PSYCHOMETRIKA—VOL. 87, NO. 4, 1422–1438
DECEMBER 2022
https://doi.org/10.1007/s11336-022-09859-5

PROCRUSTES ANALYSIS FOR HIGH-DIMENSIONAL DATA

ANGELA ANDREELLA
CA' FOSCARI UNIVERSITY OF VENICE

LIVIO FINOS
UNIVERSITY OF PADOVA

The Procrustes-based perturbation model (Goodall in J R Stat Soc Ser B Methodol 53(2):285–321, 1991) allows minimization of the Frobenius distance between matrices by similarity transformation. However, it suffers from non-identifiability, critical interpretation of the transformed matrices, and inapplicability in high-dimensional data. We provide an extension of the perturbation model focused on the high-dimensional data framework, called the ProMises (Procrustes von Mises–Fisher) model. The ill-posed and interpretability problems are solved by imposing a proper prior distribution for the orthogonal matrix parameter (i.e., the von Mises–Fisher distribution) which is a conjugate prior, resulting in a fast estimation process. Furthermore, we present the Efficient ProMises model for the high-dimensional framework, useful in neuroimaging, where the problem has much more than three dimensions. We found a great improvement in functional magnetic resonance imaging connectivity analysis because the ProMises model permits incorporation of topological brain information in the alignment's estimation process.

Key words: functional alignment, functional magnetic resonance imaging, high-dimensional data, Procrustes analysis, Von Mises–Fisher distribution.

The Procrustes problem is aimed at matching matrices using similarity transformations by minimizing their Frobenius distance. It allows comparison of matrices with dimensions defined in an arbitrary coordinate system. This method raised the interest of applied researchers hence highlighting its potentiality through a plethora of applications in several fields, such as ecology (Saito et al., 2015), biology (Rohlf & Slice, 1990), analytical chemometrics (Andrade et al., 2004), and psychometrics (Green, 1952; McCrae et al., 1996).

The interest of a large audience from applied fields stimulates, in parallel, the growth of a vast body of literature. Despite this, essentially all applications comprise spatial coordinates (i.e., two- or three-dimensional). Haxby et al. (2011) first introduced the use of this approach into a different context: align functional Magnetic Resonance Images (fMRI). The coordinates are hence substituted by voxels (i.e., three-dimensional pixels), and the problem becomes inherently high-dimensional. The approach rapidly grew in popularity in the neuroimaging community because of its effectiveness. However, the proposed solution is naive; the extension from the spatial context to a more general and high-dimensional one is a theoretical challenge that needs adequate attention.

The most serious concern is the results' interpretability. In most cases, Procrustes methods turn into an ill-posed problem. It is a barely noticeable problem with spatial coordinates because the solution is unique up to rotations; hence, the user has the freedom to choose the point of view that provides the nicest picture. When the dimensions do not have a spatial meaning, any rotation completely changes the interpretation of the results.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-022-09859-5.

Correspondence should be made to Livio Finos, Department of Developmental Psychology and Socialization, University of Padova, Via Venezia, 8, Padua, Italy. Email: livio.finos@unipd.it

© 2022 The Author(s)

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

To tackle this problem, we revise the perturbation model (Goodal, 1991), which rephrases the Procrustes problem as a statistical model. The matrices are defined as a random perturbation of a reference matrix plus an error term. The perturbation is expressed by rotation, scaling, and translation, and the matrix normal distribution (Gupta & Nagar, 2018) is assumed for the error terms. Like Green and Mardia (2013) and Mardia et al. (2006), we assume that the orthogonal matrix parameter follows the von Mises–Fisher distribution. We prove that the proposed prior distribution is conjugate, making the estimation process quite fast. Indeed, the maximum a posterior estimate of the orthogonal matrix parameters results in a minor modification of the original solution because the prior information enters into the pairwise cross-product of the matrices to be aligned. The prior distribution plays the role of regularizing term, resolving the non-identifiability of the orthogonal matrix parameter. In the application to fMRI data in Sect. 4, we further show that specification of a prior distribution permits the integration of functional and topological aspects, which largely improves the results' interpretability. We then propose a comprehensive approach to the Procrustes problem: the ProMises (Procrustes von Mises–Fisher) model.

The second problem raised by the extension to the high-dimensional framework is computational. The estimation algorithm of a Procrustes-based method involves a series of singular value decompositions of $m \times m$ matrices (m dimensions). In a typical fMRI data set, the subjects (i.e., the matrices to be aligned) have a few hundred (observations/rows) n and hundreds of thousands of voxels (dimensions/columns) m. We prove that the minimization problem can be solved by a series of singular value decompositions of $n \times n$ matrices, reducing the computation burden and making the ProMises model applicable to matrices with virtually any number of columns. We denote this approach as the Efficient ProMises model.

We emphasize here that the problem of aligning fMRI data is not three-dimensional as it could appear at first glance, although it is high-dimensional: Each voxel is one dimension of the problem. Nevertheless, in such conditions, three critical issues arise: The first is that the Procrustes method combines any voxel inside the brain without distinguishing between adjacent and distant anatomical locations. This can be questionable because the voxels have a spatial organization, and, despite the inter-subject variability, we expect some degree of spatial similarity between subjects in their functional organization. Therefore, we want the subjects' voxels of a given location to be more likely to contribute to the construction of voxels with the same location in the common space. The second issue revolves around the non-identifiability of orthogonal transformations $\boldsymbol{R}_i$, where i indexes the subjects. The Procrustes method does not return a unique solution of the maximum likelihood estimate for $\boldsymbol{R}_i$: Given a solution (i.e., a three-dimensional image), any linear combination that mixes the voxels' values is an equivalent solution to the problem. The solutions are equivalent only from a mathematical point of view because the practical consequence is the loss of results' topological interpretability. The third issue is the computational load: applying the Procrustes-based alignment to the whole brain implies the decomposition of many square matrices of dimensions roughly equal to 200.000 (i.e., the number of voxels.)

The Efficient ProMises model resolves all these three issues. The use of a properly chosen prior shrinks the estimate to the anatomical solution (i.e., no rotation), hence making the solution unique and interpretable from an anatomical point of view. Finally, as mentioned before, the Efficient implementation permits performing functional alignment on high-dimensional data such as fMRI data.

The paper is organized as follows. Section 1 introduces the perturbation model (Goodall, 1991), stressing its critical issues. Section 2 illustrates the ProMises model and its challenges: identifiability, interpretability, and flexibility. Section 3 defines the Efficient version of the ProMises model, which permits application of the functional alignment to high-dimensional data. Finally, the presented model is evaluated by analyzing task-related fMRI data in Sect. 4. The entire code used is available in https://github.com/angeella/ProMisesModel using the programming language Python (Van Rossum and Drake Jr, 1995) and in https://github.com/angeella/alignProMises using

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

the R (R Core Team, 2018) package `alignProMises`. We report the proofs of the main lemmas here, whereas the remaining proofs are included in the supplementary material.

1. Perturbation Model

1.1. Background

Let $\{\boldsymbol{X}_i \in \mathbb{R}^{n\times m}\}_{i=1,\dots,N}$ be a set of matrices to be aligned. The Procrustes-based method uses similarity transformations to match each matrix to the target one as closely as possible, according to the Frobenius distance.

Each matrix $\boldsymbol{X}_i$ could be then assumed to be a similarity transformation of a shared matrix $\boldsymbol{M} \in \mathbb{R}^{n\times m}$, which contains the common reference space's coordinates, plus a random error matrix $\boldsymbol{E}_i \in \mathbb{R}^{n\times m}$. The perturbation model proposed by Goodall (1991) is then reported.

