



**HAL**  
open science

# Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups

Xavier Pennec, Vincent Arsigny

► **To cite this version:**

Xavier Pennec, Vincent Arsigny. Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups. Frederic Barbaresco and Amit Mishra and Frank Nielsen. Matrix Information Geometry, Springer, pp.123-168, 2012, 978-3-642-30231-2 (Print) / 978-3-642-30232-9 (Online). 10.1007/978-3-642-30232-9\_7. hal-00699361

**HAL Id: hal-00699361**

**<https://inria.hal.science/hal-00699361>**

Submitted on 20 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups

Xavier Pennec<sup>1</sup> and Vincent Arsigny<sup>1</sup>

Asclepios project-team, INRIA Sophia-Antipolis Méditerranée  
2004 Route des Lucioles, BP93, F-06902 Sophia Antipolis Cedex, France  
[Xavier.Pennec@inria.fr](mailto:Xavier.Pennec@inria.fr),

**Abstract.** When performing statistics on elements of sets that possess a particular geometric structure, it is desirable to respect this structure. For instance in a Lie group, it would be judicious to have a notion of a mean which is stable by the group operations (composition and inversion). Such a property is ensured for Riemannian center of mass in Lie groups endowed with a bi-invariant Riemannian metric, like compact Lie groups (e.g. rotations). However, bi-invariant Riemannian metrics do not exist for most non compact and non-commutative Lie groups. This is the case in particular for rigid-body transformations in any dimension greater than one, which form the most simple Lie group involved in biomedical image registration.

In this paper, we propose to replace the Riemannian metric by an affine connection structure on the group. We show that the canonical Cartan connections of a connected Lie group provides group geodesics which are completely consistent with the composition and inversion. With such a non-metric structure, the mean cannot be defined by minimizing the variance as in Riemannian Manifolds. However, the characterization of the mean as an exponential barycenter gives us an implicit definition of the mean using a general barycentric equation. Thanks to the properties of the canonical Cartan connection, this mean is naturally bi-invariant. We show the local existence and uniqueness of the invariant mean when the dispersion of the data is small enough. We also propose an iterative fixed point algorithm and demonstrate that the convergence to the invariant mean is at least linear.

In the case of rigid-body transformations, we give a simple criterion for the global existence and uniqueness of the bi-invariant mean, which happens to be the same as for rotations. We also give closed forms for the bi-invariant mean in a number of simple but instructive cases, including 2D rigid transformations. For general linear transformations, we show that the bi-invariant mean is a generalization of the (scalar) geometric mean, since the determinant of the bi-invariant mean is the geometric mean of the determinants of the data.

Finally, we extend the theory to higher order moments, in particular with the covariance which can be used to define a local bi-invariant Mahalanobis distance.

## 1 Introduction

Over the last 30 years, there was an explosion of imaging modalities which allows observing both the anatomy *in vivo* and *in situ* at multiple spatial scales (from cells to the whole body), multiple time scales (beating heart, growth, aging, evolution of species), and on multiple subjects. The combination of these new observation means with the computerized methods is at the heart of *computational anatomy*, an emerging discipline at the interface of geometry, statistics and image analysis which aims at developing algorithms to model and analyze the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information.

Understanding and modeling the shape of organs is made difficult by the absence of physical models for comparing different subjects, the complexity of shapes, and the high number of degrees of freedom implied. The general method is to identify anatomically representative geometric features (points, tensors, curves, surfaces, volume transformations), and to describe and compare their statistical distribution in different populations. Although anatomical features are embedded in Euclidean spaces, the extracted geometric features most often belong to manifolds. For instance, the spine shape can be characterized by the relative position and orientation of each vertebra with respect to the previous one [2, 14]. Considering the degrees of freedom of an articulated object amounts to work in a subspace of a product of Lie groups (here rigid-body transformations) rather than in a Euclidean space. Likewise, in order to model biological shapes, D’Arcy Thompson proposed in 1917 to assume that there is a template object which represents the reference shape, and to encode the shape variations as deformations of this template [20]. With this approach, we perform statistics on deformations which naturally belong to continuous transformation groups, i.e. Lie groups. As the deformation of a smooth object should be a smooth object, the most general transformation group to consider is the group of diffeomorphisms (invertible, one-to-one mappings with smooth inverses). This formalism was developed in particular by Grenander and Miller [31, 48] based on advanced mathematical tools to compute on infinite dimensional groups of diffeomorphisms [64, 70].

Among statistics, the most fundamental is certainly the *mean*, which extracts from the data a central point, minimizing in some sense the dispersion of the data around it. In this work, we focus on a generalization of the mean to connected Lie groups. Classically, in a manifold endowed with a Riemannian metric, the natural choice of mean is called the Riemannian center of mass or *Fréchet mean*. The Riemannian structure proves to be a powerful and consistent framework for computing simple statistics [72, 50, 53, 11–13, 55] and can be extended to an effective computing framework on manifold-valued images [57]. On a Lie group, this Riemannian approach is consistent with the group operations if a bi-invariant metric exists, which is for example the case for compact groups such as

rotations [54, 49]. In this case, the bi-invariant Fréchet mean has many desirable *invariance properties*: it is invariant with respect to left- and right-multiplication, as well as inversion. Unfortunately, bi-invariant Riemannian metrics do not exist for most non compact and non-commutative Lie groups. In particular, such metrics do not exist in any dimension for *rigid-body transformations*, which form the most simple Lie group involved in biomedical image registration.

To overcome the lack of existence of bi-invariant Riemannian metrics for general Lie groups, we propose in this work to rely on a convenient affine connection structure. We first recall in Section 2 the basic properties of connected Lie groups and the notion of affine connection spaces. Among the connections that are left-invariant on a Lie group, the Cartan-Schouten connections are defined as the ones for which one-parameter subgroups are geodesics. The unique symmetric Cartan-Schouten connection is called the canonical Cartan connection of the group. We show that it provides group geodesics which are completely consistent with composition and inversion.

We turn to the definition of means in Lie groups in Section 3. In order to define bi-invariant Fréchet means, we investigate the existence conditions of bi-invariant metrics on Lie groups. It turns out that most non compact and non commutative Lie groups do not possess any bi-invariant Riemannian metric. We show that this is the case for rigid-body transformations, which constitutes one of the simplest Lie groups of interest in image analysis. However, a weaker structure of bi-invariant affine connection space exists for all connected Lie groups thanks to the canonical Cartan connection. With such a non-metric structure, the mean can obviously not be defined by minimizing the variance as there is no distance. However, the characterization of the mean as an exponential barycenter is still valid and gives us an implicit definition of the mean using a general *barycentric equation*. Thanks to the properties of the canonical Cartan connection, this mean is naturally equivariant with respect to left- and right-translations as well as inversions. We show the existence and uniqueness of the mean when the dispersion of the data is small enough by defining a mapping whose fixed point is by definition the mean. Moreover, the mapping being a contraction, it is converging to the mean at least linearly.

Section 4 focuses on the bi-invariant mean in selected matrix Lie groups. We show that a closed form can be obtained in the case of the group of scaling and translations, which constitutes one of the simplest examples of non-compact and non-commutative Lie groups which do not possess a bi-invariant Riemannian metric. There is also a closed form of the bi-invariant mean for the Heisenberg group, which can be generalized to the scaled unitriangular matrix group. For general rigid-body transformations, there is no closed form but we determine a simple criterion for the general existence and uniqueness of the bi-invariant mean, which happens to be the same as the bi-invariant Fréchet mean for rotations. For general linear transformations, we show that the bi-invariant mean is a generalization of the geometric mean of scalars, in the sense that the determinant of the mean is equal to the geometric mean of the determinants of the data.

Finally, Section 5 investigates some open points, such as the characterization of global existence and uniqueness conditions for the bi-invariant mean. It also extends the theory to higher order moments, in particular with the (2-covariant) covariance tensor which can be used to define a local bi-invariant Mahalanobis distance.

**Related works** The intuition behind such a bi-invariant mean on matrix Lie groups was present in [66] along with a practical iterative algorithm to compute it. However, no precise definition nor proof of convergence was provided. The barycentric definition of bi-invariant means on Lie groups based on one-parameter subgroups was developed in the the PhD of Vincent Arsigny [7] and in the research report [8]. In this preliminary work, the 'group geodesics' were simply defined as left translations of one-parameters subgroups without further justification. This paper extends this work by reformulating and rigorously justifying 'group geodesics' as the geodesics of the canonical Cartan-Schouten connections [17]. This allows better distinguishing the properties that are related to the connection itself (bi-invariance) from the ones that are related to the definition of the mean as an exponential barycenter in an affine connection space. A number of proofs were added or adapted in consequence.

The barycentric fixed point iteration on Lie groups that we investigate in this paper to compute the bi-invariant mean is close to the Gauss-Newton gradient descent iteration that is used on Riemannian manifolds to compute the Fréchet mean, studied in depth in [44, 45] or [32] for a generalization to zeros of vector fields on Riemannian manifolds. Indeed, both algorithms do correspond when the Lie group is provided with a bi-invariant metric. However, they differ when the Lie group is non compact and non commutative: since the canonical Cartan connections are not metric, there does not exist a Riemannian metric for which the group geodesics are Riemannian geodesics. Thus, in the general case, this algorithm does not enter into the type of iterations that are studied in [32].

In the context of optimization on Riemannian manifolds, many Newton-like methods were proposed and their convergence studied in depth, e.g. in [62, 65, 32, 21, 1] to cite just a few. The iteration we investigate in this paper does not enter into this family as it does not use the covariant derivative of the vector field of which we are trying to find the zeros. Moreover, we cannot recast it as an optimization problem on a Riemannian manifold, as stated above. Very few works deal with Newton iterations on Lie groups or affine connection spaces. Notable exceptions are [51, 47] which propose Newton algorithms to optimize general functions on non compact Lie groups based on Cartan-Schouten connections. In terms of the geometric tools used, these papers are the closest to what is present in the first Section of this paper. However, they only focuses on the optimization of functions on Lie groups without investigating any notion of bi-invariant mean.

## 2 Lie groups

### 2.1 Basics of Lie Groups

A Lie group  $\mathcal{G}$  is a smooth manifold provided with an identity element  $e$ , a smooth associative composition rule  $(g, h) \in \mathcal{G} \times \mathcal{G} \mapsto g.h \in \mathcal{G}$  and a smooth inversion rule  $\text{Inv} : f \mapsto f^{(-1)}$  which are both compatible with the manifold structure. In the sequel of this paper we always assume that the Lie group is finite dimensional and connected. Classical examples of Lie groups are vector spaces (with their commutative addition as multiplication), multiplicative matrix groups:  $GL(n)$ ,  $O(n)$ ,  $SO(n)$ , etc., with the usual matrix multiplication; and geometric transformation groups such as rigid-body transformations, similarities, affine transformations which can anyway also be looked upon as matrix groups via their ‘faithful’ representation based on homogeneous coordinates. Infinite-dimensional *Lie groups of diffeomorphisms* have also been recently gaining a considerable importance in computational anatomy [64, 70]. However, we concentrate here only on finite dimensional Lie groups.

Many details on differential geometry can be found in classical books like [63, 29, 41, 22]. Specific details on Lie groups can be found in [34]. Most results of this section can be found in the more modern (and quite comprehensive) presentation of differential geometry and Lie groups of Postnikov [58].

**Lie Bracket** We denote by  $T_g\mathcal{G}$  the tangent space at  $g \in \mathcal{G}$  and by  $T\mathcal{G}$  the tangent bundle. A section  $X : \mathcal{G} \mapsto T\mathcal{G}$  is a vector field whose value at  $g$  is denoted  $X|_g$ . The set of vector fields  $\Gamma(\mathcal{G})$  is the algebra of derivations of smooth functions  $\phi \in C^\infty(\mathcal{G})$ . Recall that a derivation  $\delta$  is a linear map from  $\Gamma(\mathcal{G})$  to  $\Gamma(\mathcal{G})$  that satisfies Leibniz’s law:  $\delta(fX) = (df)X + f(\delta X)$  for any  $f \in C^\infty(\mathcal{G})$  and  $X \in \Gamma(\mathcal{G})$ . In a local coordinate system, we can indeed write  $X\phi|_g = \partial_X\phi|_g = \frac{d}{dt}(\phi(g + tX|_g))$ .

When composing the derivations, we can see that  $XY\phi = \partial_X\partial_Y\phi$  is a second order differential and is thus not a derivation. However, we can remove the second order terms by subtracting  $\partial_Y\partial_X\phi$  (this can be checked by writing these expression in a local coordinate system). We obtain the Lie bracket

$$[X, Y](\phi) = \partial_X\partial_Y\phi - \partial_Y\partial_X\phi.$$

which is also called the Lie derivative  $\mathcal{L}_X Y$  because it is conceptually the derivative of the vector field  $X$  in the ‘direction’ generated by  $Y$ . The Lie bracket is a bilinear operator from  $\Gamma\mathcal{G} \times \Gamma\mathcal{G}$  to  $\Gamma\mathcal{G}$ . It is obviously skew symmetric ( $[X, Y] = -[Y, X]$ ), and verifies the Jacobi identity  $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$ . As it is a derivation, it also verifies  $[X, \phi Y] = \phi[X, Y] + \partial_X(\phi Y)$  for any function  $\phi$ .

**Lie Algebra** Thanks to the group structure, we can define two canonical diffeomorphisms of  $\mathcal{G}$  called the left and the right translations:  $L_g(f) = g.f$  and  $R_g(f) = f.g$  in addition to the inversion map  $\text{Inv}(f) = f^{(-1)}$ . A fourth important

diffeomorphism of  $\mathcal{G}$  is the conjugation  $C_g(f) = g.f.g^{(-1)} = L_g \circ R_{g^{(-1)}} = R_{g^{(-1)}} \circ L_g$  (this is actually an inner automorphism).

The differential  $DL_g$  of the left translations maps the tangent space  $T_h\mathcal{G}$  to the tangent space  $T_g\mathcal{G}$ . In particular,  $L_g$  maps any vector  $x \in T_e\mathcal{G}$  to the vector  $DL_g.x \in T_g\mathcal{G}$ , giving rise to the vector field  $\tilde{X}|_g = DL_g.x$ . One verifies that this vector field is left-invariant:  $\tilde{X} \circ L_h = DL_h\tilde{X}$ . Conversely, every left-invariant vector field is determined by its value at identity. Moreover, the bracket of two left-invariant vector fields  $\tilde{X} = DL.x$  and  $\tilde{Y} = DL.y$  is also left-invariant. Thus, it is determined by the vector  $[x, y] = [\tilde{X}, \tilde{Y}]|_e \in T_e\mathcal{G}$ . This allows to identify the sub-algebra of left-invariant vector fields with the tangent vector space at identity provided with the additional bracket operation:  $\mathfrak{g} = (T_e\mathcal{G}, +, \cdot, [\cdot, \cdot])$  is called the Lie algebra of the group  $\mathcal{G}$ .

Simple examples of Lie brackets are given by  $GL(n)$  and its multiplicative subgroups, like  $SL(n)$  or  $SO(n)$ . In these cases, the Lie algebra is a *vector space of square matrices*, and the Lie bracket between two elements  $M$  and  $N$  of this algebra is the *commutator* of these two matrices, i.e.  $[M, N] = M.N - N.M$ . In particular, the Lie algebra of  $GL(n)$  is  $M(n)$ , the vector space of square matrices, that of  $SL(n)$  is the vector subspace of  $M(n)$  of trace-free matrices, and the Lie algebra of  $SO(n)$  is the vector space of skew symmetric matrices.

By symmetry, we can also define the sub-algebra of right-invariant vector fields  $\tilde{X}|_g = DR_g.X$  and identify it with the tangent vector space at identity. However, one should be careful that the right-bracket is the opposite of the left-bracket.

**Adjoint Group** The adjoint is the automorphism of the Lie algebra obtained by differentiating the conjugation  $C_g(f) = g.f.g^{(-1)}$  with respect to  $f$ . More precisely, an element  $g$  of  $\mathcal{G}$  acts on an element  $x$  of  $\mathfrak{g}$  by

$$\text{Ad}(g).x = DL_g|_{g^{(-1)}}.DR_{g^{(-1)}}|_e.x = DR_{g^{(-1)}}|_g.DL_g|_e.x.$$

In the matrix case, we have the (simple this time) formula:  $\text{Ad}(R).M = R.M.R^{(-1)}$ , which only uses two matrix multiplications and one matrix inversion.

Thus, one can map each element of the group to a *linear operator* which acts on the Lie algebra:  $\mathcal{G}$  can be ‘represented’ by the adjoint operators acting on  $\mathfrak{g}$ . This is a representation in the sense of *representation theory* (see [42] for a complete treatment) as this mapping is a Lie group homomorphism from  $\mathcal{G}$  to  $GL(\mathfrak{g})$ . This means that  $\text{Ad} : \mathcal{G} \rightarrow GL(\mathfrak{g})$  is a smooth map from  $\mathcal{G}$  to  $GL(\mathfrak{g})$  which is compatible with the group structure:  $\text{Ad}(e) = Id, \forall g \in \mathcal{G}, \text{Ad}(g^{(-1)}) = \text{Ad}(g)^{(-1)}$  and  $\forall g, h \in \mathcal{G}, \text{Ad}(g.h) = \text{Ad}(g).\text{Ad}(h)$ . The subgroup  $\text{Ad}(\mathcal{G})$  of the general linear group  $GL(\mathfrak{g})$  is called the *adjoint group*. The properties of this representation and the existence of bi-invariant metrics for the group  $\mathcal{G}$  are highly linked.

