O(n)-invariant Riemannian metrics on SPD matrices

Yann Thanwerdas, Xavier Pennec

Linear Algebra and its Applications 661 (2023) 163–201

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

O(n)-invariant Riemannian metrics on SPD matrices

Yann Thanwerdas *, Xavier Pennec

Université Côte d'Azur and Inria, Epione Project Team, Sophia Antipolis, 2004 route des Lucioles, 06902 Valbonne Cedex, France

ARTICLE INFO

Article history:
Received 8 September 2021
Accepted 12 December 2022
Available online 15 December 2022
Submitted by B. Vandereycken

MSC:
15B48
53B20
15A63
53C22
58D17

Keywords:
Symmetric Positive Definite matrices
Riemannian geometry
Invariance under orthogonal transformations
Families of metrics
Log-Euclidean metric
Affine-invariant metric
Bures-Wasserstein metric
Kernel metrics

ABSTRACT

Symmetric Positive Definite (SPD) matrices are ubiquitous in data analysis under the form of covariance matrices or correlation matrices. Several O(n)-invariant Riemannian metrics were defined on the SPD cone, in particular the kernel metrics introduced by Hiai and Petz. The class of kernel metrics interpolates between many classical O(n)-invariant metrics and it satisfies key results of stability and completeness. However, it does not contain all the classical O(n)-invariant metrics. Therefore in this work, we investigate super-classes of kernel metrics and we study which key results remain true. We also introduce an additional key result called cometric-stability, a crucial property to implement geodesics with a Hamiltonian formulation. Our method to build intermediate embedded classes between O(n)-invariant metrics and kernel metrics is to give a characterization of the whole class of O(n)-invariant metrics on SPD matrices and to specify requirements on metrics one by one until we reach kernel metrics. As a secondary contribution, we synthesize the literature on the main O(n)-invariant metrics, we provide the complete formula of the sectional curvature of the affine-invariant metric and the formula of the geodesic

* Corresponding author.
E-mail addresses: yann.thanwerdas@inria.fr (Y. Thanwerdas), xavier.pennec@inria.fr (X. Pennec).

https://doi.org/10.1016/j.laa.2022.12.009

parallel transport between commuting matrices for the Bures-Wasserstein metric.

1. Introduction

Symmetric Positive Definite (SPD) matrices are ubiquitous in data analysis because in many situations, the data (signals, images, diffusion coefficients...) can be represented by their covariance matrices. This is the case in the domains of Brain-Computer Interfaces, diffusion and functional MRI, Computer Vision, Diffusion Tensor Imaging (DTI)... SPD matrices form a cone in the vector space of symmetric matrices so a first idea to compute with SPD matrices could be to perform Euclidean computations on symmetric matrices. However, this method has several drawbacks. As geodesics are straight lines, they leave the SPD cone at finite time so extrapolation methods could lead to non admissible matrices, namely with negative eigenvalues. Moreover, the trace is linearly interpolated but other invariants such as the determinant are not monotonically interpolated along geodesics. For example in DTI, where SPD matrices are represented by 3D ellipsoids, the ellipsoids along the geodesic can have a larger volume than the two ellipsoids at extremities, which leads to non realistic predictions in fiber tracking (swelling effect).

Hence, other Riemannian metrics were used in applications to solve these problems. The affine-invariant/Fisher-Rao metric [28,23,13,9,19,4,34,3] provides a Riemannian symmetric structure to the SPD manifold: it is negatively curved, geodesically complete (matrices with null eigenvalues are rejected to infinity), it is invariant under the congruence action (which, in the context of covariance matrices, corresponds to the invariance of the feature vector under affine transformations) and it is inverse-consistent. The log-Euclidean metric [2] is diffeomorphic to a Euclidean inner product: it also provides a Riemannian symmetric space, it is geodesically complete and inverse-consistent. It is not curved and it is not affine-invariant although it is still invariant under orthogonal transformations and dilations. The Bures-Wasserstein/Procrustes metric [6,7,31,16] is a positively curved quotient metric which is also invariant under orthogonal transformations. It is not geodesically complete but geodesics remain in the cone with boundaries: this means that this metric is suited for computing with Positive Semi-Definite (PSD) matrices. Many other interesting metrics exist with different properties: Bogoliubov-Kubo-Mori [24,17], polar-affine [29], Euclidean-Cholesky [35], log-Euclidean-Cholesky [14], log-Cholesky [25,15], power-Euclidean [8], and more recently power-affine [33], alpha-Procrustes [10], mixed-power-Euclidean [32].

Except those named after Cholesky, all the other Riemannian metrics cited above are invariant under orthogonal transformations. If we consider SPD matrices as covariance matrices, this transformation corresponds to a rigid-body transformation of the feature

vector $X \in \mathbb{R}^n \longmapsto RX + X_0$ where R is an orthogonal matrix. In 2009, Hiai and Petz introduced the subclass of *kernel metrics* [11], which are $\mathrm{O}(n)$-invariant metrics indexed by smooth symmetric maps $\phi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$. This class satisfies key results: it contains most of the cited $\mathrm{O}(n)$-invariant metrics, it is stable under a certain class of diffeomorphisms and it provides a sufficient condition for geodesic completeness. This sufficient condition becomes necessary if we restrict the class to the subclass of *mean kernel metrics* which is indexed by kernel maps of the form $\phi = m^\theta$ where $m : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$ is a symmetric homogeneous mean and $\theta \in \mathbb{R}$ is a power. However, the class of kernel metrics does not contain *all* the aforementioned $\mathrm{O}(n)$-invariant metrics. The main goal of this paper is to study the super-classes of kernel metrics, especially the whole class of $\mathrm{O}(n)$-invariant metrics for which we give a characterization. More precisely, our objective is to determine which key results on kernel metrics can be generalized and thus to understand better the specificity of kernel metrics within these super-classes.

1.1. Results and organization of the paper

In the remainder of the Introduction, we give the notations and conventions used in the paper. In Section 2, we introduce two preliminary concepts and one result. The first concept is the notion of $\mathrm{O}(n)$-*equivariant map* on symmetric matrices. We especially explain how to build them from a map defined on diagonal matrices via the spectral theorem because this is a procedure we need several times in the paper. Then the second concept is a particular case of the previous one, called *univariate map*. These are maps characterized by a map on positive real numbers. They are particularly interesting because their differential is known in closed form modulo eigenvalue decomposition and because the class of kernel metrics is stable under univariate diffeomorphisms. Finally the result is the characterization of $\mathrm{O}(n)$-invariant inner products on symmetric matrices. These inner products are composed of two terms, the Frobenius term and the trace term, which have different weights so they form a two-parameter family. In the proof, we give elementary tools that we reuse when we characterize $\mathrm{O}(n)$-invariant metrics on SPD matrices.

To explain why kernel metrics do not encompass all the $\mathrm{O}(n)$-invariant metrics cited above, we need to present them or at least the most important ones. One can notice that many metrics and families of metrics are actually based on five of them, namely the Euclidean, the log-Euclidean, the affine-invariant, the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics. That is why in Section 3, we synthesize the literature on these five noted metrics. For each of them, we give the fundamental Riemannian operations (squared distance, Levi-Civita connection, curvature, geodesics, logarithm map, parallel transport map) when they are known. As a secondary contribution of the paper, we give the complete formula of the sectional curvature of the affine-invariant metric and we also give, for the Bures-Wasserstein metric, the new formula of the parallel transport between commuting matrices and simpler formulae of the Levi-Civita connection, the curvature and the parallel transport equation.

In Section 4, after reviewing kernel metrics and their key properties, we give two new observations on them. Firstly, the cometric of a metric on SPD matrices can be considered itself as a metric on SPD matrices by identifying the vector space of symmetric matrices and its dual via the Frobenius inner product. Therefore we observe that the cometric of a kernel metric defined by the kernel map ϕ is a kernel metric characterized by $1/\phi$. This remarkable result has an important consequence for the numerical computation of geodesics. Indeed, the geodesic equation $\nabla_{\dot\gamma}\dot\gamma = 0$, which is a second order equation, has a Hamiltonian version which is a first order equation that only involves the cometric, not the Christoffel symbols. The Hamiltonian equation is much simpler to integrate and numerically more stable, that is why it is often preferred in numerical implementations, for instance in the Python package `geomstats` [18]. Hence knowing a simple explicit formula for the cometric helps to compute numerically the geodesics. Secondly, there is a natural extension of kernel metrics that encompasses all the aforementioned $\mathrm{O}(n)$-invariant metrics, which still satisfies the key properties of kernel metrics including the cometric stability. Roughly speaking, kernel metrics look like the Frobenius inner product on symmetric matrices where the elementary quadratic forms (the X_{ij}^2) are weighted by a coefficient involving the kernel map ϕ and depending on the point. Since the Frobenius inner product is not the only $\mathrm{O}(n)$-invariant inner product on symmetric matrices as explained above, the trace term can be added to the framework of kernel metrics to form extended kernel metrics.

In Section 5, we characterize the class of $\mathrm{O}(n)$-invariant metrics on SPD matrices by means of three multivariate maps $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}$ operating on the eigenvalues $(d_1, ..., d_n)$ of the SPD matrix and which satisfy three conditions of symmetry, compatibility and positivity (Theorem 5.1). Then, we observe that kernel metrics are characterized by two properties within this family. They are *ortho-diagonal*: it means that the metric matrix is diagonal, i.e. $\beta = 0$. They are *bivariate*: it means that the remaining functions α and γ do not depend on their $n-2$ last terms, and the compatibility condition imposes that they are equal so we can write $\gamma = \alpha = 1/\phi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$. Since the term "kernel" is quite overloaded in many different contexts (such as in Reproducing Kernel Hilbert Spaces in machine learning or in kernel density estimation/regression in statistics), we propose to designate them as Bivariate Ortho-Diagonal (BOD) metrics. Afterwards, we give key properties of $\mathrm{O}(n)$-invariant metrics in analogy with the key properties of BOD (kernel) metrics. Since we do not have a closed-form expression for the cometric anymore, we introduce the intermediate class of bivariate separable metrics which is cometric-stable and we give the expression of the cometric. A summary of the classes of metrics defined in the paper is shown on Fig. 1.

Section 6 is dedicated to the conclusion.

1.2. Notations and conventions

Manifolds Our manifold-related notations are summarized in Table 1. A chart $\varphi : \mathcal{U} \subset \mathcal{M} \longrightarrow \mathbb{R}^N$ provides a local basis of vectors $(\partial_1, ..., \partial_N)$ where $\partial_k = \frac{\partial}{\partial \varphi^k}$ is a short

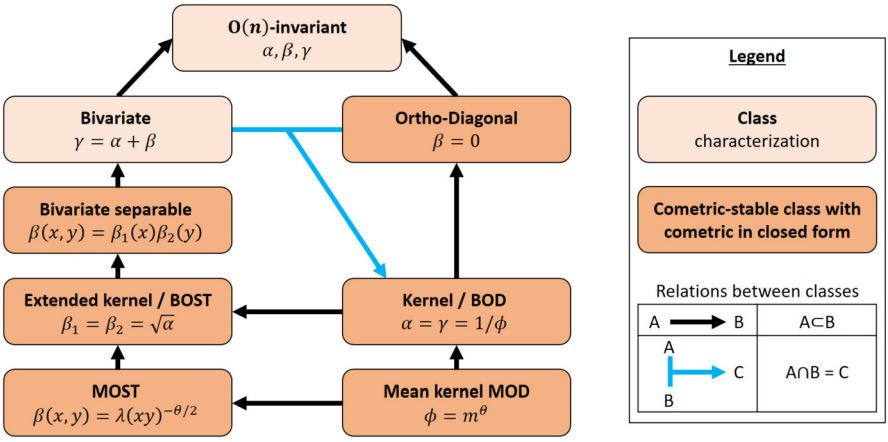

Fig. 1. Super-classes of kernel metrics.

Table 1
Notations in a manifold.

$T_x\mathcal{M}, T\mathcal{M}$	Tangent space at x, tangent bundle
$d_x f, df$	Differential of map f at x, differential of map f
f^*, f_*	Pullback via f, pushforward via f
$\dot{\gamma}$	Derivative of curve γ
g, G	Metric on $\mathrm{Sym}^+(n)$, metric on another space
d	Riemannian distance on $\mathrm{Sym}^+(n)$
∇	Levi-Civita connection
R	Curvature $R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$
$\gamma_{(\Sigma,X)}(t)$	Geodesic at time t with $\gamma(0) = \Sigma$ and $\dot{\gamma}(0) = X$
$\mathrm{Exp}, \mathrm{Log}$	Riemannian exponential and logarithm maps
$\Pi_{\gamma;\Sigma\to\Lambda}X$	Parallel transport of X along curve γ from Σ to Λ

notation defined for all differentiable maps $f : \mathcal{M} \longrightarrow \mathbb{R}$ and at each point $x \in \mathcal{U}$ by $(\partial_k f)_{|x} = \left.\frac{\partial(f\circ\varphi^{-1})}{\partial x^k}\right|_{\varphi(x)}$. A vector field X can be locally decomposed on this basis, $X = X^k\partial_k$, where $X^k : \mathcal{U} \longrightarrow \mathbb{R}$ are the coordinate functions of X and where we used Einstein's summation convention. As we deal with matrices in this paper, the coordinates often have two indices: $X = X^{ij}\partial_{ij}$.

Manifolds of matrices We denote the matrix spaces as shown in Table 2. The (i,j)-coefficient of a matrix M is denoted M_{ij}, $[M]_{ij}$ or $M(i,j)$ depending on the context. To build a matrix from its coefficients, we denote $M = [M_{ij}]_{1\leqslant i,j\leqslant n}$ or simply $M = [M_{ij}]_{i,j}$. We denote (C_{ij}) the canonical basis of matrices, $E_{ii} = C_{ii}$, $E_{ij} = \frac{1}{\sqrt{2}}(C_{ij} + C_{ji})$ and $F_{kl} = \frac{1}{2}(C_{kl} + C_{lk})$ for $i \neq j$ and $k, l \in \{1, ..., n\}$. The norms are denoted $\|M\|_1 = \sum_{i,j}|M_{ij}|$ and $\|M\|_2 = \sqrt{\mathrm{tr}(MM^\top)}$.

Table 2
Notations for matrix spaces.

Vector space of matrices		Manifold of matrices	
$\mathrm{Mat}(n)$	$n \times n$ real matrices	$\mathrm{GL}(n)$	General Linear group
		$\mathrm{GL}^+(n)$	Positive determinant
$\mathrm{Sym}(n)$	Real symmetric	$\mathrm{Sym}^+(n)$	Symmetric positive definite
$\mathrm{Skew}(n)$	Real skew-symmetric	$\mathrm{O}(n)$	Orthogonal group
		$\mathrm{SO}(n)$	Rotation group
$\mathrm{Diag}(n)$	Diagonal	$\mathrm{Diag}^+(n)$	Positive diagonal

The congruence action is the following action of the general linear group on matrices $\star : (A, M) \in \mathrm{GL}(n) \times \mathrm{Mat}(n) \longmapsto AMA^\top \in \mathrm{Mat}(n)$ which leaves stable the spaces of symmetric matrices and SPD matrices. Then, $A \in \mathrm{GL}(n)$ naturally acts on $M = M^{i_1 j_1, \ldots, i_q j_q} C_{i_1 j_1} \otimes \cdots \otimes C_{i_q j_q} \in \mathrm{Mat}(n)^{\otimes q}$ by:

$$A \star M = M^{i_1 j_1, \ldots, i_q j_q} (A \star C_{i_1 j_1}) \otimes \cdots \otimes (A \star C_{i_q j_q}) \in \mathrm{Mat}(n)^{\otimes q}.$$

$\mathrm{GL}(n)$ also acts by $\star$ on any Cartesian product $\prod_{i=1}^r \mathrm{Sym}(n)^{\otimes q_i}$ component-wise, especially on $\mathrm{Sym}(n)^p$.

Let $\mathcal{E}$ and $\mathcal{F}$ be two spaces on which $\mathrm{GL}(n)$ acts by $\star$. Let $\mathcal{G} \subseteq \mathrm{GL}(n)$ be a subgroup of $\mathrm{GL}(n)$. A map $f : \mathcal{E} \longrightarrow \mathcal{F}$ is:

- $\mathcal{G}$-equivariant if $f(A \star M) = A \star f(M)$ for all $A \in \mathcal{G}$, for all $M \in \mathcal{E}$,
- $\mathcal{G}$-invariant if $f(A \star M) = f(M)$ for all $A \in \mathcal{G}$, for all $M \in \mathcal{E}$.

In particular, a Riemannian metric $g : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \times \mathrm{Sym}(n) \longrightarrow \mathbb{R}$ (or an inner product) is $\mathcal{G}$-invariant if $g_{A\Sigma A^\top}(AXA^\top, AXA^\top) = g_\Sigma(X, X)$ for all $A \in \mathcal{G}$, $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in \mathrm{Sym}(n)$.

The symmetric group of order n is denoted by $\mathfrak{S}_n$ and the permutations by small greek letters $\sigma, \tau, \ldots$. The permutation matrix associated to the permutation σ, which sends any basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ to the permuted basis $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$, is denoted P_σ. We have $P_\sigma(i, j) = \delta_{\sigma(i),j}$ where δ is the Kronecker symbol. Given a matrix $M \in \mathrm{Mat}(n)$, we have $(P_\sigma^\top M P_\sigma)(i, j) = M(\sigma(i), \sigma(j))$.

The manifold of SPD matrices The manifold $\mathrm{Sym}^+(n)$ is an open set of the vector space of symmetric matrices $\mathrm{Sym}(n)$. Hence, the canonical immersion $\mathrm{id} : \mathrm{Sym}^+(n) \hookrightarrow \mathrm{Sym}(n)$ provides:

- An identification between the tangent space $T_\Sigma \mathrm{Sym}^+(n)$ and the vector space $\mathrm{Sym}(n)$ at any point $\Sigma \in \mathrm{Sym}^+(n)$ by $d_\Sigma \mathrm{id} : T_\Sigma \mathrm{Sym}^+(n) \xrightarrow{\sim} \mathrm{Sym}(n)$. Thus, any tangent vector $X \in T_\Sigma \mathrm{Sym}^+(n)$ is considered as a symmetric matrix: $X \equiv d_\Sigma \mathrm{id}(X) \in \mathrm{Sym}(n)$.
- A global chart $(\mathrm{id}, \mathrm{Sym}^+(n))$ of the manifold $\mathrm{Sym}^+(n)$, thus a global derivation $\partial_X Y = X^{ij}(\partial_{ij} Y^{kl})\partial_{kl}$ defined by derivation of coordinates in this global chart. More generally,

if $f : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$ is a diffeomorphism on its image, it provides a global derivation denoted ∂^f.

Another important tool is the matrix exponential $\exp(X) = \sum_{k=0}^{+\infty} \frac{X^k}{k!}$ which is a diffeomorphism between $\mathrm{Sym}(n)$ and $\mathrm{Sym}^+(n)$, and therefore its inverse, the symmetric matrix logarithm $\log : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$.

