INDEPENDENT VECTOR ANALYSIS WITH SPARSE INVERSE COVARIANCE ESTIMATION

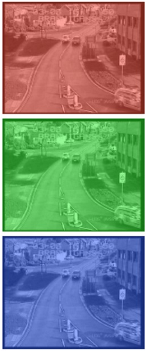
Egzona Rexhepi

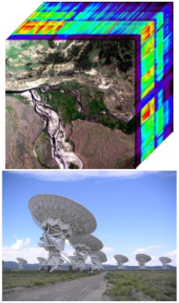
Advisor: Dr. Zois Boukouvalas

Thesis Defense
July 27, 2022

Master of Science: Statistics, Theory and Methods
American University, Washington, D.C.

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





The goal is to process the data to

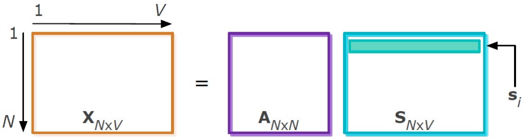
- extract meaningful information about the underlying sources
- explore the underlying structure of the data

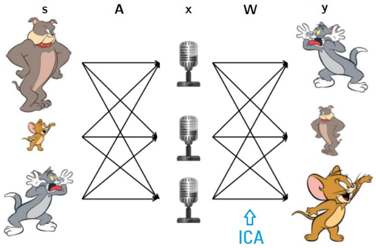
BSS starts with a simple generative model

Generative model:

$$\mathbf{x} = \mathbf{A}\mathbf{s},$$

where $\mathbf{x}$ are the **observations** and $\mathbf{s}$ are the latent **sources** linearly mixed by the **mixing matrix** $\mathbf{A}$.



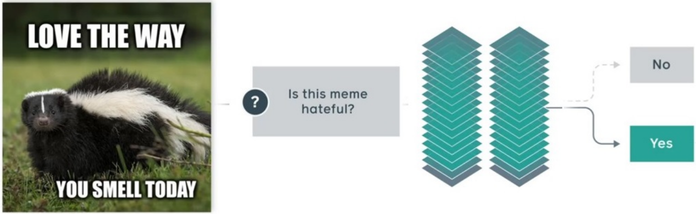


Independent component analysis (ICA) can *uniquely* identify the sources subject to scaling and permutation ambiguities by assuming source *independence*.

Joint analysis of multiple datasets arises in many applications

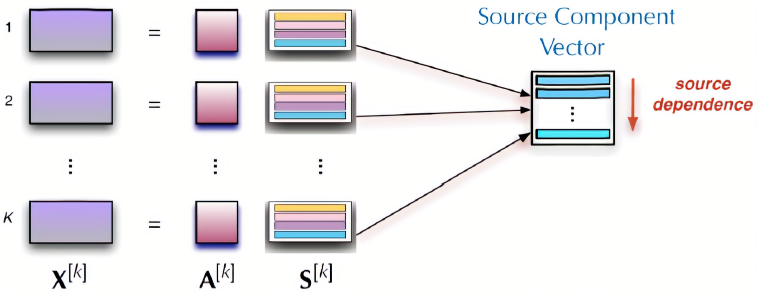
To explore the underlying structure of **multi-modal datasets**, i.e., information extraction by integrating and modeling **multiple modalities**.

Understand **content holistically**.



Different types of diversity

Statistical dependence across multiple datasets



By *performing a joint analysis*, we

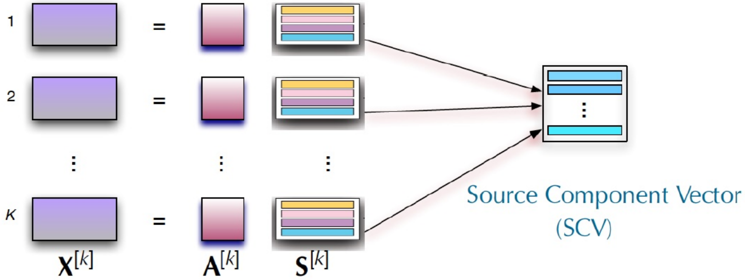
- *resolve* the permutation ambiguity across datasets
- make use of the *true multivariate nature* of the data

IVA is an effective solution for joint blind source separation

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

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

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



Development of effective IVA algorithms

The **key issues** that will enable IVA to provide effective solutions to a variety of applications include

- **Development of effective models for underlying source density and their estimation**

Density matching

- **Efficient use of available prior information**

Incorporation of prior knowledge

Summary of presentation

- **IVA framework**
- **Sparse Inverse Covariance Estimation (SPICE)**
  - **Empirical covariance**
  - **Ledoit-Wolf**
  - **Graphical lasso**
- **IVA with Sparse Inverse Covariance Estimation (IVA-SPICE)**
- **Conclusions and future directions**

Maximum Likelihood (ML) 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.

