Independent Vector Analysis Based Subgroup Identification from Multisubject fMRI data

Hanlu Yang[1], M. A. B. S Akhonda[1], Fateme Ghayem[1], Qunfang Long[1], Vince D. Calhoun[2], Tülay Adali[1]

[1]Dept. of CSEE, UMBC, Baltimore, MD 21250
[2]TReNDS, Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA 30303

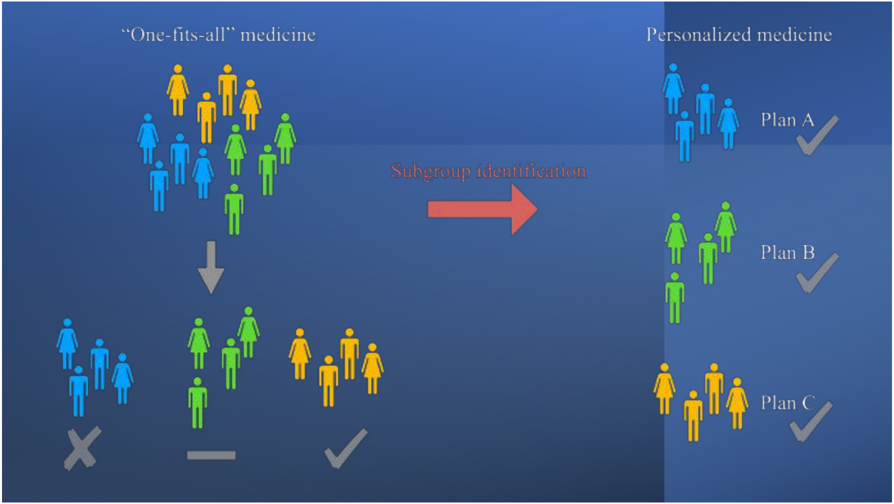
ICASSP 2022, Singapore

May 22, 2022

UMBC

MLSP
Machine Learning for Signal Processing Lab

Subgroup identification is the key for personalized medicine



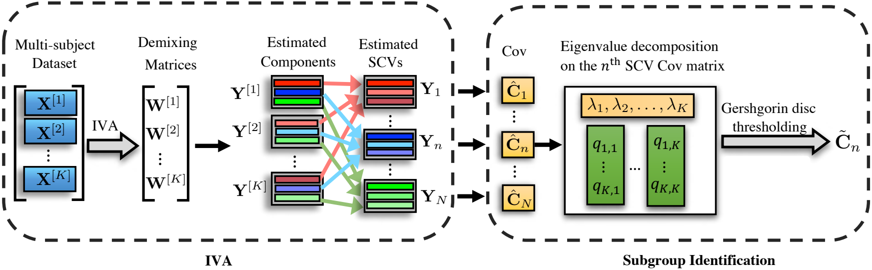
Subgroup identification is the key for personalized medicine

- Individual variability in brain functional networks are fingerprints of identifying subjects [*Finn et al., 2015*]
- Subgroup shows homogeneity

Current methods are limited by multiple factors

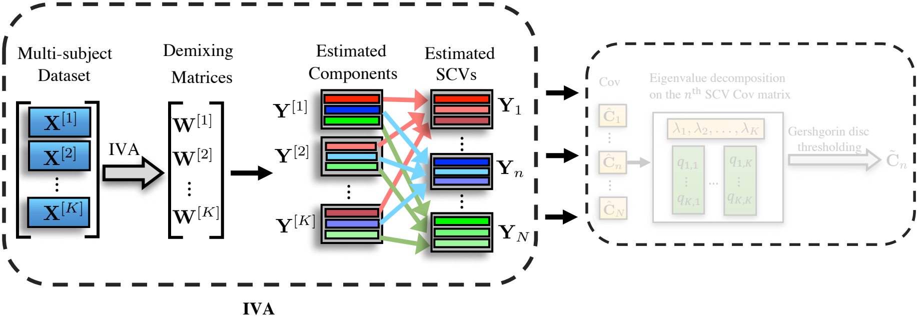
- Identify subgroups with behavioral variables, clinical, cognitive or other related scores [*Bitsika et al., 2008*], [*Veatch et al., 2014*]
- Apply independent component analysis (ICA) on individual subject fMRI data – missing multivariate information across subjects [*Durieux et al., 2019*]
- Apply independent vector analysis (IVA) to multisubject fMRI data – relying on user-defined parameters [*Long et al., 2020*]

