A ROBUST AND SCALABLE METHOD WITH AN ANALYTIC SOLUTION FOR MULTI-SUBJECT FMRI DATA ANALYSIS

Trung Vu[a], Hanlu Yang[a], Francisco Laport[a,b], Ben Gabrielson[a], Vince D. Calhoun[c], Tülay Adali[a]

[a] Department of CSEE, University of Maryland, Baltimore County, MD 21250, USA
[b] CITIC Research Center, University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain
[c] TReNDS, Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA 30303, USA

MLSP Machine Learning for Signal Processing Lab

UMBC

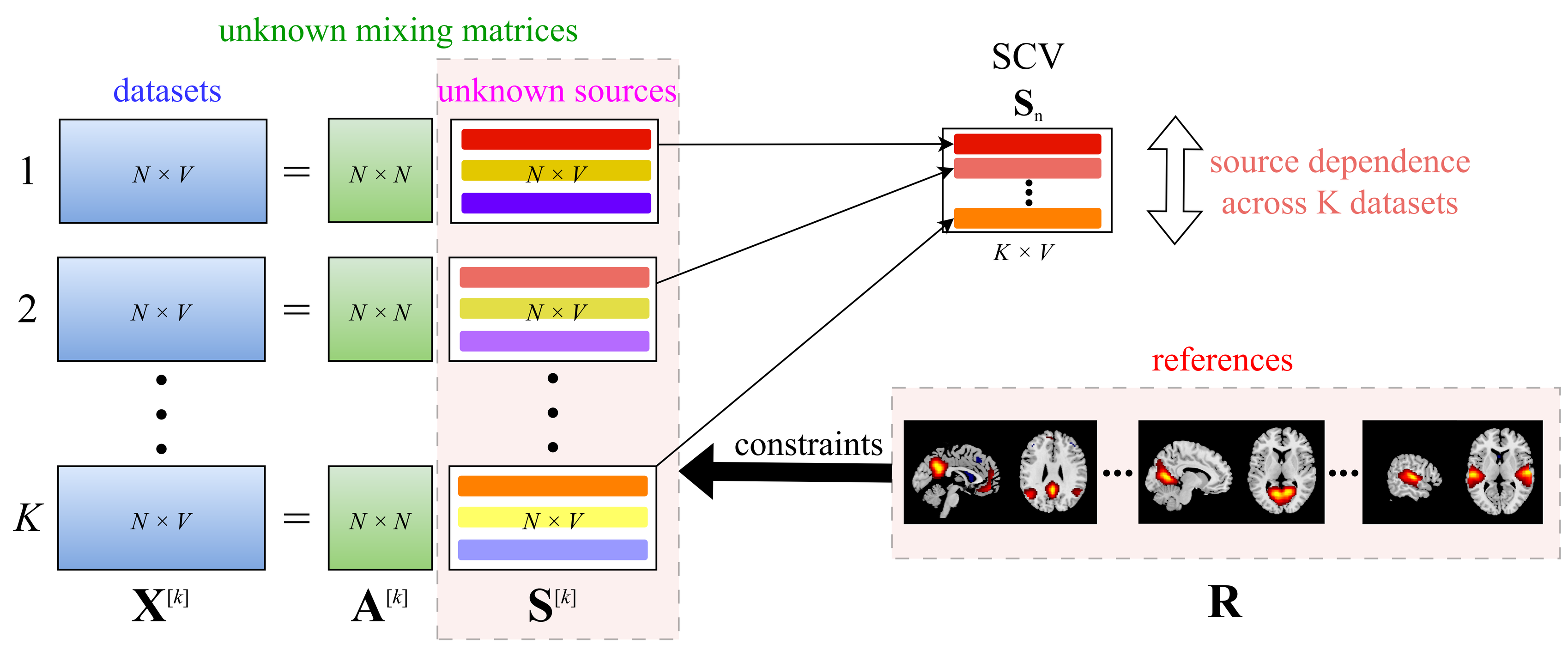
Introduction

- Source separation across multiple datasets is applied to various **neuroimaging** domains including multi-subject fMRI data analysis
- Prior knowledge about the sources or the mixing matrices can be used as references to guide the optimization to avoid sub-optimal solutions and increase the quality of source separation
- Existing approaches that use references as optimization constraints suffer from either high computational complexity or the inability to capture variability among the datasets

