Generalized Liquid Association Analysis for Multimodal Neuroimaging

Jing Zeng

Department of Statistics
Florida State University

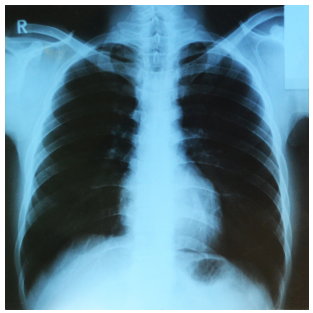
International Conference on Econometrics and Statistics
June 26, 2021
Joint work with **Dr. Lexin Li** and **Dr. Xin Zhang**
Preprint: https://arxiv.org/abs/2008.03733

Motivating application

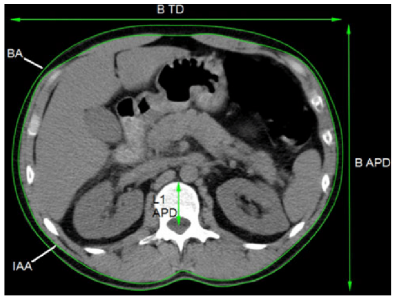
Multimodal data analysis

Backgound:

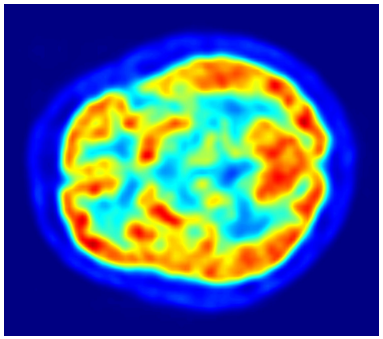
- **Imaging modality**: X-ray, Computed Tomography (CT), positron emission tomography (PET), Magnetic resonance imaging (MRI).



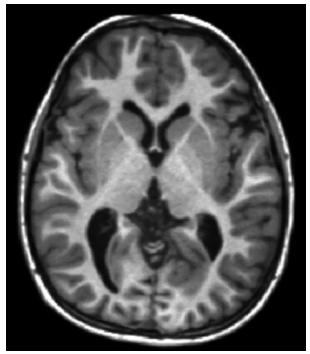
(a) X-ray



(b) CT



(c) PET



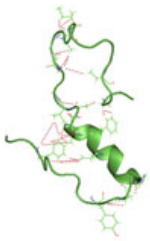
(d) MRI

- **Multimodal data analysis**: Each subject has more than one imaging modality.

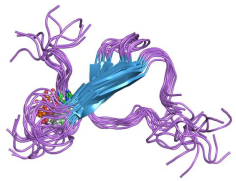
Multimodal data analysis

Backgound:

- Targets on **Alzheimer's disease (AD)** and normal **aging**.
- **Amyloid-beta** and **tau** are two hallmark proteins of AD. The **spatial patterns** of accumulations of amyloid-beta and tau are closely associated, and such association patterns are highly affected by the **subject's age**.



(a) Amyloid-beta ($\boldsymbol{X} \in \mathbb{R}^{60}$)



(b) Tau ($\boldsymbol{Y} \in \mathbb{R}^{26}$)

(c) Age ($Z \in \mathbb{R}$: average 77.5 SD 6.2)

- **Goal**: Find **how** and **where** in the brain the associations of the two proteins **change the most** as age varies. (dynamic association)
- **This talk:** Statistical framework for studying the three-way association of $\boldsymbol{X} \in \mathbb{R}^{p_1}$ and $\boldsymbol{Y} \in \mathbb{R}^{p_2}$ **given** $\boldsymbol{Z} \in \mathbb{R}^{p_3}$.

Methodology

Univariate Liquid Association (K.C. Li 2002)

- Suppose that univariates $X, Y, Z \in \mathbb{R}$ have mean zero and variance one. The association between X and Y given Z is measured by the function

$$g(z) = \mathrm{E}(XY \mid Z = z).$$

- To capture the changes in $g(z)$,

$$\mathrm{LA}(X, Y \mid Z) = \mathrm{E}\{\frac{d}{dZ} g(Z)\} \in \mathbb{R}.$$

- When Z is normal, the estimation is simple, thanks to Stein's Lemma,

$$\mathrm{E}\{\frac{d}{dZ} g(Z)\} = \mathrm{E}\{Z g(Z)\} = \mathrm{E}(XYZ).$$

- Particularly useful in discovering co-expressed gene pairs that are regulated by another gene.