The spectral theorem ensures that symmetric matrices are orthogonally congruent to a diagonal matrix. If the symmetric matrix is SPD, then the diagonal matrix has positive elements on the diagonal. Most of the time in this paper, for an SPD matrix $\Sigma \in \mathrm{Sym}^+(n)$, we denote $\Sigma = PDP^\top$ one spectral decomposition with $P \in \mathrm{O}(n)$ and $D = \mathrm{diag}(d_1, ..., d_n) \in \mathrm{Diag}^+(n)$. When we consider tangent vectors $X, Y, ... \in T_\Sigma \mathrm{Sym}^+(n)$, we denote $X' = P^\top X P$ so that every matrix expressed in the orthogonal basis given by P is denoted with a prime: $X = PX'P^\top$, $Y = PY'P^\top$, ...

Products of symmetric matrices share two nice properties with symmetric matrices. First, if $X, Y \in \mathrm{Sym}(n)$, then $\mathrm{eig}(XY) \subset \mathbb{R}$ where eig denotes the set of complex eigenvalues. Second, if $\Sigma, \Lambda \in \mathrm{Sym}^+(n)$, then $\Sigma\Lambda$ has a unique square-root matrix that represents a positive definite self-adjoint endomorphism, it is denoted $(\Sigma\Lambda)^{1/2} = \sqrt{\Sigma\Lambda} = \Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2} = \Lambda^{-1/2}(\Lambda^{1/2}\Sigma\Lambda^{1/2})^{1/2}\Lambda^{1/2}$.

2. Preliminary concepts and results

2.1. Extending maps defined on diagonal matrices

Thanks to the spectral theorem, $\mathrm{O}(n)$-equivariant maps $f : \mathrm{Sym}^+(n) \longrightarrow \mathcal{F}$ are characterized by their values on positive diagonal matrices. A question that arises several times in this paper is: are we allowed to extend a map $f : \mathrm{Diag}^+(n) \longrightarrow \mathcal{F}$ into an $\mathrm{O}(n)$-equivariant map $f : \mathrm{Sym}^+(n) \longrightarrow \mathcal{F}$ by the formula $f(PDP^\top) = P \star f(D)$? To do so, we need to show that given two eigenvalue decompositions $\Sigma = PDP^\top = Q\Delta Q^\top$, we have $P \star f(D) = Q \star f(\Delta)$. Note that (Q, Δ) is highly constrained by (P, D). The following lemma gives explicitly the possible cases, hence it tells exactly what is to be checked in such an extension process. We omit the proof.

Lemma 2.1 *(Relation between two eigenvalue decompositions of an SPD matrix). Let $D, \Delta \in \mathrm{Diag}^+(n)$ and $P, Q \in \mathrm{O}(n)$ such that $PDP^\top = Q\Delta Q^\top$. Let $\tau \in \mathfrak{S}(n)$ be a permutation that orders the values of D decreasingly, i.e. such that $D = P_\tau \mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p}) P_\tau^\top$ with $\lambda_1 > ... > \lambda_p > 0$. Then, there exists permutation $\sigma \in \mathfrak{S}(n)$ and a block-diagonal orthogonal matrix $R = \mathrm{Diag}(R_1, ..., R_p) \in \mathrm{O}(n)$ with j-th block $R_j \in \mathrm{O}(m_j)$ such that $\Delta = P_\sigma^\top P_\tau^\top D P_\tau P_\sigma$ and $Q = P P_\tau R P_\sigma$.*

Proof. Δ is clearly a permutation of D so there exists $\sigma \in \mathfrak{S}(n)$ such that $\Delta = P_\sigma^\top P_\tau^\top D P_\tau P_\sigma$. Let $R = P_\tau^\top P^\top Q P_\sigma^\top$. Then $PDP^\top = Q\Delta Q^\top$ is equivalent to $\mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p}) R = R\,\mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p})$. Decomposing R by blocks, the

off-diagonal blocks have to be null since the λ_i's are distinct. Since $RR^\top = I_n$, the diagonal blocks are orthogonal. □

This result tells what is to be checked to extend $f : \mathrm{Diag}^+(n) \longrightarrow \mathcal{F}$. In the paper, we only need to extend tensorial maps $T : \mathrm{Diag}^+(n) \longrightarrow \mathrm{Sym}(n)^{\otimes q} \otimes (\mathrm{Sym}(n)^*)^{\otimes p}$ or equivalently $T : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n)^p \longrightarrow \mathrm{Sym}(n)^{\otimes q}$ for $p, q \in \mathbb{N}$. Hence, we state the result in this particular case though it is valid for $\mathcal{F}$.

Lemma 2.2 *(Spectral extension). Let $T : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n)^p \longrightarrow \mathrm{Sym}(n)^{\otimes q}$ be a map such that for all $D_0 = \mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p})$ with $\lambda_1 > ... > \lambda_p > 0$ and for all $X \in \mathrm{Sym}(n)^p$:*

(a) $T(D_0, X) = P_\sigma \star T(P_\sigma^\top D_0 P_\sigma, P_\sigma^\top \star X)$ for all permutations $\sigma \in \mathfrak{S}(n)$,
(b) $T(D_0, X) = R \star T(D_0, R^\top \star X)$ for all block-diagonal orthogonal matrices $R \in \mathrm{O}(n)$, $R = \mathrm{Diag}(R_1, ..., R_p)$ with $R_j \in \mathrm{O}(m_j)$.

Then, $T : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n)^p \longrightarrow \mathrm{Sym}(n)^{\otimes q}$ defined by $T(PDP^\top, X) := P \star T(D, P^\top \star X)$ extends T, with $D \in \mathrm{Diag}^+(n)$, $P \in \mathrm{O}(n)$ and $X \in \mathrm{Sym}(n)^p$.

Proof. Assume that $PDP^\top = Q\Delta Q^\top$. Then by Lemma 2.1, let $\sigma, \tau \in \mathfrak{S}(n)$ and R as in (b) such that $D_0 = P_\tau^\top D P_\tau = \mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p})$, $\Delta = P_\sigma^\top D_0 P_\sigma$ and $Q = PP_\tau R P_\sigma$. Then, by applying (a) with σ, (b) with R and (a) with τ, we easily see that $Q \star T(\Delta, Q^\top \star X) = P \star T(D, P^\top \star X)$. Thus $T : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n)^p \longrightarrow \mathrm{Sym}(n)^{\otimes q}$ is well defined. □

In practice in the paper, we use Lemma 2.2 for:

- $p = 0, q = 1$ for $f : \mathrm{Diag}^+(n) \longrightarrow \mathrm{Sym}(n)$ in Section 2.2,
- $p = 1, q = 1$ for $\Phi : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ in Section 4.2.3,
- $p = 2, q = 0$ for $g : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n) \times \mathrm{Sym}(n) \longrightarrow \mathbb{R}$ in Section 5.2.

2.2. Univariate maps

We apply Lemma 2.2 to a real function $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$, extended to positive diagonal matrices $f : \mathrm{Diag}^+(n) \longrightarrow \mathrm{Diag}(n)$ by $f(\mathrm{Diag}(d_1, ..., d_n)) := \mathrm{Diag}(f(d_1), ..., f(d_n))$.

(a) Since f is defined component-wise, we have $f(D) = P_\sigma\, f(P_\sigma^\top D P_\sigma) P_\sigma^\top$.
(b) As $f(\lambda I_{m_j}) = f(\lambda) I_{m_j}$, the matrix $Rf(D)R^\top$ is a block diagonal matrix with j-th block $f(\lambda_j) R_j R_j^\top = f(\lambda_j) I_{m_j}$, which corresponds to $f(D)$'s j-th block so $R\, f(D)\, R^\top = f(D)$.

Therefore f can be extended into an $\mathrm{O}(n)$-equivariant map $f : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$ by $f(PDP^\top) = Pf(D)P^\top$. This extension is called the functional calculus of f in

Functional Analysis. In this paper, we call it a univariate map. The symmetric matrix logarithm $\log : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$, the power diffeomorphisms $\mathrm{pow}_p : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}^+(n)$ with $p \neq 0$ or the constant map $\mathrm{pow}_0 : \Sigma \in \mathrm{Sym}^+(n) \longmapsto I_n \in \mathrm{Sym}(n)$ are examples of univariate maps.

Definition 2.1 *(Univariate maps).* A univariate map is the extension of a real function $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$ into an $\mathrm{O}(n)$-equivariant map $f : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$ by the equality $f(PDP^\top) = P\,\mathrm{Diag}(f(d_1), ..., f(d_n))\,P^\top$. Moreover [5, Theorem V.3.3], if $f \in \mathcal{C}^1(\mathbb{R}^+)$, then its extension f is differentiable and the differential $df : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ is $\mathrm{O}(n)$-equivariant, thus it is characterized by its values at diagonal matrices $D \in \mathrm{Diag}^+(n)$, given by:

$$\forall X \in \mathrm{Sym}(n), [d_D f(X)]_{ij} = f^{[1]}(d_i, d_j) X_{ij}, \tag{1}$$

where $f^{[1]}$ is the first divided difference defined below. Thus, a $\mathcal{C}^1$-diffeomorphism $f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is extended into a diffeomorphism $f : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}^+(n)$.

Definition 2.2 *(First divided difference).* [5] Let $f \in \mathcal{C}^1(\mathbb{R}^+)$. The first divided difference of f is the continuous symmetric map $f^{[1]} : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}$ defined for all $x, y \in \mathbb{R}$ by:

$$f^{[1]}(x,y) = \left\{ \begin{array}{ccc} \frac{f(x)-f(y)}{x-y} & \text{if} & x \neq y \\ f'(x) & \text{if} & x = y \end{array} \right\}. \tag{2}$$

2.3. $\mathrm{O}(n)$*-invariant inner products on symmetric matrices*

To characterize the $\mathrm{O}(n)$-invariant metrics on SPD matrices, an appropriate starting point is the characterization of $\mathrm{O}(n)$-invariant inner products on the tangent space, i.e. on symmetric matrices. The following theorem states that such inner products form a two-parameter family indexed by a Scaling factor $\alpha > 0$ and a Trace factor $\beta > -\alpha/n$.

Theorem 2.1 *(Characterization of* $\mathrm{O}(n)$*-invariant inner products on symmetric matrices). Let* $\langle \cdot | \cdot \rangle : \mathrm{Sym}(n) \times \mathrm{Sym}(n) \longrightarrow \mathbb{R}$ *be an inner product on symmetric matrices. It is* $\mathrm{O}(n)$*-invariant if and only if there exists* $(\alpha, \beta) \in \mathbf{ST} := \{(\alpha, \beta) \in \mathbb{R}^2 | \min(\alpha, \alpha + n\beta) > 0\}$ *such that:*

$$\forall X \in \mathrm{Sym}(n), \langle X | X \rangle = \alpha\,\mathrm{tr}(X^2) + \beta\,\mathrm{tr}(X)^2. \tag{3}$$

Moreover, the linear isometry that pulls the Frobenius inner product back onto this one is $F_{p,q}(X) = q\,X + \frac{p-q}{n}\mathrm{tr}(X) I_n$ *with* $p = \sqrt{\alpha + n\beta}$ *and* $q = \sqrt{\alpha}$.

There are several proofs of this elementary result. We give one based on the following lemma because we reuse it to characterize $\mathrm{O}(n)$-invariant metrics on SPD matrices. This

lemma gives the characterization of inner products on symmetric matrices which are respectively invariant under two subgroups of $\mathrm{O}(n)$:

(a) the group $\mathcal{D}^{\pm}(n) := \{\varepsilon = \mathrm{Diag}(\pm 1, ..., \pm 1)\} \cong \{-1, +1\}^n$ of diagonal matrices taking their diagonal values in $\{-1, +1\}$,
(b) the group $\mathfrak{S}^{\pm}(n) := \{\varepsilon P_\sigma \in \mathrm{Mat}(n) | (\varepsilon, \sigma) \in \mathcal{D}^{\pm}(n) \times \mathfrak{S}(n)\} \cong \mathcal{D}^{\pm}(n) \times \mathfrak{S}(n)$ of signed permutation matrices.

Lemma 2.3 *(Characterization of inner products on symmetric matrices invariant under $\mathcal{D}^{\pm}(n)$ or $\mathfrak{S}^{\pm}(n)$). Let $\langle \cdot | \cdot \rangle : \mathrm{Sym}(n) \times \mathrm{Sym}(n) \longrightarrow \mathbb{R}$ be an inner product on symmetric matrices.*

(a) It is $\mathcal{D}^{\pm}(n)$-invariant if and only if there exist $\frac{n(n-1)}{2}$ positive real numbers $\alpha_{ij} = \alpha_{ji} > 0$ for $i \neq j$ and a matrix $S \in \mathrm{Sym}^+(n)$ such that:

$$\forall X \in \mathrm{Sym}(n), \langle X | X \rangle = \sum_{i \neq j} \alpha_{ij} X_{ij}^2 + \sum_{i,j} S_{ij} X_{ii} X_{jj}. \tag{4}$$

(b) It is $\mathfrak{S}^{\pm}(n)$-invariant if and only if there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ with $\alpha > 0$, $\gamma > \beta$ and $\gamma + (n-1)\beta > 0$ such that:

$$\forall X \in \mathrm{Sym}(n), \langle X | X \rangle = \gamma \sum_{i=1}^{n} X_{ii}^2 + \alpha \sum_{i \neq j} X_{ij}^2 + \beta \sum_{i \neq j} X_{ii} X_{jj}. \tag{5}$$

Proof of Lemma 2.3.

(a) We write $\langle X | X \rangle = \sum_{i,j,k,l} a_{ij,kl} X_{ij} X_{kl}$ a general inner product. Note that $a_{ij,kl} = a_{ji,kl} = a_{ji,lk} = a_{ij,lk}$ by symmetry of X and $a_{ij,kl} = a_{kl,ij}$ by symmetry of the inner product. We use the invariance under the matrix $\varepsilon_m \in \mathcal{D}^{\pm}(n)$ with -1 on the m-th component and 1 elsewhere, for $m \in \{1, ..., n\}$. We denote $\mathcal{P}$ XOR $\mathcal{Q} = 1$ if the "exclusive or" between propositions $\mathcal{P}$ and $\mathcal{Q}$ holds, and otherwise $\mathcal{P}$ XOR $\mathcal{Q} = 0$. Thus, we have $[\varepsilon_m X \varepsilon_m]_{ij} = (-1)^{(i=m)\mathrm{XOR}(j=m)} X_{ij}$ and $[\varepsilon_m X \varepsilon_m]_{ij} [\varepsilon_m X \varepsilon_m]_{kl} = \theta_{ijklm} X_{ij} X_{kl}$ with $\theta_{ijklm} = (-1)^{[(i=m)\mathrm{XOR}(j=m)]\mathrm{XOR}[(k=m)\mathrm{XOR}(l=m)]} \in \{-1, 1\}$. Then the equality $\langle X | X \rangle = \langle \varepsilon_m X \varepsilon_m | \varepsilon_m X \varepsilon_m \rangle$ leads to $a_{ij,kl} = \theta_{ijklm} a_{ij,kl}$. Therefore, if there exists $m \in \{1, ..., n\}$ such that $\theta_{ijklm} = -1$, then $a_{ij,kl} = 0$. One can easily show that $\theta_{ijklm} = -1$ if and only if m equals exactly one or exactly three index(es) among i, j, k, l. There exists such an m if:

- $\mathrm{card}(\{i, j, k, l\}) = 4$, i.e. i, j, k, l are distinct,
- $\mathrm{card}(\{i, j, k, l\}) = 3$,
- $\mathrm{card}(\{i, j, k, l\}) = 2$ and three of them are equal.

Thus we are left with $\langle X|X\rangle = \sum_{i<j} 4a_{ij,ij}X_{ij}^2 + \sum_{i,j} a_{ii,jj}X_{ii}X_{jj}$. Then, we get the expression (4) by denoting $\alpha_{ij} = 2a_{ij,ij}$ and $S_{ij} = a_{ii,jj} = S_{ji}$. Since the quadratic form splits into two quadratic forms defined on supplementary vector spaces (off-diagonal and diagonal terms), it is positive definite if and only if these two quadratic forms are positive definite, i.e. $\alpha_{ij} > 0$ for all $i \neq j$ and S is positive definite. Conversely, Equation (4) clearly defines $\mathcal{D}^{\pm}(n)$-invariant inner products.

(b) A $\mathfrak{S}^{\pm}(n)$-invariant inner product on symmetric matrices is $\mathcal{D}^{\pm}(n)$-invariant so it is of the form of Equation (4). Since it is invariant under permutations, we have $\alpha_{ij} = \alpha_{kl} =: \alpha$ and $S_{ij} = S_{kl} =: \beta$ for all $i \neq j$ and $k \neq l$ and $S_{ii} = S_{jj} =: \gamma$ for all i, j. Under these notations, Equation (4) becomes Equation (5). Since $S = (\gamma - \beta)\, I_n + \beta \mathbb{1}\mathbb{1}^\top$, then $S \in \mathrm{Sym}^+(n)$ if and only if $\gamma - \beta > 0$ and $\gamma - \beta + n\beta > 0$ as expected. Conversely, Equation (5) clearly defines $\mathfrak{S}^{\pm}(n)$-invariant inner products. □

Proof of Theorem 2.1. An $\mathrm{O}(n)$-invariant inner product on symmetric matrices is $\mathfrak{S}^{\pm}(n)$-invariant so it is of the form of Equation (5). We define the rotation matrix $R = \begin{pmatrix} R_{\pi/4} & 0 \\ 0 & I_{n-2} \end{pmatrix} \in \mathrm{O}(n)$ with $R_{\pi/4} = \frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in \mathrm{O}(2)$ and we apply it to the matrix $X = \begin{pmatrix} M & Y \\ Y^\top & Z \end{pmatrix} \in \mathrm{Sym}(n)$ with $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathrm{Sym}(2)$. Since $R_{\pi/4} M R_{\pi/4}^\top = \frac{1}{2}\begin{pmatrix} a+c+2b & c-a \\ c-a & a+c-2b \end{pmatrix}$, the coefficient of b^2 in $\langle X|X\rangle$ in Equation (5) is 2α and the coefficient of b^2 in $\langle RXR^\top|RXR^\top\rangle$ is $2\gamma - 2\beta$. Hence by invariance, $\gamma = \alpha + \beta$ and the positivity condition becomes $\alpha > 0$ and $\alpha + n\beta > 0$. Conversely, Equation (3) clearly defines $\mathrm{O}(n)$-invariant inner products. □

3. Main $\mathrm{O}(n)$-invariant metrics on SPD matrices with new formulae

The goal of this section is to describe the main $\mathrm{O}(n)$-invariant metrics on SPD matrices that can be found in the literature, namely the Euclidean (abbreviated 'E', Section 3.1), the Log-Euclidean ('LE', Section 3.2), the Affine-invariant ('A', Section 3.3), the Bures-Wasserstein ('BW', Section 3.4) and the Bogoliubov-Kubo-Mori ('BKM', Section 3.5) metrics. For each metric, we give a short explanation on the way it was introduced, some useful references and a synthetic table that summarizes its fundamental Riemannian operations: squared distance, Levi-Civita connection, curvature, geodesics, logarithm map, parallel transport map (abbreviated 'PT map').

