Noname manuscript No.
(will be inserted by the editor)

Distances and Means of Direct Similarities

Minh-Tri Pham · Oliver J. Woodford · Frank Perbet · Atsuto Maki · Riccardo Gherardi · Björn Stenger · Roberto Cipolla

Received: date / Accepted: date

Abstract The non-Euclidean nature of direct isometries in a Euclidean space, i.e. transformations consisting of a rotation and a translation, creates difficulties when computing distances, means and distributions over them, which have been well studied in the literature. Direct similarities, transformations consisting of a direct isometry and a positive uniform scaling, present even more of a challenge—one which we demonstrate and address here. In this article, we investigate divergences (a superset of distances without constraints on symmetry and sub-additivity) for comparing direct similarities, and means induced by them via minimizing a sum of squared divergences. We analyze several standard divergences: the Euclidean distance using the matrix representation of direct similarities, a divergence from Lie group theory, and the family of all left-invariant distances derived from Riemannian geometry. We derive their properties and those of their induced means, highlighting several shortcomings. In addition, we introduce a novel family of left-invariant divergences, called SRT divergences, which resolve several issues associated with the standard divergences. In our evaluation we empirically demonstrate the derived properties of the divergences and means, both qualitatively and quantitatively, on synthetic data. Finally, we compare the divergences in a real-world application: vote-based, scale-invariant object recognition. Our results show that the new divergences presented here, and their means, are both more effective and faster to compute for this task.

Keywords Direct similarity · Distance · Mean · Registration · Object Recognition

M-T Pham, O Woodford, F Perbet, R Gherardi, B Stenger
Toshiba Research Europe Ltd.
208 Cambridge Science Park, Milton Road
Cambridge, CB4 0GZ, UK
E-mail: mtpham@crl.toshiba.co.uk

A Maki
Computer Vision and Active Perception Lab
School of Computer Science and Communication
KTH Royal Institute of Technology
100 44 Stockholm, Sweden

R Cipolla
Engineering Department
University of Cambridge
Trumpington Street, Cambridge, CB2 1PZ, UK

1 Introduction

Direct isometries in Euclidean geometry, which are transformations consisting of a rotation and a translation (Coxeter, 1961, §3), are relatively common in the literature, finding use in applications such as 2D object detection (Leibe et al, 2008; Opelt et al, 2008; Shotton et al, 2008), motion segmentation (Subbarao and Meer, 2009) and interpolation (Zefran and Kumar, 1998), 3D object recognition and registration (Drost et al, 2010; Knopp et al, 2010; Pennec and Thirion, 1997; Tombari and Di Stefano, 2010), and 3D camera pose estimation e.g. (Hartley and Zisserman, 2004). The non-Euclidean nature of this transformation space, specifically due to the rotation component, though encumbering, has been addressed previously (Agrawal, 2006; Park, 1995; Pennec and Thirion, 1997). By contrast, a direct similarity (Coxeter, 1961, §5), a transformation consisting of a direct isometry with a positive uniform scaling, introduces a further challenge, due to the interactions of scale with both rotation and translation (Arsigny et al, 2006b; Poincaré, 1882), that has received little attention (Bossa and Olmos, 2006; Eade, 2011; Pham et al, 2011; Strasdat et al, 2010). Consequently, far fewer methods include variable scale (Bossa and Olmos, 2006; Khoshelham, 2007; Pham et al, 2011; Strasdat et al, 2010; Woodford et al, 2013). However, doing so enables methods to be scale-invariant, which is a desirable property when the absolute scale of data is unknown, such as in non-metric reconstructions, or when there is high intra-class scale variation.

This article concerns itself with direct similarities, and in particular, divergences[1] between such transformations. Divergences are the key component of many tasks, such as comparing, clustering, averaging, embedding, kernel density estimation and other forms of probability distributions. We compare several divergences in direct similarity space (written as $\mathcal{DS}(n)$, where n is the dimensionality of the Euclidean space on which the transformations act): the Euclidean distance using the matrix representation of direct similarities, a divergence derived from Lie group theory, which we call the Lie divergence, the family of all left-invariant distances derived from Riemannian geometry, and a new family of left-invariant divergences, called SRT divergences, which we introduce here. We enumerate the properties of all aforementioned divergences, and those of their induced means via minimizing sum of squared divergences. We show that the existing divergences have various issues, including variance to scale, bias[2] and means without closed-form solutions, which are resolved by SRT divergences. We verify these properties empirically, both qualitatively and quantitatively, on synthetic data. Lastly, we demonstrate the benefits of using SRT divergences in scale-invariant 3D object recognition, in which divergences are applied to kernel density estimation (Parzen, 1962; Rosenblatt, 1956) and mean shift optimization (Cheng, 1995).

The novel contributions of this work[3] are as follows:

1. Analysis of the Euclidean distance, including a closed form for computing means, showing they are *biased* in scale.
2. Closed forms for the matrix exponential function and its inverse, mapping between direct similarities in $\mathcal{DS}(n)$ (with $n = 2$ or 3) and tangent vectors at the identity element of $\mathcal{DS}(n)$.

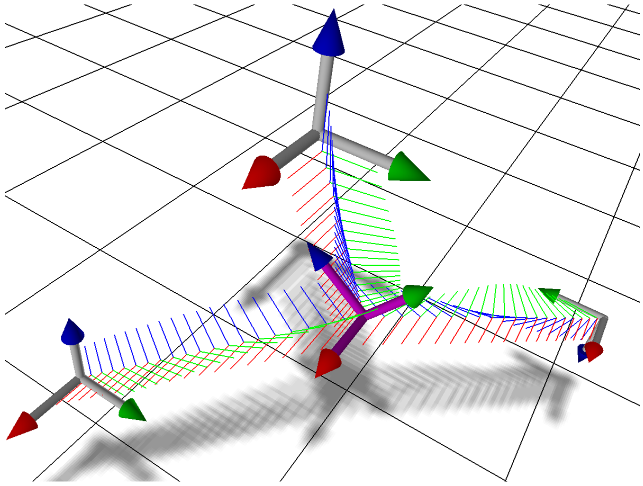


Fig. 1 A mean of three direct similarities in 3D. The origin and the orientation of a frame represent the translation and the rotation of the corresponding direct similarity, while the length of the axes represents the scale.

3. Analysis of the Lie divergence derived from the matrix exponential function, including a proof that any induced mean is scale-biased.
4. Analysis of *all* left-invariant Riemannian distances, proving any induced mean is scale-biased.
5. Proposal of a new family of efficient, left-invariant divergences, called SRT divergences, with closed form means (orders of magnitude faster to compute than means induced by intrinsic divergences). In particular, a member of the family induces *unbiased* means.
6. Experimental results on synthetic data and in 3D scale-invariant object recognition, demonstrating improved performance using the new SRT divergences.

The rest of the article is organized as follows: The next section provides background about $\mathcal{DS}(n)$, and divergences, means and distributions in it. In §3–6, we discuss the Euclidean distance, the Lie divergence, the family of left-invariant Riemannian distances, and the novel family of left-invariant SRT divergences respectively. Experimental results on synthetic data and on 3D scale-invariant object recognition are presented in §7, then §8 concludes the article. Proofs for theorems and lemmas are located in Appendix A.

2 Background

2.1 Overview of $\mathcal{DS}(n)$

A direct similarity refers to an n-dimensional transformation containing concurrently a positive uniform scaling, a rotation, and a translation. In 2D and 3D spaces, direct similarities can be visualized by Cartesian local frames, as illustrated in figure 1.

[1] Divergences are a superset of distances, defined in §2.1.1.
[2] Bias is defined in §2.1.2.
[3] Preliminary work from this article appears in (Pham et al, 2011), where contributions 1, 3, 5 and 6 are briefly reported. In addition to more detail and experiments here, we further introduce: i. closed-form Euclidean means and closed-form Lie divergences; ii. proofs that the Lie divergence and all left-invariant distances induce biased means; iii. an extension of the SRT divergence (Pham et al, 2011) to a family of SRT divergences with closed-form means.

