A Statistical approach to the alignment of fMRI data

Angela Andreella[1]

join work with

Livio Finos[2] and James V. Haxby[3]

[1]Department of Statistical Sciences, University of Padua, Italy
[2]Department of Developmental Psychology and Socialisation, University of Padua, Italy
[3]Department of Psychological and Brain Sciences, Dartmouth College, NH, United States

LUMC, Leiden, 4 June 2019

Introduction

Multi-subjects functional Magnetic Resonance Images (fMRI) studies permit to

- **compare** studies across subjects, to generalize and to validate the results;
- find **shared** cognitive process;
- analyze the **inter-subject** variability.

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

⇓

ALIGNMENT STEP can improve the analysis

Alignment Step

Talairach and Tournoux (1988)[1] proposed **Anatomical Alignment**: the images are aligned to a template by an affine transformation, often followed by spatial smoothing of the data.

⬇

It doesn't align the functional characteristics

⬇

Haxby et al. (2011)[2] proposed **Hyperalignment**: functional alignment using the principle of sequential Procrustes transformations of the images, i.e. brains.

[1] Talairach, J. J, and P. Tournoux. 1988. Co-Planar Stereotaxic Atlas of the Human Brain.3-Dimensional Proportional System: An Approach to Cerebral Imaging. Cerebral Cortex. Atlante.

[2] Haxby, J. V., et al. 2011. "A common high-dimensional model of the representational space in human ventral temporal cortex." Neuron 72 (1): 404–16.

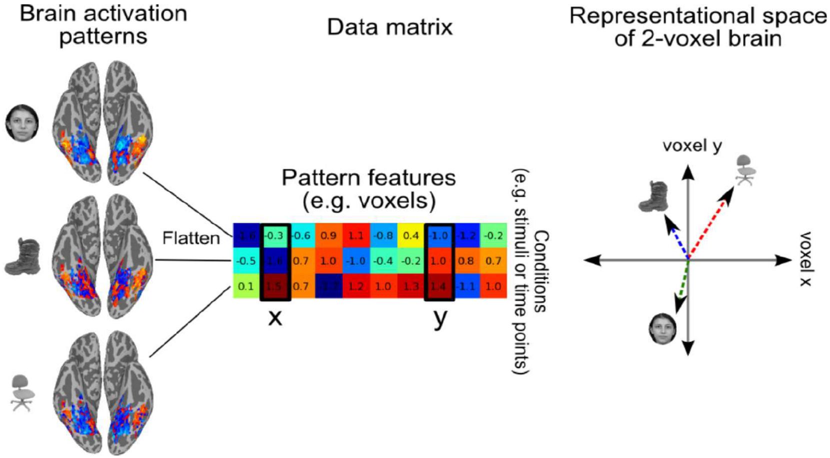


Figure: Haxby, J. V., et al. 2011. “A common high-dimensional model of the representational space in human ventral temporal cortex.” Neuron 72 (1): 404–16.

Procrustes problem in fMRI data

Let $X_i \in \mathbb{R}^{n \times v}$, where $i = 1, \dots, N$ represents the subject:

- the n rows represent the **response stimuli activation** of v voxels $\rightarrow$ the stimuli are time synchronized;
- the columns represent the **time series of activation** for each v voxel $\rightarrow$ not assumed to be in correspondence across subjects.

The Procrustes method uses similarity transformation to match matrice(s) onto the target one as close as possible.

The **Orthogonal Procrustes problem**[3] is expressed as:

$$\boxed{\min_R ||X_i - X_j R||_F^2 \quad \text{subject to} \quad R^T R = I_v}$$

[3]Schonemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. Psychometrika, 31(1):1-10.