Definition 1. (*Goodall*, 1991) Let $\{\boldsymbol{X}_i \in \mathbb{R}^{n\times m}\}_{i=1,\dots,N}$ be a set of matrices to be aligned and $\mathcal{O}(m)$ the orthogonal group in dimension m. The perturbation model is defined as

$$\boldsymbol{X}_i = \alpha_i(\boldsymbol{M} + \boldsymbol{E}_i)\boldsymbol{R}_i^\top + \mathbf{1}_n^\top \boldsymbol{t}_i \qquad \text{subject to} \quad \boldsymbol{R}_i \in \mathcal{O}(m),$$

where $\boldsymbol{E}_i \sim \mathcal{MN}_{n,m}(0, \boldsymbol{\Sigma}_n, \boldsymbol{\Sigma}_m)$ —i.e., the matrix normal distribution with $\boldsymbol{\Sigma}_n \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma}_m \in \mathbb{R}^{m\times m}$ scale parameters— $\boldsymbol{M} \in \mathbb{R}^{n\times m}$ is the shared matrix, $\alpha_i \in \mathbb{R}^+$ is the isotropic scaling, $\boldsymbol{t}_i \in \mathbb{R}^{1\times m}$ defines the translation vector, and $\mathbf{1}_n \in \mathbb{R}^{1\times n}$ is a vector of ones.

To simplify the problem, the column-centered $\boldsymbol{X}_i$ are considered. The distribution is

$$\mathrm{vec}(C_n X_i | R_i, \alpha_i, M, \Sigma_m, \Sigma_n) \sim \mathcal{N}_{nm}(\mathrm{vec}(\alpha_i C_n M R_i^\top), R_i \Sigma_m R_i^\top \otimes \alpha_i^2 C_n \Sigma_n C_n^\top),$$

where $\boldsymbol{C}_n = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n$, $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix, $\boldsymbol{J}_n$ is a $n \times n$ matrix of ones, and $\mathrm{vec}(\boldsymbol{A})$ the vectorization of the matrix $\boldsymbol{A}$. Let the singular value decomposition of $\boldsymbol{C}_n = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^\top$, where $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times(n-1)}$, then:

$$\begin{aligned}\mathrm{vec}(\Gamma^\top C_n X_i | R_i, \alpha_i, M, \Sigma_m, \Sigma_n) \sim &\mathcal{N}_{(n-1)m}(\mathrm{vec}(\alpha_i \Gamma^\top C_n M R_i^\top),\\ &R_i \Sigma_m R_i^\top \otimes \alpha_i^2 \Gamma^\top C_n \Sigma_n C_n^\top \Gamma).\end{aligned}$$

The $\boldsymbol{\Gamma}^\top$ transformation leads to independence between the rows of $\boldsymbol{E}_i$ without affecting the estimation of $\boldsymbol{R}_i$. $\boldsymbol{\Gamma}^\top\boldsymbol{C}_n\boldsymbol{X}_i$ and $\boldsymbol{\Gamma}^\top\boldsymbol{C}_n\boldsymbol{M}$ have now $n-1$ rows; however, we can simply re-project them on $\mathbb{R}^{n\times m}$ using $\boldsymbol{\Gamma}$. Because we can always write $\tilde{\boldsymbol{X}}_i = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top\boldsymbol{C}_n\boldsymbol{X}_i$, $\tilde{\boldsymbol{M}} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top\boldsymbol{C}_n\boldsymbol{M}$, and $\tilde{\boldsymbol{\Sigma}}_n = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top\boldsymbol{C}_n\boldsymbol{\Sigma}_n\boldsymbol{C}_n^\top\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ without loss of generality, we can re-write Definition 1 as follows:

Definition 2.

$$\boldsymbol{X}_i = \alpha_i(\boldsymbol{M} + \boldsymbol{E}_i)\boldsymbol{R}_i^\top \qquad \text{subject to} \quad \boldsymbol{R}_i \in \mathcal{O}(m), \tag{1}$$

where $\boldsymbol{E}_i \sim \mathcal{MN}_{nm}(0, \boldsymbol{\Sigma}_n, \boldsymbol{\Sigma}_m)$. This way, we obtain the following:

$$\mathrm{vec}(\boldsymbol{X}_i | \boldsymbol{R}_i, \alpha_i, \boldsymbol{M}, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_n) \sim \mathcal{N}_{nm}(\mathrm{vec}(\alpha_i \boldsymbol{M}\boldsymbol{R}_i^\top), R_i \Sigma_m R_i^\top \otimes \alpha_i^2 \Sigma_n). \tag{2}$$

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

The following notation is also adopted: $||\cdot||$ to indicate the Frobenius norm and $<\cdot,\cdot>$ for the Frobenius inner product (Golub & van Loan, 2013).

The main objective of this work is comparing the shapes $\boldsymbol{X}_i$ instead of the form's analysis of the matrices. For that, the parameters of interest are $\boldsymbol{R}_i$ and α_i, whereas $\boldsymbol{M}$, $\boldsymbol{\Sigma}_n$, and $\boldsymbol{\Sigma}_m$ are considered as nuisance parameters for each $i = 1, \ldots, N$. The estimation of the unknown parameters changes if these nuisance parameters are known. Section 1.2 initially presents the estimates under this assumption and then provides the more realistic case of unknown nuisance parameters.

1.2. Estimation of the Perturbation Model

We formalize some results from Theobald and Wuttke (2006) in the case of known nuisance parameters $\boldsymbol{M}$, $\boldsymbol{\Sigma}_m$, and $\boldsymbol{\Sigma}_n$, with $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\Sigma}_m$ positive definite matrices by the following theorem:

Theorem 1. (Theobald & Wuttke 2006) *Consider the perturbation model described in Definition 2, and the singular value decomposition* $\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^\top$. *The maximum likelihood estimators equal* $\hat{\boldsymbol{R}}_i = \boldsymbol{U}_i \boldsymbol{V}_i^\top$, *and* $\hat{\alpha_i}_{\hat{\boldsymbol{R}}_i} = ||\boldsymbol{\Sigma}_m^{-1/2} \hat{\boldsymbol{R}}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma_n}^{-1/2}||^2 / tr(\boldsymbol{D}_i)$.

Now consider N independent observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$. The joint log-likelihood is simply the sum of N log-likelihoods.

In the case of unknown nuisance parameters $\boldsymbol{M}$, $\boldsymbol{\Sigma}_m$, and $\boldsymbol{\Sigma}_n$, the joint likelihood cannot be written as product of separated likelihoods, one for each $\boldsymbol{X}_i$, because each of the unknown parameters is a function of the others. The solution must be found by an iterative algorithm. In particular, the two covariance matrices $\boldsymbol{\Sigma}_m$ and $\boldsymbol{\Sigma}_n$ can be estimated by a two-stage algorithm defined in Dutilleul (1999), where $\hat{\boldsymbol{\Sigma}}_n = \{\sum_{i=1}^N (\boldsymbol{X}_i - \hat{\boldsymbol{M}}) \hat{\boldsymbol{\Sigma}}_m^{-1} (\boldsymbol{X}_i - \hat{\boldsymbol{M}})^\top\}/Nm$ and $\hat{\boldsymbol{\Sigma}}_m = \{\sum_{i=1}^N (\boldsymbol{X}_i - \hat{\boldsymbol{M}})^\top \hat{\boldsymbol{\Sigma}}_n^{-1} (\boldsymbol{X}_i - \hat{\boldsymbol{M}})\}/Nn$ are maximum likelihood estimators.

The necessary and sufficient condition for the existence of $\hat{\boldsymbol{\Sigma}}_n$ and $\hat{\boldsymbol{\Sigma}}_m$ is $N \geq \frac{m}{n} + 1$, assuming $\boldsymbol{\Sigma}_m$ and $\boldsymbol{\Sigma}_n$ are positive definite matrices. In real applications, this assumption could be problematic. For example, in fMRI data analysis, m roughly equals 200, 000, and n approximately equals 200; therefore, the researcher would have to analyzing at least 1, 001 subjects, which is virtually impossible because fMRI is costly.

