Functional alignment by the "light" approach of the von Mises-Fisher-Procrustes model

Angela Andreella[1] Livio Finos[2]

[1]Department of Statistical Sciences, University of Padua
[2]Department of Developmental Psychology and Socialisation, University of Padua

SIS - 25 June 2021

Statistical methods for High dimensional data

Multi-subjects fMRI studies permit to compare studies across subjects, to generalize and to validate the results.

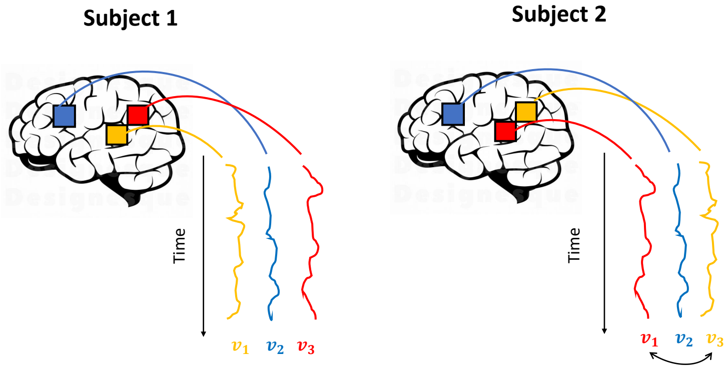
The anatomical and functional structure of brains vary across subjects even in response to identical sensory inputs.

↓

ALIGNMENT STEP

- **Anatomical Alignment** (Talairach space) [1];
- **Functional Alignment**.

[1] Talairach, J. J. and Tournoux, P. (1988)



We can assume that the neural activities in different brains are **noisy rotations of a common space**.

Most of the approaches are based on the **Procrustes** theory:

$$\min_{\boldsymbol{R}_i \in \mathcal{O}(m);\alpha_i \in \mathbb{R}_{>0}} \sum_{i=1}^{N} \|\alpha_i \boldsymbol{X}_i \boldsymbol{R}_i - \boldsymbol{M}\|_F^2$$

having n time points and m voxels:

- $\{\boldsymbol{X}_i \in \mathbb{R}^{n\times m}\}_{i=1,\dots,N}$ represent the matrices to be aligned;
- $\boldsymbol{M} \in \mathbb{R}^{n\times m}$ is the configuration reference matrix;
- $\{\boldsymbol{R}_i \in \mathcal{O}(m)\}_{i=1,\dots,N}$ is the set of orthogonal matrices of interest, and $\alpha_i \in \mathbb{R}_{>0}$ scaling factor.

The most famous are **Hyperalignment** [2] and **Generalized Procrustes Analysis**[3] (GPA). However, both of them do not return a **unique** solution of $\boldsymbol{R}_i$.

[2] Haxby, J. V., et al. (2011)

[3] Schonemann, P. H. (1966).

We proposed the **von Mises-Fisher-Procrustes model** which rephrases the Procrustes problem in terms of statistical model to insert a regularization term for $\boldsymbol{R}_i$.

$$\boldsymbol{X}_i = \alpha_i(\boldsymbol{M} + \boldsymbol{E}_i)\boldsymbol{R}_i^\top$$

where

- $\vec{\boldsymbol{E}}_i \sim \mathcal{N}_{nm}(\vec{0}, \boldsymbol{\Sigma}_n \otimes \boldsymbol{\Sigma}_m)$;
- $\boldsymbol{R}_i \sim$ **von Mises-Fisher**[4] distribution with location parameter $\boldsymbol{F} \in \mathbb{R}^{m \times m}$ and concentration parameter $k \in \mathbb{R}_{>0}$, i.e.,

$$\boxed{f(\boldsymbol{R}_i) = C \exp(k\, tr(\boldsymbol{F}^\top \boldsymbol{R}_i))}$$

[4] Downs, T. D. (1972).

Theorem

Let the singular value decomposition of $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}$ *be* $\boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^\top$*, then the vMFP model returns:*

1. $\hat{\boldsymbol{R}}_i$ *equals* $\boldsymbol{U}_i \boldsymbol{V}_i^\top$*;*
2. $\hat{\alpha}_{i\hat{\boldsymbol{R}}_i}$ *equals* $\dfrac{\|\boldsymbol{X}_i \hat{\boldsymbol{R}}_i\|_F^2}{\mathrm{Tr}(\boldsymbol{D}_i)}$.

- The von Mises-Fisher distribution is **conjugate**[5] to the matrix normal distribution, with location parameter $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}$.
- The maximum a posteriori estimators for $\boldsymbol{R}_i$ and α_i are a **slight modification** of the GPA's results. We decompose $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}$ instead of $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1}$.

[5] Mardia et al. (2013).

BUT → The **singular value decomposition** of $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1}$ has time complexity equals $O(m^3)$.

If m becomes large, as in fMRI data where m is in the order of a few under-thousands, the computation runtime is inadmissible, as well as the storing memory required.

⬇

“Light” Approach

First of all, the Procrustes minimization equals to the above maximization:

$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} \operatorname{Tr}(\boldsymbol{R}_i^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1})$$

We project the matrices $\boldsymbol{X}_i$ into a n **lower-dimensional** space with $n \ll m$ by specific **semi-orthogonal** transformations which preserve all the data's information.

