Tensor Decomposition Approaches for fMRI Classification

Vida John, Katie Keegan, Tanvi Vishwanath, Yihua Xu

Faculty Advisor: Elizabeth Newman, PhD

July 2, 2021

Emory University REU/RET in Computational Mathematics and Data Science

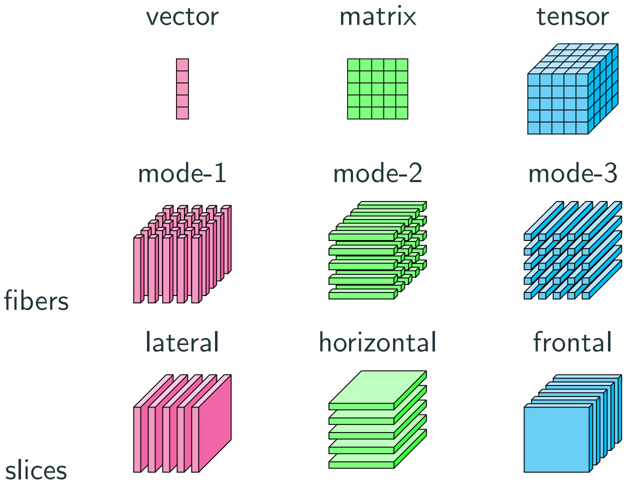
Motivation

Tensor

A multidimensional array of numbers, representing fMRI 3D brain images in response to stimuli over time.

- Exploit correlations of pixels in space and time,
- Use low rank approximations
- Better representation

Notation

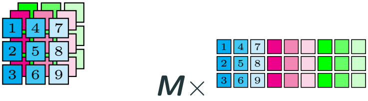

vector
matrix
tensor
mode-1
mode-2
mode-3
fibers
lateral
horizontal
frontal
slices

Mode-k Product ($\boldsymbol{\mathcal{A}} \times_k \boldsymbol{M}$):

The mode-k product of a tensor $\boldsymbol{\mathcal{A}}$ with a matrix $\boldsymbol{M}$ results in a tensor whose mode-k unfolding is $\boldsymbol{M}$ times the mode-k unfolding of $\boldsymbol{\mathcal{A}}$. In other words, $\boldsymbol{\mathcal{A}} \times_k \boldsymbol{M} = \text{fold}(\boldsymbol{M}\boldsymbol{\mathcal{A}}_{(k)})$

Facewise Product ($\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}$):

The facewise product multiplies each of the $n_1 \times n_2$ and $n_2 \times \ell$ frontal slices of two tensors in the transform domain in parallel to create a set of $n_1 \times \ell$ new slices.



Kolda and Bader, "Tensor Decompositions and Applications"

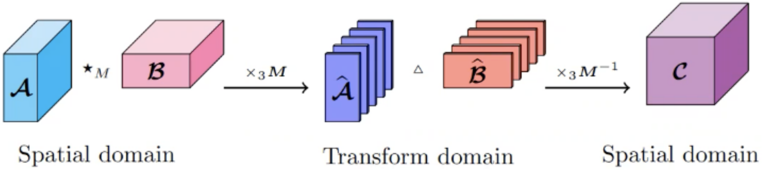
A Family of Tensor-Tensor Products

$\star_{\mathrm{M}}$-product:

Given the $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$, and the $n_2 \times \ell \times n_3$ tensor $\mathcal{B}$, with an invertible $n_3 \times n_3$ matrix $\boldsymbol{M}$:

$$\mathcal{C} = \mathcal{A} \star_{\mathrm{M}} \mathcal{B} = (\hat{\mathcal{A}} \triangle \hat{\mathcal{B}}) \times_3 \boldsymbol{M}^{-1}$$

such that $\mathcal{C}$ is an $n_1 \times \ell \times n_3$ tensor.



Kernfeld, Kilmer, and Aeron, "Tensor–tensor products with invertible linear transforms"

Our Research

Data

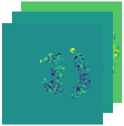
- StarPlus fMRI images of a subject's brain over time
- Multiple trials correspond to each subject
- Each trial corresponds to the subject either reading a sentence or seeing a picture

Questions

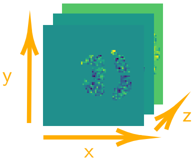
- Are tensor approaches better than matrix approaches to classifying fMRI data (i.e. picture or sentence)?
- If so, what tensor approach is best?