ML gives a natural objective function where MI is a natural measure of dependence,

Cost function

$$\mathrm{L}_{\mathrm{IVA}}(\boldsymbol{\mathcal{W}}) = \sum_{n=1}^{N} E\{\log p\,(\mathbf{y}_n)\} + \sum_{k=1}^{K} \log|\det\left\{\mathbf{W}^{[k]}\right\}| - C,$$

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

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

Optimization Framework of IVA

Gradient

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

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

Update Rule

$$\mathbf{W}^{[k]} \leftarrow \mathbf{W}^{[k]} - \mu \frac{\partial J_{IVA}}{\partial \mathbf{W}^{[k]}}.$$

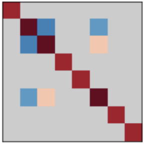
Estimation of the PDF for each SCV $\mathbf{y}_n$ *plays an important role in the development of IVA algorithms.*

Restricting to the multivariate Gaussian distribution tells important information during PDF estimation

Multivariate Gaussian PDF

$$p\left(\mathbf{y}_n \mid \mathbf{\Sigma}_n\right) = 2\pi^{-K/2}\left(\det\left(\mathbf{\Sigma}_n\right)^{-1/2}\right)\exp\left(-\frac{1}{2}\left(\mathbf{y}_n\right)^{\mathsf{T}}\mathbf{\Sigma}_n^{-1}\left(\mathbf{y}_n\right)\right)$$

- Equivalence of partial correlation and conditional independence under Gaussian assumption is shown through zero values in the inverse covariance matrix, $\mathbf{\Sigma}_n^{-1}$
- Sparsity assumption increases partial correlations, relaxing dependence assumption among SCVs



Sparse inverse covariance estimation methods provide a way to impose sparsity in PDF estimation

Empirical estimation

Maximum likelihood estimator gives a good approximation when $T > K$

Empirical Covariance

$$\mathbf{S} = \frac{1}{T}\sum_{i=1}^{T} x_i x_i^{\mathsf{T}}$$

- $\mathbf{S}^{-1}$ gives the estimate for $\boldsymbol{\Sigma}_n^{-1}$

When $K > T$, $\mathbf{S}$ is not invertible

Empirical estimation does not uniquely impose sparsity

Ledoit-Wolf (L-W)

Shrinks the empirical covariance towards a shrinkage target $\mathbf{F}$, in this case the identity matrix, through a linear combination

Ledoit-Wolf

$$\hat{\boldsymbol{\Sigma}}_{LW} = \delta \mathbf{F} + (1 - \delta)\mathbf{S},$$

where δ is the asymptotically optimal estimator found by L-W algorithm

- $\hat{\boldsymbol{\Sigma}}_{LW}^{-1}$ gives the estimate for $\boldsymbol{\Sigma}_n^{-1}$

Shrinking towards the identity matrix imposes sparsity, yet may lose off-diagonal structure

Graphical lasso (GL)

Convex Optimization Problem

$$\hat{\mathbf{\Sigma}}_{GL}^{-1} =_{\mathbf{\Sigma}^{-1}} \operatorname{tr}\left(\mathbf{S}\mathbf{\Sigma}^{-1}\right) - \log\det\mathbf{\Sigma}^{-1} - \rho\left\|\mathbf{\Sigma}^{-1}\right\|_1,$$

where the scalar parameter ρ is decided by ten fold cross validation and controls the magnitude of the penalty on the $\ell 1$ norm, $\left\|\mathbf{\Sigma}^{-1}\right\|_1$

Algorithm 1 Graphical Lasso Algorithm

1: Initialize $\boldsymbol{M} = \boldsymbol{S} + \rho\boldsymbol{I}$

2: Repeat the following steps coordinate-wise until convergence:

a: Rearrange such that target column is last

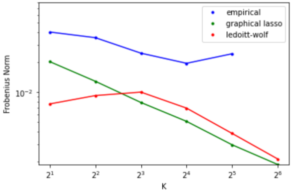
b: Solve the lasso problem (3.11)

c: Update corresponding (off-diagonal) row/column of $\boldsymbol{M}$ using $\hat{m}_{12}$ (3.9)

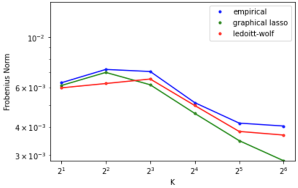
d: Store $\hat{\beta}$ for the target column in matrix $\boldsymbol{B}$

3: For each row/column, compute the diagonals $\hat{\Sigma}_{jj}^{-1}$ from (3.12). Convert $\boldsymbol{B}$ to $\hat{\mathbf{\Sigma}}^{-\mathbf{1}}$.

