EFFICIENT PROCRUSTES-BASED FUNCTIONAL ALIGNMENT FOR HIGH-DIMENSIONAL DATA

ANGELA ANDREELLA[1]   LIVIO FINOS[2]

[1]DEPARTMENT OF STATISTICAL SCIENCES, UNIVERSITY OF PADOVA
[2]DEPARTMENT OF DEVELOPMENTAL PSYCHOLOGY AND SOCIALISATION, UNIVERSITY OF PADOVA

13TH VIRTUAL CONFERENCE OF THE ITALIAN REGION OF THE INTERNATIONAL BIOMETRIC SOCIETY (IBS)

29 SEPTEMBER 2021

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.
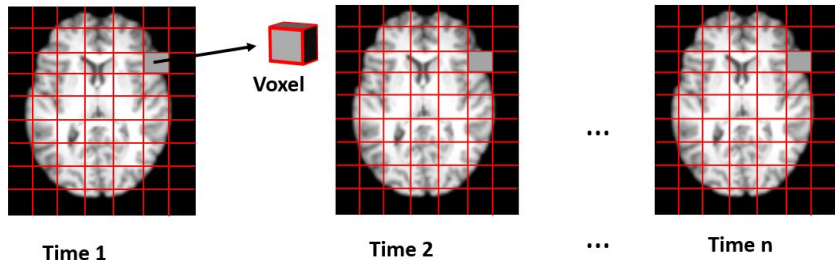
ALIGNMENT STEP

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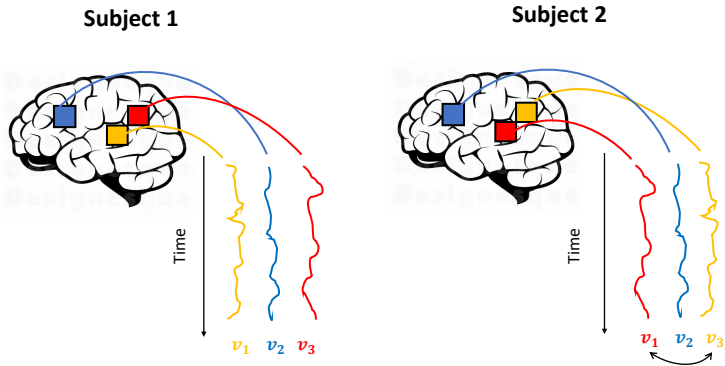- **Functional Alignment**.

---

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

Time 1   **Time 2**   ...   **Time n**

Each subject i is represented by a matrix $\mathbf{X}_i \in \mathbb{R}^{n \times m}$:

- the **rows** represent the **response stimuli activation** of voxels
  ⟶ the stimuli are time synchronized

- the **columns** represent the **time series of activation** for each m voxel
  ⟶ not assumed to be in correspondence across N subjects.

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:

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

The most famous are **Hyperalignment** [2] and **Generalized Procrustes Analysis**[3] (GPA) → **no unique** solution for $\boldsymbol{R}_i$.

---

[2]Haxby, J. V., et al. (2011)
[3]Schonemann, P. H. (1966).

We proposed the **ProMises model**:

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

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 $k \in \mathbb{R}_{>0}$:

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

---

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

Theorem

Let the SVD of $\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F} = \boldsymbol{U_i} \boldsymbol{D_i} \boldsymbol{V_i}^\top$:

1. $\hat{\boldsymbol{R}}_i = \boldsymbol{U_i} \boldsymbol{V_i}^\top$;

2. $\widehat{\alpha}_{i\hat{\boldsymbol{R}}_i} = \dfrac{\|\boldsymbol{X_i}\hat{\boldsymbol{R}}_i\|_F^2}{\mathrm{Tr}(\boldsymbol{D_i})}$.

- The von Mises-Fisher distribution is **conjugate** 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 $\boxed{\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}}$ instead of $\boxed{\boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{M} \boldsymbol{\Sigma_m}^{-1}}$.

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

If m becomes large, as in fMRI data where $m \approx 200,000$, the computation runtime and the storing memory required are inadmissible.