Dimension reduction model with sparsity

- **Dimension reduction model**: For multivariates $\boldsymbol{X} \in \mathbb{R}^{p_1}$, $\boldsymbol{Y} \in \mathbb{R}^{p_2}$, $\boldsymbol{Z} \in \mathbb{R}^{p_3}$, we seek the linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that **change the most** as (the linear combination of) $\boldsymbol{Z}$ varies. Assume that

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top \mid \boldsymbol{Z} = \boldsymbol{z}) = \boldsymbol{\Gamma}_1 \boldsymbol{f}(\boldsymbol{\Gamma}_3^\top \boldsymbol{z}) \boldsymbol{\Gamma}_2^\top,$$

  where $\boldsymbol{\Gamma}_k \in \mathbb{R}^{p_k \times r_k}$, $r_k < p_k$, are semi-orthogonal basis matrices and $\boldsymbol{f} : \mathbb{R}^{r_3} \to \mathbb{R}^{r_1 \times r_2}$ is unknown latent function.
- Reduce $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ to $\boldsymbol{\Gamma}_1^\top \boldsymbol{X}, \boldsymbol{\Gamma}_2^\top \boldsymbol{Y}, \boldsymbol{\Gamma}_3^\top \boldsymbol{Z}$ without loss of information.
- **Sparsity**: Assume that each $\boldsymbol{\Gamma}_k$ has s_k non-zero rows. The row-wise sparsity indicates that only some entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dynamically associated.

Generalized Liquid Association

- **Generalized liquid association (GLA)**:

$$\boldsymbol{\Phi} = \mathrm{GLA}(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}) = \mathrm{E}\left\{ \frac{d}{d\boldsymbol{Z}} \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top \mid \boldsymbol{Z}) \right\} \in \mathbb{R}^{p_1 \times p_2 \times p_3}.$$

- With the dimension reduction model assumption,

$$\boldsymbol{\Phi} = \boldsymbol{\Phi} \times_1 \mathbf{P}_{\boldsymbol{\Gamma}_1} \times_2 \mathbf{P}_{\boldsymbol{\Gamma}_2} \times_3 \mathbf{P}_{\boldsymbol{\Gamma}_3},$$

  where $\mathbf{P}_{\boldsymbol{\Gamma}_k} = \boldsymbol{\Gamma}_k \boldsymbol{\Gamma}_k^\top$ is the projection matrix onto the column subspace of $\boldsymbol{\Gamma}_k$.

- If $\mathbf{Z}$ is normal, then $\boldsymbol{\Phi} = \boldsymbol{\Delta} \times_3 \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}$, where

$$\boldsymbol{\Delta} = \mathrm{E}(\mathbf{X} \circ \mathbf{Y} \circ \mathbf{Z}) \in \mathbb{R}^{p_1 \times p_2 \times p_3}. \tag{1}$$

- Estimate $\boldsymbol{\Gamma}_k$'s by (sparse) tensor decomposition of $\boldsymbol{\Delta}$: (1) avoid the estimation of $\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}$; (2) avoid the normality assumption.

Tucker decomposition

- **Tucker decomposition (for three-way tensor)**:
  $\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$, where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, $\mathbf{A}_k \in \mathbb{R}^{p_k \times r_k}$.

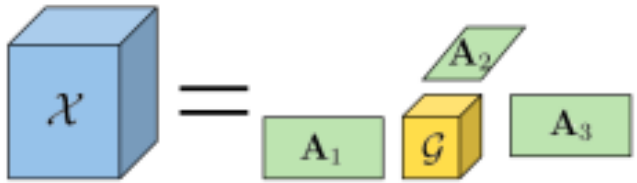


Figure 1: Tucker decomposition of $\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$.

