New Interpretable Patterns and Discriminative Features from Brain Functional Network Connectivity Using Dictionary Learning

F. Ghayem[1], H. Yang[1], F. Kantar[1], S.-J. Kim[1], V. D. Calhoun[2], T. Adali[1]

[1]MLSP-lab, CSEE Dept.
University of Maryland Baltimore County, Baltimore, USA

[2]Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS)
Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, USA

2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)

June 04-10, Rhodes Island, Greece.

The increasing importance of data-driven techniques in fMRI

Traditional neuroscience approaches

- Region of interest (ROI) analysis: Selecting specific brain regions involved in a process or behavior
- Forming hypotheses based on prior knowledge or assumptions about the brain
- Valuable in advancing our understanding of the brain
- Limited by the assumptions and biases

Data-driven techniques

- Machine Learning and ICA
- Without being constrained by pre-existing hypotheses or assumptions
- Identifying complex patterns that might be unexpected or unknown
- Analysing large amounts of data

Leveraging ICA and DL jointly being learned with a classifier for identifying interpretable patterns in fMRI

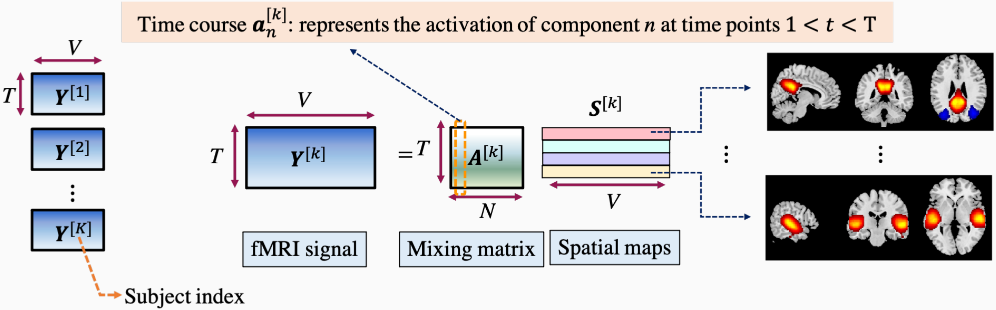
Motivation

- Developing methods for identifying interpretable patterns that can distinguish between HCs and patients
- Aim to improve understanding and diagnosis of these disorders.

Methodology

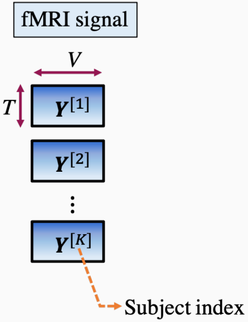
- Leveraging the advantages of ICA and DL to:
    - Extract powerful features from resting-state fMRI data
    - Identify novel, interpretable biomarkers.

Uncovering brain networks: ICA analysis of fMRI data for spatial maps and time courses



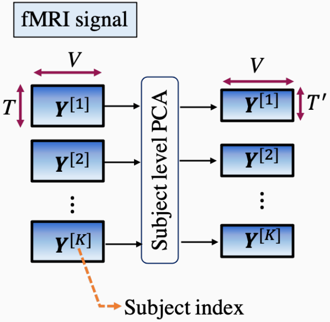
Group ICA and back-reconstruction for multi-subject brain network analysis

Group ICA + Back-reconstruction



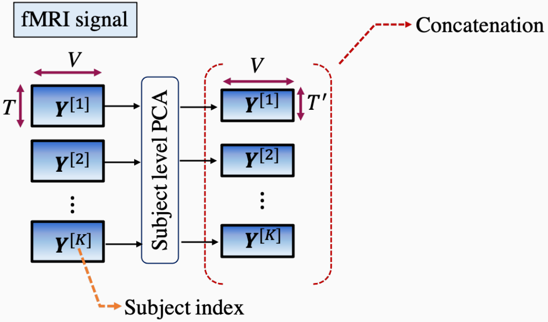
Group ICA and back-reconstruction for multi-subject brain network analysis