## 2.2 Lie Group Exponential and Logarithm

**The Matrix Case** Before we present the general group exponential and logarithm, let us recall the fundamental properties of the *matrix* exponential and logarithm, which correspond to the group exponential and logarithm of the Lie group of  $n \times n$  invertible matrices,  $GL(n)$ . They are the generalization to matrices of the well-known scalar exponential and logarithm. The exponential  $\exp(M)$  of a matrix  $M$  is given by  $\exp(M) = \sum_{n=0}^{\infty} M^n/n!$ . Let  $G \in GL(n)$ . If there exists  $M \in M(n)$  such that  $G = \exp(M)$ , then  $M$  is said to be a logarithm of  $G$ .

In general, the logarithm of a real invertible matrix may not exist, and it may not be unique if it exists. The lack of existence is a general phenomenon in connected Lie groups. One generally needs *two* exponentials to reach every element [67]. The lack of uniqueness is essentially due to the influence of rotations: rotating of an angle  $\alpha$  is the same as rotating of an angle  $\alpha + 2k\pi$  where  $k$  is an integer. Since the logarithm of a rotation matrix directly depends on its rotation angles (one angle suffices in 3D, but several angles are necessary when  $n > 3$ ), it is not unique.

When a real invertible matrix has no (complex) eigenvalue on the (closed) half line of negative real numbers, then it has a unique real logarithm whose (complex) eigenvalues have an imaginary part in  $] -\pi, \pi[$  [39, 28]. In this case this particular logarithm is well-defined and called *the principal logarithm*. We will write  $\log(M)$  for the principal logarithm of a matrix  $M$  whenever it is defined.

Thanks to their remarkable *algebraic* properties, and essentially their link with one-parameter subgroups, matrix exponential and logarithms can be quite efficiently numerically computed. In practice, we have used in this work the popular ‘Scaling and Squaring Method’ [35] to compute numerically matrix exponentials, as well as the ‘Inverse Scaling and Squaring Method’ [18] to compute matrix logarithms.

**One Parameter Subgroups and Lie Group Exponential** Let us now define the general group exponential and logarithm in Lie groups. These properties are very similar to those of the matrix exponential and logarithm, which are a particular case of such mappings. One should note that this particular case is actually quite general, since most classical Lie groups can be looked upon as matrix Lie groups anyway [33].

The flow  $\gamma_x(t)$  of a left-invariant vector field  $\tilde{X} = DL.x$  starting from  $e$  exists for all times. Its tangent vector is  $\dot{\gamma}_x(t) = DL_{\gamma_x(t)}.x$  by definition of the flow. Now fix  $s \in \mathbb{R}$  and observe that the two curves  $\gamma_x(s+t)$  and  $\gamma_x(s).\gamma_x(t)$  are going through point  $\gamma_x(s)$  with the same tangent vector. By the uniqueness of the flow, they are the same and  $\gamma_x$  is a one parameter subgroup, i.e. a group morphism from  $(\mathbb{G}, e, .)$  to  $(\mathbb{R}, 0, +)$ :

$$\gamma_x(s+t) = \gamma_x(s).\gamma_x(t) = \gamma_x(t+s) = \gamma_x(t).\gamma_x(s).$$

The group exponential is defined from these one-parameter subgroups with  $\text{Exp}(x) = \gamma_x(1)$ .

**Definition 1.** Let  $\mathcal{G}$  be a Lie group and let  $x$  be an element of the Lie Algebra  $\mathfrak{g}$ . The group exponential of  $x$ , denoted  $\text{Exp}(x)$ , is given by the value at time 1 of the unique function  $\gamma_x(t)$  defined by the ordinary differential equation (ODE)  $\dot{\gamma}_x(t) = DL_{\gamma_x(t)}.x$  with initial condition  $\gamma_x(0) = e$ .

Very much like the exponential map associated to a Riemannian metric, the group exponential is diffeomorphic locally around 0. More precisely, since the exponential is a smooth mapping, the fact that its differential map is invertible at  $e$  allows for the use of the ‘Inverse Function Theorem’, which guarantees that it is a diffeomorphism from some open neighborhood of 0 to an open neighborhood of  $\text{Exp}(0) = e$  [58, Proposition 1.3, p.13].

**Theorem 1.** The group exponential is a diffeomorphism from an open neighborhood of 0 in  $\mathfrak{g}$  to an open neighborhood of  $e$  in  $\mathcal{G}$ , and its differential map at 0 is the identity.

This theorem implies that one can define *without ambiguity* a logarithm in an open neighborhood of  $e$ : for every  $g$  in this open neighborhood, there exists a unique  $x$  in the open neighborhood of 0 in  $\mathfrak{g}$ , such that  $g = \text{Exp}(x)$ . In the following, we will write  $x = \text{Log}(g)$  for this logarithm, which is the (abstract) equivalent of the (matrix) *principal* logarithm. The absence of an inverse function theorem in infinite dimensional Fréchet manifolds prevents the straightforward extension of this property to general groups of diffeomorphisms [40].

**Baker-Campbell-Hausdorff Formula** the Baker-Campbell-Hausdorff formula (or BCH formula) is a fundamental property of the group exponential and logarithm. Intuitively, this formula shows how much  $\text{Log}(\text{Exp}(x) \cdot \text{Exp}(y))$  deviates from  $x + y$  due to the (possible) non-commutativity of the multiplication in  $\mathcal{G}$ . Remarkably, this deviation can be expressed only in terms of Lie brackets between  $x$  and  $y$  [30, Chap. VI].

**Theorem 2 (Series form of the BCH formula).** Let  $x, y$  be in  $\mathfrak{g}$ . If they are small enough, then the logarithm of the product  $\text{Exp}(x) \cdot \text{Exp}(y)$  is always well-defined and we have the following development:

$$\begin{aligned} BCH(x, y) &= \text{Log}(\text{Exp}(x) \cdot \text{Exp}(y)) \\ &= x + y + 1/2([x, y]) + 1/12([x, [x, y]] + [y, [y, x]]) \\ &\quad + 1/24([x, [x, [x, y]], y]) + O((\|x\| + \|y\|)^5). \end{aligned} \quad (1)$$

A fundamental property of this function is the following: it is not only  $\mathcal{C}^\infty$  but also *analytic* around 0, which means that  $BCH(x, y)$  (near 0) is the sum of an absolutely converging multivariate infinite series (the usual multiplication is replaced here by the Lie bracket). This implies in particular that all the (partial) derivatives of this function are also analytic.

### 2.3 Affine Connection Spaces

For each tangent vector  $x \in \mathfrak{g} \simeq T_e\mathcal{G}$ , the one parameter subgroup  $\gamma_x(t)$  is a curve that starts from identity with this tangent vector. One could question if

this curve could be seen as a geodesic. To answer this question, we first need to define what are geodesics.

**Affine Connections** When one wants to compare data in the tangent space at one point of the group with data in the tangent space at another point of the group, one needs to define a specific mapping between these two different tangent spaces: this is the notion of parallel transport. As there is generally no way to define globally a linear operator  $\Pi_g^h : T_g\mathcal{G} \rightarrow T_h\mathcal{G}$  which is consistent with the composition (i.e.  $\Pi_g^h \circ \Pi_f^g = \Pi_f^h$ ), one has to specify the path by which we connect the two points.

An (affine) connection is the specification of the parallel transport for infinitesimal displacements. It is a bilinear map  $\nabla$  from  $\Gamma\mathcal{G} \times \Gamma\mathcal{G}$  to  $\Gamma\mathcal{G}$  such that for all smooth functions  $\phi \in C^\infty(\mathcal{G})$ :

- $\nabla_{\phi X} Y = \phi \nabla_X Y$ , that is,  $\nabla$  is smooth and linear in the first variable;
- $\nabla_X(\phi Y) = \partial_X \phi Y + \phi \nabla_X Y$ , i.e.  $\nabla$  satisfies Leibniz rule in the second variable.

In a local chart, the connection is determined by its coordinates on the basis vector fields:  $\nabla_{\partial_i} \partial_j = \sum_k \Gamma_{ij}^k \partial_k$ . The  $n^3$  coordinates  $\Gamma_{ij}^k$  of the connection are functions called the Christoffel symbols. They encode how the projection from one tangent space to the neighboring one modifies the standard derivative of a vector field in a chart:

$$\nabla_X Y = \partial_X Y + \sum_{i,j,k} x^i y^j \Gamma_{ij}^k \partial_k.$$

**Geodesics** Geodesics can be defined in affine connection spaces as the generalization of straight lines: these are the curves that remain parallel to themselves (auto-parallel curves). Thus,  $\gamma(t)$  is a geodesic if its tangent vector  $\dot{\gamma}(t)$  remains parallel to itself, i.e. if the covariant derivative  $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$  of  $\gamma$  is zero. In a local coordinate system where  $\dot{\gamma} = \sum_i \dot{\gamma}^i \partial_i$ , the equation of the geodesics is thus (in Einstein notations):  $\ddot{\gamma}^k + \Gamma_{ij}^k \dot{\gamma}^i \dot{\gamma}^j = 0$ .

We retrieve here the standard equation of the geodesics in Riemannian geometry without having to rely on any particular metric. However, what is remarkable is that we still conserve many properties of the Riemannian exponential map in affine connection spaces: as geodesics are locally defined by a second order ordinary differential equation, the geodesic  $\gamma_{(p,v)}(t)$  starting at any point  $p$  with any tangent vector  $v$  is defined for a sufficiently small time, which means that we can define the exponential map  $\text{Exp}_p(v) = \gamma_{(p,v)}(1)$  for a sufficiently small neighborhood. Moreover, the *strong Whitehead theorem* still holds.

**Theorem 3 (Strong Form of Whitehead Theorem).** *Each point of an affine connection space has a normal convex neighborhood (NCN) in the sense that for any couple of points  $(p, q)$  in this neighborhood, there exists a unique geodesic  $\gamma(t)$  joining them that is entirely contained in this neighborhood. Moreover, the geodesic  $\gamma(t)$  depends smoothly on the points  $p$  and  $q$ .*

The proof of this theorem essentially involves the non-singularity of the differential of the map  $\Phi(p, v) = (p, \text{Exp}_p(v))$  and the inverse function theorem, with the use of an auxiliary Euclidean metric on the tangent spaces around the point of interest. We refer to [58, Proposition 1.3, p.13] for the detailed proof.

As geodesics control many properties of the space, it is interesting to know which affine connections lead to the same geodesics. Intuitively, a geodesic for a connection  $\nabla$  will remain a geodesic for another connection  $\bar{\nabla}$  if the parallel transport of the tangent vector *in the direction of this tangent vector* remains unchanged, i.e. if  $\nabla_X X = \bar{\nabla}_X X$  for any vector field  $X$ . However, the parallel transport of other vectors of a frame can change, hence the notion of torsion of a connection which is defined by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = -T(Y, X).$$

It measures how the skew-symmetric part differ from the Lie derivative  $\mathcal{L}_X Y = [X, Y]$ . The connection is called torsion free if the torsion vanishes. One can show that two connections have the same geodesics if they have the same symmetric part  $(\nabla_X Y + \nabla_Y X)/2$ . i.e. if they only differ by torsion. Thus, at least for the geodesics, we can restrict our attention to the torsion free connections.

## 2.4 Canonical Cartan Connection and Group Geodesics

**Cartan-Schouten Connections** Let us now consider left-invariant connections, i.e. verifying  $\nabla_{DL_g X} DL_g Y = DL_g \nabla_X Y$  for any vector fields  $X$  and  $Y$  and any group element  $g \in \mathcal{G}$ . As the connection is completely determined by its action on the sub-algebra of left-invariant vector fields, we can restrict to this sub-algebra. Let  $\tilde{X} = DL.x$  and  $\tilde{Y} = DL.y$  be two left-invariant vector fields. Stating that the covariant derivative of  $\tilde{Y}$  along  $\tilde{X}$  is left-invariant amounts to say that the field  $\nabla_{\tilde{X}} \tilde{Y} = DL(\nabla_{\tilde{X}} \tilde{Y}|_e)$  is determined by its value at identity  $\alpha(x, y) = \nabla_{\tilde{X}} \tilde{Y}|_e \in \mathfrak{g}$ . Conversely, each bilinear operator of the Lie algebra  $\alpha : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$  uniquely defines the connection at the identity and thus on all left-invariant vector fields:  $\nabla_{\tilde{X}} \tilde{Y} = \tilde{\alpha}(x, y)$ . The connection is then uniquely extended to all vector fields using the linearity in the first variable and the Leibniz rule.

**Definition 2 (Cartan-Schouten and Bi-invariant Connections).** *Among the left-invariant connections, the Cartan-Schouten connections are the ones for which geodesics going through identity are one parameter subgroups. Bi-invariant connections are both left- and right-invariant.*

The definition of Cartan-Schouten connection used here [58, Def. 6.2 p.71] generalizes the three classical  $+$ ,  $-$  and  $0$  Cartan-Schouten connections [17] (see below for their definition).

**Theorem 4.** *Cartan-Schouten connections are uniquely determined by the property  $\alpha(x, x) = 0$  for all  $x \in \mathfrak{g}$ .*

Bi-invariant connections are characterized by the condition:

$$\alpha([z, x], y) + \alpha(x, [z, y]) = [z, \alpha(x, y)] \quad \forall x, y, z \in \mathfrak{g}. \quad (2)$$

The one dimensional family of connections generated by  $\alpha(x, y) = \lambda[x, y]$  satisfy these two conditions.

Moreover, there is a unique symmetric Cartan-Schouten bi-invariant connection called the canonical Cartan connection of the Lie group (also called mean or 0-connection) defined by  $\alpha(x, y) = \frac{1}{2}[x, y]$  for all  $x, y \in \mathfrak{g}$ , i.e.  $\nabla_{\tilde{X}}\tilde{Y} = \frac{1}{2}[\tilde{X}, \tilde{Y}]$  for two left-invariant vector fields.

Indeed, let us consider the one-parameter subgroup  $\gamma_x(t)$  starting from  $e$  with initial tangent vector  $x \in \mathfrak{g}$ . As this is the integral curve of the left-invariant vector field  $\tilde{X} = DL.x$ , its tangent vector is  $\dot{\gamma}_x(t) = DL_{\gamma_x(t)}.x = \tilde{X}|_{\gamma_x(t)}$ . The curve is a geodesic if and only if it is auto-parallel, i.e. if

$$\nabla_{\dot{\gamma}_x}\dot{\gamma}_x = \nabla_{\tilde{X}}\tilde{X} = \tilde{\alpha}(x, x) = 0.$$

Thus, the one-parameter subgroup  $\gamma_x(t)$  is a geodesic if and only if  $\alpha(x, x) = 0$ .

This condition implies that the operator  $\alpha$  is skew-symmetric. However, if any skew-symmetric operator give rise to a left-invariant connection, this connection is not always right-invariant. The connection is right-invariant if  $\nabla_{DR_g X} DR_g Y = DR_g \nabla_X Y$  for any vector fields  $X$  and  $Y$  and any group element  $g$ . As we have  $(dR_g \tilde{X}) = \widetilde{\text{Ad}(g^{-1}).x}$  for any left-invariant vector field  $\tilde{X} = DL.x$ , the right-invariance is equivalent to the Ad-invariance of the operator  $\alpha$ :

$$\alpha(\text{Ad}(g^{-1})x, \text{Ad}(g^{-1})y) = \text{Ad}(g^{-1}).\alpha(x, y),$$

for any two vectors  $x, y \in \mathfrak{g}$  and  $g \in \mathcal{G}$ . We can focus on the infinitesimal version of this condition by taking the derivative at  $t = 0$  with  $g^{-1} = \text{Exp}(tz)$ . Since  $\frac{d}{dt}\text{Ad}(\text{exp}(tz)).x = [z, x]$  we obtain the requested characterization of bi-invariant connections:  $\alpha([z, x], y) + \alpha(x, [z, y]) = [z, \alpha(x, y)]$ .

The well known one-dimensional family of connections generated by  $\alpha(x, y) = \lambda[x, y]$  obviously satisfy this condition (in addition to  $\alpha(x, x) = 0$ ). It was shown by Laquer [43] that this family basically describes all the bi-invariant connections on compact simple Lie groups (the exact result is that the space of bi-invariant affine connections on  $G$  is one-dimensional) *except* for  $SU(n)$  when  $n > 3$ : in the case of  $SU(n)$  there is a two-dimensional family of bi-invariant affine connections.

The torsion of a connection can be expressed in the basis of left-invariant vector fields:  $T(\tilde{X}, \tilde{Y}) = \tilde{\alpha}(x, y) - \tilde{\alpha}(y, x) - \widetilde{[x, y]}$ . This is itself a left-invariant vector field characterized by its value at identity  $T(x, y) = \alpha(x, y) - \alpha(y, x) - [x, y]$ . Thus, the torsion of a Cartan-Schouten connection is  $T(x, y) = 2\alpha(x, y) - [x, y]$  and we are left with a unique torsion-free Cartan connection characterized by  $\alpha(x, y) = \frac{1}{2}[x, y]$ .