Our contributions are (1) the synthesis of many results scattered in the literature especially for the Bures-Wasserstein metric, (2) the complete formula of the sectional curvature of the affine-invariant metric, (3) the new formula of the parallel transport between commuting matrices and new expressions of the Levi-Civita connection, the curvature and the parallel transport equation of the Bures-Wasserstein metric.

Table 3
Riemannian operations of $\mathrm{O}(n)$-invariant Euclidean metrics on SPD matrices.

Metric	$g_\Sigma(X,X) = \alpha\|X\|_2^2 + \beta\,\mathrm{tr}(X)^2$
Sq. dist.	$d(\Sigma,\Lambda)^2 = \alpha\|\Lambda-\Sigma\|_2^2 + \beta(\mathrm{tr}(\Lambda)-\mathrm{tr}(\Sigma))^2$
Levi-Civita	$\nabla_X Y = \partial_X Y$
Curvature	$R = 0$
Geodesics	$\gamma_{(\Sigma,X)}(t) = \Sigma + tX$ for $t \in I$ where I depends on $\lambda_{\min} = \min \mathrm{eig}(\Sigma^{-1}X)$ and $\lambda_{\max} = \max \mathrm{eig}(\Sigma^{-1}X)$ as follows: • If $\lambda_{\min} < 0 < \lambda_{\max}$, then $I = (-1/\lambda_{\max}, -1/\lambda_{\min})$. • If $0 \leqslant \lambda_{\min}$, then $I = (-1/\lambda_{\max}, +\infty)$. • If $\lambda_{\max} \leqslant 0$, then $I = (-\infty, -1/\lambda_{\min})$.
Logarithm	$\mathrm{Log}_\Sigma(\Lambda) = \Lambda - \Sigma$
PT map	Does not depend on the curve: $\Pi_{\Sigma\to\Lambda} : \begin{cases} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow & T_\Lambda\mathrm{Sym}^+(n) \\ X & \longmapsto & (d_\Lambda \mathrm{id})^{-1}(d_\Sigma \mathrm{id}(X)) \equiv X \end{cases}$

3.1. $\mathrm{O}(n)$*-invariant Euclidean metrics*

A Euclidean metric on SPD matrices is the pullback of an inner product $\langle\cdot|\cdot\rangle$ on symmetric matrices by the canonical immersion $\mathrm{id} : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$. As we know $\mathrm{O}(n)$-invariant inner products on symmetric matrices from Theorem 2.1, we know all the $\mathrm{O}(n)$-invariant Euclidean metrics on SPD matrices.

Definition 3.1 *($\mathrm{O}(n)$-invariant Euclidean metrics on SPD matrices).* An $\mathrm{O}(n)$*-invariant Euclidean metric* on SPD matrices is a Riemannian metric of the following form for all $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in \mathrm{Sym}(n)$:

$$g_\Sigma^{\mathrm{E}(\alpha,\beta)}(X,X) = \alpha\,\mathrm{tr}(X^2) + \beta\,\mathrm{tr}(X)^2, \tag{6}$$

with $(\alpha,\beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. Its Riemannian operations are detailed in Table 3.

3.2. $\mathrm{O}(n)$*-invariant log-Euclidean metrics*

A log-Euclidean metric on SPD matrices [2] is the pullback of an inner product $\langle\cdot|\cdot\rangle$ on symmetric matrices by the symmetric matrix logarithm $\log : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}(n)$. Hence the SPD manifold endowed with the log-Euclidean metric is isometric to a Euclidean space, thus geodesically complete. From Theorem 2.1 and the fact that $d\log : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ is $\mathrm{O}(n)$-equivariant, we know all the $\mathrm{O}(n)$-invariant log-Euclidean metrics.

Definition 3.2 *($\mathrm{O}(n)$-invariant log-Euclidean metrics on SPD matrices).* An $\mathrm{O}(n)$*-invariant log-Euclidean metric* on SPD matrices is a Riemannian metric of the following form for all $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in \mathrm{Sym}(n)$:

Table 4
Riemannian operations of $\mathrm{O}(n)$-invariant log-Euclidean metrics on SPD matrices.

Metric	$g_\Sigma(X,X) = \alpha \|d_\Sigma \log(X)\|_2^2 + \beta \operatorname{tr}(\Sigma^{-1}X)^2$
Sq. dist.	$d(\Sigma,\Lambda)^2 = \alpha \|\log\Lambda - \log\Sigma\|_2^2 + \beta \log(\det(\Lambda)/\det(\Sigma))^2$
Levi-Civita	$\nabla_X Y = \partial_X^{\log} Y$
Curvature	$R = 0$
Geodesics	$\forall t \in \mathbb{R}, \gamma_{(\Sigma,X)}(t) = \exp(\log(\Sigma) + t\, d_\Sigma \log(X))$
Logarithm	$\mathrm{Log}_\Sigma(\Lambda) = (d_\Sigma \log)^{-1}(\log\Lambda - \log\Sigma)$
PT map	Does not depend on the curve: $\Pi_{\Sigma\to\Lambda} : \begin{cases} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow & T_\Lambda\mathrm{Sym}^+(n) \\ X & \longmapsto & (d_\Lambda \log)^{-1}(d_\Sigma \log(X)) \end{cases}$

$$g_\Sigma^{\mathrm{LE}(\alpha,\beta)}(X,X) = \alpha \operatorname{tr}(d_\Sigma \log(X)^2) + \beta \operatorname{tr}(\Sigma^{-1}X)^2, \tag{7}$$

with $(\alpha,\beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. Moreover, this metric is the pullback of the Frobenius log-Euclidean metric ($\alpha = 1$ and $\beta = 0$) by the isometry $f_{p,q} : \Sigma \in \mathrm{Sym}^+(n) \longmapsto \exp(F_{p,q}(\log\Sigma)) = \det(\Sigma)^{\frac{p-q}{n}}\Sigma^q \in \mathrm{Sym}^+(n)$ with $p = \sqrt{\alpha + n\beta}$ and $q = \sqrt{\alpha}$, where $F_{p,q}$ was defined in Theorem 2.1. Its Riemannian operations are detailed in Table 4.

3.3. Affine-invariant metrics

Affine-invariant metrics were introduced in many different ways. We adopt here the most recent viewpoint [22], which underlies the term "affine-invariant". Consider SPD matrices $\Sigma \in \mathrm{Sym}^+(n)$ as covariance matrices of a random vector $X \in \mathbb{R}^n$, namely $\Sigma = \frac{1}{n}\mathbb{E}\left((X-\bar{X})(X-\bar{X})^\top\right)$ with $\bar{X} = \mathbb{E}(X)$, where $\mathbb{E}$ denotes the expectation. Define the affine action on vectors $((A,B),X) \in (\mathrm{GL}(n) \ltimes \mathbb{R}^n) \times \mathbb{R}^n \longmapsto AX + B \in \mathbb{R}^n$. Then, the induced action on SPD matrices is $((A,B),\Sigma) \in (\mathrm{GL}(n) \ltimes \mathbb{R}^n) \times \mathrm{Sym}^+(n) \longmapsto A\Sigma A^\top \in \mathrm{Sym}^+(n)$. It is simply the congruence action of $\mathrm{GL}(n)$ on matrices. Hence an affine-invariant metric on SPD matrices simply designates a $\mathrm{GL}(n)$-invariant metric.

Historically, Siegel introduced a metric on the half space $\mathcal{S} = \{X + i\Sigma |\, X \in \mathrm{Sym}(n), \Sigma \in \mathrm{Sym}^+(n)\}$ which is invariant under the action of the symplectic group [27]. As a consequence, the restriction of this metric to SPD matrices by the immersion $\Sigma \in \mathrm{Sym}^+(n) \hookrightarrow i\Sigma \in \mathcal{S}$ was proved to be invariant under $\mathrm{GL}(n)$ and under inversion and to provide a Riemannian homogeneous structure to $\mathrm{Sym}^+(n)$. The expression of this metric is $g_\Sigma(X,Y) = \operatorname{tr}(\Sigma^{-1}X\Sigma^{-1}Y)$.

Rao considered the Fisher information of a family of densities as a Riemannian metric on the space of parameters [26] and Skovgaard detailed all the properties of the Fisher-Rao metric of the family of multivariate Gaussian densities [28]. By restriction to the family of *centered* multivariate Gaussian densities, we get the same metric as Siegel's scaled by a factor 1/2, namely $g_\Sigma(X,Y) = \frac{1}{2}\operatorname{tr}(\Sigma^{-1}X\Sigma^{-1}Y)$. In addition, Amari and

Nagaoka stated that the canonical immersion $\mathrm{id} : \Sigma \in \mathrm{Sym}^+(n) \longmapsto \Sigma \in \mathrm{Sym}(n)$ and the inversion $\mathrm{inv} : \Sigma \in \mathrm{Sym}^+(n) \longmapsto \Sigma^{-1} \in \mathrm{Sym}(n)$ give two dual coordinate systems with respect to this metric [1].

Between 2005 and 2007, this metric was used in many computational methods for Diffusion Tensor Imaging [23,13,9,19,4], in functional MRI [34] and in Brain-Computer Interfaces [3]. It was claimed to be the unique affine-invariant metric. However, Pennec showed that $\mathrm{GL}(n)$-invariant metrics are characterized by $\mathrm{O}(n)$-invariant inner products on the tangent space at I_n, that is on symmetric matrices. Hence from Theorem 2.1, there is actually a two-parameter family of affine-invariant metrics [22].

Definition 3.3 *(Affine-invariant metrics on SPD matrices).* An *affine-invariant metric* on SPD matrices is a $\mathrm{GL}(n)$-invariant Riemannian metric. It is of the following form for all $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in \mathrm{Sym}(n)$:

$$g_\Sigma^{\mathrm{A}(\alpha,\beta)}(X,X) = \alpha \,\mathrm{tr}((\Sigma^{-1}X)^2) + \beta \,\mathrm{tr}(\Sigma^{-1}X)^2, \tag{8}$$

with $(\alpha,\beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. The Fisher-Rao metric often refers to the affine-invariant metric with $(\alpha,\beta) = (1/2, 0)$. Moreover, given $\alpha > 0$, this metric is the pullback of the affine-invariant metric with $\beta = 0$ by the isometry $f_{p,1} : \Sigma \in \mathrm{Sym}^+(n) \longmapsto \det(\Sigma)^{\frac{p-1}{n}}\Sigma \in \mathrm{Sym}^+(n)$ with $p = \sqrt{\frac{\alpha+n\beta}{\alpha}}$.

The following proposition details the characteristics of homogeneity and symmetry of these Riemannian metrics. The Riemannian operations, essentially due to Skovgaard [28], are detailed in Table 5. The second term of the sectional curvature is part of our contributions as it seems to be forgotten in [28].

Proposition 3.1 *(Riemannian symmetric structure of the affine-invariant metric). The Riemannian manifold* $(\mathrm{Sym}^+(n), g^{\mathrm{A}(\alpha,\beta)})$ *is a Riemannian symmetric space, hence it is geodesically complete. The underlying homogeneous space is* $\mathrm{GL}^+(n)/\mathrm{SO}(n)$ *and* $g^{\mathrm{A}(\alpha,\beta)}$ *is a quotient metric obtained by the submersion* $\pi : A \in \mathrm{GL}^+(n) \longmapsto AA^\top \in \mathrm{Sym}^+(n)$ *from the left-invariant metric* $G_A(M,M) = 4\alpha \,\mathrm{tr}(A^{-1}M(A^{-1}M)^\top) + 4\beta \,\mathrm{tr}(A^{-1}M)^2$ *for* $A \in \mathrm{GL}^+(n)$ *and* $M \in T_A\mathrm{GL}^+(n)$*. The symmetries are* $s_\Sigma : \Lambda \in \mathrm{Sym}^+(n) \longmapsto \Sigma\Lambda^{-1}\Sigma \in \mathrm{Sym}^+(n)$*.*

Proof of sectional curvature in Table 5. Firstly, we compute the sectional curvature of the affine-invariant metrics for $\beta = 0$ at $\Sigma \in \mathrm{Sym}^+(n)$ in the orthonormal basis $(\Sigma^{1/2}E_{ij}\Sigma^{1/2})_{1\leqslant i\leqslant j\leqslant n}$, with E_{ii}, E_{ij} for $i \neq j$ defined by $E_{ii}(k,l) = \delta_{ik}\delta_{il}$ and $E_{ij}(k,l) = \frac{\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}}{\sqrt{2}}$. As $\kappa_\Sigma(X,Y) = \frac{R_\Sigma(X,Y,X,Y)}{\|X\|^2\|Y\|^2-\langle X|Y\rangle^2}$, we have $\kappa_\Sigma(\Sigma^{1/2}E_{ij}\Sigma^{1/2}, \Sigma^{1/2}E_{kl}\Sigma^{1/2}) = \frac{1}{2\alpha}\mathrm{tr}((E_{ij}E_{kl})^2 - (E_{ij}E_{kl})(E_{ij}E_{kl})^\top)$ so we only need to compute a few expressions. In the following equalities, when an elementary matrix E has two different indexes, they are assumed to be distinct:

Table 5
Riemannian operations of affine-invariant metrics on SPD matrices.

Metric	$g_\Sigma(X,X) = \alpha\|\Sigma^{-1}X\|_2^2 + \beta\,\mathrm{tr}(\Sigma^{-1}X)^2$
Sq. dist.	$d(\Sigma,\Lambda)^2 = \alpha\|\log(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})\|_2^2 + \beta\,\log(\det(\Sigma^{-1}\Lambda))^2$
Levi-Civita	$(\nabla_X Y)_{\|\Sigma} = (\partial_X Y)_{\|\Sigma} - \frac{1}{2}(X\Sigma^{-1}Y + Y\Sigma^{-1}X)$
Curvature	The sectional curvature $\kappa \in [-1/2\alpha, 0]$. More precisely, the Riemann and sectional curvatures are: $R_\Sigma(X,Y,Z,T) = \frac{\alpha}{2}(X\Sigma^{-1}Y\Sigma^{-1}(Z\Sigma^{-1}T - T\Sigma^{-1}Z)\Sigma^{-1})$ $\kappa_\Sigma(\Sigma^{1/2}E_{ii}^\beta\Sigma^{1/2}, \Sigma^{1/2}E_{ij}^\beta\Sigma^{1/2}) = -1/4\alpha$ for $i \neq j$ $\kappa_\Sigma(\Sigma^{1/2}E_{ij}^\beta\Sigma^{1/2}, \Sigma^{1/2}E_{ik}^\beta\Sigma^{1/2}) = -1/8\alpha$ for $i \neq j \neq k \neq i$ where $E_{ij}^\beta = E_{ij} - \frac{1-p}{np}\delta_{ij}I_n$. Other terms are null.
Geodesics	$\forall t \in \mathbb{R}, \gamma_{(\Sigma,X)}(t) = \Sigma^{1/2}\exp(t\,\Sigma^{-1/2}X\Sigma^{-1/2})\Sigma^{1/2}$
Logarithm	$\mathrm{Log}_\Sigma(\Lambda) = \Sigma^{1/2}\log(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})\Sigma^{1/2}$
PT map	Depends on the curve. Along a geodesic: $\Pi_{\Sigma\to\Lambda} : \begin{cases} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow & T_\Lambda\mathrm{Sym}^+(n) \\ X & \longmapsto & (\Lambda\Sigma^{-1})^{1/2}X(\Sigma^{-1}\Lambda)^{1/2} \end{cases}$

- $E_{ii}E_{jj} = \delta_{ij}C_{ij}$ hence $\|E_{ii}E_{jj}\|_2^2 = \delta_{ij}$,
- $E_{ii}E_{jk} = \frac{1}{\sqrt{2}}(\delta_{ij}C_{ik} + \delta_{ik}C_{ij})$ hence $\|E_{ii}E_{jk}\|_2^2 = \frac{1}{2}(\delta_{ij} + \delta_{ik})$,
- $E_{ij}E_{kl} = \frac{1}{2}(\delta_{jk}C_{il} + \delta_{ik}C_{jl} + \delta_{jl}C_{ik} + \delta_{il}C_{jk})$
  hence $\|E_{ij}E_{kl}\|_2^2 = \frac{1}{4}(\delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl})$,
- $(E_{ii}E_{jj})^2 = \delta_{ij}C_{ij}$ hence $\mathrm{tr}((E_{ii}E_{jj})^2) = \delta_{ij}$,
- $(E_{ii}E_{jk})^2 = 0$ hence $\mathrm{tr}((E_{ii}E_{jk})^2) = 0$,
- $(E_{ij}E_{kl})^2 = \frac{1}{4}(\delta_{jk}\delta_{il}(C_{il} + C_{jk}) + \delta_{jl}\delta_{ik}(C_{ik} + C_{jl}))$,
  hence $\mathrm{tr}((E_{ij}E_{kl})^2) = \frac{1}{2}(\delta_{jk}\delta_{il} + \delta_{jl}\delta_{ik})$.
- $\kappa_{I_n}(E_{ii}, E_{jj}) = 0$,
- $\kappa_{I_n}(E_{ii}, E_{jk}) = -\frac{1}{4\alpha}(\delta_{ij} + \delta_{ik})$,
- $\kappa_{I_n}(E_{ij}, E_{kl}) = -\frac{1}{8\alpha}((\delta_{ik} - \delta_{jl})^2 + (\delta_{il} - \delta_{jk})^2)$.

Hence the non null terms are $\kappa_{I_n}(E_{ii}, E_{ij}) = -\frac{1}{4\alpha}$ and $\kappa_{I_n}(E_{ij}, E_{ik}) = -\frac{1}{8\alpha}$.
Secondly, for $\beta \neq 0$, we use the isometry $f_{p,1}$: the values are the same if we replace $\Sigma^{1/2}E_{ij}\Sigma^{1/2}$ by $(d_\Sigma f_{p,1})^{-1}(f_{p,1}(\Sigma)^{1/2}E_{ij}f_{p,1}(\Sigma)^{1/2}) = \Sigma^{1/2}E_{ij}^\beta\Sigma^{1/2}$.