Just, *StarPlus fMRI data*

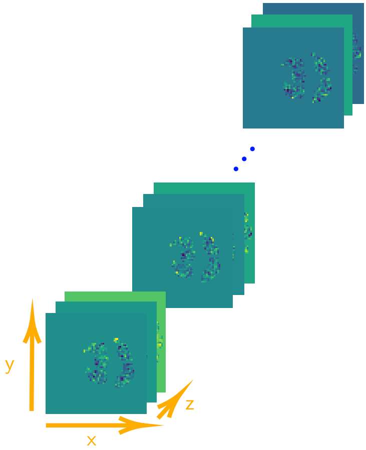
fMRI Data



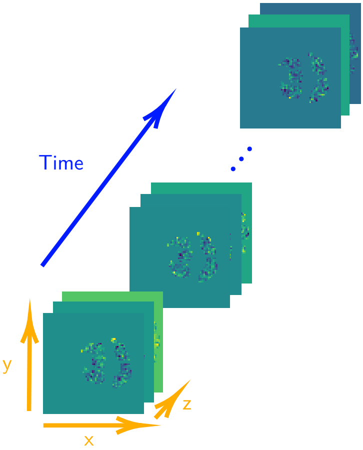
fMRI Data



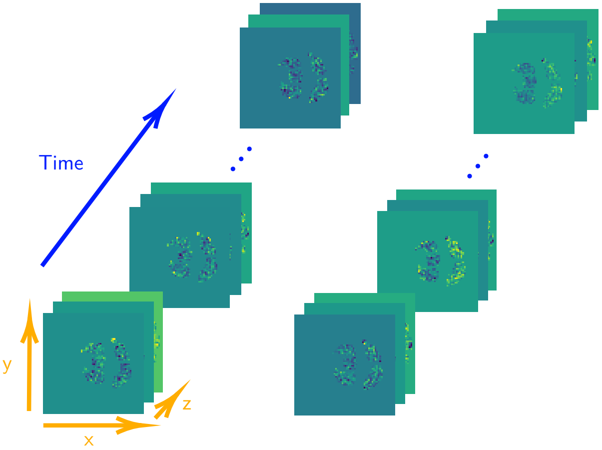
fMRI Data



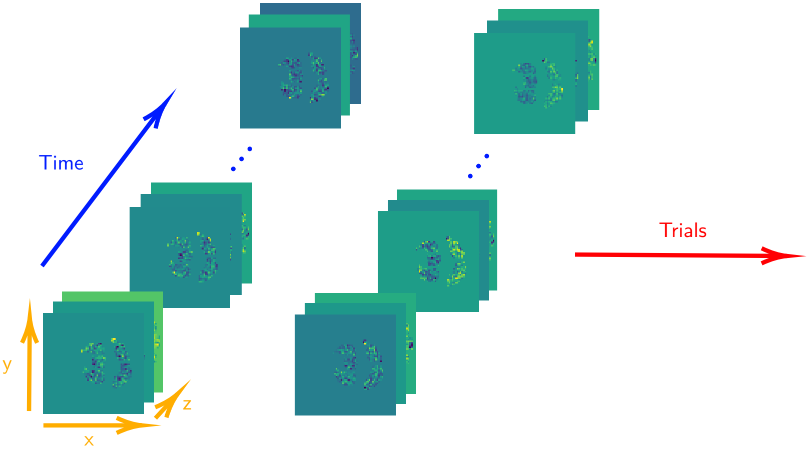
fMRI Data



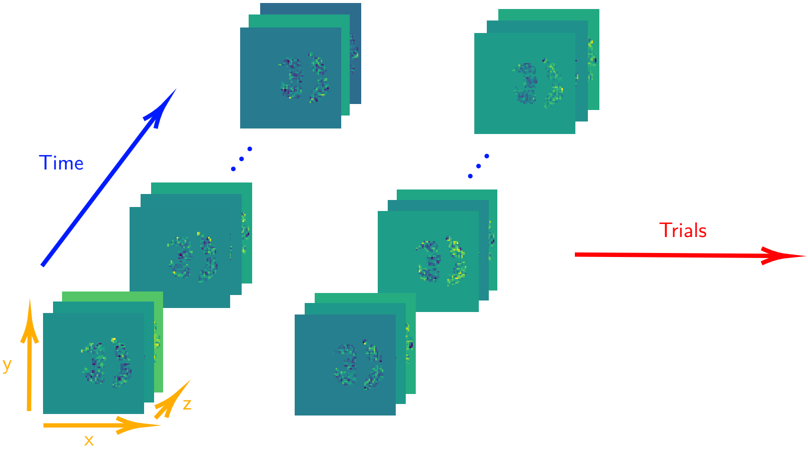
fMRI Data



fMRI Data



fMRI Data



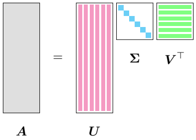
5-dimensional representation of data: (trials, x, y, z, time)

Review of Singular Value Decomposition

- SVD factorizes any matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

- Properties
  - Columns of $\boldsymbol{U}$ can be used as a basis for $\boldsymbol{A}$
  - $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal
- Highly useful in compression or extracting dominant features



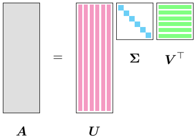
Strang, *Linear Algebra and Learning from Data*

Review of Singular Value Decomposition

- SVD factorizes any matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

- Properties
  - Columns of $\boldsymbol{U}$ can be used as a basis for $\boldsymbol{A}$
  - $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal
- Highly useful in compression or extracting dominant features



How can this be extended to higher dimensions (tensors)?