**Curvature of the Cartan-Schouten Connections** As for the torsion, the curvature tensor  $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$  can be expressed

in the basis of left-invariant vector fields and since it is left-invariant, it is characterized by its value in the Lie algebra:

$$R(x, y)z = \alpha(x, \alpha(y, z)) - \alpha(y, \alpha(x, z)) - \alpha([x, y], z).$$

For Cartan connections of the form  $\alpha(x, y) = \lambda[x, y]$ , the curvature becomes  $R(x, y)z = \lambda(\lambda - 1)[[x, y], z]$ . For  $\lambda = 0$  and  $\lambda = 1$ , the curvature is obviously null. These two flat connections are called the left and right (or + and -) Cartan connections. For the canonical Cartan connection (often called mean or 0-connection), the curvature is  $R(x, y)z = -\frac{1}{4}[[x, y], z]$ , which is generally non zero.

Among the Cartan-Schouten connection, the +, - and 0 (or left, right and mean) connections have special properties and are often called the three canonical Cartan connections. As all the Cartan connections of the form  $\alpha(x, y) = \lambda[x, y]$ , these three connections have exactly the same geodesics (left or right translations of one-parameter subgroups) because they share the same symmetric part  $\nabla_X Y + \nabla_Y X = \partial_X Y + \partial_Y X$ . However, the - connection is the unique connection for which all the left-invariant vector fields are covariantly constant; the + connection is the only connection for which all the right-invariant vector fields are covariantly constant; and the 0-connection is the only one which is torsion-free (it has curvature, but its curvature tensor is covariantly constant). Since we will only focus on geodesics in the sequel and not on the parallel transport (which differ for the three connections), we only consider from now on the (mean or 0) canonical Cartan connection, but the results would be the same for the + and - connections.

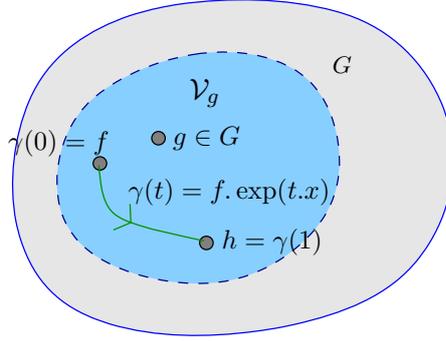
**Group Geodesics** We call *group geodesics* the geodesics of the canonical Cartan connection. We already know that the geodesics going through identity are the one-parameter subgroups (by definition of the Cartan-Schouten connections). The canonical Cartan connection being left-invariant, the curve  $\gamma(t) = g \cdot \text{Exp}(tx)$  is also a geodesic. We have indeed  $\dot{\gamma} = DL_g \cdot \dot{\gamma}_x$  and  $\nabla_{\dot{\gamma}} \dot{\gamma} = DL_g \nabla_{\dot{\gamma}_x} \dot{\gamma}_x = 0$ . As  $\gamma(0) = DL_g \cdot x$ , we finally obtain that:

**Theorem 5.** *The group geodesic starting at  $g$  with tangent vector  $v \in T_g \mathcal{G}$  is  $\gamma_{(g,v)}(t) = g \cdot \text{Exp}(t \cdot DL_{g^{(-1)}} \cdot v)$ . Thus, the (group) exponential map at point  $g$  is:*

$$\text{Exp}_g(v) = \gamma_{g,v}(1) = g \cdot \text{Exp}(DL_{g^{(-1)}} \cdot v).$$

As noted in Theorem 3, there exists for each point  $g$  of  $G$  a normal convex neighborhood (NCN) in the sense that for any couple of points  $(f, h)$  in this neighborhood, there exists a unique geodesic of the form  $\text{Exp}_f(t \cdot v)$  joining them which lies completely in this neighborhood. Furthermore, a NCN  $\mathcal{V}_e$  of the identity is transported by left-invariance into a NCN  $g \cdot \mathcal{V}_e$  of any point  $g \in \mathcal{G}$ .

Of course, we could have defined the geodesics using the right translations to obtain curves of the form  $\text{Exp}(t \cdot DR_{g^{(-1)}} \cdot v) \cdot g$ . In fact, those two types of group geodesic are the same and are related by the adjoint operator, as shown below. However, we should be careful that the left and right transport of the NCN at the identity lead to different NCN of a point  $g$ :  $g \cdot \mathcal{V} \neq \mathcal{V} \cdot g$ .



**Fig. 1. Group geodesic convexity.** For any point  $g$  of  $G$ , there exists an open neighborhood  $\mathcal{V}_g$  of  $g$ , such that any couple of points  $f$  and  $h$  in  $\mathcal{V}_g$  can be joined by a unique group geodesics of the form  $\gamma(t) = f \cdot \text{Exp}(t.x)$  satisfying  $f \cdot \text{Exp}(x) = h$  and which is entirely contained in  $\mathcal{V}_g$ .

**Theorem 6.** *Let  $x$  be in  $\mathfrak{g}$  and  $g$  in  $\mathcal{G}$ . Then we have:*

$$g \cdot \text{Exp}(x) = \text{Exp}(\text{Ad}(g).x).g.$$

For all  $g$  in  $\mathcal{G}$ , there exists an open neighborhood  $\mathcal{W}_g$  of  $e \in G$  (namely  $\mathcal{W}_g = \mathcal{V}_e \cap g \cdot \mathcal{V}_e \cdot g^{(-1)}$  where  $\mathcal{V}_e$  is any NCN of  $e$ ) such that for all  $m \in \mathcal{W}_g$  the quantities  $\text{Log}(m)$  and  $\text{Log}(g.m.g^{(-1)})$  are well-defined and are linked by the following relationship:

$$\text{Log}(g.m.g^{(-1)}) = \text{Ad}(g). \text{Log}(m).$$

Notice that in general the NCN  $\mathcal{W}_g$  depends on  $g$  unless we can find a NCN  $\mathcal{V}_e$  that is stable by conjugation.

These equations are simply the generalization to (abstract) Lie groups of the well-known matrix properties:

$$G \cdot \exp(V) \cdot G^{(-1)} = \exp(G.V.G^{(-1)}) \quad \text{and} \quad G \cdot \log(V) \cdot G^{(-1)} = \log(G.V.G^{(-1)}).$$

**Corollary 1.** *For all  $g$  in  $\mathcal{G}$ , there exists an open neighborhood  $\mathcal{V}_g$  of  $g$  such that the local Exponential and Logarithmic maps of the canonical Cartan connection are well defined and the inverse of each other. Moreover, their left and right expressions are:*

$$\text{Exp}_g(v) = g \cdot \text{Exp}(DL_{g^{(-1)}}.v) = g \cdot \text{Exp}(DR_{g^{(-1)}}.v) \quad \text{for } v \in T_g\mathcal{G};$$

$$\text{Log}_g(x) = DL_g \cdot \text{Log}(g^{(-1)}.x) = DR_g \cdot \text{Log}(x.g^{(-1)}) \quad \text{for } x \in \mathcal{V}_g.$$

### 3 Bi-invariant Means in Lie Groups

Lie groups are *not* vector spaces in general but have a more complicated structure: instead of a (commutative) addition ‘+’ and a scalar multiplication ‘.’,

they only have a (non-commutative in general) multiplication ‘ $\times$ ’ and an inversion operator (which corresponds to the scalar multiplication by  $-1$  for vector spaces). In the case of vector spaces, the arithmetic mean is obviously consistent with the translation and the multiplication by a scalar of input data points (equivariance rather than invariance). This means that the arithmetic mean is compatible with the algebraic structure of vector spaces. In the case of groups, the compatibility with the group structure requires the invariance with respect to left and right multiplications (the group can be non-commutative) and the inversion operator. When we translate (or inverse) a given set of samples or a probability measure, it is reasonable to desire that their mean be translated (or inverted) exactly in the same way.

Since one-parameter subgroups and their relationship with the Lie algebra are key properties in a Lie group, one could expect to define a mean using these features. For instance, the Log-Euclidean mean proposed in [7] maps the data points  $\{x_i\}$  for which the logarithm is well defined to the Lie algebra; takes the Euclidean mean with (non-negative and normalized) weights  $w_i$  and exponentiates the result:

$$\bar{x}_{LE} = \text{Exp} \left( \sum_i w_i \cdot \text{Log}(x_i) \right). \quad (3)$$

This definition is consistent with conjugation as  $h \cdot \bar{x}_{LE} \cdot h^{(-1)}$  is the log-Euclidean mean of the points  $h \cdot x_i \cdot h^{(-1)}$  thanks to Theorem 6. However, this definition fails to be invariant under left and right translation!

A well-established approach to define a notion of mean compatible with algebraic operations is to define first a *distance* (or metric) compatible with these operations and then to rely on this distance to define the mean. Indeed, one can generalize the classical notion of arithmetic mean by relying on the minimal variance or dispersion in the general setting of metric spaces [27]: the Fréchet mean (also called Riemannian center of mass) of the data points  $x_i$  with the non-negative weights  $w_i$  is the set of points:

$$\arg \min_{y \in E} \sum_i w_i \cdot \text{dist}(x_i, y)^\alpha. \quad (4)$$

The case  $\alpha = 2$  corresponds in vector spaces to the arithmetic mean, the case  $\alpha = 1$  to the median. The existence and uniqueness of these means on Riemannian manifolds has been studied first by Karcher (who relax the definition to local minima) [37] and then in [38, 44, 45, 3, 68, 69]. Thus, it seems natural to investigate if we can define a Riemannian metric compatible with the Lie group operations.

### 3.1 Bi-invariant Metrics on Lie Groups

A Riemannian metric is a smooth collection of positive definite bilinear forms on tangent spaces of the manifold. In the case of Lie groups, we can require the metric to be left-invariant (invariant by the left translation), or right-invariant. The left-invariance requires that for any two points  $g$  and  $h$  of  $\mathcal{G}$  and any vectors  $v$

and  $w$  of  $T_g\mathcal{G}$ , we have:  $\langle DL_h.v, DL_h.w \rangle_{T_{h.g}\mathcal{G}} = \langle v, w \rangle_{T_g\mathcal{G}}$ . In other words, a metric is left-invariant if all left translations are isometries. It is easy to see that all left-invariant metrics on a Lie group are determined by the inner product at the identity: they are thus in bijective correspondence with the Euclidean structures on the Lie algebra  $\mathfrak{g}$ .

The right-invariance case is obviously similar. In fact, *all* right-invariant metrics can be obtained from left-invariant metrics by ‘inversion’, and *vice versa*. This comes from the fact that for any two elements  $g, h$  of  $\mathcal{G}$ , we have  $g.h = (h^{(-1)}.g^{(-1)})^{(-1)}$ . This shows that the left-translation can be obtained smoothly from one right-translation and two inversions. Using the operator  $\text{Inv} : f \mapsto f^{(-1)}$ , we have:  $L_g = \text{Inv} \circ R_{g^{(-1)}} \circ \text{Inv}$ . We put forward this fact here to simplify the comparisons between the geodesics of the left- and right-invariant metrics in Section 4.

**Proposition 1.** *Let  $\langle, \rangle$  be a left-invariant Riemannian metric defined on  $\mathcal{G}$ . Then the ‘inverted’ metric  $\ll, \gg$*

$$\ll v, w \gg_g \stackrel{\text{def}}{=} \langle D\text{Inv}|_g.v, D\text{Inv}|_g.w \rangle_{T_{g^{(-1)}\mathcal{G}}} .$$

*is right-invariant with  $\langle, \rangle_e = \ll, \gg_e$ .*

*Proof.* Differentiating the equalities  $(h.g)^{(-1)} = g^{(-1)}.h^{(-1)}$  gives  $D\text{Inv}|_{h.g} \circ DL_h|_g = DR_{h^{(-1)}}|_{g^{(-1)}} \circ D\text{Inv}|_g$ , which shows directly that:  $\ll DR_h|_g.v, DR_h|_g.w \gg_{T_{h.g}\mathcal{G}} = \ll v, w \gg_{T_m\mathcal{G}}$ , i.e.  $\ll, \gg$  is right-invariant. The equality  $\langle, \rangle_e = \ll, \gg_e$  comes from the fact that  $D\text{Inv}|_e = -Id$ , where  $Id$  is the identity operator in  $T_e\mathcal{G}$ . This can be easily seen from the classical result  $D\text{Exp}|_e = (D\text{Log}|_0)^{(-1)} = Id$  and the equality  $g^{(-1)} = \text{Exp}(-\text{Log}(g))$  which is valid in an open neighborhood of  $e$ .

Riemannian metrics which are simultaneously left- and right-invariant are called *bi-invariant*. For these special metrics, we have the very interesting result:

**Theorem 7.** *A left-invariant metric on a Lie group is bi-invariant if and only if for all  $g \in \mathcal{G}$ , the adjoint operator  $\text{Ad}(g)$  is an isometry of the Lie algebra  $\mathfrak{g}$ , or equivalently if and only if for all elements  $x, y, z \in \mathfrak{g}$ :*

$$\langle [x, y], z \rangle + \langle y, [x, z] \rangle = 0. \quad (5)$$

*Moreover, a bi-invariant metric is also invariant w.r.t. inversion. Group geodesics of  $\mathcal{G}$  (including one-parameter subgroups) are the geodesics of such metrics.*

The proof is given in [63, Chap. V] and [58, Chap. 25]. Equation 5 is the infinitesimal version of  $\text{Ad}(g)$  being an isometry. It actually specifies that the Levi-Civita connection of the metric considered is the canonical symmetric Cartan connection of the Lie group.

An interesting consequence is that any Lie group with a bi-invariant metric has non-negative sectional curvature. Indeed, the sectional curvature in the two-plane  $\text{span}(x, y)$  for  $x, y \in \mathfrak{g}$  can be computed using left-invariant vector fields:

$$K(x, y) = \frac{\langle R(x, y)y, x \rangle}{\|x\|^2\|y\|^2 - \langle x, y \rangle^2} = \frac{1}{4} \frac{\|[x, y]\|^2}{\|x\|^2\|y\|^2 - \langle x, y \rangle^2}, \quad (6)$$

where we used the expression  $R(x, y)z = -\frac{1}{4}[[x, y], z]$  of the Riemannian curvature and Eq. 5 to move one bracket from left to right in the inner product. Thus, taking two orthonormal vectors of the Lie algebra, the section curvature reduces to  $K(x, y) = \frac{1}{4}\|[x, y]\|^2$  which is non-negative.

Since the metric inducing the notion of mean is bi-invariant, so is the mean, which is then fully compatible with the algebraic properties of the Lie group. As a consequence, this notion of mean is particularly well-adapted to Lie groups. However, contrary to left- or right-invariant metrics, which always exists, bi-invariant metrics fail to exist for some Lie groups.

**Compactness, Commutativity and Bi-invariant Metrics** From Theorem 7, we see that if a bi-invariant metric exists for the Lie group, then  $\text{Ad}(g)$  is an isometry of  $\mathfrak{g}$  and can thus be looked upon as an element of the orthogonal group  $O(n)$  where  $n = \dim(\mathfrak{G})$ . As  $O(n)$  is a *compact* group, the adjoint group  $\text{Ad}(\mathfrak{G}) = \{\text{Ad}(g)/g \in \mathfrak{G}\}$  is necessarily *included in a compact set*, a situation called *relative compactness*. This notion actually provides an excellent criterion, since the theory of differential forms and their integration can be used to explicitly construct a bi-invariant metric on relatively compact subgroups [63, Theorem V.5.3.].

**Theorem 8.** *The Lie group  $\mathfrak{G}$  admits a bi-invariant metric if and only if its adjoint group  $\text{Ad}(\mathfrak{G})$  is relatively compact.*

In the case of *compact* Lie groups, the adjoint group is the image of a compact set by a continuous mapping and is thus also compact. Thus, Theorem 8 implies that bi-invariant metrics exist in such a case. This is the case of *rotations*, for which bi-invariant Fréchet means have been extensively studied and used in practical applications, for instance in [53, 54, 49]. In the case of commutative Lie groups, left and right translations are identical and any left-invariant metric is trivially bi-invariant. Direct products of compact Abelian groups obviously admit bi-invariant metrics but Theorem 8 shows that in the general case, non-compact and non-commutative Lie groups which are not the direct product of such groups may fail to admit a bi-invariant metric.

### 3.2 There is No Bi-invariant Metric for Rigid Transformations

In biomedical imaging, the simplest possible registration procedure between two images uses rigid-body transformations which are characterized by a rotation matrix and a translation vector. Since there exists bi-invariant metrics on rotations and on translations, one could hope for the existence of bi-invariant metrics. We show below that this is not the case.

The Lie group of rigid-body transformations in the  $n$ -dimensional Euclidean space, written here  $SE(n)$ , is the semi-direct product of  $(SO(n), \times)$  (rotations)

and  $(\mathbb{R}^n, +)$  (translations). An element of  $SE(n)$  is uniquely represented by a couple  $(R, t) \in SO(n) \times \mathbb{R}^n$  with the action on a point  $x$  of  $\mathbb{R}^n$  being defined by  $(R, t).x = R.x + t$ . The multiplication is then  $(R', t').(R, t) = (R'.R, R'.t + t')$ , the neutral element  $(Id, 0)$  and the inverse  $(R^T, -R^T.t)$ . The fact that the product between rotations and translations is semi-direct and not direct (there is a coupling between rotation and translation multiplication) is at the heart of the the non-existence of a bi-invariant metric on the product group.