To prove that $\kappa \in [-1/2\alpha, 0]$, it suffices to note that for normed and orthogonal $X, Y \in \mathrm{Sym}(n)$, we have $\kappa_{I_n}(X,Y) = -\frac{1}{4\alpha}\|[X,Y]\|_2^2$. Diagonalizing $X = P\Delta P^\top$ and denoting $Z = P^\top YP$, from $(d_i - d_j)^2 \leqslant 2(d_i^2 + d_j^2) \leqslant 2\|D\|^2$, we get $\kappa_{I_n}(X,Y) = \kappa_{I_n}(\Delta, Z) = -\frac{1}{4\alpha}\sum_{i\neq j}(d_i - d_j)^2Z_{ij}^2 \geqslant -\frac{1}{2\alpha}\|D\|^2\|Z\|^2 = -\frac{1}{2\alpha}$. This bound is reached for $X = \frac{1}{\sqrt{2}}(E_{ii} - E_{jj})$ and $Y = E_{ij}$. □

Another metric that also provides a Riemannian symmetric structure on $\mathrm{Sym}^+(n)$ was used in [29,36]. It was introduced directly by the quotient structure detailed in Proposition 3.1 but with the submersion $\sqrt{\pi} : A \in \mathrm{GL}^+(n) \longmapsto \sqrt{AA^\top} \in \mathrm{Sym}^+(n)$ based on the polar decomposition of A (and without the coefficient 4). We called it the polar-affine metric in [32]. It is $\mathrm{GL}(n)$-invariant with respect to the action $(A, \Sigma) \in \mathrm{GL}(n) \times$

$\mathrm{Sym}^+(n) \longmapsto \sqrt{A\Sigma^2 A^\top} \in \mathrm{Sym}^+(n)$. Hence it is $\mathrm{O}(n)$-invariant in the usual sense. It is the pullback metric of the affine-invariant metric via the square diffeomorphism $\mathrm{pow}_2 : \Sigma \longmapsto \Sigma^2$ [32].

3.4. Bures-Wasserstein metric

The L^2-Wasserstein distance between multivariate centered Gaussian distributions is given by $d(\Sigma,\Lambda)^2 = \mathrm{tr}\Sigma + \mathrm{tr}\Lambda - 2\mathrm{tr}((\Sigma\Lambda)^{1/2})$. It corresponds to the Procrustes distance between square-root matrices, namely $d(\Sigma,\Lambda)^2 = \inf_{U\in\mathrm{O}(n)} \|\Sigma^{1/2} - \Lambda^{1/2}U\|^2_{\mathrm{Frob}}$. The second order approximation of this squared distance defines a Riemannian metric called the Bures metric (or the Helstrom metric) in quantum physics. All these viewpoints are explained in details with modern notations in [6]. In particular, the expression of the Riemannian metric is derived in [6] and we take it as a definition.

Definition 3.4 *(Bures-Wasserstein metric).* The *Bures-Wasserstein metric* is the Riemannian metric associated to the Bures-Wasserstein distance. It is $\mathrm{O}(n)$-invariant and given an eigenvalue decomposition $\Sigma = PDP^\top \in \mathrm{Sym}^+(n)$ with $P \in \mathrm{O}(n)$ and $D = \mathrm{diag}(d_1, ..., d_n)$ and $X = PX'P^\top$, its expression is:

$$g^{\mathrm{BW}}_\Sigma(X,X) = g^{\mathrm{BW}}_D(X',X') = \frac{1}{2}\sum_{i,j}\frac{1}{d_i+d_j}X'^2_{ij}. \tag{9}$$

The Bures-Wasserstein metric can also be expressed by means of the linear map $\mathcal{S}_\Sigma : \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ implicitly defined by the Sylvester equation $X = \Sigma\mathcal{S}_\Sigma(X) + \mathcal{S}_\Sigma(X)\Sigma$ for $X \in \mathrm{Sym}(n)$. More explicitly with the previous notations, we have $\mathcal{S}_\Sigma(X) = P\left[\frac{X'_{ij}}{d_i+d_j}\right]_{i,j}P^\top$. Then we have $g^{\mathrm{BW}}_\Sigma(X,Y) = \frac{1}{2}\mathrm{tr}(X\mathcal{S}_\Sigma(Y)) = \mathrm{tr}(\mathcal{S}_\Sigma(X)\Sigma\mathcal{S}_\Sigma(Y))$, where $X, Y \in T_\Sigma\mathrm{Sym}^+(n)$ are canonically identified with $d_\Sigma\mathrm{id}(X), d_\Sigma\mathrm{id}(Y) \in \mathrm{Sym}(n)$, as explained in the introduction. This is a common expression in recent papers [16,21]. However, in [31] which is a reference paper on the Bures-Wasserstein metric, Takatsu gives the expression $g_\Sigma(X,Y) = \mathrm{tr}(\mathbf{X}\Sigma\mathbf{Y})$. The trick comes from the identification $\mathcal{S}_\Sigma(X) \equiv \mathbf{X} \in \mathrm{Sym}(n)$ that differs from the canonical one $d_\Sigma\mathrm{id}(X) \equiv \mathbf{X} \in \mathrm{Sym}(n)$. As this could be confusing when the formula is written without this precision (and without bold letters), we adopt the same formalism as in [16,6,21].

We recall the quotient structure of the Bures-Wasserstein metric [6] in Table 6. The Riemannian operations are detailed in Table 7. Let us precise what was known and what is new in Table 7.

The proofs of the formulae of the distance and the logarithm can be found in [6]. The Levi-Civita connection and the exponential map were computed in [16]. We computed the Levi-Civita connection independently using a more geometric proof that we provide in Appendix A. We get a simpler formula.

Takatsu computed the curvature in [30] in a basis of vectors and gave a general formula in [31]. However, we argued above that the notations of [31] could be confusing

Table 6
Quotient structure of the Bures-Wasserstein metric.

Bundle	$\mathrm{GL}(n)$
Group action	$\rho : (A, U) \in \mathrm{GL}(n) \times \mathrm{O}(n) \longmapsto AU \in \mathrm{GL}(n)$
Submersion	$\pi : A \in \mathrm{GL}(n) \longmapsto \Sigma := AA^\top \in \mathrm{Sym}^+(n)$
Vertical space	$\mathcal{V}_A = \ker d_A\pi = \mathrm{Skew}(n)\, A^{-\top}$
Bundle metric	$G_A(M, M) = \mathrm{tr}(MM^\top)$
Hor. space	$\mathcal{H}_A = \mathcal{V}_A^{\perp_G} = \mathrm{Sym}(n)A$
Hor. isometry	$(d_A\pi)_{\vert\mathcal{H}_A} : \begin{cases} \mathcal{H}_A = \mathrm{Sym}(n)A & \longrightarrow \quad T_\Sigma\mathrm{Sym}^+(n) \\ X^h = X^0 A & \longmapsto \quad X = \Sigma X^0 + X^0\Sigma \end{cases}$
Sym. lift X^0	$\mathcal{S}_\Sigma : \begin{cases} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow \quad \mathcal{H}_{I_n} = \mathrm{Sym}(n) \\ X & \longmapsto \quad X^0 = PX^{0\prime}P^\top \text{with } X^{0\,\prime}_{ij} = \frac{X'_{ij}}{d_i+d_j} \end{cases}$
Hor. lift X^h	$X \in T_\Sigma\mathrm{Sym}^+(n) \longmapsto X^h = X^0 A \in \mathcal{H}_A$

Table 7
Riemannian operations of the Bures-Wasserstein metric on SPD matrices.

Metric	$g_\Sigma(X, X) = g_{\Sigma^{1/2}}(X^h, X^h) = \frac{1}{2}\sum_{i,j} \frac{1}{d_i+d_j} X'^2_{ij}$
Sq. dist.	$d(\Sigma, \Lambda)^2 = \mathrm{tr}\Sigma + \mathrm{tr}\Lambda - 2\mathrm{tr}((\Sigma\Lambda)^{1/2})$
Levi-Civita	$(\nabla_X Y)_{\vert\Sigma} = (\partial_X Y)_{\vert\Sigma} - (X^0\Sigma Y^0 + Y^0\Sigma X^0)$
Curvature	The sectional curvature is non-negative. More precisely $R_\Sigma(X, Y, X, Y) = \frac{3}{2}\sum_{i,j} \frac{d_i d_j}{d_i+d_j} \left[X^{0\prime}, Y^{0\prime}\right]^2_{ij}$ where $[V, W] = VW - WV$ is the Lie bracket of matrices.
Geodesics	$\gamma_{(\Sigma,X)}(t) = \Sigma + tX + t^2 X^0\Sigma X^0$ for $t \in I$ where I depends on $\lambda_{\max} = \max \mathrm{eig}(X^0)$ and $\lambda_{\min} = \min \mathrm{eig}(X^0)$ as follows: • If $\lambda_{\min} < 0 < \lambda_{\max}$, then $I = (-1/\lambda_{\max}, -1/\lambda_{\min})$. • If $0 \leqslant \lambda_{\min}$, then $I = (-1/\lambda_{\max}, +\infty)$. • If $\lambda_{\max} \leqslant 0$, then $I = (-\infty, -1/\lambda_{\min})$.
Logarithm	$\mathrm{Log}_\Sigma(\Lambda) = (\Sigma\Lambda)^{1/2} + (\Lambda\Sigma)^{1/2} - 2\Sigma$
PT map	Depends on the curve. Along a geodesic between commuting matrices $\Sigma = PDP^\top$ and $\Lambda = P\Delta P^\top$: $\Pi_{\Sigma\to\Lambda} : \begin{cases} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow \quad T_\Lambda\mathrm{Sym}^+(n) \\ X & \longmapsto \quad P\left[\sqrt{\frac{\delta_i+\delta_j}{d_i+d_j}}[P^\top X P]_{ij}\right]P^\top \end{cases}$

because of the chosen identification. Moreover, the expression of the curvature given there is a bit implicit since it is $R_\Sigma(X, Y, X, Y) = \frac{3}{4}\mathrm{tr}((\vert Y, X] - S)\Sigma([Y, X] - S)^\top)$ where $S = \mathcal{S}_\Sigma([X, Y]\Sigma + \Sigma[Y, X]) \in \mathrm{Sym}(n)$. Therefore, we prove in Appendix A the compact and explicit formula provided in Table 7 using the same method: O'Neill's equations of submersions [20].

Finally, the geodesic parallel transport between commuting SPD matrices is new. We provide a new formulation of the equation of the parallel transport between any two SPD matrices in the following proposition. It is used in the Python package `geomstats` [18] to compute the parallel transport. The proofs are given in Appendix A.

Proposition 3.2 *(Parallel transport equation of Bures-Wasserstein metric). Let $\gamma(t)$ the geodesic between $\gamma(0) = \Sigma$ and $\gamma(1) = \Lambda$, and a vector $X \in T_\Sigma\mathrm{Sym}^+(n)$. We denote*

$\gamma^h(t) = (1-t)\Sigma^{1/2} + t\Sigma^{-1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}$ *the horizontal lift of the geodesic* γ*. The two following statements are equivalent.*

(i) The vector field $X(t)$ *defined along* $\gamma(t)$ *is the parallel transport of* X*.*
(ii) $X(t) = \gamma(t)X^0(t) + X^0(t)\gamma(t)$ *where* $X^0(t)$ *is a curve in* $\mathrm{Sym}(n)$ *satisfying the following ODE:*

$$\gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \gamma^h(t)\dot{\gamma}^{h\top}X^0(t) + X^0(t)\dot{\gamma}^h\gamma^h(t)^\top = 0. \tag{10}$$

3.5. Bogoliubov-Kubo-Mori metric

The Bogoliubov-Kubo-Mori metric is a Riemannian metric used in quantum physics [24], given by $g^{\mathrm{BKM}}_\Sigma(X,X) = \mathrm{tr}(\int_0^\infty (\Sigma + t\,I_n)^{-1}X(\Sigma + t\,I_n)^{-1}X dt)$. It can be seen as the integration of the affine-invariant metric on a half-line included in the SPD cone. It can be rewritten thanks to the differential of the logarithm and we take this other expression as a definition.

Definition 3.5 *(Bogoliubov-Kubo-Mori (BKM) metric).* The *Bogoliubov-Kubo-Mori metric* is the $\mathrm{O}(n)$-invariant Riemannian metric defined for $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in T_\Sigma\mathrm{Sym}^+(n)$ by:

$$g^{\mathrm{BKM}}_\Sigma(X,X) = \mathrm{tr}(X\, d_\Sigma \log(X)). \tag{11}$$

Important functions related to this metric are defined by [17] to get simple expressions of the Levi-Civita connection and the curvature. Given $\Sigma = PDP^\top \in \mathrm{Sym}^+(n)$, they define $m_{ij} = \int_0^\infty (d_i+t)^{-1}(d_j+t)^{-1}dt$ which is symmetric in (i,j) and $m_{ijk} = \int_0^\infty (d_i+t)^{-1}(d_j+t)^{-1}(d_k+t)^{-1}dt$ which is symmetric in (i,j,k). They also denote $g_\Sigma(X) = d_\Sigma \log(X)$ whose expression is $g_\Sigma(X) = P\,g_D(X')\,P^\top$ and $[g_D(X')]_{ij} = m_{ij}X'_{ij}$ where $X' = P^\top XP$. This g_Σ is defined so that $g_\Sigma(X,Y) = \mathrm{tr}(X\,g_\Sigma(Y))$. By differentiating this equality and using the definition of the BKM metric, they get the differential of $\Sigma \longmapsto g_\Sigma$:

$$\begin{aligned} d_\Sigma g(PF_{ij}P^\top)(PF_{kl}P^\top) &= d_D g(F_{ij})(F_{kl}) \\ &= -\frac{1}{2}(\delta_{jk}m_{ilj}F_{il} + \delta_{jl}m_{ikj}F_{ik} + \delta_{il}m_{jki}F_{jk} + \delta_{ik}m_{jli}F_{jl}), \end{aligned}$$

or more compactly $[d_\Sigma g(PXP^\top)(PXP^\top)]_{ij} = -2\sum_{k=1}^n m_{ijk}X_{ik}X_{jk}$. The Levi-Civita connection and the curvature can be expressed in closed forms by means of g and dg, as shown in Table 8. Note that the sign of the sectional curvature is not known. The distance, exponential, logarithm and parallel transport maps are not known either.

In this section, we reviewed five of the mainly used $\mathrm{O}(n)$-invariant Riemannian metrics and we contributed new formulae. We also highlighted that the $\mathrm{O}(n)$-invariant

Table 8
Riemannian operations of the BKM metric on SPD matrices.

Metric	$g_\Sigma(X,X) = \mathrm{tr}(X\, d_\Sigma \log(X))$
Levi-Civita	$(\nabla_X Y)_{\vert\Sigma} = (\partial_X Y)_{\vert\Sigma} + \frac{1}{2} g_\Sigma^{-1}(d_\Sigma g(X)(Y))$
Curvature	$R_\Sigma(X,Y)Z = -\frac{1}{4} g_\Sigma^{-1}(d_\Sigma g(X)(g_\Sigma^{-1}(d_\Sigma g(Y)(Z))))$ $+ \frac{1}{4} g_\Sigma^{-1}(d_\Sigma g(Y)(g_\Sigma^{-1}(d_\Sigma g(X)(Z))))$

Euclidean, the $\mathrm{O}(n)$-invariant log-Euclidean and the affine-invariant metrics are actually two-parameter families of Riemannian metrics indexed by $(\alpha,\beta) \in \mathbf{ST}$ while this extra term weighted by the trace factor β is never defined in the literature for the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics. Actually, there does not seem to exist a natural way of extending them with a trace term. Indeed, under the Bures-Wasserstein metric, there is a choice of an $\mathrm{O}(n)$-right-invariant inner product on $\mathrm{GL}(n)$ but they differ from $\mathrm{O}(n)$-invariant inner products on symmetric matrices given in Theorem 2.1. Indeed, any inner product on $\mathrm{GL}(n)$ of the form $\langle X|X\rangle = \mathrm{tr}(X^\top S X)$ with $S \in \mathrm{Sym}^+(n)$ is $\mathrm{O}(n)$-right-invariant. As for the BKM metric, we could change the inner product in the integral but after computation, we would obtain this metric: $\alpha\, g_\Sigma^{\mathrm{BKM}}(X,X) + \beta \sum_{i,j} \log^{[1]}(d_i,d_j) X'_{ii} X'_{jj}$. The fact that we cannot separate the indices i and j in the trace term differs from the previous situations.

In the next section, we recall the definition of the class of kernel metrics [11,12] and a selection of its key properties. Since this class of Riemannian metrics contains all the previously introduced metrics without trace term, we show that this is the right framework to define the trace term extension. We show that this new class of extended kernel metrics still satisfies the key results on kernel metrics we selected. We also prove another property of these two classes: the stability under the cometric.

4. The interpolating class of kernel metrics: new observations

Kernel metrics were introduced by Hiai and Petz in 2009 [11]. It is a family of $\mathrm{O}(n)$-invariant metrics indexed by smooth bivariate functions $\phi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$ called kernels. It has several key properties and it encompasses all the $\mathrm{O}(n)$-invariant metrics introduced in Section 3 without trace factor ($\beta = 0$). After recalling these key results (Section 4.1), we provide new observations on kernel metrics (Section 4.2), especially the trace term extension and the stability under the cometric.

4.1. The general class of kernel metrics

Definition 4.1 *(Kernel metrics, mean kernel metrics).* [11] A *kernel metric* is an $\mathrm{O}(n)$-invariant metric for which there is a smooth bivariate map $\phi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$ such that $g_\Sigma(X,X) = g_D(X',X') = \sum_{i,j} \frac{1}{\phi(d_i,d_j)} X'^2_{ij}$, where $\Sigma = PDP^\top$ with $P \in \mathrm{O}(n)$ and $D = \mathrm{Diag}(d_1,...,d_n)$, and $X = PX'P^\top$.

Table 9
Bivariate functions of all the O(n)-invariant metrics of Section 3.

Metric	$\phi(x,y)$	Mean m	θ
Euclidean	1	Any mean	0
Log-Euclidean	$(\frac{x-y}{\log(x)-\log(y)})^2$	Logarithmic mean	2
Affine-invariant	xy	Geometric mean	2
Polar-affine	$(\frac{2xy}{x+y})^2$	Harmonic mean	2
Bures-Wasserstein	$4\,\frac{x+y}{2}$	Arithmetic mean	1
BKM	$\frac{x-y}{\log(x)-\log(y)}$	Logarithmic mean	1

A *mean kernel metric* is a kernel metric characterized by a bivariate map ϕ of the form $\phi(x,y) = a\,m(x,y)^\theta$ where $a > 0$ is a positive coefficient, $\theta \in \mathbb{R}$ is a homogeneity power and $m : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$ is a symmetric homogeneous mean, that is:

1. symmetric, i.e. $m(x,y) = m(y,x)$ for all $x, y > 0$,
2. homogeneous, i.e. $m(\lambda x, \lambda y) = \lambda\, m(x,y)$ for all $\lambda, x, y > 0$,
3. non-decreasing in both variables,
4. $\min(x,y) \leqslant m(x,y) \leqslant \max(x,y)$ for all $x, y > 0$. It implies $m(x,x) = x$.