Proposed method: subgroup identification using independent vector analysis (SI-IVA)



- SI-IVA:
  - ✓ captures subgroup structures
  - ✓ estimates the number of subgroups in each SCV
  - ✓ identifies the corresponding subjects in each subgroup

SI-IVA estimates correlated components across subjects



- Generative model: $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}$
- Demixing matrices $\mathbf{W}^{[k]}$, and estimated signals $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$
- source component vectors (SCV), $\mathbf{Y}_n = [\mathbf{y}_n^{[1]}, \mathbf{y}_n^{[2]}, \ldots, \mathbf{y}_n^{[K]}]^\top \in \mathbb{R}^{K\times V}$

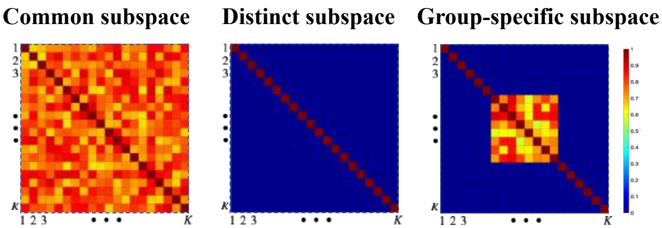
[Kim *et al.*, 2006], [Anderson *et al.*, 2012]

Individual SCV preserves the correlation structure of a component across subjects

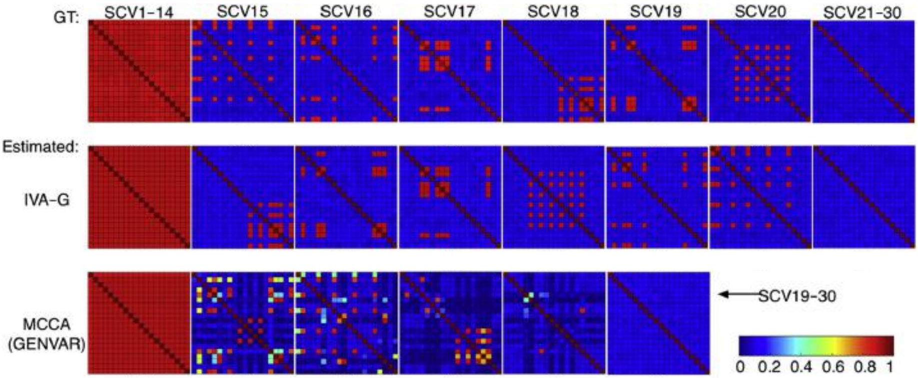
- Sample covariance matrix of n^{th} SCV: $\hat{\mathbf{C}}_n = \begin{bmatrix} 1 & \rho_n^{[1,2]} & \cdots & \rho_n^{[1,K]} \\ \rho_n^{[2,1]} & 1 & \cdots & \rho_n^{[2,K]} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_n^{[K,1]} & \rho_n^{[K,2]} & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{K \times K}$

  $\rho_n^{[k_1,k_2]}$ is the correlation between k_1^{th} and k_2^{th} subjects of n^{th} component

- SCVs construct three subspaces:

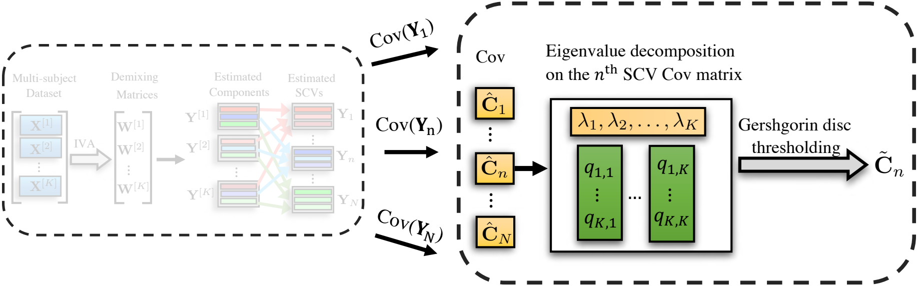


IVA-G preserves correlation structures effectively



[Long *et al.*, 2020], [Anderson *et al.*, 2012]

