Sparse Independent Component Analysis with an Application to Cortical Surface fMRI Data in Autism

Zihang Wang[1], Irina Gaynanova[2], Aleksandr Aravkin[3] and Benjamin Risk[1]

[1] *Department of Biostatistics and Bioinformatics, Emory University,*

[2] *Department of Biostatistics, University of Michigan,* [3] *Department of Applied Mathematics, University of Washington*

thebrisklab.org

Independent Component Analysis (ICA)

- Let $X \in \mathbb{R}^{P \times T}$ represent the fMRI time series for a single subject, where each column is a vectorized image of dimension P, and T is the number of time points.
- The noisy ICA model with isotropic noise is defined as

$$X = SM + N.$$

- $S \in \mathbb{R}^{P \times Q}$: matrix of non-Gaussian components (sources) with $Q < T$.
- $M \in \mathbb{R}^{Q \times T}$ is the mixing matrix.
- $N \in \mathbb{R}^{P \times T}$ is a matrix of normal random variables.

Background

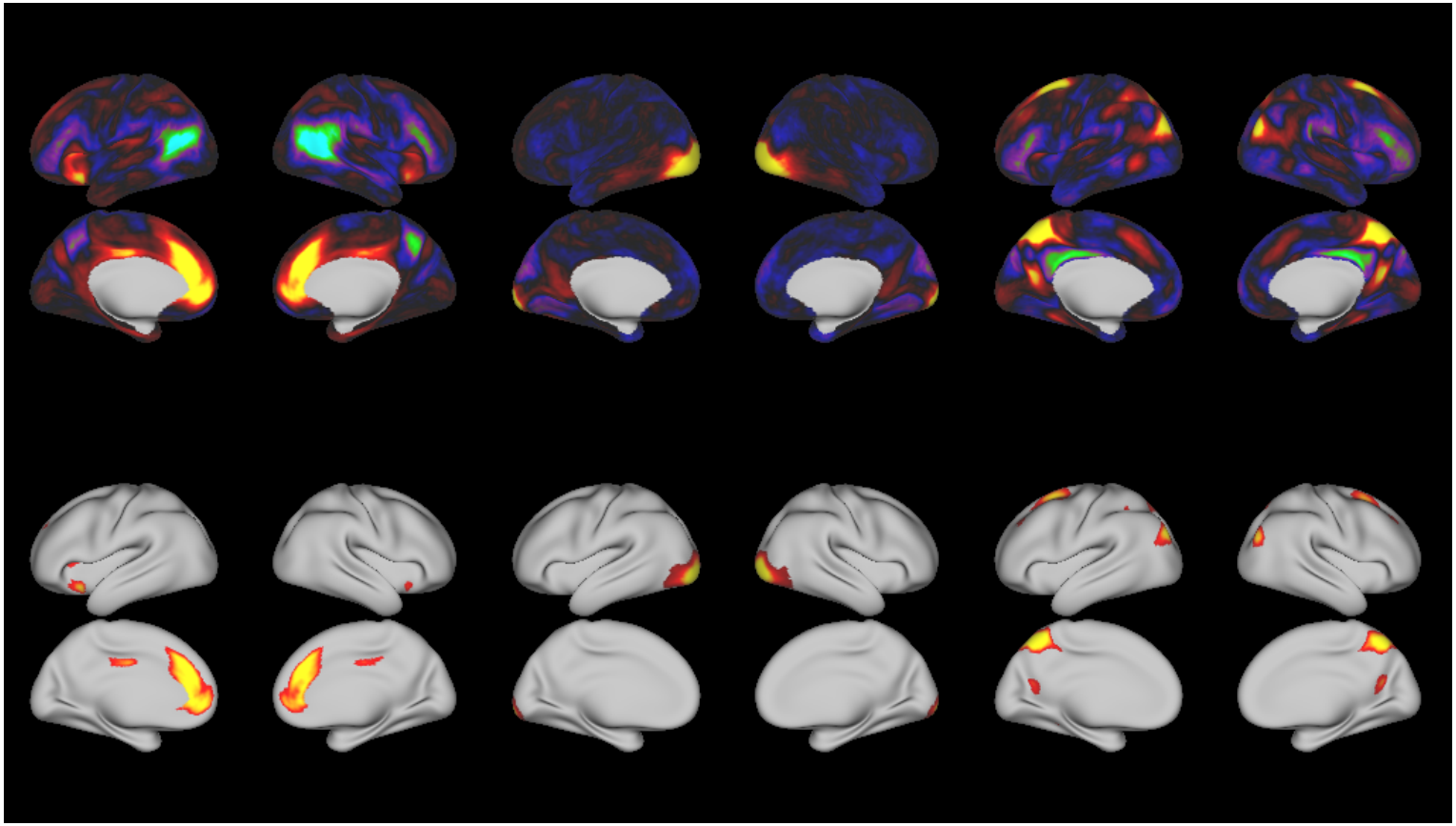
- ICA is commonly applied to resting-state functional magnetic resonance imaging (fMRI) data (Smith et al. 2009).
- The independent components (ICs) are known as "resting-state networks," which are sparse brain regions that tend to co-activate while the brain is not engaged in any tasks (Biswal et al. 2010).
- The correlations between the time courses of the components (functional connectivity) reveal brain communication patterns (Smith et al. 2015).
- Thresholding is commonly applied to components after estimation to aid visualization and interpretation (Smith et al. 2009).
- Subject-specific time courses are back reconstructed using the original unthresholded components (Calhoun et al. 2001).