As the goal of this paper is to extend the class of kernel metrics, we selected from [11,12] the results that we found simple and powerful to be able to generalize them later on. It would be interesting to study other properties such as monotonicity and comparison properties but it is beyond the scope of this paper. Our selection of results is in Proposition 4.1.

Proposition 4.1 *(Key results on kernel metrics). [11]*

1. *(Generality) The Euclidean, log-Euclidean and affine-invariant metrics without trace term ($\beta = 0$), the polar-affine, the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics are mean kernel metrics. The kernels and the names of the corresponding means are given in Table 9.*
2. *(Stability) The class of kernel metrics is stable under univariate diffeomorphisms. More precisely, if g is a kernel metric with kernel function ϕ and if f is a univariate diffeomorphism (defined in Section 2.2), then the pullback metric f^*g is a kernel metric with bivariate function $(x,y) \longmapsto \frac{\phi(f(x),f(y))}{f^{[1]}(x,y)^2}$. Note that the class of mean kernel metrics is not stable under univariate diffeomorphisms because of the non-decreasing property required for mean kernel metrics.*
3. *(Completeness) A mean kernel metric with homogeneity power θ is geodesically complete if and only if $\theta = 2$. Therefore this result provides a sufficient condition for kernel metrics to be geodesically complete.*

Another property that we left for a different reason is the attractivity of the Log-Euclidean metric, i.e. the fact that the log-Euclidean metric is the limit when p tends

to 0 of $\frac{1}{p^2}\mathrm{pow}_p^* g$, that is the pullback of a kernel metric by a power diffeomorphism $\mathrm{pow}_p : \Sigma \in \mathrm{Sym}^+(n) \longmapsto \Sigma^p \in \mathrm{Sym}^+(n)$, scaled by $\frac{1}{p^2}$. However, it is not specific to kernel metrics since this is the case for any metric g.

4.2. New observations on kernel metrics

4.2.1. Kernel metrics form a cone

The class of kernel metrics is a sub-cone of the cone of Riemannian metrics on the SPD manifold. Indeed, it is stable by positive scaling and it is convex because if g, g' are kernel metrics associated to ϕ, ϕ', then $(1-t)g + tg'$ is a kernel metric associated to $\phi\phi'/((1-t)\phi' + t\phi) > 0$ for $t \in [0,1]$.

4.2.2. Cometric stability of the class of kernel metrics

A Riemannian metric $g : T\mathcal{M} \times T\mathcal{M} \longrightarrow \mathbb{R}$ on a manifold $\mathcal{M}$ defines a cometric $g^* : T^*\mathcal{M} \times T^*\mathcal{M} \longrightarrow \mathbb{R}$ defined for all covectors $\omega, \omega' \in T^*\mathcal{M}$ by $g^*(\omega, \omega') = \omega(x')$ where $x' \in T\mathcal{M}$ is the unique vector such that for all vectors $x \in T\mathcal{M}$, $g(x, x') = \omega'(x)$ (Riesz's theorem).

On the manifold of SPD matrices $\mathcal{M} = \mathrm{Sym}^+(n)$, we have a canonical identification of $T_\Sigma\mathcal{M}$ with $\mathrm{Sym}(n)$ given by $d_\Sigma \mathrm{id}$. Hence by duality, we also have a canonical identification between $T^*_\Sigma\mathcal{M}$ and $\mathrm{Sym}(n)^*$. So to identify $T_\Sigma\mathcal{M}$ with $T^*_\Sigma\mathcal{M}$, we only need an identification between $\mathrm{Sym}(n)$ and $\mathrm{Sym}(n)^*$. This is provided by the Frobenius inner product. To summarize, there is a natural identification between the tangent space and the cotangent space given by:

$$\left\{ \begin{array}{ccc} T_\Sigma\mathrm{Sym}^+(n) & \longrightarrow & T^*_\Sigma\mathrm{Sym}^+(n) \\ X & \longmapsto & (Y \in T_\Sigma\mathrm{Sym}^+(n) \longmapsto \mathrm{tr}(d_\Sigma\mathrm{id}(X) d_\Sigma\mathrm{id}(Y))) \end{array} \right. . \tag{12}$$

Hence, a cometric on SPD matrices can be seen as a metric.

Back to kernel metrics, it is interesting to note that this class is stable under taking the cometric and that the cometric has a simple expression.

Proposition 4.2 *(Cometric stability of kernel metrics). Let g be a kernel metric with kernel function ϕ. Then the cometric g^* seen as a metric through the identification explained above is a kernel metric with kernel function $\phi^* = 1/\phi$.*

This elementary fact is interesting from a numerical point of view. Indeed, to compute numerically the geodesics, one can either integrate the geodesic equation involving the Christoffel symbols (which is of second order) or integrate its Hamiltonian version involving the cometric (which is of first order). Hence, the fact that the cometric of a kernel metric is available is a quite important result that appeared to be previously unnoticed. More precisely, the geodesic equation writes $\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0$ where $x(t)$ is a curve on the manifold $\mathcal{M}$ and Γ^k_{ij} are the Christoffel symbols related to the metric

by $\Gamma_{ij}^k = \frac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$. By considering a curve $p(t)$ on the cotangent bundle $T^*\mathcal{M}$ instead, and $x(t)$ the curve on the manifold $\mathcal{M}$ such that $p(t) \in T^*_{x(t)}\mathcal{M}$, the geodesic equation admits the following Hamiltonian formulation:

$$\begin{cases} \dot{x}^k = g^{kl}p_l \\ \dot{p}_l = -\frac{1}{2}\frac{\partial g^{ij}}{\partial x^l}p_i p_j \end{cases}. \tag{13}$$

The Hamiltonian equation is often preferred to compute the geodesics numerically since the integration is simpler and more stable. It only involves the cometric $g^* = (g^{ij})_{i,j}$, which is very easy to compute for a kernel metric.

4.2.3. Canonical Frobenius-like expression of a kernel metric

An expression of kernel metrics was given in [11] by means of the operators $\mathbb{L}_\Sigma : X \longmapsto \Sigma X$, $\mathbb{R}_\Sigma : X \longmapsto X\Sigma$ and $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma) : \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ defined for $\Sigma = PDP^\top \in \mathrm{Sym}^+(n)$ by $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma)X = P\left([\phi(d_i, d_j)]_{i,j} \circ (P^\top X P)\right)P^\top$, where $\circ$ denotes the Schur (entry-wise) product. This expression is $g^\phi_\Sigma(X, X) = \mathrm{tr}(X\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma)^{-1}(X))$. The existence of the map $\Phi : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ hidden in $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma) := \Phi_\Sigma$ is ensured by extending the $\mathrm{O}(n)$-equivariant map $\Phi : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ defined by $[\Phi_D(X)]_{ij} = \phi(d_i, d_j)X_{ij}$. Indeed, one can easily check that Φ satisfies the two hypotheses of Lemma 2.2. In this work, we even prefer to define the bivariate map $\psi = \phi^{-1/2}$ and define in a analogous way the map $\Psi : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ so that we can write the kernel metric with a suitable Frobenius-like expression:

$$g^\phi_\Sigma(X, X) = \mathrm{tr}(\Psi_\Sigma(X)^2). \tag{14}$$

We can give explicitly Ψ in some particular cases:

1. Euclidean metric: $\Psi^{\mathrm{E}}_\Sigma(X) = X$;
2. log-Euclidean metric: $\Psi^{\mathrm{LE}}_\Sigma(X) = d_\Sigma \log(X)$;
3. affine-invariant metric: $\Psi^{\mathrm{A}}_\Sigma(X) = \Sigma^{-1/2}X\Sigma^{-1/2}$.

This is an important step towards the trace term extension.

4.2.4. Kernel metrics with a trace term

The class of kernel metrics does not encompass the $\mathrm{O}(n)$-invariant Euclidean, $\mathrm{O}(n)$-invariant log-Euclidean and affine-invariant metrics with a trace factor $\beta \neq 0$. However, thanks to the previous canonical expression, we can define a natural extension of a kernel metric with a trace term.

Definition 4.2 *(Extended kernel metrics).* Let g^ϕ be a kernel metric associated to the kernel function $\phi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$. We define the map $\psi = \phi^{-1/2}$ and the map $\Psi : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$ as described above so that $g_\Sigma(X, X) = \mathrm{tr}(\Psi_\Sigma(X)^2)$. We

define a two-parameter family which extends the kernel metric g^{ϕ} for all $\Sigma \in \mathrm{Sym}^+(n)$ and $X \in \mathrm{Sym}(n)$ by:

$$g_{\Sigma}^{\phi,\alpha,\beta}(X,X) = \alpha \operatorname{tr}(\Psi_{\Sigma}(X)^2) + \beta \operatorname{tr}(\Psi_{\Sigma}(X))^2, \tag{15}$$

where $(\alpha, \beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\alpha + n\beta > 0$.

We can apply this definition to the Bures-Wasserstein and the BKM metrics. One can show that the trace term such defined is $\beta \operatorname{tr}(\Sigma^{-1/2}X)^2$. Contrarily to the log-Euclidean and the affine-invariant cases, there is no isometry a priori between two metrics of the family. It is interesting to note that Propositions 4.1 and 4.2 are still valid for these extended kernel metrics. We omit the proofs since they are analogous to the ones given for kernel metrics in [11].

Proposition 4.3 *(Key results on extended kernel metrics).*

1. *(Generality) All the metrics in Section 3 are extended kernel metrics.*
2. *(Stability) The class of extended kernel metrics is stable under univariate diffeomorphisms and the transformation is the same as in Proposition 4.1.*
3. *(Completeness) An extended mean kernel metric with homogeneity power θ is geodesically complete if and only if $\theta = 2$.*
4. *(Cometric) The class of extended kernel metrics is cometric-stable and the corresponding transformation is* $(\phi, \alpha, \beta) \longmapsto (\frac{1}{\phi}, \frac{1}{\alpha}, -\frac{\beta}{\alpha(\alpha+n\beta)})$.

In this section, we recalled the definition of kernel metrics and three key properties. We added the property of stability under the cometric with an explicit expression and we argued that it is an interesting property from a numerical point of view to compute geodesics. We found a wider class of metrics which satisfies the same key properties and which encompasses all the $\mathrm{O}(n)$-invariant metrics defined in Section 3. It is now tempting to look for wider classes of $\mathrm{O}(n)$-invariant metrics and to determine if these properties are still valid.

In the next section, we characterize $\mathrm{O}(n)$-invariant metrics by means of three multivariate functions satisfying conditions of symmetry, compatibility and positivity. This result allows to understand better the specificity of kernel metrics and extended kernel metrics within the whole class of $\mathrm{O}(n)$-invariant metrics. Then we give a counterpart of Proposition 4.3 and we propose a new intermediate class of $\mathrm{O}(n)$-invariant metrics which is cometric stable.

5. Characterization of $\mathrm{O}(n)$-invariant metrics

In this section, we give a characterization of $\mathrm{O}(n)$-invariant metrics on SPD matrices. We present it as an extension of Theorem 2.1 characterizing $\mathrm{O}(n)$-invariant inner

products on symmetric matrices. Instead of two parameters α, β which satisfy a positivity condition, an $\mathrm{O}(n)$-invariant metric is characterized by three multivariate functions $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}$ which satisfy a positivity condition plus a symmetry condition and a compatibility condition. This is explained in Section 5.1. We also give two corollary results which characterize two subclasses of $\mathrm{O}(n)$-invariant metrics with additional invariances: scaling invariance and inverse-consistency. Section 5.2 is dedicated to the proof of the theorem. In Section 5.3, we reinterpret kernel metrics in light of the theorem. In Section 5.4, we give key results on $\mathrm{O}(n)$-invariant metrics and we compare them to those on kernel metrics given in Proposition 4.1. In particular, we state that the cometric can be difficult to compute. Hence in Section 5.5, we introduce the class of bivariate separable metrics which is an intermediate class between $\mathrm{O}(n)$-invariant and extended kernel metrics, which is cometric-stable and for which the cometric is known in closed-form.

5.1. Theorem and corollaries

Let us rephrase the characterization of $\mathrm{O}(n)$-invariant inner products on $\mathrm{Sym}(n)$ (Theorem 2.1). An inner product $\langle \cdot | \cdot \rangle$ on $\mathrm{Sym}(n)$ is $\mathrm{O}(n)$-invariant if and only if there exist real numbers $\gamma, \alpha > 0$ and $\beta \in \mathbb{R}$ such that:

$$\langle X | X \rangle = \gamma \sum_i X_{ii}^2 + \alpha \sum_{i \neq j} X_{ij}^2 + \beta \sum_{i \neq j} X_{ii} X_{jj}, \tag{16}$$

1. (Compatibility) $\gamma = \alpha + \beta$,
2. (Positivity) the symmetric matrix S defined by $S_{ii} = \gamma$ and $S_{ij} = \beta$ is positive definite.

The characterization of $\mathrm{O}(n)$-invariant metrics on $\mathrm{Sym}^+(n)$ has an analogous form where real numbers are replaced by n-multivariate functions and where there is an additional property of symmetry of these functions. We introduce this notion of symmetry before stating the theorem. The proof is in Section 5.2.

Definition 5.1 *($(k, n-k)$-symmetric functions).* We say that a function $f : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}$ is $(k, n-k)$-symmetric if it is symmetric in its k first variables and symmetric in its $n-k$ last variables. In other words, f is invariant under permutations $\sigma = \sigma_1 \sigma_2$ where σ_1 has support in $\{1, ..., k\}$ and σ_2 has support in $\{k+1, ..., n\}$. Hence, given a set $I \subseteq \{1, ..., n\}$ of cardinal k and $d \in (\mathbb{R}^+)^n$, we denote $f(d_{i \in I}, d_{i \notin I}) := f(\sigma \cdot d)$ where $\sigma(\{1, ..., k\}) = I$ and $(\sigma \cdot d)_i = d_{\sigma(i)}$.

Theorem 5.1 *(Characterization of $\mathrm{O}(n)$-invariant metrics). Let g be a Riemannian metric on $\mathrm{Sym}^+(n)$. If g is $\mathrm{O}(n)$-invariant, then there exist three maps $\gamma, \alpha : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}^+$ and $\beta : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}$ such that for all $\Sigma = PDP^\top \in \mathrm{Sym}^+(n)$ and $X = PX'P^\top \in T_\Sigma \mathrm{Sym}^+(n)$:*

$$g_\Sigma(X,X) = g_D(X',X') \tag{17}$$
$$= \sum_i \gamma(d_i, d_{k\neq i}) X'^2_{ii} + \sum_{i\neq j} \alpha(d_i,d_j,d_{k\neq i,j}) X'^2_{ij} + \sum_{i\neq j} \beta(d_i,d_j,d_{k\neq i,j}) X'_{ii}X'_{jj},$$

0. (Symmetry) γ is $(1,n-1)$-symmetric and α,β are $(2,n-2)$-symmetric,
1. (Compatibility) γ equals $\alpha+\beta$ on the set $\mathcal{D} = \{d \in (\mathbb{R}^+)^n | d_1 = d_2\}$,
2. (Positivity) for all $d \in (\mathbb{R}^+)^n$, the symmetric matrix $S(d)$ defined by $S_{ii}(d) = \gamma(d_i, d_{k\neq i})$ and $S_{ij}(d) = \beta(d_i,d_j,d_{k\neq i,j})$ is positive definite.

Conversely, if there exist such maps α,β,γ, then Equation (17) correctly defines an $\mathrm{O}(n)$-invariant Riemannian metric that we denote $g^{\alpha,\beta,\gamma}$ or equivalently $g^{\alpha,S}$. Moreover, g is continuous if and only if α,β,γ are continuous.

Before giving the proof, we observe that this theorem allows to characterize subclasses of $\mathrm{O}(n)$-invariant metrics as well. Here we give the general form of $\mathrm{O}(n)$-invariant metrics that are invariant under scaling and under inversion respectively. We omit the proof.

Proposition 5.1 *(Characterizations of subclasses of $\mathrm{O}(n)$-invariant metrics). Let g be an $\mathrm{O}(n)$-invariant metric characterized by the maps α,β,γ.*

1. g is invariant under scaling if and only if $f(\lambda d) = \frac{1}{\lambda^2} f(d)$ for $f \in \{\alpha,\beta,\gamma\}$, for all $d \in (\mathbb{R}^+)^n$ and for all $\lambda > 0$.
2. g is invariant under inversion if and only if $\gamma(d_1^{-1},...,d_n^{-1}) = d_1^4 \gamma(d_1,...,d_n)$ and $f(d_1^{-1},...,d_n^{-1}) = d_1^2 d_2^2 f(d_1,...,d_n)$ for $f \in \{\alpha,\beta\}$, for all $d \in (\mathbb{R}^+)^n$.

5.2. Proof of the theorem

Proof of Theorem 5.1 (Characterization of $\mathrm{O}(n)$-invariant metrics). Let g be an $\mathrm{O}(n)$-invariant metric on $\mathrm{Sym}^+(n)$. Since any diagonal matrix D is invariant under the subgroup $\mathcal{D}^\pm(n)$, the inner product g_D is $\mathcal{D}^\pm(n)$-invariant. Hence, Lemma 2.3 (a) ensures that there are positive coefficients $\alpha_{ij}(D) = \alpha_{ji}(D)$ and a matrix $S(D) \in \mathrm{Sym}^+(n)$ s.t. $g_D(X,X) = \sum_{i\neq j} \alpha_{ij}(D) X_{ij}^2 + \sum_{i,j} S_{ij}(D) X_{ii}X_{jj}$. Then, we define the three maps:

- $\alpha : d \in (\mathbb{R}^+)^n \longmapsto \alpha_{12}(\mathrm{Diag}(d)) > 0$,
- $\beta : d \in (\mathbb{R}^+)^n \longmapsto S_{12}(\mathrm{Diag}(d))$,
- $\gamma : d \in (\mathbb{R}^+)^n \longmapsto S_{11}(\mathrm{Diag}(d)) > 0$.

Following the same idea as in the proof of Lemma 2.3 (b), we use the invariance under permutations since $\mathrm{Diag}^+(n)$ is stable under this action. Then, one easily checks that α,β are $(2,n-2)$-symmetric and γ is $(1,n-1)$-symmetric and that we can express the other coefficients in function of α,β,γ by permuting the d_i's. We get for $i \neq j$:

- $\alpha_{ij}(\mathrm{Diag}(d)) = \alpha(d_i, d_j, d_{k\neq i,j})$,
- $S_{ij}(\mathrm{Diag}(d)) = \beta(d_i, d_j, d_{k\neq i,j})$,
- $S_{ii}(\mathrm{Diag}(d)) = \gamma(d_i, d_{k\neq i})$.

So we get the expression (17), the symmetry and the positivity conditions. We only miss the compatibility condition so let $d = (d_1, ..., d_n) \in (\mathbb{R}^+)^n$ such that $d_1 = d_2$. Since $D = \mathrm{Diag}(d)$ is stable under any block-diagonal orthogonal matrix $R = \mathrm{Diag}(R_\theta, I_{n-2}) \in \mathrm{O}(n)$ with $R_\theta \in \mathrm{O}(2)$, with the same computations as in the proof of Theorem 2.1, we get $\gamma(d) = \alpha(d) + \beta(d)$.

