Alignment of SPD matrices to a common reference

Jacopo Schiavon[*]

March 9, 2021

1   INTRODUCTION AND NOTATION

Let $\mathbb{O}_p$ be the space of p-dimensional orthogonal matrix and $\mathcal{S}_p^+$ be the space of Symmetric and Positive Definite (SPD) matrices. In the following, unless needed, we will omit the subscript p.

Let $\Sigma_1, \Sigma_2 \in \mathcal{S}^+$ be two SPD matrices, we define the geodesic distance induced by the *affine-invariant* metric [2] as

$$d(\Sigma_1, \Sigma_2)^2 = \left\| \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right\|_{AI}^2 \tag{1}$$

where the norm of $\Sigma \in \mathcal{S}^+$ is defined as ($\lambda_i(\Sigma)$ are the eigenvalues of the matrix Σ)

$$\|\Sigma\|_{AI}^2 = \sum_{h=1}^p \log^2 \lambda_h(\Sigma) \tag{2}$$

$$= \|\mathrm{Log}\, \Sigma\|_F^2$$

where $\mathrm{Log}(\cdot)$ is the matrix logarithm (which for SPD matrices can be computed analytically via the singular value decomposition) and $\|\cdot\|_F$ is the frobenius norm (the sum of the eigenvalues).

It is useful to note that the fact that $\lambda_i(\Sigma) > 0$ for all i if $\Sigma \in \mathcal{S}^+$ implies that the norm in equation (2) is well defined. Moreover, it can be helpful to observe that $\lambda_i(\Sigma_1^{-1}\Sigma_2) = \lambda_i(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})$. Using this fact, it is immediate to prove that the following identity, which will be useful later, holds:

$$\left\| \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right\|_{AI}^2 = \left\| \Sigma_1^{-1} \Sigma_2 \right\|_{AI}^2. \tag{3}$$

Finally, we introduce a particular notation from [1]. Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we define $x^\uparrow$ and $x^\downarrow$ to be the vectors obtained by arranging the components of x respectively in increasing and decreasing order. Moreover, we say that $x \prec y$ (x is *majorised* by y) if

$$\sum_{i=1}^k x_i \leqslant \sum_{i=1}^k y_i \quad \text{for } k < n \qquad \text{and} \qquad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i.$$

[*]Department of Statistical Sciences, University of Padova. Contact: `jschiavon@stat.unipd.it`

II   ORTHOGONAL PROCRUSTES PROBLEM

The orthogonal Procrustes problem consists in finding the orthogonal matrix Ω that most closely maps A to B. The problem is stated for generic matrices with the distance induced by the Frobenius norm and its solution is given by the following lemma:

LEMMA II.I: Given two matrices A and B with $BA^\top = U\Sigma V^\top$ the singular value decomposition of their product, then

$$\underset{\Omega \in \mathbb{O}}{\operatorname{argmin}} \|\Omega A - B\|_F^2 = UV^\top \tag{4}$$

Proof. The simplest proof relies on the properties of the inner product related to the Frobenius norm:

$$
\begin{aligned}
R &= \underset{\Omega \in \mathbb{O}}{\operatorname{argmin}} \|\Omega A - B\|_F^2 \\
&= \underset{\Omega \in \mathbb{O}}{\operatorname{argmin}} \langle \Omega A - B, \ \Omega A - B \rangle \\
&= \underset{\Omega \in \mathbb{O}}{\operatorname{argmin}} \|\Omega A\|_F^2 + \|B\|_F^2 - 2 \langle \Omega A, \ B \rangle \\
&= \underset{\Omega \in \mathbb{O}}{\operatorname{argmin}} \|A\|_F^2 + \|B\|_F^2 - 2 \langle \Omega A, \ B \rangle \\
&= \underset{\Omega \in \mathbb{O}}{\operatorname{argmax}} \langle \Omega A, \ B \rangle = \underset{\Omega \in \mathbb{O}}{\operatorname{argmax}} \langle \Omega, \ BA^\top \rangle \\
&= \underset{\Omega \in \mathbb{O}}{\operatorname{argmax}} \langle \Omega, \ U\Sigma V^\top \rangle \\
&= U \left(\underset{\Omega \in \mathbb{O}}{\operatorname{argmax}} \langle U^\top \Omega V, \ \Sigma \rangle \right) V^\top
\end{aligned}
$$