Direct similarities of an n-dimensional Euclidean space span a space called the direct similarity group (Schramm and Schreck, 2003), denoted by:

$$\mathcal{DS}(n) := \left\{ \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} : s \in \mathbb{R}^+, \mathbf{R} \in \mathcal{SO}(n), \mathbf{t} \in \mathbb{R}^n \right\}, \tag{1}$$

where $\mathcal{SO}(n)$ is the group of n-dimensional rotation matrices and $\mathbf{0}$ denotes an all zero vector (or matrix) of the appropriate size. $\mathcal{DS}(n)$ belongs to the family of matrix Lie groups under matrix multiplication. Its dimension is $n_{ds} := \frac{n(n+1)}{2}+1$. Since $\mathcal{DS}(n)$ is a subgroup of the affine transformation group, we write $m\left(\mathbf{A},\mathbf{b}\right) := \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix}$ as a direct similarity and $M\left(\mathbf{A},\mathbf{b}\right) := \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 0 \end{bmatrix}$ as a tangent vector in $\mathcal{DS}(n)$ represented as a matrix. For notational convenience, given $\mathbf{X} \in \mathcal{DS}(n)$, let the scale, rotation and translation components of $\mathbf{X}$ be denoted by $\mathbf{X}_s$, $\mathbf{X}_r$ and $\mathbf{X}_t$ respectively, such that $\mathbf{X} = m\left(\mathbf{X}_s\mathbf{X}_r, \mathbf{X}_t\right)$.

$\mathcal{DS}(n)$ is a subgroup of the general linear group $\mathcal{GL}(n+1)$, with identity element $\boldsymbol{I}_{n+1}$ ($\boldsymbol{I}_n$ being the n-by-n identity matrix). It can be constructed from three subgroups: the group of direct dilations (Coxeter, 1961, §5) $\mathcal{D}(n) := \{m\left(s\boldsymbol{I}_n, \mathbf{0}\right) : s \in \mathbb{R}^+\}$, the rotation group $\mathcal{R}(n) := \{m\left(\mathbf{R},\mathbf{0}\right) : \mathbf{R} \in \mathcal{SO}(n)\}$ and the translation group $\mathcal{T}(n) := \{m\left(\boldsymbol{I}_n, \mathbf{x}\right) : \mathbf{x} \in \mathbb{R}^n\}$, via the following equation:

$$\mathcal{DS}(n) = \left(\mathcal{D}(n) \times \mathcal{R}(n)\right) \rtimes \mathcal{T}(n), \tag{2}$$

where $\times$ and $\rtimes$ denote the direct product and right semidirect product of groups. It is oriented and connected, making integration definable.

Although $\mathcal{R}(n)$ is compact, the other spaces $\mathcal{D}(n)$ and $\mathcal{T}(n)$ are not, making $\mathcal{DS}(n)$ locally compact (since it is a matrix Lie group) but not compact. As will be shown in subsequent sections, this fact causes a problem when working with the Lie divergence and left-invariant distances.

When discussing differentiating a function defined in $\mathcal{DS}(n)$, a map that parameterizes elements of $\mathcal{DS}(n)$ minimally by n_{ds}-dimensional vectors is required. Consider the following map $\phi : \mathbb{R}^{n_{ds}} \to \mathcal{DS}(n)$:

$$\phi(\mathbf{x}) := m\left(\mathrm{e}^{\mathbf{x}_s}\mathrm{e}^{\mathbf{x}_r^\times}, \mathbf{x}_t\right), \tag{3}$$

where $\mathbf{x}_s := \mathbf{x}_1$ denotes the first component of $\mathbf{x}$, i.e. the log-scale, $\mathbf{x}_r$ denotes the vector containing the next $n_r := \frac{n(n-1)}{2}$ components of $\mathbf{x}$, i.e. the rotation vector, $\mathbf{x}_t$ denotes the vector containing the last n components of $\mathbf{x}$, i.e. the translation vector, and the cross operator $\times$, converts a n_r-dimensional vector into a n-by-n skew-symmetric matrix (see §A.1 for the definition). In $\mathbb{R}^3$, the cross operator resembles the cross product of 3D vectors, i.e. $\mathbf{a}^\times\mathbf{b} = \mathbf{a} \times \mathbf{b}$. Function $\mathrm{e}^{\mathbf{x}_r^\times}$ denotes the matrix exponential series, given by:

$$e^{\mathbf{Z}} := \sum_{k=0}^{\infty} \mathbf{Z}^k/k!, \tag{4}$$

for any square matrix $\mathbf{Z}$. Let $\mathcal{J}_s := \{1\}$ be the set of scale coordinates, $\mathcal{J}_r := \{2,..,n_r+1\}$ be the set of rotation coordinates, and $\mathcal{J}_t := \{n_r+2,..,n_{ds}\}$ be the set of translation coordinates.

2.1.1 Divergences in $\mathcal{DS}(n)$

A distance function $d : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ for some space $\mathcal{G}$ defines a notion of dissimilarity between elements of $\mathcal{G}$. Typically, a distance satisfies the following four axioms which define a *metric*. For all $x, y, z \in \mathcal{G}$:

1. non-negativity: $d(x,y) \geq 0$,
2. coincidence: $d(x,y) = 0 \Leftrightarrow x = y$,
3. symmetry: $d(x,y) = d(y,x)$,
4. sub-additivity: $d(x,y) + d(y,z) \geq d(x,z)$.

In this paper we use "distance" to refer to a metric, i.e. a function that *does* satisfy all four conditions above. *Semimetrics* are a superset of metrics which drop the requirement of symmetry, requiring only conditions 1,2 and 4. *Divergences*, also sometimes referred to as quasisemimetrics, form a further superset of metrics which also drop the requirement of sub-additivity, requiring only conditions 1 and 2.

As well as determining which of the above conditions are met for each divergence considered, we further determine whether each divergence is left-invariant. Left-invariance implies invariance to scaling, rotation and translation altogether, and vice versa, and is thus an important property in many applications. Indeed, we focus our study on left-inavriant divergences.

Whilst rotation and translation invariance are almost always desirable, scale invariance is only necessary when the absolute scale of the coordinate system is unknown, e.g. in non-metric 3D reconstruction, or when scale is known to vary, e.g. in 3D object recognition where each class of objects contains instances of differing size.

2.1.2 Means in $\mathcal{DS}(n)$

Means can summarize a set. They are involved in many methods and models, such as k-means, mixture of Gaussians, mean shift etc.

A note on notation: we use a variant of *Einstein summation convention*, whereby if an iterable index is underlined, we take the smallest expression containing all occurences of that index, iterate the index over all possible values, and sum up the terms. For example, an arithmetic mean can be written as:

$$\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i}}/\mathsf{w}_{\underline{j}} := \frac{\sum_{i=1}^{N}\mathsf{w}_i\boldsymbol{X}_i}{\sum_{j=1}^{N}\mathsf{w}_j}. \tag{5}$$

This notation simplifies many expressions in the article.

In a Euclidean space, the arithmetic mean minimizes

$$\mathcal{E}(\mathbf{X}) := \mathsf{w}_{\underline{i}}d(\boldsymbol{X}_{\underline{i}}, \mathbf{X})^2, \tag{6}$$

where $d(\cdot,\cdot)$ is the Euclidean distance. In a differentiable manifold $\mathcal{G}$, arithmetic means may not exist. However, the term $\mathcal{E}(\mathbf{X})$, a measure of variance, is still a valid function if $d(\boldsymbol{X}_i, \mathbf{X})$ are well-defined for all i. When d is a metric, the most common way to proceed is to define a mean as an element that minimizes $\mathcal{E}(\mathbf{X})$ (Fréchet, 1948; Karcher, 1977):

$$\bar{\mathbf{X}} := \operatorname*{argmin}_{\mathbf{X}\in\mathcal{G}} \mathcal{E}(\mathbf{X}). \tag{7}$$

Strictly speaking, there may be cases where $\bar{\mathbf{X}}$ is not unique. However, a recent study (Arnaudon and Miclo, 2014) shows that in complete manifolds such cases occur in a subset of a null measure, implying that $\bar{\mathbf{X}}$ is almost surely unique.