Various solutions can be found in the literature: Theobald and Wuttke (2006) proposed a regularization for the covariance matrix, whereas Lele (1993) estimated $\boldsymbol{\Sigma}_n$ using the distribution of $\boldsymbol{X}_i \boldsymbol{X}_i^\top$. In this work, we maintain a general formulation of the estimator for $\hat{\boldsymbol{\Sigma}}_m = g(\hat{\boldsymbol{\Sigma}}_n, \hat{\boldsymbol{M}}, \boldsymbol{X}_i)$ and $\hat{\boldsymbol{\Sigma}}_n = g(\hat{\boldsymbol{\Sigma}}_m, \hat{\boldsymbol{M}}, \boldsymbol{X}_i)$, because we aim to find a proper estimate of $\boldsymbol{R}_i$ rather than $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\Sigma}_m$. The shared matrix $\boldsymbol{M}$ is estimated by the element-wise arithmetic mean of $\{\hat{\boldsymbol{X}}_i\}_{i=1,\ldots,N}$, where $\hat{\boldsymbol{X}}_i = \hat{\alpha}_{\hat{\boldsymbol{R}}_i}^{-1} \boldsymbol{X}_i \hat{\boldsymbol{R}}_i$. We then modified the iterative algorithm of Gower (1975) (i.e., generalized Procrustes analysis), to estimate $\boldsymbol{R}_i$.

Groisser (2005) proved the convergence of the generalized Procrustes analysis algorithm. However, it leads to non-identifiable estimators of $\boldsymbol{R}_i$, $i = 1, \ldots, N$, proved by the following lemma:

Lemma 1. *Let* $\{\hat{\boldsymbol{R}}_i\}_{i=1,\ldots,N}$ *be the maximum likelihood solutions for* $\{\boldsymbol{R}_i\}_{i=1,\ldots,N}$ *with* $\boldsymbol{M}$, $\boldsymbol{\Sigma}_n$, *and* $\boldsymbol{\Sigma}_m$ *as unknown parameters. If* $\boldsymbol{Z} \in \mathcal{O}(m)$, *then* $\{\hat{\boldsymbol{R}}_i \boldsymbol{Z}\}_{i=1,\ldots,N}$ *are still valid maximum likelihood solutions for* $\{\hat{\boldsymbol{R}}_i\}_{i=1,\ldots,N}$.

Proof Consider $\boldsymbol{Z} \in \mathcal{O}(m)$, and

$$\frac{1}{\alpha_i} \boldsymbol{X}_i \boldsymbol{R}_i \boldsymbol{Z} - \boldsymbol{M}\boldsymbol{Z} = \boldsymbol{E}_i \boldsymbol{Z} \sim \mathcal{MN}(0, \boldsymbol{\Sigma}_n, \boldsymbol{Z}^\top \boldsymbol{\Sigma}_m \boldsymbol{Z}).$$

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

Algorithm 1

Generalized Procrustes analysis (Gower, 1975) with the two-stage algorithm from Dutilleul (1999), where $\mathtt{T}$ is the threshold for $||\hat{\boldsymbol{M}} - \hat{\boldsymbol{M}}_{old}||^2$, $\mathtt{maxIt}$ is the maximum number of iterations allowed, and ϵ_1 and ϵ_2 denote two infinitesimal positive quantities.

Require: $\boldsymbol{X}_i, \mathtt{T}, \mathtt{maxIt}, \forall i = 1, \ldots, N$
Ensure: $\hat{\boldsymbol{X}}_i \ \forall i = 1, \ldots, N$
$\hat{\boldsymbol{M}} = \sum_{i=1}^{N} \boldsymbol{X}_i / N, \hat{\alpha}_i = 1, \hat{\boldsymbol{\Sigma}}_{n;\,\mathrm{old}} = \hat{\boldsymbol{\Sigma}}_n = \boldsymbol{I}_n, \hat{\boldsymbol{\Sigma}}_{m;\,\mathrm{old}} = \hat{\boldsymbol{\Sigma}}_m = \boldsymbol{I}_m$
$\mathtt{count} = 0, \mathtt{dist} = \mathtt{Inf}$
while $\mathtt{dist} > \mathtt{T}$ OR $\mathtt{count} < \mathtt{maxIt}$ **do**
  for $i = 1$ to N **do**
    $\boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^\top = \mathtt{SVD}(\boldsymbol{X_i}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}{}^{-1} \hat{\boldsymbol{M}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}{}^{-1})$ ▷ Singular value decomposition
    $\hat{\boldsymbol{R}}_i = \boldsymbol{U}_i \boldsymbol{V}_i^\top$
    $\hat{\boldsymbol{X}}_i = \boldsymbol{X}_i \hat{\boldsymbol{R}}_i$
    $\hat{\alpha_i}_{\hat{\boldsymbol{R}}_i} = ||\hat{\boldsymbol{\Sigma}}_m^{-1/2} \hat{\boldsymbol{R}}_i^\top \boldsymbol{X}_i^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}{}^{-1/2}||^2 / tr(\boldsymbol{D}_i)$
    $\hat{\boldsymbol{X}}_i = \hat{\alpha_i}_{\hat{\boldsymbol{R}}_i}^{-1} \hat{\boldsymbol{X}}_i$ ▷ Update $\boldsymbol{X}_i$
  end for
  $\hat{\boldsymbol{M}}_{\mathrm{old}} = \hat{\boldsymbol{M}}$ ▷ Save $\hat{\boldsymbol{M}}$
  $\hat{\boldsymbol{M}} = \sum_{i=1}^{N} \hat{\boldsymbol{X}}_i / N$ ▷ Update $\hat{\boldsymbol{M}}$
  $\hat{\boldsymbol{\Sigma}}_n = g(\hat{\boldsymbol{\Sigma}}_m, \hat{\boldsymbol{M}}, \boldsymbol{X}_i), \hat{\boldsymbol{\Sigma}}_m = g(\hat{\boldsymbol{\Sigma}}_n, \hat{\boldsymbol{M}}, \boldsymbol{X}_i)$
  while $||\hat{\boldsymbol{\Sigma}}_n - \hat{\boldsymbol{\Sigma}}_{n;\,\mathrm{old}}||^2 > \epsilon_1$ OR $||\hat{\boldsymbol{\Sigma}}_m - \hat{\boldsymbol{\Sigma}}_{m;\,\mathrm{old}}||^2 > \epsilon_2$ **do**
    $\hat{\boldsymbol{\Sigma}}_{n;\,\mathrm{old}} = \hat{\boldsymbol{\Sigma}}_n, \hat{\boldsymbol{\Sigma}}_{m;\,\mathrm{old}} = \hat{\boldsymbol{\Sigma}}_m$
    $\hat{\boldsymbol{\Sigma}}_n = g(\hat{\boldsymbol{\Sigma}}_m, \hat{\boldsymbol{M}}, \boldsymbol{X}_i), \hat{\boldsymbol{\Sigma}}_m = g(\hat{\boldsymbol{\Sigma}}_n, \hat{\boldsymbol{M}}, \boldsymbol{X}_i)$
  end while
  $\mathtt{dist} = ||\hat{\boldsymbol{M}} - \hat{\boldsymbol{M}}_{\mathrm{old}}||^2, \mathtt{count} = \mathtt{count} + 1.$
end while

Consider the proof of Theorem 1 placed in the supplementary material, we have

$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} \sum_{i=1}^{N} < \boldsymbol{R}_i, \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} >= \max_{\boldsymbol{R}_i \in \mathcal{O}(m)} \sum_{i=1}^{N} tr(\boldsymbol{Z}^\top \boldsymbol{R}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{Z} \boldsymbol{Z}^\top \boldsymbol{\Sigma}_m^{-1} \boldsymbol{Z}). \quad (3)$$

Since $\boldsymbol{Z}^\top \boldsymbol{Z} = \boldsymbol{Z}\boldsymbol{Z}^\top = \boldsymbol{I}_m$, the solutions $\{\hat{\boldsymbol{R}}_i \boldsymbol{Z}\}_{i=1,\ldots,N}$ are still valid solutions for the maximization (3). □

To sum up, in the more realistic case, when we must estimate the nuisance parameters, the Procrustes solutions are infinite in general, in both in the high-dimensional case and the low-dimensional by Lemma 1. We emphasize here that, to resolve the non-identifiability of $\boldsymbol{R}_i$, the proposed ProMises model imposes a prior distribution for the parameter $\boldsymbol{R}_i$.