- **Our optimization problem**:

$$(\widehat{\mathbf{\Gamma}}_1, \widehat{\mathbf{\Gamma}}_2, \widehat{\mathbf{\Gamma}}_3) = \underset{\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3}{\operatorname{argmin}} \|\widetilde{\mathbf{\Delta}} - \widetilde{\mathbf{\Delta}} \times_1 \mathbf{P}_{\mathbf{G}_1} \times_2 \mathbf{P}_{\mathbf{G}_2} \times_3 \mathbf{P}_{\mathbf{G}_3}\|_F^2,$$

  where $\widetilde{\mathbf{\Delta}} = n^{-1} \sum_{i=1}^n \mathbf{X}_i \circ \mathbf{Y}_i \circ \mathbf{Z}_i$ is the sample estimator.
- Challenges in theory and algorithm: non-convex and high-dimensional.

Algorithm: Sparse HOSVD

1. **Input**: The Tucker ranks $r_k \leq p_k$, and the sparsity parameters $(\eta_k, \widetilde{\eta}_k)$, $k = 1, 2, 3$.
2. **Initialization**:
   - **2.1** Compute the sample estimate $\widetilde{\mathbf{\Delta}} = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \circ \mathbf{Y}_i \circ \mathbf{Z}_i$. Obtain the initial active set: $\widehat{I}_k^{(0)} = \{j : \|(\widetilde{\mathbf{\Delta}}_{(k)})_{[j,:]}\|_{\max} > \eta_k\}$.
   - **2.2** Let $\widetilde{\mathbf{\Delta}}^{(0)} = \widetilde{\mathbf{\Delta}} \times_1 \mathbf{D}_{\widehat{I}_1^{(0)}} \times_2 \mathbf{D}_{\widehat{I}_2^{(0)}} \times_3 \mathbf{D}_{\widehat{I}_3^{(0)}}$, compute the initial basis matrices by $\widehat{\mathbf{\Gamma}}_k^{(0)} = \mathrm{SVD}\{\widetilde{\mathbf{\Delta}}_{(k)}^{(0)}\}$, $k = 1, 2, 3$.
3. **Repeat until the stopping criterion is met.**
   When $k = 1$, let $\mathbf{W} = \widetilde{\mathbf{\Delta}} \times_2 (\widehat{\mathbf{\Gamma}}_2^{(t-1)})^\top \times_3 (\widehat{\mathbf{\Gamma}}_3^{(t-1)})^\top$.
   - **3.1** Update the active set: $\widehat{I}_1^{(t)} = \{j : \|\mathbf{W}_{(1)}\|_2^2 > \widetilde{\eta}_1\}$.
   - **3.2** Perform SVD: $\widehat{\mathbf{\Gamma}}_1^{(t)} = \mathrm{SVD}\{\mathbf{D}_{\widehat{I}_1^{(t)}} \mathbf{W}_{(1)}\} \in \mathbb{R}^{p_k \times r_k}$.

   (The updates are similar when $k = 2, 3$.)
4. **Output**: The estimated basis matrices $\widehat{\mathbf{\Gamma}}_k$, $k = 1, 2, 3$, and $\widehat{\mathbf{\Delta}} = \widetilde{\mathbf{\Delta}} \times_1 \mathbf{P}_{\widehat{\mathbf{\Gamma}}_1} \times_2 \mathbf{P}_{\widehat{\mathbf{\Gamma}}_2} \times_3 \mathbf{P}_{\widehat{\mathbf{\Gamma}}_3}$.

Consistency results

Theorem 1

Under mild assumptions, let $s = s_1 s_2 s_3$ *and* $p = p_1 p_2 p_3$. *Assume that* $\sqrt{s \log p/n} = o(1)$, *with probability tending to one,*

1. $\|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_F \to 0$;
2. $\max_{k=1,2,3} \|\mathbf{P}_{\widehat{\boldsymbol{\Gamma}}_k} - \mathbf{P}_{\boldsymbol{\Gamma}_k}\|_F \to 0$;
3. $\widehat{I}_k^{(t)} = I_k$, $k = 1, 2, 3$, *and* $t = 0, 1, \ldots, t_{\max}$.