Toy example

Example Source Mixture

ICA and thresholding

- Thresholding: interpreted as if only the most extreme locations are activated.



ICA in ASD

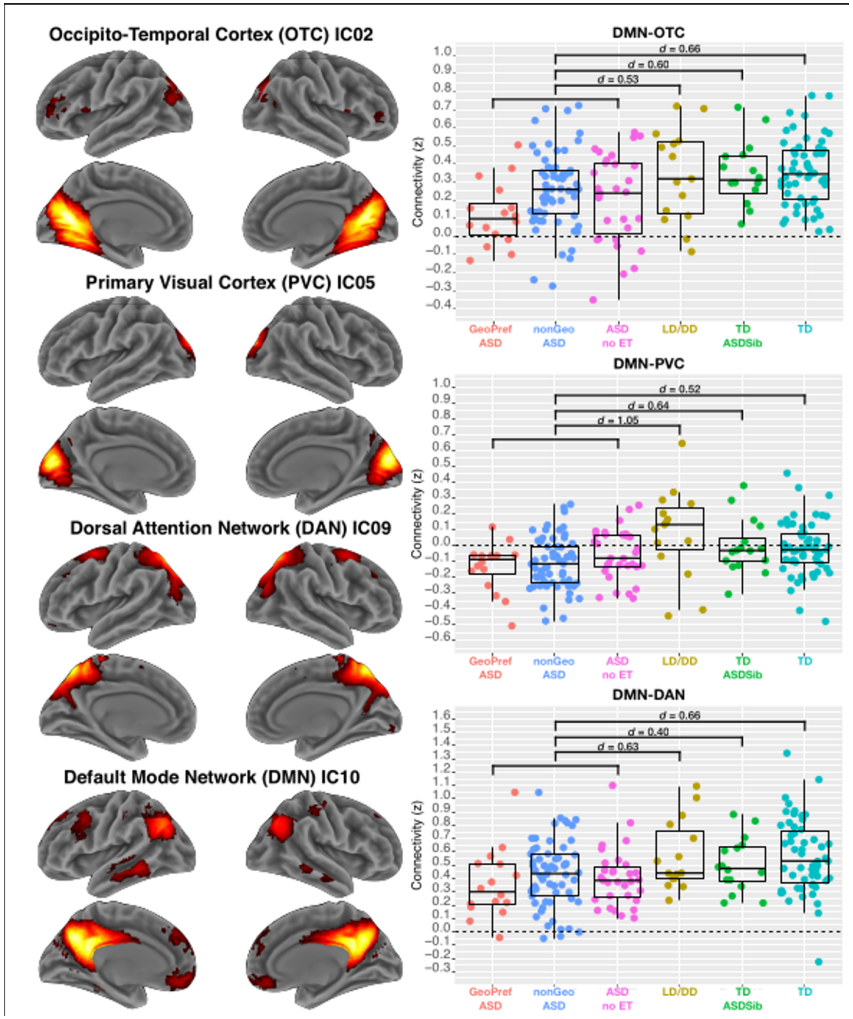


Figure: Example application of ICA with post-hoc thresholding in a functional connectivity study of autism (Lombardo et al. 2019).

Previous methods

- Fast ICA and Infomax ICA are popular (Hyvarinen 1999, Bell and Sejnowski 1995): dense estimation.
- Sparse PCA (Zou et al. 2006, Shen and Huang 2008): independence is not enforced.
- Sparse Fast ICA (Ge et al. 2016): multiple tuning parameters, not exact sparsity.
- Sparse ICA entropy-bound minimization (SICA-EBM) (Boukouvalas et al. 2018): two tuning parameters, computationally costly, not exact sparsity.
- Hierarchical ICA models for multiple subjects also don't estimate sparse components (Lukemire et al. 2020, Mejia et al. 2020).

Challenges

1. Existing ICA algorithms are only capable of handling smooth objective functions.
2. Solving the ICA problem involves non-convex optimization.
3. An ICA method should be fast for neuroimaging data.