We obtain a faithful representation of  $SE(n)$  and its Lie algebra using homogeneous coordinates:

$$(R, t) \simeq \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad (\Omega, v) \simeq \begin{pmatrix} \Omega & v \\ 0 & 0 \end{pmatrix},$$

where  $\Omega$  is any skew  $n \times n$  matrix and  $v$  any vector of  $\mathbb{R}^n$ . In the homogeneous representation, the Lie bracket  $[\cdot, \cdot]$  is simply the matrix commutator, which gives the following Lie bracket for the Lie algebra  $\mathfrak{se}(n) = \mathfrak{so}(n) \times \mathbb{R}^n$ :  $[(\Omega, v), (\Omega', v')] = (\Omega.\Omega' - \Omega'.\Omega, \Omega.v' - \Omega'.v)$ .

**Proposition 2.** *The action of the adjoint operator  $Ad$  of the group of rigid-body transformations  $SE(n)$  at the point  $(R, t)$  on an infinitesimal displacement  $(\Omega, v) \in \mathfrak{se}(n)$  is given by:*

$$Ad(R, t).(\Omega, v) = (R.\Omega.R^T, -R.\Omega.R^T.t + R.v).$$

*As a consequence, no bi-invariant Riemannian metric exists on the space of rigid-body transformations (for  $n > 1$ , of course).*

Such a result was already known for  $SE(3)$  [71]. It is established it here for all dimensions. In the case of matrix Lie groups, we have the following formula [33]  $Ad(M).X = M.X.M^{(-1)}$  for  $X \in \mathfrak{g}$ . Using the classical faithful representation of rigid-body transformations in homogeneous coordinates, we get the announced expression of the adjoint: the translation ‘ $t$ ’ introduces a unbounded term which prevents the adjoint group from being bounded. Following Theorem 8, it is clear that no bi-invariant metric exists for rigid-body transformations in  $n$ -D ( $n > 1$ ), and *a fortiori* for affine transformations. Other examples of non-compact and non-commutative groups with no bi-invariant metrics can be found in Section 4.

Our result contradicts a statement of [66] which claimed that a bi-invariant metric exists on affine transformations when  $n = 2$ . The reference backing this claim was [60], in which it is only stated that though  $SE(2)$  is non-compact, it has a bi-invariant *measure* (Chapter 7, page 92). But whereas the existence of an invariant metric determines an invariant measure (see e.g. [55, p.131]), the converse is false.

### 3.3 A Barycentric Definition of the Mean?

Since bi-invariant Riemannian metrics can fail to exist on Lie groups, we have to rely on a different basis to define a general notion of bi-invariant means.

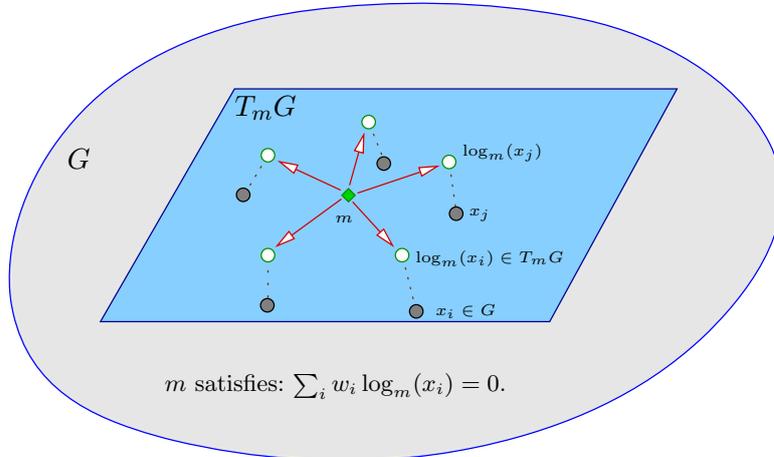
Let us recall the classical definition of a mean in an *affine* space: the mean (or *barycenter*) of a set of points  $\{x_i\}$  with (non-negative and normalized) weights  $w_i$  is the unique point  $m$  that verifies the *barycentric* equation  $\sum_i w_i \overrightarrow{mx_i} = \vec{0}$ . Since the space is 'flat', we get the closed form expression:  $m = x_1 + \sum_i w_i \cdot \overrightarrow{x_1 x_i}$ . At the mean, the sum of the weighted displacements to each of the sample points is null, i.e. the mean is at the center of the data.

In a Riemannian manifold with Riemannian metric  $\|\cdot\|_m$  at point  $m$ , the Fréchet mean of a set of points  $\{x_i\}$  with non-negative normalized weights  $\{w_i\}$  are the absolute minima of the variance (the Karcher means being the local minima):

$$\sigma^2(m) = \sum_i w_i \cdot \text{dist}(m, x_i)^2 = \sum_i w_i \cdot \|\log_m(x_i)\|_m^2,$$

where  $\log_m$  is the Riemannian logarithmic map at the point  $m$ . The existence of Karcher means is ensured when the variance is finite at one point. The uniqueness of the Fréchet/Karcher mean was investigated by Karcher, Kendall and more recently by Le, Afsari and Yang [37, 38, 45, 3, 69]. We give here a simplified result of [38] revisited by [3]:

**Theorem 9 (Local uniqueness of the karcher mean [3]).** *Let  $\kappa$  be an upper bound of sectional curvatures and  $\text{inj}(\mathcal{M})$  be the radius of injection (which can be infinite) of the Riemannian manifold. When the point set is contained in a geodesic ball  $\mathcal{B}(x, r)$  with  $r \leq r_* = \frac{1}{2} \min\{\text{inj}(\mathcal{M}), \pi/\sqrt{\kappa}\}$ , then there exists a*



**Fig. 2. Geometric Property of the Karcher mean.** When well-defined, the Karcher mean of a set of points  $\{x_i\}$  with non-negative (normalized) weights  $w_i$  satisfies a *barycentric equation*. This has a geometric interpretation: in the tangent space at the mean  $m$ , 0 (i.e.  $m$ ) is precisely the barycenter of the vectors  $\log_m(x_i)$  associated to the weights  $w_i$ . In this geometrical sense,  $m$  is at the center of the points  $\{x_i\}$ .

unique Karcher mean which lies in this ball. Moreover, it is characterized by:

$$\sum_{i=1}^N w_i \cdot \log_m(x_i) = 0. \quad (7)$$

This 'critical condition' equation can also be taken as the definition of the mean, which leads to the notion of exponential barycenter [23, 19].

### 3.4 A Fixed Point Iteration to Compute the Karcher Mean

An efficient iterative strategy to optimize in Riemannian manifolds is to use a Newton-like gradient descent algorithm. Newton algorithms on Riemannian manifolds were first proposed in the general context of optimization on manifolds [62, 65]. Their convergence has been studied in depth in [47, 32, 1] to cite just a few of the important works.

However, when the goal is to compute the Karcher mean, which is a non-linear least-squares problem, it is more efficient to use the Gauss-Newton variant which does not require the computation of the Hessian function of the Riemannian distance: this avoids implementing the connection. Although the Gauss-Newton algorithm is often presented as an approximation of the Newton method, it is more interesting to see it as the exact minimization of approximated criteria obtained at each step by linearizing the residuals [52, 2]. When applied to the variance to compute the Karcher mean, this leads to the very simple iteration which does not require the implementation of the connection:

$$m_{t+1} = \exp_{m_t} \left( \frac{1}{N} \sum_{i=1}^N w_i \cdot \log_{m_t}(x_i) \right). \quad (8)$$

This algorithm has been regularly used in the literature with varying justifications but always excellent numerical efficiency (see e.g. [53, 55] for homogeneous manifolds including  $SO(3)$  and  $SE(3)$ , [44] for shape spaces). The study of the convergence of this specific algorithm was performed in [44, 45] in the context of the Fréchet mean, while [32] investigated more generally the convergence of algorithms of the type  $m_{t+1} = \exp_{m_t}(Y(m_t))$  to the zeroes of the vector field  $Y$  on a Riemannian manifold.

**Theorem 10 (Convergence of the Gauss-Newton Iteration on Riemannian Manifolds [45]).** *Assuming that the support of the the probability is contained in a geodesic ball  $\mathcal{B}(x, r)$  with  $r \leq r_*$  as in Theorem 9. Then the iterates define at Eq. 8 starting at any point in  $\mathcal{B}(x, r)$  converges to the unique Karcher mean lying within the ball.*

In the particular case of Lie groups provided with a bi-invariant metric (but only in this case), the metric geodesics correspond to group geodesics and the group logarithm and Riemannian logarithm are the same. Equation (7) can thus be simplified to:

$$\sum_{i=1}^N w_i \cdot \text{Log}(m^{(-1)} \cdot x_i) = 0.$$

Moreover, this equation is left-, right- and inverse-invariant, since it derives from a bi-invariant metric. The corresponding Gauss-Newton iteration can be written as follows.

**Algorithm 1 (Barycentric Fixed Point Iteration on Lie Groups.)**

- 1 Initialize  $m_0$ , for example with  $m_0 := x_1$ .
- 2 Update the estimate of the mean by:

$$m_{t+1} := m_t \cdot \text{Exp} \left( \sum_i w_i \cdot \text{Log}(m_t^{(-1)} \cdot x_i) \right). \quad (9)$$

- 3 Test convergence: if  $\| \text{Log}(m_t^{(-1)} \cdot m_{t+1}) \|_e > \epsilon \cdot \sigma(m_t)$ , go to step 2.

The properties of this algorithm for general Lie groups without a bi-invariant metric will be studied in depth in Section 3.5. Notice that for such Lie groups, the group geodesics generally cannot be seen as Riemannian geodesics (as the canonical Cartan connection is non metric) so that this algorithm cannot be written  $m_{t+1} = \exp_{m_t}(Y(m_t))$  for some Riemannian metric. Thus, it does not enter into the type of iterations that are studied in [32].

It was suggested in [66] to compute empirically bi-invariant means of invertible matrices with the same algorithm, even though no bi-invariant Riemannian metrics exist for such transformations. This works well in practice, but no precise definition of bi-invariant means was given in this work. Furthermore, the existence and uniqueness of bi-invariant means was not established, and no proof of convergence of the iterative strategy was given.

A similar algorithm was proposed to compute the Karcher mean on the Lie group of similarity transformations in dimension 3 in [25, Algorithm 1]. However, the algorithm was based on a confusion between the Riemannian and the group logarithm functions when defining the Riemannian distance [25, Eq. 9]. The same confusion was made in [25, Section 3] for the definition of the principal geodesic curves. Since similarity transformations include rigid body transformations (for which we know that there does not exist a bi-invariant metric), left- or right-invariant Riemannian geodesics generally differ from group geodesics. Actually, the proposed algorithm was computing a bi-invariant mean as we will see below instead of a Karcher mean! The confusion was later corrected in [26] with the proper use of Riemannian logarithms.

### 3.5 Bi-invariant Means with Exponential Barycenters

The key idea developed in this work is the following: although bi-invariant metrics may fail to exist, the group geodesics always exists in a Lie group and one can define a bi-invariant mean implicitly as an exponential barycenter, at least locally. As will be shown in the sequel, this definition has all the desirable invariance properties, even when bi-invariant metrics do not exist. Moreover, we can show the existence and uniqueness of the bi-invariant mean provided the dispersion of the data is small enough.

**Definition 3 (Bi-invariant Means).** Let  $\{x_i\}$  be a finite set of data points belonging to the an open set  $\mathcal{V}$  such that  $\text{Log}(g^{(-1)}.x_i)$  and  $\text{Log}(x_i.g^{(-1)}) = \text{Ad}(g). \text{Log}(g^{(-1)}.x_i)$  exists for any point  $g \in \mathcal{V}$ , and  $\{w_i\}$  be associated (normalized) non-negative weights. The points  $m \in \mathcal{V}_g$  which are solutions of the following group barycentric equation (if there are some) are called bi-invariant means.

$$\sum_i w_i. \text{Log}(m^{(-1)}.x_i) = 0 \quad (10)$$

This definition is close to the Riemannian center of mass (or more specifically the Riemannian average) of [32] but uses the group logarithm instead of the Riemannian logarithm. As in [32], the definition implicitly depends on the open set  $\mathcal{V}$ . Using the convex hulls of the points as proposed in [32, 3] could be a solution to solve this issue.

**Theorem 11 (Left, Right and Inverse Invariance of Bi-invariant Means).**

*The bi-invariant means are left-, right- and inverse-invariant: if  $m$  is a mean of  $\{x_i\}$  and  $h \in G$  is any group element, then  $h.m$  is a mean of  $\{h.x_i\}$ ,  $m.h$  is a mean of the points  $\{x_i.h\}$  and  $m^{(-1)}$  is a mean of  $\{x_i^{(-1)}\}$ .*

*Proof.* If  $m$  is a mean of the points  $\{x_i\}$  and  $h \in G$  is any group element, then  $\text{Log}((h.g)^{(-1)}.h.x_i) = \text{Log}(g^{(-1)}.x_i)$  obviously exists for all points  $h.x_i$  and the point  $h.m \in \mathcal{V}_{h.g} = h.\mathcal{V}_g$  is a solution of the barycentric equation

$$\sum_i w_i. \text{Log}((h.m)^{(-1)}.h.x_i) = 0,$$

which shows that  $h.m$  is a mean of the points  $\{h.x_i\}$ . For the right-invariance, we have to apply Theorem 6:

$$\text{Ad}(m). \left( \sum_i w_i. \text{Log}(m^{(-1)}.x_i) \right) = \sum_i w_i. \text{Log}(x_i.m^{(-1)}).$$

Since  $\text{Ad}(m)$  is invertible, the usual barycentric equation, which is left-invariant, is equivalent to a right-invariant barycentric equation, and the same argument as for left-invariance can be used to show that  $m.h$  is a mean of the points  $\{x_i.h\}$ . Now, to prove the invariance with respect to inversion, note that since  $\text{Log}(x^{(-1)}) = -\text{Log}(x)$ :

$$(-1) \times \left( \sum_{i=1}^N w_i. \text{Log}(m^{(-1)}.x_i) \right) = \sum_{i=1}^N w_i. \text{Log}(x_i^{(-1)}.(m^{(-1)})^{(-1)}).$$

This shows that whenever  $m$  is a bi-invariant mean of  $\{x_i\}$ ,  $m^{(-1)}$  is a mean of  $\{x_i^{(-1)}\}$ .

Since there exists a normal convex neighborhood of each point in which the logarithmic map is well defined in any affine connection space (and in particular

in a Lie group equipped with the canonical Cartan connection, see Sec. 2.4), this definition is well posed locally. In fact we have much more: the bi-invariant mean is locally unique. To define the 'locality' we rely in the following any norm  $\|\cdot\|$  on  $\mathfrak{g}$  such that for all  $x, y$  in  $\mathfrak{g}$ , we have:  $\|[x, y]\| \leq \|x\| \cdot \|y\|$ . Since  $\text{Log}_x(y) = \text{Log}(x^{-1} \cdot y)$  is well defined for all  $x, y$  in a NCN of  $m$ , the function  $d(x, y) = \|\text{Log}(x^{-1} \cdot y)\|$  can be used to measure distances in this neighborhood.

**Theorem 12 (Local Existence and Uniqueness of the Bi-invariant Mean).**

*If the data points  $\{x_i\}$  belong to a sufficiently small normal convex neighborhood  $\mathcal{V}$  of some point  $g \in \mathcal{G}$ , then there exists a unique solution of Eq. (10) in  $\mathcal{V}$ . Moreover, the iterated fixed point strategy of Algorithm 1 converges at least at a linear rate towards this unique solution, provided the initialization is close enough to  $g$ .*

The proof will be given in Sections 3.6 and 3.7.

As in the case of the Karcher mean, there is a closed form for the bi-invariant mean of two points since this point is on the geodesic joining them.

**Proposition 3.** *Let  $x$  be in  $G$  and  $y$  be in a normal convex neighborhood of  $x$ . Then the bi-invariant mean of  $x$  and  $y$  (with weights  $1 - \alpha$  and  $\alpha$ ) is given by:*

$$m = x \cdot \text{Exp}(\alpha \text{Log}(x^{(-1)} \cdot y)) = x \cdot (x^{(-1)} \cdot y)^\alpha. \quad (11)$$

Notice that the explicit formula given by Eq. (11) is quite exceptional. In general, there will be *no* closed form for the bi-invariant mean, as soon as there are more than 2 points. However, there are some specific groups where a closed form exists for the bi-invariant mean in *all cases*, and we will detail in Section 4 some examples of this rare phenomenon.

### 3.6 Existence of the Bi-invariant Mean

Let us now turn to the proof of Theorem 12. Taking  $y_i = g^{(-1)} \cdot x_i$ , we can focus on the proof of this theorem only around the identity  $e$ . The fixed point mapping  $\Phi(m) = m \cdot \text{Exp}\left(\sum_{i=1}^N w_i \cdot \text{Log}(m^{(-1)} \cdot x_i)\right)$  of Algorithm 1 plays a central role in our approach.

**Proposition 4.** *Let  $\{w_i\}$  be a set of fixed non-negative weights. Then the mapping  $\Psi : \mathfrak{g}^{N+1} \rightarrow \mathfrak{g}$  defined by*

$$\Psi(v_1, \dots, v_N, z) = \text{Log}\left(\text{Exp}(z) \cdot \text{Exp}\left(\sum_{i=1}^N w_i \cdot \text{Log}(\text{Exp}(-z) \cdot \text{Exp}(v_i))\right)\right)$$

*is analytic near 0.*

*Proof.* This comes from the fact that  $\Psi$  is a composition of other analytic mappings: namely the *BCH* mapping defined in Subsection 2.2, the mapping  $v \mapsto -v$  and the weighted sum  $(v_1, \dots, v_N) \mapsto \sum_i w_i \cdot v_i$ . This suffices to ensure that near 0,  $\Psi$  is the sum of an absolutely converging infinite multivariate series whose variables are the  $v_1, \dots, v_N$  and  $z$ .

