Procrustes analysis for high dimensional data

Angela Andreella & Livio Finos

Statistical methods and models for complex data — Poster Session: 21 September 2022

INTRODUCTION

Procrustes problem: matches matrices using **similarity transformations** by minimizing their Frobenius distance.

Most of the applications comprise **spatial coordinates**. The extension to **high dimensional** context lits:

- **Ill-posed problem**;
- **Polynomial time complexity** in no. of variables to align.

PERTURBATION MODEL (Goodall, 1991)

Let $\{X_i \in \mathbb{R}^{n\times m}\}_{i=1,\dots,N}$ be a set of matrices to be aligned:

$$X_i = \alpha_i(M + E_i)R_i^\top + 1_n^\top t_i \quad \text{subject to} \quad R_i \in O(m)$$

- $E_i \sim MN_{n,m}(0, \Sigma_n, \Sigma_m)$;
- $R_i \in O(m)$;
- $M \in \mathbb{R}^{n\times m}$ is the **shared matrix**;
- $\alpha_i \in \mathbb{R}^+$ is the isotropic **scaling**;
- $t_i \in \mathbb{R}^{1\times m}$ defines the **translation** vector.

Consider the SVD of $X_i^\top \Sigma_n^{-1} M \Sigma_m^{-1} = U_i^\top D_i V_i$:

- $\hat{R}_i = U_i^\top V_i$;
- $\hat{\alpha}_{i R_i} = ||\Sigma_m{}^{1/2} \hat{R}_i^\top X_i \Sigma_n^{-1/2}||_F^2 / tr(D_i)$.

Consider N independent X_i. If M, Σ_m, Σ_n are unknown:

Require: X_i, T, maxIt, $\forall i = 1, \dots, N$
Ensure: $\hat{X}_i$ $\forall i = 1, \dots, N$

1: $\hat{M} = \sum_{i=1}^N X_i/N$, $\hat{\alpha}_i = 1$, $\hat{\Sigma}_{n;\,\text{old}} = \hat{\Sigma}_n = I_n$, $\hat{\Sigma}_{m;\,\text{old}} = \hat{\Sigma}_m = I_m$, count = 0, dist = Inf
2: **while** dist > T OR count < maxIt **do**
3:   **for** $i = 1$ to N **do**
4:     $U_i D_i V_i^\top = \text{SVD}(X_i^\top \hat{\Sigma}_n{}^{-1} \hat{M} \hat{\Sigma}_m{}^{-1})$
5:     $\hat{R}_i = U_i V_i^\top$
6:     $\hat{X}_i = X_i \hat{R}_i$
7:     $\hat{\alpha}_{i\hat{R}_i} = ||\hat{\Sigma}_m^{-1/2} \hat{R}_i^\top X_i^\top \hat{\Sigma}_n{}^{-1/2}||^2 / tr(D_i)$
8:     $\hat{X}_i = \hat{\alpha}_{i\hat{R}_i}^{-1} \hat{X}_i$ ▷ Update X_i
9:   **end for**
10:   $\hat{M}_{\text{old}} = \hat{M}$ ▷ Save $\hat{M}$
11:   $\hat{M} = \sum_{i=1}^N \hat{X}_i/N$ ▷ Update $\hat{M}$
12:   $\hat{\Sigma}_n = g(\hat{\Sigma}_m, \hat{M}, X_i)$, $\hat{\Sigma}_m = g(\hat{\Sigma}_n, \hat{M}, X_i)$
13:   **while** $||\hat{\Sigma}_n - \hat{\Sigma}_{n;\,\text{old}}||^2 > \epsilon_1$ | $||\hat{\Sigma}_m - \hat{\Sigma}_{m;\,\text{old}}||^2 > \epsilon_2$ **do**
14:     $\hat{\Sigma}_{n;\,\text{old}} = \hat{\Sigma}_n$, $\hat{\Sigma}_{m;\,\text{old}} = \hat{\Sigma}_m$
15:     $\hat{\Sigma}_n = g(\hat{\Sigma}_m, \hat{M}, X_i)$, $\hat{\Sigma}_m = g(\hat{\Sigma}_n, \hat{M}, X_i)$
16:   **end while**
17:   dist = $||\hat{M} - \hat{M}_{\text{old}}||^2$, count = count + 1.
18: **end while**

ISSUES:

- If $Z \in O(m)$, then $\{\hat{R}_i Z\}_{i=1,\dots,N}$ still valid ML solutions.
- If $n < m$, then the ML estimate for R_i is **not unique**.

