# Procrustes problems in Riemannian manifolds of positive definite matrices

Rajendra Bhatia, Marco Congedo

# Procrustes Problems in Riemannian Manifolds of Positive Definite Matrices

Rajendra Bhatia[a], Marco Congedo[b*].


a Ashoka University, Sonepat, Haryana, 131029, India.

b GIPSA-lab, CNRS, Grenoble Alpes University, Grenoble INP, Grenoble, France

* corresponding author: marco.congedo@gmail.com

**Abstract**

We consider the manifold of positive definite matrices endowed with the Fisher Riemannian metric and some other distances commonly used in information theory. We show that for each of them the best approximant to $A$ from the unitary orbit of another matrix $B$ commutes with $A$.

**Keywords:** Procrustes Problem, Riemannian Manifold, Positive Definite Matrices, Matrix Approximation

# 1. Introduction

The classical Procrustes problem in statistics is the problem of finding $\min_U \|A - BU\|_2$, where $A$ and $B$ are fixed $n \times n$ matrices and $U$ varies over the unitary matrices. (See e.g., [1].) Here $\|X\|_2$ denotes the Euclidean (Frobenius) norm

$$\|X\|_2 = \left(\operatorname{tr} X^*X\right)^{1/2} = \left(\sum_{j=1}^n s_j^2(X)\right)^{1/2}, \tag{1}$$

$s(X) = (s_1(X), \ldots, s_n(X))$ being the singular values of $X$. There is a variant of the problem in which the Euclidean norm is replaced by a unitarily invariant norm $\|\cdot\|_\Phi$ corresponding to a symmetric gauge function $\Phi$:

$$\|X\|_\Phi = \Phi(s(X)). \tag{2}$$

See Chapter IV of [2] for the basic facts about such norms used in this paper.

In recent years there has been considerable interest in the manifold **P** of positive definite matrices with the Riemannian distance

$$\delta_2(A, B) = \left\|\log\left(A^{-1/2}BA^{-1/2}\right)\right\|_2. \tag{3}$$

If $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$ are the eigenvalues of $X$, then

$$\delta_2(A, B) = \left[\sum_{j=1}^n \log^2 \lambda_j\left(A^{-1}B\right)\right]^{1/2}. \tag{4}$$

This is the matrix version of the Fisher-Rao metric on probability distributions. (See Chapter 6 of [3].) The analogue of the Procrustes problem on **P** is solving

$$\min_U \delta_2\left(A, U^*BU\right), \tag{5}$$

where $U$ varies over the set of unitary matrices $\mathbf{U}$. Let $\mathcal{U}_B = \left\{ U^* B U : U \in \mathbf{U} \right\}$ be the *unitary orbit* of $B$.

Then problem (5) asks for the best $\delta_2$-approximant of $A$ from the set $\mathcal{U}_B$.

Each symmetric gauge function $\Phi$ on $\mathbb{R}^n$ gives rise to a metric on $\mathbf{P}$ defined as

$$\delta_\Phi (A, B) = \left\| \log \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right) \right\|_\Phi = \Phi \left( \log \lambda \left( A^{-1} B \right) \right). \tag{6}$$

(See [4].) The metric $\delta_2$ is the special case when $\Phi$ is the symmetric gauge function corresponding to the Euclidean norm. Besides these, there are other distances on $\mathbf{P}$ that are widely used in statistics, signal processing, machine learning and other areas. The Bures-Wasserstein distance

$$d_W (A, B) = \left[ \operatorname{tr} (A + B) - 2 \operatorname{tr} \left( A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} \right]^{\frac{1}{2}} \tag{7}$$

is of great interest in the theory of optimal transport, quantum information and Riemannian geometry. See [5] for a recent exposition of this topic and for earlier references. Widely used are also the Kullback-Leibler divergence (a Bregman divergence) defined as

$$d_{KL}^2 (A, B) = \operatorname{tr} \left( A B^{-1} - I \right) - \log \det \left( A B^{-1} \right) \tag{8}$$

and the Bhattacharyya divergence

$$d_B^2 (A, B) = \log \det \left( \frac{A + B}{2} \right) - \log \det \left( A^{\frac{1}{2}} B^{\frac{1}{2}} \right). \tag{9}$$