**Corollary 2.** *If the points  $x_i$  and  $m$  belong to a sufficiently small neighborhood of  $e$ , we have the following development:*

$$\text{Log}(\Phi(m)) = \sum_i w_i \cdot \text{Log}(x_i) + O\left(\left(\sum_{i=1}^N \|\text{Log}(x_i)\| + \|\text{Log}(m)\|\right)^2\right). \quad (12)$$

*Proof.* Successive applications of the BCH formula (Section 2.2) yield the first term of the infinite series of  $\Psi$ , which is intuitively the usual arithmetic mean obtained when all the data and  $m$  commute. The bound obtained is a direct consequence of the fact that  $\Psi$  is analytic: the order of any remaining term of the infinite series is equal or larger to two and as a consequence the other terms can be bounded by a  $O\left(\left(\sum_{i=1}^N \|\text{Log}(x_i)\| + \|\text{Log}(m)\|\right)^2\right)$ .

**Corollary 3.** *For all  $\alpha$  in  $]0, 1[$ , there exists a  $R > 0$  such that whenever  $\|\text{Log}(x_i)\| \leq \alpha.R$  and  $\|\text{Log}(m)\| \leq R$  then we also have  $\|\text{Log}(\Phi(m))\| \leq R$ .*

*Proof.* Notice in Eq. (12) that the norm of the first order term is less than or equal to  $\alpha.R$ . Since the second-order term is of  $O\left(\left(\sum_{i=1}^N \|\text{Log}(x_i)\| + \|\text{Log}(m)\|\right)^2\right)$ , there exists a constant  $C$  such that the second-order term is bounded in the following way:

$$O\left(\left(\sum_{i=1}^N \|\text{Log}(x_i)\| + \|\text{Log}(m)\|\right)^2\right) \leq C.(N.\alpha^N + 1).R^2.$$

Since  $R^2$  is a  $O(R)$ ,  $C.(N.\alpha^N + 1).R^2 \leq (1 - \alpha).R$  provided that  $R$  is sufficiently small. From this we obtain  $\|\text{Log}(\Phi(m))\| \leq \alpha.R + (1 - \alpha).R = R$ , which concludes the proof.

Corollary 3 shows that provided the  $x_i$  and  $m$  are close enough to  $e$ , we can iterate indefinitely  $\Phi$  over the successive estimates of the ‘mean’ of the  $x_i$ . This shows that the barycentric fixed point iteration of Algorithm 1 is *stable* and remains indefinitely well-defined when the data is close enough to  $e$ .

**Existence of the Bi-invariant Mean** Let  $\alpha$  be in  $]0, 1[$  and  $R > 0$  such that for all  $i$   $\|\text{Log}(x_i)\| \leq \alpha.R$  and  $\|\text{Log}(m)\| \leq R$ . Then, we know from Corollary 3 that  $\|\text{Log}(\Phi(m))\| \leq R$ . Now, let us define  $\mathcal{B}_R = \{m \in G : \|\text{Log}(m)\| \leq R\}$ . From Corollary 3, we know that  $\Phi$  defines a mapping from  $\mathcal{B}_R$  to  $\mathcal{B}_R$ . A point  $\tilde{m} \in \mathcal{B}_R$  is a solution of Eq. (10) if and only if  $\tilde{m}$  is a *fixed point* of  $\Phi$ , i.e.  $\Phi(\tilde{m}) = \tilde{m}$ .

**Theorem 13 (Brouwer’s Fixed Point Theorem [59]).** *Let  $\Psi : B^n \rightarrow B^n$  be a continuous mapping, where  $B^n$  is the  $n$ -dimensional Euclidean closed ball, i.e.  $B^n = \{x \in \mathbb{R}^n : \sum_i (x_i)^2 \leq 1\}$ . Then  $\Psi$  has at least one fixed point.*

**Corollary 4.** *With the assumptions made at the beginning of this subsection, Eq. (10) has at least one solution in  $\mathcal{B}_R$ .*

*Proof.* Let us define  $\Psi : \text{Log}(\mathcal{B}_R) \rightarrow \text{Log}(\mathcal{B}_R)$  by  $\Psi(v) = \text{Log}(\Phi(\text{Exp}(v)))$ . Since  $\text{Log}(\mathcal{B}_R)$  is precisely a closed ball, and thus homeomorphic to the Euclidean closed ball, then Brouwer's theorem applies and guarantees the existence of at least one fixed point of  $\Psi$ , which is also a fixed point of  $\Phi$  and therefore a solution of Eq. (10).

### 3.7 The Barycentric Fixed Point Iteration is Converging

In order to prove the convergence of the iterative strategy to a fixed point of  $\Phi$ , we need a more powerful fixed point theorem.

**Theorem 14 (Banach Fixed Point Theorem [9]).** *Let  $(E, d)$  be a complete metric space and  $f : E \rightarrow E$  be a  $K$ -contraction, i.e. for all  $x, y$  of  $E$ ,  $d(f(x), f(y)) \leq K.d(x, y)$ , with  $0 < K < 1$ . Then  $f$  has a unique fixed point  $p$  in  $E$  and for all sequence  $(x_n)_{n>0}$  verifying  $x_{n+1} = f(x_n)$ , then  $x_n \rightarrow p$  when  $n \rightarrow +\infty$ , with at least a  $K$ -linear speed of convergence.*

Here,  $(\mathcal{B}_R, d)$  is the complete metric space in which the successive evaluations of the 'mean' live. The distance  $d$  is simply given by  $d(m, n) = \|\text{Log}(m) - \text{Log}(n)\|$ . To obtain the existence, uniqueness of a solution of Eq. (10) and linear convergence of our iterative scheme to this point, it only remains to show that  $\Phi$  is a contraction. This leads to the following Proposition:

**Proposition 5.** *When the  $R$  in Corollary 3 is chosen small enough,  $\Phi$  is a contraction.*

*Proof.* Let us consider  $E = \text{Log}(\mathcal{B}_R)$  with  $\Theta : E \rightarrow E$  defined as in the proof of Corollary 4 by  $\Theta(v) = \text{Log}(\Phi(\text{Exp}(v)))$ . The key idea is to see that  $\Theta$  is smooth with respect to  $\text{Log}(m)$  and the  $(\text{Log}(x_i))$ , with the property that the norm of the differential of  $\Theta$  is uniformly bounded in the following way:

$$\|D_{\text{Log}(m)}\Theta\| \leq O(\|\text{Log}(m)\| + \sum_i \|\text{Log}(x_i)\|). \quad (13)$$

In fact, Eq. (13) is a simple consequence of the fact that  $\Psi$  is analytic: the partial derivative  $D_{\text{Log}(m)}\Theta$  is therefore also analytic. Its value at 0 is precisely 0, and therefore all the terms of its infinite series are of order one or larger, which yields the bound in  $O(\|\text{Log}(m)\| + \sum_i \|\text{Log}(x_i)\|)$ .

With the bound given by Eq. (13), we can ensure that when  $R$  is small enough, there exists  $\beta$  in  $]0, 1[$  such that  $\|D_{\text{Log}(m)}\Theta\| \leq \beta$  for all  $m$ . Then we have the classical bound:

$$\|\Theta(v) - \Theta(w)\| \leq (\sup_{z \in E} \|D_z\Theta\|) \|v - w\| \leq \beta \|v - w\|.$$

Since  $\beta < 1$ ,  $\Theta$  is by definition a contraction, and so is  $\Phi$ .

**Corollary 5.** *As a consequence, when the data  $\{x_i\}$  are given close enough to  $e$ , there exists an open neighborhood of  $e$  in which there exists a unique solution to Eq. (10). Moreover, the iterative strategy given above always converges towards this solution, provided that the initialization to this algorithm is chosen sufficiently close to the data (so that Corollary 2 can apply). Last but not least, the speed of convergence is at least linear.*

*Proof.* Simply apply the Banach fixed point theorem to  $\Phi$  and recall that being a fixed point of  $\Phi$  is equivalent to being a solution of Eq. (10).

We have rigorously generalized to any real Lie group the notion of bi-invariant mean normally associated to bi-invariant Riemannian metrics, even in the case where such metrics fail to exist. This novel mean has all the desirable invariance properties, and can be iteratively computed in a very efficient way. From a theoretical point of view, the methods used in this Section, based on contraction theorems to show the convergence of an iteration strategy, are quite close to the ones used in [45, 32] to analyze the existence and uniqueness of the Fréchet/Karcher mean in Riemannian manifolds. However, the use of the canonical Cartan connection brings additional properties since we can treat the problem in the Lie algebra only. It would be interesting to carefully compare the proofs to better understand what is specific to the use of a Riemannian metric, what is due to the use of the canonical Cartan connection, and what is generic for affine connection spaces.

As usual with means in manifolds, the bi-invariant mean is unique only if the data are close enough to one another: the dispersion should not be too large. In the next section, we will see more precisely in various situations which practical limitation is imposed on the dispersion of the data. One does not seem to lose much in this regard with respect to existing Riemannian bi-invariant means: we will show for example that the bi-invariant mean of rigid-body transformations exists if and only if the bi-invariant mean of their rotation parts exists.

## 4 Bi-invariant Means in Selected Matrix Lie Groups

Let us now detail several insightful cases where the algebraic mean can be *explicitly* or directly computed, without using an iterative scheme.

### 4.1 Scalings and Translations in $n$ -D

Here, we will devote some time to a very instructive group: *the group of scalings and translations in  $n$ -D*. The study of this (quite) simple group is relevant in the context of this work, because it is one of the most simple cases of non-compact and non-commutative Lie groups which does not possess any bi-invariant Riemannian metric. This group has many of the properties of rigid-body or affine transformations, but with only  $n + 1$  degrees of freedom, which simplifies greatly the computations, and allows a direct 2D geometric visualization in the plane

for  $n = 1$ . For these reasons, this is a highly pedagogical case. In the rest of this Subsection, we will let this group be written  $ST(n)$ .

Following the notations of Section 3.2, an element  $g$  of  $ST(n)$  can be uniquely represented by a couple  $(\lambda, t)$  in  $\mathbb{R}_+^* \times \mathbb{R}^n$  where  $\lambda$  corresponds to the scaling factor and  $t$  to the translation part. We define the action of  $(\lambda, t) \in ST(n)$  on a vector of  $x \in \mathbb{R}^n$  by:  $(\lambda, t).x = \lambda.x + t$ . Accordingly, the composition in  $ST(n)$  is:  $(\lambda', t').(\lambda, t) = (\lambda'.\lambda, \lambda'.t + t')$  and inversion is  $(\lambda, t)^{(-1)} = (1/\lambda, -t/\lambda)$ . The commutator is  $[(\mu, t), (\mu', t')] = (0, \mu.v' - \mu'.v)$ . This example shows that the semi-direct product of two commutative groups may lead to a non-commutative one.

The elements of the Lie algebra are of the form  $(\mu, v)$ , where  $\mu \in \mathbb{R}$  and  $v \in \mathbb{R}^n$  and the group exponential  $\text{Exp}(\mu, v)$  has the form

$$\text{Exp}(\mu, v) = (e^\mu, v.(e^\mu - 1)/\mu),$$

where  $e^\lambda$  is the scalar exponential of  $\lambda$ . The Taylor expansion  $(e^x - 1)/x = 1 + x/2 + O(x^2)$  shows that the formula is well posed for  $\mu$  close or equal to zero. Thus, we see that the group exponential is simply given by the scalar exponential on the scaling part, whereas the translation part mixes the multiplicative and additive influences of both components. Moreover, we see geometrically that in the upper half space  $\mathbb{R}_+ \times \mathbb{R}^n$ , the curve given by  $\text{Exp}(s.(\mu, v))$  with  $s$  varying in  $\mathbb{R}$  is on a *straight line*, whose equation is  $t = v.(\lambda - 1)/\mu$  ( $\mu$  and  $v$  are parameters).

The entire space  $ST(n)$  is a normal convex neighborhood: any two points can be joined by a unique group geodesic. In particular, the group logarithm is always well-defined and given by:

$$\text{Log}(\lambda, t) = \left( \ln(\lambda), t. \frac{\ln(\lambda)}{1 - \lambda} \right)$$

where  $\ln(\lambda)$  is the natural (scalar) logarithm of  $\lambda$ . The Taylor expansion  $\ln(\lambda)/(\lambda - 1) = 1 + (1 - \lambda)/2 + O((1 - \lambda)^2)$  shows that the above formula and the following ones are numerically stable around  $\lambda = 1$ . As for the exponential, we get the classical logarithm on the scaling part and a mixture of the multiplicative and additive logarithms on the translation part. We recall that in the case of an additive group such as  $(\mathbb{R}^n, +)$ , both additive exponential and logarithm are simply the identity, which is also what we get here when there is no scaling.

The unique group geodesic joining  $(\lambda, t)$  and  $(\lambda', t')$  has the form  $\gamma(s) = (\lambda, t). \text{Exp}(s.(\mu, v))$  with  $s$  in  $[0, 1]$ , where the parameters  $(\mu, v)$  are given by:

$$(\mu, v) = \left( \ln \left( \frac{\lambda'}{\lambda} \right), \left( \frac{t' - t}{\lambda} \right) \cdot \left( \frac{\ln(\lambda'/\lambda)}{\lambda'/\lambda - 1} \right) \right). \quad (14)$$

**Absence of Bi-invariant Metrics.**  $ST(n)$  is one of the most simple non-compact and non-commutative Lie groups. As expected for such Lie groups, it has no bi-invariant metric. To show this, we have to analyze the boundedness

of the adjoint group of  $ST(n)$ . The group can be faithfully represented by the subgroup of triangular matrices of the form  $\begin{pmatrix} \lambda Id_n & t \\ 0 & 1 \end{pmatrix}$ . The adjoint is:

$$\text{Ad}((\lambda, t)).(\mu, v) \sim \begin{pmatrix} \lambda Id_n & t \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu Id_n & v \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\lambda} Id_n & -\frac{t}{\lambda} \\ 0 & 1 \end{pmatrix} = (\mu, \lambda.v - t.\mu).$$

Both factors ‘ $t$ ’ and ‘ $\lambda$ ’ in  $\lambda.v - t.\mu$  are not bounded and thus  $\text{Ad}(ST(n))$  cannot be bounded. As a consequence,  $ST(n)$  has no bi-invariant metric. Both  $(R_+^*, \times)$  and  $(\mathbb{R}^n, +)$  are commutative and thus have bi-invariant metrics, but their semi-direct product has no such metric.

**A Closed Form for the Bi-invariant Mean.** Here, since we have explicit formulas for the group exponential and logarithm, one can use these formulas to try to solve directly the barycentric Equation (10).

**Proposition 6.** *Let  $\{(\lambda_i, t_i)\}$  be a set of points in  $ST(n)$  and  $\{w_i\}$  be associated non-negative (normalized) weights. Then the bi-invariant mean  $(\bar{\lambda}, \bar{t})$  is given explicitly by:*

$$\bar{\lambda} = \exp\left(\sum_i w_i \cdot \ln(\lambda_i)\right), \text{ (weighted geometric mean of scalings),}$$

$$\bar{t} = \frac{1}{Z} \sum_i w_i \cdot \alpha_i \cdot t_i, \text{ (scalings reweighed arithmetic mean of translations),}$$

with  $\alpha_i = \frac{\ln(\lambda_i/\bar{\lambda})}{\lambda_i/\bar{\lambda} - 1} = 1 + \frac{1}{2} (1 - \lambda_i/\bar{\lambda}) + O\left((1 - \lambda_i/\bar{\lambda})^2\right)$  and  $Z = \sum_i w_i \cdot \alpha_i$ .

*Proof.* Just replace in the barycentric equation the exponentials and logarithms by the formulas given above. Since the scaling component is independent from the translation one, we simply obtain the geometric mean, which is the bi-invariant mean for positive numbers. The translation part can be handled simply by using directly Eq. (14), which yields this simplified expression for the barycentric equation:

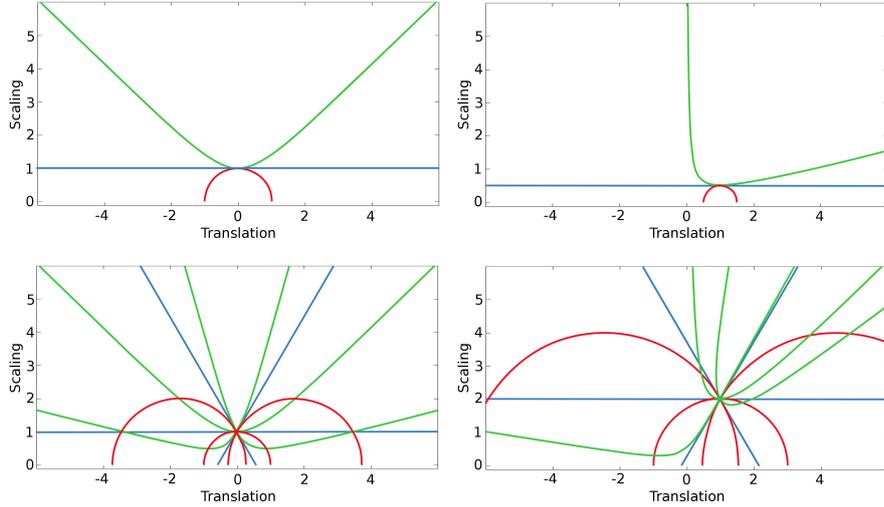
$$\sum_i w_i \left( \frac{t_i - \bar{t}}{\bar{\lambda}} \right) \cdot \left( \frac{\ln(\lambda_i/\bar{\lambda})}{\lambda_i/\bar{\lambda} - 1} \right) = 0.$$

**Comparison Between Group and Metric Geodesics.** In Figure 3, one can visually compare the group geodesics to some of their left-invariant and right-invariant (metric) counterparts for the group  $ST(1)$ . Interestingly, one of the left-invariant metrics induces an isometry between this group and the Poincaré half-plane model for hyperbolic geometry. The scalar product of this scalar metric is the most simple at the point  $(1, 0)$ : it is the usual Euclidean scalar product. Geodesics take a very particular form in this case: they are the set of all the *semicircles perpendicular to the axis of translations* and of all (truncated below

the axis of translations) lines perpendicular to the axis of translations (these lines can be seen as semicircles of infinite diameter anyway).