There are other ways to define a mean in the literature. The distance in equation (6) can be raised by any power $\rho \geq 1$, leading to ρ-means (e.g. (Fréchet, 1948)). Instead of minimizing a function, a mean can be defined as a solution to a barycenter equation, leading to bi-invariant means (Arsigny et al, 2006b). We restrict this article to means defined by equations (6) and (7) and refer to any solution of them as a mean induced by the distance d. This definition guarantees *convergence* of Gaussian mean shift to a minimal energy level (Carreira Perpinan, 2007), since at each step of mean shift $\mathcal{E}(\mathbf{X})$ (for fixed weights w_i) is decreased. Such a property is not readily available with other definitions. Although the definition does not guarantee uniqueness of the mean, with respect to the distances to be discussed in the article, it almost surely does.

Pennec and Ayache (1998) proved that if d is left-invariant, means induced by d are (left-)*equivariant:* if all input $\boldsymbol{X}_i$ are multiplied by a same matrix $\mathbf{Z}$, the new mean equals the old mean multiplied by $\mathbf{Z}$.

In addition, all these properties still hold if d is relaxed from a distance to a divergence.

While left-invariance ensures that means behave consistently, it does not reveal whether a mean is biased, where bias is defined as:

Definition 1 Let $\overline{\mathbf{X}}$ be a mean of a set $\{(\boldsymbol{X}_i, \mathsf{w}_i) \in \mathcal{DS}(n)\times\mathbb{R}^+\}_{i=1}^N$ of N direct similarities $\boldsymbol{X}_i$ with weights $\mathsf{w}_i > 0$ induced by a given divergence function d.

– If $\boldsymbol{X}_{1;s} = \boldsymbol{X}_{2;s} = \ldots = \boldsymbol{X}_{N;s} \neq \overline{\mathbf{X}}_s$, $\overline{\mathbf{X}}$ is biased in scale, or *scale-biased.*
– If $\boldsymbol{X}_{1;r} = \boldsymbol{X}_{2;r} = \ldots = \boldsymbol{X}_{N;r} \neq \overline{\mathbf{X}}_r$, $\overline{\mathbf{X}}$ is biased in rotation, or *rotation-biased.*
– If $\boldsymbol{X}_{1;t} = \boldsymbol{X}_{2;t} = \ldots = \boldsymbol{X}_{N;t} \neq \overline{\mathbf{X}}_t$, $\overline{\mathbf{X}}$ is biased in translation, or *translation-biased.*

As we shall see in subsequent sections, some divergences induce means which are scale-biased, e.g. figures 2 and 3, biased due to variation in rotation and translation components respectively. However, analyzing the behaviour of means for each component, as we do in Theorems 1, 2 and 3, is most easily achieved by limiting variation in input direct similarities to that component alone. To enable this, we introduce a further set of definitions regarding *compatibility*:

Definition 2 Let $\overline{\mathbf{X}}$ be a mean of a set $\{(\boldsymbol{X}_i, \mathsf{w}_i) \in \mathcal{DS}(n)\times\mathbb{R}^+\}_{i=1}^N$ of N direct similarities $\boldsymbol{X}_i$ with weights $\mathsf{w}_i > 0$ induced by a given divergence function d.

– $\overline{\mathbf{X}}$ is compatible to a scale mean, or *scale-compatible*, if it is neither biased in rotation nor in translation. In other words, when all input rotations are the same and all input translations are the same, the mean does not have a different rotation and translation from these.
– Likewise, $\overline{\mathbf{X}}$ is compatible to a rotation mean, or *rotation-compatible*, if it is neither biased in scale nor in translation.
– Finally, $\overline{\mathbf{X}}$ is compatible to a translation mean, or *translation-compatible*, if it is neither biased in scale nor in rotation.

2.1.3 Distributions in $\mathcal{DS}(n)$

Probability distributions play a fundamental role in many inference techniques. The parameterization of such distributions often relies on a combination of kernels around particular points in the space in question. The classic example of such a kernel $K(\cdot)$ is the Gaussian kernel, $\exp\left(-\frac{\cdot}{2\sigma^2}\right)$, where σ is the bandwidth of the kernel, a single instance of which defines the normal distribution. Other distributions may be constructed from sums of kernels, e.g. mixture of Gaussians, kernel density estimates, or products of kernels, e.g. product of experts. These kernels often take as input a divergence between two points in the space. For example, a kernel density estimate (Parzen, 1962; Rosenblatt, 1956) PDF can be written as

$$\hat{f}(\mathbf{X}) := \mathsf{w}_{\underline{i}}\zeta(\boldsymbol{X}_{\underline{i}})^{-1}K(d(\boldsymbol{X}_{\underline{i}}, \mathbf{X})^2). \tag{8}$$

where $\mathbf{X}$ is the random variable, $(\boldsymbol{X}_i, \mathsf{w}_i)_{i=1}^N$ is a collection of input points $\boldsymbol{X}_i$ with weight $\mathsf{w}_i \geq 0$, $\mathsf{w}_{\underline{i}} = 1$, and $\zeta(\boldsymbol{X}_i)$ is the volume density function which normalizes $K(d(\boldsymbol{X}_i, \cdot)^2)$.

While $\zeta(\boldsymbol{X}_i)$ is a constant in a Euclidean space, it is *not* generally constant in a non-Euclidean space (Pelletier, 2005; Subbarao and Meer, 2009). A common solution to this problem is to simply assume that $\zeta(\boldsymbol{X}_i)$ is a constant, i.e. independent of $\boldsymbol{X}_i$, allowing many standard inference algorithms to be used, for example mean shift on kernel density estimates over Riemannian manifolds (Subbarao and Meer, 2009). However, Pham et al (2011) proved that if the divergence $d(\cdot, \cdot)$ is left-invariant, $\zeta(\boldsymbol{X}_i)$ are constant and can be ignored.

3 Euclidean Distance

Any matrix Lie group $\mathcal{G}$ is by definition a subgroup of a general matrix Lie group $\mathcal{GL}(n) \approx \mathbb{R}^{n^2}$ of n-by-n non-singular matrices. The most straightforward divergence in $\mathcal{G}$ is the Euclidean distance in the ambient Euclidean space $\mathbb{R}^{n^2}$, given by:

$$d_{\mathrm{E}}(\mathbf{X}, \mathbf{Y}) := \|\mathbf{X} - \mathbf{Y}\|_{\mathrm{F}}, \tag{9}$$

where $\|\mathbf{Z}\|_{\mathrm{F}} := \sqrt{\mathrm{trace}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})}$ denotes the Frobenius norm. Under d_{E}, the mean of matrices $(\boldsymbol{X}_i)_{i=1}^N$ with weights $(\mathsf{w}_i)_{i=1}^N$, becomes an arithmetic mean $\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i}}/\mathsf{w}_{\underline{j}}$ in the space $\mathbb{R}^{n^2}$. However, since we only consider matrices in $\mathcal{G}$, we rely on equation (7) to define and to find the mean. This mean is known in the literature as the *extrinsic* Euclidean mean (Bhattacharya and Patrangenaru, 2003).

As the Euclidean distance is not invariant to scaling (Pham et al, 2011), the weights w_i computed at every iteration of mean shift become scale-biased, leading to output poses with smaller scales. In (Pham et al, 2011), extrinsic Euclidean means are computed by minimizing equation (6). In contrast, the following lemma presents efficient formulæ for computing them.