Group ICA + Back-reconstruction



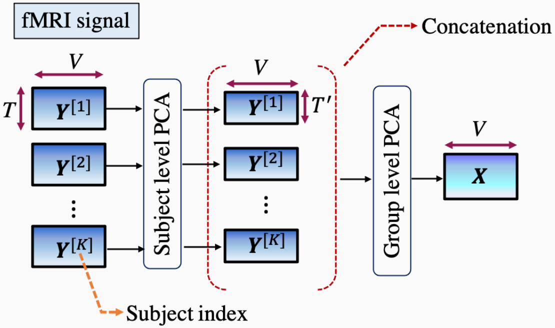
Group ICA and back-reconstruction for multi-subject brain network analysis

Group ICA + Back-reconstruction

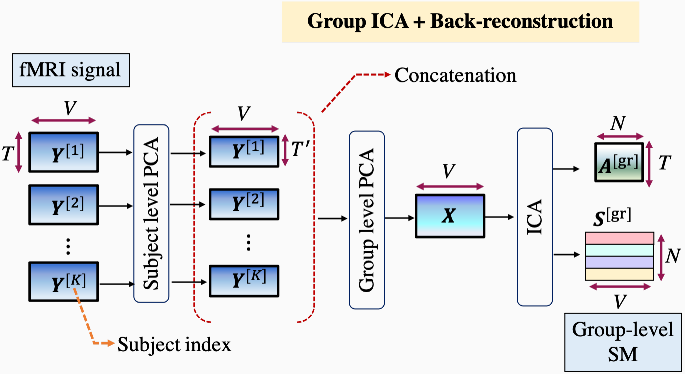


Group ICA and back-reconstruction for multi-subject brain network analysis

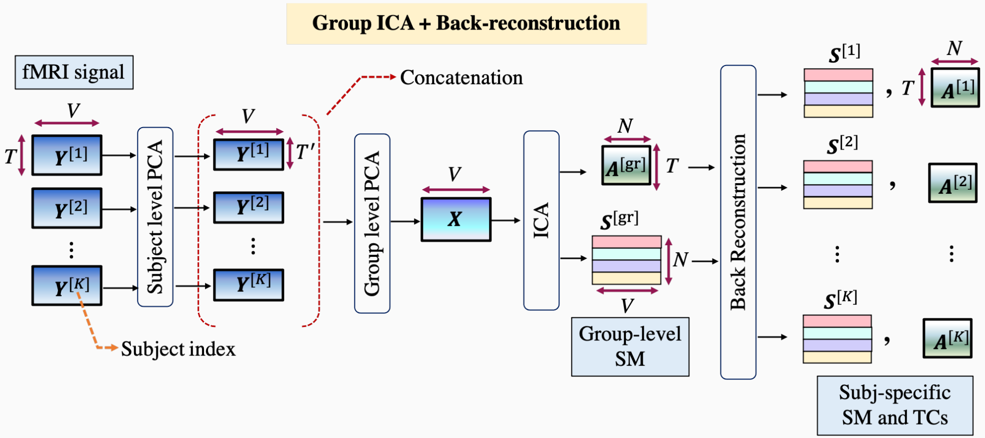
Group ICA + Back-reconstruction



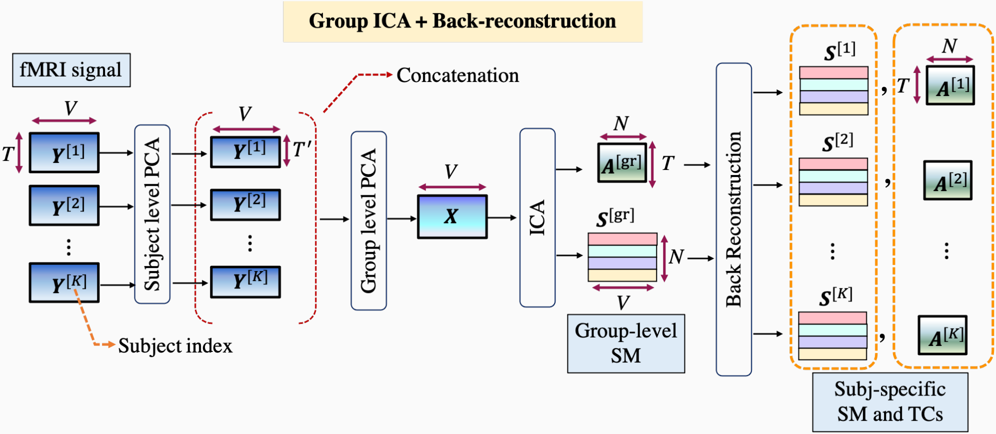
Group ICA and back-reconstruction for multi-subject brain network analysis



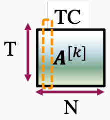
Group ICA and back-reconstruction for multi-subject brain network analysis



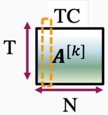
Group ICA and back-reconstruction for multi-subject brain network analysis