Laplace ICA Model

- X is centered and whitened, $\tilde{X} \in \mathbb{R}^{P \times Q}$ such that $\tilde{X}^\top \tilde{X} = I_Q$, PCA+ICA used to approximate noisy ICA (Beckmann and Smith 2004):
- Let $U \in \mathcal{O}^{Q \times Q}$ be an orthogonal transformation.
- We aim to solve

$$\underset{U}{\text{minimize}} \quad J(\tilde{X}U) \quad \text{s.t.} \quad U^\top U = I_Q,$$

where

$$J(\tilde{X}U) = -\sum_{i=1}^{P}\sum_{j=1}^{Q} \log p_j\left\{(\tilde{X}U)_{ij}\right\} = -\sum_{i=1}^{P}\sum_{j=1}^{Q} \log p_j(\tilde{x}_i^\top u_j),$$

$p_j(s) = e^{-\frac{|s-\mu|}{\lambda}}/2\lambda$ be the density of a component, where $\mu = 0$ and $\lambda = \sqrt{2}/2$.

Relax-and-Split Framework

Algorithm challenges:

- Non-smooth objective function, can't use Newton-type algorithms like ICA fixed point algorithm.
- Orthogonality constraint on U, which is non-convex.

These challenges can be solved through the Relax-and-Split framework (Zheng and Aravkin 2020).

The relaxed problem

$$\underset{U,V}{\text{minimize}} \quad J(V) + \frac{1}{2\nu}\|V - \tilde{X}U\|_F^2 \quad \text{s.t.} \quad U^\top U = I_Q,$$

where the auxiliary V explicitly models the independent components S, and parameter ν controls the level of relaxation.

Alternating Updates

Algorithm 1 Relax and split ICA

1: **Input:** U^0, V^0
2: Initialize: $k = 0$
3: **while** not converged **do**
4: $\quad V^{k+1} \leftarrow \arg\min_V J(V) + \frac{1}{2\gamma}\|V - \tilde{X}U^k\|_F^2$
5: $\quad U^{k+1} \leftarrow \arg\min_{U^\top U = I} \|V^{k+1} - \tilde{X}U\|_F^2$
6: $\quad k \leftarrow k+1$
7: **Output:** U^k, V^k

It only requires two ingredients:

- Algorithm for V update: element-wise gradient descent.
- Algorithm for U update: orthogonal Procrustes problem.

The Full Algorithm

Algorithm 2 Pseudo Code for Sparse ICA

Inputs: $X \in \mathbb{R}^{P \times T}$, $Q \in \mathbb{Z}^+$, $\nu > 0$, $\max_{iter} \in \mathbb{Z}^+$, $\epsilon \in \mathbb{R}$.
Preprocess: X is centered, whitened, SVD to get $\tilde{X} \in \mathbb{R}^{P \times Q}$.
Initialize:
$U^0 \in \mathcal{O}^{Q \times Q} \leftarrow$ A random orthogonal matrix;
$V^0 \in \mathbb{R}^{P \times Q} \leftarrow \tilde{X} U^0$
while $iter < \max_{iter}$ and $\max(|(|\text{diag}\left(U^{iter+1}(U^{iter})'\right)|) - 1|) > \epsilon$ **do**
    Update V**:** For $1 \le i \le P$, $1 \le j \le Q$,
    $V_{ij}^{iter+1} \leftarrow \left(\left|(\tilde{X}U^{iter})_{ij}\right| - \sqrt{2}\nu\right)_+ \cdot \text{sign}\left((\tilde{X}U^{iter})_{ij}\right)$
    Update U**:** Apply SVD decomposition $\tilde{X}^T V^{iter+1} = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$.
    $U^{iter+1} \leftarrow \tilde{U}\tilde{V}^\top$
    $iter \leftarrow iter + 1$
Outputs: $\hat{S} = V^{iter}$, $\hat{M} = \left(\hat{S}^\top \hat{S}\right)^{-1} \hat{S}^\top X$.

Repeat over a grid of ν with warm starts and select ν using a BIC-like criterion (Allen and Maletić-Savatić 2011).

Group ICA

- Group ICA for multiple subjects (Calhoun et al. 2001).
- SVD on each subject, save Q_i left singular vectors, concatenate, conduct another SVD and retain Q left singular vectors.
- Conduct sparse ICA on group singular vectors to obtain group independent components $\hat{S}_g$.
- Estimate subject-specific time courses by projecting subject data onto the group ICA basis: $\hat{M}_i = \left(\hat{S}_g^\top \hat{S}_g\right)^{-1} \hat{S}_g^\top X_i$.