Theorem

Let consider the thin singular value decompositions of $\boldsymbol{X}_i = \boldsymbol{L}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top$, where $\boldsymbol{Q}_i \in \mathbb{R}^{n \times m}$. The following holds:

$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} \mathrm{Tr}\left(\boldsymbol{R}_i^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1}\right) = \max_{\boldsymbol{R}_i^\star \in \mathcal{O}(n)} \mathrm{Tr}\left(\boldsymbol{R}_i^{\top\star} \boldsymbol{Q}_i^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1} \boldsymbol{Q}_j\right).$$

The same idea of Theorem 2 can be exploited to define the “light” version of the von Mises-Fisher-Procrustes model:

Lemma

Let consider the von Mises-Fisher-Procrustes model, and the assumptions of Theorem 2, then:

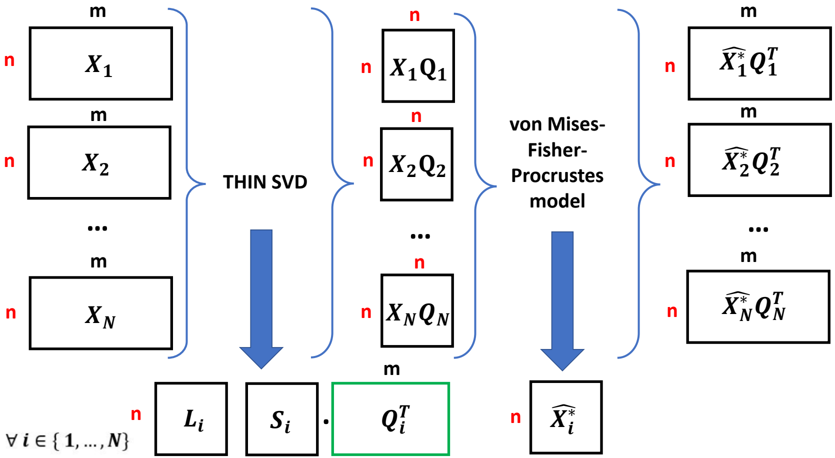
$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} \mathrm{Tr}(\boldsymbol{R}_i^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}) =$$
$$\max_{\boldsymbol{R}_i^\star \in \mathcal{O}(n)} \mathrm{Tr}(\boldsymbol{R}_i^{\top\star} \boldsymbol{Q}_i^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1} \boldsymbol{Q}_j + k\boldsymbol{F}^\star)$$

.

where $\boldsymbol{F} \in \mathbb{R}^{m \times m}$ *and* $\boldsymbol{F}^\star \in \mathbb{R}^{n \times n}$.

- The "light" approach reaches the **same maximum** while working in the **reduced space** of the first n eigenvectors (which contains all the information) instead of the full set of data;
- The original problem estimates orthogonal matrices of size $m \times m$: $\boldsymbol{R}_i \in \mathcal{O}(m)$, while the "light" solution provides a set of orthogonal matrices of size $n \times n$: $\boldsymbol{R}_i^\star \in \mathcal{O}(n)$;
- The "light" approach reaches the same fit to the data, but under a different set of **constraints** (i.e., $n \times n$ orthogonal matrices instead of $m \times m$), hence the solutions of the two algorithms will not be identical.

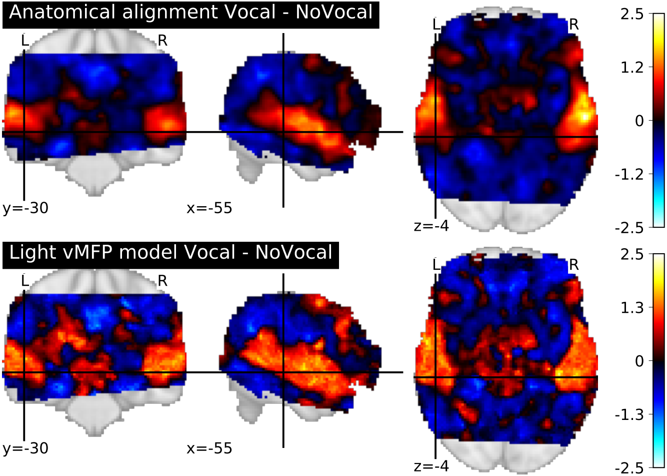
- The theoretical **time complexity** of the algorithm reduces from $O(m^3)$ to $O(mn^2)$, while the space complexity, i.e., memory, from $O(m^2)$ to $O(mn)$;
- The proposed model easily applicable to **high-dimensional data** such as the application presented in the next section, where the data dimension moves from roughly $200,000$ to 200.


m
n
X_1
m
n
X_2
...
m
n
X_N
THIN SVD
n
n
X_1Q_1
n
n
X_2Q_2
...
n
n
X_NQ_N
von Mises-Fisher-Procrustes model
m
n
$\widehat{X_1^*}Q_1^T$
m
n
$\widehat{X_2^*}Q_2^T$
...
m
n
$\widehat{X_N^*}Q_N^T$
$\forall\, i \in \{1, \dots, N\}$
n
L_i
S_i
m
Q_i^T
n
$\widehat{X_i^*}$

We align the images of of 18 subjects passively listening to vocal, i.e., speech, and non-vocal sounds.

After the $\boldsymbol{X}_i$ matrices' alignment, **the one-sample t-test** was performed to study the significance of the group's mean activation concerning the difference between the neural activation during the two stimuli.

The inferential analysis is performed on the **whole brain**. The "light" von Mises-Fisher-Procrustes model is compared with the **anatomical alignment**, being the only method applicable to the entire brain.



The “light” version returns tests 65.67% higher than those returned by the anatomical alignment, with baseline 50%.

- The "light" version permits to speed up the **computation time** in performing the SVD step of the estimation process;
- and to apply the functional alignment on **high-dimensional data** as fMRI data;
- The algorithm proposed allows a faster and more accessible shape analysis **without loss of information** in the case of $n \ll m$;
- The alignment using the "light" approach takes approximately 5 minutes while ≈ 1 hour was required to run the original vMFP model having $m \approx 10,000$;
- In the whole-brain analysis, the improvement with respect to the **anatomical alignment** is noticeable, and the computational effort remains affordable (≈ 2 hours).