Contributions

- Propose a simple yet efficient method, named RGCA, for source separation of multiple datasets that uses source templates as references
- Establish an analytic solution that enables an efficient implementation of RGCA
- Demonstrate RGCA obtains competitive performance while having a runtime superior to other regression methods

Multi-Dataset Source Separation with References

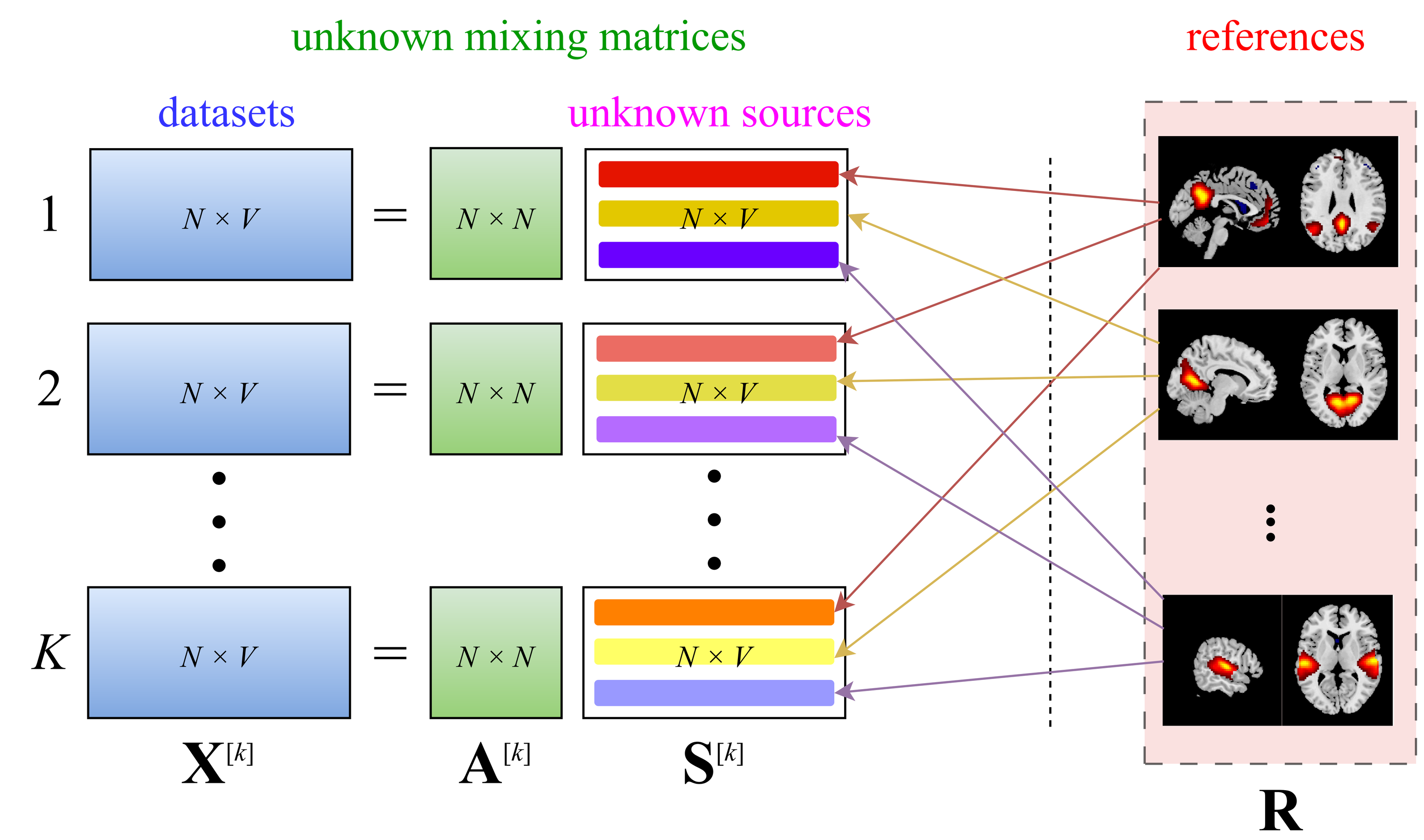


Constrained independent vector analysis (cIVA)