Conversely, if α, β, γ are three maps satisfying the conditions of symmetry, compatibility and positivity, then we define $g_D(X,X) = \sum_i \gamma(d_i, d_{k\neq i}) X_{ii}^2 + \sum_{i\neq j} \alpha(d_i, d_j, d_{k\neq i,j}) X_{ij}^2 + \sum_{i\neq j} \beta(d_i, d_j, d_{k\neq i,j}) X_{ii} X_{jj}$. In other words, we define a map $g : \mathrm{Diag}^+(n) \times \mathrm{Sym}(n) \times \mathrm{Sym}(n) \longrightarrow \mathbb{R}$ and we would like to extend it by defining $g_{PDP^\top}(X,X) = g_D(P^\top XP, P^\top XP)$. According to Lemma 2.2, we have two cases to study. One can easily show that the first condition with permutations is satisfied. The non-trivial condition is the second one, involving a diagonal matrix $D = \mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p})$ with sorted diagonal values $\lambda_1 > ... > \lambda_p > 0$ and a block-diagonal orthogonal matrix $R = \mathrm{Diag}(R_1, ..., R_p) \in \mathrm{O}(n)$ with $R_k \in \mathrm{O}(m_k)$. So we have to show that $g_D(R^\top XR, R^\top XR) = g_D(X,X)$ for all matrix $X \in \mathrm{Sym}(n)$, since $R^\top DR = D$. We denote $\bar{X}^{kl} \in \mathrm{Mat}(m_k, m_l)$ the (k,l) block matrix defined by $\bar{X}^{kl}_{ij} = X_{n_{k-1}+i, n_{l-1}+j}$ where $n_k = \sum_{j=1}^k m_j$. Note that $\bar{X}^{kk} \in \mathrm{Sym}(m_k)$ is the k-th diagonal block of X and $\bar{X}^{lk} = (\bar{X}^{kl})^\top$. Therefore $\overline{R^\top XR}^{kl} = R_k^\top \bar{X}^{kl} R_l$. In the following, we split the sums between the blocks with multiplicity 1 and the blocks with higher multiplicity and we use the compatibility condition. The notation $\alpha(\lambda_k, \lambda_l, ...)$ stands for $\alpha(d_i, d_j, d_{m\neq i,j})$ where $\lambda_k = d_i$ and $\lambda_l = d_j$, i.e. $n_{k-1} + 1 \leqslant i \leqslant n_k$ and $n_{l-1} + 1 \leqslant j \leqslant n_l$. We compute the difference:

$$
\begin{aligned}
&g_D(R^\top XR, R^\top XR) - g_D(X,X) \\
&\quad = \sum_{k:m_k=1} \gamma(d_{n_k}, d_{m\neq n_k}) \underbrace{((R^\top XR)^2_{n_k n_k} - X^2_{n_k n_k})}_{0} \\
&\qquad + \sum_{\substack{k\neq l \\ m_k=m_l=1}} \alpha(\lambda_k, \lambda_l, ...) \underbrace{((R^\top XR)^2_{n_k n_l} - X^2_{n_k n_l})}_{0} \\
&\qquad + \sum_{\substack{k\neq l \\ m_k=m_l=1}} \beta(\lambda_k, \lambda_l, ...) \underbrace{((R^\top XR)_{n_k n_k} (R^\top XR)_{n_l n_l} - X_{n_k n_k} X_{n_l n_l})}_{0} \\
&\qquad + \sum_{k:m_k>1} \underbrace{\gamma(\lambda_k, \lambda_k, ...)}_{\alpha(\lambda_k,\lambda_k,...)+\beta(\lambda_k,\lambda_k,...)} \sum_{i=n_{k-1}+1}^{n_k} ((R^\top XR)^2_{ii} - X^2_{ii})
\end{aligned}
$$

$$+ \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \alpha(\lambda_k, \lambda_l, ...) \sum_{\substack{n_{k-1}+1 \leqslant i \leqslant n_k \\ n_{l-1}+1 \leqslant j \leqslant n_l \\ i \neq j}} ((R^\top X R)_{ij}^2 - X_{ij}^2)$$

$$+ \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \beta(\lambda_k, \lambda_l, ...) \sum_{\substack{n_{k-1}+1 \leqslant i \leqslant n_k \\ n_{l-1}+1 \leqslant j \leqslant n_l \\ i \neq j}} ((R^\top X R)_{ii}(R^\top X R)_{jj} - X_{ii}X_{jj}).$$

Hence the missing term $i = j$ in the two last sums is provided by the sum weighted by γ. After a change of indices based on the equality $\overline{R^\top X R}^{kl} = R_k^\top \bar{X}^{kl} R_l$, we get:

$$\begin{aligned}
&g_D(R^\top X R, R^\top X R) - g_D(X, X) \\
&= \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \alpha(\lambda_k, \lambda_l, ...) \underbrace{\sum_{i=1}^{m_k} \sum_{j=1}^{m_l} ((R_k^\top \bar{X}^{kl} R_l)_{ij}^2 - (\bar{X}^{kl})_{ij}^2)}_{\mathrm{tr}(R_k^\top \bar{X}^{kl} R_l (R_k^\top \bar{X}^{kl} R_l)^\top) - \mathrm{tr}(\bar{X}^{kl}(\bar{X}^{kl})^\top) = 0} \\
&\quad + \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \beta(\lambda_k, \lambda_l, ...) \underbrace{\sum_{i=1}^{m_k} \sum_{j=1}^{m_l} ((R_k^\top \bar{X}^{kk} R_k)_{ii}(R_l^\top \bar{X}^{ll} R_l)_{jj} - \bar{X}_{ii}^{kk} \bar{X}_{jj}^{ll})}_{\mathrm{tr}(R_k^\top \bar{X}^{kk} R_k)\mathrm{tr}(R_l^\top \bar{X}^{ll} R_l) - \mathrm{tr}(\bar{X}^{kk})\mathrm{tr}(\bar{X}^{ll}) = 0} \\
&= 0.
\end{aligned}$$

This proves that g_Σ is well defined for all $\Sigma \in \mathrm{Sym}^+(n)$ and $\mathrm{O}(n)$-invariant by construction. The positivity condition ensures that g is a metric.

Finally, it is clear that α, β, γ have at least the same regularity as the metric g since they are coordinates of the map $D \in \mathrm{Diag}^+(n) \longmapsto g_D$. Let us prove that if α, β, γ are continuous, then g is continuous. The main argument is in the following lemma (proved after the proof of the theorem).

Lemma 5.1 *(Eigenvalues and eigenvectors of close symmetric matrices). Let $\Sigma, \Lambda \in \mathrm{Sym}^+(n)$. Let $D, \Delta \in \mathrm{Diag}^+(n)$ be their matrices of ordered eigenvalues, i.e. $D = \mathrm{Diag}(d_1, ..., d_n)$ and $\Delta = \mathrm{Diag}(\delta_1, ..., \delta_n)$ with $d_1 \leqslant ... \leqslant d_n$ and $\delta_1 \leqslant ... \leqslant \delta_n$. We denote $m = m_\Sigma = \min_{\lambda \neq \mu \in \mathrm{eig}(\Sigma)} (\lambda - \mu)^2$ if $\Sigma \notin \mathbb{R}^+ I_n$, and $m = +\infty$ otherwise. Then:*

1. *[5, (IV.62)] $\|D - \Delta\|_2 \leqslant \|\Sigma - \Lambda\|_2$,*
2. *there exists $(P, Q) \in \mathrm{O}(n) \times \mathrm{O}(n)$ such that $\Sigma = PDP^\top$, $\Lambda = Q\Delta Q^\top$ and $\|P - Q\|_2 \leqslant 2\sqrt{\frac{2}{m}} \|\Sigma - \Lambda\|_2$.*

Let us prove that g is continuous by showing that for all ε, for all $\Sigma, \Lambda \in \mathrm{Sym}^+(n)$, there exists η such that if $\|\Sigma - \Lambda\|_2 \leqslant \eta$, then for all $X \in \mathrm{Sym}^+(n)$, $|g_\Sigma(X, X) - g_\Lambda(X, X)| \leqslant \varepsilon \|X\|_2^2$. Let $\varepsilon > 0$ and $\Sigma, \Lambda \in \mathrm{Sym}^+(n)$. Given Lemma 5.1, let $D, \Delta \in$

$\mathrm{Diag}^+(n)$ and $P, Q \in \mathrm{O}(n)$ such that $\Sigma = PDP^\top$, $\Lambda = Q\Delta Q^\top$, $\|D-\Delta\|_2 \leqslant \|\Sigma - \Lambda\|_2$ and $\|P-Q\|_2^2 \leqslant 4\sqrt{\frac{n}{m}}\|\Sigma-\Lambda\|_2$. For all $X \in \mathrm{Sym}(n)$:

$$\begin{aligned}
|g_\Sigma(X,X) - g_\Lambda(X,X)| &\leqslant \sum_i |\gamma(d_i, d_{k\neq i})[P^\top XP]_{ii}^2 - \gamma(\delta_i, \delta_{k\neq i})[Q^\top XQ]_{ii}^2| \\
&+ \sum_{i\neq j} |\alpha(d_i, d_j, d_{k\neq i,j})[P^\top XP]_{ij}^2 - \alpha(\delta_i, \delta_j, \delta_{k\neq i,j})[Q^\top XQ]_{ij}^2| \\
&+ \sum_{i\neq j} |\beta(d_i, d_j, d_{k\neq i,j})[P^\top XP]_{ii}[P^\top XP]_{jj} - \beta(\delta_i, \delta_j, \delta_{k\neq i,j})[Q^\top XQ]_{ii}[Q^\top XQ]_{jj}|.
\end{aligned}$$

To use Lemma 5.1, we separate the eigenvalues and eigenvectors by introducing $0 = -\gamma(d_i, d_{k\neq i})[Q^\top XQ]_{ii}^2 + \gamma(d_i, d_{k\neq i})[Q^\top XQ]_{ii}^2$ in the absolute value on the first line:

$$\begin{aligned}
&\sum_i |\gamma(d_i, d_{k\neq i})[P^\top XP]_{ii}^2 - \gamma(\delta_i, \delta_{k\neq i})[Q^\top XQ]_{ii}^2| \\
&\leqslant \sum_i |\gamma(d_i, d_{k\neq i})|\,|[P^\top XP]_{ii}^2 - [Q^\top XQ]_{ii}^2| + \sum_i |\gamma(d_i, d_{k\neq i}) - \gamma(\delta_i, \delta_{k\neq i})|\,|[Q^\top XQ]_{ii}|^2 \\
&\leqslant C\sum_i |[P^\top XP]_{ii}^2 - [Q^\top XQ]_{ii}^2| + \|X\|_2^2 \sum_i |\gamma(d_i, d_{k\neq i}) - \gamma(\delta_i, \delta_{k\neq i})| \\
&\leqslant C\sum_{i,j,k,l} |[P^\top XP]_{ij}[P^\top XP]_{kl} - [Q^\top XQ]_{ij}[Q^\top XQ]_{kl}| + \|X\|_2^2 \sum_{\sigma\in\mathfrak{S}(n)} |\gamma\circ\sigma(D) - \gamma\circ\sigma(\Delta)|,
\end{aligned}$$

where $C = \max_{f\in\{\alpha,\beta,\gamma\},\sigma\in\mathfrak{S}(n)} |f\circ\sigma(D)|$. We can get analogous bounds for α, β. Therefore, we get:

$$\begin{aligned}
|g_\Sigma(X,X) - g_\Lambda(X,X)| \leqslant 3C &\sum_{i,j,k,l} |[P^\top XP]_{ij}[P^\top XP]_{kl} - [Q^\top XQ]_{ij}[Q^\top XQ]_{kl}| \\
&+ 3\|X\|_2^2 \max_{f\in\{\alpha,\beta,\gamma\}} \sum_{\sigma\in\mathfrak{S}(n)} |f\circ\sigma(D) - f\circ\sigma(\Delta)|.
\end{aligned}$$

Since α, β, γ and permutations are continuous, the term $\max_f \sum_\sigma |f\circ\sigma(D) - f\circ\sigma(\Delta)|$ can be made inferior to $\frac{\varepsilon}{6}$ for Δ sufficiently close to D, let's say $\|D-\Delta\| \leqslant \eta_1$ for a given $\eta_1 > 0$. On the other hand:

$$\begin{aligned}
&\sum_{i,j,k,l} |[P^\top XP]_{ij}[P^\top XP]_{kl} - [Q^\top XQ]_{ij}[Q^\top XQ]_{kl}| \\
&\leqslant \sum_{i,j,k,l} |[P^\top XP]_{ij}|(|[P^\top XP]_{kl} - [P^\top XQ]_{kl}| + |[P^\top XQ]_{kl} - [Q^\top XQ]_{kl}|) \\
&\quad + (|[P^\top XP]_{ij} - [P^\top XQ]_{ij}| + |([P^\top XQ]_{ij} - [Q^\top XQ]_{ij})|)|[Q^\top XQ]_{kl}| \\
&\leqslant (\|P^\top XP\|_1 + \|Q^\top XQ\|_1)(\|(P^\top X(P-Q)\|_1 + \|(P-Q)^\top XQ\|_1) \\
&\leqslant 4n^2\|P-Q\|_2\|X\|_2^2.
\end{aligned}$$

So for $\|P-Q\|_2 \leqslant \frac{\varepsilon}{24n^2C}$ and $\|D-\Delta\|_2 \leqslant \eta_1$, we have $|g_\Sigma(X,X) - g_\Lambda(X,X)| \leqslant \varepsilon\|X\|_2^2$. Thus if we choose $\eta := \min(\eta_1, \frac{\sqrt{m}}{2\sqrt{2}}\frac{\varepsilon}{24n^2C})$, then if $\|\Sigma-\Lambda\|_2 \leqslant \eta$, we have $|g_\Sigma(X,X) - g_\Lambda(X,X)| \leqslant \varepsilon\|X\|_2^2$, which proves the continuity. □

Proof of Lemma 5.1 (Eigenvalues and eigenvectors of close symmetric matrices). The first point comes from [5, (IV.62)]. Let us prove the second point. Let $P, Q \in \mathrm{O}(n)$ such that $\Sigma = PDP^\top$ and $\Lambda = Q\Delta Q^\top \in \mathrm{Sym}^+(n)$, where D, Δ are sorted by increasing order.

We denote $R = \mathrm{Diag}(R_1, ..., R_p) \in \mathrm{O}(n)$ a block-diagonal orthogonal matrix with $R_j \in \mathrm{O}(m_j)$, where $m_1, ..., m_p \in \mathbb{N}$ are the multiplicities of the eigenvalues of $D = \mathrm{Diag}(\lambda_1 I_{m_1}, ..., \lambda_p I_{m_p})$. We are looking for R such that $\|PR - Q\|_2 \leqslant 2\sqrt{\frac{2}{m}}\|\Sigma - \Lambda\|_2$ with $m = \min_{i\neq j}(\lambda_i - \lambda_j)^2$. We denote $U = P^\top Q$ and $W = \mathrm{Diag}(W_1, ..., W_p)$ the block-diagonal pinching of U where $W_j \in \mathrm{Mat}(m_j)$. We choose R as the orthogonal factor in a polar decomposition of $W = SR$ where $S = \sqrt{WW^\top}$ is a symmetric positive semi-definite matrix. Since for all $j \in \{1, ..., p\}$, $W_j W_j^\top \leqslant I_{m_k}$ for the Loewner order (because W_j is a principal block of the orthogonal matrix U), we have $WW^\top \leqslant I_n$. Thus $S = \sqrt{WW^\top} \geqslant WW^\top$ since $\sqrt{x} \geqslant x$ for all $x \in [0,1]$. So $\mathrm{tr}(WR^\top) = \mathrm{tr}(S) \geqslant \mathrm{tr}(WW^\top)$. Thus:

$$\begin{aligned}
\|PR - Q\|_2^2 &= 2\mathrm{tr}(I_n - R^\top U) = 2\mathrm{tr}(I_n - R^\top W) \\
&\leqslant 2\mathrm{tr}(I_n - WW^\top) = 2\sum_{d_i \neq d_j} U_{ij}^2 \\
&\leqslant \frac{2}{m}\sum_{d_i \neq d_j}(d_i - d_j)^2 U_{ij}^2 = \frac{2}{m}\|DU - UD\|_2^2 = \frac{2}{m}\|\Sigma - QDQ^\top\|_2^2, \\
\|PR - Q\|_2 &\leqslant \sqrt{\frac{2}{m}}(\|\Sigma - \Lambda\|_2 + \|Q(\Delta - D)Q^\top\|_2) \leqslant 2\sqrt{\frac{2}{m}}\|\Sigma - \Lambda\|_2,
\end{aligned}$$

which proves the result. □

The smoothness seems to be more complicated to study. We suspect additional conditions of compatibility on the derivatives of the smooth maps α, β, γ at the singular set of SPD matrices with repeated eigenvalues in order to make the metric g is smooth.

5.3. Reinterpretation of kernel metrics

Theorem 5.1 allows to reinterpret kernel metrics. The curiosity of this theorem is the function γ because we have no information on it as soon as the d_i's are distinct. If α, β, γ do not depend on their $n-2$ last arguments, i.e. if they are *bivariate*, then γ does not depend on its second argument and $\gamma(d_1)$ must be equal to $\alpha(d_1, d_1) + \beta(d_1, d_1)$. Hence $g_\Sigma(X,X) = \sum_{i,j}\alpha(d_i, d_j)X_{ij}'^2 + \sum_{i,j}\beta(d_i, d_j)X_{ii}'X_{jj}'$ with $\alpha > 0$ and $\alpha + n\beta > 0$, which is much more tractable. Moreover, if $\beta = 0$, then the quadratic form has a *diagonal*

Table 10
Name correspondences for kernel metrics and sub/super-classes.

Previous description	New designation
Kernel metric	BOD metric
Mean kernel metric	MOD metric
Extended kernel metric	BOST metric
Extended mean kernel metric	MOST metric

expression (sum of squares $X_{ij}'^2$, no mixed terms $X_{ii}'X_{jj}'$) in the basis of matrices induced by the *orthogonal* matrix $P \in \mathrm{O}(n)$ in the eigenvalue decomposition of Σ. In this case, we say that the metric is *ortho-diagonal*.

To sum up, the subclass of kernel metrics has two fundamental properties: it is bivariate ($\alpha = \gamma - \beta = 1/\phi$) and ortho-diagonal ($\beta = 0$). This is the reason why we propose to designate kernel (resp. mean kernel) metrics as Bivariate Ortho-Diagonal or BOD metrics (resp. Mean Ortho-Diagonal or MOD metrics), as summarized in Table 10. We say that the metric is Bivariate Ortho-ST when it is the extension by Definition 4.2 of a Bivariate Ortho-Diagonal metric with the Scaling and Trace factors $\alpha > 0$ and $\beta > -\alpha/n$. Hence, the extended (mean) kernel metrics can also be designated as BOST (and MOST) metrics.

