Tensor partition regression models with applications in imaging biomarker detection
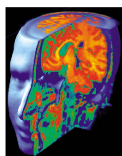
Michelle F. Miranda
joint work with Hongtu Zhu and Joseph G. Ibrahim

May 6, 2015

GENERAL GOAL

The development of new statistical tools that allow us to look at the whole brain and to establish associations between what do we see inside and the external world.
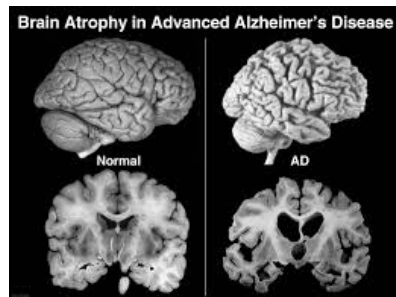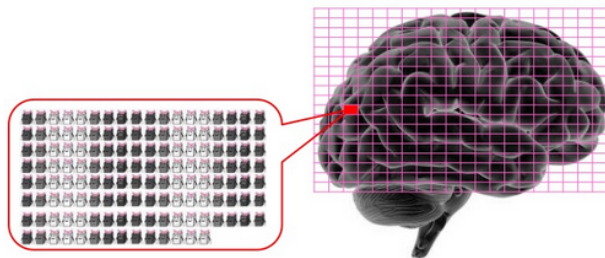
SPECIFIC GOAL

Identify imaging biomarkers that are relevant to predict disease status.
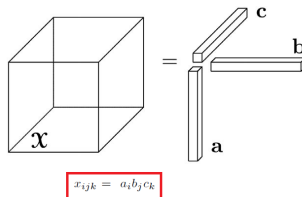

Brain Atrophy in Advanced Alzheimer's Disease

WHAT ARE THE CHALLENGES?

- Typical image size $256 \times 256 \times 256 \approx 17$ million voxels
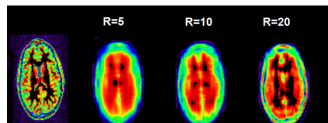- After downsizing + cropping $96 \times 96 \times 96 \approx 885.000$ voxels

UNDERSTANDING THE TENSOR DECOMPOSITION

- Tensor=multidimensional array
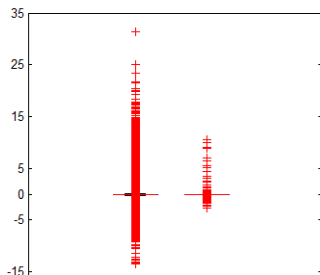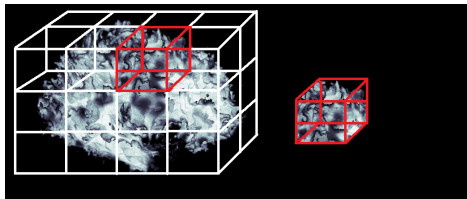- Order= dimension of the tensor
- Rank one tensors

$$x_{ijk} = a_i b_j c_k$$

$$x_{ijk} \approx \sum_{r=1}^{R} \lambda_r \, a_{ir} \, b_{jr} \, c_{kr} \qquad \mathcal{X} = [\![\boldsymbol{\lambda} \, ; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$
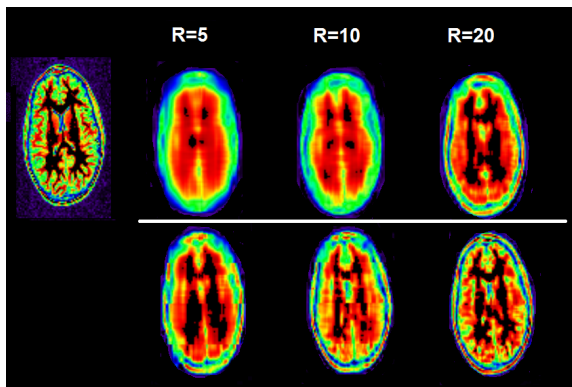
WHAT DO WE PROPOSE?

Strategy adopted: partitions!!

We increase the chances that small regions are captured!!

ARE THE PARTITIONS NECESSARY?