Inter-network relationships are studied through temporal functional network connectivity (tFNC)



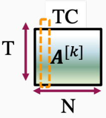
Inter-network relationships are studied through temporal functional network connectivity (tFNC)



tFNC: Pearson correlations between pairs of time courses (TCs)

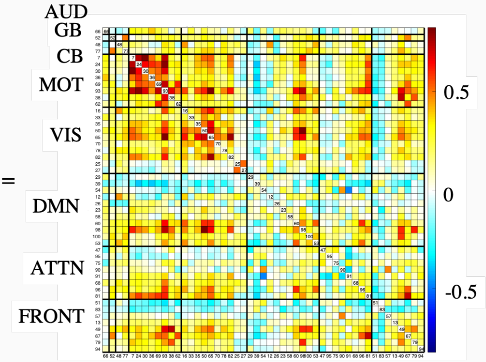
tFNC: $(A^{[k]})^T A^{[k]} =$

Inter-network relationships are studied through temporal functional network connectivity (tFNC)

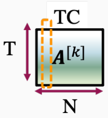


tFNC: Pearson correlations between pairs of time courses (TCs)

tFNC: $(\boldsymbol{A}^{[k]})^T \boldsymbol{A}^{[k]} =$

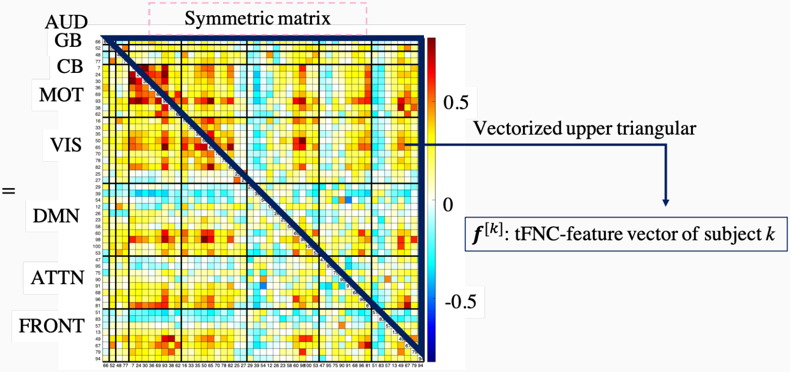


Inter-network relationships are studied through temporal functional network connectivity (tFNC)