2. ProMises Model

2.1. Background

In the previous section, we justified how the perturbation model could be problematic because Lemma 1 proves the non-identifiability of the parameter $\boldsymbol{R}_i$ in the realistic case of unknown nuisance parameters. This scenario returns to be critical in several applications, where the m columns of the matrix $\boldsymbol{X}_i$ do not express the three-dimensional spatial coordinates, such as in the

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

fMRI data framework illustrated in Sect. 4. In the high-dimensional case, the final orientations of the aligned data can be relevant for interpretation purposes.

For that, we propose its Bayesian approach: the ProMises model. We stress here that a proper prior distribution leads to a closed-form and interpretable, unique point estimate of $\boldsymbol{R}_i$. The specification of the prior parameters is essential, especially in our high-dimensional context.

2.2. *Interpretation of the Prior Parameters*

Because $\boldsymbol{R}_i \in \mathcal{O}(m)$, a proper prior distribution must take values in the Stiefel manifold $V_m(\mathbb{R}^m)$. The matrix von Mises–Fisher distribution is a non-uniform distribution on $V_m(\mathbb{R}^m)$, which describes a rigid configuration of m distinct directions with fixed angles. It was proposed by Downs (1972) and investigated by many authors (e.g., Chikuse, 1979; Jupp & Mardia 2003).

We report below the formal definition of the von Mises–Fisher distribution.

Definition 3. (*Downs*, 1972) The von Mises–Fisher distribution for $\boldsymbol{R}_i \in \mathcal{O}(m)$ is

$$f(\boldsymbol{R}_i) = C(\boldsymbol{F}, k) \exp\left\{tr(k\boldsymbol{F}^\top \boldsymbol{R}_i)\right\}, \tag{4}$$

where $C(\boldsymbol{F}, k)$ is a normalizing constant, $\boldsymbol{F} \in \mathbb{R}^{m\times m}$ is the location matrix parameter, and $k \in \mathbb{R}^+$ is the concentration parameter.

The parameter k defined in (4) balances the amount of concentration of the distribution around $\boldsymbol{F}$. If $k \to 0$, the prior distribution is near a uniform distribution (i.e., unconstrained). If $k \to +\infty$, the prior distribution tends toward a Dirac distribution (i.e., maximum constraint).

A proper specification of the prior distribution leads to improved estimation of $\boldsymbol{R}_i$. Therefore, the core of the ProMises model is the specification of $\boldsymbol{F}$ defined in (4). Consider the polar decomposition and singular value decomposition of $\boldsymbol{F} = \boldsymbol{P}\boldsymbol{K} = \boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{B}^\top = \boldsymbol{L}\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top$, where $\boldsymbol{P}, \boldsymbol{L}, \boldsymbol{B} \in \mathcal{O}(m)$, and $\boldsymbol{K} \in \mathbb{R}^{m\times m}$ are symmetric positive semi-definite matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ diagonal matrix with non-negative real numbers on the diagonal. The mode of the density defined in (4) equals $\boldsymbol{P}$ (Jupp & Mardia, 1979), so the most plausible rotation matrix depends on the orientation characteristic of $\boldsymbol{F}$. Merging the two decompositions, $\boldsymbol{P} = \boldsymbol{L}\boldsymbol{B}^\top$ describes the orientation part of $\boldsymbol{F}$, and $\boldsymbol{K} = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top$ defines the concentration part. The mode is specified by the product of the left and right singular vectors of $\boldsymbol{F}$. These decompositions are useful to understand when the density (4) is uni-modal. If $\boldsymbol{F}$ has full rank, $\boldsymbol{\Sigma}$ does too, the polar decomposition is unique, and thus the mode of the density (i.e., $\boldsymbol{P}$ is the global maximum). Let $\boldsymbol{F}$ be a full rank matrix, then the maximum equals $\max_{\boldsymbol{R}_i\in\mathcal{O}(m)} tr(\boldsymbol{F}\boldsymbol{R}_i^\top) = tr\{\boldsymbol{L}\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top(\boldsymbol{L}\boldsymbol{B}^\top)^\top\} = tr(\boldsymbol{\Sigma})$.

To sum up, the prior specification allows us to include a priori information about the optimal orientation in the perturbation model. We anticipate here the result of Lemma 3: If $\boldsymbol{F}$ is defined as a full rank matrix, the maximum a posteriori solution $\hat{\boldsymbol{R}}'_i$ will be unique with the orientation structure of $\boldsymbol{F}$.

Two simple examples of $\boldsymbol{F}$ are delineated below.

Example 1. The most simple definition of $\boldsymbol{F}$ is $\boldsymbol{I}_m$ (Lee, 2018). The eigenvalues are all 1, and $\boldsymbol{L}$ and $\boldsymbol{B}$ are equal to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$, where $\boldsymbol{e}_i$ is the standard basis forming an orthonormal basis of $\mathbb{R}^m$. The prior distribution shrinks the possible solutions for $\boldsymbol{R}_i$ toward orthogonal matrices that consider only the combination of variables with the same location.

Alternatively, considering the fMRI scenario, the hyperparameter $\boldsymbol{F}$ can be defined as an Euclidean similarity matrix using the $3D$ anatomical coordinates x, y, and z of each voxel:

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

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

where $i, j = 1, \ldots m$. In this way, $\boldsymbol{F}$ is a symmetric matrix with ones in the diagonal, which means that voxels with the same spatial location are combined with weights equalling 1, and the weights decrease as the voxels to be combined are more spatially distant.

Example 2. Consider N matrices, one for each plant, describing the three-dimensional spatial trajectories of a climbing plant, Pisum sativum, having wooden support as a stimulus (Guerra et al., 2019) across time. The spatiotemporal trajectories of the plants are analyzed until they come to grasp the stick. The aim is to functionally align the three time series, one for each coordinate (x, y, z); then, we have $m = 3$. In this case, we could suppose that the rotation along the z axis is not of interest because the functional misalignment between plants can be along the x-axis and y-axis with the z-axis being the one that reflects the growth of the plants and the x-axis and y-axis describing the elliptical movement (circumnutation) of the plants (Guerra et al., 2019). Therefore, the $\boldsymbol{F}$ location matrix parameter can be described as follows:

$$\boldsymbol{F} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

The axes x and y have the same probability of entering in the calculation of the first two dimensions of the final common space, whereas the z axis is not considered.

We use the kinematic plant data from Guerra et al. (2019), consisting of five matrices/plants. Figure 1 shows the elliptical movement expressed by the axes x and y, in the case of unaligned and aligned plant trajectories. The rotation transformations are estimated by the ProMises model with $\boldsymbol{F}$ expressed as (5). We do not go into detail about the meaning of the results because we have introduced this example to explain the usefulness of $\boldsymbol{F}$. However, we can note how the ProMises model aligns the final coordinates of the tendrils (i.e., when the plant touches the wooden support).

2.3. Von Mises–Fisher Conjugate Prior

The von Mises–Fisher distribution (4) was proved by Khatri and Mardia (1977) to be a member of the standard exponential family (Barndorff–Nielsen, 2014). Green and Mardia (2006) mentioned that the von Mises–Fisher distribution is a conjugate prior for the matrix normal distribution, which we formally prove in the following lemma under the perturbation model's assumptions.

Lemma 2. *Consider the perturbation model of Definition 2, with $\boldsymbol{R}_i$ distributed according to (4), then the posterior distribution $f(\boldsymbol{R}_i|k, \boldsymbol{F}, \boldsymbol{X}_i)$ is a conjugate distribution to the von Mises–Fisher prior distribution with location posterior parameter equalling the following:*

$$\boldsymbol{F}^{\star} = \boldsymbol{X_i}^{\top} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} + k\boldsymbol{F}. \tag{6}$$

The posterior location parameter is the sum of $\boldsymbol{X}_i^{\top} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1}$ and the prior location parameter $\boldsymbol{F}$ multiplied by k. Consider the singular value decomposition of $\boldsymbol{X}_i^{\top} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1}$:

$$\boldsymbol{X}_i^{\top} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^{\top} = \boldsymbol{U}_i \boldsymbol{V}_i^{\top} \boldsymbol{V}_i \boldsymbol{D}_i \boldsymbol{V}_i^{\top}. \tag{7}$$

The right part of (7) $\boldsymbol{V}_i \boldsymbol{D}_i \boldsymbol{V}_i^{\top}$ is the elliptical part of $\boldsymbol{X}_i^{\top} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1}$, which is a measure of variation relative to the decomposition $\boldsymbol{U}_i \boldsymbol{V}_i^{\top}$ (i.e., the maximum likelihood estimator of $\boldsymbol{R}_i$).