Equivalency between group ICA SVD steps and JIVE

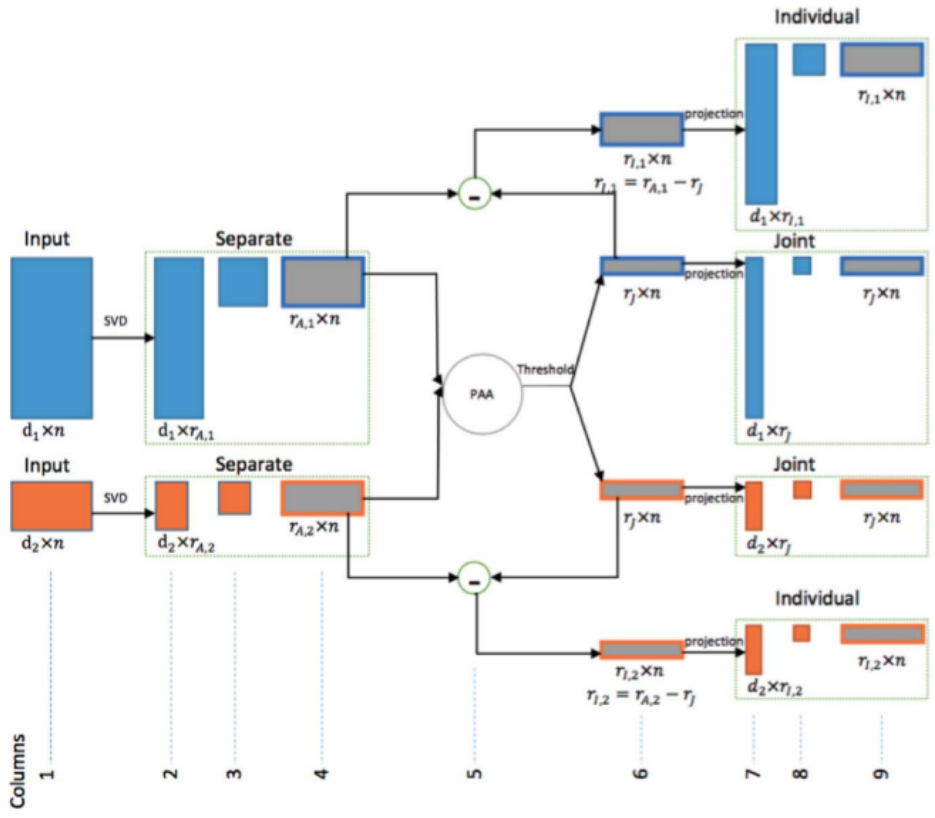


Fig. 1. Analytical steps of the JIVE algorithm. Raw data on the left are input to separate SVD low rank approximations (green boxes). Next steps use the separate SVD scores (all score matrices are shown as gray boxes), first thresholding the Principal Angle Analysis to obtain joint scores, then performing basis subtraction to obtain individual scores. Finally, projection gives the JIVE loadings.

Figure: JIVE schematic from Yu et al. (2017).

Spatio-temporal Simulation Design

- Three true source signals ($Q = 3$) mixed over 50 time points ($T = 50$).
- Each IC is a 33×33 image, a vector of length $P = 1089$.
- Active pixels are in the shape of “1", “2 2", or “3 3 3" with values varying from 0.5 to 1. All inactive pixels are set to 0.
- Noise level: low, medium, and high variance signal to variance noise ratio (SNR). Gaussian random fields.
- Comparisons: Fast ICA with tanh non-linearity (Hyvarinen 1999), Infomax ICA (Cardoso 1997), SICA-EBM (Boukouvalas et al. 2018), and Sparse Fast ICA with tanh non-linearity (Ge et al. 2016).
- Scale and permutation invariant root mean squared error (PRMSE):

$$\text{PRMSE}(S, \hat{S}) = \left(\frac{1}{PQ} \underset{P \in \mathcal{P}}{\text{argmin}} \| S - P\hat{S} \|_F^2 \right)^{1/2}.$$

