

## Bi-Invariant Dissimilarity Measures for Sample Distributions in Lie Groups\*

Martin Hanik<sup>†</sup>, Hans-Christian Hege<sup>†</sup>, and Christoph von Tycowicz<sup>‡</sup>

**Abstract.** Data sets sampled in Lie groups are widespread, and as with multivariate data, it is important for many applications to assess the differences between the sets in terms of their distributions. Indices for this task are usually derived by considering the Lie group as a *Riemannian* manifold. Then, however, compatibility with the group operation is guaranteed only if a bi-invariant metric exists, which is not the case for most noncompact and noncommutative groups. We show here that if one considers an affine connection structure instead, one obtains bi-invariant generalizations of well-known dissimilarity measures: *Hotelling T<sup>2</sup> statistic*, *Bhattacharyya distance*, and *Hellinger distance*. Each of the dissimilarity measures matches its multivariate counterpart for Euclidean data and is translation invariant, so that biases, e.g., through an arbitrary choice of reference, are avoided. We further derive nonparametric two-sample tests that are bi-invariant and consistent. We demonstrate the potential of these dissimilarity measures by performing group tests on data of knee configurations and epidemiological shape data. Significant differences are revealed in both cases.

**Key words.** nonmetric statistics, two-sample test, Lie group, shape analysis

**MSC codes.** 62R30, 22E70, 53Z50

**DOI.** 10.1137/21M1410373

**1. Introduction.** Manifold-valued data occur in many applications [18, 52, 56, 59] and it can often be conceived of as elements of a Lie group. Examples include representations of skeletal systems (e.g., in robotics [51]), inferences of anatomical structures by evaluating diffusion tensor fields in medicine [6–8, 55], position- and motion-independent recognition of objects in computer vision [38, 72–74], or analysis of covariance matrices (e.g., in feature-based image analysis [71]). The algorithmic tasks involved in Lie group-based computations are correspondingly diverse: They range from computation of geometric means [44, 45], to function approximation and regression [33, 61, 76], numerical solution of differential equations [40], numerical minimization [66], signal processing [9, 10, 16, 28], image analysis [11, 24], as well as computer vision [15, 48, 49, 57, 68, 69].

Another area where Lie-group-valued measurements are regularly performed is shape analysis: The idea is to represent shapes of objects as deformations of a common reference, as introduced by D'Arcy Thompson over 100 years ago [67]; the deformations are usually el-

\* Received by the editors April 7, 2021; accepted for publication (in revised form) March 9, 2022; published electronically November 15, 2022. This work is an extension of the conference paper [*Shape in Medical Imaging*, Springer, Cham, Switzerland, 2020, pp. 44–54].

<https://doi.org/10.1137/21M1410373>

**Funding:** This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) through BIFOLD - The Berlin Institute for the Foundations of Learning and Data (01IS18025A and 01IS18037A). The first author is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy. The Berlin Mathematics Research Center MATH+ (EXC-2046/1, 390685689).

<sup>†</sup> Zuse Institute Berlin, 14195 Berlin, Germany ([hanik@zib.de](mailto:hanik@zib.de), [hege@zib.de](mailto:hege@zib.de)).

<sup>‡</sup> Freie Universität Berlin, 14195 Berlin, Germany ([vontycowicz@zib.de](mailto:vontycowicz@zib.de)).

ements of a Lie group. For example in [1, 13], configurations of the human spine are encoded in a product group consisting of translations and rotations. Classical matrix groups are also used in physically motivated shape spaces [4, 75] as well as in the characterization of volume [77] and surface [3] deformations. Important algorithmic tasks are matching, analysis, and statistics of shapes [29, 39, 64, 65, 70].

For data analysis in Lie groups, it is necessary to generalize methods of multivariate statistics to the Lie group setting. Keeping in mind that Lie groups possess symmetries since they act on themselves via translations, it is desirable to look for generalizations that respect these symmetries. The task, then, is to derive methods that are invariant/equivariant under in-group translations (from both left and right). This property is not only an established theoretical criterion for selection of statistical methods [32], but it also has practical advantages: In shape analysis, for example, it avoids a bias due to the choice of reference and, at the same time, different data configurations [1, 3, 77]. In machine learning applications, equivariant convolutional networks have been shown to perform very well on Lie groups [22, 27].

Although geometrically defined statistical methods from Riemannian geometry have long been used, they respect the symmetries only if a bi-invariant Riemannian metric exists on the Lie group at hand. However, this is often not the case; particularly relevant examples for applications are the group of rigid body transformations and the general linear group in dimensions greater than one.

To overcome the problems of the Riemannian approach, Pennec and Arsigny generalized the notions of mean, covariance, and Mahalanobis distance to Lie groups using a (non-metric) affine structure [53]. The quantities defined in this way have the desired invariance/equivariance properties under translations. We build on this work to derive bi-invariant (i.e., invariant under translation from both left and right) generalizations of the Hotelling  $T^2$  statistic and Bhattacharyya distance for data in Lie groups that follow a generalized normal distribution. Both are well-known indices from multivariate statistics and are used to measure dissimilarities between two probability distributions. As such they have found numerous applications, for example, in statistical hypothesis testing [50], feature extraction [19], and image processing [31]. In contrast to generalizations based on Riemannian structures [37, 47], the proposed quantities are not only bi-invariant, but also reduce to the multivariate formulations in the case of Euclidean spaces. To illustrate the potential, we use the newly defined notions to construct hypothesis tests for two prospective applications. First, we detect differences in the configuration of human knees under osteoarthritis when compared to healthy controls. Second, we test for differences in shape of the right hippocampus of patients in early Alzheimer's disease; thereby, both a local and a global test confirm effects known from the literature. Python implementations of our proposed methods, which were also used for the tests, are available online as part of the Morphomatics library [2].

This article is based in part on the workshop paper [34], but includes both theoretical and experimental extensions: We provide a rigorous mathematical formulation of the underlying concepts, including new discussions of the symmetry properties of the proposed indices under an exchange of the data sets and on the behavior of the indices under inversion of the data. Furthermore, we derive the connection between the bi-invariant Bhattacharyya distance and its version for densities and propose a bi-invariant Hellinger distance. Finally, we extend the previous experiments to include a global hypothesis test for the equality of shape distributions

and a local test based on the Bhattacharyya distance and perform a new hypothesis test for differences in the SE(3)-valued configuration of the knee under osteoarthritis.

**2. Theoretical background.** Let  $(p_1, \dots, p_m)$  and  $(q_1, \dots, q_n)$  be two data sets in  $\mathbb{R}^d$ , each with independent and identically normally distributed elements, and let  $\bar{p}$  and  $\bar{q}$  denote the respective sample means. Assuming homoscedasticity (i.e., that both distributions share the same covariance matrix) the data's pooled sample covariance is given by

$$(2.1) \quad \widehat{S} = \frac{1}{m+n-2} \left( \sum_{i=1}^m (p_i - \bar{p})(p_i - \bar{p})^T + \sum_{j=1}^n (q_j - \bar{q})(q_j - \bar{q})^T \right).$$

Since the covariances are the same,  $\widehat{S}$  is an unbiased estimator of the pooled covariance. The multivariate Hotelling  $T^2$  statistic is then defined as the square of the Mahalanobis distance scaled by  $mn/(m+n)$ :

$$(2.2) \quad t^2((p_i), (q_i)) = \frac{mn}{m+n} (\bar{p} - \bar{q})^T \widehat{S}^{-1} (\bar{p} - \bar{q}).$$

It measures the difference of  $\bar{p}$  and  $\bar{q}$  weighted by the inverse of the pooled covariance. Therefore, directions in which high variability was observed are weighted less than those with little spreading around the corresponding components of the mean.

If additionally the variances of both distributions differ, the Bhattacharyya distance [41] is a suggested index for assessing the dissimilarity between them. Denoting the matrix determinant by  $\det$ , the respective sample covariance matrices by  $S_{p_i}$  and  $S_{q_i}$ , and setting  $\bar{S} = (S_{p_i} + S_{q_i})/2$ , it is defined by

$$(2.3) \quad D_B((p_i), (q_i)) = \frac{1}{8} (\bar{p} - \bar{q})^T \bar{S}^{-1} (\bar{p} - \bar{q}) + \frac{1}{2} \ln \left( \frac{\det(\bar{S})}{\sqrt{\det(S_{p_i}) \det(S_{q_i})}} \right).$$

Note that the first summand (without the constant factor) differs from Hotelling's  $T^2$  statistic only by the use of the averaged covariance matrix instead of the pooled one.

A fundamental property of both the Hotelling's  $T^2$  statistic and the Bhattacharyya distance is invariance under data translations (e.g.,  $t^2((p_i + v), (q_i + v)) = t^2((p_i), (q_i))$  for all  $v \in \mathbb{R}^d$ ). It is thus highly desirable that generalizations of these quantities also exhibit this invariance property. We achieve this for Lie groups by considering them as affine manifolds.

**2.1. Affine manifolds and Lie groups.** In the following, we summarize the relevant facts of affine manifolds and (finite-dimensional) Lie group theory; we also explain how the Riemannian is integrated in the affine setting. For more information on Lie groups see [35] and [58]; the latter is also a good reference on affine manifolds. Additional information on differential geometry can, for example, be found in [23]. In the following we use “smooth” synonymously with “infinitely often differentiable.”

Let  $M$  be a smooth manifold. We denote the tangent space at  $p \in M$  by  $T_p M$ , and the sets of smooth functions and vector fields on  $M$  by  $\mathcal{C}^\infty(M)$  and  $\Gamma(TM)$ , respectively. If  $X \in \Gamma(TM)$ , then  $X_p$  denotes its value in  $T_p M$ . Given  $X, Y \in \Gamma(TM)$ , an affine connection  $\nabla$  yields a way to differentiate  $Y$  along  $X$ ; the result  $\nabla_X Y \in \Gamma(TM)$  being again a vector field. If

$M$  is *additionally* endowed with a Riemannian metric (i.e., a smoothly varying inner product on the tangent spaces), then (for every such metric) there is a unique affine connection called a Levi–Civita connection. Whenever we speak of a *Riemannian* manifold in the following, we mean a smooth manifold that is endowed with the Levi–Civita connection of some arbitrary Riemannian metric.

With  $\gamma' = \frac{d\gamma}{dt}$ , we can define a geodesic  $\gamma : [0, 1] \rightarrow M$  by  $\nabla_{\gamma'} \gamma' = 0$  as a curve without acceleration. An important fact is that every point  $p \in M$  has a so-called normal convex neighborhood  $U$ . Each pair  $q, r \in U$  can be joined by a unique geodesic  $[0, 1] \ni t \mapsto \gamma(t; p, q)$  that lies completely in  $U$ . Furthermore, with  $\gamma'(0; p, q) = v$ , this defines the *exponential*  $\text{Exp}_p : T_p M \rightarrow M$  of the connection  $\nabla$  at  $p$  by

$$\text{Exp}_p(v) = \gamma(1; p, q).$$

It is a local diffeomorphism with local inverse

$$\text{Log}_p(q) = \gamma'(0; p, q).$$

In the case of Riemannian manifolds,  $\text{Exp}$  and  $\text{Log}$  are called *Riemannian exponential* and *logarithm*, respectively.

A Lie group  $G$  is a smooth manifold that (in addition) has a compatible group structure; that is, there is a smooth (not necessarily commutative) group operation  $G \times G \ni (g, h) \mapsto gh \in G$  with corresponding identity element  $e \in G$  such that the inversion map  $g \mapsto g^{-1}$  is also smooth. Vector spaces (with addition) are instances of Lie groups. Another example is the general linear group  $\text{GL}(n)$ , that is, the set of all bijective linear mappings on an  $n$ -dimensional vector space  $V$ , where the group operation is the composition of mappings (i.e., a matrix multiplication) with  $e$  being the identity map. Whenever we speak of matrix groups in the following, arbitrary closed subgroups of  $\text{GL}(n)$  are meant.