Lemma 1 (Proof in §A.2) *The mean* $\overline{\mathbf{X}}$ *of a set* $\{(\boldsymbol{X}_i, \mathsf{w}_i) \in \mathcal{DS}(n) \times \mathbb{R}^+\}_{i=1}^N$ *of* N *direct similarities* $\boldsymbol{X}_i$ *with weights* w_i *induced by the Euclidean distance is given by:*

$$\overline{\mathbf{X}}_s = \mathrm{trace}(\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i};s}\boldsymbol{X}_{\underline{i};r}^{\mathrm{T}}\overline{\mathbf{X}}_r)/(n\mathsf{w}_{\underline{j}}), \tag{10}$$

$$\overline{\mathbf{X}}_r = \mathrm{sop}(\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i};s}\boldsymbol{X}_{\underline{i};r}), \tag{11}$$

$$\overline{\mathbf{X}}_t = \mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i};t}/\mathsf{w}_{\underline{j}}, \tag{12}$$

where $\boldsymbol{X}_{i;s}$, $\boldsymbol{X}_{i;r}$ *and* $\boldsymbol{X}_{i;t}$ *respectively are the scale, rotation, and translation components of direct similarities* $\boldsymbol{X}_i$ *for* $i = 1, \ldots, N$.

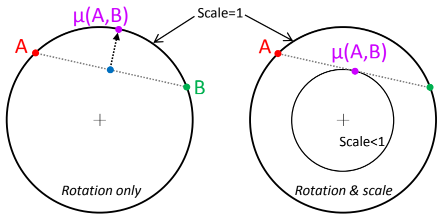


Fig. 2 Scale bias of an extrinsic Euclidean mean. Let us consider $\mathcal{DS}(2)$ (without translation): on a plane, a rotation can be represented as a point on a circle, the radius being the scale. *Left*: with rotation only, the arithmetic mean of **A** and **B** leads to a smaller scale but the reprojection onto the manifold (i.e. the unit circle) gives a reasonable result. *Right*: with rotation and scale, the mean is already on the manifold, but with a smaller scale.

In lemma 1, $\mathrm{sop}(\mathbf{X})$ is defined as the matrix function that returns the rotation matrix closest to matrix $\mathbf{X}$ with respect to the Frobenius norm, i.e. $\mathrm{sop}(\mathbf{X}) := \mathrm{argmin}_{\mathbf{Y} \in \mathcal{SO}(n)} \|\mathbf{Y} - \mathbf{X}\|_{\mathrm{F}}$. If $\mathbf{X}$ is decomposed into $\mathbf{X} = \mathbf{U}\mathrm{diag}(\mathsf{a}_1 \ldots, \mathsf{a}_n)\mathbf{V}^{\mathrm{T}}$ for some orthogonal matrices $\mathbf{U}, \mathbf{V} \in O(n, \mathbb{R})$ and singular values $\mathsf{a}_1 \geq \ldots \geq \mathsf{a}_n \geq 0$ via singular-value decomposition, the function $\mathrm{sop}(\mathbf{X})$ computes (Schönemann, 1966):

$$\mathrm{sop}(\mathbf{X}) = \mathbf{U}\mathrm{diag}(1, \ldots, 1, \det(\mathbf{U}\mathbf{V}))\mathbf{V}^{\mathrm{T}}. \tag{13}$$

Simply by inspecting (10), we see that the scale component of a mean is biased due to rotations, as illustrated in figure 2.

4 Lie Divergence

Another divergence arises from Lie group theory (Subbarao and Meer, 2009). Given a matrix Lie group $\mathcal{G} \subset \mathcal{GL}(n)$ for some dimension n, the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$ is the tangent space at $\boldsymbol{I}_n$, i.e. $\mathfrak{g} := T_{\boldsymbol{I}_n}\mathcal{G}$, which is a vector space. At any point $\mathbf{X} \in \mathcal{G}$, there is a (left-)Lie exponential map (Hall, 2003) which sends elements of $\mathfrak{g}$ to nearby points around $\mathbf{X}$ via integral curves, given by:

$$\mathrm{Exp}_{\mathbf{X}} : \mathfrak{g} \to \mathcal{G} : \mathrm{Exp}_{\mathbf{X}}(\mathbf{Z}) := \mathbf{X}e^{\mathbf{Z}}, \tag{14}$$

where $e^{\mathbf{Z}}$ is defined in equation (4). Since $\mathfrak{g}$ is a linear algebra, one can use the inverse map $\mathrm{Exp}_{\mathbf{X}}^{-1}$ which works locally (to be discussed in §4.1) to send a point $\mathbf{Y}$ near $\mathbf{X}$ back to $\mathfrak{g}$, and define the norm of the returned element as the divergence from $\mathbf{X}$ to $\mathbf{Y}$, given by:

$$d_{\mathrm{L}}(\mathbf{X}, \mathbf{Y}) := \left\|\ln(\mathbf{X}^{-1}\mathbf{Y})\right\|_{\mathrm{F}}, \tag{15}$$

where $\ln(\cdot)$ is the principal matrix logarithm (defined in §4.1). We refer to this divergence as the *Lie divergence*[4,5,6].

4.1 Matrix Exponential and Logarithm in $\mathcal{DS}(n)$

Working with the Lie divergence involves working with the matrix exponential and its inverse, called the principal matrix logarithm. Strictly speaking, only the principal matrix logarithm is required to compute the Lie divergence. However, if both the matrix exponential and logarithm can be computed efficiently, one can generally map a local neighborhood of the manifold to a vector space efficiently, and operate on the vector space instead. This idea has been applied to domains other than $\mathcal{DS}(n)$, e.g. (Begelfor and Werman, 2006; Cetingul and Vidal, 2009; Park, 1995; Pennec, 2006; Subbarao and Meer, 2009), where in most cases the Lie divergence is replaced by a Riemannian distance which happens to make use of the matrix exponential and logarithm.

The matrix exponential $\mathrm{e}^{\mathbf{Y}}$ of any matrix $\mathbf{Y}$, defined in equation (4), always converges. When $\mathbf{Y} \in \mathfrak{g}$, $\mathrm{e}^{\mathbf{Y}}$ is an element of $\mathcal{G}$ (Hall, 2003, §2). In contrast, since the map $\mathbf{Y} \to \mathrm{e}^{\mathbf{Y}}$ is many-to-one, its inverse may not converge. However, if we restrict the domain to a sufficiently small open neighborhood around $\mathbf{0}$, the matrix exponential becomes invertible, in which case its inverse, the principal matrix logarithm, denoted by $\ln\mathbf{X}$, is given by (Hall, 2003, §2):

$$\ln\mathbf{X} := -\sum_{k=1}^{\infty} \left(\boldsymbol{I}_n - \mathbf{X}\right)^k / k. \tag{16}$$

It is known that this series converges when $\|\mathbf{I} - \mathbf{X}\|_{\mathrm{F}} < 1$, but does not necessarily converge when $\|\mathbf{I} - \mathbf{X}\|_{\mathrm{F}} \geq 1$ (Hall, 2003). In addition, closed forms for $e^{\mathbf{Y}}$ and $\ln\mathbf{X}$ do not always exist, and finding closed forms for them in special cases are open research problems (Cheng et al, 2000; Gallier and Xu, 2002).

We are not aware of any previously reported closed forms for the two series, applied to $\mathcal{DS}(n)$ for the general case, nor for any particular n, except for $e^{\mathbf{Y}}$ in $\mathcal{DS}(2)$ (Eade, 2011) and $\mathcal{DS}(3)$ (Eade, 2011; Strasdat et al, 2010). We derive closed forms for both $e^{\mathbf{Y}}$ *and* $\ln\mathbf{X}$ in $\mathcal{DS}(2)$ and $\mathcal{DS}(3)$ below, providing for the first time a bidirectional mapping between $\mathcal{DS}(n)$ and its Lie algebra, denoted by $\mathfrak{ds}(n)$, for $n \in \{2,3\}$. Furthermore, we show in §A.3 that $\ln\mathbf{X}$ converges if and only if the rotation angle of $\mathbf{X}_r$ is not 180°. Closed forms in higher dimensional $\mathcal{DS}(n)$ are out the article's scope, as discussed in §A.3.