Simulation Results: Components

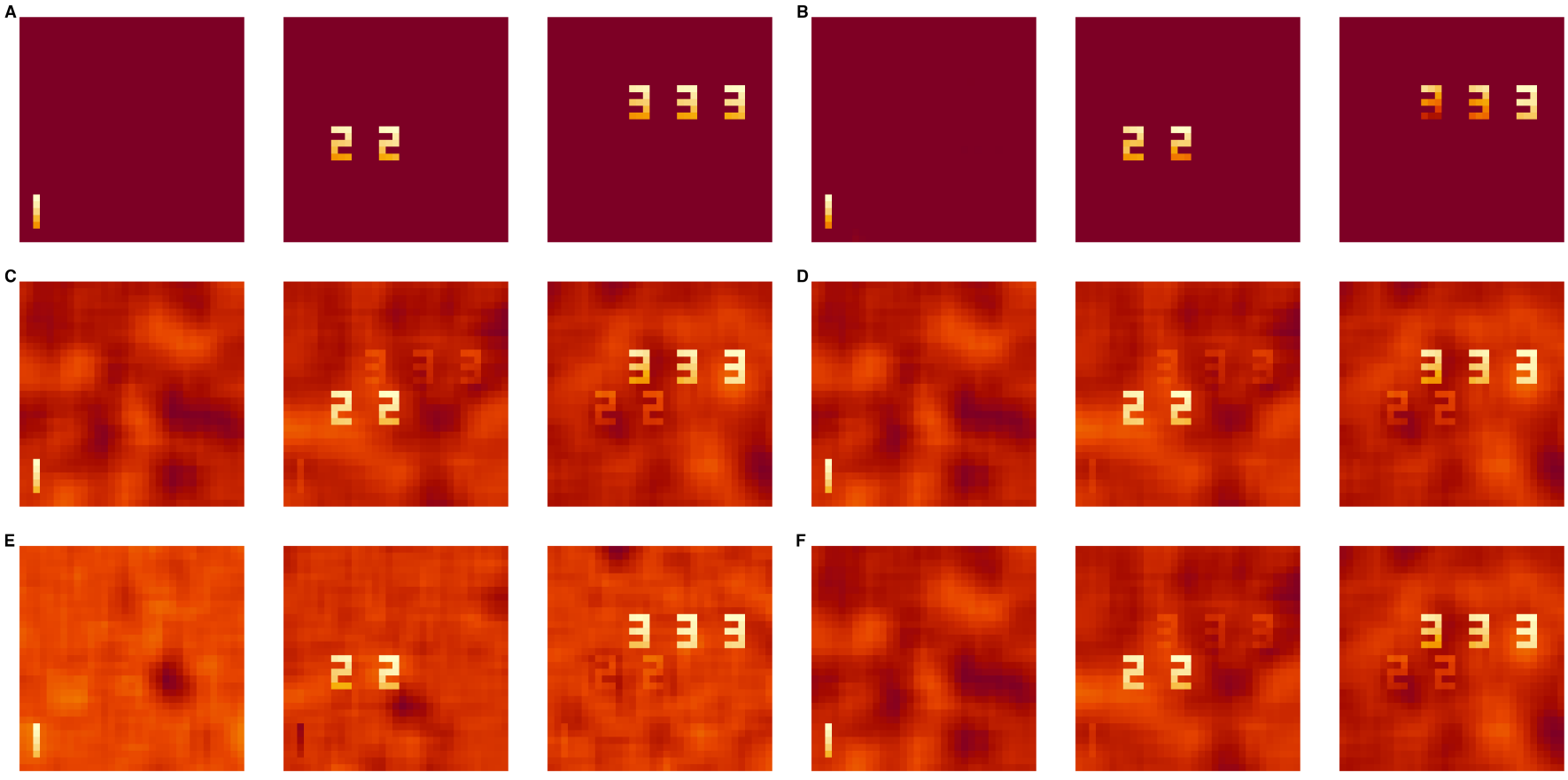
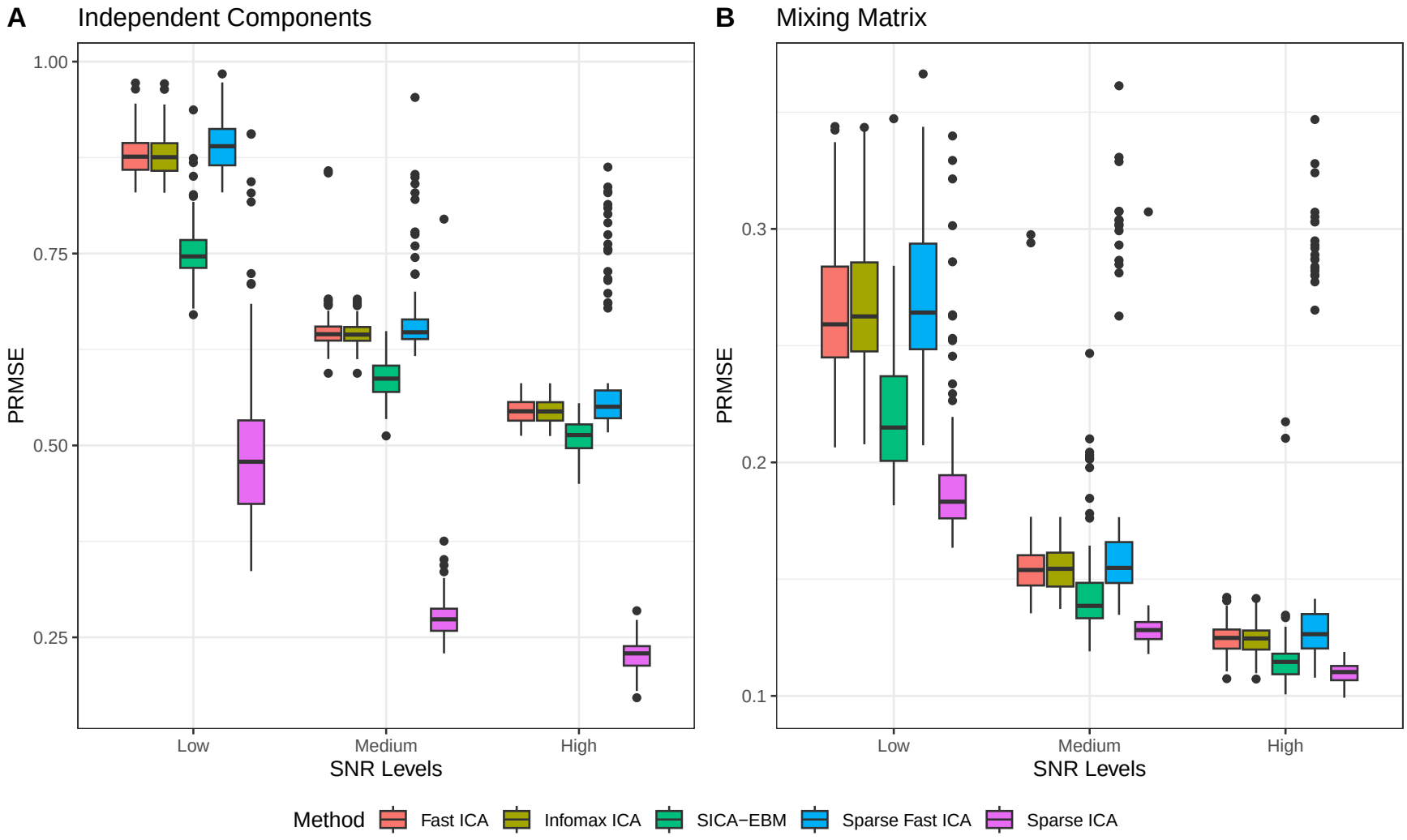