For each  $g \in G$  the group operation defines two automorphisms on  $G$ : the *left* and *right* translation  $L_g : h \mapsto gh$  and  $R_g : h \mapsto hg$ . Their derivatives  $d_h L_g$  and  $d_h R_g$  at  $h \in G$  map tangent vectors  $v \in T_h G$  bijectively to the tangent spaces  $T_{gh} G$  and  $T_{hg} G$ , respectively. In particular, it holds that

$$T_g G = \{d_e L_g(v) : v \in T_e G\} = \{d_e R_g(w) : w \in T_e G\}.$$

Thus, each  $X_e \in T_e G$  determines a smooth vector field  $X \in \Gamma(TG)$  by  $X_g = d_e L_g(X_e)$  for all  $g \in G$ . It is called left invariant because  $X_{L_g(h)} = d_h L_g(X_h)$  for all  $h \in G$ , that is, the value at a left translated point is the left translated vector. (For matrix groups with identity matrix  $I$  we get the simple equation  $d_I L_A(M) = AM$  for an element  $A$  and a matrix  $M$  in the tangent space at  $I$ .) Together with the Lie bracket they form the so-called Lie algebra  $\mathfrak{g} \subset \Gamma(TG)$  of  $G$ . (Remember that the *Lie bracket* of  $X, Y \in \Gamma(TG)$  is defined by  $[X, Y] = XY - YX \in \Gamma(TG)$ .) Furthermore, the converse also holds: Every left invariant

---

<sup>1</sup>This is to be understood as a differential operator, i.e., if  $\psi \in \mathcal{C}^\infty(G)$ , then  $[X, Y]\psi \in \mathcal{C}^\infty(G)$ , where  $([X, Y]\psi)(g)$  measures the failure to commute when taking directional derivatives of  $\psi$  at  $g$  in directions  $X_g$  and  $Y_g$ .

vector field is uniquely determined by its value at the identity. Consequently,  $T_e G$  and  $\mathfrak{g}$  are isomorphic.<sup>2</sup>

Since  $d_e L_g$  is bijective for every  $g \in G$  (and  $L_g$  smooth in  $g$ ), any basis of  $T_e G$  determines a unique global frame (a smooth choice of basis for each tangent space) of left invariant vector fields. In particular, for each  $g \in G$  any basis of  $T_e G$  can be smoothly transported to a basis of  $T_g G$  without dependence on the “path” (“take the basis from the corresponding global frame”). Because of this, we can view  $T_e G$  as the *reference tangent space* of  $G$ .

Of course, right invariant vector fields are defined analogously and have parallel properties to left invariant fields.

The integral curve  $\alpha_X : \mathbb{R} \rightarrow G$  through  $e$  of an invariant (left or right) vector field  $X \in \Gamma(TG)$  determines a unique 1-parameter subgroup of  $G$  since  $\alpha_X(s+t) = \alpha_X(s)\alpha_X(t)$  for all  $s, t \in \mathbb{R}$ . The *group exponential*  $\exp : T_e G \rightarrow G$  is then defined by

$$\exp(X_e) = \alpha_X(1).$$

It is also a diffeomorphism in a neighborhood  $V$  of  $e$  and, hence, we can define the *group logarithm*  $\log$  as its inverse there. In the case of matrix groups they coincide with the matrix exponential and logarithm. Important for us will be the inverse consistency of the group logarithm, that is, for all  $g \in V$

$$(2.4) \quad \log(g^{-1}) = -\log(g).$$

Central to our constructions will be the *canonical Cartan-Shouten (CCS) connection*  $\nabla$  of  $G$ . On  $\mathfrak{g}$ , it is defined by

$$\nabla_X Y = \frac{1}{2}[X, Y], \quad X, Y \in \mathfrak{g};$$

see [58, Chap. 6]. We can extend it to general vector fields, since  $\mathfrak{g}$  provides global frames. If we endow  $G$  with its CCS connection, then geodesics and left (or right) translated 1-parameter subgroups coincide; that is, for any geodesic  $\gamma$  in  $G$  with  $\gamma(0) = g$  both  $t \mapsto g^{-1}\gamma(t)$  and  $t \mapsto \gamma(t)g^{-1}$  are 1-parameter subgroups of  $G$ . Thus, on a Lie group  $G$  endowed with the CCS connection, the local exponential and logarithm of the connection are given by

$$(2.5) \quad \begin{aligned} \text{Exp}_g(v) &= g \exp(d_g L_{g^{-1}}(v)) = \exp(d_g R_{g^{-1}}(v)) g, & v \in T_g G, \\ \text{Log}_g(h) &= d_e L_g \log(g^{-1}h) = d_e R_g \log(hg^{-1}), & h \in U; \end{aligned}$$

see [55, Cor. 5.1]. On the other hand, the CCS connection is the Levi–Civita connection of a Riemannian metric if and only if the latter is bi-invariant, that is, invariant under left *and* right translations. This is the only case in which  $\text{Exp}_e \equiv \exp$  and  $\text{Log}_e \equiv \log$  on a Lie group with Riemannian structure; only then do geodesics and translated 1-parameter subgroups coincide. Unfortunately, there are many Lie groups that do not have a bi-invariant metric. A well-known example is the group of rigid body transformations  $\text{SE}(3)$  (or, more generally, the special Euclidean group  $\text{SE}(n)$  for  $n > 1$ ). Indeed, any Lie group that is not a direct product of

---

<sup>2</sup>To be exact,  $T_e G$  with the bracket  $[X_e, Y_e] := [X, Y]_e$  is isomorphic to  $\mathfrak{g}$  with its Lie bracket.

compact and commutative groups does not have a bi-invariant metric [43], including  $\mathrm{GL}(n)$ ,  $\mathrm{GL}^+(n)$ , the special linear group  $\mathrm{SL}(n)$  (i.e., all linear, orientation and volume preserving transformations of  $\mathbb{R}^n$ ), and the Heisenberg group [54] (again, all for  $n > 1$ ).

Another fundamental automorphism of  $G$  is the conjugation  $C_g : h \mapsto ghg^{-1}$ . Important to us will be its differential at the identity, which we call *group adjoint* and denote by  $\mathrm{Ad}(g)$ . It acts bijectively on vectors  $v \in T_e G$  by

$$\mathrm{Ad}(g)v = d_{g^{-1}}L_g(d_e R_{g^{-1}}(v)) = d_g R_{g^{-1}}(d_e L_g(v)).$$

For matrix groups this reduces to  $\mathrm{Ad}(A)(M) = AMA^{-1}$  for elements  $A$  and matrices  $M$  in the tangent space at the identity. The group adjoint yields the following crucial relation [53, Thm. 6]: For  $f, g \in G$  such that  $\log(fg^{-1})$  exists, it links logarithms of left and right translated points according to

$$(2.6) \quad \log(gf^{-1}) = \mathrm{Ad}(f)\log(f^{-1}g).$$

**2.2. The group mean.** The fact that bi-invariant metrics generally do not exist on Lie groups makes the construction of (statistical) notions that are invariant/equivariant under translations from both left and right very difficult in the Riemannian setting. On the other hand, Lie groups that are endowed with their CCS connections possess the natural prerequisites. Following Pennec and Arsigny, we define the mean in a Lie group as an exponential barycenter [53, Def. 3].

**Definition 2.1 (group mean).** Let  $G$  be a Lie group endowed with its CCS connection and let  $g_1, \dots, g_m \in V \subseteq G$ ,  $V$  being simply connected, such that  $\log(g_i^{-1}g_i)$  exists for all  $g \in V$  and  $i = 1, \dots, m$ . Then, we call  $\bar{g} \in G$  the group mean of  $g_1, \dots, g_m$  if

$$(2.7) \quad \sum_{i=1}^m \mathrm{Log}_{\bar{g}}(g_i) = 0.$$

Note that, because of (2.5), (2.7) is equivalent to both

$$\sum_{i=1}^m \log(\bar{g}^{-1}g_i) = 0 \quad \text{and} \quad \sum_{i=1}^m \log(g_i\bar{g}^{-1}) = 0.$$

From this we can see that the elements  $g_1, \dots, g_m$  must be “sufficiently localized” to obtain a mean value, because if they are too far apart, the logarithm may not be defined. The most general result known so far on existence and uniqueness is the following: If  $U \subseteq G$  is a CSLCG (convex with semilocal convex geometry) neighborhood and  $g_1, \dots, g_m \in U$ , then their group mean exists and is unique. The details can be found in [Appendix A](#).

Central properties of the group mean are summarized in the following theorem [54, Thm. 5.13].

**Theorem 2.2 (equivariance of the group mean).** Let  $G$  be a Lie group endowed with its CCS connection and  $\bar{g}$  be a group mean of  $g_1, \dots, g_m \in V \subseteq G$ . Then, for any  $f \in G$ , the group means of the left translated data  $(fg_1, \dots, fg_m)$ , right translated data  $(g_1f, \dots, g_mf)$ , and inverted data  $(g_1^{-1}, \dots, g_m^{-1})$  are  $f\bar{g}$ ,  $\bar{g}f$ , and  $\bar{g}^{-1}$ , respectively.

According to the theorem, group means are equivariant under the group operations. These are very favorable properties that make (statistical) data analysis more robust.

Finally, group means can be computed efficiently with a fixed point iteration [54, Thm. 5.14].

**2.3. Sample covariance and bi-invariant Mahalanobis distance.** In [53], Pennec and Arsigny also define the sample covariance of Lie-group-valued data  $(g_1, \dots, g_m)$  with group mean  $\bar{g}$ . Denoting the vector of coordinates of a tangent vector as well as the representing matrix of a (bi)linear map in any given basis by  $[\cdot]$ , it is the (2,0)-tensor

$$\Sigma_{g_i}^* = \frac{1}{m} \sum_{l=1}^m \text{Log}_{\bar{g}}(g_l) \otimes \text{Log}_{\bar{g}}(g_l) \in T_{\bar{g}}G \otimes T_{\bar{g}}G,$$

where the tensor product  $\otimes$  means that in any basis of  $T_{\bar{g}}G$ , the representing matrix is given by  $[\Sigma_{g_i}^*] = 1/m \sum_l [\text{Log}_{\bar{g}}(g_l)][\text{Log}_{\bar{g}}(g_l)]^T$ . From this, they define the *bi-invariant Mahalanobis distance* of  $f \in G$  to the distribution of the  $g_i$  by

$$(2.8) \quad \mu_{(\bar{g}, \Sigma_{g_i}^*)}^2(f) = \left[ \text{Log}_{\bar{g}}(f) \right]^T \left[ \Sigma_{g_i}^{*-1} \right] \left[ \text{Log}_{\bar{g}}(f) \right],$$

if  $\text{Log}_{\bar{g}}(f) = \log(\bar{g}^{-1}f)$  exists. (Here,  $\Sigma_{g_i}^{*-1}$  denotes the inverse of  $\Sigma_{g_i}^*$ .) The distance (2.8) is left and right invariant, that is,

$$\mu_{(\bar{g}, \Sigma_{g_i}^*)}^2(f) = \mu_{(h\bar{g}, \Sigma_{h g_i}^*)}^2(hf) = \mu_{(\bar{g}h, \Sigma_{g_i h}^*)}^2(fh)$$

for all  $h \in G$  and all  $f$  such that  $\text{Log}_{\bar{g}}(f)$  exists; see [54, p. 181].

**3. Bi-invariant dissimilarity measures for sample distributions on Lie groups.** In this section we derive bi-invariant extensions of the Hotelling  $T^2$  statistic and Bhattacharyya distance to a Lie group  $G$  that is endowed with its CCS connection. To this end, we always consider a CSLCG neighborhood  $U$  of a Lie group  $G$ .

**3.1. Extending the Hotelling  $T^2$  statistic.** To generalize Hotelling's  $T^2$  statistic, we define a slightly modified sample covariance. This is motivated by the fact that  $T_eG$  is the reference tangent space for the whole group.