The Lie algebra of $\mathcal{DS}(n)$, $\mathfrak{ds}(n) := T_{\boldsymbol{I}_{n+1}}\mathcal{DS}(n)$, is the set of following matrices:

$$\mathfrak{ds}(n) = \{M\left(a\boldsymbol{I}_n + \mathbf{W}, \mathbf{u}\right) \\ : a \in \mathbb{R}, \mathbf{W} \in \mathfrak{so}(n), \mathbf{u} \in \mathbb{R}^n\}, \tag{17}$$

where $\mathfrak{so}(n) := T_{\boldsymbol{I}_n}\mathcal{SO}(n)$ represents the space of skew-symmetric matrices. Note that $a\boldsymbol{I}_n + \mathbf{W}$ uniquely identifies both a and $\mathbf{W}$, since the diagonal elements of $\mathbf{W}$ are zeros.

Substituting $\mathbf{Y} = M\left(a\boldsymbol{I}_n + \mathbf{W}, \mathbf{u}\right) \in \mathfrak{ds}(n)$ into equation (4) yields

$$\begin{aligned} e^{\mathbf{Y}} &= m\left(\sum_{k=0}^{\infty} \frac{(a\boldsymbol{I}_n + \mathbf{W})^k}{k!}, \sum_{k=0}^{\infty} \frac{(a\boldsymbol{I}_n + \mathbf{W})^k}{(k+1)!}\mathbf{u}\right) \\ &= m\left(e^a e^{\mathbf{W}}, \xi(a\boldsymbol{I}_n + \mathbf{W})\mathbf{u}\right), \end{aligned} \tag{18}$$

where $\mathrm{e}^{\mathbf{W}}$ is the matrix exponential in $\mathcal{SO}(n)$ and

$$\xi(\mathbf{Z}) := \sum_{k=0}^{\infty} \frac{\mathbf{Z}^k}{(k+1)!} \tag{19}$$

is a matrix series.

Similarly, substituting $\mathbf{X} = m\left(s\mathbf{R}, \mathbf{t}\right) \in \mathcal{DS}(n)$ into equation (16) yields

$$\begin{aligned} \ln\mathbf{X} &= M\left(-\sum_{k=1}^{\infty} \frac{(\boldsymbol{I}_n - s\mathbf{R})^k}{k}, \sum_{k=0}^{\infty} \frac{(\boldsymbol{I}_n - s\mathbf{R})^k}{k+1}\mathbf{t}\right) \\ &= M\left((\ln s)\boldsymbol{I}_n + \ln\mathbf{R}, \eta(s\mathbf{R})\mathbf{t}\right), \end{aligned} \tag{20}$$

where $\ln\mathbf{R}$ is the principal matrix logarithm in $\mathcal{SO}(n)$ and

$$\eta(\mathbf{Z}) := \sum_{k=0}^{\infty} \frac{(\boldsymbol{I}_n - \mathbf{Z})^k}{k+1} \tag{21}$$

is a matrix series.

To turn these series in $\mathcal{DS}(n)$ into closed forms, we need closed forms for the matrix exponential and logarithm series in $\mathcal{SO}(n)$, and closed forms for equations (19) and (21). While the former is available when $n \in \{2,3\}$ (Gallier and Xu, 2002), the latter has not been previously reported. We present the final results when $n \in \{2,3\}$ in the following four lemmas.

Lemma 2 (Proof in §A.3.1) *The matrix exponential of* $\mathbf{Y} = M\left(a\boldsymbol{I}_2 + \mathbf{W}, \mathbf{u}\right) \in \mathfrak{ds}(2)$ *is*

$$\mathrm{e}^{\mathbf{Y}} = m\left(\mathrm{e}^a \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \boldsymbol{E}_2\mathbf{u}\right) \tag{22}$$

[4] The Lie divergence in $\mathcal{DS}(n)$ is first discussed in (Pham et al, 2011), where it is called the intrinsic distance.

[5] There is also a right-Lie exponential map, $\mathrm{Exp}_{\mathbf{X}}(\mathbf{Z}) = e^{\mathbf{Z}}\mathbf{X}$. However, the resulting divergence, $d_{\mathrm{L}}\left(\mathbf{X}, \mathbf{Y}\right) = \left\|\ln(\mathbf{Y}\mathbf{X}^{-1})\right\|_{\mathrm{F}}$, would be right-invariant, not left-invariant.

[6] There is a similar divergence in the literature called the *Log-Euclidean distance*, defined by $d_{\mathrm{LE}}(\mathbf{X}, \mathbf{Y}) := \|\ln\mathbf{X} - \ln\mathbf{Y}\|_{\mathrm{F}}$ in the space of symmetric positive-definite matrices (Arsigny et al, 2006a). This divergence, if applied to $\mathcal{DS}(n)$, becomes inverse-invariant but not left-invariant.

where $\theta = \mathbf{W}_{2,1} = -\mathbf{W}_{1,2}$ *and*

$$\boldsymbol{E}_2 = \begin{bmatrix} \xi_{\mathfrak{r}} & -\xi_{\mathfrak{i}} \\ \xi_{\mathfrak{i}} & \xi_{\mathfrak{r}} \end{bmatrix}, \tag{23}$$

$$\xi_{\mathfrak{r}} := \frac{a(e^a \cos\theta - 1) + \theta e^a \sin\theta}{a^2+\theta^2}, \tag{24}$$

$$\xi_{\mathfrak{i}} := \frac{a e^a \sin\theta - \theta(e^a\cos\theta - 1)}{a^2+\theta^2}. \tag{25}$$

Lemma 3 (Proof in §A.3.1) *The principal matrix logarithm of* $\mathbf{X} = m\left(s\mathbf{R}, \mathbf{t}\right) \in \mathcal{DS}(2)$ *is*

$$\ln\mathbf{X} = M\left(\begin{bmatrix} \ln s & -\theta \\ \theta & \ln s \end{bmatrix}, \boldsymbol{L}_2\mathbf{t}\right) \tag{26}$$

where $\theta = \arctan(\mathbf{R}_{2,1}/\mathbf{R}_{1,1})$*, and*

$$\boldsymbol{L}_2 = \begin{bmatrix} \eta_{\mathfrak{r}} & -\eta_{\mathfrak{i}} \\ \eta_{\mathfrak{i}} & \eta_{\mathfrak{r}} \end{bmatrix}, \tag{27}$$

$$\eta_{\mathfrak{r}} := \frac{\ln s\,(s\cos\theta - 1) + \theta\,(s\sin\theta)}{(s\cos\theta-1)^2 + (s\sin\theta)^2}, \tag{28}$$

$$\eta_{\mathfrak{i}} := \frac{\theta\,(s\cos\theta - 1) - \ln s\,(s\sin\theta)}{(s\cos\theta-1)^2 + (s\sin\theta)^2}. \tag{29}$$