SI-IVA identifies subgroup structures

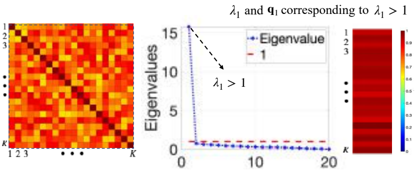


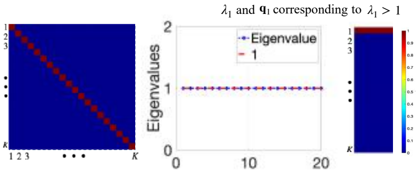
Subgroup Identification

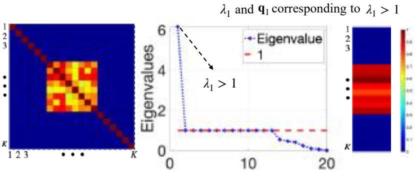
- Decompose each $\hat{\mathbf{C}}_n = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}$
- Estimates the number of subgroups from SCVs, $\mathbf{Y}_1 \ldots \mathbf{Y}_N$
- Identify subjects that are correlated within a SCV, $\rho_n^{[k_1,k_2]} \neq 0$

Eigenvalues and eigenvectors of $\hat{c}_n$ reveal subgroups

- $\#(\lambda > 1) \rightarrow \#$ subgroups
- $q_{k,1} \neq 0 \in \mathbf{q}_1 \rightarrow$ subject index





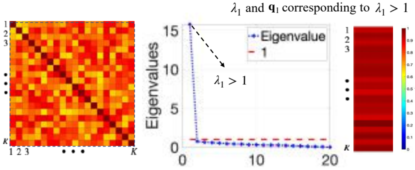


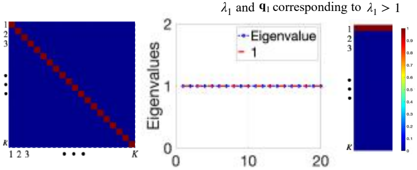
The variation in eigenvalues can lead to overestimation of subgroups

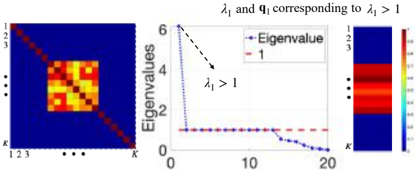
[Hasija *et al.*, 2020], [Akhonda *et al.*, 2021]

Eigenvalues and eigenvectors of $\hat{c}_n$ reveal subgroups

- $\#(\lambda > 1) \rightarrow$ # subgroups
- $q_{k,1} \neq 0 \in \mathbf{q}_1 \rightarrow$ subject index



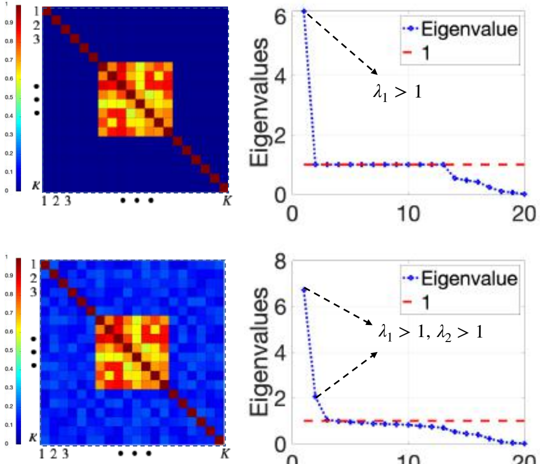




The variation in eigenvalues can lead to overestimation of subgroups

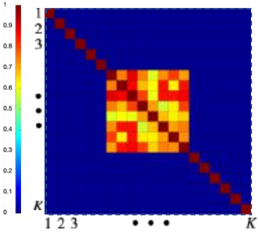
[Hasija *et al.*, 2020], [Akhonda *et al.*, 2021]