**Definition 3.1 (centralized sample covariance).** Given data  $(g_1, \dots, g_m)$  in  $U$  with group mean  $\bar{g}$ , its left-centralized sample covariance is defined as the (2,0)-tensor

$$\Sigma_{g_i} = \frac{1}{m} \sum_{l=1}^m \log(\bar{g}^{-1}g_l) \otimes \log(\bar{g}^{-1}g_l) \in T_eG \otimes T_eG,$$

which in local coordinates is

$$\left[ \Sigma_{g_i} \right] = \frac{1}{m} \sum_{l=1}^m \left[ \log(\bar{g}^{-1}g_l) \right] \left[ \log(\bar{g}^{-1}g_l) \right]^T.$$

Since we only use the centralized covariance that is defined under left translations of the data, we will drop the word "left" in the following. The centralized covariance will be fundamental for our constructions since it allows us to compare the variability of data sets with different means in a single tangent space; this, in turn, makes it possible to transfer

notions from multivariate statistics. Taking advantage of the bi-invariance property we also have

$$\mu_{(\bar{g}, \Sigma_{g_i}^*)}^2(f) = \mu_{(e, \Sigma_{g_i})}^2(\bar{g}^{-1}f).$$

Moreover, we can now define the pooled covariance at the following identity.

**Definition 3.2 (pooled sample covariance).** *Given data sets  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  in  $U$  with group means  $\bar{g}$  and  $\bar{h}$ , their left-pooled sample covariance is defined by*

$$\widehat{\Sigma}_{g_i, h_i} = \frac{1}{m+n-2} (m\Sigma_{g_i} + n\Sigma_{h_i}).$$

As for the centralized covariance, we drop the word left for the pooled covariance. Furthermore, because there is no danger of confusion, we only write  $\widehat{\Sigma}$  from now on.

In order to obtain bi-invariant indices from the newly defined notions, we need to understand how they transform under joint translations of the data. For this, we denote the centralized covariance of data that was jointly left or right translated with  $f \in G$  by  $f \bullet \Sigma_{g_i}$  and  $\Sigma_{g_i} \bullet f$ , respectively (i.e.,  $f \bullet \Sigma_{g_i} = \Sigma_{fg_i}$  and  $\Sigma_{g_i} \bullet f = \Sigma_{g_i f}$ ); furthermore,  $\text{inv} \bullet \Sigma_{g_i} = \Sigma_{g_i^{-1}}$  denotes the centralized covariance of the inverted data points. We also extend these definitions linearly to weighted sums of covariances. Then, we have the following lemma.

**Lemma 3.3.** *Given data  $(g_1, \dots, g_m)$  in  $U$ , its centralized covariance is invariant under left translations, while under right translations with  $f \in G$  it transforms according to*

$$[\Sigma_{g_i} \bullet f] = [\text{Ad}(f^{-1})] [\Sigma_{g_i}] [\text{Ad}(f^{-1})]^T.$$

Furthermore, if the data points are inverted, then

$$[\text{inv} \bullet \Sigma_{g_i}] = [\text{Ad}(\bar{g})] [\Sigma_{g_i}] [\text{Ad}(\bar{g})]^T.$$

*Proof.* Using the equivariance of the mean gives

$$[f \bullet \Sigma_{g_i}] = \frac{1}{m} \sum_{l=1}^m [\log(\bar{g}^{-1} f^{-1} f g_l)] [\log(\bar{g}^{-1} f^{-1} f g_l)]^T = [\Sigma_{g_i}],$$

and applying (2.6) yields

$$\begin{aligned} [\Sigma_{g_i} \bullet f] &= \frac{1}{m} \sum_{l=1}^m [\log(f^{-1} \bar{g}^{-1} g_l f)] [\log(f^{-1} \bar{g}^{-1} g_l f)]^T \\ &= \frac{1}{m} \sum_{l=1}^m [\text{Ad}(f^{-1})] [\log(\bar{g}^{-1} g_l)] [\log(\bar{g}^{-1} g_l)]^T [\text{Ad}(f^{-1})]^T \\ &= [\text{Ad}(f^{-1})] [\Sigma_{g_i}] [\text{Ad}(f^{-1})]^T. \end{aligned}$$

After inverting the data points we find, using [Theorem 2.2](#), [\(2.4\)](#), and [\(2.6\)](#), that

$$\begin{aligned} \left[ \text{inv} \bullet \Sigma_{g_i} \right] &= \frac{1}{m} \sum_{l=1}^m \left[ \log(\bar{g}g_l^{-1}) \right] \left[ \log(\bar{g}g_l^{-1}) \right]^T \\ &= \frac{1}{m} \sum_{l=1}^m (-1)^2 \left[ \log(g_l\bar{g}^{-1}) \right] \left[ \log(g_l\bar{g}^{-1}) \right]^T \\ &= \left[ \text{Ad}(\bar{g}) \right] \left[ \Sigma_{g_i} \right] \left[ \text{Ad}(\bar{g}) \right]^T. \end{aligned}$$

■

We can translate this directly to the pooled covariance:

**Corollary 3.4.** *The pooled covariance is invariant under left translations of the data and transforms under right translations with  $f \in G$  according to*

$$\left[ \hat{\Sigma} \bullet f \right] = \left[ \text{Ad}(f^{-1}) \right] \left[ \hat{\Sigma} \right] \left[ \text{Ad}(f^{-1}) \right]^T.$$

If the data are inverted, then we have

$$\left[ \text{inv} \bullet \hat{\Sigma} \right] = \frac{1}{m+n-2} \left( m \left[ \text{Ad}(\bar{g}) \right] \left[ \Sigma_{g_i} \right] \left[ \text{Ad}(\bar{g}) \right]^T + n \left[ \text{Ad}(\bar{h}) \right] \left[ \Sigma_{h_i} \right] \left[ \text{Ad}(\bar{h}) \right]^T \right).$$

In particular, if  $\bar{g} = \bar{h}$ , then

$$\left[ \text{inv} \bullet \hat{\Sigma} \right] = \left[ \text{Ad}(\bar{g}) \right] \left[ \hat{\Sigma} \right] \left[ \text{Ad}(\bar{g}) \right]^T.$$

With this, we propose the following generalization of the Hotelling  $T^2$  statistic for Lie groups, using the abbreviated notation  $((g_i), (h_i))$  for  $((g_1, \dots, g_m), (h_1, \dots, h_n))$ .

**Definition 3.5 (bi-invariant Hotelling  $T^2$  statistic).** Let  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  be data sets in  $U$  with group means  $\bar{g}$  and  $\bar{h}$  such that  $\log(\bar{g}^{-1}\bar{h})$  exists, and let  $\Sigma$  be their pooled sample covariance. Then, the bi-invariant Hotelling  $T^2$  statistic is defined by

$$t^2((g_i), (h_i)) = \frac{mn}{m+n} \mu_{(e, \hat{\Sigma})}^2(\bar{g}^{-1}\bar{h}).$$

This definition is independent of the chosen basis of  $T_e G$  [53, p. 39]. Furthermore, it reduces to [\(2.2\)](#) when  $G = \mathbb{R}^n$ . This is a consequence of the existence of a common reference tangent space; as the latter is missing in Riemannian manifolds, reduction to the multivariate case is problematic using the Riemannian approach. Indeed, the only generalization of Hotelling's  $T^2$  statistic to Riemannian manifolds known to the authors does not have this property; see [Appendix B](#) for more on this.

The following theorem shows that the newly defined notion has the invariance properties we aimed for and does not depend on the order of the data sets.

**Theorem 3.6 (properties of the bi-invariant Hotelling  $T^2$  statistic).** Let  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  be data sets in  $U$  with group means  $\bar{g}$  and  $\bar{h}$ , respectively. The bi-invariant Hotelling  $T^2$  statistic has the following properties:

- (i) it is symmetric;
- (ii) it is invariant under left and right translations;
- (iii) if  $\bar{g} = \bar{h}$ , then it is invariant under inversion.

*Proof.* Note first that the pooled covariance is independent of the order of the data groups. Thus, using (2.4), property (i) follows from

$$t^2((g_i), (h_i)) = \frac{mn}{m+n}(-1)^2 \left[ \log \left( \bar{h}^{-1} \bar{g} \right) \right]^T \left[ \widehat{\Sigma} \right]^{-1} \left[ \log \left( \bar{h}^{-1} \bar{g} \right) \right] = t^2((h_i), (g_i)).$$

Next, we show (ii). Invariance under left translations follows directly since  $f$  and  $f^{-1}$  cancel out as in the proof of Lemma 3.3. Let  $f \in G$ . Then, using Theorem 2.2, (2.6), and Corollary 3.4, we find

$$\begin{aligned} t^2((g_i f), (h_i f)) &= \frac{mn}{m+n} \left[ \log \left( f^{-1} \bar{g}^{-1} \bar{h} f \right) \right]^T \left[ \widehat{\Sigma} \bullet f \right]^{-1} \left[ \log \left( f^{-1} \bar{g}^{-1} \bar{h} f \right) \right] \\ &= \frac{mn}{m+n} \left[ \log \left( \bar{g}^{-1} \bar{h} \right) \right]^T \left[ \text{Ad}(f^{-1}) \right]^T \left[ \text{Ad}(f^{-1}) \right]^{-T} \left[ \widehat{\Sigma} \right]^{-1} \\ &\quad \left[ \text{Ad}(f^{-1}) \right]^{-1} \left[ \text{Ad}(f^{-1}) \right] \left[ \log \left( \bar{g}^{-1} \bar{h} \right) \right] \\ &= t^2((g_i), (h_i)). \end{aligned}$$

Finally, (iii) follows immediately from  $\log(e) = 0$ . ■

Note that although the statistic is not invariant under inversion in general, the fact that equality of means is invariantly detected makes it also interesting (e.g., to perform hypothesis tests for equality of means) when invariance under inversion is of interest. (Nevertheless, the authors are not aware of any such application.)

**Remark 3.7.** Observe that we could replace left- by right-centralized sample covariances in all definitions of this section. The Hotelling  $T^2$  statistic will then be different in general, but it will have the same invariance properties as Definition 3.5.

**3.2. Extending the Bhattacharyya distance.** Since the first summand of the multivariate Bhattacharyya distance (2.3) coincides with a Hotelling  $T^2$  statistic, we can generalize it using the bi-invariant Hotelling  $T^2$  statistic. To this end, we begin with the definition of the averaged sample covariance.

**Definition 3.8 (averaged sample covariance).** Given data sets  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  in  $U$  with group means  $\bar{g}$  and  $\bar{h}$ , their left-averaged sample covariance is defined by

$$\overline{\Sigma}_{g_i, h_i} = \frac{1}{2}(\Sigma_{g_i} + \Sigma_{h_i}).$$

Again, in the following we drop the word left and the subscripts. Since it only differs in the weighting, the averaged covariance has the same properties as the pooled covariance.

**Corollary 3.9.** The averaged covariance is invariant under left translations of the data and transforms under right translations with  $f \in G$  according to

$$\left[ \overline{\Sigma} \bullet f \right] = \left[ \text{Ad}(f^{-1}) \right] \left[ \overline{\Sigma} \right] \left[ \text{Ad}(f^{-1}) \right]^T.$$

If the data are inverted, then we have

$$[\text{inv} \bullet \bar{\Sigma}] = \frac{1}{2} \left( [\text{Ad}(\bar{g})] [\Sigma_{g_i}] [\text{Ad}(\bar{g})] + [\text{Ad}(\bar{h})] [\Sigma_{h_i}] [\text{Ad}(\bar{h})] \right).$$

In particular, if  $\bar{g} = \bar{h}$ , then

$$[\text{inv} \bullet \bar{\Sigma}] = [\text{Ad}(\bar{g})] [\bar{\Sigma}] [\text{Ad}(\bar{g})]^T.$$

With this, we can generalize the Bhattacharyya distance to Lie groups.

**Definition 3.10 (bi-invariant Bhattacharyya distance).** Let  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  be data sets in  $U$  with group means  $\bar{g}$  and  $\bar{h}$ . Then, the bi-invariant Bhattacharyya distance is defined by

$$D_B((g_i), (h_i)) = \frac{1}{8} \mu_{(e, \bar{\Sigma})}^2 (\bar{g}^{-1} \bar{h}) + \frac{1}{2} \ln \left( \frac{\det([\bar{\Sigma}])}{\sqrt{\det([\Sigma_{g_i}]) \det([\Sigma_{h_i}])}} \right).$$

The following theorem shows that the attribute “bi-invariant” is indeed justified.