Thanks to Proposition 1, we know that the right-invariant Riemannian metric whose scalar product at  $(1, 0)$  is the same as the previous metric can be obtained simply by ‘inverting’ this left-invariant metric. As a consequence, its geodesics can be computed simply by inverting the initial conditions, computing the associated left-invariant geodesic and finally inverting it. As a consequence, the geodesics of the right-invariant metric visualized in Fig. 3 are somehow ‘inverted semicircles’. In fact, simple algebraic computations show that these geodesics are all *half hyperbolas*.

The simple form taken by left-invariant geodesics is exceptional. In general, there are no closed form solutions for neither the left- nor the right-invariant geodesics, and group geodesics are simpler to compute, since in most practical cases they only involve the computation of a matrix exponential and a matrix logarithm, for which very efficient methods exist [35, 18].



**Fig. 3. Examples of geodesics in the group of scalings and translations in 1D.** In each of the sub-figures, we plot the left-invariant geodesic (red), the right-invariant geodesic (green) and the group geodesic (blue) starting from the same point with the same tangent vector. **Top row:** two examples of one left-invariant, one right-invariant and one group geodesic starting from two different points. **Bottom row:** two examples of left-invariant, right-invariant and group geodesics with three different initial tangent vectors. Note the particular form taken by group geodesics, which are parts of straight lines and of the left-invariant geodesics, which are semicircles perpendicular to the horizontal axis. Right-invariant geodesics are also given in a closed form and are in fact half hyperbolas.

## 4.2 The Heisenberg Group

This is the group of 3D upper triangular matrices  $M$  of the form:

$$M = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}.$$

To simplify notations, we will write an element of this group  $(x, y, z)$ . The multiplication is  $(x_1, y_1, z_1) \cdot (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + x_1 \cdot y_2)$ . Thus the first two parameters live in a 2D additive group which is independent of the third parameter, whereas the third additive parameter is influenced by the first two. The Heisenberg group is thus a semi-direct product of  $(\mathbb{R}^2, +)$  and  $(\mathbb{R}, +)$ , which is not commutative. The inversion is  $(x, y, z)^{(-1)} = (-x, -y, -z + x \cdot y)$  with neutral element  $(0, 0, 0)$ .

As in the  $ST(n)$  case, the entire Heisenberg group is a normal convex neighborhood and we have:

$$\begin{cases} \text{Exp}(u, v, w) = (u, v, w + \frac{1}{2} \cdot u \cdot v), \\ \text{Log}(x, y, z) = (x, y, z - \frac{1}{2} x \cdot y). \end{cases}$$

The unique group geodesic joining  $(x, y, z)$  and  $(x', y', z')$  is of the form  $\gamma(s) = (x, y, z) \cdot \text{Exp}(s \cdot (u, v, w))$ ,  $s \in [0, 1]$  where the parameters  $(u, v, w)$  are given by:

$$(u, v, w) = \left( x' - x, y' - y, z' - z + \frac{1}{2} \cdot (x \cdot y - x' \cdot y' + x \cdot y' - x' \cdot y) \right). \quad (15)$$

**Bi-invariant Metrics and Bi-invariant Means.** As in the  $ST(n)$  case, no bi-invariant metric exists and one has the closed form for the bi-invariant mean. Interestingly, the bi-invariant mean yields a simple arithmetic averaging of the first two parameters. The third parameter is also averaged arithmetically, except that this arithmetic mean is ‘corrected’ by a quadratic function of the first two parameters of the data.

**Proposition 7.** *The action of the adjoint operator  $Ad$  of the Heisenberg group at a point  $(x, y, z)$  on an infinitesimal displacement  $(u, v, w)$  is given by:*

$$Ad(x, y, z) \cdot (u, v, w) = (u, v, -y u + x v + w).$$

*As a consequence, no bi-invariance metric exists for the Heisenberg group.*

*Proof.* The last coordinate is unbounded, which prevents compactness.

**Proposition 8.** *Let  $\{(x_i, y_i, z_i)\}$  be a finite set of points in the Heisenberg group and  $\{w_i\}$  be associated non-negative (normalized) weights. Then the bi-invariant mean  $(\bar{x}, \bar{y}, \bar{z})$  is given explicitly by:*

$$(\bar{x}, \bar{y}, \bar{z}) = \left( \sum_i w_i x_i, \sum_i w_i y_i, \sum_i w_i z_i + \frac{1}{2} \left( (\sum_i w_i x_i) \cdot (\sum_i w_i y_i) - \sum_i w_i x_i \cdot y_i \right) \right).$$

*Proof.* Just replace in the barycentric equation the exponentials and logarithms by the formulas given above. Since the first two components are additive and independent from the third one, their bi-invariant mean is simply their arithmetic mean. The third coefficient case can be handled simply using directly Eq. (15), which yields this simplified expression for the barycentric equation:

$$\sum_i w_i \left( z_i - \bar{z} + \frac{1}{2} \cdot (\bar{x} \cdot \bar{y} - x_i \cdot y_i + \bar{x} \cdot y_i - x_i \cdot \bar{y}) \right) = 0.$$

### 4.3 Scaled Upper Unitriangular Matrix Group

We can generalize the results obtained on the Heisenberg group to the following subgroup of triangular matrices:

**Definition 4.** Let  $UT(n)$  be the group of  $n \times n$  scaled upper unitriangular matrices (upper triangular with scaled unit diagonal). Such matrices have the form:

$$M = \lambda \cdot Id + N,$$

where  $\lambda$  is any positive scalar,  $Id$  the identity matrix and  $N$  an upper triangular nilpotent matrix ( $N^n = 0$ ) with only zeros in its diagonal.

The Heisenberg group is the subgroup of matrices of  $UT(3)$  whose  $\lambda$  is always equal to 1. The situation in this case is particularly nice, since thanks to the fact that  $N$  is nilpotent, one can perform *exactly* all the usual algebraic operations in  $UT(n)$ . The group multiplication and inversion are:

$$M' \cdot M = (\lambda' \cdot Id + N') \cdot (\lambda \cdot Id + N) = (\lambda' \cdot \lambda) \cdot Id + (\lambda' \cdot N + \lambda \cdot N' + N' \cdot N).$$

$$M^{(-1)} = (\lambda \cdot Id + N)^{(-1)} = \lambda^{(-1)} \cdot \left( Id + \frac{N}{\lambda} \right)^{(-1)} = \lambda^{(-1)} \cdot \sum_{k=0}^{n-1} (-1)^k \cdot \frac{N^k}{\lambda^k}.$$

The group exponential and logarithm are:

$$\text{Exp}(X) = \exp(\mu \cdot Id + Y) = \exp(\mu \cdot Id) \cdot \exp(Y) = e^\mu \cdot \sum_{k=0}^{n-1} \frac{(Y)^k}{k!}.$$

$$\text{Log}(M) = \text{Log} \left( (\lambda \cdot Id) \cdot \left( Id + \frac{1}{\lambda} \cdot N \right) \right) = \ln(\lambda) \cdot Id + \sum_{k=1}^{n-1} \frac{(-1)^{k+1}}{k} \cdot \frac{N^k}{\lambda^k}.$$

Using these closed forms, one can derive the following equation:

$$\text{Log}(M' \cdot M) = \ln(\lambda' \cdot \lambda) \cdot Id + \sum_{k=1}^{n-1} \frac{(-1)^{k+1}}{k} \cdot \left( \frac{1}{\lambda} \cdot N + \frac{1}{\lambda'} \cdot N' + \frac{1}{\lambda' \cdot \lambda} \cdot N \cdot N' \right)^k,$$

which in turn allows to compute the equation satisfied by the bi-invariant mean  $\bar{M} = \bar{\lambda} \cdot Id + \bar{N}$  in  $UT(n)$ :

$$\begin{aligned} 0 &= \sum_i w_i \text{Log}(\bar{M} \cdot M_i^{(-1)}) = - \sum_i w_i \text{Log}(\bar{M}^{(-1)} \cdot M_i) \\ 0 &= \sum_i w_i \left( \ln(\bar{\lambda} \cdot \lambda_i^{(-1)}) \cdot Id \right. \\ &\quad \left. + \sum_i w_i \left( \sum_{k=1}^{n-1} \frac{(-1)^{k+1}}{k} \cdot \left( \frac{1}{\bar{\lambda} \cdot \lambda_i^{(-1)}} \cdot N_i^{(-1)} + \frac{1}{\bar{\lambda}} \cdot \bar{N} + \frac{1}{\bar{\lambda} \cdot \lambda_i^{(-1)}} \cdot N_i^{(-1)} \cdot \bar{N} \right)^k \right) \right), \end{aligned} \quad (16)$$

where  $N_i^{(-1)}$  is the nilpotent part of  $M_i^{(-1)}$ . From Eq. (16), we see that  $\bar{\lambda}$  is simply the geometric mean of the  $\lambda_i$ , and that the coefficient of  $\bar{N}$  can be recursively computed, starting from coefficients above the diagonal. The key idea is that the  $k^{\text{th}}$  power of a nilpotent matrix  $N$  will have non-zero coefficients only in its  $k^{\text{th}}$  upper diagonal.

As a consequence, to compute the coefficients of  $\bar{M}$  above the diagonal, one only needs to take into account the following terms:  $N_i^{(-1)}/\lambda_i^{(-1)} + \bar{N}/\bar{\lambda}$ . These coefficients will simply be a weighted arithmetic mean of the coefficients in the data, the weights being equal to  $(w_i \cdot \lambda_i / \bar{\lambda}) / S$  with  $S = \sum_j w_j \cdot \lambda_j / \bar{\lambda}$ . Using this result, then one can compute the coefficients above, which are a weighted arithmetic mean of the corresponding coefficients in the data, with a quadratic correction involving the previous coefficients. The same phenomenon appears for the next set of coefficients above, with an even more complex correction involving all the previously computed coefficients. One can continue this way until all the coefficients of the mean have been effectively computed.

#### 4.4 General Rigid-Body Transformations

We use in this section the notations previously introduced in Section 3.2. The group exponential can be computed using directly the matrix representation, or by identifying the one-parameter subgroups of  $SE(n)$ . It is given by:

$$\text{Exp}(\Omega, v) = \left( e^{\Omega}, e^{\Omega} \cdot \left( \int_0^1 e^{-u \cdot \Omega} du \right) \cdot v \right),$$

where  $e^{\Omega}$  is the matrix exponential of  $\Omega$ .

**Decomposition of  $n$ -D Rotations into 2-D Rotations** From classical linear algebra, we know that the spectral decomposition of a rotation matrix  $R$  has a very special form. Indeed, the characteristic polynomial  $P(\lambda) = \det(R - \lambda \cdot Id_n)$  is a real polynomial of degree  $n$ . Thus the  $n$  complex eigenvalues are actually conjugate by pairs (or real) and the polynomial can be factored into at most  $\lfloor n/2 \rfloor$  quadratic terms (potentially with multiplicity) and real linear terms. The conservation of the norm by the rotation ( $\|R \cdot x\| = \|x\|$ ) shows that the modulus of all the eigenvalues is 1. Thus, eigenvalues are  $e^{\pm i\theta_j}$  or 1. Since a rotation is a normal matrix, it can be diagonalized and we conclude that every rotation matrix, when expressed in a suitable coordinate system, partitions into  $\lfloor n/2 \rfloor$  independent 2D rotations [36]:

$$\mathcal{R}(\theta_j) = \begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{pmatrix} = \exp \left( \theta_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right).$$

Conversely, each skew symmetric matrix  $\Omega$  decomposes the space  $\mathbb{R}^n$  in a direct sum of *mutually orthogonal* subspaces, which are all invariant under  $\Omega$  [36]. The decomposition is of this form:

- $k$  (possibly equal to zero) 2-dimensional vector subspaces  $E_j$  on which  $\Omega$  acts non trivially.
- A single subspace  $F$  of dimension  $n - 2k$  (the orthogonal complement of the span of other subspaces), which is the kernel of  $\Omega$ .

For any  $E_j$ , there exists an orthonormal basis of  $E_j$  such that  $\Omega$  restricted to  $E_j$  is in this basis of the following matrix form:  $\theta_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  where  $\theta_j$  ( $\neq 0$ ) is the  $j^{\text{th}}$  angle of rotation of the  $n$ -dimensional rotation  $e^\Omega$ .

**Existence of the Logarithm.** Since we have a faithful representation of  $SE(n)$  in terms of matrices, we can use the matrix criterion for the existence of the principal logarithm: from Subsection 2.2, we know that an invertible matrix with no (complex) eigenvalue on the closed half-line of negative real numbers has unique matrix logarithm with eigenvalues having imaginary parts in  $] -\pi, \pi[$ . This means that the angles of the previous  $\lfloor n/2 \rfloor$  2D rotations decomposing the rotation  $R$  should not go outside  $] -\pi, \pi[$  if we want the logarithm of  $R$  to be well-defined. Otherwise, one cannot define a unique logarithm. This is only the case for 2D rotations of 180 degrees, whose two ‘smallest’ real logarithms are the following:

$$\begin{pmatrix} 0 & -\pi \\ \pi & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & \pi \\ -\pi & 0 \end{pmatrix}.$$

Going back to  $SE(n)$ , we have the following result:

**Proposition 9.** *The logarithm of a rigid-body transformation  $(R, t)$  is well-defined if and only if the logarithm of its rotation part  $R$  is well-defined, i.e. if the angles of the 2D rotations of its decomposition are less than  $\pi$  in absolute value.*

*Proof.* The logarithm of  $(R, t)$  is well-defined if and only if the matrix representing  $(R, t)$  has a principal logarithm, which is equivalent to the fact that it has no eigenvalue on the closed negative line. Then, this is equivalent to the fact that  $R$  has no eigenvalue on the closed negative line, since the eigenvalues of the upper triangular matrix (in terms of blocks)  $\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}$  depend only on the blocks in its diagonal, i.e. only on  $R$ , and *not*  $t$ . As a consequence, the logarithm of a rigid-body transformation is well-defined if and only if the logarithm of its rotation part is well-defined.

**Criterion for the Existence of the Bi-invariant Mean.** We have seen in Section 3.2 that no bi-invariant metric exists in the rigid-body case. One may now ask the question: is there a simple criterion for the existence/uniqueness of the bi-invariant mean of rigid-body transformations? When bi-invariant metrics exist, one has the criterion given by Theorem 9: the mean exists and is unique as long as the data are include in a geodesic ball of radius strictly less than  $r_* = \frac{1}{2} \min\{\text{inj}(\mathcal{M}), \pi/\sqrt{\kappa}\}$ .

Let us investigate first the  $n$ -D rotations part with the bi-invariant Froebenius metric  $\langle X, Y \rangle = \frac{1}{2}\text{Tr}(XY^T)$ , which lead to the Riemannian (squared) distance  $\text{dist}(Id_n, \exp(\Omega))^2 = \frac{1}{2}\text{Tr}(\Omega\Omega^T)$  for any skew-symmetric matrix  $\Omega \in \mathfrak{se}(n)$ . Since any rotation can be decomposed into independent 2D rotations and since the trace does not depend on the basis in which we express our rotation, the Riemannian (squared) distance of a rotation  $R$  to the identity is the sum the squares of the 2D angles:  $\text{dist}(R, Id)^2 = \sum_i \theta_i^2$ . The minimal distance to the cut locus is obtained for each 2D rotation independently, so that we have  $\text{inj}(Id) = \text{inj}(SO(n)) = \pi$ .

The sectional curvature in the two-plane spanned by two orthonormal vectors  $x, y \in \mathfrak{g}$  is  $K(x, y) = \frac{1}{4}\| [x, y] \|^2$  according to Eq. 6. We can get the sectional curvature everywhere else by left or right translation. Let  $E_{ij} = e_i e_j^T - e_j e_i^T$  be the matrix with 1 for the coefficient  $ij$ ,  $-1$  for the coefficient  $ji$  and 0 everywhere else. One verifies that the  $n(n-1)/2$  vectors  $E_{ij}$  with  $j > i$  form an orthonormal basis of the Lie algebra of skew-symmetric matrices. Now, the commutator of these basis vectors

$$[E_{ij}, E_{kl}] = \delta_{jk}E_{il} + \delta_{il}E_{jk} + \delta_{jl}E_{ki} + \delta_{ki}E_{lj},$$

is non-zero if one of the indices  $i, j$  is equal to one of the indices  $k, l$ . In the case  $j = l$  for instance then we get  $[E_{ij}, E_{kj}] = E_{ki}$  and thus the sectional curvature is  $K(E_{ij}, E_{kj}) = 1/4$ . We obtain the same result for the other cases where the commutator is non-zero. As a conclusion, the sectional curvature is bounded above by  $\kappa = 1/4$ , and the maximal radius is  $r_* = \frac{1}{2} \min\{\pi, 2\pi\} = \pi/2$ .