Considering that Σ is diagonal by construction and that $U^\top \Omega V$ is orthogonal, this inner product is maximum [4] when $U^\top \Omega V$ is the identity, thus leaving $R = UV^\top$.   □

III   PROCRUSTES-LIKE PROBLEM FOR SPD MATRICES

It is easy to see that the product $\Omega \Sigma$ with $\Omega \in \mathbb{O}$ does not preserve the positive definiteness $\Sigma \in \mathcal{S}^+$. It can be proved [5], fortunately, that it does indeed exists a class of transformations that preserve the positive definiteness of a matrix. This class can be parameterized by an invertible matrix $W \in GL$ as

$$
\begin{aligned}
T_W \colon \mathcal{S}^+ &\to \mathcal{S}^+ \\
\Sigma &\mapsto T_W(\Sigma) = W\Sigma W^\top.
\end{aligned}
\tag{5}
$$

The orthogonal group of matrices is still of interest in that it preserves the norm of a matrix, and as such we can restate the Procrustes problem as the search for an orthogonal matrix $\Omega \in \mathbb{O}$ that, when applied with transformation (5) to a matrix Σ_2, minimizes the affine-invariant distance between the transformed Σ_2 and a reference Σ_1.

LEMMA III.I: Let $\Sigma_1, \Sigma_2 \in \mathcal{S}^+$ be two SPD matrices and let $\Sigma_i = \Gamma_i \Lambda_i \Gamma_i^\top$ the eigenvalue decomposition for each matrix. Then

$$\operatorname*{argmin}_{\Omega \in \mathbb{O}} d\left(\Sigma_1, T_\Omega(\Sigma_2)\right)^2 = \operatorname*{argmin}_{\Omega \in \mathbb{O}} d(\Sigma_1, \Omega\Sigma_2\Omega^\top)^2 = \Gamma_1 \Gamma_2^\top \tag{6}$$

This lemma is proved with more generality by BHATIA and CONGEDO in [3], and we repeat the proof here with the notation slightly changed for the sake of consistency.

Proof. First we recall the substitution (3), which allows us to write

$$d(\Sigma_1, \Sigma_2)^2 = \left[\sum_h^p \log^2 \lambda_h \left(\Sigma_1^{-1}\Sigma_2 \right) \right]$$

Then we use a result on majorization for eigenvalues from Gel'fand, Naimark and Lidskii [theorem III.4.6 in 1, p. 73], which states that

$$\log \lambda^\downarrow(\Sigma_1) + \log \lambda^\uparrow(\Sigma_2) \prec \log \lambda(\Sigma_1\Sigma_2) \prec \log \lambda^\downarrow(\Sigma_1) + \log \lambda^\downarrow(\Sigma_2)$$

or equivalently (if we replace Σ_2 with its inverse):

$$\log \lambda^\downarrow(\Sigma_1) - \log \lambda^\downarrow(\Sigma_2) \prec \log \lambda(\Sigma_1\Sigma_2^{-1}) \prec \log \lambda^\downarrow(\Sigma_1) - \log \lambda^\uparrow(\Sigma_2).$$

This set of inequalities holds also for any gauge function of the sequence of the logarithms of the eigenvalues (such as the sum of squares of the elements, see [1, § IV] on the general definition of *gauge function*) and thus implies that

$$d(\Lambda_1^\downarrow, \Lambda_2^\downarrow)^2 \leqslant d(\Sigma_1, \Sigma_2)^2 \leqslant d(\Lambda_1^\downarrow, \Lambda_2^\uparrow)^2$$