**Theorem 3.11 (properties of the bi-invariant Bhattacharyya distance).** Let  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  be data sets in  $U$  with group means  $\bar{g}, \bar{h}$ , respectively. The bi-invariant Bhattacharyya distance has the following properties:

- (i) it is symmetric;
- (ii) it does not depend on the chosen basis;
- (iii) it is invariant under left and right translations;
- (iv) if  $\bar{g} = \bar{h}$ , then it is invariant under inversion.

*Proof.* Note that we can focus on the second summand, since for the first all results can be shown as in the proof of [Theorem 3.6](#) (replacing the pooled by the averaged covariance). Thus, (i) follows immediately, since  $\bar{\Sigma}$  is invariant under an exchange of the data sets.

Next we show (iii). Let  $f \in G$  and  $\rho_f = \det(\text{Ad}(f^{-1}))$ ; note that  $\rho_f \neq 0$ . Invariance of the second summand under left translations follows from [Lemma 3.3](#) and [Corollary 3.9](#). They further imply that

$$\frac{\det([\bar{\Sigma} \bullet f])}{\sqrt{\det([\Sigma_{g_i} \bullet f]) \det([\Sigma_{h_i} \bullet f])}} = \frac{\rho_f^2 \det([\bar{\Sigma}])}{\rho_f^2 \sqrt{\det([\Sigma_{g_i}]) \det([\Sigma_{h_i}])}},$$

which yields invariance under right translations. The proofs of (ii) and (iv) work analogously (using [Corollary 3.9](#) for (iv) and replacing group adjoints with basis transformations for (ii)). ■

Note that in [Definition 3.10](#) we take determinants of matrix representations of (2,0)-tensors instead of (0,2)-tensors like in the Euclidean case (for background information see, e.g., [23, Chap. 4]). This does not make a difference since lowering indices with any (auxiliary) metric does not have an influence on the ratio of the determinants in [Definition 3.10](#); the additional terms (i.e., the matrix representation of the metric) cancel out analogously to those introduced

by translations (i.e., matrix representations of group adjoints) in the proofs of Theorems 3.6 and 3.11.

*Remark 3.12.* A notion from multivariate statistics that is intimately related to the Bhattacharyya distance is the *Hellinger distance*  $H$  [36] (see also [50, p. 51]):

$$(3.1) \quad H((p_i), (q_i)) = \sqrt{1 - 1/\exp(D_B((p_i), (q_i)))}$$

for data sets  $(p_1, \dots, p_m)$  and  $(q_1, \dots, q_n)$  in  $\mathbb{R}^d$ . It has found various applications: for example, in decision trees [21], internet telephony [63], and data visualization [60].

By using the bi-invariant Bhattacharyya distance in (3.1), we directly obtain a *bi-invariant Hellinger distance* on Lie groups with the same invariance properties.

**3.3. Connection of the bi-invariant Bhattacharyya distance to densities.** In the multivariate setting, the Bhattacharyya distance has a more general integral definition for distributions with a probability density function (pdf). We show in this section that the bi-invariant Bhattacharyya distance is compatible with this view. Like in Euclidean space, we obtain it for Gaussian-like distributions from the generalized integral definition. On the other hand, in contrast to our definition, the integral version has the drawback that it is not bi-invariant in general.

Let  $P$  and  $Q$  be two random variables on  $\mathbb{R}^d$  with pdfs  $p$  and  $q$ , respectively. The Bhattacharyya distance then takes the form

$$D_B(p, q) = -\ln \left( \int_{\mathbb{R}^d} \sqrt{pq} \, dx \right),$$

which reduces to (2.3) for two normal distributions. In this section we investigate classes of Lie groups and distributions for which the above (or more precisely an analog) integral reduces to the bi-invariant Bhattacharyya distance.

On a Lie group  $G$  of dimension  $d$ , integrals of functions are defined via differential  $d$ -forms;<sup>3</sup> see, for example, [58, Chap. 26]. For our purpose, it suffices to recount how one integrates over the domain of a single chart. (In general, integrals are defined via so-called partitions of unity.)

Let  $\Phi : V \rightarrow U$  be a diffeomorphism between  $V \subseteq \mathbb{R}^d$  and a neighborhood  $U \subseteq G$  (i.e., an inverse chart) and  $\omega$  a  $d$ -form on  $U$ . Then, there is a unique  $\varphi \in \mathcal{C}^\infty(\Phi(V))$  such that the pullback<sup>4</sup> of  $\omega$  along the diffeomorphism  $\Phi$  is given by  $\Phi^*(\omega) = \varphi \, dx$  ( $dx$  being the ordinary volume form on  $\mathbb{R}^d$ ). The integral of  $\omega$  over  $U$  is then defined by

$$\int_U \omega = \int_V \Phi^*(\omega) = \int_V \varphi \, dx$$

if the right-hand side exists.

---

<sup>3</sup>Remember that [35, Sec. 1.3] on a smooth manifold  $M$ , a differential  $d$ -form  $\omega$  smoothly assigns to each  $p \in M$  an alternating  $d$ -linear map  $\omega_p : (T_p M)^d \rightarrow \mathbb{R}$ . If  $\psi \in \mathcal{C}^\infty(M)$ , then  $\psi \omega$  is also a  $d$ -form on  $M$ .

<sup>4</sup>Let  $F : M \rightarrow N$  be a diffeomorphism of two smooth  $d$ -dimensional manifolds  $M, N$  and  $\omega$  a  $d$ -form on  $N$ . The pullback  $F^*(\omega)$  of  $\omega$  along  $F$  is a  $d$ -form on  $M$  defined by  $F^*(\omega)(X_1, \dots, X_d) = \omega(dF(X_1), \dots, dF(X_d))$  for  $X_1, \dots, X_d \in \Gamma(TM)$ . Note that  $F^*(\psi \omega) = (\psi \circ F) F^*(\omega)$  for all  $\psi \in \mathcal{C}^\infty(N)$ .

In the affine setting there is no canonical volume form on  $G$ . Instead, we must choose a reference measure, that is, a nonzero left invariant  $d$ -form  $dg$  on  $G$  (a *left Haar measure*) consistent with the orientation of  $G$  (which we choose such that  $\exp$  is orientation preserving). Luckily, a different choice of reference measure only leads to multiplying each integral with the same constant, which is why the actual choice of  $dg$  depends (up to considerations of machine precision when working on a computer) only on one's taste. Thus, from now on we assume that  $dg$  is arbitrary but fixed. Then, the integral of a function  $\psi \in \mathcal{C}^\infty(U)$  is defined by

$$(3.2) \quad \int_U \psi \, dg.$$

This, in turn, allows one to define pdfs on  $G$  as (not necessarily smooth) functions  $\mu \geq 0$  on  $G$  such that  $\int_G \mu \, dg = 1$  (see, e.g., [18]). Consequently, we get the integral version of the Bhattacharyya distance on Lie groups:

$$(3.3) \quad D_B^{\text{int}}(\mu, \nu) = -\ln \left( \int_G \sqrt{\mu \nu} \, dg \right)$$

for pdfs  $\mu, \nu$  on  $G$ .

Note that, in general, (3.2) is left invariant (i.e.,  $\int_G \psi \circ L_h \, dg = \int_G \psi \, dg$  for each  $h \in G$  and  $\psi \in \mathcal{C}^\infty(G)$  whose integral exists) but not right invariant.<sup>5</sup> Thus, in contrast to Definition 3.10, (3.3) cannot generally be bi-invariant. For this to hold, we need a *unimodular* Lie group, that is, a Lie group with bi-invariant volume form. (A well-known example of a nonunimodular Lie group is the group of invertible affine transformations of  $\mathbb{R}^n$ .) Therefore, we will only be interested in the unimodular case. Since then distributions can be “moved around via translations” without changing integrals, we can consider without loss of generality those centered at  $e$ .

In the following, let  $E : \mathbb{R}^d \rightarrow T_e G$  be any orientation-preserving vector space isomorphism (which we obtain from choosing a basis of  $T_e G$ ). Further, let  $\epsilon \subseteq T_e G$  be a neighborhood of  $0 \in T_e G$  such that *first*  $\exp|_\epsilon$  is a diffeomorphism onto the neighborhood  $W = \exp(\epsilon)$  of  $e$ , and *second* there is an orientation-preserving diffeomorphism  $\Phi : T_e G \rightarrow \epsilon$  between the whole of  $T_e G$  and  $\epsilon$ . Lemma A.5 shows that we can always choose  $\epsilon := \log(W)$ , where  $W$  is any CSLCG neighborhood of  $e$ . Now, we can define  $\widetilde{\exp} = \exp \circ \Phi \circ E$  and  $\log = E^{-1} \circ \Phi^{-1} \circ \log$ . Note that the latter is a chart of  $W$  with image  $\mathbb{R}^d$ . In order to identify densities on  $G$  that connect both versions of the Bhattacharyya distance we first need the lemma below.

**Lemma 3.13.** *Let  $G$  be a  $d$ -dimensional Lie group and  $g \in G$ . Further, let  $p, q$  be pdfs on  $\mathbb{R}^d$  and  $\widetilde{\exp}^*(dg|_W) = \phi \, dx$  with  $0 < \phi \in \mathcal{C}^\infty(\mathbb{R}^d)$ . Define*

$$(3.4) \quad \mu(g) = \begin{cases} (p/\phi \circ \widetilde{\log})(g), & g \in W, \\ 0, & g \in G \setminus W, \end{cases}$$

and

$$(3.5) \quad \nu(g) = \begin{cases} (q/\phi \circ \widetilde{\log})(g), & g \in W, \\ 0, & g \in G \setminus W. \end{cases}$$

---

<sup>5</sup>We have a similar problem when we start with a right-invariant volume form.

Then,  $\mu$  and  $\nu$  are pdfs on  $G$  with

$$\int_G \sqrt{\mu\nu} dg = \int_{\mathbb{R}^d} \sqrt{pq} dx.$$

*Proof.* Clearly,  $\mu, \nu \geq 0$ . Furthermore, since  $\mu$  and  $\nu$  are only supported in  $W$  and  $\widetilde{\log} \circ \widetilde{\exp} = \text{Id}_{\mathbb{R}^d}$ , we find  $\int_G \mu dg = \int_G \nu dg = 1$  and

$$\int_G \sqrt{\mu\nu} dg = \int_W \sqrt{\mu\nu} dg = \int_{\mathbb{R}^d} \widetilde{\exp}^*(\sqrt{\mu\nu} dg) = \int_{\mathbb{R}^d} (\sqrt{\mu\nu} \circ \widetilde{\exp})\phi dx = \int_{\mathbb{R}^d} \sqrt{pq} dx. \quad \blacksquare$$

We can also write the function  $\phi$  more explicitly. Denoting  $dy_i = \widetilde{\log}^*(dx_i)$ ,  $i = 1, \dots, d$ , there is  $0 < \psi \in \mathcal{C}^\infty(G)$  such that  $dg|_W = \psi dy_1 \wedge \dots \wedge dy_d$ <sup>6</sup>. Hence, the transformation rule for forms of maximal degree (i.e., the dimension of the underlying manifold) implies

$$(3.6) \quad \widetilde{\exp}^*(dg|_W) = \det(d\widetilde{\exp})(\psi \circ \widetilde{\exp}) dx_1 \wedge \dots \wedge dx_d$$

and, thus,  $\phi = \det(d\widetilde{\exp})(\psi \circ \widetilde{\exp})$ .

In the following,  $\mathcal{N}(x, [\Sigma])$  denotes the multivariate normal distribution on  $\mathbb{R}^d$  with mean  $x$  and covariance matrix  $[\Sigma]$ . The following proposition establishes a connection between the bi-invariant Bhattacharyya distance and the integral version.