5.4. Key results on $\mathrm{O}(n)$*-invariant metrics*

In Section 4, we gave four key results on BOD/MOD metrics in Propositions 4.1 and 4.2, and four key results on BOST/MOST metrics in Proposition 4.3. Here we give the counterpart of these propositions for $\mathrm{O}(n)$-invariant metrics.

Proposition 5.2 *(Key results on* $\mathrm{O}(n)$*-invariant metrics).*

1. *(Generality) The class of* $\mathrm{O}(n)$*-invariant metrics obviously contains the classes of BOD, MOD, BOST, MOST metrics, hence it contains all the metrics in Section 3.*
2. *(Stability) The class of* $\mathrm{O}(n)$*-invariant metrics is obviously stable by* $\mathrm{O}(n)$*-equivariant diffeomorphisms of* $\mathrm{Sym}^+(n)$*. Hence it is stable by univariate diffeomorphisms* $f : \mathrm{Sym}^+(n) \longrightarrow \mathrm{Sym}^+(n)$ *and in this case, the pullback metric* $f^* g^{\alpha,\beta,\gamma}$ *is characterized by the three maps:*

   (a) $\alpha_f : d \in (\mathbb{R}^+)^n \longmapsto \frac{\alpha(f(d))}{f^{[1]}(d_1,d_2)^2}$,
   (b) $\beta_f : d \in (\mathbb{R}^+)^n \longmapsto \frac{\beta(f(d))}{f^{[1]}(d_1,d_2)^2}$,
   (c) $\gamma_f : d \in (\mathbb{R}^+)^n \longmapsto \frac{\gamma(f(d))}{f'(d_1)^2}$.

3. *(Completeness) Let* $g = g^{\alpha,\beta,\gamma}$ *be an* $\mathrm{O}(n)$*-invariant metric. We assume that* α, β, γ *satisfy a homogeneity property which is similar to the one assumed for mean kernel*

metrics: there exists $\theta \in \mathbb{R}$ *such that for* $f \in \{\alpha, \beta, \gamma\}$, $x \in (\mathbb{R}^+)^n$ *and* $\lambda > 0$, *we have* $f(\lambda x) = \lambda^{-\theta} f(x)$. *If the metric* g *is geodesically complete, then* $\theta = 2$.

4. *(Cometric) The class of* $\mathrm{O}(n)$*-invariant metrics is obviously cometric-stable. The cometric is characterized by* $\alpha^* = 1/\alpha$ *and* $S^* = S^{-1}$ *where* $S(d) \in \mathrm{Sym}^+(n)$ *is defined by* $S_{ij}(d) = \beta(d_i, d_j, d_{k\neq i,j})$ *and* $S_{ii}(d) = \gamma(d_i, d_{k\neq i})$ *for all* $d \in \mathbb{R}^+$ *and* $i \neq j$.

We omit the proof since it consists in elementary verifications for all but the third statement, whose proof is analogous to the one given in [11].

About completeness, the result is much weaker for general $\mathrm{O}(n)$-invariant metrics. Indeed, we lost the converse implication: "if $\theta = 2$, then the metric is geodesically complete". According to the proof of [11], the key element to prove this converse implication is exactly the bivariance, plus the fact that a symmetric homogeneous mean satisfies $m(x, x) = x$. It is worth noticing that $\theta = 2$ is still necessary though.

About the cometric, we lost the closed-form expression we had for BOD and BOST metrics. Computing the cometric is numerically quite heavy in general because it is equivalent to invert the matrix $S(d)$ for all $d \in (\mathbb{R}^+)^n$. However, note that when $\beta = 0$, the cometric is obviously given by the triple $(1/\alpha, 0, 1/\gamma)$. These ortho-diagonal metrics can be seen as the multivariate generalization of BOD metrics. In the next section, we give a cometric-stable extension of the class of BOST metrics for which the cometric can be computed in closed form: the class of bivariate separable metrics.

5.5. Bivariate separable metrics

We argued in Section 5.3 that *bivariate* metrics are of the form $g_\Sigma(X, X) = \sum_{i,j} \alpha(d_i, d_j) X'^2_{ij} + \sum_{i,j} \beta(d_i, d_j) X'_{ii} X'_{jj}$ with $\alpha > 0$ and $\alpha + n\beta > 0$. Then, the first term corresponds to a BOD metric and it can be rewritten $\mathrm{tr}(\Psi_\Sigma(X)^2)$, but it is still difficult to write the second term in a more compact way. If the function β is *separable*, i.e. if β can be written $\beta(x, y) = \psi^{(1)}(x)\psi^{(2)}(y)$, then the second term is simply $\mathrm{tr}(\Psi^{(1)}_\Sigma(X))\mathrm{tr}(\Psi^{(2)}_\Sigma(X))$. Indeed, we can define $\Psi^{(k)}_D(X) = \mathrm{Diag}(\psi^{(k)}(d_i)X_{ii})$ and extend it into $\Psi^{(k)}_\Sigma$ as explained in Section 4.2.3. In particular, BOST metrics correspond to the case when $\beta(x, y) = \lambda\sqrt{\alpha(x, x)\alpha(y, y)}$ with $1 + n\lambda > 0$. The wider class of *bivariate separable* metrics is actually cometric-stable and the cometric can be computed quite easily. This is stated in Proposition 5.3.

Proposition 5.3 *(Cometric of bivariate separable metrics). Let* $\psi : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}^+$ *be a symmetric map and let* $\psi^{(1)}, \psi^{(2)} : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ *be two maps on positive real numbers. As explained above, we define their extensions* $\Psi, \Psi^{(1)}, \Psi^{(2)} : \mathrm{Sym}^+(n) \times \mathrm{Sym}(n) \longrightarrow \mathrm{Sym}(n)$. *The quadratic form defined by* $g_\Sigma(X, X) = \mathrm{tr}(\Psi_\Sigma(X)^2) + \mathrm{tr}(\Psi^{(1)}_\Sigma(X))\mathrm{tr}(\Psi^{(2)}_\Sigma(X))$ *automatically satisfies the symmetry and compatibility conditions of Theorem 5.1. Then* g *is positive definite if and only if the vectors* $x = x(d) = \left(\frac{\psi^{(1)}(d_i)}{\psi(d_i, d_i)}\right)_{1\leqslant i\leqslant n}$ *and* $y = y(d) = \left(\frac{\psi^{(2)}(d_i)}{\psi(d_i, d_i)}\right)_{1\leqslant i\leqslant n}$ *satisfy the inequality* $\|x\|\|y\| - \langle x|y\rangle < 2$ *for all* $d \in (\mathbb{R}^+)^n$.

In this case, we say that g is a Bivariate Separable metric. As an $\mathrm{O}(n)$-invariant metric, it is characterized by $\alpha(d) = \psi(d_1, d_2)^2$ and the matrix $S = S(d) = \Delta(I_n + \frac{1}{2}(xy^\top + yx^\top))\Delta$ with $\Delta = \mathrm{Diag}(\psi(d_i, d_i))$. This class of metrics is cometric-stable. If $x = 0$ or $y = 0$, the cometric at Σ is simply characterized by $S^{-1} = \Delta^{-2}$. Otherwise, the cometric is given by:

$$S^{-1} = \Delta^{-1}\left[I_n - \frac{1}{4c}(2 + \langle x|y\rangle)(xy^\top + yx^\top) + \frac{1}{4c}(\|y\|^2 xx^\top + \|x\|^2 yy^\top)\right]\Delta^{-1}, \quad (18)$$

with $c = 1 + \langle x|y\rangle - \frac{1}{4}(\|x\|^2\|y\|^2 - \langle x|y\rangle^2) > 0$.

Proof of Proposition 5.3. To determine when g is a metric, we express the functions α, β, γ, S of Theorem 5.1 in function of $\psi, \psi^{(1)}, \psi^{(2)}$:

1. $\alpha(d_1, ..., d_n) = \psi(d_1, d_2)^2 > 0$,
2. $\beta(d_1, ..., d_n) = \frac{1}{2}(\psi^{(1)}(d_1)\psi^{(2)}(d_2) + \psi^{(1)}(d_2)\psi^{(2)}(d_1))$,
3. $\gamma(d_1, ..., d_n) = \psi(d_1, d_1)^2 + \psi^{(1)}(d_1)\psi^2(d_1)$,
4. hence $S_{ij}(d) = \Delta_{ij}^2 + \frac{1}{2}(\psi^{(1)}(d_i)\psi^{(2)}(d_j) + \psi^{(2)}(d_i)\psi^{(1)}(d_j))$, so we have $S = \Delta(I_n + \frac{1}{2}(xy^\top + yx^\top))\Delta$ with the notations of the proposition.

The symmetry and compatibility conditions of Theorem 5.1 are trivially satisfied. The positivity condition reduces to $S \in \mathrm{Sym}^+(n)$, i.e. $I_n + \frac{1}{2}(xy^\top + yx^\top) \in \mathrm{Sym}^+(n)$. As the eigenvalues of $M = xy^\top + yx^\top$ are 0 (with multiplicity $n - 2$) and $\langle x|y\rangle \pm \|x\|\|y\|$, S is positive definite if and only if $2 + \langle x|y\rangle \pm \|x\|\|y\| > 0$. But $\langle x|y\rangle + \|x\|\|y\| \geqslant 0$ so there is only one condition: $2 > \|x\|\|y\| - \langle x|y\rangle (\geqslant 0)$, as announced.

Now, we want to compute S^{-1}. If $x = 0$ or $y = 0$, the result is obvious so we assume that $x, y \neq 0$. As M is of rank 2 at most, there exists a polynomial P of degree 3 at most such that $P(I_n + \frac{1}{2}M) = 0$. Let us find such a polynomial to compute S^{-1}. Since $M^2 = \langle x|y\rangle M + N$ with $N = \|y\|^2 xx^\top + \|x\|^2 yy^\top$ and $NM = \|x\|^2\|y\|^2 M + \langle x|y\rangle N$, we have:

$$\begin{aligned}
\left(I_n + \frac{1}{2}M\right)^2 &= I_n + \left(1 + \frac{\langle x|y\rangle}{4}\right)M + \frac{1}{4}N, \\
\left(I_n + \frac{1}{2}M\right)^3 &= \left(I_n + \frac{1}{2}M\right)^2 + \frac{1}{2}\left(I_n + \frac{1}{2}M\right)^2 M \\
&= \left(I_n + \frac{1}{2}M\right)^2 + \frac{1}{2}M + \frac{1}{2}\left(1 + \frac{\langle x|y\rangle}{4}\right)M^2 + \frac{1}{8}NM \\
&= \left(I_n + \frac{1}{2}M\right)^2 + \frac{4 + 4\langle x|y\rangle + \langle x|y\rangle^2 + \|x\|^2\|y\|^2}{8}M + \frac{1}{4}(2 + \langle x|y\rangle)N \\
&= a\left(I_n + \frac{1}{2}M\right)^2 + \frac{b}{2}M - (2 + \langle x|y\rangle)I_n
\end{aligned}$$

$$= a\left(I_n + \frac{1}{2}M\right)^2 + b\left(I_n + \frac{1}{2}M\right) + c\, I_n,$$

with $\begin{cases} a = 3 + \langle x|y\rangle \\ b = \frac{-12-8\langle x|y\rangle - \langle x|y\rangle^2 + \|x\|^2\|y\|^2}{4} \\ c = 1 + \langle x|y\rangle + \frac{\langle x|y\rangle^2 - \|x\|^2\|y\|^2}{4} = 1 - a - b > 0 \end{cases}$. Indeed, $c > 1 + \langle x|y\rangle - \frac{1}{2}(\|x\|\|y\| + \langle x|y\rangle) = 1 - \frac{1}{2}(\|x\|\|y\| - \langle x|y\rangle) > 0$. Hence, denoting $S_0 := I_n + \frac{1}{2}M$, we have $S_0^{-1} = \frac{1}{c}\left(S_0^2 - a\, S_0 - bI_n\right) = I_n + \frac{1}{4c}(N - (2 + \langle x|y\rangle)M)$ and $S^{-1} = \Delta^{-1}\big(I_n + \frac{1}{4c}(N - (2 + \langle x|y\rangle)M)\big)\Delta^{-1}$ which is exactly Equation (18).

Finally, we want to prove that the cometric is bivariate separable. Regarding Equation (18), we look for $x' = \frac{Ax+By}{4c}$ and $y' = Cx + Dy$ for $A, B, C, D \in \mathbb{R}$ such that:

$$x'y'^{\top} + y'x'^{\top} = -\frac{1}{2c}(2 + \langle x|y\rangle)(xy^{\top} + yx^{\top}) + \frac{1}{2c}(\|y\|^2 xx^{\top} + \|x\|^2 yy^{\top}) \tag{19}$$

It is satisfied if $AC = \|y\|^2$, $BD = \|x\|^2$ and $AD + BC = -2(2 + \langle x|y\rangle)$, or equivalently $(AX+B)(CX+D) = \|y\|^2X^2 - 2(2+\langle x|y\rangle)X + \|x\|^2$. This is a second-order polynomial with roots $\lambda = \frac{2+\langle x|y\rangle+\sqrt{\delta}}{\|y\|^2}$ and $\mu = \frac{2+\langle x|y\rangle-\sqrt{\delta}}{\|y\|^2}$ where $\delta = (2 + \langle x|y\rangle + \|x\|\|y\|)(2 + \langle x|y\rangle - \|x\|\|y\|) > 0$ is the discriminant. Hence, it suffices to define $A = \|y\|$, $B = -\lambda\|y\|$, $C = \|y\|$ and $D = -\mu\|y\|$, so that $S^{-1} = \Delta^{-1}\left(I_n + \frac{1}{2}(x'y'^{\top} + y'x'^{\top})\right)\Delta^{-1}$. Hence, the cometric is bivariate separable and this class of metrics is cometric-stable. □

6. Conclusion

To encompass all the $\mathrm{O}(n)$-invariant metrics summarized in Section 3, including the ones with a trace term ($\beta \neq 0$), we defined the class of extended kernel metrics. This class satisfies the key results of stability and completeness we selected from [11] plus the cometric-stability with cometric in closed form, which is important to compute geodesics numerically via the Hamiltonian formulation. Then, from the characterization of $\mathrm{O}(n)$-invariant metrics in terms of three continuous maps $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \longrightarrow \mathbb{R}^+$ satisfying properties of symmetry, compatibility and positivity, we were able to characterize kernel metrics as Bivariate Ortho-Diagonal (BOD) metrics. Among the key results on mean kernel metrics, the sufficient condition of completeness and the closed-form expression of the cometric disappear for general $\mathrm{O}(n)$-invariant metrics. We finally defined the intermediate class of bivariate separable metrics which is cometric-stable and for which the cometric has a simple expression.

Since kernel metrics encompass very different metrics regarding curvature and completeness, it would be nice to introduce some more requirements on metrics to perform the opposite work of defining principled sub-classes of (mean) kernel metrics. There is actually a companion paper entitled "The geometry of mixed-Euclidean metrics on symmetric positive definite matrices" where we propose some principled subfamilies of kernel metrics. It would also be interesting to rely on the cometric-stability of kernel metrics

or super-classes to effectively compute the geodesics numerically and to investigate their properties regarding statistical analyses.

Another interesting direction would be to consider other properties of kernel metrics that were described in the original paper, namely monotonicity and comparison properties. It would be challenging to understand how they could be generalized to BOST metrics or even to $\mathrm{O}(n)$-invariant metrics. Furthermore, to our knowledge there is no trace of families of non $\mathrm{O}(n)$-invariant metrics in the literature. However, there exist some situations where the $\mathrm{O}(n)$-invariance is not relevant, for example on correlation matrices because the space is not stable under this group action. This a promising perspective for future works.

Declaration of competing interest

None declared.

Data availability

No data was used for the research described in the article.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant G-Statistics agreement No 786854). This work has been supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01 and through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. The authors warmly thank the two anonymous reviewers for their very careful proofreading of this manuscript and Nicolas Guigui, Dimbihery Rabenoro and James Benn for insightful discussions and accurate suggestions to improve the paper.

Appendix A. Proofs on the Bures-Wasserstein metric

Proof of Levi-Civita connection in Table 7. Let X, Y be vector fields on $\mathrm{Sym}^+(n)$. The Levi-Civita connection is computed in [16]. With our notation $X^0 = \mathcal{S}_\Sigma(X)$ defined by $X = \Sigma X^0 + X^0\Sigma$, their result writes $\nabla_X Y = \partial_X Y - \{X^0 Y + Y^0 X\}_S + \{\Sigma X^0 Y^0 + \Sigma Y^0 X^0\}_S$ where $\{A\}_S = \frac{1}{2}(A + A^\top)$ is the symmetric part of the matrix A. It is easy to see that it rewrites $\nabla_X Y = \partial_X Y - (X^0\Sigma Y^0 + Y^0\Sigma X^0)$ which is a simpler expression.

We would like to give a different proof that relies on the geometry of the horizontal distribution. According to [20], Lemma 1, $d\pi(\nabla^G_{X^h} Y^h) = \nabla_X Y$, where $\nabla^G = \partial$

is the Levi-Civita connection of the Frobenius metric G on $\mathrm{GL}(n)$, i.e. the derivative of coordinates in the canonical basis of matrices. We differentiate the equality $X^h = (X^0 \circ \pi) \times \mathrm{Id}_{\mathrm{GL}(n)}$ on $\mathrm{GL}(n)$:

$$\begin{aligned}
(\nabla^G_{X^h} Y^h)_{|A} &= \partial_{X^h_A}(Y^0 \circ \pi)A + Y^0_{\pi(A)} X^h_A \\
&= (\partial_{X_{\pi(A)}} Y^0)A + Y^0_{\pi(A)} X^0_{\pi(A)} A, \\
(\nabla_X Y)_{|AA^\top} &= d_A\pi((\nabla^G_{X^h} Y^h)_{|A}) \\
&= AA^\top(\partial_{X_{\pi(A)}} Y^0) + (\partial_{X_{\pi(A)}} Y^0)AA^\top \\
&\quad + AA^\top X^0_{\pi(A)} Y^0_{\pi(A)} + Y^0_{\pi(A)} X^0_{\pi(A)} AA^\top, \\
(\partial_X Y)_{|\Sigma} &= \Sigma(\partial_{X_\Sigma} Y^0) + (\partial_{X_\Sigma} Y^0)\Sigma + X_\Sigma Y^0_\Sigma + Y^0_\Sigma X_\Sigma \\
&= \Sigma(\partial_{X_\Sigma} Y^0) + (\partial_{X_\Sigma} Y^0)\Sigma + \Sigma X^0_\Sigma Y^0_\Sigma + Y^0_\Sigma X^0_\Sigma \Sigma \\
&\quad + X^0_\Sigma \Sigma Y^0_\Sigma + Y^0_\Sigma \Sigma X^0_\Sigma \\
&= (\nabla_X Y)_{|\Sigma} + X^0_\Sigma \Sigma Y^0_\Sigma + Y^0_\Sigma \Sigma X^0_\Sigma.
\end{aligned}$$

Finally, we find $\nabla_X Y = \partial_X Y - (X^0 \Sigma Y^0 + Y^0 \Sigma X^0)$ as expected. □

Proof of curvature in Table 7. Let $X, Y \in T_\Sigma \mathrm{Sym}^+(n)$ be tangent vectors at $\Sigma \in \mathrm{Sym}^+(n)$. We would like to compute the sectional curvature $\kappa(X,Y) = \frac{R(X,Y,X,Y)}{\|X\|^2\|Y\|^2 - \langle X|Y\rangle^2}$, i.e. $R(X,Y,X,Y)$. Let $X^h, Y^h \in \mathcal{H}_{\Sigma^{1/2}}$ be the horizontal lifts of X, Y at $\Sigma^{1/2}$ and $X^0, Y^0 \in \mathrm{Sym}(n)$ defined as explained above. We extend X^h, Y^h into vector fields by $X^h_A := X^0 A$ and $Y^h_A := Y^0 A$. We do so because the formula we use to compute the curvature is based on a Lie bracket and can only be computed with fields. As the curvature is a tensor, it only depends on the values of X and Y at Σ so the way we extend the fields does not influence the result (but it simplifies the computation).