Hyperalignment

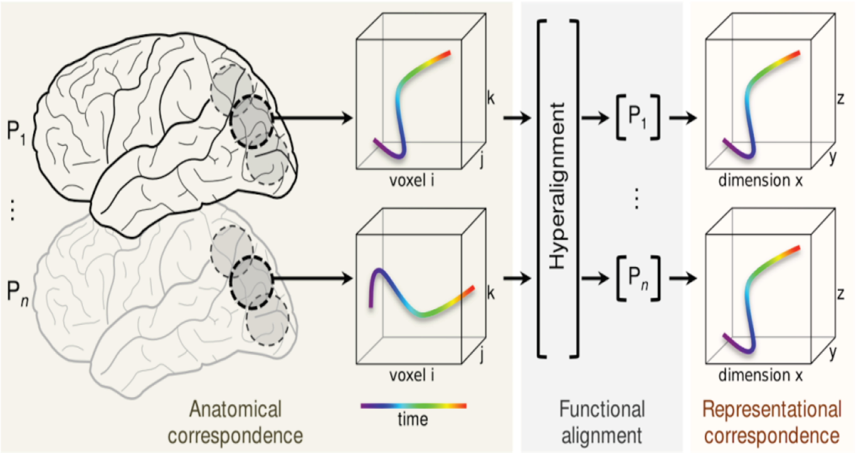


Figure: Nastase, S. A. et al., Attention selectively reshapes the geometry of distributed semantic representation, Cereb. Cortex., 2017, 27(8):4277-4291

Hyperalignment

It a **sequential** application of the Procrustes transformations. Let $X_1, \ldots, X_N$:

1. X_1 is aligned into X_2 space, X_3 is aligned into the mean of the two first transformed matrices and so on;
2. X_i are aligned into the mean of the transformed X_i from the first step;
3. R_i are recomputed considering the alignment of X_i into the mean of the transformed matrices from the third step.

⇓

No statistical approach and optimization criteria.

Statistical model

The main idea is to rephrase the Procrustes problem in terms of a **statistical model**.

Let X_i a matrix having dimension $n \times v$, with $i = 1, \dots, N$, where N is the number of subjects analyzed, n the number of time points and v the number of voxels.

The **Orthogonal Procrustes problem** can be defined as

$$\boxed{X_i = MR_i^\top + \varepsilon \quad \text{subject to} \quad R_i R_i^\top = R_i^\top R_i = I_v}.$$

The **alignment** process is the **rotation** of each dataset X_i to the common mean M.

Statistical model

- ε is the **error term** having a **Multivariate Normal Matrix** distribution with dimension $n \times v$, each row having $\sim N(0, \Sigma)$. In this case, the error terms are considered independent and having equal variance $\rightarrow$ **further extension** could be assuming dependent errors with different variance, estimating Σ from the data.

- M a matrix parameter or **mean** with dimension $n \times v$.

The maximum likelihood estimate for the parameter R_i equals to the Procrustes solution. Let the Singular Value Decomposition of $X_i^\top M = UDV^\top$, Schonemann (1966)[4] founds that $\hat{R}_i = UV^\top$.

[4] Schonemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. Psychometrika, 31(1):1-10.

Let the corresponding **likelihood function**:

$$\mathcal{L}(R_i; M) \propto \exp\{-\frac{1}{2}\operatorname{tr}((X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top)\}.$$

and therefore the **log-likelihood function** is equal to:

$$\begin{aligned}\ell(R_i; M) &\propto \operatorname{tr}(-\frac{1}{2}(X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top) \\ &= \frac{1}{2\sigma^2}(-\operatorname{tr}((X_i^\top - R_i M^\top)^\top(X_i^\top - R_i M^\top)).\end{aligned}$$

Assuming $\Sigma = \sigma^2 I$, i.e the voxels have independent errors with equal variance.

Then, the **maximization** problem to estimate R_i can be defined as:

$$\begin{aligned} R_i &= \arg\max_{R_i}(-\|X_i^\top - R_i M^\top\|_F^2) = \arg\max_{R_i}(< X_i^\top M, R_i >_F) \\ &= (\text{Let the SVD: } X_i^\top M = UDV^\top) = \arg\max_{R_i}(< UDV^\top, R_i >) \\ &= \arg\max_{R_i}(\text{tr}(V^\top DUR_i)) = \arg\max_{R_i}(\text{tr}(D \underbrace{UR_iV^\top}_{R_i^\circ})) \\ &= U \underbrace{\left(\arg\max_{R_i^o}(< D, R_i^\circ >)\right)}_{=I_v} V^\top = UV^\top \end{aligned} \tag{1}$$

The (1) step is proved by Gower and Dijksterhuis [5], due to the fact that $< D, R_i^o >$ is maximum when $R_i^o = I_v$.

[5] Gower, J. C. and Dijksterhuis, B. G. (2004). Procrustes Problems. Oxford University Press.