Thus, if all the data are included in a geodesic ball of radius  $r < r_* = \pi/2$ , then the largest 2D angle of rotation of  $R_j^{(-1)}.R_i$  (and thus all of them) is less than  $2r < \pi$  for any couple of data  $R_i$  and  $R_j$ . In consequence, the principal logarithm of  $R_j^{(-1)}.R_i$  is well-defined. Thus, in the general conditions of existence and uniqueness of the Riemannian mean of rotations, the rotation part of the bi-invariant mean is well-defined. Remarkably, this is sufficient to guarantee the existence and uniqueness of the bi-invariant mean of rigid-body transformations.

**Theorem 15.** *Let  $\{R_i, t_i\}$  be a set of rigid-body transformations belonging to a geodesic ball of radius  $r < \pi/2$ . Then (according to Theorem 9) the bi-invariant Riemannian mean of their rotation parts is well-defined and there exists a unique bi-invariant mean for on  $SE(n)$ .*

*Proof.* Let  $\bar{R}$  be the bi-invariant Riemannian mean of the rotation parts of the data. The bi-invariant mean of the data is necessarily of the form  $(\bar{R}, \bar{t})$ , since in an open neighborhood of the rotation parts,  $\bar{R}$  is the only solution of the rotation part of the bi-invariant barycentric equation, which does not depend on translations.  $\bar{R}$  is included in the same geodesic ball as the rotations and therefore, for any  $R_i$  of the data, the 2D angles of rotation of  $\bar{R}^{(-1)}.R_i$  are all smaller than or equal to  $\pi - C$ , where  $C$  is a positive constant smaller than  $\pi$ .

Let us now check whether there exists a unique translation  $\bar{t}$ , which satisfies the barycentric equation of bi-invariant means, which writes here:

$$\sum_i w_i \cdot \text{Log}((\bar{R}, \bar{t}).(R_i, t_i)^{(-1)}) = 0. \quad (17)$$

From Proposition 9, we know that the logarithm of  $(\bar{R}, t) \cdot (R_i, t_i)^{(-1)}$  is well-defined for *any value of  $t$* , since the logarithm of  $\bar{R} \cdot R_i^T$  is well-defined for all  $i$ . Now, does there exist a unique value of  $t$  (by definition  $\bar{t}$ ) satisfying Eq. (17)?

We have:  $(\bar{R}, t) \cdot (R_i, t_i)^{(-1)} = (\bar{R} \cdot R_i^T, \bar{R} \cdot (-R_i^T \cdot t_i) + t)$ . Let us write  $M(\Omega) = e^\Omega \cdot \int_0^1 e^{-u \cdot \Omega} du$ . In terms of translations, Eq. (17) writes:

$$\begin{aligned} \sum_i w_i \cdot M(\log(\bar{R} \cdot R_i^T))^{(-1)} \cdot (\bar{R} \cdot (-R_i^T \cdot t_i) + \bar{t}) &= 0. \\ \iff \\ \left( \sum_i w_i \cdot M(\log(\bar{R} \cdot R_i^T))^{(-1)} \right) \cdot \bar{t} &= \sum_i w_i \cdot M(\log(\bar{R} \cdot R_i^T))^{(-1)} \cdot R_i^T \cdot t_i. \end{aligned} \quad (18)$$

Thus, we see that the existence and uniqueness of  $\bar{t}$  resorts to the invertibility of the matrix  $\sum_i w_i \cdot M(\log(\bar{R} \cdot R_i^T))^{(-1)}$ . Under the assumptions described above on rotations, this matrix is invertible by Lemma 1 below, which concludes the proof.

**Lemma 1.** *Let  $\{\Omega_i\}$  be a set of skew symmetric matrices such that the norm of their largest (complex) eigenvalue is smaller than  $\pi - C$ , with  $C > 0$ . Let  $M(\Omega)$  be equal to  $e^\Omega \cdot \int_0^1 e^{-u \cdot \Omega} du$  for any skew symmetric matrix. Then for all  $\Omega_i$ ,  $M(\Omega_i)$  is invertible, and for any non-negative weights  $\{w_i\}$ ,  $\sum_i w_i \cdot M(\Omega_i)^{(-1)}$  is also invertible.*

*Proof.* In an appropriate orthonormal basis, the skew symmetric matrix  $\Omega$  can be decomposed into a zero matrix in the kernel of  $\Omega$  and  $k$  (possibly equal to zero) 2-dimensional matrices of the form  $\theta_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ , in the mutually orthogonal subspaces  $E_j$ , where  $\theta_j$  ( $\neq 0$ ) is the  $j^{\text{th}}$  angle of rotation of the  $n$ -dimensional rotation  $e^\Omega$ .

We can explicitly compute  $M(\Omega)$  in the above subspaces. First, in the kernel  $F$  of  $\Omega$ ,  $M(\Omega)$  is simply the identity. In the subspace  $E_j$ , we have:

$$\exp(\Omega)|_{E_j} \sim \begin{pmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{pmatrix}.$$

A few extra manipulations yield:

$$M(\Omega)|_{E_j} = \left( \exp(\Omega) \cdot \int_0^1 \exp(-u \cdot \Omega) du \right) |_{E_j} \sim \begin{pmatrix} \frac{\sin(\theta_j)}{\theta_j} & \frac{\cos(\theta_j) - 1}{\theta_j} \\ -\frac{\cos(\theta_j) - 1}{\theta_j} & \frac{\sin(\theta_j)}{\theta_j} \end{pmatrix}.$$

Thus  $M(\Omega)$  is invertible whenever  $0 < |\theta_j| < 2\pi$  for all  $j$  (which is more than we need), since the determinant of the latter matrix is equal to  $2(1 - \cos(\theta_j))/\theta_j^2$ , which is positive for  $|\theta_j| < 2\pi$ . Furthermore, a direct computation shows that the inverse of  $M(\Omega)$  takes the following form in  $E_j$ :

$$M(\Omega)^{(-1)}|_{E_j} \sim \begin{pmatrix} \frac{\theta_j \cdot \sin(\theta_j)}{2 \cdot (1 - \cos(\theta_j))} & \frac{\theta_j}{2} \\ -\frac{\theta_j}{2} & \frac{\theta_j \cdot \sin(\theta_j)}{2 \cdot (1 - \cos(\theta_j))} \end{pmatrix}.$$

For  $|\theta_j| < \pi - C$ , some elementary calculus shows that there exists a constant  $K > 0$ , such that  $\frac{\theta_j \cdot \sin(\theta_j)}{2(1-\cos(\theta_j))} > K$ . As a consequence, we have:

$$M(\Omega)^{(-1)}|_{E_j} \sim \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

with  $a > K > 0$ . Under the assumption that  $|\theta_j| < \pi - C$  for all  $j$ , this implies that  $M(\Omega)^{(-1)} = S + A$ , where  $S$  is a symmetric positive-definite matrix with all its eigenvalues larger than  $K$  and  $A$  is a skew symmetric matrix. Then let us take a set of skew symmetric matrices  $\{\Omega_i\}$  whose eigenvalues are smaller than  $\pi - C$ . Any convex combination of the  $M(\Omega_i)^{(-1)}$  writes:

$$\sum_i w_i \cdot M(\Omega_i)^{(-1)} = \left( \sum_i w_i \cdot S_i \right) + \left( \sum_i w_i \cdot A_i \right) = \tilde{S} + \tilde{A},$$

where  $\tilde{S}$  is still symmetric positive-definite and  $\tilde{A}$  is skew symmetric. To see that this quantity is invertible, remark that  $(\tilde{S} + \tilde{A}) \cdot x = 0$  implies  $x^T \cdot \tilde{S} \cdot x + x^T \cdot \tilde{A} \cdot x = 0$ . But since  $x^T \cdot \tilde{A} \cdot x = (x^T \cdot \tilde{A} \cdot x)^T = -x^T \cdot \tilde{A} \cdot x = 0$ , then  $(\tilde{S} + \tilde{A}) \cdot x = 0$  implies  $x^T \cdot \tilde{S} \cdot x = 0$ , which is equivalent ( $\tilde{S}$  is symmetric positive-definite) to  $x = 0$ . Consequently  $\tilde{S} + \tilde{A}$  is invertible and this ends the proof.

#### 4.5 2D Rigid Transformations

Contrary to the general case, the bi-invariant mean of 2D rigid-body transformations have a *closed form*. The underlying reason is that  $SO(2)$ , the group of 2D rotations, is *commutative*. As a consequence, one can compute explicitly the bi-invariant mean of the rotation parts of the data and deduce from it the translation part using the barycentric equation, like in the proof of Theorem 15. More precisely, we have:

**Proposition 10.** *Let  $\{(R_i, t_i)\}$  be a set of 2D rigid-body transformations, such that the angles of rotation of the rotations  $R_i \cdot R_j^T$  are all strictly less than  $\pi$ . Then the bi-invariant mean  $(\bar{R}, \bar{t})$  associated to the weights  $\{w_i\}$  is given explicitly by:*

$$\begin{cases} \bar{R} = R_1 \cdot \exp\left(+ \sum_i w_i \cdot \log(R_1^T \cdot R_i)\right), \\ \bar{t} = \sum_i w_i \cdot Z^{(-1)} \cdot M\left(\log(\bar{R} \cdot R_i^T)\right)^{(-1)} \cdot R_i^T \cdot t_i, \end{cases} \quad (19)$$

with the following formulas for  $M$  and  $Z$ :

$$M\left(\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}\right)^{(-1)} \stackrel{def}{=} \begin{pmatrix} \frac{\theta \cdot \sin(\theta)}{2(1-\cos(\theta))} & \frac{\theta}{2} \\ -\frac{\theta}{2} & \frac{\theta \cdot \sin(\theta)}{2(1-\cos(\theta))} \end{pmatrix}, Z \stackrel{def}{=} \sum_i w_i \cdot M\left(\log(\bar{R} \cdot R_i^T)\right)^{(-1)}.$$

**Example of Bi-invariant Mean.** Let us take a look at the example chosen in [56, p.31]. Let  $T_1 = (\mathcal{R}(\pi/4), [-\sqrt{2}/2, \sqrt{2}/2]^T)$ ,  $T_2 = (Id_2, [\sqrt{2}, 0]^T)$  and  $T_3 = (\mathcal{R}(-\pi/4), [-\sqrt{2}/2, -\sqrt{2}/2]^T)$  be three rigid-body transformations in 2D.

We can compute exactly the bi-invariant mean of these rigid-body transformations with Eq. (19). A left-invariant Fréchet mean can also be computed explicitly in this case thanks to the simple form taken by the corresponding geodesics. And finally, thanks to Proposition 1, the analogous right-invariant Fréchet mean can be computed by inverting the data, computing their left-invariant mean and then inverting this Fréchet mean. The log-Euclidean mean can also easily be computed in closed form. This yields (after a number of simple but tedious algebraic manipulations):

- Left-invariant Fréchet mean:  $(Id_2, [0, 0]^T)$ ,
- Log-Euclidean mean:  $(Id_2, [\frac{\sqrt{2}-\frac{\pi}{4}}{3}, 0]^T) \simeq (Id_2, [0.2096, 0]^T)$ ,
- Bi-invariant mean:  $(Id_2, [\frac{\sqrt{2}-\frac{\pi}{4}}{1+\frac{\pi}{4}(\sqrt{2}+1)}, 0]^T) \simeq (Id_2, [0.2171, 0]^T)$ ,
- Right-invariant Fréchet mean:  $(Id_2, [\frac{\sqrt{2}}{3}, 0]^T) \simeq (Id_2, [0.4714, 0]^T)$ .

Interestingly, we thus see that the mean rotation angle is exactly the same in all cases. But the mean translations are different, and the bi-invariant mean is located nicely *between* the left- and right-invariant Fréchet means. This is quite intuitive, since the bi-invariant mean can be looked upon as an in-between alternative with regard to left- and right-invariant Fréchet means.

Although the Log-Euclidean mean is not left- nor right-invariant, it is actually quite close to the bi-invariant mean. In fact both means do correspond when the mean transformation is the identity, but differ at other points due to the curvature of the canonical Cartan connection.

#### 4.6 General Linear Transformations

In the linear group  $GL(n)$ , the determinant of the bi-invariant mean is equal to the *scalar geometric mean* of the determinant of the data. Our bi-invariant mean can thus be looked upon as a generalization of the geometric mean to invertible linear transformations. This generalization is not the only possible one. For instance, the Log-Euclidean mean has the same property. However, the Log-Euclidean mean is neither left- nor right-invariant (and is restricted to linear transformations whose principal logarithm is well-defined), which is *not* the case for the bi-invariant mean.

**Proposition 11.** *Let  $\{T_i\}$  be a set of linear transformations in  $GL(n)$  and let  $\{w_i\}$  be (normalized) non-negative weights, such that their bi-invariant mean  $\bar{T}$  uniquely exists. Then, we have:*

$$\begin{cases} \text{if } \det(S_i) > 0, \text{ for all } i, \text{ then } \det(\bar{T}) = \exp(\sum_i w_i \cdot \ln(\det(S_i))), \\ \text{if } \det(S_i) < 0, \text{ for all } i, \text{ then } \det(\bar{T}) = -\exp(\sum_i w_i \cdot \ln(-\det(S_i))), \end{cases}$$

*Proof.* Here, we extend the scope of our work to a Lie group which has two distinct sheets (positive or negative determinant) and it is thus not connected. The mean obviously does not exist when there are some data on both sheets

as there is no curve (thus no geodesic) to join the points on both sides. When the bi-invariant mean is well-defined, then all of the determinants of the data have the same sign, which is also the sign of  $\bar{T}$ , and we can actually restrict the analysis to this connected leaf of the manifold.

To prove our result, we will rely only on two ingredients: the barycentric equation (10) and the property:  $\det(M) = \exp(\text{Trace}(\log(M)))$ , which holds for any square matrix with a principal logarithm. This (classical) equality can be shown for example using the Jordan (or Schur) decomposition of the matrix  $M$ . Taking the trace of the barycentric equation and then the (scalar) exponential, we get:

$$1 = \prod_i \exp(w_i \cdot \text{Trace}(\ln(\det(\bar{T}^{(-1)} \cdot T_i)))) .$$

Then, using  $\det(A \cdot B) = \det(A) \cdot \det(B)$  and  $\det(\bar{T}^{(-1)} \cdot T_i) = |\det(\bar{T})|^{(-1)} \cdot |\det(S_i)|$ , we get the geometrical interpolation of determinants:

$$1 = |\det(\bar{T})|^{(-1)} \cdot \exp\left(\sum_i w_i \cdot \ln(|\det(S_i)|)\right),$$

which yields the result.

**Symmetric Positive Definite (SPD) Matrices.** SPD matrices are called tensors in medical image analysis. They are used for instance to encode the covariance matrix of the Brownian motion (diffusion) of water in Diffusion Tensor Imaging (DTI) [10, 46] or to encode the joint variability at different places (Green function) in shape analysis [24].

A number of teams in medical image processing proposed independently to endow this space with the *affine-invariant metric*  $\langle X, Y \rangle_{\Sigma} = \text{Tr}(V \cdot \Sigma^{-1} \cdot W \cdot \Sigma^{-1})$  which is completely independent of the choice of the coordinate system. This allowed to generalize to SPD-valued images a number of image processing algorithms [57]. The same metric was previously introduced in statistics to model the geometry of the multivariate normal family (the Fisher information metric) [15, 61, 16]. We showed in [56] that there is actually a one-parameter family of such affine-invariant metrics

$$\langle X, Y \rangle_{\Sigma} = \text{Tr}(V \cdot \Sigma^{-1} \cdot W \cdot \Sigma^{-1}) + \beta \cdot \text{Tr}(V \cdot \Sigma^{-1}) \cdot \text{Tr}(W \cdot \Sigma^{-1}) \quad \text{with} \quad \beta > -\frac{1}{\dim}$$

that share the same invariant connection  $\nabla_X Y = -\frac{1}{2}(V \cdot \Sigma^{-1} \cdot W + W \cdot \Sigma^{-1} \cdot V)$ .

The SPD manifold can be seen as a sub-manifold of  $GL(n)$ , but it is not a subgroup. Interestingly, the Fréchet mean associated to affine-invariant Riemannian metrics on the SPD space coincides with the bi-invariant mean of SPD matrices, looked upon as elements of  $GL(n)$ . Indeed, the affine-invariant Fréchet mean  $\bar{\Sigma}$  of a set of SPD matrices  $\Sigma_1, \dots, \Sigma_N$  is defined implicitly by the following barycentric equation:

$$\sum_i w_i \cdot \text{Log}(\bar{\Sigma}^{-\frac{1}{2}} \cdot \Sigma_i \cdot \bar{\Sigma}^{-\frac{1}{2}}) = 0, \quad (20)$$

which happens to be exactly equivalent to our general equation (10) for bi-invariant means (just multiply (20) on the left by  $\bar{\Sigma}^{-\frac{1}{2}}$  and on the right by  $\bar{\Sigma}^{+\frac{1}{2}}$  to obtain (10)). Intuitively, this means that our bi-invariant mean naturally unifies into a very general framework a number of well-established notions of means for various types of data living in geometric spaces (e.g, tensors, rotations, translations). The underlying reason is that tensors are normal matrices.