A first strategy to compute the curvature is to use the Levi-Civita connection via the definition $R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$. It is tedious but doable. Another one consists in using the relation between the curvatures of the quotient metric (here, Bures-Wasserstein) and the original metric (here, Frobenius) found in [20], formula {4}. According to this formula, since the Euclidean metric is flat, the formula is $R_\Sigma(X,Y,X,Y) = \frac{3}{4}\|\mathrm{ver}[\![X^h, Y^h]\!]_{\Sigma^{1/2}}\|^2$ where $\mathrm{ver} : X^v + X^h \in T\mathrm{GL}(n) \longmapsto X^v \in \mathcal{V}$ is the vertical projection and $[\![\cdot,\cdot]\!]$ denotes the Lie bracket on vector fields of $\mathrm{GL}(n)$, which must be distinguished from the matrix Lie bracket $[V,W] = VW - WV$. Note that the right term only depends on $X^h_{\Sigma^{1/2}}$ and $Y^h_{\Sigma^{1/2}}$ because if $f : \mathrm{GL}^+(n) \longrightarrow \mathbb{R}$ is a map, then $\mathrm{ver}[\![fX^h, Y^h]\!]_{\Sigma^{1/2}} = f(\Sigma^{1/2})\mathrm{ver}[\![X^h, Y^h]\!]_{\Sigma^{1/2}} + d_{\Sigma^{1/2}} f(Y^h) \underbrace{\mathrm{ver}(X^h)}_{0}$.

The rest of the proof consists in computing $\mathrm{ver}[\![X^h, Y^h]\!] = [\![X^h, Y^h]\!] - \mathrm{hor}[\![X^h, Y^h]\!]$. On the one hand, $[\![X^h, Y^h]\!]_A = Y^0 X^h_A - X^0 Y^h_A = -[X^0, Y^0]A$. On the other hand, let $Z^h_A := \mathrm{hor}[\![X^h, Y^h]\!]_A =: Z^0_{AA^\top} A \in \mathcal{H}_A$. Now, we can fix $\Sigma \in \mathrm{Sym}^+(n)$ and $A = \Sigma^{1/2}$. We

take a spectral decomposition $\Sigma = PDP^\top$ and we denote with a prime all the previous matrices taken in the basis P of eigenvectors of Σ, e.g. $X^{0\prime} = P^\top X^0 P$. Then:

$$\begin{aligned}
Z_\Sigma &:= d_{\Sigma^{1/2}}\pi([\![X^h, Y^h]\!]_{\Sigma^{1/2}}) = \Sigma[X^0, Y^0] - [X^0, Y^0]\Sigma,\\
Z^h_{\Sigma^{1/2}} &= (d_{\Sigma^{1/2}}\pi_{|\mathcal{H}_{\Sigma^{1/2}}})^{-1}(d_{\Sigma^{1/2}}\pi([\![X^h, Y^h]\!]_{\Sigma^{1/2}}))\\
&= (d_{\Sigma^{1/2}}\pi_{|\mathcal{H}_{\Sigma^{1/2}}})^{-1}(Z_\Sigma),\\
[Z^{0\prime}_\Sigma]_{ij} &= \frac{1}{d_i + d_j}(D[X^{0\prime}, Y^{0\prime}] - [X^{0\prime}, Y^{0\prime}]D)_{ij}\\
&= \frac{d_i - d_j}{d_i + d_j}[X^{0\prime}, Y^{0\prime}]_{ij},\\
[Z^{h\prime}_{\Sigma^{1/2}}]_{ij} &= \sqrt{d_j}[Z^{0\prime}_\Sigma]_{ij} = \sqrt{d_j}\frac{d_i - d_j}{d_i + d_j}[X^{0\prime}, Y^{0\prime}]_{ij},\\
(\mathrm{ver}[\![X^h, Y^h]\!]_{\Sigma^{1/2}})'_{ij} &= ([\![X^h, Y^h]\!]_{\Sigma^{1/2}})'_{ij} - [Z^{h\prime}_{\Sigma^{1/2}}]_{ij}\\
&= -[X^{0\prime}, Y^{0\prime}]_{ij}\sqrt{d_j} - \sqrt{d_j}\frac{d_i - d_j}{d_i + d_j}[X^{0\prime}, Y^{0\prime}]_{ij}\\
&= -\frac{2d_i\sqrt{d_j}}{d_i + d_j}[X^{0\prime}, Y^{0\prime}]_{ij},\\
R_\Sigma(X, Y, X, Y) &= \frac{3}{4}\|(\mathrm{ver}[\![X^h, Y^h]\!]_{\Sigma^{1/2}})'\|^2\\
&= 3\sum_{i,j}\frac{d_i^2 d_j}{(d_i + d_j)^2}\left[X^{0\prime}, Y^{0\prime}\right]^2_{ij}\\
&= \frac{3}{2}\sum_{i,j}\frac{d_i d_j}{d_i + d_j}\left[X^{0\prime}, Y^{0\prime}\right]^2_{ij},
\end{aligned}$$

where $P^\top X P = DX^{0\prime} + X^{0\prime}D$ and $P^\top Y P = DY^{0\prime} + Y^{0\prime}D$. □

Proof of geodesic parallel transport between commuting matrices in Table 7. We want to prove that the geodesic parallel transport of the Bures-Wasserstein metric between two commuting matrices is $\Pi_{\Sigma\to\Lambda}X = P\left[\sqrt{\frac{\delta_i+\delta_j}{d_i+d_j}}[P^\top X P]_{ij}\right]_{i,j}P^\top$ where $\Sigma = PDP^\top$ and $\Lambda = P\Delta P^\top \in \mathrm{Sym}^+(n)$. The geodesic parallel transport is $\mathrm{O}(n)$-invariant so we only need to prove that $[\Pi_{D\to\Delta}X]_{ij} = \sqrt{\frac{\delta_i+\delta_j}{d_i+d_j}}X_{ij}$. The geodesic from D to Δ is $\gamma(t) = ((1-t)\sqrt{D} + t\sqrt{\Delta})^2$. Let us define $X(t) = \left[\sqrt{\frac{((1-t)d_i+t\delta_i)^2+((1-t)d_j+t\delta_j)^2}{d_i+d_j}}X_{ij}\right]_{i,j}$ and let us check that $\nabla_{\dot\gamma}X = 0$. We compute:

$$\begin{aligned}
[X^0(t)]_{ij} &= \frac{[X(t)]_{ij}}{\gamma_i(t) + \gamma_j(t)} = \frac{1}{\sqrt{d_i + d_j}\sqrt{((1-t)d_i + t\delta_i)^2 + ((1-t)d_j + t\delta_j)^2}}X_{ij},\\
\dot\gamma(t) &= 2(\sqrt{\Delta} - \sqrt{D})((1-t)\sqrt{D} + t\sqrt{\Delta}),
\end{aligned}$$

$$\begin{aligned}
\dot\gamma^0(t) &= \frac{1}{2}\dot\gamma(t)\gamma^{-1}(t) = \frac{1}{2}\gamma^{-1}(t)\dot\gamma(t) = (\sqrt{\Delta}-\sqrt{D})((1-t)\sqrt{D}+t\sqrt{\Delta})^{-1},\\
[\dot X(t)]_{ij} &= \frac{2(\sqrt{\delta_i}-\sqrt{d_i})((1-t)d_i+t\delta_i)+2(\sqrt{\delta_j}-\sqrt{d_j})((1-t)d_j+t\delta_j)}{2\sqrt{d_i+d_j}\sqrt{((1-t)d_i+t\delta_i)^2+((1-t)d_j+t\delta_j)^2}}X_{ij}\\
&= (\sqrt{\delta_i}-\sqrt{d_i})((1-t)d_i+t\delta_i)[X^0(t)]_{ij}\\
&\quad + [X^0(t)]_{ij}(\sqrt{\delta_j}-\sqrt{d_j})((1-t)d_j+t\delta_j)\\
&= [\dot\gamma^0(t)\gamma(t)X^0(t)+X^0(t)\gamma(t)\dot\gamma^0(t)]_{ij},\\
\nabla_{\dot\gamma(t)}X &= \dot X(t)-(\dot\gamma^0(t)\gamma(t)X^0(t)+X^0(t)\gamma(t)\dot\gamma^0(t)) = 0.
\end{aligned}$$

So the geodesic parallel transport from $\Sigma = PDP^\top$ to $\Lambda = P\Delta P^\top$ is $\Pi_{\Sigma\to\Lambda}X = P\left[\sqrt{\frac{\delta_i+\delta_j}{d_i+d_j}}[P^\top XP]_{ij}\right]_{i,j}P^\top$. □

Proof of equation of the geodesic parallel transport in Table 7. The geodesic parallel transport equation is $\nabla_{\dot\gamma(t)}X = 0$ along the geodesic $\gamma(t) = \gamma^h(t)\gamma^h(t)^\top$ between Σ and $\Lambda \in \mathrm{Sym}^+(n)$, where $\gamma^h(t) = (1-t)\Sigma^{1/2} + t\Sigma^{-1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}$. For a vector field $X(t)$ on $\mathrm{Sym}^+(n)$ defined along $\gamma(t)$, we can define the horizontal lift $X^h(t) = X^0(t)\gamma^h(t) \in \mathcal{H}_{\gamma^h(t)}$ where $X^0(t)$ is defined by $X(t) = \gamma(t)X^0(t) + X^0(t)\gamma(t)$. We are going to prove that $X(t)$ is the geodesic parallel transport of $X \in T_\Sigma\mathrm{Sym}^+(n)$ if and only if $X^0(t)$ satisfies the following ODE:

$$\gamma(t)\dot X^0(t) + \dot X^0(t)\gamma(t) + \gamma^h(t)\dot\gamma^{h\top}X^0(t) + X^0(t)\dot\gamma^h\gamma^h(t)^\top = 0. \tag{A.1}$$

To rewrite the geodesic parallel transport equation $\nabla_{\dot\gamma(t)}X = 0$, we need to compute the following derivatives:

$$\begin{aligned}
\dot X(t) &= \gamma(t)\dot X^0(t) + \dot X^0(t)\gamma(t) + \dot\gamma(t)X^0(t) + X^0(t)\dot\gamma(t),\\
\dot\gamma(t) &= \dot\gamma^h\gamma^h(t)^\top + \gamma^h(t)\dot\gamma^{h\top} \text{ where } \dot\gamma^h = \dot\gamma^0(t)\gamma^h(t).
\end{aligned}$$

Now, we simply rewrite the equation:

$$\begin{aligned}
\nabla_{\dot\gamma(t)}X = 0 &\iff \dot X(t) - (\dot\gamma^0(t)\gamma(t)X^0(t) + X^0(t)\gamma(t)\dot\gamma^0(t)) = 0\\
&\iff \gamma(t)\dot X^0(t) + \dot X^0(t)\gamma(t)\\
&\qquad + (\dot\gamma(t) - \dot\gamma^h\gamma^h(t)^\top)X^0(t) + X^0(t)(\dot\gamma(t) - \gamma^h(t)\dot\gamma^{h\top}) = 0\\
&\iff \gamma(t)\dot X^0(t) + \dot X^0(t)\gamma(t) + \gamma^h(t)\dot\gamma^{h\top}X^0(t) + X^0(t)\dot\gamma^h\gamma^h(t)^\top = 0. \quad \square
\end{aligned}$$

References

[1] S.i. Amari, H. Nagaoka, Methods of Information Geometry, vol. 191, Oxford University Press, 2000.
[2] V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Log-Euclidean metrics for fast and simple calculus on diffusion tensors, Magn. Reson. Med. 56 (2006) 411–421.

[3] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a Riemannian-based kernel for BCI applications, Neurocomputing 112 (2013) 172–178.
[4] P.G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, A. Connelly, A rigorous framework for diffusion tensor calculus, Magn. Reson. Med. 53 (2005) 221–225.
[5] R. Bhatia, Matrix Analysis, Graduate Texts in Mathematics, vol. 169, Springer New York, New York, NY, 1997.
[6] R. Bhatia, T. Jain, Y. Lim, On the Bures–Wasserstein distance between positive definite matrices, Expo. Math. 37 (2019) 165–191.
[7] I.L. Dryden, A. Koloydenko, D. Zhou, Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging, Ann. Appl. Stat. 3 (2009) 1102–1123.
[8] I.L. Dryden, X. Pennec, J.M. Peyrat, Power Euclidean metrics for covariance matrices with application to diffusion tensor imaging, 2010, ArXiv e-prints.
[9] P.T. Fletcher, S. Joshi, Riemannian geometry for the statistical analysis of diffusion tensor data, Signal Process. 87 (2007) 250–262.
[10] M. Ha Quang, A unified formulation for the Bures-Wasserstein and Log-Euclidean/Log-Hilbert-Schmidt distances between positive definite operators, in: Proceedings of GSI 2019 - 4th Conference on Geometric Science of Information, Springer International Publishing, Toulouse, France, 2019, pp. 475–483.
[11] F. Hiai, D. Petz, Riemannian metrics on positive definite matrices related to means, Linear Algebra Appl. 430 (2009) 3105–3130.
[12] F. Hiai, D. Petz, Riemannian metrics on positive definite matrices related to means. II, Linear Algebra Appl. 436 (2012) 2117–2136.
[13] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing, J. Math. Imaging Vis. 25 (2006) 423–444.
[14] P. Li, Q. Wang, H. Zeng, L. Zhang, Local log-Euclidean multivariate Gaussian descriptor and its application to image classification, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 803–817.
[15] Z. Lin, Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition, SIAM J. Matrix Anal. Appl. 40 (2019) 1353–1370.
[16] L. Malagò, L. Montrucchio, G. Pistone, Wasserstein Riemannian geometry of Gaussian densities, Inf. Geom. 1 (2018) 137–179.
[17] P.W. Michor, D. Petz, A. Andai, The curvature of the Bogoliubov-Kubo-Mori scalar product on matrices, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 3 (2000) 1–14.
[18] N. Miolane, N. Guigui, A. Le Brigant, J. Mathe, B. Hou, Y. Thanwerdas, S. Heyder, O. Peltre, N. Koep, H. Zaatiti, H. Hajri, Y. Cabanes, T. Gerald, P. Chauchat, C. Shewmake, D. Brooks, B. Kainz, C. Donnat, S. Holmes, X. Pennec, Geomstats: a Python package for Riemannian geometry in machine learning, J. Mach. Learn. Res. 21 (2020) 1–9.
[19] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (2005) 735–747.
[20] B. O'Neill, The fundamental equations of a submersion, Mich. Math. J. 13 (1966) 459–469.
[21] J. van Oostrum, Bures-Wasserstein geometry for positive-definite Hermitian matrices and their trace-one subset, Inf. Geom. 5 (2022) 405–425.
[22] X. Pennec, Statistical computing on manifolds: from Riemannian geometry to computational anatomy, in: Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, November 18-20, 2008, in: Lecture Notes in Computer Science, vol. 5416, Springer, Berlin, Heidelberg, 2009, pp. 347–386. Revised Invited Papers.
[23] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis. 66 (2006) 41–66.
[24] D. Petz, G. Toth, The Bogoliubov inner product in quantum statistics, Lett. Math. Phys. 27 (1993) 205–216.
[25] J.C. Pinheiro, D.M. Bates, Unconstrained parametrizations for variance-covariance matrices, Stat. Comput. 6 (1996) 289–296, https://doi.org/10.1007/BF00140873.
[26] C.R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37 (1945) 81–91.
[27] C.L. Siegel, Symplectic geometry, Am. J. Math. 65 (1943) 1–86.
[28] L.T. Skovgaard, A Riemannian geometry of the multivariate normal model, Scand. J. Stat. 11 (1984) 211–223.
[29] J. Su, I.L. Dryden, E. Klassen, H. Le, A. Srivastava, Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds, Image Vis. Comput. 30 (2012) 428–442.

[30] A. Takatsu, On Wasserstein geometry of Gaussian measures, in: M. Kotani, M. Hino, T. Kumagai (Eds.), Probabilistic Approach to Geometry, Mathematical Society of Japan, Kyoto University, Japan, 2010, pp. 463–472.
[31] A. Takatsu, Wasserstein geometry of Gaussian measures, Osaka J. Math. 48 (2011) 1005–1026.
[32] Y. Thanwerdas, X. Pennec, Exploration of balanced metrics on symmetric positive definite matrices, in: Proceedings of GSI 2019 - 4th Conference on Geometric Science of Information, Springer International Publishing, Toulouse, France, 2019, pp. 484–493.
[33] Y. Thanwerdas, X. Pennec, Is affine-invariance well defined on SPD matrices? A principled continuum of metrics, in: Proceedings of GSI 2019 - 4th Conference on Geometric Science of Information, Springer International Publishing, Toulouse, France, 2019, pp. 502–510.
[34] G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, B. Thirion, Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling, in: D. Shen, A. Frangi, G. Szekely (Eds.), Medical Image Computing and Computer Added Intervention, Tianzi Jiang, Springer, Beijing, China, 2010, pp. 200–208.
[35] Z. Wang, B.C. Vemuri, Y. Chen, T.H. Mareci, A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI, IEEE Trans. Med. Imaging 23 (2004) 930–939.
[36] Z. Zhang, J. Su, E. Klassen, H. Le, A. Srivastava, Rate-invariant analysis of covariance trajectories, J. Math. Imaging Vis. 60 (2018) 1306–1323.