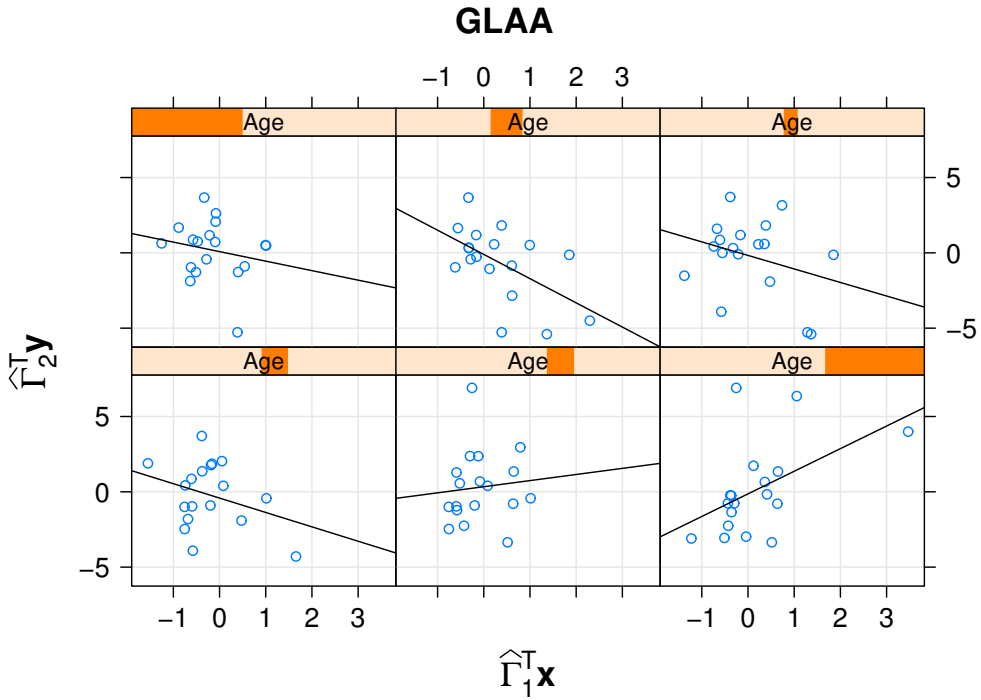
- **Remark**: In ultra-high dimensional setting, our method achieves consistency in **variable selection** and in the estimation of **GLA tensor** and the **dimension reduction subspaces**.