Now it is sufficient to observe that the left and right hand sides of this inequality do not change if we replace Σ_2 with $\Omega\Sigma_2\Omega^\top$, which means that to obtain the minimum of $d(\Sigma_1, \Omega\Sigma_2\Omega^\top)$ we need that

$$\lambda\left(\Gamma_1^{-\top}\Lambda_1^{-1}\Gamma_1^{-1}\Omega\Gamma_2\Lambda_2\Gamma_2^\top\Omega^\top\right) = \lambda\left(\Lambda_1^{-1}\Lambda_2\right)$$

which implies (using the cyclic property of the trace)

$$\Omega = \Gamma_1\Gamma_2^\top$$

as we wanted to show. $\square$

IV EXTENSION TO MULTIPLE MATRICES

Let assume that we have n SPD matrices $\{\Sigma_i\}_{i=1}^n$ of fixed rank p and that we are interested in finding the n orthogonal matrices $\vec{\Omega} = \Omega_i$ that *most closely maps all the matrices together*. We can state this problem in two formal ways:

1. We are interested in finding the set of matrices that minimize the pairwise distances:

$$\underset{\vec{\Omega}\in\mathbb{O}}{\operatorname{argmin}} \sum_{i}^{n} \sum_{j>i}^{n} d\left(T_{\Omega_i}(\Sigma_i), T_{\Omega_j}(\Sigma_j)\right)^2$$

2. We are interested in finding the set of matrices that minimize the distances with respect to a fixed reference M:

$$\underset{\vec{\Omega}\in\mathbb{O}}{\operatorname{argmin}} \sum_{i}^{n} d\left(M, T_{\Omega_j}(\Sigma_j)\right)^2$$

The first problem of course does not admit a unique solution, as one can rotate every set of matrices in infinite ways, but we can modify it slightly by fixing a matrix Σ_k as reference

$$\underset{\vec{\Omega}\in\mathbb{O}}{\operatorname{argmin}} \sum_{i\neq k}^{n} d\left(\Sigma_k, T_{\Omega_i}(\Sigma_i)\right)^2 + \sum_{i,j\neq k; j>i}^{n} d\left(T_{\Omega_i}(\Sigma_i), T_{\Omega_j}(\Sigma_j)\right)^2.$$

Rewritten in this way, the first problem is actually a particular case of the second one, so we will only provide the solution to the latter in the following lemma.

LEMMA IV.I: Let $\{\Sigma_i \in \mathcal{S}^+\}_{i=1}^n$ a set of SPD matrices and let $M \in \mathcal{S}^+$ a reference SPD matrix. Let $\Sigma_i = \Gamma_i \Lambda_i \Gamma_i^\top$ and $M = \Gamma_M \Lambda_M \Gamma_M$ the eigenvalue decomposition respectively of Σ_i and of M and let $T_\Omega \Sigma_i = \Omega_i \Sigma_i \Omega_i^\top$ the transformed matrix Σ_i and $\vec{\Omega}$ the set of orthogonal transformations. Then,

$$\underset{\vec{\Omega}\in\mathbb{O}}{\operatorname{argmin}} = \sum_{i}^{n} d(M, T_{\Omega_i}(\Sigma_i))^2 = \left\{\Gamma_M \Gamma_i^\top\right\}_{i=1}^n \tag{7}$$

Proof. To prove this result it is sufficient to observe that each $d(M, T_{\Omega_i}(\Sigma_i))^2 \geqslant 0$ for every i and every choice of Ω and moreover there are not cross-dependencies between the terms, thus the minimum of the sum is achieved at the minimum of each term. Then the result follows by repeatedly applying Lemma III.I. □