Lemma 4 (Proof in §A.3.2) *The matrix exponential of* $\mathbf{Y} = M\left(a\boldsymbol{I}_3 + \mathbf{W}, \mathbf{u}\right) \in \mathfrak{ds}(3)$ *is*

$$\mathrm{e}^{\mathbf{Y}} = m\left(\mathrm{e}^a \mathrm{e}^{\mathbf{W}}, \boldsymbol{E}_3\mathbf{u}\right) \tag{30}$$

where

$$\mathrm{e}^{\mathbf{W}} = \boldsymbol{I}_3 + \frac{\sin\theta}{\theta}\mathbf{W} + \frac{1-\cos\theta}{\theta^2}\mathbf{W}^2 \tag{31}$$

is Rodrigues' formula (Gallier and Xu, 2002) for computing the matrix exponential of $\mathbf{W} \in \mathfrak{so}(3)$*,* $\theta = \frac{\|\mathbf{W}\|_{\mathrm{F}}}{\sqrt{2}}$*, and*

$$\boldsymbol{E}_3 = \frac{e^a-1}{a}\boldsymbol{I}_3 + \frac{\xi_{\mathfrak{i}}}{\theta}\mathbf{W} + \left(\frac{e^a-1}{a\theta^2} - \frac{\xi_{\mathfrak{r}}}{\theta^2}\right)\mathbf{W}^2, \tag{32}$$

with $\xi_{\mathfrak{r}}$ *and* $\xi_{\mathfrak{i}}$ *defined in lemma 2.*

Lemma 5 (Proof in §A.3.2) *The matrix principal logarithm of* $\mathbf{X} = m\left(s\mathbf{R}, \mathbf{t}\right) \in \mathcal{DS}(3)$ *is*

$$\ln\mathbf{X} = M\left((\ln s)\boldsymbol{I}_3 + \mathbf{W}, \boldsymbol{L}_3\mathbf{t}\right) \tag{33}$$

where $\theta = \arccos((\mathrm{trace}(\mathbf{R}) - 1)/2)$*,*

$$\mathbf{W} = \ln\mathbf{R} = \frac{0.5\theta}{\sin\theta}\left(\mathbf{R} - \mathbf{R}^{\mathrm{T}}\right) \tag{34}$$

computes the logarithm of rotation matrix $\mathbf{R}$*, and*

$$\boldsymbol{L}_3 = \frac{\ln s}{s-1}\boldsymbol{I}_3 + \frac{\eta_{\mathfrak{i}}}{\theta}\mathbf{W} + \left(\frac{\ln s}{(s-1)\theta^2} - \frac{\eta_{\mathfrak{r}}}{\theta^2}\right)\mathbf{W}^2, \tag{35}$$

with $\eta_{\mathfrak{r}}$ *and* $\eta_{\mathfrak{i}}$ *defined in lemma 3.*

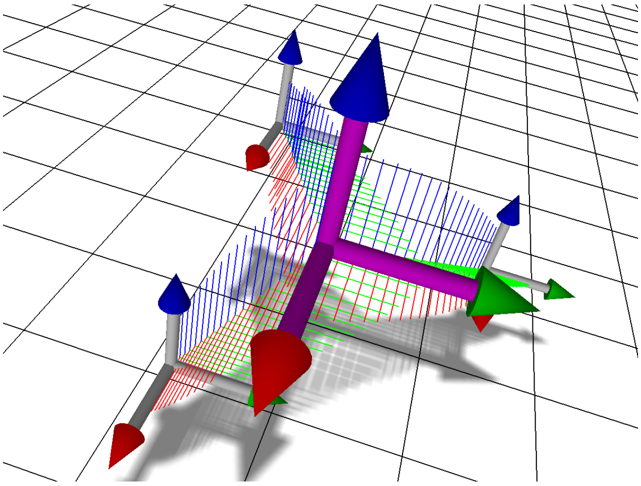


Fig. 3 Scale bias of a Lie mean. The Lie mean of three direct similarities in $\mathcal{DS}(3)$ with the same scale and rotation is a direct similarity with a larger scale.

4.2 Properties of the Lie Divergence and Means

The Lie divergence is symmetric since $\ln\mathbf{A} = -\ln\left(\mathbf{A}^{-1}\right)$ for all matrices $\mathbf{A}$ for which $\ln\mathbf{A}$ converges (Hall, 2003, §2). However, sub-additivity does not always hold. As an example, pick three 3D direct similarities: $\mathbf{A} := m\left(\boldsymbol{I}_3, \hat{\boldsymbol{e}}_1\right)$, $\mathbf{B} := m\left(\boldsymbol{I}_3, -\hat{\boldsymbol{e}}_1\right)$, and $\mathbf{C} := m\left(1.005\boldsymbol{I}_3, \mathbf{0}\right)$ in $\mathcal{DS}(3)$, where $\hat{\boldsymbol{e}}_i$ represents a unit vector having a 1 at row i and 0 everywhere else. Using lemma 5, the pairwise Lie divergences among them are $d_{\mathrm{L}}(\mathbf{A},\mathbf{B}) = 2$, $d_{\mathrm{L}}(\mathbf{A},\mathbf{C}) \approx 0.99752$, and $d_{\mathrm{L}}(\mathbf{C},\mathbf{B}) \approx 0.99752$. Hence, $d_{\mathrm{L}}(\mathbf{A},\mathbf{B}) > d_{\mathrm{L}}(\mathbf{A},\mathbf{C}) + d_{\mathrm{L}}(\mathbf{C},\mathbf{B})$. The Lie divergence is therefore a semimetric.

It is left-invariant since for all $\mathbf{X},\mathbf{Y},\mathbf{Z} \in \mathcal{G}$:

$$\begin{aligned} d_{\mathrm{L}}(\mathbf{ZX},\mathbf{ZY}) &= \left\|\ln(\mathbf{X}^{-1}\mathbf{Z}^{-1}\mathbf{ZY})\right\|_{\mathrm{F}} \\ &= \left\|\ln(\mathbf{X}^{-1}\mathbf{Y})\right\|_{\mathrm{F}} = d_{\mathrm{L}}(\mathbf{X},\mathbf{Y}). \end{aligned} \tag{36}$$

The mean induced by the Lie divergence is equivariant to scaling, rotation and translation altogether since the divergence is left-invariant. However, as shown by the following theorem, it is not translation-compatible, i.e. it is biased in scale due to translations, illustrated in figure 3. To our knowledge, no mean induced by the Lie divergence in $\mathcal{DS}(n)$ has a closed form; it is a solution to a non-convex optimization problem.

Theorem 1 (Proof in §A.4) *Any mean induced by the Lie divergence is scale-compatible, rotation-compatible, but not translation-compatible.*

4.3 Computing Lie Means and Mean Shift

We use the L-BFGS optimization algorithm[7] (Liu and Nocedal, 1989) to estimate the mean induced by the Lie divergence, using the closed forms presented in lemmas 3 and 5 to evaluate equation (15). We also use this approach to compute means within mean shift, in §7.2.2.

Mean shift in matrix Lie groups has been addressed before in the the literature: Subbarao and Meer (2009) propose a general mean shift method for any Riemmanian manifold, which they call *intrinsic mean shift*. They claim that intrinsic mean shift can be applied to any matrix Lie group, such as $\mathcal{DS}(n)$, and that in these cases the underlying Riemannian distance is given by $d(\mathbf{X}, \mathbf{Y}) = \left\|\ln(\mathbf{X}^{-1}\mathbf{Y})\right\|_{\mathrm{F}}$, i.e. the Lie divergence. Indeed, this approach was evaluated in $\mathcal{DS}(n)$ by Pham et al (2011). However, when applied to $\mathcal{DS}(n)$, this claim is incorrect—while Riemannian distances satisfy sub-additivity, the Lie divergence in $\mathcal{DS}(n)$ does not. The claim is therefore not entirely true, and results regarding intrinsic mean shift in (Pham et al, 2011) are also invalid.

The source of this confusion is worth investigating. Subbarao and Meer (2009) give two different formulæ for an inverse exponential map: the inverse of a *Riemannian* exponential map (equation (8) of that paper), which is defined from constant-velocity geodesics with the same origin (e.g. see (Lee, 1997)), and the inverse of the *Lie* exponential map (equation (34) of that paper, here given in (14)), defined from integral curves with the same origin. When the underlying metric tensor of a Riemannian manifold is bi-invariant, the Riemannian exponential map coincides with the Lie exponential map (O'Neill, 1983, §11.10); the names Riemannian and Lie can be dropped, the Lie divergence becomes a Riemannian distance, and means induced by the Lie divergence can be computed iteratively via Riemannian gradients. This is the case for several well known spaces, such as Euclidean spaces and rotation groups $\mathcal{SO}(n)$ (Pennec, 1998). However, bi-invariant metric tensors exist if and only if the adjoint representation of the group is compact (Sternberg, 1999, theorem V.5.3), a fact that was not addressed in (Subbarao and Meer, 2009). The adjoint representation of $\mathcal{DS}(n)$, $\mathrm{Ad}(\mathcal{DS}(n))$, contains linear operators that are not bounded by translations and scalings, i.e. a non-compact group, therefore there is no bi-invariant metric tensor in $\mathcal{DS}(n)$, implying the two maps do not coincide.