Sra [6] has recently shown that the square root of $d_B^2$ is a metric. This is not true for $d_{KL}^2$. Chebbi and Moakher [7] have defined a parametrized family of log-determinant divergence functions that include (8) and (9). These are:

$$d_\alpha^2 = \begin{cases} \frac{4}{1-\alpha^2} \log \det \left( \frac{1-\alpha}{2} A + \frac{1+\alpha}{2} B \right) - \log \left( \det A^{\frac{1-\alpha}{2}} \det B^{\frac{1+\alpha}{2}} \right), & -1 < \alpha < 1 \\ \operatorname{tr} \left( A^{-1} B - I \right) - \log \det \left( A^{-1} B \right), & \alpha = -1 \\ \operatorname{tr} \left( B^{-1} A - I \right) - \log \det \left( B^{-1} A \right), & \alpha = 1 \end{cases} \tag{10}$$

For α=0 this is $d_B^2$ (up to a constant factor) and for α=1 this is $d_{KL}^2$. We will see that for any of the distances in (6)-(10) the Procrustes problem on **P** has the same solution as for the metric (4).

Let $\lambda_j^\downarrow(A)$, $1 \le j \le n$, be the eigenvalues of $A$ arranged in decreasing order and $\lambda_j^\uparrow(A)$ be an enumeration of the same numbers in increasing order. (Thus $\lambda_j^\downarrow(A) = \lambda_{n-j+1}^\uparrow(A)$ for all $1 \le j \le n$.) We denote by $\lambda^\downarrow(A)$ both the vector $\left(\lambda_1^\downarrow(A),\dots,\lambda_n^\downarrow(A)\right)$ and the diagonal matrix $\mathrm{diag}\left(\lambda_1^\downarrow(A),\dots,\lambda_n^\downarrow(A)\right)$. It will be clear from the context which one of the two is meant.

By the Spectral Theorem there exist unitary matrices $V^\downarrow$ and $V^\uparrow$ such that

$$V^{\downarrow *} A V^\downarrow = \lambda^\downarrow(A) \quad \text{and} \quad V^{\uparrow *} A V^\uparrow = \lambda^\uparrow(A) \tag{11}$$

and unitary matrices $W^\downarrow$ and $W^\uparrow$ such that

$$W^{\downarrow *} B W^\downarrow = \lambda^\downarrow(B) \quad \text{and} \quad W^{\uparrow *} B W^\uparrow = \lambda^\uparrow(B). \tag{12}$$

Our main result is the following:

*Theorem*

Let $A$ and $B$ any two elements of **P**. Consider the function $f(U) = d\left(A, U^* B U\right)$ defined on the manifold **U** of unitary matrices with $d(\cdot, \cdot)$ any of the metrics given in (6)-(10). Then the global extrema of $f(U)$ occur at points of $\mathcal{U}_B$ commuting with $A$. Further,

$$\underset{U}{\mathrm{argmin}}\, f(U) = W^\downarrow V^{\downarrow *}, \qquad \min f(U) = d\left(\lambda^\downarrow(A), \lambda^\downarrow(B)\right) \tag{13}$$

and

$$\underset{U}{\mathrm{argmax}}\, f(U) = W^\uparrow V^{\downarrow *}, \qquad \max f(U) = d\left(\lambda^\downarrow(A), \lambda^\uparrow(B)\right). \tag{14}$$

*Comment*

The extreme points being the same for a variety of different metrics has a parallel in the celebrated Lidskii-Wielandt Theorem in matrix analysis. This says that for Hermitian matrices *A* and *B* we have

$$\left\| \lambda^{\downarrow}(A) - \lambda^{\downarrow}(B) \right\|_{\Phi} \leq \left\| A - B \right\|_{\Phi} \leq \left\| \lambda^{\downarrow}(A) - \lambda^{\uparrow}(B) \right\|_{\Phi}, \tag{15}$$

for every unitarily invariant norm $\left\| \cdot \right\|_{\Phi}$. This statement can be reinterpreted to say that the distance $\left\| A - U^{*}BU \right\|_{\Phi}$ is minimized when $U_{\min} = W^{\downarrow}V^{\downarrow *}$ and maximized when $U_{\max} = W^{\uparrow}V^{\downarrow *}$.