## 5 Perspectives

In this paper, we have presented a general framework to define bi-invariant means in Lie groups as exponential barycenters of the canonical Cartan connection. This new mean is invariant with respect to left- and right-multiplication, as well as inversion. We provided an iterative fixed point algorithm which converges at least linearly to this mean. From this optimization point of view, it would be interesting to extend the barycentric fixed point iteration into a Newton iteration on Lie groups [47], although the computation of the covariant derivative of the vector field could be a computational issue.

**Global Uniqueness Conditions?** In this work, we only showed the local existence and uniqueness of the bi-invariant mean for sufficiently concentrated data. However, we have no clear way to identify whenever data are sufficiently concentrated or not, contrarily to Riemannian manifolds where we now have fairly tight conditions to ensure the existence and uniqueness of the Riemannian barycenter [3, 68, 69]. Given an inner product on the Lie algebra of the group, we have two canonical Riemannian metrics which are respectively left- and right-invariant. One could conjecture that whenever the Fréchet mean is unique for some left- or right-invariant metric, then the bi-invariant mean exists and is unique as well, as we showed here for the particular case of rigid-body transformations.

An idea to investigate this link is the following. The bi-invariant mean defined in this work is a special instance of the exponential barycenters proposed in [23, 5, 6] for Riemannian manifolds. The existence and uniqueness of the exponential barycenters was recently established in affine connection manifolds which are convex with semi-local convex geometry (CSLCG) by Arnaudon and Li [4]. In our case, the whole group equipped with the canonical Cartan connection cannot be CSLCG, but it should be possible to define a separating function using left- or right-invariant Riemannian metrics such that the regular geodesic balls ensuring the uniqueness of the Fréchet mean actually define CSLCG neighborhoods with the canonical Cartan connection.

**Higher Order Moments** The mean is an important statistic which indicates the location of the distribution in the group, but higher order moments are also needed to characterize the dispersion of the population around this central value. The existence of a unique bi-invariant mean  $m$  of  $N$  samples  $\{x_i\}$  is based on the existence of a convex normal neighborhood of  $m$  containing the data

points. In this neighborhood, there exists a unique group geodesic joining any point  $x$  to each data point  $x_i$ , parameterized by its tangent vector  $\text{Log}_x(x_i) = DL_x|_e \cdot \text{Log}(x^{(-1)} \cdot x_i)$  at  $x$ . Thus, exactly as in Riemannian manifolds [55], we can locally linearize the Lie group around the mean using the map  $\text{Log}_x$  and define the 2-covariant tensor:

$$\Sigma(x) = \frac{1}{N} \sum_k \text{Log}_x(x_k) \otimes \text{Log}_x(x_k) \in T_x \mathcal{G} \otimes T_x \mathcal{G},$$

where the direct product  $\otimes$  means that in any basis of  $T_x \mathcal{G}$ , the coordinates are  $[\Sigma(x)]^{ij} = \frac{1}{N} \sum_k [\text{Log}_x(x_k)]^i \cdot [\text{Log}_x(x_k)]^j$ . Higher order empirical moments could be computed in the same way.

However, one should be careful that this definition is not the equivalent of the empirical covariance matrix without an auxiliary (Riemannian) metric to lower the indices of the tensor and turn it into a bi-linear form (an element of  $T_x^* \mathcal{G} \otimes T_x^* \mathcal{G}$ ). In particular, the usual interpretation of the coordinates of the covariance matrix using the scalar products of data with the basis vectors ( $\Sigma_{ij} = \mathbf{E}[\langle x, e_i \rangle \cdot \langle x, e_j \rangle]$ ) is valid only if we have a reference metric that allows us to define the orthonormality of the basis vectors  $e_i$ . Likewise, diagonalizing the  $\Sigma$  to extract the main modes of variability only makes sense with respect to a local metric: changing the metric of  $T_m \mathcal{G}$  will not only change the eigenvectors and the eigenvalues but also potentially the order of the eigenvalues. This means that Principle Component Analysis (PCA) cannot be generalized to the affine connection space setting.

**Bi-invariant Mahalanobis Distance** Despite the absence of a canonical reference metric, some interesting tools can be defined from the 2-covariant tensor in a bi-invariant way without having to rely on an auxiliary metric. One of them is the Mahalanobis distance of a point  $y$  (in the normal convex neighborhood specified above), which can be defined as:

$$\mu_{(m, \Sigma)}^2(y) = [\text{Log}_m(y)]^i \Sigma_{ij}^{(-1)} [\text{Log}_m(y)]^j. \quad (21)$$

In this formula,  $m$  is the bi-invariant mean and  $\Sigma_{ij}^{(-1)}$  are the coefficients of the inverse of  $\Sigma^{ij}$  in a given basis. One verifies that this definition does not depend on the basis chosen for  $T_m \mathcal{G}$ . Furthermore, the Mahalanobis distance is invariant by left and right translation. Indeed, if we have  $x'_i = g \cdot x_i$  and  $y' = g \cdot y$ , then we have  $m' = g \cdot m$ . Using the equality  $\text{Log}_{m'}(x') = DL_g|_m \cdot \text{Log}_m(x)$ , we see that the left translation just amounts to jointly change the coordinate system of both the covariance  $\Sigma'$  and the coordinates of  $\text{Log}_{m'}(y')$ . Thus, we have the left-invariance of the Mahalanobis distance:  $\mu_{(m', \Sigma')}^2(y') = \mu_{(m, \Sigma)}^2(y)$ . The invariance by right translation is obtained similarly from the right expression of the local logarithmic map (Corollary 1). This simple extension of the Mahalanobis distance suggests that it might be possible to extend much more statistical definitions and tools on Lie groups in a consistent way.

## Acknowledgments

The authors would like to thank the reviewers for their very detailed and insightful comments which considerably improved the manuscript. However, as mentioned in the text, some of their questions remains open for future investigations.

## References

1. P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
2. Roy L. Adler, Jean-Pierre Dedieu, Joseph Y. Margulies, Marco Martens, and Mike Shub. Newton’s method on riemannian manifolds and a geometric model for the human spine. *IMA Journal of Numerical Analysis*, 22(3):359–390, 2002.
3. Bijam Afsari. Riemannian  $l^p$  center of mass: existence, uniqueness, and convexity. *Proc. of the AMS*, 180(2):655–673, February 2010.
4. Marc Arnaudon and Xue-Mei Li. Barycenters of measures transported by stochastic flows. *The annals of probability*, 33(4):1509–1543, 2005.
5. M. Arnaudon. Espérances conditionnelles et  $C$ -martingales dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXVIII*, volume 1583 of *Lect. Notes in Math.*, pages 300–311. Springer-Verlag, 1994.
6. M. Arnaudon. Barycentres convexes et approximations des martingales continues dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXIX*, volume 1613 of *Lect. Notes in Math.*, pages 70–85. Springer-Verlag, 1995.
7. Vincent Arsigny. *Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI*. Thèse de sciences (phd thesis), École polytechnique, November 2006.
8. Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Bi-invariant means in lie groups. application to left-invariant polyaffine transformations. Research report rr-5885, INRIA Sophia-Antipolis, April 2006.
9. S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fund. Math.*, 3:133–181, 1922.
10. P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
11. R. Bhattacharya and V. Patrangenaru. Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, 108:23–36, 2002.
12. R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds, I. *Annals of Statistics*, 31(1):1–29, 2003.
13. R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds, II. *Annals of Statistics*, 33(3):1225–1259, 2005.
14. Jonathan Boisvert, Farida Cheriet, Xavier Pennec, Hubert Labelle, and Nicholas Ayache. Geometric variability of the scoliotic spine using statistics on articulated shape models. *IEEE Transactions on Medical Imaging*, 27(4):557–568, 2008.
15. J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.
16. M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9:119–138, 1991.

17. E. Cartan and J.A. Schouten. On the geometry of the group-manifold of simple and semi-simple groups. *Proc. Akad. Wetensch, Amsterdam*, 29:803–815, 1926.
18. S. Hun Cheng, N. J. Higham, C. S. Kenney, and A. J. Laub. Approximating the logarithm of a matrix to specified accuracy. *SIAM J. Matrix Anal. Appl.*, 22(4):1112–1125, 2001.
19. J.M. Corcuera and W.S. Kendall. Riemannian barycentres and geodesic convexity. *Math. Prob. Camb. Phil. Soc.*, 127:253–269, 1999.
20. W. D’Arcy Thompson. *On Growth and Form*. Cambridge University Press, England., 1917.
21. JeanPierre Dedieu, Pierre Priouret, and Gregorio Malajovich. Newton’s method on Riemannian manifolds: covariant alpha theory. *IMA Journal of Numerical Analysis*, 23(3):395–419, 2003.
22. M. do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.
23. M. Emery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXV*, volume 1485 of *Lect. Notes in Math.*, pages 220–233. Springer-Verlag, 1991.
24. Pierre Fillard, Vincent Arsigny, Xavier Pennec, Kiralee M. Hayashi, Paul M. Thompson, and Nicholas Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34(2):639–650, January 2007. Also as INRIA Research Report 5887, April 2006. PMID: 17113311.
25. P.T. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal component analysis on Lie groups. In *Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR’2003), Volume 1*, pages 95–101. IEEE Computer Society, 2003.
26. P.T. Fletcher, C. Lun, S.M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. on Medical Imaging*, 2004.
27. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’Institut Henri Poincaré*, 10:215–310, 1948.
28. Jean Gallier. Logarithms and square roots of real matrices. Technical reports (cis), UPENN, May 2008. arXiv:0805.0245.
29. S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer-Verlag, 2nd edition edition, 1993.
30. R. Godement. *Introduction à la Théorie des Groupes de Lie, Tomes I et II*. Publications Mathématiques de l’Université Paris VII, 1982.
31. U. Grenander. *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford University Press Inc., New York, NY., 1993.
32. David Groisser. Newton’s method, zeroes of vector fields, and the Riemannian center of mass. *Adv. in Applied Math*, 33:95–135, 2004.
33. B. C. Hall. *Lie Groups, Lie Algebras, and Representations: an Elementary Introduction*, volume 222 of *Graduate Texts in Mathematics*. Springer Verlag, 2003.
34. S. Helgason. *Differential Geometry, Lie groups, and Symmetric Spaces*. Academic Press, 1978.
35. N. J. Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM J. Matrix Anal. Appl.*, 26(4):1179–1193, 2005.
36. R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, 1990.
37. H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications in Pure and Applied Mathematics*, 30:509–541, 1977.
38. W.S. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
39. Kenney and Laub. Condition estimates for matrix functions. *SIAM J. Matrix Anal. Appl.*, 10:191–209, 1989.

40. Boris A. Khesin and Robert Wendt. *The Geometry of Infinite Dimensional Lie groups*, volume 51 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Springer Verlag, 2009.
41. W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.
42. S. Lang. *Algebra*. Graduate Texts in Mathematics. Springer, 3rd rev. ed. 2002. corr. 4th printing edition, 2004.
43. H. Turner Laquer. Invariant affine connections on Lie groups. *Transactions of the American Mathematical Society*, 331(2):pp. 541–551, 1992.
44. Huiling Le. Locating Fréchet means with application to shape spaces. *Advances in Applied Probabilities*, 33:324–338, 2001.
45. Huiling Le. Estimation of Riemannian barycenters. *LMS J. Comput. Math.*, 7:193–200, 2004.
46. D. Le Bihan, J.-F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriet. Diffusion tensor imaging: Concepts and applications. *Journal Magnetic Resonance Imaging*, 13(4):534–546, 2001.
47. R. Mahony and R. Manton. The geometry of the Newton method on non-compact Lie groups. *Journal of Global Optimization*, 23:309–327, 2002.
48. M.I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.
49. M. Moakher. Means and averaging in the group of rotations. *SIAM Journal of Matrix Analysis and Applications*, 24(1):1–16, 2002.
50. J.M. Oller and J.M. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.
51. B. Owren and B. Welfert. The Newton iteration on Lie groups. *BIT Numerical Mathematics*, 40(1):121–145, 2000.
52. R.L.M. Peeters and B. Hanzon. The Riemannian interpretation of Gauss-Newton and scoring, with application to system identification. FEWEB research memorandum 1992-22, Vrije Universiteit Amsterdam, 1992.
53. Xavier Pennec. *L'incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire*. Thèse de sciences (phd thesis), Ecole Polytechnique, Palaiseau (France), December 1996.
54. Xavier Pennec. Computing the mean of geometric features - application to the mean rotation. Research Report RR-3371, INRIA, March 1998.
55. Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, July 2006. A preliminary appeared as INRIA RR-5093, January 2004.
56. Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006.
57. Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, January 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
58. M. M. Postnikov. *Geometry VI: Riemannian Geometry*. Encyclopedia of mathematical science. Springer, 2001.
59. H. Samelson. On the Brouwer fixed point theorem. *Portugal. Math.*, 22:264–268, 1963.
60. D.H. Sattinger and O.L. Weaver. *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*, volume 61 of *AMS*. Springer Verlag, 1986.
61. L.T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scand. J. Statistics*, 11:211–223, 1984.

62. Steven T. Smith. Optimization techniques on Riemannian manifolds, Hamiltonian and gradient flows, algorithms and control. *Fields Inst. Commun.*, 3:113136, 1994.
63. Shlomo Sternberg. *Lectures on Differential Geometry*. Prentice Hall Mathematics Series. Prentice Hall Inc., 1964.
64. Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.
65. Constantin Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.
66. R. P. Woods. Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *Neuroimage*, 18(3):769–88, 2003.
67. M. Wüstner. A connected lie group equals the square of the exponential image. *Journal of Lie Theory*, 13:307–309, 2003.
68. Le Yang. Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479, 2010.
69. Le Yang. *Medians of probability measures in Riemannian manifolds and applications to radar target detection*. PhD thesis, Poitier University, December 2011.
70. Laurent Younes. *Shapes and Diffeomorphisms*, volume 171 of *Applied Mathematical Sciences*. Springer-Verlag, 2011.
71. Milos Zefran, Vijay Kumar, and Croke Christopher. Metrics and connections for rigid-body kinematics. *International Journal of Robotics Research*, 18(2):243–258, 1999.
72. H. Ziezold. On expected figures in the plane. In A. Hübler, W. Nagel, B.D. Ripley, and G. Werner, editors, *Geobild '89*, volume 51 of *Math. Res.*, pages 105–110. Akademie-Verlag, Berlin, 1989.

# Index

- Adjoint
  - Group, 6, 16
  - Operator, 6, 12, 15
- Affine connection, *see* Connection
- Baker-Campbell-Hausdorff Formula, 8
- Christoffel symbols, 9
- Computational anatomy, 2
- Conjugation, 5
- Connection, 9
  - Bi-invariant, 10, 11
  - Canonical Cartan, 10, 12
  - Cartan-Schouten, 4, 10
  - Flat, 12
  - Left-invariant, 10
  - Right-invariant, 10
  - Torsion free, 10
- Contraction, 24
- Covariance
  - 2-covariant tensor, 39
  - Mahalanobis distance, 39
- Curvature
  - Sectional, 16, 18, 33
  - Tensor, 11
- Exponential
  - Canonical Cartan connection, 13
  - Group, 7
  - Matrix, 7
- Exponential barycenter, 19
  - Canonical Cartan connection, 21
- Fixed point iteration
  - Convergence, 24
  - Gauss-Newton, 4, 19
  - Newton, 4, 19
  - on Lie groups, 20
  - on Riemannian manifolds, 19
- General linear group
  - Bi-invariant mean, 37
- Geodesics
  - Auto-parallel curves, 9
  - Group, 12, 20
  - Riemannian, 20
- Heisenberg group, 29–30
  - Bi-invariant mean, 29
  - Riemannian metric, 29
- Injection radius, 18
- Left translation, 5
- Lie
  - algebra, 5
  - bracket, 5
  - derivative, 5
  - group, 1–40
- Linear transformations
  - Bi-invariant mean (determinant), 36
- Logarithm
  - Canonical Cartan connection, 13
  - Group, 8
  - Matrix, 7
- Mahalanobis distance, 39
- Mean
  - Bi-invariant, 4, 21
  - Fréchet, 14, 18
  - Karcher, 18
  - Log-Euclidean, 14, 36
- Metric
  - Bi-invariant, 15–17, 19
  - Left-invariant, 14, 15
  - Right-invariant, 15
- Normal convex neighborhood, 9, 12
- One-parameter subgroups, 7, 10, 12, 15
  - Scaling and squaring algorithm, 7
- Parallel transport, *see* Connection
- Riemannian center of mass, *see* Mean
- Riemannian metric, *see* Metric
- Right translations, 5
- Rigid-body transformations, 16, 31–36
  - Bi-invariant mean, 32, 33, 35, 36
  - Left-invariant mean, 36
  - Log-Euclidean mean, 36
  - Logarithm, 32
- Rotation
  - Decomposition into 2-D rotations, 31

- Riemannian Mean, 33
- Sectional curvature, 33
  
- Scaled unitriangular matrices, 30–31
  - Bi-invariant mean, 30
- Scalings and translations group, 25–28
  - Bi-invariant mean, 27
  - Metric geodesics, 27
  - Riemannian metric, 26
- Symmetric positive definite matrices, 37
  - Invariant connection, 37
  - Invariant mean, 37
  - Invariant metrics, 37
  
- Torsion, 10, 11
  
- Vector field
  - Left invariant, 6
  - Right-invariant, 6
  
- Whitehead theorem, 9