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

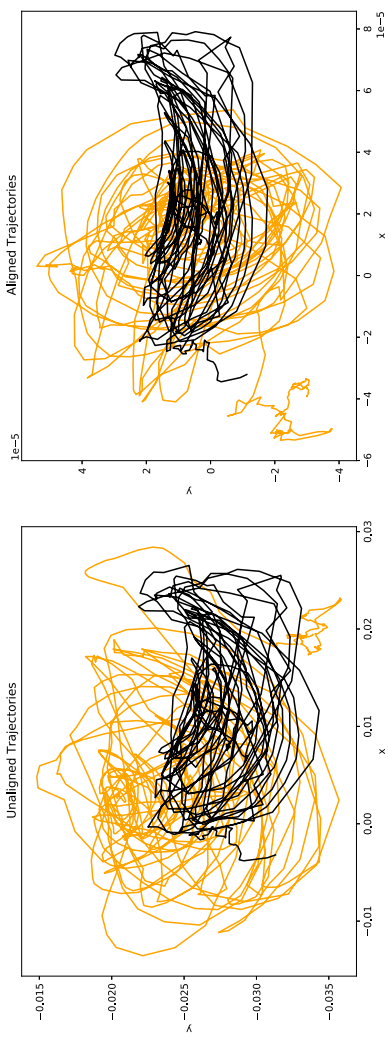


FIGURE 1.
Left panel: Unaligned spatial trajectories of the tendrils of two plants. Right panel: Aligned spatial trajectories of the tendrils of two plants.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

Focus on the right part of (6), $\boldsymbol{F} = \boldsymbol{P}\boldsymbol{K}$, which is the polar decomposition of $\boldsymbol{F}$, where $\boldsymbol{P}$ is the mode of the von Mises–Fisher distribution, and $\boldsymbol{K}$ is its measure of variation. Therefore, $\boldsymbol{F}^\star$ is expressed as a combination of the maximum likelihood estimate $\hat{\boldsymbol{R}}_i = \boldsymbol{U}_i \boldsymbol{V}_i^\top$ and the prior mode $\boldsymbol{P}$, multiplied by corresponding measures of variation.

Thanks to the conjugacy, the estimation process remains simple; with a small modification we keep the previous algorithm without increasing the computational burden.

2.4. Estimation of the ProMises Model

This section delineates the estimation process for $\boldsymbol{R}_i$ using the ProMises method. First of all, $f(\boldsymbol{X_i}|\alpha_i, \boldsymbol{R_i})$ depends only on the product $\alpha_i \boldsymbol{R_i}$, we thus refer to the distribution $f(\boldsymbol{X_i}|\alpha_i \boldsymbol{R_i})$ instead of $f(\boldsymbol{X_i}|\alpha_i, \boldsymbol{R_i})$ defined in (2). The following density is then considered as prior distribution for the product $\alpha_i \boldsymbol{R_i}$:

$$f(\alpha_i \boldsymbol{R_i}) \sim \exp\left\{\frac{k}{\alpha_i} tr(\boldsymbol{F}^\top \boldsymbol{R}_i)\right\} \alpha_i^{-1}. \tag{8}$$

The following theorem delineates the estimation of $\boldsymbol{R}_i$ with known nuisance parameters:

Theorem 2. *The ProMises model is defined as the perturbation model specified in Definition 2 imposing the prior distribution (8) for $\alpha_i \boldsymbol{R}_i$. 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 maximum a posteriori estimators equal $\hat{\boldsymbol{R}}'_i = \boldsymbol{U}_i \boldsymbol{V}_i^\top$ and $\hat{\alpha_i}'_{\hat{\boldsymbol{R}}'_i} = ||\boldsymbol{\Sigma}_m^{-1/2} \hat{\boldsymbol{R}}'^\top_i \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1/2}||^2 / tr(\boldsymbol{D}_i)$.*

The prior information about $\boldsymbol{R}_i$'s structure is directly entered in the singular value decomposition step; the maximum a posteriori estimator turns out to be a slight modification of the solution given in Theorem 1. We decompose $\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} + k\boldsymbol{F}$ instead of $\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1}$.

Let $\{\boldsymbol{X}_i \in \mathbb{R}^{n \times m}\}_{i=1,\ldots,N}$ be a set of independent matrices. Then, the joint posterior distribution is simply the product of the single posterior distribution.

If $\boldsymbol{M}$, $\boldsymbol{\Sigma}_n$, and $\boldsymbol{\Sigma}_m$ are unknown, the maximization problem has no closed-form solution, like in Sect. 1.2. Because we proved that the prior specification modifies only the singular value decomposition step of Theorem 1, we then modify the Line 5 of Algorithm 1, as follows:

Algorithm 2
ProMises model.

Use Algorithm 1, only modify:

5: $\boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^\top = \texttt{SVD}(\boldsymbol{X_i}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}^{-1} \hat{\boldsymbol{M}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}^{-1} + k\boldsymbol{F})$ ▷ Singular value decomposition

---

2.5. On the Choice of the Parameter of the von Mises–Fisher Distribution

We choose the von Mises–Fisher distribution as prior distribution for $\boldsymbol{R}_i$ for its useful and practical properties. First, as shown, it is a conjugate prior distribution, leading to a direct calculation and interpretation of $\hat{\boldsymbol{R}}'_i$. Second, it expresses the orthogonality constraint imposed by the Procrustes problem. Finally, the definition of $\boldsymbol{F}$ does not require strong assumptions (Downs, 1972); nevertheless, if we specify it as a full-rank matrix, we guarantee solution's uniqueness. This permits formulation of the below lemma.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

Lemma 3. *If $\boldsymbol{F}$ has full rank, the maximum a posteriori estimates for $\boldsymbol{R}_i$ given by Theorem 2 are unique.*

Proof Consider the proof of Lemma 1. Multiplying by $\boldsymbol{Z}$ leads to the following maximization:

$$\sum_{i=1}^{N} tr(\boldsymbol{Z}^\top \boldsymbol{R}_i^\top (\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{Z} \boldsymbol{Z}^\top \boldsymbol{\Sigma}_m^{-1} \boldsymbol{Z} + k\boldsymbol{F})) \neq \sum_{i=1}^{N} tr(\boldsymbol{R}_i^\top (\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} + k\boldsymbol{F}))$$

since the cyclic permutation invariance property of the trace does not work as Lemma 1 having the additional term $ktr(\boldsymbol{F}\boldsymbol{R}_i)$.

In addition, recalling Lemma 4, the solution for $\boldsymbol{R}_i$ is unique if and only if $\boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{M} \boldsymbol{\Sigma}_m^{-1} + k\boldsymbol{F}$ has full rank. If $\boldsymbol{F}$ is defined with full rank, so $\tilde{\boldsymbol{X}}_i^\top \tilde{\boldsymbol{M}} + k\boldsymbol{F}$, and the solution for R_i is unique. Furthermore, recalling Jupp and Mardia (1979), the mode of the von Mises–Fisher is the orientation part of location matrix parameter. Because the polar decomposition of $\boldsymbol{F}^\star$ is unique, the maximum a posteriori estimate is unique. □

To sum up, the ProMises model enables resolving the non-identifiability of $\boldsymbol{R}_i$ that characterizes the perturbation model. The prior information inserted in the model permits guidance of the estimation process, computing a unique and interpretable data orthogonal transformation. Finally, all these properties are reached without complicating the estimation process of the perturbation model; we only modify the singular value decomposition step of Algorithm 1.