**Proposition 3.14.** *Let  $G$  be a  $d$ -dimensional unimodular Lie group with CCS connection, and let  $p$  be the pdf of  $\mathcal{N}(0, [\Sigma_{g_i}])$  and  $q$  the pdf of  $\mathcal{N}([\log(\bar{g}^{-1}\bar{h})], [\Sigma_{h_i}])$  on  $\mathbb{R}^d$ . From  $p$  and  $q$ , define pdfs  $\mu$  and  $\nu$  on  $G$  by (3.4) and (3.5). Further, let  $(g_1, \dots, g_m)$  and  $(h_1, \dots, h_n)$  be data sets in a CSLCG neighborhood  $U \subseteq \exp(\mathfrak{e}) \subseteq G$  with group means  $\bar{g}$  and  $\bar{h}$ , respectively. Then,*

$$D_B^{\text{int}}(\mu, \nu) = D_B((g_i), (h_i)).$$

*Proof.* The assertion follows from Lemma 3.13 and the fact that the Bhattacharyya distance between two multivariate normal distributions is of the form (2.3).  $\blacksquare$

According to the proposition, “scaled push-forwards” of the multivariate normal distribution provide a connection between both versions of the Bhattacharyya distance. Densities connected to the these were investigated for  $\text{SE}(n)$  in [17]—also with an emphasis on bi-invariance. Note that the requirement  $U \subseteq \exp(\mathfrak{e})$  is not needed in the proof but is introduced to ensure that the data samples have nonvanishing probabilities under  $p$  and  $q$ . If one wants to choose  $p$  and  $q$ , for example, to build a model, setting  $\mathfrak{e} = \log(U)$  seems sensible because thereby the locality condition that we impose on the data for  $D_B$  is met. On the other hand, larger domains could be used if “outliers” should be possible. An interesting special case are Lie groups whose exponentials are *global* diffeomorphisms. Examples are simply connected, nilpotent Lie groups (like the Heisenberg group, which allows for a unique mean in the entire group [55, pp. 198–199]); they are unimodular and the pullback of the Lebesgue measure under the logarithm is a bi-invariant Haar measure [20]. Thus, in a simply-connected, nilpotent Lie group,  $dg$  can be chosen such that  $\psi \equiv 1$  in (3.6).

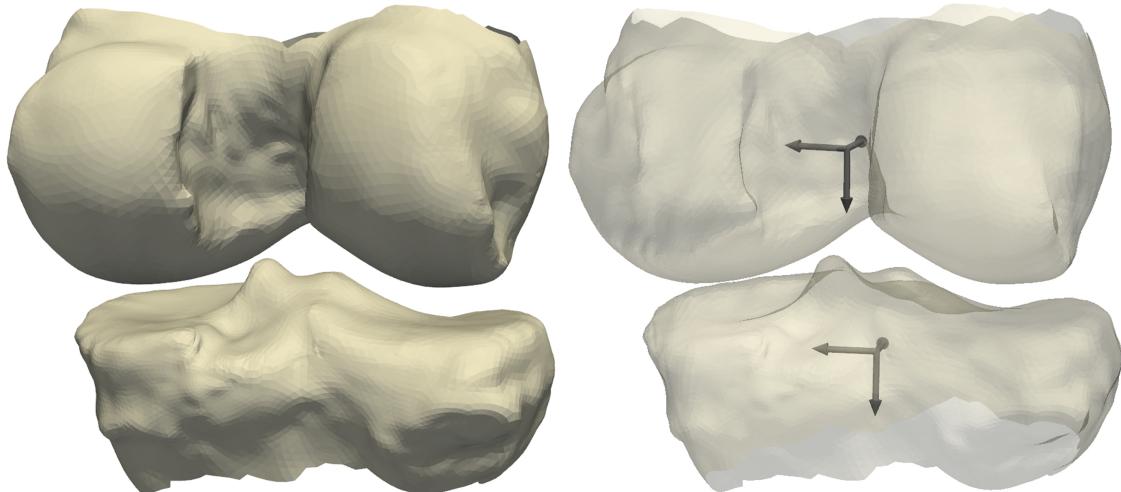
---

<sup>6</sup>Here,  $\wedge$  denotes the “wedge product” of differential forms. Any form of maximal degree  $d$  can be written as the wedge product of  $d$  coordinate 1-forms multiplied by a smooth function; see, e.g., [35, Sec. 1.3].

**4. Experiments.** In this section, we use the proposed indices for hypothesis tests in two application fields. Although we have chosen two specific contexts, the testing procedures described below can be used in many other application fields.

**4.1. Configurations of the knee joint under osteoarthritis.** In the first part, we investigate configurations of the human knee joint under osteoarthritis (OA). The relative position of the femur with respect to the tibia is described by a rigid-body transformation. Using both the Hotelling  $T^2$  statistic and Bhattacharyya distance in a two-sample test, we infer a significant difference between knee configurations of people with severe OA and healthy controls; we thus find the well-known joint space narrowing that is indicative for OA.

**4.1.1. Data description.** The first data set, which was also used in [75], is derived from the OA Initiative<sup>7</sup> (OAI). The OAI is a longitudinal study of knee OA that provides (among others) publicly accessible clinical evaluation data and radiological images from 4796 men and women of ages 45–79. By means of a special support plate it was ensured that the patients' legs were constrained to the same posture. From the baseline data set (i.e., the images from the initial visits), we chose 58 severely diseased subjects and 58 healthy subjects according to their Kellgren–Lawrence score [42]. The sets were balanced in order to maximize the statistical power of our hypothesis test. For the 116 subjects, surfaces in correspondence of the distal femora and proximal tibiae were extracted from the respective 3-dimensional DESS-WE (double-echo steady-state sequence with water excitation) MR images ( $0.37 \times 0.37$  mm matrix, 0.7 mm slice thickness). The correspondence was then used to consistently extract (i.e., cut out) the condylar regions of the femora and tibiae; results from one subject can be seen on the left of Figure 1. A supervised post-processing procedure was used to ensure the



**Figure 1.** Visualization of the knee bones and the computed orthonormal frames. On the left, the distal femur (top) and the proximal tibia (below) are shown; the meshes (and their relative position) were reconstructed from OA data. The space between both (i.e., the joint space) is filled with cartilage, ligaments, and menisci (all not shown). On the right, we depict the orthonormal frames that correspond to both meshes.

<sup>7</sup><http://www.oai.ucsf.edu/L>

quality of the segmentations and the correspondence of the resulting meshes (8988 vertices for each femur and 8320 for each tibia). During the process, the relative positions of femur and tibia, which are present in the raw data through the global coordinate system of the scanner, are maintained.

**4.1.2. Encoding relative positions in SE(3).** Remember that, with  $O(n)$  denoting the group of orthogonal matrices in dimension  $n$ , the Euclidean group  $E(3) = O(3) \ltimes \mathbb{R}^3$  is a semidirect product with group operation  $(R, v)(Q, w) = (RQ, v+Rw)$  for  $(R, v), (Q, w) \in E(3)$ . The group of rigid-body motions is the subgroup  $SE(3) = SO(3) \ltimes \mathbb{R}^3$ .

Since for each femur-tibia pair the relative position is encoded in the coordinates of the meshes, we can capture it by calculating the rigid body transformation in  $SE(3)$  that moves the distal femur onto the proximal tibia. More precisely, let  $(O_f, o_f), (O_t, o_t) \in E(3)$  be local orthonormal reference frames for the femur and tibia, respectively. (The vector part yields the origin, while the matrix columns represent the coordinate axes.) Then, after translating the meshes so that  $o_f$  is moved to the origin (i.e., applying  $(0, -o_f) \in SE(3)$  to both frames), the rotation-translation pair  $P = (R, v) \in SE(3)$  that maps the reference frame  $(O_f, o_f)$  of the femur to the reference frame  $(O_t, o_t)$  of the tibia encodes the relative position of the bones. Now, since

$$(O_t O_f^T, o_t - o_f)(O_f, 0) = (O_t, o_t - o_f),$$

we have

$$P = (R, v) = (O_t O_f^T, o_t - o_f).$$

Now, in order to compute this we need to choose frames. We determine each reference frame from the principal component analysis of the vertices of the corresponding triangle mesh: While we pick the center of gravity as origin, the axes are chosen as unit vectors pointing along the principal directions; the latter is done in a consistent way throughout the population for both bone types; the resulting frames from one subject are visualized on the right of Figure 1.

Note that, in this context, left and right translations of  $P$  with elements from  $SE(3)$  correspond to different choices of reference frames (with the same orientations) for the femur and tibia, respectively.

Finally, we want to point out that this approach of representing relative positions by elements of the special Euclidean group is also used in many other applications, for example, medical image analysis [13], robotics [51], human action recognition [73], radar detection [10, 16], and state description of molecules [12]. Thus, the permutation test outlined below can directly be applied to data from these domains.

**4.1.3. Bi-invariant permutation test.** We use a nonparametric permutation setup to test whether there is a difference between the knee configurations of people with OA and healthy controls. The reason is that we do not want to make assumptions about how both groups are distributed, but rather learn this from the data. After the 116 femur-tibia pairs have been processed as described in subsection 4.1.2, we obtain a set  $(P_1^{(H)}, \dots, P_{58}^{(H)})$  in  $SE(3)$  of transformations derived from healthy controls and a set  $(P_1^{(OA)}, \dots, P_{58}^{(OA)})$  coming from

patients with OA. We consider the null hypothesis  $H_0$  of equal distributions underlying both groups, that is,  $P^{(H)} \stackrel{H_0}{\sim} P^{(OA)}$ . As test statistic  $T$  we use (and thus compare) both the bi-invariant Hotelling  $T^2$  statistic  $t^2$  and the bi-invariant Bhattacharyya distance  $D_B$ .

We initialize the test by computing the baseline

$$T_0 = T((P_i^{(H)}), (P_i^{(OA)})).$$

Then, we perform 10000 random permutations of the joint set

$$(Z_1, \dots, Z_{116}) = \left( P_1^{(H)}, \dots, P_{58}^{(H)}, P_1^{(OA)}, \dots, P_{58}^{(OA)} \right);$$

that is, denoting the  $l$ th permutation by  $\sigma_l$ , we compute the values

$$T_l = T((Z_{\sigma_l(1)}, \dots, Z_{\sigma_l(58)}), (Z_{\sigma_l(59)}, \dots, Z_{\sigma_l(116)})), \quad l = 1, \dots, 10,000.$$

With  $\mathbf{1}_{a \geq b}$  being 1 if  $a \geq b$  and 0 else, the  $p$ -value for the statistic  $T$  is then given by

$$p_T = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{T_l \geq T_0};$$

it is the proportion of test statistics that are greater than the one computed for the original (unpermuted) groups. A standard level to reject the null hypothesis is  $p_T < 0.05$ ; in this case, the difference between the distributions is called *significant*. Performing the test with each statistic, we obtain

$$p_{t^2} \approx 0.00019 \quad \text{and} \quad p_{D_B} < 10^{-5}.$$

Thus, the differences are significant for both statistics. This result is in accordance with what is known about OA: Part of the disease is the so-called joint space narrowing—a decrease in distance between femur and tibia as a consequence of degenerating cartilage. Note that we can deduce a posteriori that the sample size was sufficient to detect significant differences with an error probability smaller than 0.05.

**4.2. Hippocampi shapes under Alzheimer's.** In the second part, we analyze hippocampal atrophy patterns due to mild cognitive impairment (MCI) by applying and thereby comparing the derived Hotelling  $T^2$  statistic and Bhattacharyya distance. MCI is a common condition in the elderly and often represents an intermediate stage between normal cognition and Alzheimer's disease. As is consistently reported in neuroimaging studies, atrophy of the hippocampal formation is a characteristic early sign of MCI. Using both a local and global two-sample test, we infer significant differences in distribution of shapes of right hippocampi between a cognitive normal and the MCI group, in agreement with the literature.