PROMISES MODEL (Andreella and Finos, 2022)

HOW TO SOLVE THE ILL-POSED PROBLEM?

→ Prior on R_i: **von Mises-Fisher distribution**:

$$f(R_i) = C(F, k)\exp\{tr(kF^\top R_i)\}$$

- $F \in \mathbb{R}^{m\times m}$ is the **location parameter**,
- $k \in \mathbb{R}^+$ is the **concentration parameter**.

The **mode** equals the polar part of **F**.

- the posterior distribution is **conjugate** with location parameter:

$$F^\star = X_i{}^\top \Sigma_n^{-1} M \Sigma_m^{-1} + kF.$$

- Consider the SVD of $F^\star$:
  - $\hat{R}_i = U_i^\top V_i$;
  - $\hat{\alpha}_{i R_i} = ||\Sigma_m{}^{1/2} \hat{R}_i^\top X_i \Sigma_n^{-1/2}||_F^2 / tr(D_i)$.
- If F has full rank, the MAP estimates for R_i are unique.

Consider N independent X_i. If M, Σ_m, Σ_n are unknown → we modify line 5 of the perturbation model's algorithm.

EFFICIENT PROMISES MODEL

HOW TO SOLVE POLYNOMIAL TIME COMPLEXITY PROBLEM?

The ProMises model provides unique solutions but at each step, we perform N **SVD of** $m \times m$ **matrices**, i.e., $O(m^3)$.

Let Q_i be the semi-orthogonal matrix from the **thin** SVD of X_i. Then:

$$\max_{R_i \in O(m)} tr(R_i^\top X_i^\top \Sigma_n{}^{-1} X_j \Sigma_m^{-1} + kF) =$$
$$\max_{R_i^\star \in O(n)} tr\{R_i^{\star\top}(Q_i^\top X_i^\top \Sigma_n^{-1} X_j \Sigma_m^{-1} Q_j^\top + kF^\star)\},$$

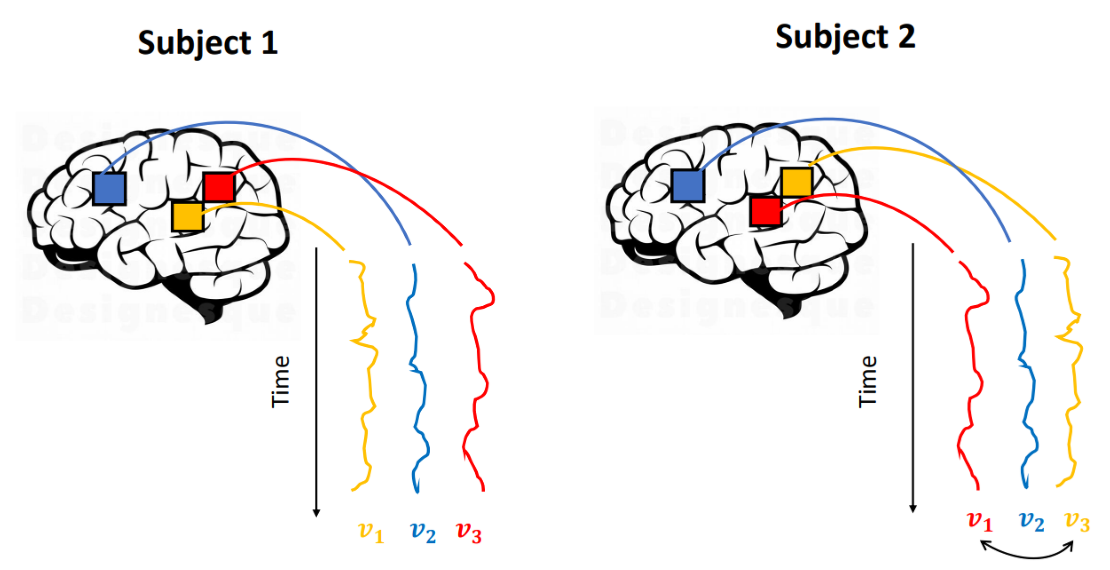
$\{X_i Q_i \in \mathbb{R}^{n\times n}\}_{i=1,\dots,N}$ are aligned by the perturbation or ProMises model. Then, $\{Q_i^\top\}_{i=1,\dots,N}$ projects back the aligned matrices to the original $n \times m$-size matrices.

The lines to reduce the dimensions of X_i are added on the perturbation model's algorithm → **Time complexity** from $O(m^3)$ to $O(mn^2)$, and the **space** one from $O(m^2)$ to $O(mn)$.