V   THE OPTIMAL REFERENCE MATRIX

Given the result obtained in the previous section, it is interesting to ask if it exists a (unique) solution to the problem

$$\underset{M\in\mathcal{S}^+}{\operatorname{argmin}} \sum_{i}^{n} d\left(M, T_{\Omega_i}(\Sigma_i)\right)^2 \qquad \text{with } \Omega_i = \Gamma_M \Gamma_i^\top \tag{8}$$

which asks if there is an *optimal* matrix to use as a reference for the rotations, and considers optimal the matrix that will be the closest to the rotated elements.

v.i   Relations with the geometrical mean (or Karcher mean)

This problem resembles the expression of the geometrical mean of a set of SPD matrices, known in literature as the *Karcher* mean, which is defined as the matrix M that minimizes

$$\underset{M \in \mathcal{S}^+}{\operatorname{argmin}} = \sum_i^n d\left(M, \Sigma_i\right)^2 = \sum_i^n \left[\sum_h^p \log^2 \lambda_h \left(M^{-1}\Sigma_i\right) \right].$$

Even if this and problem (8) looks very similar, the only difference being the rotation transformation T_{Ω_i}, this apparently small difference together with the expression of the rotations (7), will allow various simplifications in the expression, providing the possibility of an elegant and simple analytical solution which is not available in the general Karcher mean problem.

v.ii   Eigenvalues of the optimal reference matrix

By writing explicitly this expression and using expression (7) for the rotations we obtain

$$\begin{aligned}
\sum_i^n d\left(M, T_{\Omega_i}(\Sigma_i)\right)^2 &= \sum_i^n \left[\sum_h^p \log^2 \lambda_h \left(M^{-1} T_{\Omega_i}(\Sigma_i)\right) \right] \\
&= \sum_i^n \left[\sum_h^p \log^2 \lambda_h \left(M^{-1} \Gamma_M \Gamma_i^\top \Sigma_i \Gamma_i \Gamma_M^\top\right) \right] \\
&= \sum_i^n \left[\sum_h^p \log^2 \lambda_h \left(\Gamma_M \Lambda_M^{-1} \Gamma_M^\top \Gamma_M \Gamma_i^\top \Gamma_i \Lambda_i \Gamma_i^\top \Gamma_i \Gamma_M^\top\right) \right] \\
&= \sum_i^n \left[\sum_h^p \log^2 \lambda_h \left(\Lambda_M^{-1} \Lambda_i\right) \right]
\end{aligned} \tag{9}$$

From this we can observe that the final expression depends only on the eigenvalues of M. This implies two things:

1. Since $M = \Gamma_M \Lambda_M \Gamma_M^\top$ but equation (9) depends only on the eigenvalues Λ_M, there are infinite solutions M that solves the problem, as one can chose any matrix $\Gamma_M \in \mathbb{O}$ as the eigenvector matrix.

2. On the other side, the solution for the eigenvalues can be obtained as a standard optimization in an euclidean constrained setting. Indeed, it can be rewritten as

$$\underset{\vec{\lambda} \in \mathbb{R}_+^p}{\operatorname{argmin}} = \sum_i^n \sum_h^p \log^2 \left(\frac{\alpha_{ih}}{\lambda_h} \right)$$

   where α_{ji} is the ith eigenvalue of the j-th matrix and λ_i the ith eigenvalue of the objective matrix. This problem has a simple and elegant analytical solution, as the minimum of this function is achieved in

$$\lambda_h = \left[\prod_{i=1}^n \alpha_{ih} \right]^{\frac{1}{n}} \tag{10}$$

as can be seen by computing the gradient and noticing that each component is independent and there is only one zero. From the continuity of the function and its limits for $\lambda_h \to 0$ and $\lambda_h \to +\infty$ we see that this is indeed a minimum.

V.III   EIGENVECTORS OF THE OPTIMAL REFERENCE MATRIX

In order to select a reference matrix M, as discussed in the previous paragraph, we then need to choose a matrix Γ_M, which is the eigenvector matrix. From a formal standpoint the only requirement is that this matrix need to be orthogonal, which does not restrict much the possible choices, thus we need to add an heuristic constraint that can limit the choice.