The joint likelihood considering N subjects is defined as:

$$\prod_{i=1}^{N} \mathcal{L}(R_i; M) \propto \prod_{i=1}^{N} \exp\{-\frac{1}{2} \operatorname{tr}((X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top)\}$$
$$= \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \operatorname{tr}((X_i - MR_i^\top)(X_i - MR_i^\top)^\top)\}.$$

The **maximization problem** to find the maximum likelihood estimation for R_i is expressed as:

$$\boxed{R_i = \arg\max_{R_i} \Big(- \sum_{i=1}^{N} ||X_i^\top - R_i M^\top||_F^2 \Big)}.$$

In this case, the solution hasn't a closed form, and so we must use **the Generalized Procrustes Solution**, an iterative algorithm.

Maximum likelihood estimation
Multi subjects case

Require: X_i, k, Q, $\texttt{T}$, $\texttt{maxIt}$,

$M \leftarrow \bar{X}$ ▷ Reference = global mean
$\texttt{count} \leftarrow 0$
$\texttt{dist} \leftarrow \texttt{Inf}$
while $\texttt{dist} > \texttt{T}$ & $\texttt{count} < \texttt{maxIt}$ **do**
  for $i = 1$ to N **do**
    $U, D, V^\top \leftarrow \texttt{SVD}(X_i^\top M)$
    $\hat{R}_i \leftarrow UV^\top$
    $\hat{X}_i \leftarrow X_i \hat{R}_i$ ▷ Update X_i
  end for
  $M_{\text{old}} = M$; ▷ Save M
  $M = \bar{\hat{X}}$; ▷ Update M
  $\texttt{dist} \leftarrow ||M - M_{\text{old}}||_F^2$
  $\texttt{count} \leftarrow \texttt{count} + 1$
end while
return $\hat{X}_i$ ▷ $\forall i = 1, \dots, N$

Prior distribution
Matrix Fisher - Von Mises distribution

We want to put **prior information** into R_i, and so analyze the most plausible rotation. The matrix R_i is orthogonal, then the prior distribution must take values in a **Stiefel manifold**.
The Matrix Fisher - Von Mises distribution was introduced by Down(1972)[6] to represent **orthogonal matrices**, as $R_i(v \times v)$, i.e:

$$f(R_i) \sim C \exp(k_0 \, tr(Q^\top R_i))$$

where

- C: normalizing constant;
- k_0: **concentration** parameter;
- Q: matrix **location** parameter with dimension $v \times v$.

[6] Downs, T. D. (1972). Orientation statistics. Biometrika, pages 665-676

Orthogonal matrix estimation
Two subjects case

Having the prior distribution for R_i

$$\begin{aligned}\mathcal{L}(R_i; M, k_0, Q) = f(X_i \cap R_i) = f(X_i|R_i)f(R_i)\\ \propto \exp\{-\frac{1}{2}\operatorname{tr}((X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top)\}\\ \cdot \exp\{k_0 \operatorname{tr}(Q^\top R_i)\}.\end{aligned}$$

and therefore the **log-likelihood function** is equal to:

$$\begin{aligned}\ell(R_i; M, k, Q) \propto \operatorname{tr}(-\frac{1}{2}(X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top) + k_0\ tr(Q^\top R_i).\\ = \frac{1}{2\sigma^2}(-\operatorname{tr}((X_i^\top - R_i M^\top)^\top(X_i^\top - R_i M^\top))\\ + k \operatorname{tr}(Q^\top R_i)).\end{aligned}$$

Let $k = 2\sigma^2 k_0$ w.l.o.g., excluding the constant term and assuming $\Sigma = \sigma^2 I$ as previously.

Orthogonal matrix estimation
Two subjects case

Then, the **maximization** problem to estimate R_i can be defined as:

$$\begin{aligned} R_i &= \arg\max_{R_i}(-||X_i^\top - R_i M^\top||_F^2 + k \ \text{tr}(Q^\top R_i)) \\ &= \arg\max_{R_i}(< X_i^\top M + k \cdot Q, R_i >) \\ &= (\text{Let the SVD: } X_i^\top M + k \cdot Q = UDV^\top) \\ &= \arg\max_{R_i}(< UDV^\top, R_i >) \\ &= UV^\top \end{aligned}$$

We only decompose $\boxed{X_i^\top M + k \cdot Q}$ instead of $X_i^\top M$.