3. Efficient ProMises Model

The framework depicted above can be applied both in low- and high-dimensional settings. However, the extension to the high-dimensional case does not come for free if the perturbation model is used. When $n < m$, rank equals n, the identifiability of the solution is lost even when the nuisance parameters are known. The lemma below formally states:

Lemma 4. *Consider $\boldsymbol{X}_i \in \mathbb{R}^{n \times m}$, if $n < m$, then the maximum likelihood estimate for $\boldsymbol{R}_i$ defined in Theorem 1 is not unique.*

Although the ProMises model provides unique solutions even in high-dimensional frameworks, a second issue remains prominent: the computational load. At each step, the presented algorithms perform N singular values decompositions of $m \times m$ matrices, which have a polynomial-time complexity $O(m^3)$. When m becomes large, as in fMRI data where m is a few hundred thousands, the computation runtime, and the required storing memory, becomes inadmissible.

This section proposes the Efficient ProMises model, which resolves the two above points. The method is efficient in terms of space and time complexity and fixes the non-identifiability of $\boldsymbol{R}_i$. The algorithm allows a faster and more accessible shape analysis without loss of information in the case of $n \ll m$. It essentially merges the thin singular value decomposition (Bai et al., 2000) with the Procrustes problem.

In practice, the Efficient ProMises approach projects the matrices $\boldsymbol{X}_i$ into an n-lower-dimensional space using a specific semi-orthogonal transformation (Abadir & Magnus, 2005; Groß et al., 1999) $\boldsymbol{Q}_i$, with dimensions $m \times n$, which preserve all the data's information. It aligns, then, the reduced $n \times n$ matrices $\{\boldsymbol{X}_i \boldsymbol{Q}_i \in \mathbb{R}^{n \times n}\}_{i=1,\dots,N}$ by the perturbation or ProMises model. Finally, it projects the aligned matrices back to the original $n \times m$-size matrices $\{\boldsymbol{X}_i \in \mathbb{R}^{n \times m}\}_{i=1,\dots,N}$ using the transpose of $\{\boldsymbol{Q}_i\}_{i=1,\dots,N}$.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

The following theorem proves that the maximum defined in Eq. (3) using $\{\boldsymbol{X}_i \boldsymbol{Q}_i \in \mathbb{R}^{n\times n}\}_{i=1,\ldots,N}$ equals the original maximum because the Procrustes problem analyzes the first $n \times n$ dimensions of $\boldsymbol{R}_i$. The maximum remains the same if we multiply $\{\boldsymbol{X}_i \in \mathbb{R}^{n\times m}\}_{i=1,\ldots,N}$ by $\boldsymbol{Q}_i$.

Theorem 3. *Consider the perturbation model in Definition 2 with* $\boldsymbol{\Sigma}_m = \sigma^2 \boldsymbol{I}_m$ *and the thin singular value decompositions of* $\boldsymbol{X}_i = \boldsymbol{L}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top$ *for each* $i = 1, \ldots, N$, *where* $\boldsymbol{Q}_i$ *has dimensions* $m \times n$. *The following holds*

$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} tr(\boldsymbol{R}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{X}_j \boldsymbol{\Sigma}_m^{-1}) = \max_{\boldsymbol{R}_i^\star \in \mathcal{O}(n)} tr(\boldsymbol{R}_i^{\star\top} \boldsymbol{Q}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{X}_j \boldsymbol{\Sigma}_m^{-1} \boldsymbol{Q}_j^\top).$$

Additionally, the condition for the existence of $\hat{\boldsymbol{\Sigma}}_n$ by Dutilleul (1999) is satisfied because $\boldsymbol{X}_i$ now has dimensions $n \times n$, and the functional alignment needs at least $N = 2$ observations.

So, Theorem 3 is used to define an Efficient version of the ProMises model.

Lemma 5. *Consider the assumptions of Theorem 3, then*

$$\max_{\boldsymbol{R}_i \in \mathcal{O}(m)} tr(\boldsymbol{R}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma_n}^{-1} \boldsymbol{X}_j \boldsymbol{\Sigma}_m^{-1} + k\boldsymbol{F}) = \max_{\boldsymbol{R}_i^\star \in \mathcal{O}(n)} tr\{\boldsymbol{R}_i^{\star\top} (\boldsymbol{Q}_i^\top \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{X}_j \boldsymbol{\Sigma}_m^{-1} \boldsymbol{Q}_j^\top + k\boldsymbol{F}^\star)\},$$

where $F \in \mathbb{R}^{m\times m}$ *and* $F^\star = Q_i^\top F Q_j \in \mathbb{R}^{n\times n}$.

Proofs of Theorem 3 and Lemma 5 are shown in the supplementary material, whereas here, we make some further considerations about the proposed method.

At first glance, the assumption $\boldsymbol{\Sigma}_m = \sigma^2 \boldsymbol{I}_m$ may dilute the resul's impact. However, this assumption does not imply that the data are column-wise independent because this dependence is modeled by $\boldsymbol{R}_i$. Additionally, the joint and accurate estimate of $\boldsymbol{\Sigma}_m$, $\boldsymbol{\Sigma}_n$ and $\boldsymbol{R}_i$ requires a large number of observations. So, it is common in real applications to set $\boldsymbol{\Sigma}_m$ and $\boldsymbol{\Sigma}_n$ to be proportional to the identities in Procrustes-like problems (Haxby et al., 2020). When the model is high-dimensional, this problem becomes even more pronounced because of the huge number of parameters to be estimated.

The Efficient ProMises 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 data set. Therefore, the original problem estimates orthogonal matrices of size $m \times m$: $\boldsymbol{R}_i \in \mathcal{O}(m)$, whereas the Efficient solution provides a set of orthogonal matrices of size $n \times n$: $\boldsymbol{R}_i^\star \in \mathcal{O}(n)$. Even when the solution is projected back into the $m \times m$ space through $\boldsymbol{Q}_i \boldsymbol{R}_i^\star \boldsymbol{Q}_i^\top$, the rank remains n, whereas the matrices of the original solutions have rank m. This should clarify that the Efficient approach reaches the same fit to the data under a different set of constraints that is $n \times n$ orthogonal matrices instead of $m \times m$ matrices; hence, the solutions of the two algorithms will not be identical.

Then, we add on Algorithm 1 the lines used to reduce the dimensions of $\boldsymbol{X}_i$, and we modify Line 5 to insert the prior information:

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

Algorithm 3
Efficient ProMises model.

Use Algorithm 1, only add these three lines at the beginning:

1: **for** $i = 1$ to N **do**
2: $\quad \boldsymbol{L}_i \boldsymbol{S}_i \boldsymbol{Q}_i^\top = \texttt{SVD}(\boldsymbol{X}_i)$ ▷ Thin singular value decomposition
3: $\quad \boldsymbol{X}_i := \boldsymbol{X}_i \boldsymbol{Q}_i$
4: **end for**

and modify:

5: $\boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^\top = \texttt{SVD}({\boldsymbol{X}_i}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}{}^{-1} \hat{\boldsymbol{M}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}{}^{-1} + k\boldsymbol{F})$ ▷ Singular value decomposition

The Efficient approach reduces the time complexity from $\mathcal{O}(m^3)$ to $\mathcal{O}(mn^2)$ and the space complexity from $\mathcal{O}(m^2)$ to $\mathcal{O}(mn)$.

4. Functional Magnetic Resonance Imaging Data Application

4.1. Motivation

The alignment problem is recognized in fMRI multi-subject studies because the brain's anatomical and functional structures vary across subjects. The most used anatomical alignments are the Talairach normalization (Talairach & Tournoux, 1988) and the Montréal Neurological Institute (MNI) space normalization (Jenkinson et al., 2002), where the brain images are aligned to an anatomical template by affine transformations using a set of major anatomical landmarks. However, this alignment does not explore the between-subjects variability in anatomical positions of the functional loci. The functional brain regions are not consistently placed on the anatomical landmarks defined by the Talairach and MNI templates. The anatomical alignment is then an approximate inter-subject registration of the functional cortical areas. Haxby et al. (2011) proved that functional brain anatomy exhibits a regular organization at a fine spatial scale shared across subjects.

Therefore, we can assume that anatomical and functional structures are subject-specific (Conroy et al., 2009; Sabuncu et al., 2010) and that the neural activities in different brains are noisy rotations of a common space (Haxby et al., 2011). Functional alignments (e.g., Procrustes methods) attempt to rotate the neural activities to maximize similarity across subjects.