Again, we will follow a minimization approach: given the expression (7), we see that the rotations depends only on the eigenvector matrix, thus it might be interesting to select the matrix Γ_M that minimizes the *rotational effort*, or in other words the matrix Γ_M that generates the smallest rotations. This can be achieved by minimizing the term

$$\operatorname*{argmin}_{\Gamma \in \mathbb{O}} \sum_i^n \left\| \Gamma \Gamma_i^\top - \mathbb{I} \right\|_F^2 \tag{11}$$

It is immediate to observe the resemblance between this equation and equation (4). Indeed, this problem might be seen as a generalization to the Procrustes problem to the case of n matrices. Such problem does not have a solution in general, but the particular case that we are analyzing allows a simple expression.

Indeed, by following the same steps of Lemma II.I, we obtain:

$$\begin{aligned}
\Gamma_M &= \operatorname*{argmin}_{\Gamma \in \mathbb{O}} \sum_i^n \left\| \Gamma \Gamma_i^\top - \mathbb{I} \right\|_F^2 \\
&= \operatorname*{argmin}_{\Gamma \in \mathbb{O}} \sum_i^n \left\langle \Gamma \Gamma_i^\top - \mathbb{I}, \ \Gamma \Gamma_i^\top - \mathbb{I} \right\rangle \\
&= \operatorname*{argmin}_{\Gamma \in \mathbb{O}} \sum_i^n \left[\left\| \Gamma \Gamma_i^\top \right\|_F^2 + \left\| \mathbb{I} \right\|_F^2 - 2 \left\langle \Gamma \Gamma_i^\top, \ \mathbb{I} \right\rangle \right] \\
&= \operatorname*{argmin}_{\Gamma \in \mathbb{O}} \sum_i^n \left[2 \left\| \mathbb{I} \right\|_F^2 - 2 \left\langle \Gamma \Gamma_i^\top, \ \mathbb{I} \right\rangle \right] \\
&= \operatorname*{argmax}_{\Gamma \in \mathbb{O}} \sum_i^n \left\langle \Gamma \Gamma_i^\top, \mathbb{I} \right\rangle = \operatorname*{argmax}_{\Gamma \in \mathbb{O}} \sum_i^n \left\langle \Gamma, \Gamma_i \right\rangle \\
&= \operatorname*{argmax}_{\Gamma \in \mathbb{O}} \sum_i^n \operatorname{Tr} \left[\Gamma^\top \Gamma_i \right] = \operatorname*{argmax}_{\Gamma \in \mathbb{O}} \operatorname{Tr} \left[\Gamma^\top \left(\sum_i^n \Gamma_i \right) \right] \\
&= \operatorname*{argmax}_{\Gamma \in \mathbb{O}} \left\langle \Gamma, \sum_i^n \Gamma_i \right\rangle = U V^\top \tag{12}
\end{aligned}$$

where U and V are obtained from the singular value decomposition of $\sum_i^n \Gamma_i$ and in the second-to-last we used the distributive and linearity properties of the trace.

References

[1]   Rajendra Bhatia
      Matrix analysis. Springer-Verlag New York, 1997.

[2]   Rajendra Bhatia
      Positive definite matrices. Princeton University Press, 2007.

[3]   Rajendra Bhatia and Marco Congedo
      "Procrustes problems in Riemannian manifolds of positive definite matrices". In: *Linear Algebra and its Applications* 563 (2019), pp. 440–445.

[4]   John C. Gower and Garmt B. Dijksterhuis
      Procrustes Problems. OUP Oxford, 2004. 248 pp.

[5]   Huynh Dinh Tuan, Tran Thi Nha Trang, and Doan The Hieu
      "Positive definite preserving linear transformations on symmetric matrix spaces". In: *ArXiv preprint* (2010). arXiv: 1008.1347.