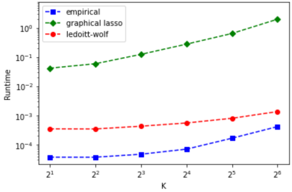
- Blockwise updating scheme guarantees invertible $\hat{\mathbf{\Sigma}}_{GL}$



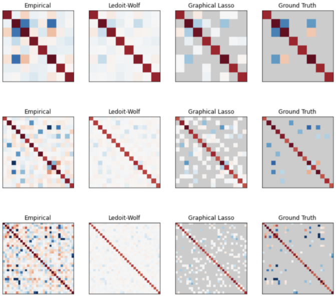
T=64



T=10,000



T=1,000



Performance comparison in terms of Frobenius norm, runtime, and structure for varying parameters.

Incorporation into the IVA model

Derivative of density log-likelihood w.r.t. $\mathbf{y}_n$

$$\frac{\partial \log p\left(\mathbf{y}_n^{[k]} \mid \boldsymbol{\Sigma}_n\right)}{\partial \mathbf{y}_n} = \boldsymbol{\Sigma}_n^{-1}\mathbf{y}_n^{[k]}$$

IVA-SPICE

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

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

$$\frac{\partial J_{\mathrm{IVA}}(\mathbf{W}^{[k]})}{\partial \mathbf{W}^{[k]}} = -E\left\{\hat{\boldsymbol{\Sigma}}_{GL}^{-1}\mathbf{y}_n^{[k]}\left[\mathbf{x}^{[k]}\right]^{-\mathrm{T}}\right\} - \left(\mathbf{W}^{[k]}\right)^{-\mathrm{T}}$$

where $\hat{\boldsymbol{\Sigma}}_{GL}^{-1}$ is given by the graphical lasso algorithm.

IVA-SPICE takes advantage of the accurate estimation capability of graphical lasso to improve source separation performance when underlying structure of SCVs is sparse.

Experimental results using simulated data

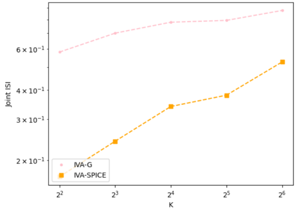


Figure: N = 10, T = 1,000

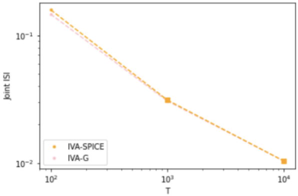


Figure: Non-sparse data

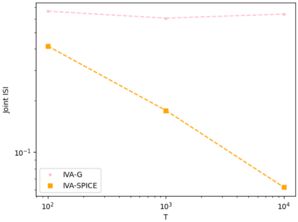


Figure: N = 10, K = 4

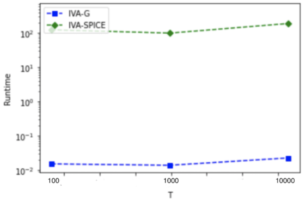


Figure: Mean Runtime

Performance comparison in terms of Joint ISI and average CPU time for varying parameters.

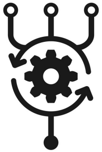
Summary of contributions

- **Present a new multivariate Gaussian probability density estimator using sparse inverse covariance estimation**
  - ✓ **uses a proved efficient technique, graphical lasso**
  - ✓ **holds the structure of true inverse covariance matrix**

- **Derive an efficient IVA algorithm that**
  - ✓ **accurately separates sources defined by a wide range of parameters and underlying structures**
  - ✓ **enables the application of IVA to new applications where underlying sparsity exists**

Motivated by the success of IVA-SPICE, we suggest expanding our work in two major directions

- **Theoretical developments**



- **Application to medical imaging data**

Theoretical Developments

Expanding *beyond assumptions* in current IVA-SPICE framework

Non-Gaussian IVA-SPICE

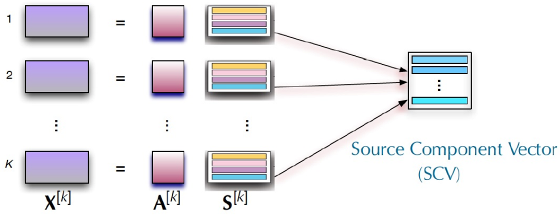
- Real world data may be better modeled by other distributions, motivating expansion. Potential challenges include
  - Distributions not parameterized by inverse covariance
  - No equivalence of partial correlation and conditional independence

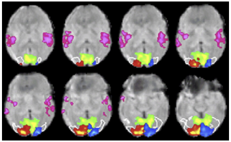
Accounting for sample-to-sample dependence

- Samples which are i.i.d are not always realistic. Potential extensions to IVA-SPICE include
  - Second-order blind identification
  - Entropy rate bound minimization

IVA for medical imaging

Medical imaging with IVA-SPICE





Through the estimation of the SCVs and their associated sparse inverse covariance matrices, we could

- ✓ isolate strongest covariate relationships among signals representing neural activity
- ✓ enhance understanding of brain activity and improve separation performance of mixed multivariate data

Thank you!

AMERICAN UNIVERSITY

WASHINGTON, DC