Figure: A: The true ICs. Zeros appear as maroon. B: Sparse ICA. C: Infomax ICA. D: Fast ICA. E: SICA-EBM. F: Sparse Fast ICA.

Simulation Results: Mixing matrix



Computation Time

	Sparse ICA	Fast ICA	Infomax ICA	SICA-EBM	Sparse Fast ICA
Low SNR	0.0080	0.0022	0.0141	16.0053	0.0055
Medium SNR	0.0079	0.0023	0.0186	22.9243	0.0024
High SNR	0.0081	0.0023	0.0177	19.9347	0.0019
High-dimensional	1.9147	0.4323	2.3359	2233.4951	0.0253

Table: Average computation times in seconds over 100 replications for Sparse ICA and other methods. SICA-EBM and Sparse Fast ICA are implemented in Matlab, other ICA methods are in R and RCpp.

Sparse ICA competitive, much faster than Infomax ICA and SICA-EBM.

ABIDE Study

- Resting-state fMRI data from school-age children selected from the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al. 2014; 2017).
- 144 Autistic (ASD) children (8-13 years old), 252 non-ASD children.
- $n = 312$ passed motion QC criteria.
- 30 group components estimated.
- Subject-specific time courses were estimated to construct correlation matrices with the dimension of 30×30.
- Account for selection bias from motion QC using augmented inverse probability of inclusion weighted estimation (Robins et al. 1994, Robins 2000), with SuperLearner for outcome and propensity models, follows Nebel et al. (2022).
- FDR-corrected p-values (Benjamini and Hochberg 1995).