**4.2.1. Data description.** For our experiments we prepared a data set consisting of 26 subjects showing MCI and 26 cognitive normal (CN) controls from the open access Alzheimer's Disease Neuroimaging Initiative<sup>8</sup> (ADNI) database. ADNI provides, among others, 1632

---

<sup>8</sup><http://adni.loni.usc.edu/>

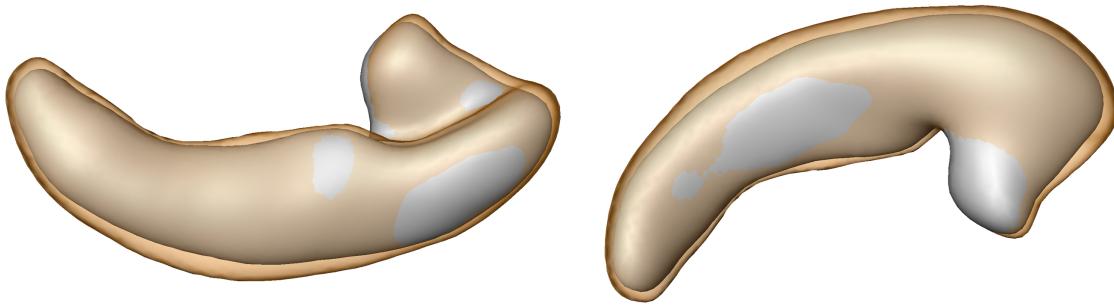
brain magnetic resonance images collected on four different time points with segmented hippocampi. We established surface correspondence (2280 vertices, 4556 triangles) in a fully automatic manner employing the deblurring and denoising of functional maps approach [26] for isosurfaces extracted from the available segmentations. The data set was randomly assembled from shapes for which segmentations were simply connected and remeshed surfaces were well-approximating ( $\leq 10^{-5}$  mm root mean square surface distance to the isosurface).

**4.2.2. Lie-group-based shape space.** To model the shapes of the hippocampi, we use the representation from [3]. Given  $n$  homogeneous objects in the form of triangle meshes  $T_i \subset \mathbb{R}^3$  that are in correspondence and Procrustes aligned [30], their shapes are described in terms of (a power of) the Lie group  $\text{GL}^+(3)$  of real 3-by-3 matrices with positive determinant. To obtain this representation, we view each mesh  $T_i$  as the result of a deformation  $\phi_i$  of a common reference  $\bar{T} \subset \mathbb{R}^3$ ; that is, each  $\phi_i : \bar{T} \rightarrow T_i$  is an orientation-preserving, simplicial isomorphism that yields a semantic correspondence. The underlying idea is that the derivatives of these deformations comprise the shape information, because they provide a local characterization of the deformation without being influenced by the location in space. Thus, the Jacobian  $D\phi_i$  constitutes the shape representation of the  $i$ th object. Let  $m$  be the (necessarily same) number of triangles of each mesh. Since each  $D\phi_i$  is constant on the faces  $F_j$  of  $\bar{T}$ , there are 3-by-3 matrices  $G_j^{(i)}$  with  $\det(G_j^{(i)}) > 0$  such that  $D\phi_i|_{F_j} = G_j^{(i)} \in \text{GL}^+(3)$  for all  $i = 1, \dots, n$ , and  $j = 1, \dots, m$ . The  $i$ th shape is thus given by  $(G_1^{(i)}, \dots, G_m^{(i)}) \in \text{GL}^+(3)^m$ . Note that the above  $\text{GL}^+(3)$  model stems from a continuous formulation [75, Sec. 2] that admits consistent and convergent discretizations.

The task of obtaining a surface immersion whose Jacobian is closest to given differential coordinates leads to a variational problem. Its minimizer is given by the solution of the well-known Poisson equation for which fast numerical methods exist. Furthermore, as a global variational approach, the minimizer given by the Poisson equation tends to distribute errors uniformly such that local gradient field inconsistencies are attenuated.

**4.2.3. Hippocampal atrophy patterns in CN versus MCI.** To test for a difference in distribution between the CN and MCI groups we extend the permutation test setup from subsection 4.1.3 and perform a local (i.e., trianglewise) and a global test. After encoding the shapes as described in subsection 4.2.2, the group means for the CN and MCI set are computed with the iterative algorithm from [53, Alg. 1]. A well-known [46] phenomenon in patients that develop MCI is a loss of total hippocampal volume; it can be clearly observed in the qualitative comparison of both mean shapes shown in Figure 2.

**4.2.4. Local test.** In the following, when we speak of the  $i$ th differential coordinate we mean the  $i$ th element of the differential coordinates (vector). For  $i = 1, \dots, 4556$  let  $G_i^{(\text{CN})}$  and  $G_i^{(\text{MCI})}$  denote the  $i$ th differential coordinate of the mean hippocampi from the CN and MCI group, respectively. (We choose the first mesh of the CN group as reference object.) In order to identify subregions that contribute to differences in mean shape between the groups, we perform trianglewise, partial tests. For every triangle we *independently* consider the null hypothesis  $H_0$  that the distribution of its differential coordinate is the same for both groups, that is,  $G_i^{(\text{CN})} \stackrel{H_0}{\sim} G_i^{(\text{MCI})}$ .



**Figure 2.** Group means of right hippocampi for CN (orange, transparent) and impaired (grey) subjects overlaid.

To test this hypothesis, we perform a permutation and compare again the bi-invariant Hotelling  $T^2$  statistic and Bhattacharyya distance as test statistics. First, we perform independent tests for each triangle using the testing procedure from subsection 4.1.3. Therefore, denoting the test statistic by  $T$  again, we start by computing

$$T_0^i = T((G_i^{(\text{CN})}), (G_i^{(\text{MCI})})).$$

We then perform 10000 (random) permutations of the full set

$$(Z_{i,1}, \dots, Z_{i,52}) = \left( G_{i,1}^{(\text{CN})}, \dots, G_{i,26}^{(\text{CN})}, G_{i,1}^{(\text{MCI})}, \dots, G_{i,26}^{(\text{MCI})} \right);$$

here the second subscript enumerates the subjects of the groups. Then, we compute the statistics

$$T_l^i = T((Z_{i,\sigma_l(1)}, \dots, Z_{i,\sigma_l(26)}), (Z_{i,\sigma_l(27)}, \dots, Z_{i,\sigma_l(52)})), \quad l = 1, \dots, 10000.$$

The  $p$ -value of the  $i$ th triangle for test statistic  $T$  is then given by

$$p_T^{(i)} = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{T_l^i \geq T_0^i}.$$

For each triangle, we reject  $H_0$ , if  $p_T^{(i)} < 0.05$  and call the difference between both groups significant. Using the bi-invariant Hotelling  $T^2$  statistic and Bhattacharyya distance as test statistics, the described test is bi-invariant (in this case in  $\text{GL}^+(3)$ ). In particular, because of right-invariance, the results are independent from the reference that was chosen to compute the differential coordinates. (Left-invariance would allow one to *jointly* transform the differential coordinates of the subjects by elements from  $\text{GL}^+(3)$ , which, e.g., could be of interest for a better numerical performance of an algorithm.)

Because of the large number of tests we apply the Benjamini–Hochberg false discovery correction at the level  $\alpha = 0.05$  to identify triangles with significant differences, for which we reject  $H_0$ . In Figure 3, we visualize the triangles with significant differences, showing the respective  $p$ -values for both statistics.



**Figure 3.** Group tests for differences in the distribution of right hippocampi for CN and impaired subjects. Results for the bi-invariant Hotelling  $T^2$  statistic are shown at the top and for the Bhattacharyya distance at the bottom. Each triangle of the CN mean is color coded according to its  $p$ -value (false discovery corrected) using the colormap 0.0 0.05.

In line with literature on MCI [46], the obtained results suggest more differentiated morphometric changes beyond homogeneous volumetric decline of the hippocampi. The Bhattacharyya distance detects more significant differences, which is expected because it also takes covariances into account. Interestingly, there are also some areas where the Hotelling  $T^2$  statistic seems to be more sensitive. This is probably the case when, for the Bhattacharyya distance, small but significant differences in the mean become insignificant due to largely coinciding covariances.

**4.2.5. Global test.** Complementary to the local test is a global test that includes the overall pattern of the covariance structures thus being sensitive to spatial dependencies across the shape. In the construction of the test we follow the approach from [62].

For each triangle we first map  $T_l^i$  to an approximate uniform distribution in  $[0, 1]$  by applying the corresponding empirical cumulative distribution function (cdf)  $C_i$ . More precisely, we compute

$$C_i(T_l^i) = \frac{1}{10000} \sum_{r=1}^{10000} \mathbf{1}_{T_r^i \leq T_l^i}, \quad i = 1, \dots, 4556, \quad l = 0, \dots, 10000.$$

Then, setting  $\tilde{C}_i(T_l^i) = 0.9998 C_i(T_l^i) - 0.00001$  and denoting the cdf of the standard normal distribution by  $\phi$ , the data are mapped to  $U_l^i = \phi^{-1}(\tilde{C}_i(T_l^i))$ . The latter follow an approximate standard normal distribution for each triangle.

Using the sample covariance matrix  $\Sigma = 1/9999 UU^T$  (where  $U = [U_1, \dots, U_{10000}] = [U_l^i]$ ), a suitable test statistic is now given by the squared Mahalanobis distance; this yields

$$(4.1) \quad M_0 = U_0^T \Sigma^{-1} U_0, \quad M_l = U_l^T \Sigma^{-1} U_l, \quad l = 1, \dots, 10000.$$

We then call the difference in global shape significant when

$$p = \frac{1}{10000} \sum_{l=1}^{10000} \mathbf{1}_{M_l \geq M_0} < 0.05.$$

In order to account for the irregularity in surface triangulations, the sample covariance operator and thus Mahalanobis distance in (4.1) can be extended in terms of an adapted inner product that weights each component by the corresponding triangle's area. In this experiment, we adhere to the standard formulation (4.1), as the studied hippocampi surfaces are meshed uniformly.

Applying this test to the values obtained from the local test yielded  $p$ -values smaller than  $1/10^5$  for the bi-invariant Hotelling  $T^2$  statistic and Bhattacharyya distance revealing that both are sensitive to the underlying morphological changes.

**5. Discussion.** In this work, we generalized established dissimilarity measures between sample distributions to Lie groups. These new indices are, in particular, invariant under translations, which yield analyses that are unbiased from the position of the data sets in the group. As an application, we constructed nonparametric two-sample tests based on the proposed measures for both local and global hypothesis tests of shapes and used them for group tests on malformations of right hippocampi due to MCI.

An interesting question, both theoretically and practically, is given a Lie group, to what extent is there smeariness of the group mean [25]? An answer would give an indication on the size a data set should have for reliable results.

In order for the Mahalanobis distances to be well-defined, the sample covariance operator needs to be invertible; this is frequently violated, most importantly when the number of observations is lower than the number of variables. A common approach in such situations is to resort to the pseudoinverse of the covariance. Doing this in (2.8), however, will not result in a bi-invariant notion of Mahalanobis distance; the reason for this lies in the fact that taking pseudoinverses of matrix products is in general more difficult than inverses. Since high-dimensional data are common today, extending the proposed expressions to such scenarios poses another interesting direction for future work.

**Appendix A. Existence and uniqueness of the group mean.** Let  $G$  be a Lie Group with affine connection. If  $U \subseteq G$  is a normal convex neighborhood and  $g_1, \dots, g_m \in U$ , then there always exists a group mean  $\bar{g} \in U$  [54, Thm. 5.3]. Nevertheless, it need not be unique. For the latter to be ensured we need a stronger notion of convexity that was proposed by Arnaudon and Li in [5]. In the following, we summarize their ideas; for more details, including examples, we refer to their article or [54] and the references therein.

First, we need the following definitions. (Remember that a real-valued function on a smooth manifold is called convex, if its restriction to any geodesic is convex.)

**Definition A.1.** *A function  $\rho : U \times U \rightarrow \mathbb{R}_{\geq 0}$  that is convex with respect to the product structure is called a separating function if it vanishes on the diagonal of  $U \times U$  and only there.*

**Definition A.2 ( $p$ -convexity).** *Let  $\rho$  and  $d$  be a smooth separating function and an auxiliary Riemannian distance function on  $U$ , respectively. We say that  $U$  has a  $p$ -convex geometry if there are constants  $c, C \in \mathbb{R}$  with  $0 < c < C$  and an even integer  $p \geq 2$  such that*

$$cd(f, g)^p \leq \rho(f, g) \leq Cd(f, g)^p$$

for all  $f, g \in U$ .

For general smooth manifolds it is known that not all normal convex neighborhoods have a  $p$ -convex geometry. On the other hand, Whiteheads theorem ensures that each point in  $G$  has a 2-convex neighborhood. Importantly, if the normal convex neighborhood  $U$  has a  $p$ -convex geometry for any  $p \in 2\mathbb{N}$ , then this is enough to ensure the uniqueness of the group mean. Nevertheless, there is a weaker condition that still yields uniqueness.