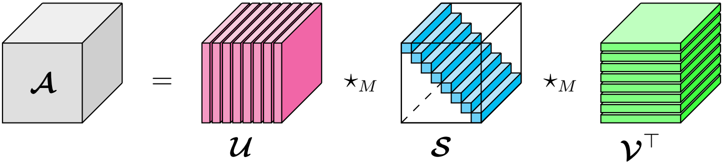
Strang, *Linear Algebra and Learning from Data*

t-SVDM (third order)

Given $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and an invertible $\boldsymbol{M} \in \mathbb{R}^{n_3 \times n_3}$, the t-SVDM of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{U} \star_{\mathrm{M}} \mathcal{S} \star_{\mathrm{M}} \mathcal{V}^T$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$.



Kilmer et al., *Tensor-Tensor Products for Optimal Representation and Compression*

Local t-SVDM Algorithm

- Pre-processing
  1. Separate the training dataset $\boldsymbol{\mathcal{A}}$ into distinct classes

$$\boldsymbol{\mathcal{A}}_1, \boldsymbol{\mathcal{A}}_2, \ldots, \boldsymbol{\mathcal{A}}_{\#\text{ of classes}}$$

  2. For each class i, compute a truncated local t-SVDM and store the first k basis elements

$$\boldsymbol{\mathcal{A}}_i = \boldsymbol{\mathcal{U}}_i \star_{\mathrm{M}} \boldsymbol{\mathcal{S}}_i \star_{\mathrm{M}} \boldsymbol{\mathcal{V}}_i^\top \qquad \boldsymbol{\mathcal{U}}_{i,k} = \boldsymbol{\mathcal{U}}_i(:, 1:k, :)$$

The Frobenius norm for third-order tensors is $\|\boldsymbol{\mathcal{B}}\|_F = \sqrt{\sum_{i,j,k} b_{ijk}^2}$.

Newman, Kilmer, and Horesh, *Image classification using local tensor singular value decompositions*

Local t-SVDM Algorithm

- Pre-processing
  1. Separate the training dataset $\mathcal{A}$ into distinct classes

$$\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{\#\text{ of classes}}$$

  2. For each class i, compute a truncated local t-SVDM and store the first k basis elements

$$\mathcal{A}_i = \mathcal{U}_i \star_{\mathrm{M}} \mathcal{S}_i \star_{\mathrm{M}} \mathcal{V}_i^\top \qquad \mathcal{U}_{i,k} = \mathcal{U}_i(:, 1:k, :)$$

- For each test image $\mathcal{T}$
  1. For each basis, project a test image onto the space spanned by the class basis

$$\mathcal{P}_i = \mathcal{U}_{i,k} \star_{\mathrm{M}} \mathcal{U}_{i,k}^\top \star_{\mathrm{M}} \mathcal{T}$$

  2. Categorize the test image as the class whose projection was "closest" to the original image

$$i^* = \arg\min_i \|\mathcal{T} - \mathcal{P}_i\|_F$$

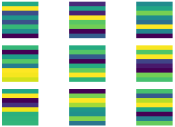
The Frobenius norm for third-order tensors is $\|\mathcal{B}\|_F = \sqrt{\sum_{i,j,k} b_{ijk}^2}$.

Newman, Kilmer, and Horesh, *Image classification using local tensor singular value decompositions*

Intuition - Stripe Data Example

$\mathcal{A}_1$

$\mathcal{A}_2$



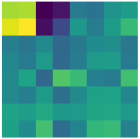
$\mathcal{U}_1$

$\mathcal{U}_2$

$\mathcal{T}$

$\mathcal{P}_1$

$\mathcal{P}_2$



Choice of M

To use t-SVDM, following $\boldsymbol{M}$s are selected and implemented:

Matrix Type	Advantage	Orthogonal
Banded	Time Series	No
Haar	Capture Data Structure	Yes
Random Orthogonal	Base Line	Yes
Data-Dependent	Data Matching	Yes

Special choices of $\boldsymbol{M}$ give us the following products:

Method Type	Abbreviation	Transformation
Tensor-tensor	t	Fast Fourier Transform
Cosine	c	Discrete Cosine Transform
Facewise	f	No Transform

Malik et al., *Tensor Graph Convolutional Networks for Prediction on Dynamic Graphs*

Kernfeld, Kilmer, and Aeron, "Tensor–tensor products with invertible linear transforms"