FMRI DATA APPLICATION

Each subject's **brain activation** is represented by a matrix:

- **Rows** represent the **stimuli/time points** → correspondence btw subject (time-synchronized),
- **Columns** represent the **voxels** → not correspondence btw subject:



AIM: Represent the neural responses to stimuli into a **common high-dimensional space**, instead of canonical anatomical space → variability of functional topographies loci.

SOLUTION: **Procrustes-based** models.

High-dimensional problem → each voxel is one dimension:

1. Perturbation model combines any voxel inside the brain without considering **anatomical locations**;
2. non-identifiability of R_i → loss of **topological interpretability** of the results;
3. **computational load** → SVD of many square matrices of dimensions ≈ 200.000, i.e., the number of voxels.

We solve these issues using the efficient ProMises model!

The anatomical information are inserted into **F** → defined as an **Euclidean similarity matrix** using the 3*D* **anatomical coordinates** x, y and z of each voxel:

$$F = \left[\exp\left\{-\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}\right\}\right],$$

where $i, j = 1, \dots m$. So:

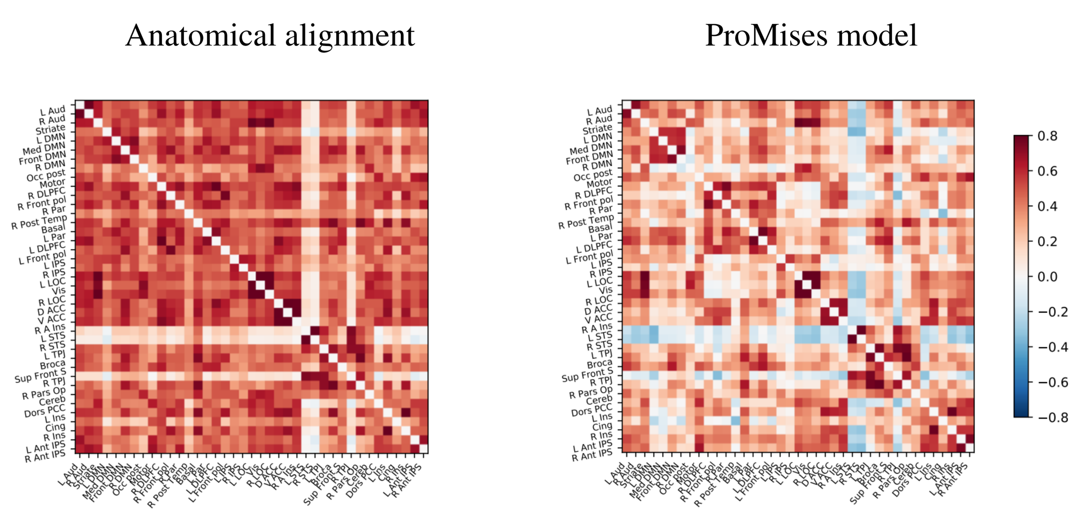
- Voxels with the **same spatial location** combined with weights 1;
- weights **decrease** as the voxels are more **spatially distant**.

1. FUNCTIONAL CONNECTIVITY

We align brain images consisting of neural activations of 18 subjects passively listening to vocal and non-vocal sounds.