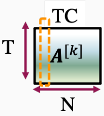
tFNC: Pearson correlations between pairs of time courses (TCs)

tFNC: $(A^{[k]})^T A^{[k]} =$

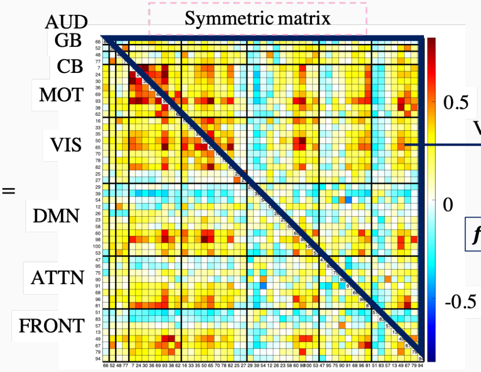


Inter-network relationships are studied through temporal functional network connectivity (tFNC)

tFNC: Pearson correlations between pairs of time courses (TCs)



tFNC: $(\boldsymbol{A}^{[k]})^T \boldsymbol{A}^{[k]} =$

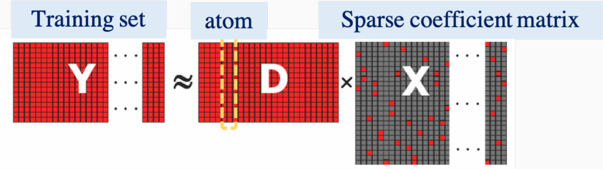


Vectorized upper triangular

$\boldsymbol{f}^{[k]}$: tFNC-feature vector of subject k

Strength and directionality of inter-network relationships

Sparse dictionary learning for FNC feature extraction



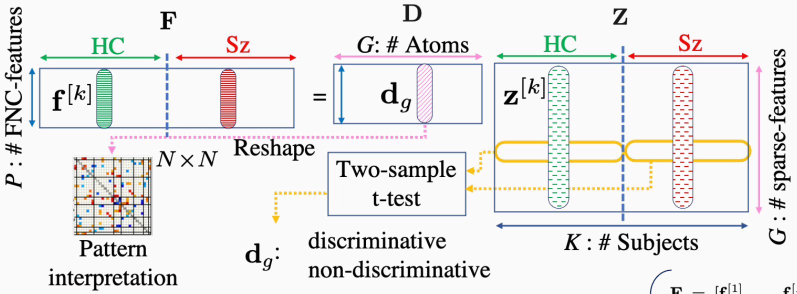
Dictionary learning problem: $\min_{\mathbf{D},\mathbf{X}} \sum_{i=1}^{L} \frac{1}{2}||\mathbf{y}_i - \mathbf{D}\mathbf{x}_i||_2 = \frac{1}{2}||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2$ s.t. $\mathbf{D} \in \mathcal{D}$, $\boxed{\mathbf{X} \in \mathcal{X}}$

Alternating Minimization

Start with $(\mathbf{D}^{(0)}, \mathbf{X}^{(0)})$. Alternate between two steps:

1. Sparse representation: $\mathbf{X}^{(k+1)} = \mathrm{argmin}_{\mathbf{X} \in \mathcal{X}} \frac{1}{2}||\mathbf{Y} - \mathbf{D}^{(k)}\mathbf{X}||_F^2$
   - OMP, IST, SL0
2. Dictionary update: $\mathbf{D}^{(k+1)} = \mathrm{argmin}_{\mathbf{D} \in \mathcal{D}} \frac{1}{2}||\mathbf{Y} - \mathbf{D}\mathbf{X}^{(k+1)}||_F^2$
   - MOD, KSVD

Sparse representation of tFNCs can reveal new interpretable patterns and discriminant features



$$
\begin{cases}
\mathbf{F} = [\mathbf{f}^{[1]}, \ldots, \mathbf{f}^{[K]}] \in \mathbb{R}^{P \times K} \\
\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_G] \in \mathbb{R}^{P \times G} \\
\mathbf{Z} = [\mathbf{z}^{[1]}, \ldots, \mathbf{z}^{[K]}] \in \mathbb{R}^{G \times K} \\
\mathbf{f}^{[k]} = \sum_{i=1}^{G} z_g \mathbf{d}_g
\end{cases}
$$

Jointly learn a linear classifier and a dictionary for tFNC sparse representation to improve discrimination