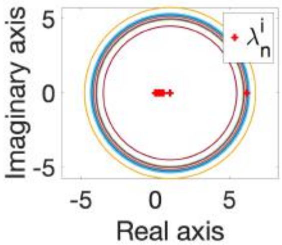
The variation in eigenvalues leads to overestimation of subgroups



Gershgorin disc transforms eigenvalues' variation

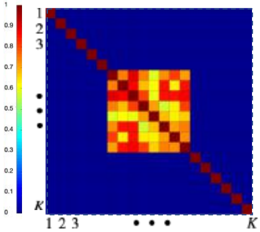
- Gershgorin disc: $\left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$, where $R_i = \sum_{j \neq i} \left|\rho_n^{[i,j]}\right|$
- $\text{eig}(\hat{\mathbf{C}}_n) \in \bigcup_{i=1}^{K} \left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$
- Gershgorin discs of $\hat{\mathbf{C}}_n$ are located at (1,0)

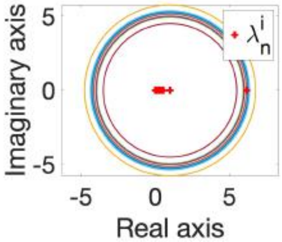




Gershgorin disc transforms eigenvalues' variation

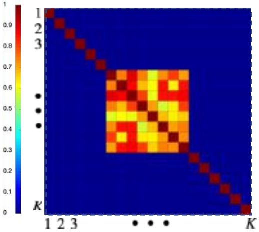
- Gershgorin disc: $\left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$, where $R_i = \sum_{j \neq i} \left|\rho_n^{[i,j]}\right|$
- $\text{eig}(\hat{\mathbf{C}}_n) \in \bigcup_{i=1}^{K} \left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$
- Gershgorin discs of $\hat{\mathbf{C}}_n$ are located at (1,0)

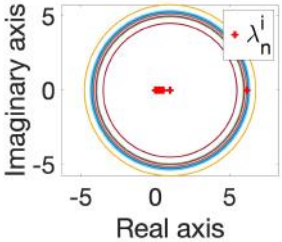




Gershgorin disc transforms eigenvalues' variation

- Gershgorin disc: $\left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$, where $R_i = \sum_{j \neq i} \left|\rho_n^{[i,j]}\right|$
- $\text{eig}(\hat{\mathbf{C}}_n) \in \bigcup_{i=1}^{K} \left\{z \in \mathbb{R} : \left|z - \rho_n^{[i,i]}\right| \leq R_i\right\}$
- Gershgorin discs of $\hat{\mathbf{C}}_n$ are located at (1,0)

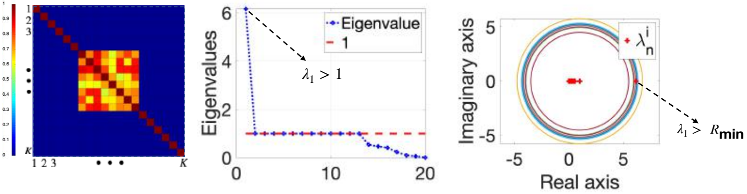




Gershgorin disc solves the overestimating issue

- Instead of looking for eigenvalues $\lambda > 1$, SI-IVA looks for $\lambda > (R_{\min} + 1)$

- Eigenvalue decomposition with Gershgorin disc (EGD) incorporates eigenvalue decomposition with hard thresholding (EHT)

