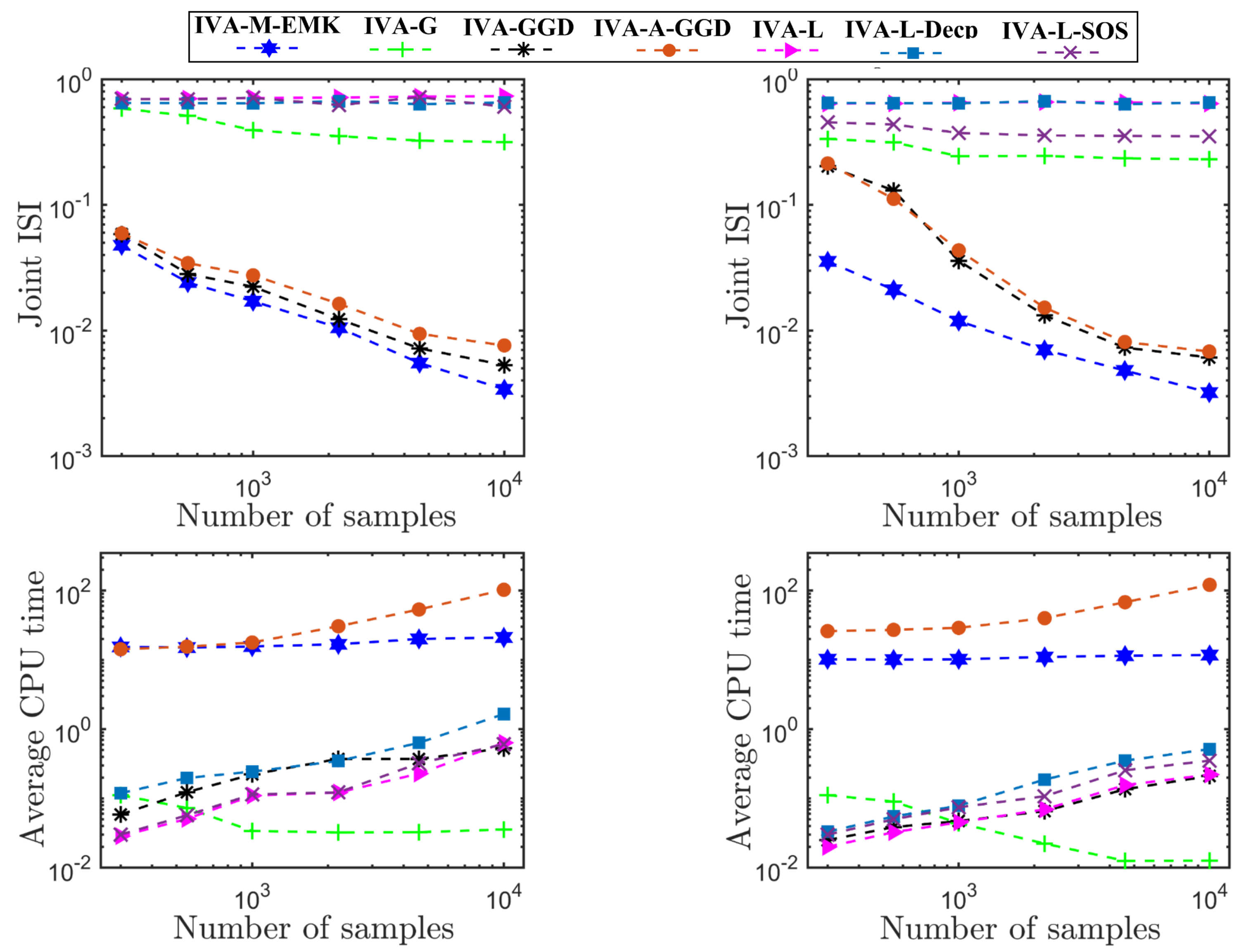


We evaluate the multidimensional integrals by $Q_{T,K}(p(\mathbf{y})) = \Omega\left(\frac{1}{T}\sum_{i=0}^{T-1} p(s)\right)$ where Ω, T, p and s denote the dimensional measure, sample size, density function and the generated points sequence, respectively.



M-EMK outperforms two popular density estimation algorithms[2,3].

IVA-M-EMK experimental results



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

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

Summary of contributions

- Present a new multivariate density estimator based on MEP that provides flexible multivariate PDFs while keeping complexity low.
- Derive an efficient IVA algorithm that accurately separates sources from a wide range of multivariate PDFs outperforming widely used IVA algorithms[4].