[7] We initialize the algorithm at one of the input direct similarities, the highest weighted one if computing a weighted mean.

This raises the question of how Riemannian distances and means in fact behave in $\mathcal{DS}(n)$, which we address in the next section.

5 Left-Invariant Riemannian Distances

Using Riemannian geometry (e.g. see (Lee, 1997; Sternberg, 1999)), one can convert any connected space $\mathcal{G}$ into a connected Riemannian manifold by endowing on it a metric tensor g; examples include (Begelfor and Werman, 2006; Dubbelman et al, 2012; Pennec, 2006; Subbarao and Meer, 2009). It can be shown that the metric tensor g induces a distance d_{R} restricted to $\mathcal{G}$ called a Riemannian distance, defined as the length of the shortest g-geodesic (geodesic induced by g) connecting two elements of the manifold. Conversely, given any Riemannian distance d_{R} on $\mathcal{G}$, one can obtain g at any element $\mathbf{X}$ by taking the derivatives of $d_{\mathrm{R}}(\mathbf{X}, \cdot)$. In other words, g and d_{R} can be used interchangeably.

In this section, we investigate the family of all left-invariant Riemannian distances in $\mathcal{DS}(n)$, leaving the study of the more general family of all Riemannian distances in $\mathcal{DS}(n)$ for future work. When the space $\mathcal{G}$ is a Lie group, if g is left-invariant then so is d_{R}, and vice versa (Arnold et al, 1989, §A.2).

Additionally, letting $f_{\mathbf{X}}(\mathbf{Y}) := d_{\mathrm{R}}(\mathbf{X}, \mathbf{Y})^2$, the following result holds whenever $f_{\mathbf{X}}$ is differentiable (i.e. not at the cut locus):

$$\mathrm{grad} f_{\mathbf{X}}(\mathbf{Y}) = -2\exp_{\mathbf{X}}^{-1}(\mathbf{Y}), \tag{37}$$

where $\exp_{\mathbf{X}}(\cdot)$ is the Riemannian exponential function at $\mathbf{X}$ (e.g. see (Lee, 1997, §4-6)).

The partial derivatives of the map ϕ, defined in (3), w.r.t. the local coordinates $\mathbf{x}$ at $\mathbf{X} = \phi(\mathbf{x})$, denoted by $\partial_{k;\mathbf{X}} := \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}_k}$ for $k = 1..n_{ds}$, provide a basis for the tangent space at $\mathbf{X}$, denoted by $T_{\mathbf{X}}\mathcal{DS}(n)$. These partial derivatives are given[8] by:

$$\partial_{k;\mathbf{X}} = \begin{cases} M\left(\mathrm{e}^{\mathbf{x}_s}\mathrm{e}^{\mathbf{x}_r^\times}, \mathbf{0}\right) & k \in \mathcal{J}_s, \\ M\left(\mathrm{e}^{\mathbf{x}_s}\mathrm{e}^{\mathbf{x}_r^\times}(\tilde{\Psi}_{\mathbf{x}_r}\hat{\boldsymbol{e}}_{k-1})^\times, \mathbf{0}\right) & k \in \mathcal{J}_r, \\ M\left(\mathbf{0}, \hat{\boldsymbol{e}}_{k-n_r-1}\right) & k \in \mathcal{J}_t, \end{cases} \tag{38}$$

where the matrix function $\tilde{\Psi}_{\mathbf{x}_r}$ is defined in (72).

Let $\varphi_{\mathbf{X}} : \mathfrak{ds}(n) \to \mathbb{R}^{n_{ds}}$ be the coordinate map for $T_{\mathbf{X}}\mathcal{DS}(n)$ under the basis $(\partial_{k;\mathbf{X}})_{k=1}^{n_{ds}}$. Any left-invariant metric tensor g can be expressed uniquely as:

$$\begin{aligned} g_{\mathbf{X}}(\mathbf{U}, \mathbf{V}) &= g_{\boldsymbol{I}_{n+1}}(dL_{\mathbf{X}^{-1}}(\mathbf{U}), dL_{\mathbf{X}^{-1}}(\mathbf{V})) \\ &= \varphi_{\mathbf{0}}(\mathbf{X}^{-1}\mathbf{U})^{\mathrm{T}}\tilde{\mathbf{G}}\varphi_{\mathbf{0}}(\mathbf{X}^{-1}\mathbf{V}), \end{aligned} \tag{39}$$

[8] These formulæ are derived from the partial derivatives of the matrix exponential in $\mathcal{SO}(n)$, given in §A.1.

where dL is the differential of the left-translation operator, i.e. $dL_{\mathbf{X}}(\mathbf{U}) := \mathbf{XU}$ for any matrices $\mathbf{X}$ and $\mathbf{U}$, and $\tilde{\mathbf{G}} \in \mathcal{GL}(n_{ds})$ is a constant, symmetric and positive-definite matrix that identifies the metric tensor g.

Means induced by left-invariant Riemannian distances are equivariant to scaling, rotation and translation altogether. Different values of $\tilde{\mathbf{G}}$ correspond to different left-invariant distances. In the next subsection, we determine if the means are biased or not.

5.1 Means induced by Left-Invariant Riemannian Distances

Recall that $\mathcal{D}(n)$ is the direct dilation group, $\mathcal{R}(n)$ is the rotation group, and $\mathcal{T}(n)$ is the translation group. The following theorem connects scale-compatibility, rotation-compatibility, and translation-compatibility to a concept called *totally geodesic in $\mathcal{DS}(n)$*, meaning any geodesic starting from a given element of a submanifold of $\mathcal{DS}(n)$ with a given tangent vector resides entirely in the submanifold[9] (e.g. see (Lee, 1997, §8)).

Theorem 2 (Proof in §A.5) *Let $\mathcal{DS}(n)$ be equipped with a left-invariant metric tensor g.*

1. *Every unique g-mean (mean induced by g) is scale-compatible if and only if $\mathcal{D}(n)$ is totally geodesic in $\mathcal{DS}(n)$.*
2. *Every unique g-mean is rotation-compatible if and only if $\mathcal{R}(n)$ is totally geodesic in $\mathcal{DS}(n)$.*
3. *Every unique g-mean is translation-compatible if and only if $\mathcal{T}(n)$ is totally geodesic in $\mathcal{DS}(n)$.*

Therefore, to find out if means induced by a given distance are scale-compatible, translation-compatible, and/or rotation-compatible, we can equivalently find out whether $\mathcal{D}(n)$, $\mathcal{R}(n)$, $\mathcal{T}(n)$ respectively are totally geodesic in $\mathcal{DS}(n)$ or not. The next theorem is a key finding of this section.

Theorem 3 (Proof in §A.6) *For every left-invariant metric tensor g on $\mathcal{DS}(n)$, $\mathcal{T}(n)$ is not totally geodesic in $\mathcal{DS}(n)$.*

Thus, combining the two theorems, we can firmly state that every mean induced by a left-invariant metric tensor is not translation-compatible. In contrast, the scale-compatibility or the rotation-compatibility of the mean depend on the choice of the metric tensor. This point is demonstrated in the next subsection.

[9] As an example, a great circle is totally geodesic in a sphere.