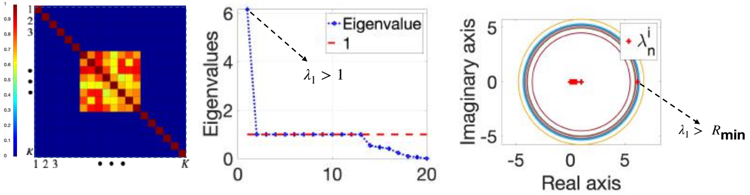


- EGD solves the overestimating issue that is caused by hard thresholding

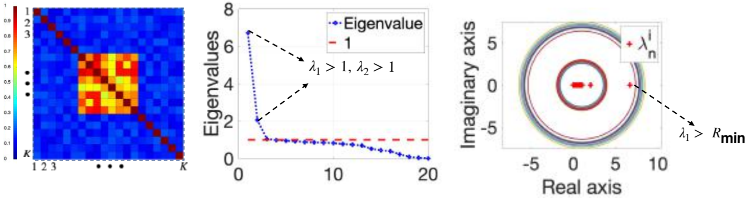
Gershgorin disc solves the overestimating issue

- Instead of looking for eigenvalues $\lambda > 1$, SI-IVA looks for $\lambda > (R_{\min} + 1)$

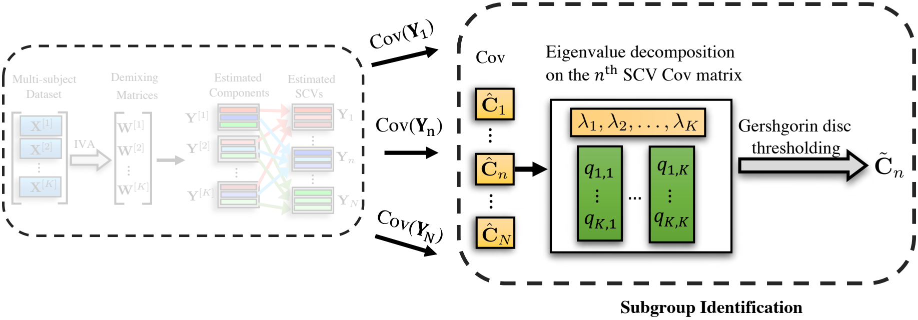
- Eigenvalue decomposition with Gershgorin disc (EGD) incorporates eigenvalue decomposition with hard thresholding (EHT)



- EGD solves the overestimating issue that is caused by hard thresholding

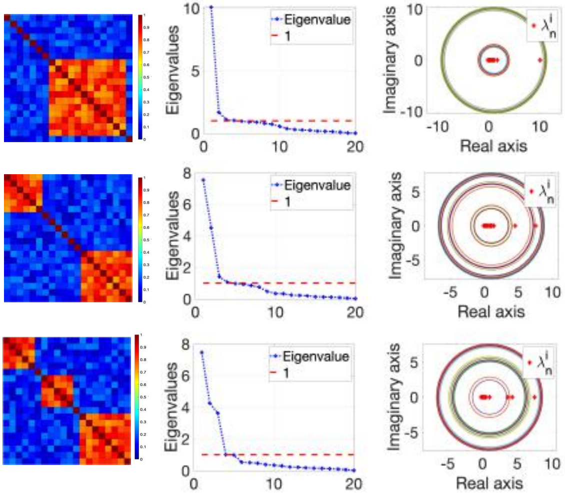


SI-IVA: Identification of subgroup structure



- # $(\lambda_n^1 \ldots \lambda_n^M) > (R_{min} + 1) \rightarrow$ # subgroups
- $\mathbf{q}_n^1 \ldots \mathbf{q}_n^M \rightarrow$ subject index