$$\min_{\boldsymbol{W}} \mathcal{J}_{\text{IVA}}(\boldsymbol{W}) \text{ s.t. } \epsilon(\boldsymbol{r}_m, \boldsymbol{y}_m^{[k]}) \geq \rho_m^{[k]}, \quad \forall m = 1, \dots, M \text{ and } k = 1, \dots, K$$

where $\epsilon(\boldsymbol{r}_m, \boldsymbol{y}_m^{[k]})$ is similarity between the mth reference $\boldsymbol{r}_m$ and the mth estimated source in the kth dataset $\boldsymbol{y}_m^{[k]}$; and $\rho_m^{[k]}$ is the corresponding threshold value (constraint parameter)

Reference-Guided Component Analysis (RGCA)



- References is used to guide the identification of sources in each dataset separately

$$\min_{\boldsymbol{W} \in \mathbb{R}^{M \times N}} \frac{1}{2V} \|\boldsymbol{R} - \boldsymbol{W}\boldsymbol{X}\|_F^2 \qquad \text{s.t. } \boldsymbol{W}\boldsymbol{W}^\top = \boldsymbol{I}_M,$$

- Orthogonality constraint on demixing matrix aids in interpretability

- Relaxing the orthogonality constraint with regularization

$$\min_{\boldsymbol{W} \in \mathbb{R}^{M \times N}} \frac{1}{2V} \|\boldsymbol{R} - \boldsymbol{W}\boldsymbol{X}\|_F^2 + \frac{\lambda}{4} \|\boldsymbol{W}\boldsymbol{W}^\top - \boldsymbol{I}_M\|_F^2$$

- A simple analytic solution exists by setting the gradient to $\mathbf{0}$

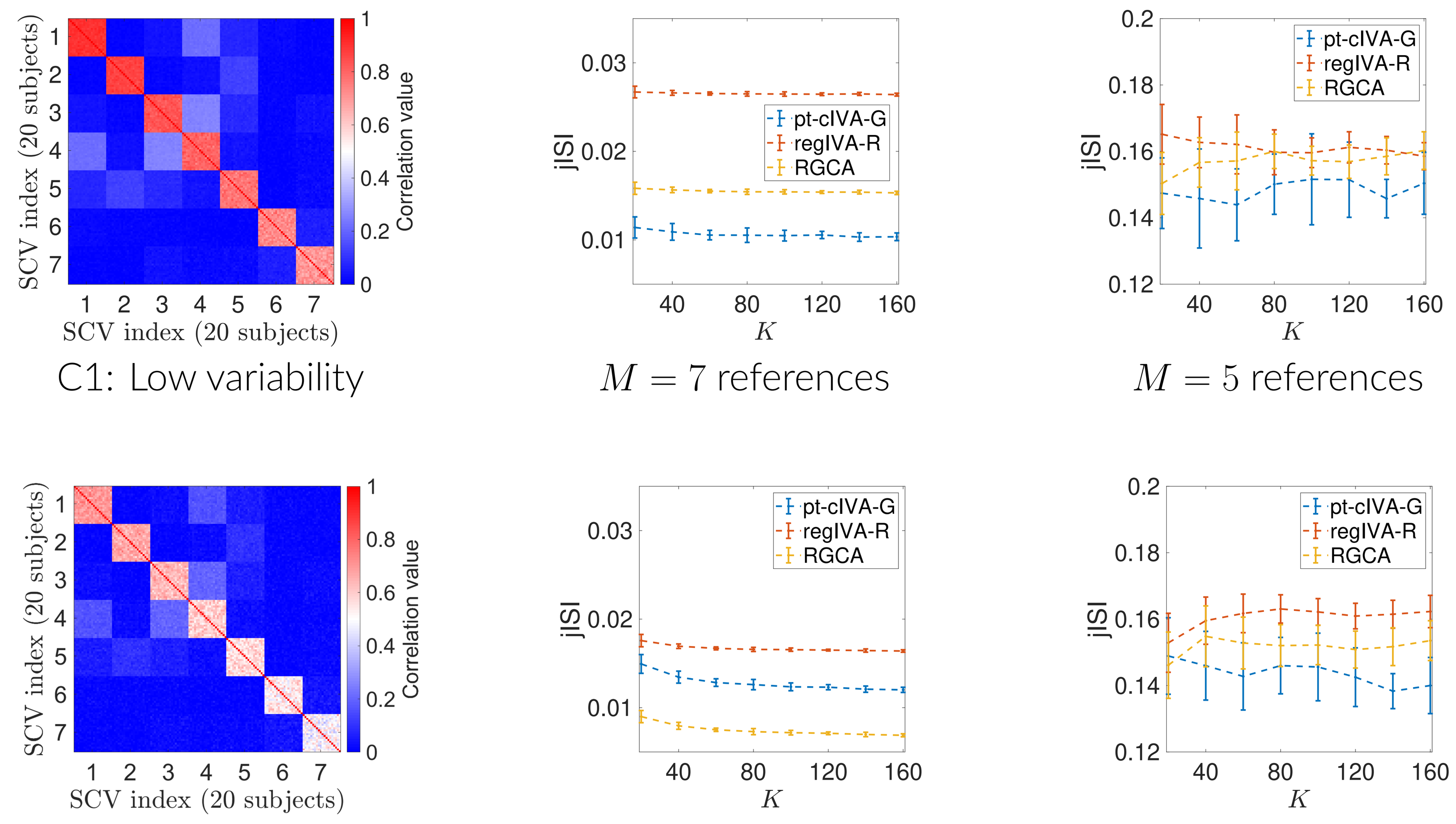
$$\nabla f(\boldsymbol{W}) = \lambda \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{W} + (1 - \lambda)\boldsymbol{W} - \frac{1}{V} \boldsymbol{R}\boldsymbol{X}^\top = \mathbf{0}$$