Results

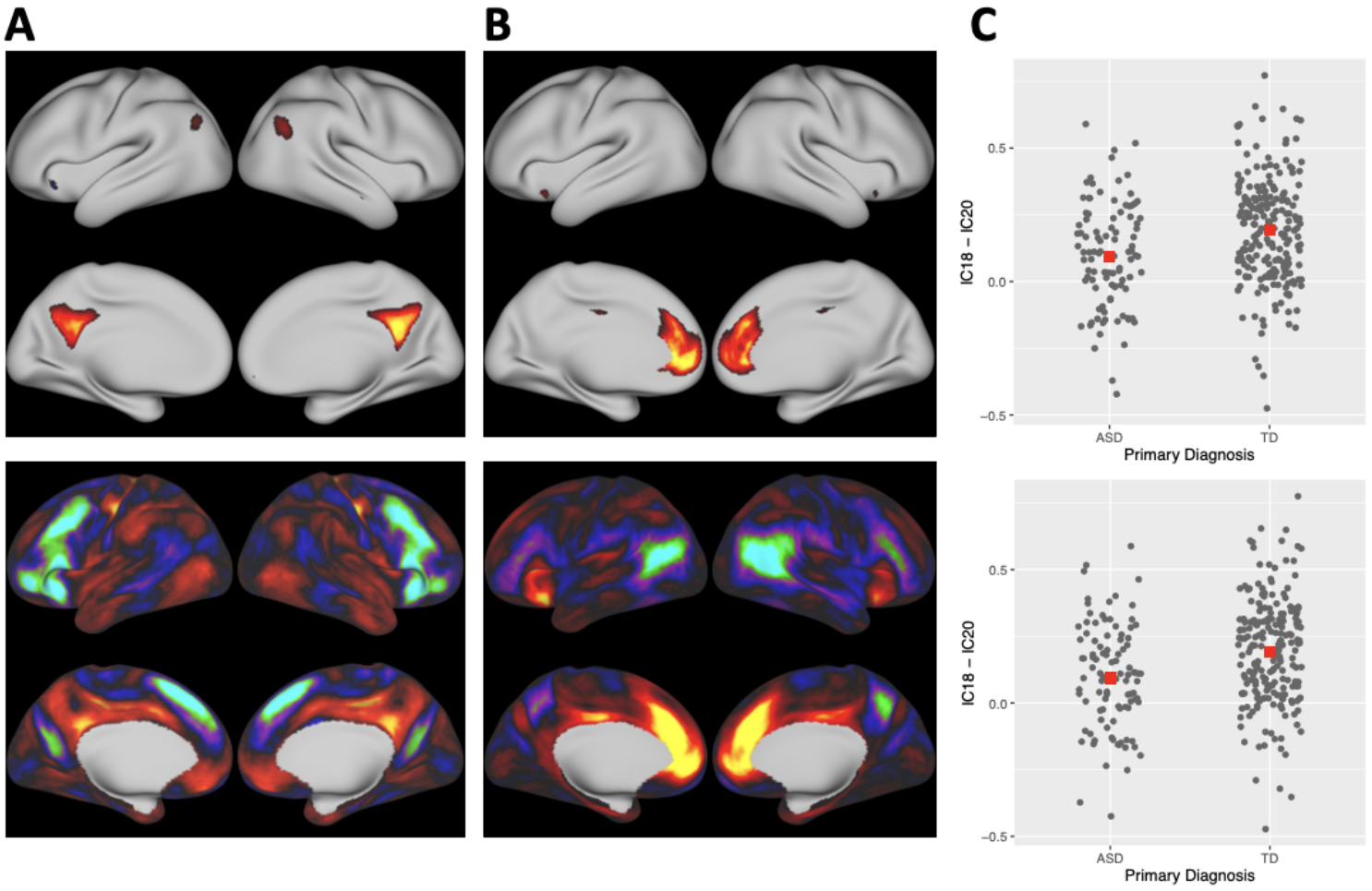


Figure: Panel A: medial posterior default mode network. Panel B: anterior part of the default mode network.