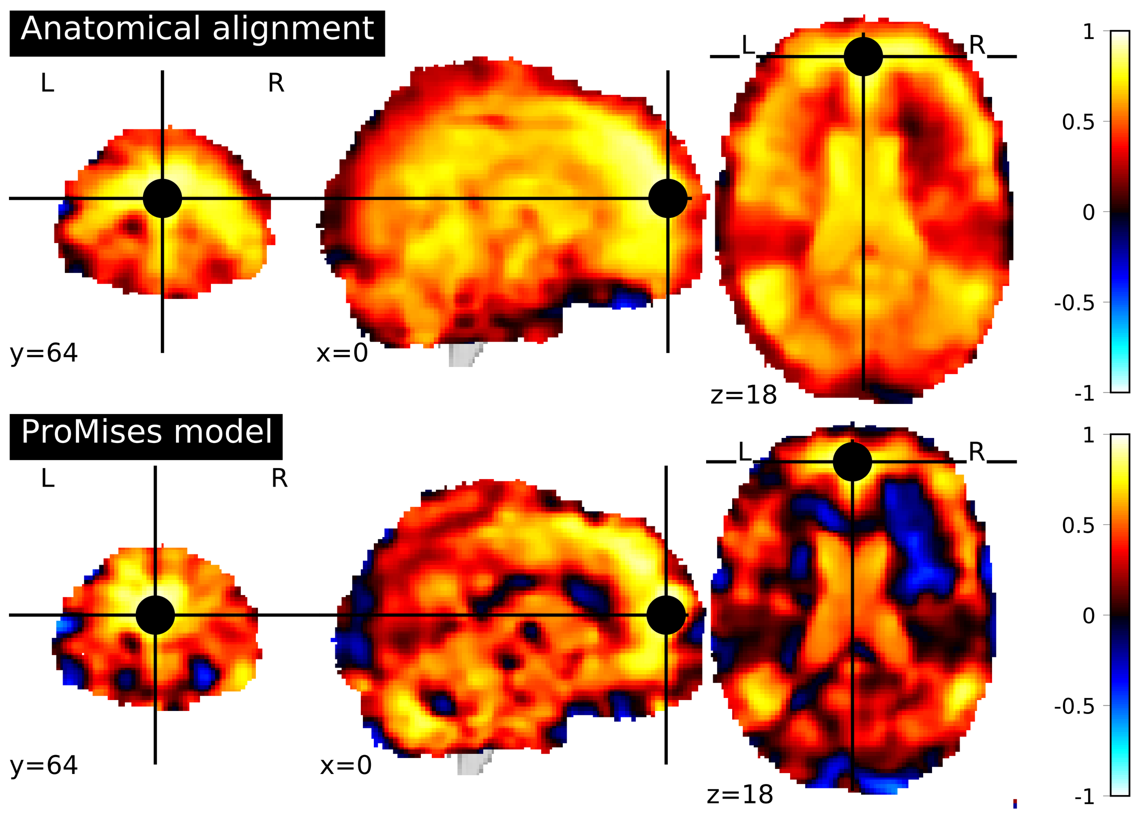
a. Region of interest correlation analysis

Functional correlation are computed btw the region of interest from a standard atlas.



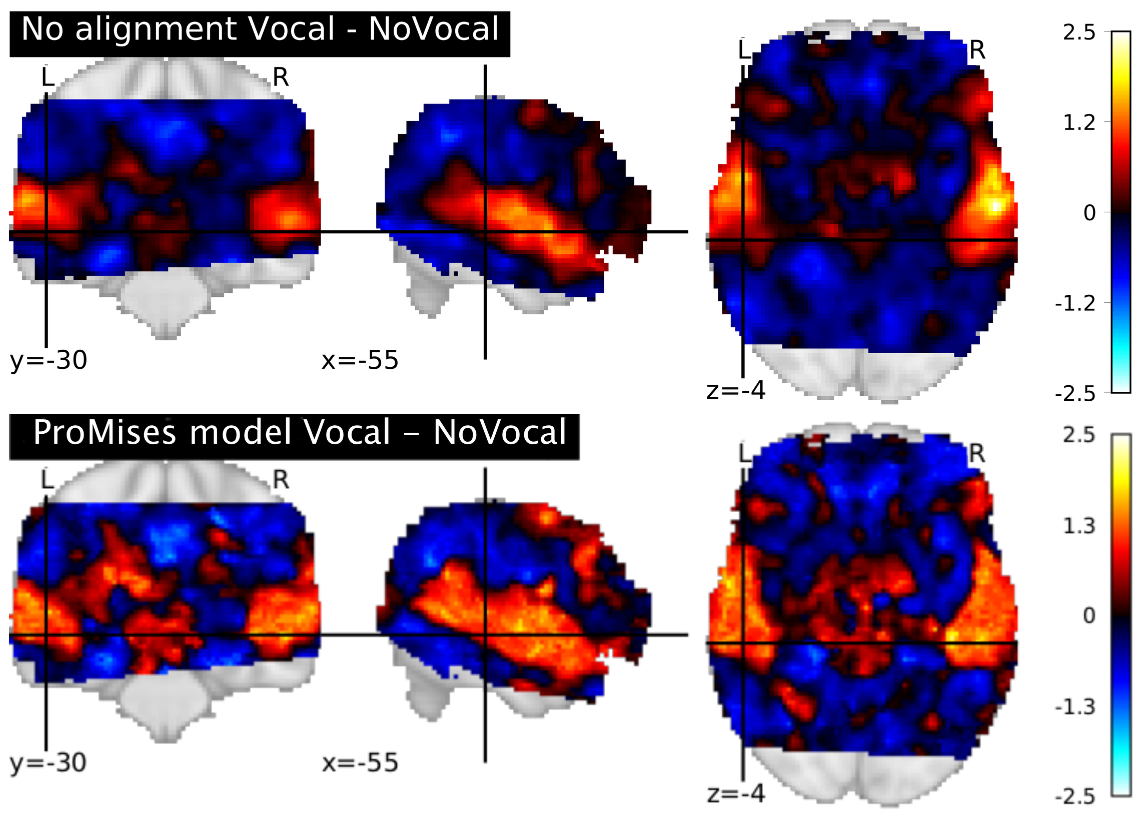
b. Seed-based correlation analysis

Functional correlation are computed btw a seed (Frontal Pole/ black dot) and any voxel.