## 2. Proof of the Theorem

We will use some standard results on majorization, log-majorization and eigenvalues. A standard reference for these topics is [2], Chapters II and III. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be real vectors and let $x_1^{\downarrow}, \ldots, x_n^{\downarrow}$ be the components of *x* arranged in decreasing order. The same numbers arranged in increasing order are denoted as $x_1^{\uparrow}, \ldots, x_n^{\uparrow}$. We say that *x* is majorized by *y*, is symbols $x \prec y$, if for all $1 \leq k \leq n$ we have $\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}$, with equality when *k=n*. Let *A* and *B* be two positive definite matrices. According to a famous Theorem of Gel'fand, Naimark and Lidskii we have

$$\log \lambda^{\downarrow}(A) + \log \lambda^{\uparrow}(B) \prec \log \lambda(AB) \prec \log \lambda^{\downarrow}(A) + \log \lambda^{\downarrow}(B). \tag{16}$$

See p.73 in [2] or p. 228 in [3]. Changing *B* to $B^{-1}$ leads to

$$\log \lambda^{\downarrow}(A) - \log \lambda^{\downarrow}(B) \prec \log \lambda(AB^{-1}) \prec \log \lambda^{\downarrow}(A) - \log \lambda^{\uparrow}(B). \tag{17}$$

Hence for every symmetric gauge function $\Phi$

$$\Phi\left(\log \lambda^{\downarrow}(A) - \log \lambda^{\downarrow}(B)\right) \leq \Phi\left(\log \lambda(AB^{-1})\right) \leq \Phi\left(\log \lambda^{\downarrow}(A) - \log \lambda^{\uparrow}(B)\right). \tag{18}$$

This is the same as saying

$$\delta_\Phi\left(\lambda^\downarrow(A),\lambda^\downarrow(B)\right)\le\delta_\Phi(A,B)\le\delta_\Phi\left(\lambda^\downarrow(A),\lambda^\uparrow(B)\right). \tag{19}$$

The left and right hand sides of this inequality do not change when $B$ is replaced by any other matrix in its unitary orbit. This shows that the minimum of $\delta_\Phi\left(A,U^*BU\right)$ is attained with $U_{\min}=W^\downarrow V^{\downarrow*}$ and the maximum with $U_{\max}=W^\uparrow V^{\downarrow*}$. This proves the Theorem for the metrics $\delta_\Phi$ defined in (6).

Next, we turn to the Bures-Wasserstein metric defined in (7). Since $\mathrm{tr}(A+B)=\mathrm{tr}\left(A+U^*BU\right)$ for every unitary $U$, we need to consider only the term

$$\mathrm{F}(A,B)=\mathrm{tr}\left(A^{\frac{1}{2}}BA^{\frac{1}{2}}\right)=\mathrm{tr}(AB)^{\frac{1}{2}}. \tag{20}$$

This is called the *fidelity* between $A$ and $B$ in physics literature and the problem of finding its extreme values on unitary orbits has been studied, for example, in [8, 9]. For the reader's convenience we give a proof following the same arguments as above. Since $\log\lambda\left(A^{\frac{1}{2}}\right)=\frac{1}{2}\log\lambda(A)$ and log is a monotone function, we obtain from (16)

$$\log\lambda^\downarrow\left(A^{\frac{1}{2}}\right)+\log\lambda^\uparrow\left(B^{\frac{1}{2}}\right)\prec\log\lambda^\downarrow(AB)^{\frac{1}{2}}\prec\log\lambda^\downarrow\left(A^{\frac{1}{2}}\right)+\log\lambda^\downarrow\left(B^{\frac{1}{2}}\right). \tag{21}$$

As a corollary

$$\left\langle\lambda^\downarrow\left(A^{\frac{1}{2}}\right),\lambda^\uparrow\left(B^{\frac{1}{2}}\right)\right\rangle\le\mathrm{tr}(AB)^{\frac{1}{2}}\le\left\langle\lambda^\downarrow\left(A^{\frac{1}{2}}\right),\lambda^\downarrow\left(B^{\frac{1}{2}}\right)\right\rangle. \tag{22}$$

It follows that

$$\max\mathrm{F}\left(A,U^*BU\right)=\mathrm{F}\left(\lambda^\downarrow(A),\lambda^\downarrow(B)\right)=\left\langle\lambda^\downarrow\left(A^{\frac{1}{2}}\right),\lambda^\downarrow\left(B^{\frac{1}{2}}\right)\right\rangle, \tag{23}$$

$$\min\mathrm{F}\left(A,U^*BU\right)=\mathrm{F}\left(\lambda^\downarrow(A),\lambda^\uparrow(B)\right)=\left\langle\lambda^\downarrow\left(A^{\frac{1}{2}}\right),\lambda^\uparrow\left(B^{\frac{1}{2}}\right)\right\rangle. \tag{24}$$

This proves the Theorem for the distance $d_W$.