Results

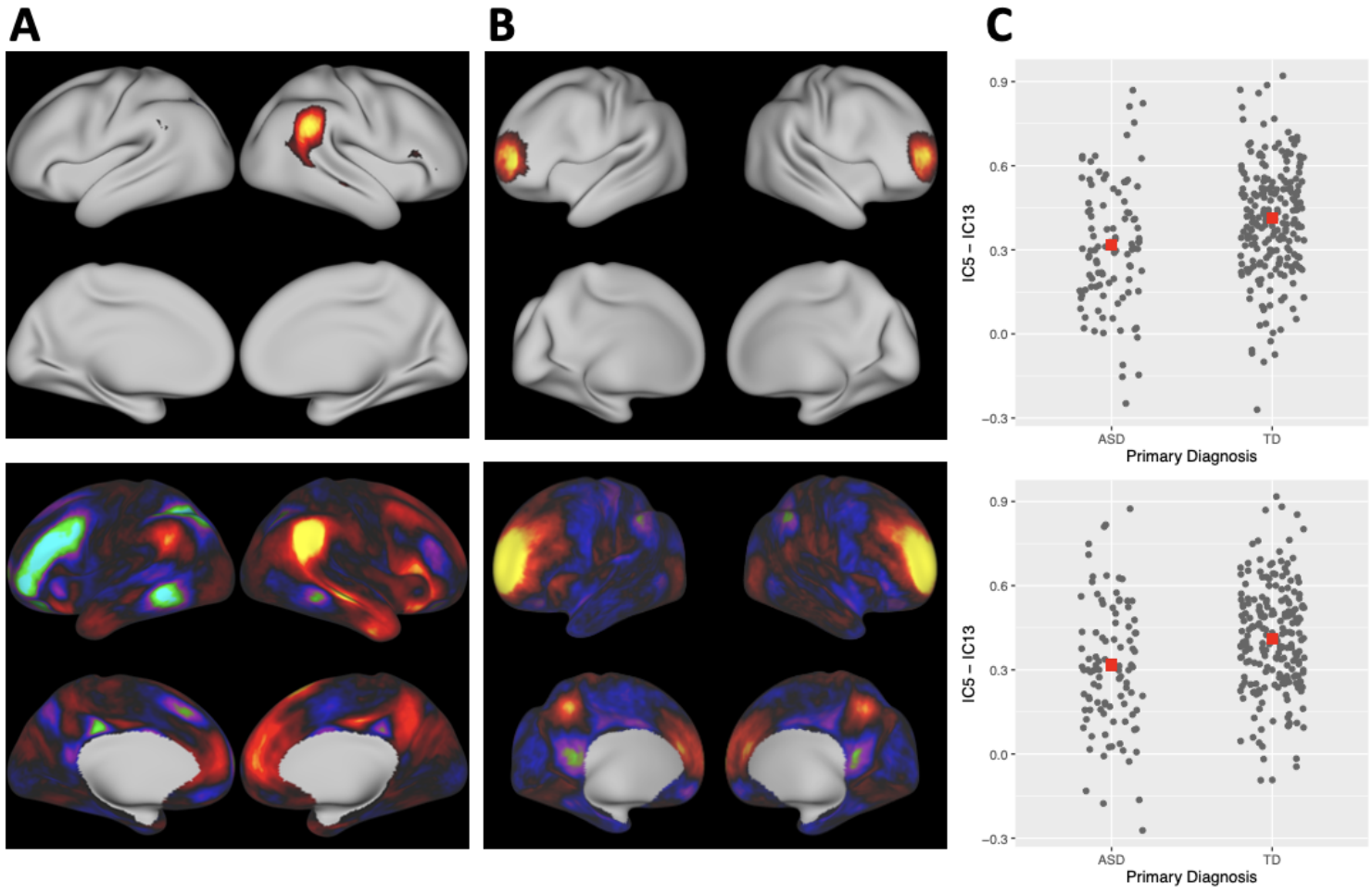


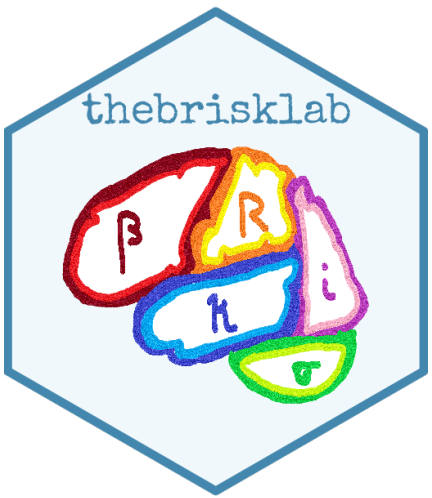
Figure: Panel A: temporal parietal junction: information processing. Panel B: frontoparietal network: executive function

Summary

- A novel ICA method that estimates sparse independent components using a relax-and-split algorithm and is computationally fast.
- Sparse features improve interpretation and accuracy.
- Important for downstream functional connectivity analyses.
- In the real data application, we found:
  - decreased connectivity between the anterior and posterior parts of the default mode network in ASD versus non-ASD children.
  - decreased connectivity between components corresponding to brains regions associated with executive function and information processing in ASD.

Paper and R Package

- Wang Z., Gaynanova, I., Aravkin, A., Risk, B. B. (2024). Sparse Independent Component Analysis with an Application to Cortical Surface fMRI Data in Autism. *Journal of the American Statistical Association*, 1–22.
- R package `SparseICA` available on: https://github.com/thebrisklab/SparseICA.



Acknowledgments

We thank Peng Zheng (University of Washington), Lei Zhou (Emory University), Sijian Fan (University of South Carolina).

BBR and ZW supported by the National Institute of Mental Health of the National Institutes of Health under award number R01 MH129855. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References I

G. I. Allen and M. Maletić-Savatić. Sparse non-negative generalized pca with applications to metabolomics. *Bioinformatics*, 27(21):3029–3035, 2011.

C. F. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the national academy of sciences*, 107(10):4734–4739, 2010.

Z. Boukouvalas, Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı. Sparsity and independence: Balancing two objectives in optimization for source separation with application to fmri analysis. *Journal of the Franklin Institute*, 355(4):1873–1887, 2018.

V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151, 2001.

J.-F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal processing letters*, 4(4):112–114, 1997.

A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

A. Di Martino, D. O'connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific data*, 4 (1):1–15, 2017.

R. Ge, Y. Wang, J. Zhang, L. Yao, H. Zhang, and Z. Long. Improved fastica algorithm in fmri data analysis using the sparsity property of the sources. *Journal of neuroscience methods*, 263:103–114, 2016.

References II

A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3):626–634, 1999.

M. V. Lombardo, L. Eyler, A. Moore, M. Datko, C. Carter Barnes, D. Cha, et al. Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *elife*, 8:e47427, 2019.

J. Lukemire, Y. Wang, A. Verma, and Y. Guo. Hint: A hierarchical independent component analysis toolbox for investigating brain functional networks using neuroimaging data. *Journal of neuroscience methods*, 341:108726, 2020.

A. F. Mejia, M. B. Nebel, Y. Wang, B. S. Caffo, and Y. Guo. Template independent component analysis: targeted and reliable estimation of subject-level brain networks using big data population priors. *Journal of the American Statistical Association*, 115(531):1151–1177, 2020.

M. B. Nebel, D. E. Lidstone, L. Wang, D. Benkeser, S. H. Mostofsky, and B. B. Risk. Accounting for motion in resting-state fmri: What part of the spectrum are we characterizing in autism spectrum disorder? *NeuroImage*, 257:119296, 2022.

J. M. Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association*, volume 1999, pages 6–10. Indianapolis, IN, 2000.

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

H. Shen and J. Z. Huang. Sparse principal component analysis via regularized low rank matrix approximation. *Journal of multivariate analysis*, 99(6):1015–1034, 2008.

S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, 106(31):13040–13045, 2009.

S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. Behrens, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature neuroscience*, 18(11):1565–1567, 2015.

Q. Yu, B. B. Risk, K. Zhang, and J. Marron. Jive integration of imaging and behavioral data. *NeuroImage*, 152:38–49, 2017.

P. Zheng and A. Aravkin. Relax-and-split method for nonconvex inverse problems. *Inverse Problems*, 36(9):095013, 2020.

H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.