**Definition A.3 (CSLCG neighborhood [5]).** *We say that  $U$  is CSLCG if every compact subset  $K \subset U$  has a relatively compact neighborhood  $U_K$  with  $p_K$ -convex geometry for some  $p_K \in 2\mathbb{N}$  depending on  $K$ .*

Observe that if  $U$  has a  $p$ -convex geometry for some  $p \in 2\mathbb{N}$ , then it is CSLCG. We have the following result, which is a special case of [5, Prop. 2.4].

**Proposition A.4.** *Let  $G$  be a Lie group with affine connection. Further, let  $U \subseteq G$  be a CSLCG neighborhood and  $g_1, \dots, g_m \in U$ . Then, there exists a unique group mean  $\bar{g} \in U$  of  $g_1, \dots, g_m$ .*

The following lemma provides a property of CSLCG neighborhoods that is needed in this work.

**Lemma A.5.** *Let  $G$  be a  $d$ -dimensional Lie group with affine connection and  $U \subseteq G$  be a CSLCG neighborhood. Then,  $U$  is diffeomorphic to  $\mathbb{R}^d$ .*

*Proof.* Let  $g \in U$ . Since  $U$  is a normal convex neighborhood,  $\text{Log}_g(U) \subseteq T_g G$  is well-defined and star-shaped about  $0 \in T_g G$ . The claim now follows since any star-shaped domain in a  $d$ -dimensional vector space is diffeomorphic to  $\mathbb{R}^d$  [14, Thm. 5.1]. ■

**Appendix B. Hotelling's  $T^2$  statistic for Riemannian manifolds.** In [47, Sec. 3.3], Muralidharan and Fletcher introduce a generalization of Hotelling's  $T^2$  statistic to Riemannian manifolds  $M$ , i.e., for samples  $(p_1, \dots, p_m), (q_1, \dots, q_n)$  in  $M$ . The centers of the data sets are then given by the Fréchet means  $\bar{p}, \bar{q} \in M$ , respectively. Assuming that  $\bar{p}, \bar{q}$  are unique, the difference between the means can be replaced by the Riemannian logarithms  $v_{\bar{p}} = \text{Log}_{\bar{p}}(\bar{q}) \in T_{\bar{p}} M$  or  $v_{\bar{q}} = \text{Log}_{\bar{q}}(\bar{p}) \in T_{\bar{q}} M$ . Depending on the choice, the vectors are from different tangent spaces. Then, sample covariance matrices can be defined by

$$[W_{p_i}] = \frac{1}{m} \sum_{i=1}^m [\text{Log}_{\bar{p}}(p_i)][\text{Log}_{\bar{p}}(p_i)]^T, \quad [W_{q_i}] = \frac{1}{n} \sum_{i=1}^n [\text{Log}_{\bar{q}}(q_i)][\text{Log}_{\bar{q}}(q_i)]^T.$$

Since there is no canonical way to compare vectors from different tangent spaces, Muralidharan and Fletcher propose to calculate a generalized  $T^2$  statistic at both means and average the results. This leads to the generalized Hotelling  $T^2$  statistic

$$(B.1) \quad t^2((p_i), (q_i)) = \frac{1}{2} \left( [\text{Log}_{\bar{p}}(\bar{q})]^T [W_{p_i}^{-1}] [\text{Log}_{\bar{p}}(\bar{q})] + [\text{Log}_{\bar{q}}(\bar{p})]^T [W_{q_i}^{-1}] [\text{Log}_{\bar{q}}(\bar{p})] \right)$$

for Riemannian manifolds.

Note, however, that in the Euclidean case the statistics (B.1) and (B.1) do not coincide. To see this, observe first that in this case  $\text{Log}_{\bar{p}}(\bar{q}) = -\text{Log}_{\bar{q}}(\bar{p}) = \bar{q} - \bar{p}$  by the standard identification of tangent spaces. Hence, we get from (B.1) that

$$\mathbf{t}^2((p_i), (q_i)) = (\bar{p} - \bar{q})^T \frac{1}{2} (W_{p_i}^{-1} + W_{q_i}^{-1}) (\bar{p} - \bar{q}).$$

Thus, it is enough to show that  $\widehat{S}^{-1} \neq 1/2 (W_{p_i}^{-1} + W_{q_i}^{-1})$  in general for the pooled sample covariance  $\widehat{S}$ . But this is true—not only because of scaling—since for symmetric positive definite matrices  $P_1, P_2$  (of the same size) we have  $P_1^{-1} + P_2^{-1} \neq (P_1 + P_2)^{-1}$  in general (and any such matrix can be a pooled sample covariance as can be seen with the help of the singular value decomposition).

## REFERENCES

- [1] R. L. ADLER, J. DEDIEU, J. Y. MARGULIES, M. MARTENS, AND M. SHUB, *Newton's method on Riemannian manifolds and a geometric model for the human spine*, IMA J. Numer. Anal., 22 (2002), pp. 359–390, <https://doi.org/10.1093/imanum/22.3.359>.
- [2] F. AMBELLAN, M. HANIK, AND C. VON TYCOWICZ, *Morphomatics: Geometric Morphometrics in Non-Euclidean Shape Spaces*, (2021), <https://doi.org/10.12752/8544>, <https://morphomatics.github.io/>.
- [3] F. AMBELLAN, S. ZACHOW, AND C. VON TYCOWICZ, *An as-invariant-as-possible  $GL^+(3)$ -based statistical shape model*, in Proceedings 7th MICCAI workshop on Mathematical Foundations of Computational Anatomy, Springer, Cham, Switzerland, 2019, pp. 219–228, <https://doi.org/10.1007/978-3-030-33226-6>.
- [4] F. AMBELLAN, S. ZACHOW, AND C. VON TYCOWICZ, *Rigid motion invariant statistical shape modeling based on discrete fundamental forms*, Med. Image Anal., 73 (2021), pp. 102178, <https://doi.org/10.1016/j.media.2021.102178>.
- [5] M. ARNAUDON AND X.-M. LI, *Barycenters of measures transported by stochastic flows*, Ann. Probab., 33 (2005), pp. 1509–1543, <https://doi.org/10.1214/009117905000000071>.
- [6] V. ARSIGNY, P. FILLARD, X. PENNÉC, AND N. AYACHE, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magn. Reson. Med., 56 (2006), pp. 411–421, <https://doi.org/10.1002/mrm.20965>.
- [7] V. ARSIGNY, P. FILLARD, X. PENNÉC, AND N. AYACHE, *Geometric means in a novel vector space structure on symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 328–347, <https://doi.org/10.1137/050637996>.
- [8] M. BACÁK, R. BERGMANN, G. STEIDL, AND A. WEINMANN, *A second order nonsmooth variational model for restoring manifold-valued images*, SIAM J. Sci. Comput., 38 (2016), pp. A567–A597, <https://doi.org/10.1137/15M101988X>.
- [9] F. BARBARESCO, *Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry*, in 2008 IEEE Radar Conference, IEEE, Piscataway, NJ, 2008, pp. 1–6, <https://doi.org/10.1109/RADAR.2008.4720937>.
- [10] F. BARBARESCO, *Lie group statistics and Lie group machine learning based on Souriau Lie groups thermodynamics & Koszul-Souriau-Fisher metric: New entropy definition as generalized Casimir invariant function in coadjoint representation*, Entropy, 22 (2020), 642, <https://doi.org/10.3390/e22060642>.
- [11] E. J. BEKKERS, M. W. LAFARGE, M. VETA, K. A. EPPENHOF, J. P. PLUIM, AND R. DUTTS, *Rototranslation covariant convolutional networks for medical image analysis*, in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, Switzerland, 2018, pp. 440–448, [https://doi.org/10.1007/978-3-030-00928-1\\_50](https://doi.org/10.1007/978-3-030-00928-1_50).
- [12] S. BENOIT, D. D. HOLM, AND V. PUTKARADZE, *Helical states of nonlocally interacting molecules and their linear stability: A geometric approach*, J. Phys. A, 44 (2011), 055201, <https://doi.org/10.1088/1751-8113/44/5/055201>.

- [13] J. BOISVERT, F. CHERIET, X. PENNEC, H. LABELLE, AND N. AYACHE, *Geometric variability of the scoliotic spine using statistics on articulated shape models*, IEEE Trans. Med. Imaging, 27 (2008), pp. 557–568, <https://doi.org/10.1109/TMI.2007.911474>.
- [14] R. BOTT AND L. W. TU, *Differential Forms in Algebraic Topology*, Grad. Texts in Math. 82, Springer, Berlin, 1982, <https://doi.org/10.1007/978-1-4757-3951-0>.
- [15] E. CALABI, P. J. OLVER, C. SHAKIBAN, A. TANNENBAUM, AND S. HAKER, *Differential and numerically invariant signature curves applied to object recognition*, Int. J. Comput. Vis., 26 (1998), pp. 107–135, <https://doi.org/10.1023/A:1007992709392>.
- [16] J. ČESIĆ, I. MARKOVIĆ, I. CVIŠIĆ, AND I. PETROVIĆ, *Radar and stereo vision fusion for multitar-target tracking on the special Euclidean group*, Robotics Autonom. Syst., 83 (2016), pp. 338–348, <https://doi.org/10.1016/j.robot.2016.05.001>.
- [17] E. CHEVALLIER AND N. GUIGUI, *A bi-invariant statistical model parametrized by mean and covariance on rigid motions*, Entropy, 22 (2020), 432, <https://doi.org/10.3390/e22040432>.
- [18] G. S. CHIRIKJIAN, *Stochastic Models Information Theory and Lie Groups Volume 2: Analytic Methods and Modern Applications*, Birkhäuser Boston, Cambridge, MA, 2012, <https://doi.org/10.1007/978-0-8176-4944-9>.
- [19] E. CHOI AND C. LEE, *Feature extraction based on the Bhattacharyya distance*, Pattern Recognit., 36 (2003), pp. 1703–1709, [https://doi.org/10.1016/S0031-3203\(03\)00035-9](https://doi.org/10.1016/S0031-3203(03)00035-9).
- [20] M. CHRIST, *The strong maximal function on a nilpotent group*, Trans. Amer. Math. Soc., 331 (1992), pp. 1–13, <https://doi.org/10.2307/2153994>.
- [21] D. CIESLAK, T. R. HOENS, N. V. CHAWLA, AND W. P. KEGELMEYER, *Hellinger distance decision trees are robust and skew-insensitive*, Data Min. Knowl. Discov., 24 (2011), pp. 136–158, <https://doi.org/10.1007/s10618-011-0222-1>.
- [22] T. S. COHEN AND M. WELLING, *Group equivariant convolutional networks*, in Proceedings of the 33rd International Conference on Machine Learning, Vol. 48, Curran Associates, Red Hook, NY, 2016, pp. 2990–2999.
- [23] M. P. DO CARMO, *Riemannian Geometry*, 2nd ed., Birkhäuser Boston, Cambridge, MA, 1992.
- [24] R. DURITS AND B. BURGETH, *Scale spaces on Lie groups*, in International Conference on Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2007, pp. 300–312, [https://doi.org/10.1007/978-3-540-72823-8\\_26](https://doi.org/10.1007/978-3-540-72823-8_26).
- [25] B. ELTZNER AND S. F. HUCKEMANN, *A smearsy central limit theorem for manifolds with application to high-dimensional spheres*, Ann. Statist., 47 (2019), pp. 3360–3381, <https://doi.org/10.1214/18-AOS1781>.
- [26] D. EZUZ AND M. BEN-CHEN, *Deblurring and denoising of maps between shapes*, Comput. Graph. Forum., 36 (2017), pp. 165–174, <https://doi.org/10.1111/cgf.13254>.
- [27] M. FINZI, S. STANTON, P. IZMAILOV, AND A. G. WILSON, *Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data*, in Proceedings of the 37th International Conference on Machine Learning, Curran Associates, Red Hook, NY, 2020, pp. 3165–3176.
- [28] S. FIORI AND T. TANAKA, *An algorithm to compute averages on matrix Lie groups*, IEEE Trans. Signal Process., 57 (2009), pp. 4734–4743, <https://doi.org/10.1109/TSP.2009.2027754>.
- [29] P. T. FLETCHER, C. LU, AND S. JOSHI, *Statistics of shape via principal geodesic analysis on Lie groups*, in 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, IEEE Computer Society, Los Alamitos, CA, 2003, I-I, <https://doi.org/10.1109/CVPR.2003.1211342>.
- [30] C. GOODALL, *Procrustes methods in the statistical analysis of shape*, J. R. Stat. Soc. Ser. B Stat. Methodol., 53 (1991), pp. 285–321, <https://doi.org/10.1111/j.2517-6161.1991.tb01825.x>.
- [31] F. GOUDAIL, P. RÉFRÉGIER, AND G. DELYON, *Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images*, J. Opt. Soc. Amer. A, 21 (2004), pp. 1231–1240, <https://doi.org/10.1364/JOSAA.21.001231>.
- [32] M. HALLIN AND J. JUREČKOVÁ, *Equivariant Estimation: Theory*, American Cancer Society, Atlanta, GA, 2014, <https://doi.org/doi.org/10.1002/9781118445112.stat07378>.
- [33] M. HANIK, H.-C. HEGE, A. HENNEMUTH, AND C. VON TYCOWICZ, *Nonlinear regression on manifolds for shape analysis using intrinsic Bézier splines*, in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer, Cham, Switzerland, 2020, pp. 617–626, [https://doi.org/10.1007/978-3-030-59719-1\\_60](https://doi.org/10.1007/978-3-030-59719-1_60).