Problem formulation

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \frac{1}{2}\|\mathbf{F}-\mathbf{D}\mathbf{Z}\|_F^2+\lambda\cdot r(\mathbf{Z})+\frac{\beta}{2}\|\mathbf{L}_{\text{tr}}-\mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2,$$

$$\text{s.t. } \mathbf{D}\in\mathcal{D}:=\{\mathbf{D}:\|\mathbf{d}_g\|_2=1, g=1,2,\ldots,G\}.$$

Binary group labels:

$$\begin{cases} \mathbf{L}_{\text{tr}} = [\mathbf{l}_{\text{tr}}^{[1]},\ldots,\mathbf{l}_{\text{tr}}^{[K_{\text{tr}}]}] \\ \mathbf{l}^{(\text{HC})} = [0,1]^T \text{ and } \mathbf{l}^{(\text{Sz})} = [1,0]^T \end{cases}$$

Jointly learn a linear classifier and a dictionary for tFNC sparse representation to improve discrimination

Problem formulation

Sparse representation error

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \frac{1}{2}\overbrace{\|\mathbf{F}-\mathbf{DZ}\|_F^2}+\lambda\cdot r(\mathbf{Z})+\frac{\beta}{2}\|\mathbf{L}_{\text{tr}}-\mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2,$$
$$\text{s.t.}\ \mathbf{D}\in\mathcal{D}:=\{\mathbf{D}:\|\mathbf{d}_g\|_2=1, g=1,2,\dots,G\}.$$

Binary group labels:
$$\begin{cases}\mathbf{L}_{\text{tr}} = [\mathbf{l}_{\text{tr}}^{[1]},\dots,\mathbf{l}_{\text{tr}}^{[K_{\text{tr}}]}] \\ \mathbf{l}^{(\text{HC})} = [0,1]^T \text{ and } \mathbf{l}^{(\text{Sz})} = [1,0]^T\end{cases}$$

Jointly learn a linear classifier and a dictionary for tFNC sparse representation to improve discrimination

Problem formulation

Sparsity promoting function
- convex/non-convex
- Smooth/non-smooth

Sparse representation error

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \; \frac{1}{2}\overbrace{\|\mathbf{F}-\mathbf{DZ}\|_F^2}+\lambda\cdot r(\mathbf{Z})+\frac{\beta}{2}\|\mathbf{L}_{\text{tr}}-\mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2,$$

$$\text{s.t. } \mathbf{D}\in\mathcal{D}:=\{\mathbf{D}:\|\mathbf{d}_g\|_2=1, g=1,2,\ldots,G\}.$$

Binary group labels:

$$\begin{cases} \mathbf{L}_{\text{tr}} = [\mathbf{l}_{\text{tr}}^{[1]},\ldots,\mathbf{l}_{\text{tr}}^{[K_{\text{tr}}]}] \\ \mathbf{l}^{(\text{HC})} = [0,1]^T \text{ and } \mathbf{l}^{(\text{Sz})} = [1,0]^T \end{cases}$$

Jointly learn a linear classifier and a dictionary for tFNC sparse representation to improve discrimination

Problem formulation

Sparsity promoting function

- convex/non-convex
- Smooth/non-smooth

Sparse representation error

Linear classifier (only on training set)

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \frac{1}{2}\overbrace{\|\mathbf{F}-\mathbf{D}\mathbf{Z}\|_F^2}+\lambda\cdot r(\mathbf{Z})+\frac{\beta}{2}\|\mathbf{L}_{\text{tr}}-\mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2,$$

$$\text{s.t. } \mathbf{D}\in\mathcal{D}:=\{\mathbf{D}:\|\mathbf{d}_g\|_2=1, g=1,2,\ldots,G\}.$$

Binary group labels:
$$\begin{cases} \mathbf{L}_{\text{tr}} = [\mathbf{l}_{\text{tr}}^{[1]},\ldots,\mathbf{l}_{\text{tr}}^{[K_{\text{tr}}]}] \\ \mathbf{l}^{(\text{HC})} = [0,1]^T \text{ and } \mathbf{l}^{(\text{Sz})} = [1,0]^T \end{cases}$$