RGCA

- effectively retains source dependence across datasets via the common use of reference guidance
- *automatically aligns* source components across datasets
- *balances* the trade-off between fitting to references and orthogonality of the demixing matrix through λ

fMRI-like Data Simulation Results

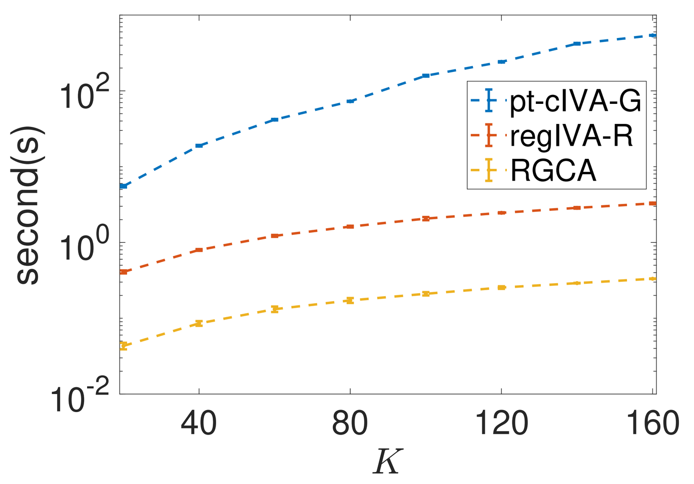
Joint inter-symbol-interference (jISI)



C1: Low variability — $M = 7$ references — $M = 5$ references

C2: High variability — $M = 7$ references — $M = 5$ references

Runtime



- pt-cIVA-G performs the best regarding jISI score thanks to the powerful properties of *the IVA framework*
- regIVA-R performs the worst since the estimated sources are not necessarily independent or uncorrelated
- RGCA performs relatively well compared with pt-cIVA-G, achieving the highest jISI when the *variability is high*
- pt-cIVA-G has the slowest runtime (quadratic in K) while RGCA yields the fastest (linear in K), almost 20 times faster

Summary

- RGCA leverages prior information on the sources (templates) as references for the solution and can be solved analytically, facilitating fast implementations
- Simulation with fMRI-like data illustrates the separation capability of RGCA while capturing well the variability among the datasets
- RGCA offers a robust and scalable solution to group fMRI studies, enabling fast joint analysis of thousands of subjects

This work is supported in part by the grants NIH R01MH118695, NIH R01MH123610, NIH R01AG073949, NSF 2316420, and Xunta de Galicia ED481B 2022/012