- ► T1-weighted $64 \times 108 \times 99$
- ► Top: no-partition
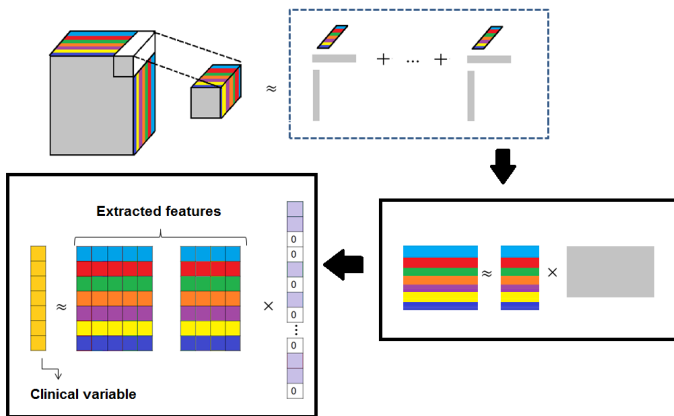- ► Bottom: 24 partitions of size $32 \times 27 \times 33$
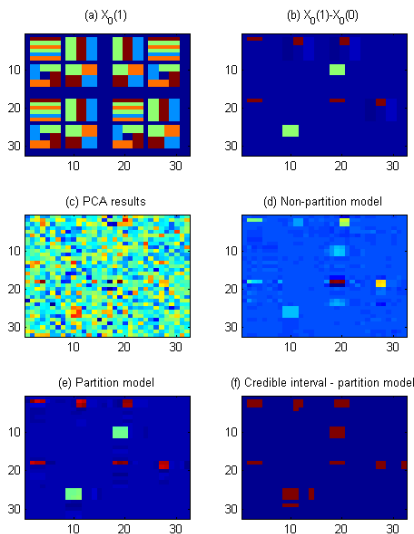
SOLUTION- HIERARCHICAL MODEL

1 Partition the images

2 Tensor decomposition

3 Factor Model

4 GLM on extracted features with sparse priors

DOES IT IMPROVE PREDICTIONS?

(a) $X_0(1)$

(b) $X_0(1)-X_0(0)$

(c) PCA results

(d) Non-partition model

(e) Partition model

(f) Credible interval - partition model

REAL DATA ANALYSIS- ADNI

Data: 402 MRI scans from ADNI1, 181 of them were diagnosed with AD, and 221 healthy controls.

RESULTS

- Based on a 95 % credible interval 36 basis are important to predict AD outcome
- Posterior mean of the projection $\mathcal{P} = \|\Lambda; A^{(1)}, A^{(2)}, A^{(3)}, \tilde{P}\|$
- 95% credible interval for $\mathcal{P}$ reveals the regions of the biomarkers selected to predict the AD outcome
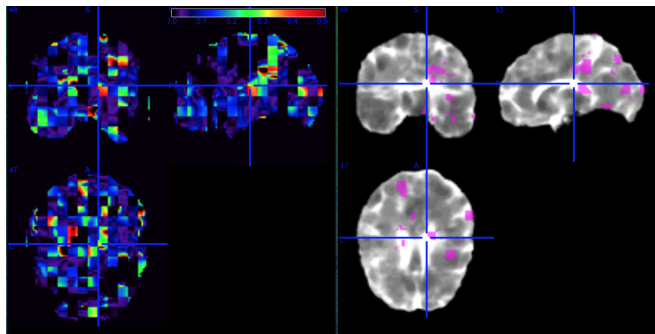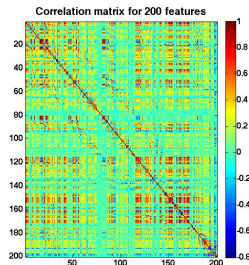
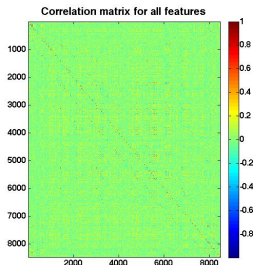# BIOMARKERS

- White matter: cingulum, fascicle, fornix
- Temporal lobe: hippocampus
- Parietal lobe: superior parietal lobe
- Frontal lobe: premotor cortex, primary motor cortex

Vemuri and Jack Jr. Role of structural MRI in Alzheimers disease. *Alzheimers Research Therapy* 2010, 2:23

WHY THE FACTOR MODEL IS NECESSARY?

- ▶ To reduce features
- ▶ To account for multicolinearity

PARTITION SELECTION

Partition size	$24 \times 24 \times 24$	$12 \times 12 \times 12$	$6 \times 6 \times 6$
$R = 5$	—	—	**0.7813**
$R = 10$	0.6714	0.7311	0.7613
$R = 20$	0.6937	0.7587	0.7588
$R = 30$	0.5770	0.6544	—
No partition	$1 \times 1 \times 1$		
$R = 100$	0.7498	—	—

Table: Mean prediction accuracy for a 10-fold cross-validation procedure. There is a smaller error measurement with an increase of the rank R.