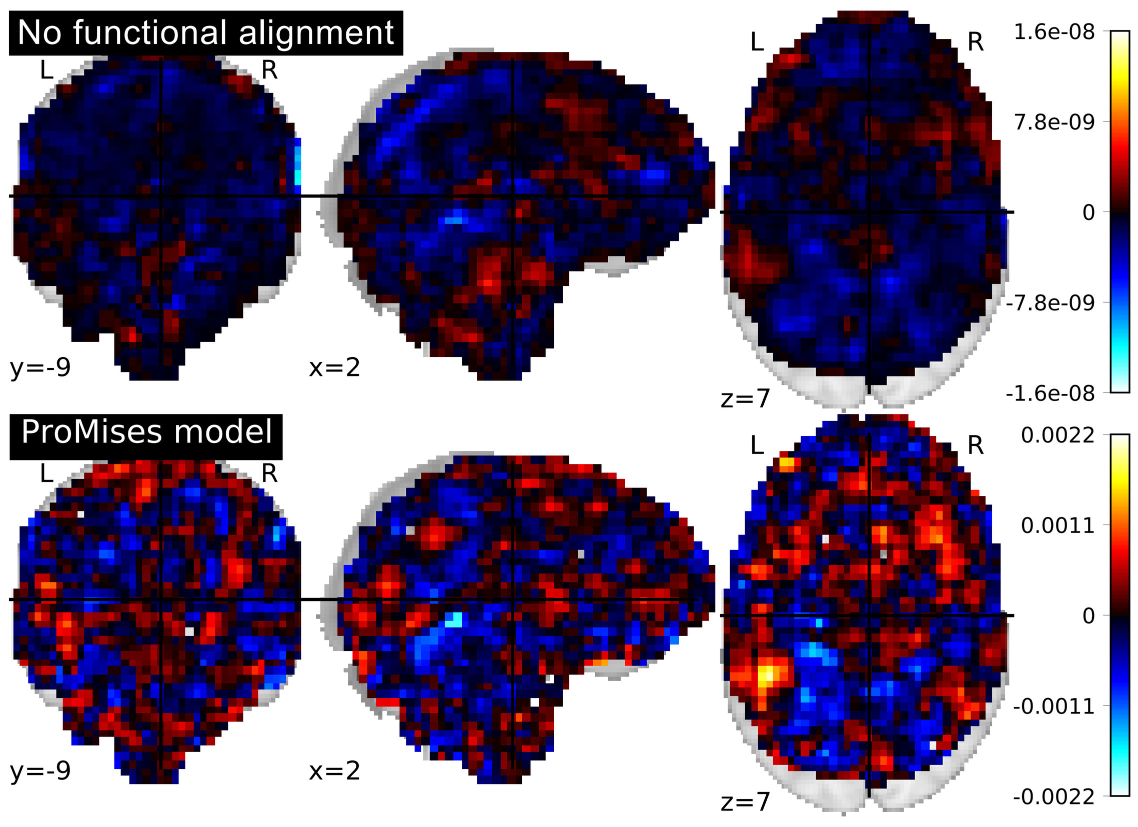
2. INFERENCE ANALYSIS

Same data as before, but we compute **t-tests** to analyze the difference of neural activation during the two stimuli.



3. CLASSIFICATION ANALYSIS

We align the brain images of 10 subjs watching images of faces and objects, and we **classify** it by multi-class SVM.



TAKE HOME MESSAGES

The issues of the perturbation model are solved:

- **Unique** and **interpretable** $\hat{R}_i$;
- Efficient approach permits the **applicability when** $m >>$;
- F leads R_i' s estimation (anatomical position in fMRI).

References

Andreella, A., Finos, L. Procrustes analysis for high-dimensional data. *Psychometrika*, 2022, 1 – 17.

Andreella, A., Finos, L., Lindquist, M. (2022) Enhanced hyperalignment via spatial prior information. *arXiv:2209.07960*

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *JRSS: Series B* **53** (2), 285 – 321.

Contact information

Angela Andreella, Assistant Professor (RTD-A)
Ca' Foscari University of Venice
angela.andreella@unive.it
https://github.com/angeella