Specifically, each subject's brain activation can be represented by a matrix, where the rows represent the stimuli/time points, and the columns represent the voxels. The stimuli are time-synchronized among subjects, so we have correspondence among the matrices' rows. However, the columns are not assumed to be in correspondence among subjects, as explained before. Each time series of brain activation (i.e., each of the matrices' columns) represents the voxels' functional characteristics that the anatomical normalization fails to align. We aim to represent the neural responses to stimuli into a common high-dimensional space, rather than in a canonical anatomical space that does not consider the variability of functional topographies loci.

Figure 2 shows three voxels' neural activities (i.e., v_1, v_2, and v_3) three columns of the data matrix in two subjects recorded across time. The functional pattern of v_2 is equal across subjects, whereas v_1 and v_3 are swapped. A rotation matrix can resolve this misalignment, with the swap being a particular case of the rotation matrix. For further details about the motivation in using functional Procrustes-based alignment in fMRI studies, see Haxby et al. (2020).

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

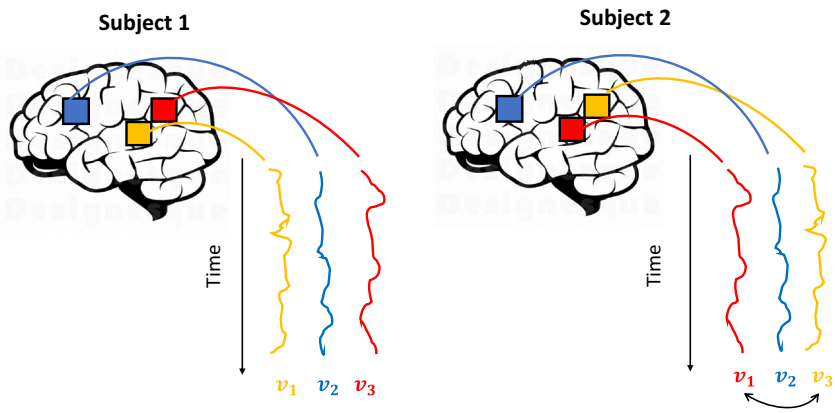


FIGURE 2.
Illustration of functional misalignment between fMRI images, where three voxels' time series are plotted considering two subjects. The time series of voxels v_1 and v_3 of the second subject are swapped with respect to the first subject.

4.2. Data Description

We apply the proposed method to data from Pernet et al. (2015), available at https://openneuro.org/datasets/ds000158/versions/1.0.0. The study consists of neural activations of 218 subjects passively listening to vocal (i.e., speech) and nonvocal sounds. Because the application has had a mere illustrative purpose, we choose to use a small number of subjects (18) to facilitate the example's reproducibility by the readers. We preprocessed the data using the Functional MRI of the Brain Software Library (FSL) (Jenkinson et al., 2012) using a standard processing procedure (i.e., high-pass filtering, brain extraction, spatially smoothing, registration to standard MNI space, dealing with motion and differences in slice acquisition time). Anatomical and functional alignment (based on the ProMises model) is compared, having images preprocessed in the same way, but in one case, the functional alignment is applied, while in the other case not. For details about the experimental design and data acquisition, please see Pernet et al. (2015).

4.3. Functional Connectivity

We performed region of interest and seed-based correlation analysis (Cordes et al., 2000). The seed-based correlation map shows the level of functional connectivity between a seed and every voxel in the brain, whereas the region of interest analysis expresses the functional correlation between predefined regions of interest coming from a standard atlas. The analysis process is defined as follows: First, the subject images are aligned using Algorithm 3, then the element-wise arithmetic mean across subjects is calculated, and finally, the functional connectivity analysis is developed on this average matrix.

We take the frontal pole as seed, being a region with functional diversity (Liu et al., 2013). The anatomical alignment considered here refers to the MNI space normalization (Jenkinson et al., 2002). Figure 3 shows the correlation values between the seed and each voxel in the brain using data without functional alignment (top of Fig. 3) and with functional alignment using the Efficient ProMises model (bottom of Fig. 3). The first evidence is that the functional alignment produces more interpretable maps, where the various regions, such as the superior temporal gyrus, are delineated by marked spatial edges, while the non-aligned map produces more spread regions, hence being less interpretable. It is interesting to evaluate the regions more correlated with the frontal pole, for example, the superior temporal gyrus. This region is associated with the processing

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

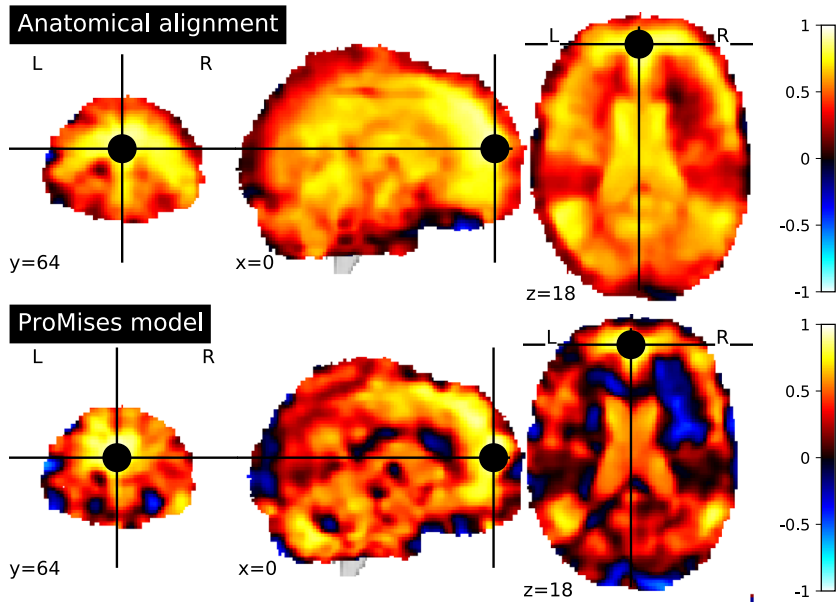


FIGURE 3.
Seed-based correlation map for $\boldsymbol{M}$, using data only aligned anatomically (top figure), and data also functionally aligned by the Efficient ProMises model (bottom figure). The black point refers to the seed used (i.e., frontal pole with MNI coordinates (0, 64, 18)). So, the brain map indicates the level of correlation between each voxel and the frontal pole.

of auditory stimuli. The correlation of the superior temporal gyrus with the seed is clear in the bottom part of Fig. 3, where functionally aligned images are used.

In contrast, the region of interest correlations analysis shows the integration mechanisms between specialized brain areas. Figure 4 indicates the correlation matrices of time-series extracted from the 39 main regions of the atlas of Varoquaux et al. (2011). Using functionally aligned data (right side of Fig. 4), we can see delineated blocks of synchronized regions that can be interpreted as large-scale functional networks. Instead, using data without functional alignment (left side of Fig. 4) the distinctions between blocks are clearly worse. Using functionally aligned data, the left and right visual systems, composed of the dorsolateral prefrontal cortex (DLPFC), frontal pole (Front pol), and parietal (Par), are clearly visible, whereas in the analysis using functionally nonaligned data, this distinction is hidden by noise.

The preprocessed data are available on the GitHub repository: http://github.com/angeella/fMRIdata, as well as the code used to perform functional connectivity: http://github.com/angeella/ProMisesModel/Code/Auditory.