Iterative proximal-projection as a flexible approach to a range of sparsity-promoting functions including non-convex and non-smooth

$$\min_{\mathbf{D},\mathbf{Z},\mathbf{W}} \frac{1}{2}\|\mathbf{F}-\mathbf{D}\mathbf{Z}\|_F^2 + \lambda \cdot r(\mathbf{Z}) + \frac{\beta}{2}\|\mathbf{L}_{\text{tr}} - \mathbf{W}\mathbf{Z}_{\text{tr}}\|_F^2,$$
$$\text{s.t. } \mathbf{D} \in \mathcal{D} := \{\mathbf{D} : \|\mathbf{d}_g\|_2 = 1, g = 1, 2, \ldots, G\}.$$

1. Perform alternating minimization over **D**, **Z**, and **W**
2. Using iterative proximal-projection approach
    - Flexible to a range of sparsity-promoting functions, including non-convex and non-smooth scenarios

Experimental setup

Data preparation

- Bipolar-schizophrenia network on intermediate phenotypes resting-state fMRI dataset (five sites)
- 179 HC and 179 Sz patients
- To obtain subject-specific tFNC-feature vectors $\mathbf{f}^{[k]}$:
- Group ICA-EBM with order $N = 55$
- Selecting $N = 32$ functionally relevant components

DL setup

- $\mathbf{D}$ and $\mathbf{W}$ are initialized with DCT dictionary
- $\mathbf{Z}$ initialized with a null matrix
- Complete dictionary
- Sparsity level 50%
- Two scenarios:
  - $\beta = 0$: DL without learning a linear classifier
  - $\beta = 0.05$: linear classifier jointly learned with $\mathbf{D}$

Jointly learned sparse features improve different classification metrics

- Training SVM classifiers with polynomial kernels of order 3 using tFNC-features and sparse-features
- Repeating the experiment 100 times and reporting the average results

Table 1: Average classification rates [%].

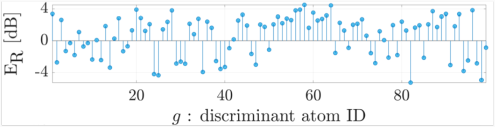
Metric\Feature	tFNC	Sparse ($\beta = 0$)	Sparse ($\beta = 0.05$)
Recall	74.75 ± 0.61	73.56 ± 0.65	**75.19** ± 0.65
Specificity	73.78 ± 0.70	74.14 ± 0.70	**74.47** ± 0.68
Precision	74.35 ± 0.50	74.27 ± 0.53	**74.93** ± 0.51
Accuracy	74.26 ± 0.40	73.85 ± 0.45	**74.83** ± 0.43
F1-score	74.35 ± 0.40	73.72 ± 0.46	**74.87** ± 0.45

- Sparse features outperform tFNC features

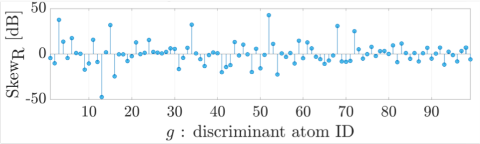
Sparse coefficients give statistics on each atom's contribution

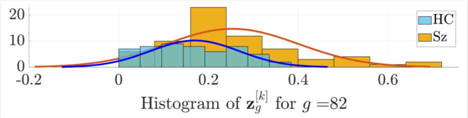
Two-sample t-test on sparse coefficients: 99 atoms discriminate HC and Sz groups

$$\mathrm{E_R} = 10\log(\|\mathbf{z}_g^{[k_{\mathrm{HC}}]}\|_2^2 / \|\mathbf{z}_g^{[k_{\mathrm{Sz}}]}\|_2^2)$$



$$\mathrm{Skew_R} = 10\log(|\mathrm{skew}(\mathbf{z}_g^{[k_{\mathrm{HC}}]})| / |\mathrm{skew}(\mathbf{z}_g^{[k_{\mathrm{Sz}}]})|)$$