Preliminary Results

Attributes	Matrix Approach	Tensor Approach
Dimensions	(x·y·z·time, trials)	(x, trials, y, z, time)
Shape	(524288,26)	(64,26,64,8,16)
Computation	Expensive	Parallelizable
Transformation	N/A	$\boldsymbol{M}$ can be varied
Best Accuracy	78.6%	100%
(With Parameters)	$k = 4$	$\boldsymbol{M} =$ 'ddm', $k = 4/5$

- Accuracy = number of correct classification in one class/all data points in one class

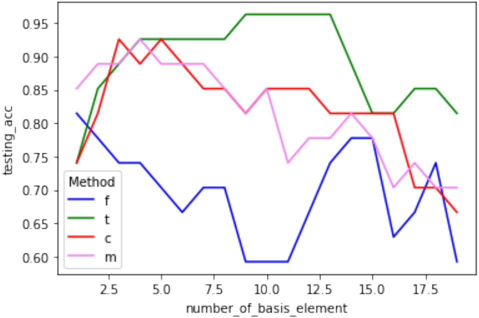
Hyperparameter Tuning

Number of basis elements: Choose the k largest singular values and their corresponding tensors to be our basis element.

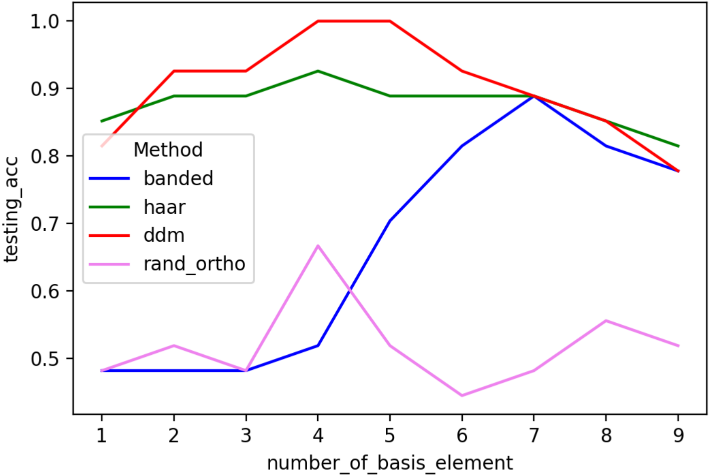
Punchline:

low representation power ⟷ high representation power

small k ⟷ large k



Choice of M



testing_acc
number_of_basis_element
1.0
0.9
0.8
0.7
0.6
0.5
1
2
3
4
5
6
7
8
9
Method
banded
haar
ddm
rand_ortho

Conclusions

Conclusion: Tensor approach outperforms matrix approach

Future Work: Experiment with various parameters:

- Transformations ($\boldsymbol{M}$)
- Bases (k)
- Distance metrics

Thank you!

References

Just, Marcel. *StarPlus fMRI data*. http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/. Center for Cognitive Brain Imaging at Carnegie Mellon University.

Kernfeld, Eric, Misha Kilmer, and Shuchin Aeron. "Tensor–tensor products with invertible linear transforms". In: *Linear Algebra and its Applications* 485 (Nov. 2015), pp. 545–570. DOI: 10.1016/j.laa.2015.07.021.

Kilmer, Misha et al. *Tensor-Tensor Products for Optimal Representation and Compression*. 2019. arXiv: 2001.00046 [math.NA].

Kolda, Tamara G. and Brett W. Bader. "Tensor Decompositions and Applications". In: *SIAM Review* 51.3 (2009), pp. 455–500. DOI: 10.1137/07070111X.

Malik, Osman Asif et al. *Tensor Graph Convolutional Networks for Prediction on Dynamic Graphs*. 2020. URL: https://openreview.net/forum?id=rylVTTVtvH.

Newman, Elizabeth, Misha Kilmer, and Lior Horesh. *Image classification using local tensor singular value decompositions*. 2017. arXiv: 1706.09693 [stat.ML].

Strang, G. *Linear Algebra and Learning from Data*. Wellesley-Cambridge Press, 2019, pp. 56–74. ISBN: 9780692196380. URL: https://books.google.com/books?id=L0Y_wQEACAAJ.

Choice of M

- Banded Matrix
  - Lower triangular and banded matrix. Entries in M on the diagonal and a specific number of subdiagonals are set to 1, while all others are 0. Finally normalization along the row.
  - Example (4*4, bandwidth = 1): $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$
- Normalized Haar Matrix
- Random Orthogonal Matrix
- Data-Dependent Matrix
  - Unfold the tensor to a matrix along an axis (whose dimension later becomes the dimension of $\boldsymbol{M}$)
  - Conduct SVD on the matrix, and $\boldsymbol{U}$ is the $\boldsymbol{M}$ we need here.

Malik et al., *Tensor Graph Convolutional Networks for Prediction on Dynamic Graphs*

EMORY UNIVERSITY