Simulation: EGD predicts correct number of subgroups

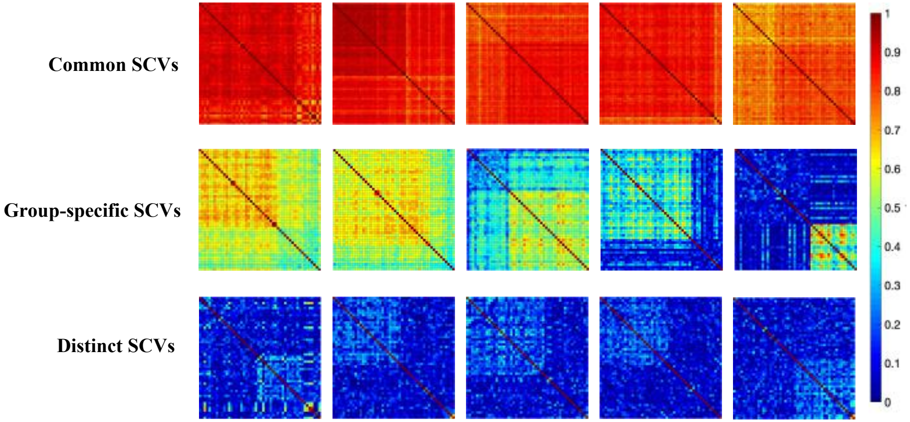


- $\rho_c \sim \mathcal{U}[0.7, 0.9]$, $\rho_d \sim \mathcal{U}[0.05, 0.25]$
- SCVs: MGGD, $\beta \sim \mathcal{U}[0.1, 0.8]$

SI-IVA identifies subgroup structures with multisubject resting-state fMRI data

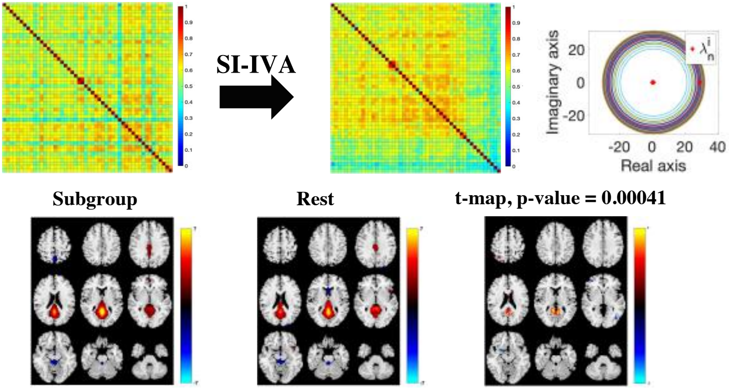
- 50 schizophrenia patients' resting-state fMRI data were collected from the Center of Biomedical Research Excellence (COBRE) (https://coins.trendscenter.org)
- IVA-G was implemented with order number as 85
- Cross intersymbol interference (Cross-ISI) was used to select the most consistent run for 10 runs with random initialization [Long and Jia et al., 2018]

SI-IVA identifies subgroup structures with multisubject resting-state fMRI data



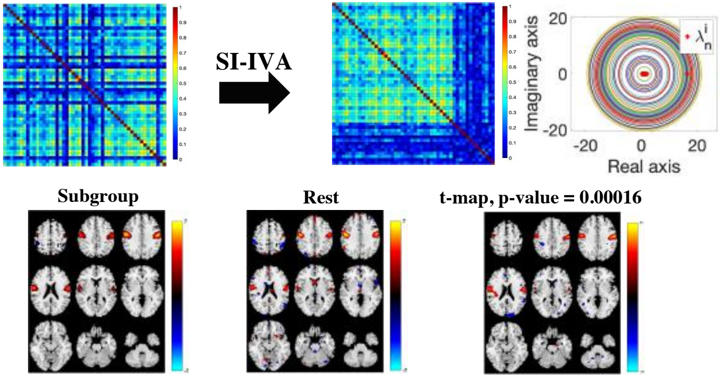
SI-IVA identifies subgroup structures with multisubject resting-state fMRI data

- Mean components are thresholded at $Z = 2$
- Two-sample t-test used to identify the discriminative components ($p < 0.05$)



SI-IVA identifies subgroup structures with multisubject resting-state fMRI data

- Mean components are thresholded at $Z = 2$
- Two-sample t-test used to identify the discriminative components ($p < 0.05$)



Summary

SI-IVA

- identifies subgroup structures
- yields meaningful components and subgroups
- requires no user-defined parameters
- can be adopted to other multi-set data

Acknowledgements

- This work was supported in part by NSF grants CCF 1618551, NCS 1631838 and HRD 2112455, and NIH grants R01MH123610 and R01MH118695

http://mlsp.umbc.edu