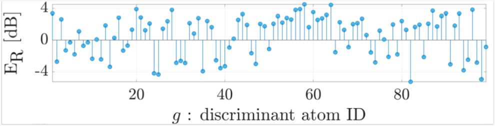




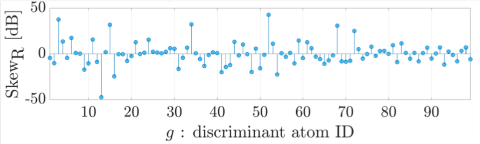
Sparse coefficients give statistics on each atom's contribution

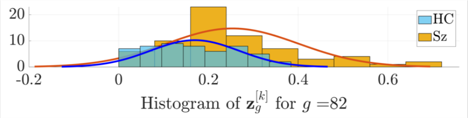
Two-sample t-test on sparse coefficients: 99 atoms discriminate HC and Sz groups

$$\mathrm{E_R} = 10\log(||\mathbf{z}_g^{[k_{\mathrm{HC}}]}||_2^2/||\mathbf{z}_g^{[k_{\mathrm{Sz}}]}||_2^2)$$



$$\mathrm{Skew_R} = 10\log(|\mathrm{skew}(\mathbf{z}_g^{[k_{\mathrm{HC}}]})|/|\mathrm{skew}(\mathbf{z}_g^{[k_{\mathrm{Sz}}]})|)$$

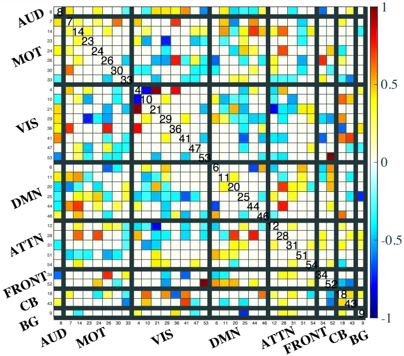




- Atoms with energy ratios $E_R > 0$ are dominant in HC, and atoms with $E_R < 0$ are dominant in Sz
- Skewness differs in one group compared with the other

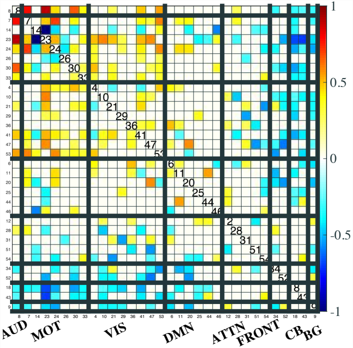
Discriminant atoms are interpretable

- Reshaped discriminant atoms reveal different network interaction patterns in HC and Sz.



$\sum_g \bar{\mathbf{z}}_g^{[Sz]}$ for g's dominant in Sz

- Less structured
- With extreme values



$\sum_g \bar{\mathbf{z}}_g^{[Sz]}$ for g's dominant in HC

- More modularity
- More anatomical organization

Conclusion and perspectives

- Sparse representation of brain temporal functional network connectivity (tFNC) is presented
- A dictionary and linear classifier were jointly learned to classify HC and Sz subjects using sparse coefficients
- Sparse features improved classification and identified new discriminative patterns in brain network interaction
- The approach offers new perspectives to study fMRI dynamics and can be extended to multiple fMRI datasets
- A non-linear classifier learned with the dictionary can improve classification rates

Conclusion and perspectives

- Sparse representation of brain temporal functional network connectivity (tFNC) is presented
- A dictionary and linear classifier were jointly learned to classify HC and Sz subjects using sparse coefficients
- Sparse features improved classification and identified new discriminative patterns in brain network interaction
- The approach offers new perspectives to study fMRI dynamics and can be extended to multiple fMRI datasets
- A non-linear classifier learned with the dictionary can improve classification rates

---

Thank you!