Efficient Approach

First of all, the Procrustes minimization equals

$$\max_{R_i \in \mathcal{O}(m)} \mathrm{Tr}(R_i^\top X_i^\top \Sigma_n^{-1} X_j \Sigma_m^{-1}).$$

Project $\{\boldsymbol{X}_i\}_{i=1,\ldots,N}$ 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 SVD $\boldsymbol{X}_i = \boldsymbol{L}_i \boldsymbol{S}_i \boldsymbol{Q}_i^{\top}$, where $\boldsymbol{Q}_i \in \mathbb{R}^{n \times m}$:

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

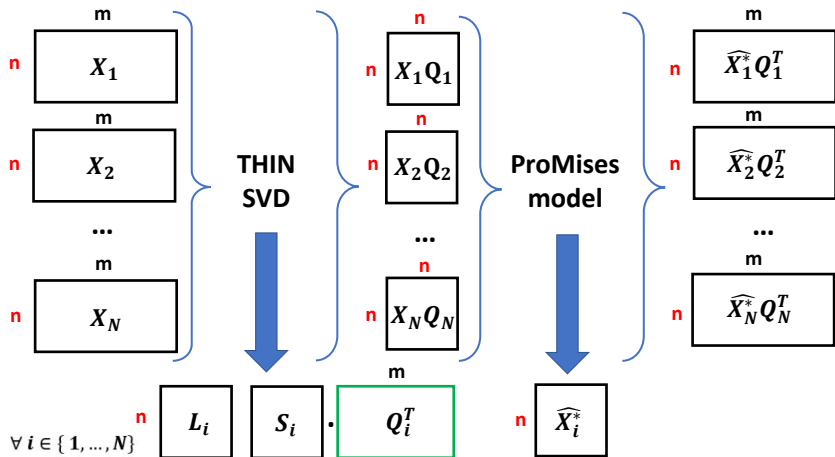The same idea is used to define the efficient version of the
ProMises model.

Lemma

Let consider the assumptions of the ProMises model:

$$\max_{\boldsymbol{R_i} \in \mathcal{O}(m)} \text{Tr}(\boldsymbol{R_i}^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1} + k\boldsymbol{F}) =$$

$$\max_{\boldsymbol{R_i^\star} \in \mathcal{O}(n)} \text{Tr}(\boldsymbol{R_i}^{\top \star} \boldsymbol{Q_i}^\top \boldsymbol{X_i}^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X_j} \boldsymbol{\Sigma_m}^{-1} \boldsymbol{Q_j} + k\boldsymbol{F^\star})$$

.

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

- We align the images of 18 subjects passively listening to vocal, i.e., speech, and non-vocal sounds from Pernet et al. (2015)[5].

- After the X_i matrices' alignment → **One-sample t-tests** 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 efficient ProMises model is compared with the **anatomical alignment**, being the only method applicable to the entire brain.

---

[5]https://openneuro.org/datasets/ds000158/versions/1.0.0
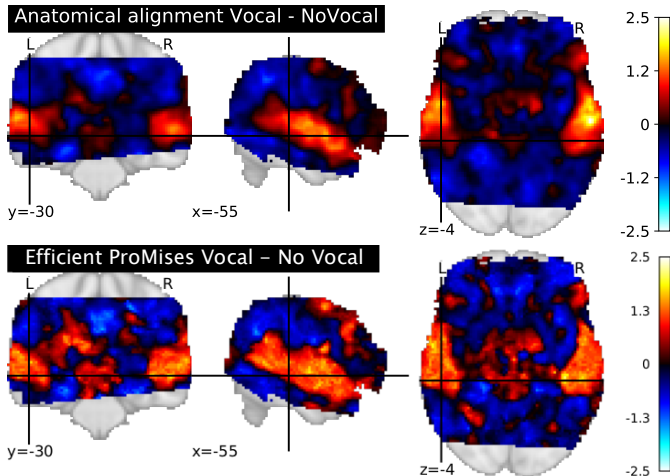
The efficient version returns tests 65.67% higher than those returned by the anatomical alignment, with baseline 50%.

- **Same maximum** working in the **reduced space** of the first n eigenvectors (different set of constraints);

- **Time complexity** reduces from $O(m^3)$ to $O(mn^2)$, while the **space complexity** from $O(m^2)$ to $O(mn)$;

- Having $m \approx 10,000$, the efficient approach takes ≈ 5 **minutes** while ≈ 1 **hour** was required for the original ProMises model;

- In the **whole-brain analysis**, the improvement with respect to the **anatomical alignment** is noticeable, and the computational effort remains **affordable** (≈ 2 hours).