- [34] M. HANIK, H.-C. HEGE, AND C. VON TYCOWICZ, *Bi-invariant two-sample tests in Lie groups for shape analysis*, in Shape in Medical Imaging, Springer, Cham, Switzerland, 2020, pp. 44–54, [https://doi.org/10.1007/978-3-030-61056-2\\_4](https://doi.org/10.1007/978-3-030-61056-2_4).
- [35] S. HELGASON, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Grad. Stud. Math. 34, American Mathematical Society, Providence, RI, 2001, <https://doi.org/10.1090/gsm/034>.
- [36] E. HELLINGER, *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen*, J. Reine Angew. Math., 1909 (1909), pp. 210–271, <https://doi.org/10.1515/crll.1909.136.210>.
- [37] Y. HONG, N. SINGH, R. KWITT, AND M. NIETHAMMER, *Group testing for longitudinal data*, in International Conference on Information Processing in Medical Imaging, Springer, Cham, Switzerland, 2015, pp. 139–151, <https://doi.org/10.1007/978-3-319-19992-4>.
- [38] Z. HUANG, C. WAN, T. PROBST, AND L. VAN GOOL, *Deep learning on Lie groups for skeleton-based action recognition*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2017, pp. 6099–6108, <https://doi.org/10.1109/CVPR.2017.137>.
- [39] S. HUCKEMANN, T. HOTZ, AND A. MUNK, *Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions*, Stat. Sinica, 20 (2010), pp. 1–58.
- [40] A. ISERLES, H. MUNTHE-KAAS, S. NØRSETT, AND A. ZANNA, *Lie-group methods*, Acta Numer., 9 (2005), pp. 215–365, <https://doi.org/10.1017/S0962492900002154>.
- [41] T. KAILATH, *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Trans. Commun., 15 (1967), pp. 52–60, <https://doi.org/10.1109/TCOM.1967.1089532>.
- [42] J. H. KELLGREN AND J. S. LAWRENCE, *Radiological assessment of osteo-arthrosis*, Ann. Rheum. Dis., 16 (1957), pp. 494–502, <https://doi.org/10.1136/ard.16.4.494>.
- [43] N. MIOLANE AND X. PENNÉC, *Computing bi-invariant pseudo-metrics on Lie groups for consistent statistics*, Entropy, 17 (2015), pp. 1850–1881, <https://doi.org/10.3390/e17041850>.
- [44] M. MOAKHER, *Means and averaging in the group of rotations*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 1–16, <https://doi.org/10.1137/S0895479801383877>.
- [45] M. MOAKHER, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 735–747, <https://doi.org/10.1137/S0895479803436937>.
- [46] S. G. MUELLER, N. SCHUFF, K. YAFFE, C. MADISON, B. MILLER, AND M. W. WEINER, *Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer’s disease*, Human Brain Mapp., 31 (2010), pp. 1339–1347, <https://doi.org/10.1002/hbm.20934>.
- [47] P. MURALIDHARAN AND P. FLETCHER, *Sasaki metrics for analysis of longitudinal data on manifolds*, in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2012, pp. 1027–1034, <https://doi.org/10.1109/CVPR.2012.6247780>.
- [48] P. OLVER, G. SAPIRO, AND A. TANNENBAUM, *Differential invariant signatures and flows in computer vision: A symmetry group approach*, in Geometry-Driven Diffusion in Computer Vision, Springer, Berlin, 1994, pp. 255–306, <https://doi.org/10.1007/978-94-017-1699-4>.
- [49] G. PAI, M. JOSEPH-RIVLIN, R. KIMMEL, AND N. SOCHEN, *On geometric invariants, learning, and recognition of shapes and forms*, in Handbook of Variational Methods for Nonlinear Geometric Data, Springer, Berlin, 2020, pp. 443–461, [https://doi.org/10.1007/978-3-030-31351-7\\_16](https://doi.org/10.1007/978-3-030-31351-7_16).
- [50] L. PARDO, *Statistical Inference Based on Divergence Measures*, Chapman and Hall, London, 2005, <https://doi.org/10.1201/9781420034813>.
- [51] F. PARK, J. BOBROW, AND S. PLOEN, *A Lie group formulation of robot dynamics*, Int. J. Robot. Res., 14 (1995), pp. 609–618, <https://doi.org/10.1177/027836499501400606>.
- [52] X. PENNÉC, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, J. Math. Imaging Vision, 25 (2006), pp. 127–154, <https://doi.org/10.1007/s10851-006-6228-4>.
- [53] X. PENNÉC AND V. ARSIGNY, *Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups*, in Matrix Information Geometry, Springer, Berlin, 2013, pp. 123–166, <https://doi.org/10.1007/978-3-642-30232-9>.
- [54] X. PENNÉC AND M. LORENZI, *Beyond Riemannian geometry: The affine connection setting for transformation groups*, in Riemannian Geometric Statistics in Medical Image Analysis, Academic Press, London, 2020, pp. 169–229, <https://doi.org/10.1016/B978-0-12-814725-2.00012-1>.
- [55] X. PENNÉC, S. SOMMER, AND T. FLETCHER, eds., *Riemannian Geometric Statistics in Medical Image Analysis*, Elsevier, Amsterdam, 2020, <https://doi.org/10.1016/C2017-0-01561-6>.

- [56] G. PEYRÉ, *Manifold models for signals and images*, Comput. Vis. Image Underst., 113 (2009), pp. 249–260, <https://doi.org/10.1016/j.cviu.2008.09.003>.
- [57] F. PORIKLI, O. TUZEL, AND P. MEER, *Covariance tracking using model update based on Lie algebra*, in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 1, IEEE Computer Society, Los Alamitos, CA, 2006, pp. 728–735, <https://doi.org/10.1109/CVPR.2006.94>.
- [58] M. POSTNIKOV, *Geometry VI: Riemannian Geometry*, Springer, Berlin, 2013, <https://doi.org/10.1007/978-3-662-04433-9>.
- [59] I. U. RAHMAN, I. DRORI, V. C. STODDEN, D. L. DONOHO, AND P. SCHRÖDER, *Multiscale representations for manifold-valued data*, Multiscale Model. Simul., 4 (2005), pp. 1201–1232, <https://doi.org/10.1137/050622729>.
- [60] C. R. RAO, *A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance*, Qüestiió, 19 (1995), pp. 23–63.
- [61] C. SAMIR, P.-A. ABSIL, A. SRIVASTAVA, AND E. KLASSEN, *A gradient-descent method for curve fitting on Riemannian manifolds*, Found. Comput. Math., 12 (2012), pp. 49–73, <https://doi.org/10.1007/s10208-011-9091-7>.
- [62] J. SCHULZ, S. PIZER, J. MARRON, AND F. GODTLIEBSEN, *Non-linear hypothesis testing of geometric object properties of shapes applied to hippocampi*, J. Math. Imaging Vision, 54 (2016), pp. 15–34, <https://doi.org/10.1007/s10851-015-0587-7>.
- [63] H. SENGAR, H. WANG, D. WIJESEKERA, AND S. JAJODIA, *Detecting VoIP floods using the Hellinger distance*, IEEE Trans. Parallel Distrib. Syst., 19 (2008), pp. 794–805, <https://doi.org/10.1109/TPDS.2007.70786>.
- [64] A. SRIVASTAVA AND E. P. KLASSEN, *Functional and Shape Data Analysis*, Vol. 1, Springer, Berlin, 2016, <https://doi.org/10.1007/978-1-4939-4020-2>.
- [65] H. TABIA, H. LAGA, D. PICARD, AND P.-H. GOSELIN, *Covariance descriptors for 3D shape matching and retrieval*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2014, pp. 4185–4192, <https://doi.org/10.1109/CVPR.2014.533>.
- [66] C. J. TAYLOR AND D. J. KRIEGMAN, *Minimization on the Lie Group  $SO(3)$  and Related Manifolds*, Technical report 9405, Yale University, 1994.
- [67] D. W. THOMPSON, *On Growth and Form*, Cambridge University Press, Cambridge, 1992, <https://doi.org/10.1017/cbo9781107325852>.
- [68] D. TOSATO, M. FARENZENA, M. SPERA, V. MURINO, AND M. CRISTANI, *Multi-class classification on Riemannian manifolds for video surveillance*, in European Conference on Computer Vision, Springer, Berlin, 2010, pp. 378–391, [https://doi.org/10.1007/978-3-642-15552-9\\_28](https://doi.org/10.1007/978-3-642-15552-9_28).
- [69] A. TROUVÉ, *Diffeomorphisms groups and pattern matching in image analysis*, Int. J. Comput. Vis., 28 (1998), pp. 213–221, <https://doi.org/10.1023/A:1008001603737>.
- [70] A. TROUVÉ AND L. YOUNES, *Metamorphoses through Lie group action*, Found. Comput. Math., 5 (2005), pp. 173–198, <https://doi.org/10.1007/s10208-004-0128-z>.
- [71] O. TUZEL, F. PORIKLI, AND P. MEER, *Region covariance: A fast descriptor for detection and classification*, in European Conference on Computer Vision, Springer, Berlin, 2006, pp. 589–600, [https://doi.org/10.1007/11744047\\_45](https://doi.org/10.1007/11744047_45).
- [72] V. VEERIAH, N. ZHUANG, AND G.-J. QI, *Differential recurrent neural networks for action recognition*, in Proceedings of the IEEE International Conference on Computer Vision, IEEE, Piscataway, NJ, 2015, pp. 4041–4049, <https://doi.org/10.1109/iccv.2015.460>.
- [73] R. VEMULAPALLI, F. ARRATE, AND R. CHELLAPPAA, *Human action recognition by representing 3d skeletons as points in a Lie group*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2014, pp. 588–595, <https://doi.org/10.1109/CVPR.2014.82>.
- [74] R. VEMULAPALLI AND R. CHELLAPPAA, *Rolling rotations for recognizing human actions from 3d skeletal data*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2016, pp. 4471–4479, <https://doi.org/10.1109/CVPR.2016.484>.
- [75] C. VON TYCOWICZ, F. AMBELLAN, A. MUKHOPADHYAY, AND S. ZACHOW, *An efficient Riemannian statistical shape model using differential coordinates*, Med. Image Anal., 43 (2018), pp. 1–9, <https://doi.org/10.1016/j.media.2017.09.004>.

- [76] J. WALLNER, E. N. YAZDANI, AND P. GROHS, *Smoothness properties of Lie group subdivision schemes*, Multiscale Model. Simul., 6 (2007), pp. 493–505, <https://doi.org/10.1137/060668353>.
- [77] R. P. WOODS, *Characterizing volume and surface deformations in an atlas framework: Theory, applications, and implementation*, NeuroImage, 18 (2003), pp. 769–788, [https://doi.org/10.1016/s1053-8119\(03\)00019-3](https://doi.org/10.1016/s1053-8119(03)00019-3).