The joint likelihood considering N subjects is defined as:

$$
\begin{aligned}
\prod_{i=1}^{N} \mathcal{L}(R_i; M, k, Q) &= \prod_{i=1}^{N} f(X_i \cap R_i) = \prod_{i=1}^{N} f(X_i | R_i) f(R_i) \\
&\propto \prod_{i=1}^{N} \exp\{-\frac{1}{2} \operatorname{tr}((X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top)\} \\
&\cdot \exp\{k_0 \operatorname{tr}(Q^\top R_i)\} \\
&= \exp\{-\frac{1}{2} \sum_{i=1}^{N} \operatorname{tr}((X_i - MR_i^\top)\Sigma^{-1}(X_i - MR_i^\top)^\top)\} \\
&\cdot \exp\{k_0 \sum_{i=1}^{N} \operatorname{tr}(Q^\top R_i)\}.
\end{aligned}
$$

Considering $k = 2\sigma^2 k_0$ w.l.o.g, the **maximization problem** to find the maximum likelihood estimation for R_i is expressed as:

$$\boxed{R_i = \arg\max_{R_i} \Big(-\sum_{i=1}^{N} ||X_i^\top - R_i M^\top||_F^2 + k \sum_{i=1}^{N} \mathrm{tr}(Q^\top R_i) \Big)}.$$

As previously, the solution hasn't a closed form, and so we must use the iterative algorithm, i.e. **the Generalized Procrustes Solution**.

We modify the Generalized Procrustes Solution in the **Singular Value Decomposition** step, we decompose $\boxed{X_i^\top M + k \cdot Q}$ instead of $X_i^\top M$.

Orthogonal matrix estimation

Multi subjects case

Require: X_i, k, Q, $\texttt{T}$, $\texttt{maxIt}$,

$M \leftarrow \bar{X}$ $\quad \triangleright$ Reference = global mean

$\texttt{count} \leftarrow 0$

$\texttt{dist} \leftarrow \texttt{Inf}$

while $\texttt{dist} > \texttt{T}$ & $\texttt{count} < \texttt{maxIt}$ **do**

  for $i = 1$ to N **do**

    $\texttt{svd} \leftarrow \texttt{SVD}(X_i^\top M + k \cdot Q)$

    $\hat{R}_i \leftarrow UV^\top$

    $\hat{X}_i \leftarrow X_i \hat{R}_i$ $\quad \triangleright$ Update X_i

  end for

  $M_{\text{old}} = M$; $\quad \triangleright$ Save M

  $M = \bar{\hat{X}}$; $\quad \triangleright$ Update M

  $\texttt{dist} \leftarrow ||M - M_{\text{old}}||_F^2$

  $\texttt{count} \leftarrow \texttt{count} + 1$

end while

return $\hat{X}_i$ $\quad \triangleright$ $\forall i = 1, \dots, N$

How to choose the prior parameters?

The matrix Q can be expressed as a **similarity matrix** considering the **coordinates** of the voxels.

We expect that **closer** voxels have **similar** rotation loadings, while voxels very **far** each other should have a **less** similar rotation loadings.

Therefore, the idea is to cast a prior distribution, i.e. information, on the **most plausible rotations**.

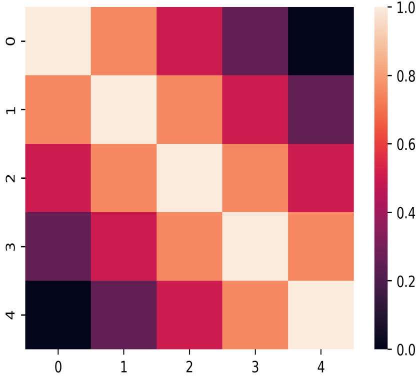
An example of the Q matrix could be:

$$\boxed{q_{ij} = 1 - \frac{d_{ij}}{\max(d_{ij})}}$$

where $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$.

How to choose the prior parameters?

For example, considering 5 voxels of 1 dimension having coordinates values between 0 and 4, we have the following Q matrix:



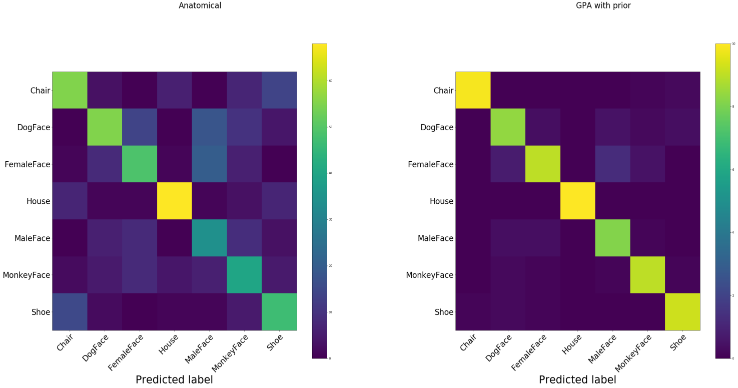
Experiments

The performance of the method proposed is assessed using two fMRI datasets used in the Haxby et al. (2011) paper:

- **Faces and Object**: analyze the Ventral Temporal Cortex of 10 subjects watching static, gray-scale images of faces and objects;
- **Raiders**: analyze the Ventral Temporal Cortex, the Occipital Lobe, the Early Visual Cortex of 31 subjects watching the movie *Raiders of the Lost Ark* (1981).

The **Linear Support Vector Machine** is used as classifier with leave one out subject cross-validation. To avoid circularity problem, the alignment and concentration parameters are fitted in leave one out runs using nested cross-validation.

- The **accuracy** classification equals to **0.67** using the data aligned by the method proposed. Instead, the anatomical alignment leads to an accuracy equals to **0.31** .
- The method proposed reduces the **error** of classification by **10%** respect to the Hyperalignment method and by **17%** respect to the classical Generalized Procrustes solution.



Anatomical
Chair
DogFace
FemaleFace
House
MaleFace
MonkeyFace
Shoe
Predicted label
GPA with prior
Chair
DogFace
FemaleFace
House
MaleFace
MonkeyFace
Shoe
Predicted label

The classifier is a **one nearest neighbor classifier** based on vector correlation having as unit time 18s segment of the movie. The alignment and concentration parameters are computed using half of the movie and using nested cross-validation.

ROI	Anatomical	GPA with prior
VT	0.232	0.413
LO	0.238	0.568
EV	0.534	0.709

The method proposed reduces the **error** of classification by **3.75%** respect to the Hyperalignment method, in VT and EV ROIs, and by **1.45%** respect to the classical Generalized Procrustes solution in VT ROI. In the rest, the reduction is less than 1.2%.

By construction, the GPA **minimizes** the sum of the distance between all the matrices and also between the matrices and the global mean:

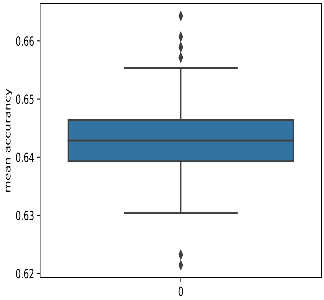
$$\sum_{i<j}^{N} ||X_i R_i - X_j R_j||_F^2 = N \sum_{i=1}^{N} ||X_i R_i - M||_F^2$$

where $M = N^{-1} \sum_{i=1}^{N} (X_i R_i)$.

The sum of the **pairwise distance** of the aligned matrices after Hyperalignment is equal to **10707.803**, while after the GPA is equal to **9207.464**.

After Hyperalignment, the brains are more **different**.

The Hyperalignment method is a **sequential** application of the Procrustes transformation. It depends on the **order of the subjects** in contrast to the method proposed.



Also, the Hyperalignment method does not reach the **global minimum** imposed by the GP.

Connections with related methods
Generalized Procrustes

Let M group average configuration and $D^\top D = I_v$:

$$\min_{R_i} N \sum_{i=1}^{N} ||X_i R_i D - M D||_F^2$$

$$= \min_{R_i} N \sum_{i=1}^{N} \mathrm{tr}(D^\top (X_i R_i - M)^\top (X_i R_i - M) D)$$

$$= \min_{R_i} N \sum_{i=1}^{N} \mathrm{tr}((X_i R_i - M)^\top (X_i R_i - M))$$

If the matrices R_i are rotated, the minimal condition is maintained. In the method proposed, the $-k\,\mathrm{tr}(Q^\top R_i)$ quantity is considered in the minimization step, so the solution is **unique**.

The solution mantains an anatomical meaning.

Conclusion

The method proposed

- doesn't depend on the **order of the subjects** as Hyperalignment;
- returns a **unique solution**, having anatomical meaning, of the rotation matrix;
- reaches the **global minimum** imposed by GP;
- improves the **classification** analysis between subjects, the functional alignment captures the fine-grained patterns of neural activity;
- uses the prior distribution to restrict the range of possible transformations $\rightarrow$ **anatomical information** $\rightarrow$ rotation matrices are more understandable.