Multimodal PET analysis

Multimodal PET analysis

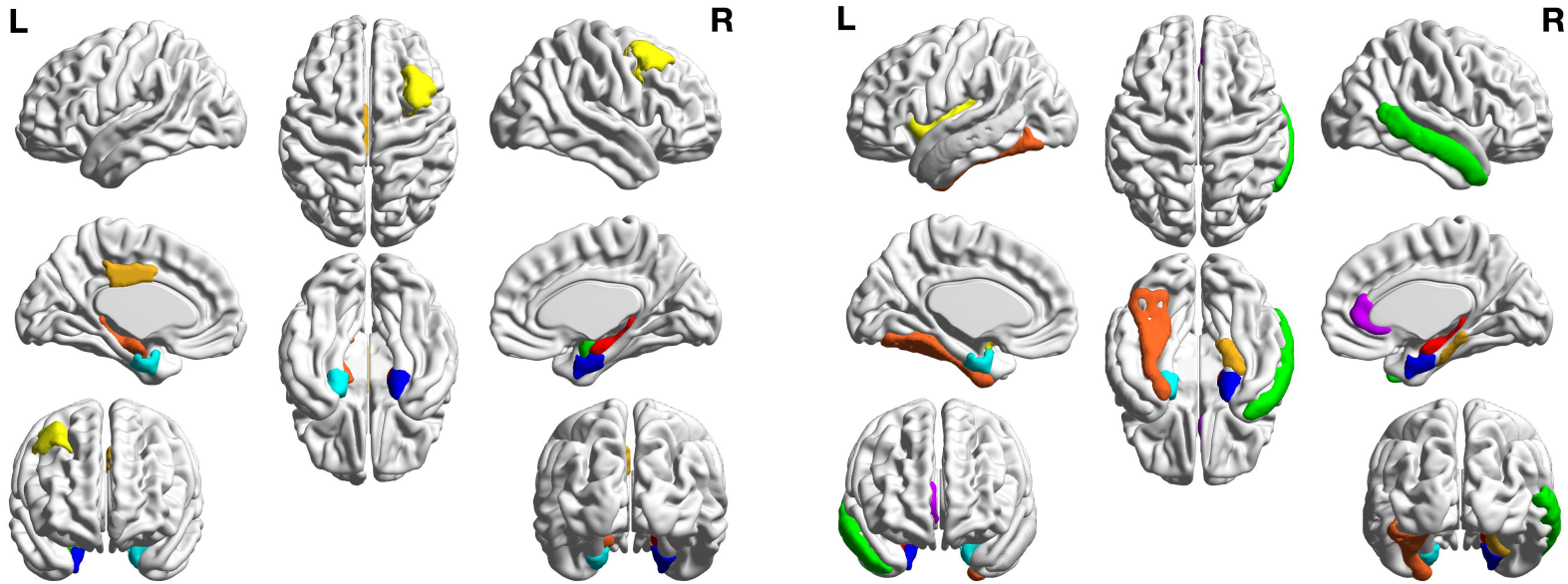
- **Data description**
  - Part of the Berkeley Aging Cohort Study (ongoing project).
  - $n = 81$: sample size.
  - $\boldsymbol{X} \in \mathbb{R}^{60}$: the amount of **amyloid-beta** at $p_1 = 60$ brain ROIs.
  - $\boldsymbol{Y} \in \mathbb{R}^{26}$: the amount of **tau** at $p_2 = 26$ brain ROIs.
  - $Z \in \mathbb{R}$: age; average 77.5 with SD 6.2.
- We use Tucker ranks $r_1 = r_2 = 1$ to identify the most age-dependent linear combinations in multimodal PET association.

Dynamic association plots



- GLAA provides a useful dimension reduction tool to help visualize the dynamic patterns.
- The association changes from negative to positive in later years.
- The spread of tau out of medial temporal lobes is accelerated by the presence of amyloid-beta at elder age.

Selected variables (regions)



(a) amyloid-beta (b) tau

Figure 2: Identified brain regions for amyloid-beta and tau by GLAA.

Findings

Modality	Identified regions				
amyloid-beta	Entorhinal R Orbitofrontal L	Entorhinal L Posterior Cingulate L	Hippocampus R Middle Frontal R	Hippocampus L	Amygdala R
tau	Entorhinal R Middle Temporal R	Entorhinal L Middle Temporal L	Hippocampus R Insula L	Parahippocampal R Rostral Anterior Cingulate R	Fusiform L

Table 1: Regions in the left hemisphere are denoted by "L", and regions in the right hemisphere are denoted by "R".

- Many of these regions are known to be closely related to AD.
- For both amyloid-beta and tau, the identified regions include **hippocampus** and **entorhinal cortex**.
- **Hippocampus** is one of the first brain regions to suffer damage from AD; and hippocampus atrophy is a well-known biomarker for AD.
- **Entorhinal cortex**, together with hippocampus, plays an important role in memories, and the atrophy in the entorhinal cortex is consistently reported in AD.

Conclusion

- **Scientifically**, GLAA offers a unique angle for understanding the age-dependent patterns between amyloid-beta and tau in AD and normal aging.
- **Statistically**, we propose a new framework in the association analysis among three sets of variables. Specifically, our method has
  - a population dimension reduction model
  - a computationally scalable algorithm
  - solid theoretical properties in high dimensions
- **Future research directions**: handling discrete/categorical variables; extensions to non-linear relationships; etc.

Thank you!