5. Discussion

The ProMises model provides a methodologically grounded approach to the Procrustes problem allowing functional alignment on high-dimensional data in a computationally efficient way. The issues of the perturbation model (Goodall, 1991)—non-uniqueness, critical interpretation, and inapplicability when $n \ll m$—are completely surpassed thanks to our Bayesian extension. Indeed, the ProMises method returns unique and interpretable orthogonal transformations, and its efficient approach extends the applicability to high-dimensional data. The presented method is particularly useful in fMRI data analysis because it allows the functional alignment of images having roughly $200 \times 200{,}000$ dimensions, obtaining a unique representation of the aligned images in the brain space and a unique interpretation of the related results.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

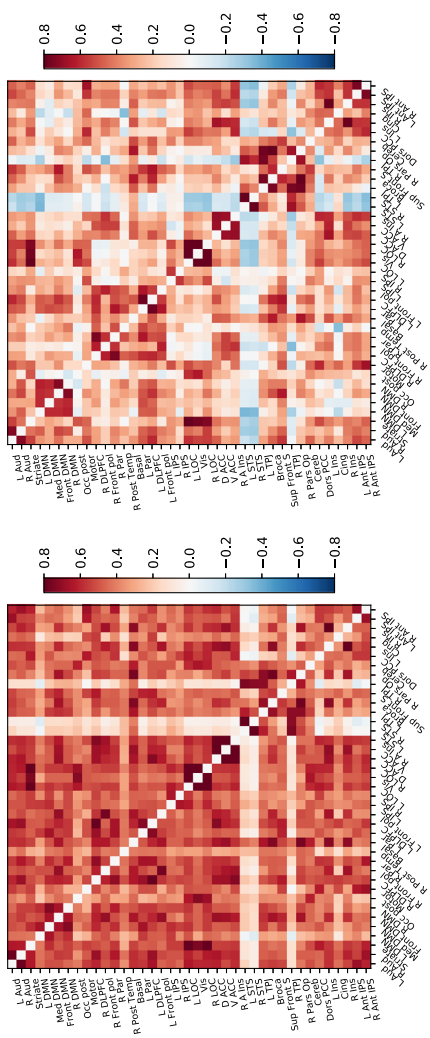


FIGURE 4.
Correlation matrix for $\boldsymbol{M}$, using data only aligned anatomically (left figure) and data also functionally aligned by the Efficient ProMises model (right figure). The cells of the matrix represent the correlation between the regions (represented by the row/column labels) of the Varoquaux et al. (2011)'s atlas.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

In the application example presented in Sect. 4, a subsample was analyzed. However, the algorithm has a linear growth in N, and therefore, it is not a problem to work with larger samples. Also, the algorithm permits a parallel computation for the subjects.

The Bayesian framework gives the user the advantage and the duty to insert prior information into the model through k and $\boldsymbol{F}$. The parameter k plays the role of regularization parameter, which is rarely known a priori. We estimated it by cross-validation, although it may be interesting to adapt approximations-based methods (e.g., generalized cross-validation) to reduce the computational burden. Alternatively, we could assume a prior distribution taking values in $\mathbb{R}^+$ for the regularization parameter k and proceed to jointly estimate this parameter as well. More interestingly, the matrix $\boldsymbol{F}$ addresses the estimate of the optimal rotations, which is favorable in the analysis of fMRI data because, in this context, the variables have a spatial anatomical location. In the example in Sect. 4, our definition of $\boldsymbol{F}$ favors the combination of voxels with equal location. However, a more thoughtful specification can entirely exploit the voxels' specific spatial position in the anatomical template. This opens up the possibility to explore various specifications of $\boldsymbol{F}$ and will be the subject of further research.

Acknowledgments

The authors thank the Editor, the Associate Editor as well as the two anonymous referees for helpful comments that greatly improved the paper. Angela Andreella gratefully acknowledges funding from the grant BIRD2020/SCAR_ASEGNIBIRD2020_01 of the University of Padova, Italy, and PON 2014-2020/DM 1062 of the Ca' Foscari University of Venice, Italy. Some of the computational analyses done in this manuscript were carried out using the University of Padova Strategic Research Infrastructure Grant 2017: "CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l'Innovazione", http://capri.dei.unipd.it. The authors thank Prof. Umberto Castiello and Dr. Silvia Guerra for sharing the cinematic plant data.

Funding Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Abadir, K. M., & Magnus, J. R. (2005). *Matrix algebra* (Vol. 1). Cambridge: Cambridge University Press.

Andrade, J. M., Gómez-Carracedo, M. P., Krzanowski, W., & Kubista, M. (2004). Procrustes rotation in analytical chemistry, a tutorial. *Chemometrics and Intelligent Laboratory Systems, 72*(2), 123–132.

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., & van der Vorst, H. (2000). *Templates for the solution of algebraic eigenvalue problems: a practical guide*. Society for Industrial and Applied Mathematics.

Barndorff-Nielsen, O. (2014). *Information and exponential families: In statistical theory*. New York: Wiley.

Chikuse, Y. (2003). *Statistics on special manifolds* (Vol. 174). Berlin: Springer.

Conroy, B. R., Singer, B. D., Haxby, J. V., & Ramadge, P. J. (2009). Fmri-based inter-subject cortical alignment using functional connectivity. *Advances in Neural Information Processing systems, 22,* 378.

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., et al. (2000). Mapping functionally related regions of brain with functional connectivity mr imaging. *American Journal of Neuroradiology, 21*(9), 1636–1644.
Downs, T. D. (1972). Orientation statistics. *Biometrika, 59*(3), 665–676.
Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation, 64*(2), 105–123.
Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press.
Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological), 53*(2), 285–321.
Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*(1), 33–51.
Green, B. F. (1952). The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika, 17*(4), 429–440.
Green, P. J., & Mardia, K. V. (2006). Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika, 93*(2), 235–254.
Groisser, D. (2005). On the convergence of some procrustean averaging algorithms. *Stochastics an International Journal of Probability and Stochastic Processes, 77*(1), 31–60.
Groß, J., Trenkler, G., & Troschke, S. O. (1999). On semi-orthogonality and a special class of matrices. *Linear Algebra and its Applications, 289*(1–3), 169–182.
Guerra, S., Peressotti, A., Peressotti, F., Bulgheroni, M., Baccinelli, W., D'Amico, E., et al. (2019). Flexible control of movement in plants. *Scientific Reports, 9*(1), 1–9.
Gupta, A. K., & Nagar, D. K. (2018). *Matrix variate distributions* (Vol. 104). Chapman and Hall/CRC.
Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., et al. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron, 72*(2), 404–416.
Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: modeling shared information encoded in idiosyncratic cortical topographies. *Elife, 9,* e56601.
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825–841.
Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). *Fsl. Neuroimage,62*(2), 782–790.
Jupp, P. E., & Mardia, K. V. (1979). Maximum likelihood estimators for the matrix von mises-fisher and bingham distributions. *The Annals of Statistics, 7*(3), 599–606.
Khatri, C., & Mardia, K. V. (1977). The von mises-fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1), 95–106.
Lee, T. (2018). Bayesian attitude estimation with the matrix fisher distribution on so(3). *IEEE Transactions on Automatic Control, 63*(10), 3377–3392.
Lele, S. (1993). Euclidean distance matrix analysis (edma): Estimation of mean form and mean form difference. *Mathematical Geology, 25*(5), 573–602.
Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *Journal of Neuroscience, 33*(16), 6782–6790.
Mardia, K. V., Fallaize, C. J., Barber, S., Jackson, R. M., & Theobald, D. L. (2013). Bayesian alignment of similarity shapes. *The Annals of Applied Statistics, 7*(2), 989.
McCrae, R., Zonderman, A., Costa, P., Bond, M., & Paunonen, S. (1996). Evaluating replicability of factors in the revised neo personality inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology, 70*(3), 552–566.
Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., et al. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage, 119,* 164–174.
R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
Rohlf, F. J., & Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Biology, 39*(1), 40–59.
Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., & Haxby, J. V. (2010). Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex, 20*(1), 130–140.
Saito, V. S., Fonseca-Gessner, A. A., & Siqueira, T. (2015). How should ecologists define sampling effort? the potential of procrustes analysis for studying variation in community composition. *Biotropica, 47*(4), 399–402.
Talairach, J. J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain 3-dimensional proportional system: An approach to cerebral imaging*. Thieme Medical Publishers.
Theobald, D. L., & Wuttke, D. S. (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian procrustes problem. *Proceedings of the National Academy of Sciences, 103*(49), 18521–18527.
Van Rossum, G. & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., & Thirion, B. (2011). Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Biennial international conference on information processing in medical imaging* (pp. 562–573). Springer.

Manuscript Received: 5 AUG 2021
Final Version Received: 1 MAR 2022
Published Online Date: 18 MAY 2022

https://doi.org/10.1007/s11336-022-09859-5 Published online by Cambridge University Press