Now we consider the Kullback-Leibler divergence defined in (8) and as the limit cases $\alpha = \pm 1$ in (10) . Since $\det\left(AU^*B^{-1}U\right) = \det\left(AB^{-1}\right)$ for every unitary $U$, we need to consider only the term $\operatorname{tr}\left(AU^*B^{-1}U\right)$ in (8). The statement of the Theorem follows from von Neumann's trace inequality ([2], p. 78):

$$\left\langle \lambda^\downarrow(A), \lambda^\uparrow(B) \right\rangle \leq \operatorname{tr}(AB) \leq \left\langle \lambda^\downarrow(A), \lambda^\downarrow(B) \right\rangle. \tag{25}$$

This inequality is a special corollary of (16).

Finally, we look at the log-determinant α-divergence for $-1 < \alpha < 1$, which reduces to the divergence (9) when $\alpha = 0$. In this case for every unitary $U$, $\log\det\left(A^{\frac{1-\alpha}{2}}\left(U^*BU\right)^{\frac{1+\alpha}{2}}\right) = \log\det\left(A^{\frac{1-\alpha}{2}}B^{\frac{1+\alpha}{2}}\right)$. The statement of the Theorem follows from an inequality due to Fiedler that says ([2] p. 182)

$$\prod_{j=1}^{n}\left(\lambda_j^\downarrow(A) + \lambda_j^\downarrow(B)\right) \leq \det(A+B) \leq \prod_{j=1}^{n}\left(\lambda_j^\downarrow(A) + \lambda_j^\uparrow(B)\right), \tag{26}$$

upon replacing $A$ and $B$ with $A^{\frac{1-\alpha}{2}}$ and $B^{\frac{1+\alpha}{2}}$.

## 3. Remarks

1) We present another argument for solving the minimization problem in our theorem for the metrics $\delta_\Phi$. This invokes the fundamental *exponential metric increasing inequality* (EMI):

$$\delta_\Phi(A, B) \geq \|\log A - \log B\|_\Phi. \tag{27}$$

(See [4].) Combining this with the first of the Lidskii-Wielandt inequalities (15) one sees that

$$\begin{aligned}
\delta_\Phi\left(A, U^*BU\right) &\geq \left\|\log A - U^*(\log B)U\right\|_\Phi \\
&\geq \left\|\lambda^\downarrow(\log A) - \lambda^\downarrow(\log B)\right\|_\Phi \\
&= \left\|\log \lambda^\downarrow(A) - \log \lambda^\downarrow(B)\right\|_\Phi \\
&= \delta_\Phi\left(\lambda^\downarrow(A), \lambda^\downarrow(B)\right)
\end{aligned}$$

2)        In another approach, one may study the extremal problem for the function $f(U) = d(A, U^*BU)$ on $\mathcal{U}_B$ as a calculus problem. One can write out the derivatives $Df(U)$ for the different metrics in (6)-(10) and show that if $U_0$ is any local extremum for $f(U)$, then $U_0^*BU_0$ commutes with $A$. In particular this is true for the global minimum $U_{min}$ and for the global maximum $U_{max}$.

## Acknowledgements

## References

[1]  N. Higham, Functions of Matrices: Theory and Computations, SIAM, Philadelphia, 2008.
[2]  R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.
[3]  R. Bhatia, Positive Definite Matrices, Princeton University Press, New Jersey, 2007.
[4]  R. Bhatia, On the exponential metric increasing property, Linear Algebra Appl., 375 (2003) 211-220.
[5]  R. Bhatia, T. Jain, Y. Lim, On the Bures-Wasserstein distance between positive definite matrices, Expo. Math., to appear.
[6]  S. Sra, Positive definite matrices and the S-divergence, Proc. Amer. Math. Soc, 144 (2016), 2787-2797.
[7]  Z. Chebbi, M. Moakher, Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function, Linear Algebra Appl., 436 (7), (2012), 1872-1889.
[8]  D. Markham, J.A. Miszczak, Z. Puchala, K. Zyczkowski, Quantum state discrimination: a geometric approach, Phys. Rev. A, 77, 042111 (2008).
[9]  L. Zhang, Fei S.-M., Quantum fidelity and relative entropy between unitary orbits, Journal of Phys. A Math. Theoret., 47(5), (2014), 055301