Backup slides

PROMISES MODEL ALGORITHM

Data: $X_i, k, F, \mathsf{T}, \texttt{maxIt}\ \forall i = 1, \dots, N$
Result: $\hat{X}_i\ \forall i = 1, \dots, N$

$\hat{X} = X, \hat{\alpha}_i = 1, \hat{\Sigma}_n = I_n, \hat{\Sigma}_m = I_m, \hat{\Sigma}_{n;\text{old}} = I_n, \hat{\Sigma}_{m;\text{old}} = I_m, M = \bar{\hat{X}}$
$\texttt{count} = 0, \texttt{dist} = \texttt{Inf}$
while $\textit{dist} > \mathsf{T}\ \textit{OR}\ \textit{count} < \textit{maxIt}$ **do**
    for $i = 1$ *to* N **do**
        $U_i D_i V_i^\top = \texttt{SVD}(X_i^\top \hat{\Sigma}_n^{-1} M \hat{\Sigma}_m^{-1} + k \cdot F)$ // Singular Value Decomposition
        $\hat{R}_i = U_i V_i^\top$
        $\hat{X}_i = X_i \hat{R}_i$
        $\hat{\alpha}_{i\hat{R}_i} = \frac{\|\hat{X}_i^\top \hat{\Sigma}_n^{-1} M \hat{\Sigma}_m^{-1} + k \cdot F\|^2}{\text{Tr}(D_i)}$
        $\hat{X}_i = \hat{\alpha}_i^{-1} X_i \hat{R}_i$ // Update X_i
    end
    $M_{\text{old}} = M, M = \bar{\hat{X}}$ // Save and Update M
    $\hat{\Sigma}_n = g(\hat{\Sigma}_m, \hat{X}_i, M), \hat{\Sigma}_m = g(\hat{\Sigma}_n, \hat{X}_i, M)$
    while $\|\hat{\Sigma}_n - \hat{\Sigma}_{m;old}\|^2 > \epsilon_1\ \textit{OR}\ \|\hat{\Sigma}_n - \hat{\Sigma}_{m;old}\|^2 > \epsilon_1$ **do**
        $\hat{\Sigma}_{n;\text{old}} = \hat{\Sigma}_n, \hat{\Sigma}_{m;\text{old}} = \hat{\Sigma}_m$
        $\hat{\Sigma}_n = g(\hat{\Sigma}_m, \hat{X}_i, M), \hat{\Sigma}_m = g(\hat{\Sigma}_n, \hat{X}_i, M)$
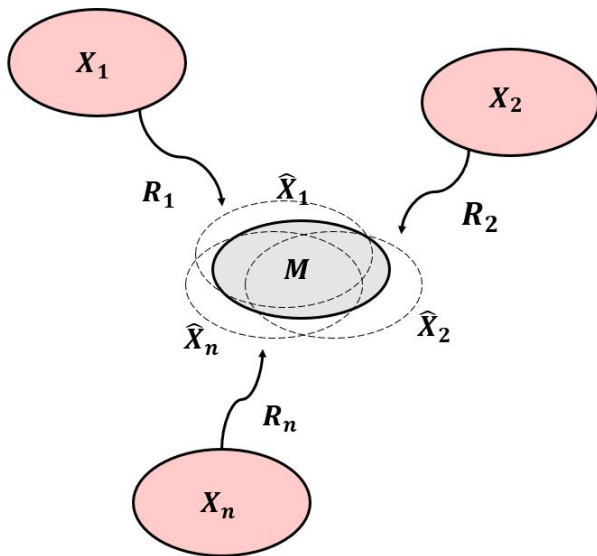   **end**
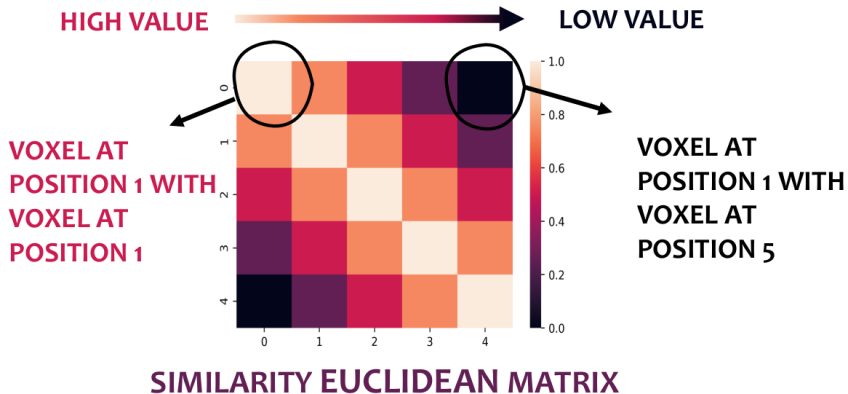 $\texttt{dist} \leftarrow \|M - M_{\text{old}}\|^2, \texttt{count} \leftarrow \texttt{count} + 1$
end

HIGH VALUE ⟶ LOW VALUE

VOXEL AT POSITION 1 WITH VOXEL AT POSITION 1

VOXEL AT POSITION 1 WITH VOXEL AT POSITION 5

SIMILARITY EUCLIDEAN MATRIX

$$F = \exp(-d_{ij})$$