5.2 Natural Riemannian Distance

In this subsection, we analyze a left-invariant metric tensor that corresponds to the case that $\tilde{\mathbf{G}} = \boldsymbol{I}_{n_{ds}}$ in (39). Let us refer to it as $\breve{g}$, satisfying

$$\breve{g}_{\mathbf{X}}(\mathbf{U}, \mathbf{V}) := \left\|(\mathbf{X}^{-1}\mathbf{U})^{\mathrm{T}}(\mathbf{X}^{-1}\mathbf{V})\right\|_{\mathrm{F}}^{2} \tag{40}$$

for any $\mathbf{X} \in \mathcal{DS}(n)$ and any $\mathbf{U}, \mathbf{V} \in T_{\mathbf{X}}\mathcal{DS}(n)$. We refer to the Riemannian distance induced by $\breve{g}$ as the natural Riemannian distance.

Using the basis for $T_{\mathbf{X}}\mathcal{DS}(n)$ defined in (38), the metric tensor $\breve{g}$ leads to the following line element[10]:

$$ds^2 = d\mathbf{x}_s^2 + \left\|\tilde{\Psi}_{\mathbf{x}_r} d\mathbf{x}_r\right\|^2 + \mathrm{e}^{-2\mathbf{x}_s}\left\|d\mathbf{x}_t\right\|^2. \tag{41}$$

This equation enables us to write $(\mathcal{DS}(n), \breve{g})$ conveniently as the product space of two Riemannian manifolds (disregarding the group structure of $\mathcal{DS}(n)$):

$$(\mathcal{DS}(n), \breve{g}) = (\mathcal{R}(n), \breve{g}_r) \times (\mathcal{DT}(n), \breve{g}_{st}), \tag{42}$$

where $\mathcal{DT}(n) := \mathcal{D}(n) \rtimes \mathcal{T}(n)$ is the set of *direct dilatations* (Coxeter, 1961, §5), i.e. transformations consisting of a direct dilation and a translation, and $\breve{g}_r$ and $\breve{g}_{st}$ are respectively the restricted versions of $\breve{g}$ on $\mathcal{R}(n)$ and $\mathcal{DT}(n)$. Their corresponding line elements are:

$$ds_r^2 = \left\|\tilde{\Psi}_{\mathbf{x}_r} d\mathbf{x}_r\right\|^2, \tag{43}$$

$$ds_{st}^2 = d\mathbf{x}_s^2 + \mathrm{e}^{-2\mathbf{x}_s}\left\|d\mathbf{x}_t\right\|^2. \tag{44}$$

This decomposition has several corollaries. First, it divides the problem of finding the mean of direct similarities into two problems: finding the mean of rotations induced by $\breve{g}_r$, and finding the mean of direct dilatations induced by $\breve{g}_{st}$. This simplifies the original problem substantially. The squared Riemannian distance $d_{\mathrm{R}}(\mathbf{x}, \mathbf{y})^2$ becomes the sum of two squared distances:

$$d_{\mathrm{R}}(\mathbf{x}, \mathbf{y})^2 = d_r(\mathbf{x}, \mathbf{y})^2 + d_{st}(\mathbf{x}, \mathbf{y})^2, \tag{45}$$

where d_r and d_{st} are induced $\breve{g}_r$ and $\breve{g}_{st}$ respectively.

Second, we can see that the restricted metric tensor $\breve{g}_r$ is left-invariant and the corresponding matrix $\tilde{\mathbf{G}}$ in equation (39) is $\tilde{\mathbf{G}} = \boldsymbol{I}_{n_r}$. Thus, it coincides with the natural Riemannian metric tensor for rotation matrices in the literature (Park and Ravani, 1997), the geodesics and distances of which are well understood. Hence, the Riemannian rotation distance between two direct similarities parameterized by $\mathbf{x}$ and $\mathbf{y}$ is given by:

$$d_r(\mathbf{x}, \mathbf{y}) = \left\|\ln(\mathrm{e}^{-\mathbf{x}_r^{\times}}\mathrm{e}^{\mathbf{y}_r^{\times}})\right\|_{\mathrm{F}}, \tag{46}$$

the gradient of which, according to equation (37), is given by:

$$\mathrm{grad}(f_{r;\mathbf{x}})(\mathbf{y}) = 2\phi(\mathbf{x})M\left(\ln(\mathrm{e}^{-\mathbf{x}_r^{\times}}\mathrm{e}^{\mathbf{y}_r^{\times}}), \mathbf{0}\right), \tag{47}$$

[10] i.e. A line element is an infinitesimal arc-length.

enabling one to compute the rotation mean via Karcher's gradient-descent procedure (Karcher, 1977), as shown in (Pennec, 1998).

Third, by reparameterizing $(\mathbf{x}_s, \mathbf{x}_t) \to (\mathrm{e}^{\mathbf{x}_s}, \mathbf{x}_t)$, we see that the Riemannian manifold $(\mathcal{DT}(n), g_{st})$ is isometric to Poincaré's upper half-space model $\mathbb{R}^+ \times \mathbb{R}^n$ with line element $\|d\mathbf{x}\|^2 / \mathbf{x}_1^2$. Poincaré's model (originally due to Beltrami (1868)) is a classic example of a hyperbolic space, whose geodesics are half circles orthogonal to the hyperplane $\mathbf{x}_1 = 0$. Thus, the Poincarédistance between two direct dilatations parameterized by $(\mathbf{x}_s, \mathbf{x}_t)$ and $(\mathbf{y}_s, \mathbf{y}_t)$ is given as (Beltrami, 1868; Vaccaro, 2012):

$$d_{st}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left(1 + \frac{(\mathrm{e}^{\mathbf{x}_s} - \mathrm{e}^{\mathbf{y}_s})^2 + \|\mathbf{x}_t - \mathbf{y}_t\|^2}{2\mathrm{e}^{\mathbf{x}_s}\mathrm{e}^{\mathbf{y}_s}}\right), \tag{48}$$

where $\cosh(\cdot)$ is the hyperbolic cosine function. In order to find the mean of direct dilatations induced by $\breve{g}_{st}$, we can again use Karcher's gradient-descent procedure (Karcher, 1977). The missing element is a formula for the gradient of the squared Poincaré distance function. However, substituting equation (38) into equation (37), we obtain:

$$\begin{aligned} &\mathrm{grad}(f_{st;\mathbf{x}})(\mathbf{y}) = 2\mathrm{e}^{-\mathbf{x}_s}\beta(z) \times \\ &\quad M\left(\left((\mathrm{e}^{\mathbf{y}_s} - \mathrm{e}^{\mathbf{x}_s}) - \mathrm{e}^{\mathbf{x}_s}z/2\right)\boldsymbol{I}_n, \mathbf{y}_t - \mathbf{x}_t\right), \end{aligned} \tag{49}$$

where

$$z := \frac{(\mathrm{e}^{\mathbf{x}_s} - \mathrm{e}^{\mathbf{y}_s})^2 + \|\mathbf{x}_t - \mathbf{y}_t\|^2}{\mathrm{e}^{\mathbf{x}_s}\mathrm{e}^{\mathbf{y}_s}}, \tag{50}$$

$$\beta(z) := \frac{2\cosh^{-1}(1 + z/2)}{\sqrt{z(z+4)}}. \tag{51}$$

Fourth, due to the decomposition, the injectivity radius of any direct similarity in $(\mathcal{DS}(n), \breve{g})$ is the minimum of the injectivity radius of any point in $(\mathcal{R}(n), \breve{g}_r)$ and any point in $(\mathcal{DT}(n), \breve{g}_{st})$. The former is known to be $\frac{1}{4}$ (Moakher, 2002), at which the rotation angle between two rotations is 180°. The latter turns out to be ∞ since it is isometric to Poincaré's model, which has ∞ injectivity radius (Karcher, 1977; Vaccaro, 2012). Therefore, as long as the rotation angle between two direct similarities in $\mathcal{DS}(n)$ is less than 180°, the unit-speed geodesic between them under $\breve{g}$ is unique.

Finally, the decomposition leads to the fact that both $\mathcal{DT}(n)$ and $\mathcal{R}(n)$ are totally geodesic in $\mathcal{DS}(n)$ under $\breve{g}$, because the line elements of equations (43) and (44) are parameterized separately. Also note that $\mathcal{D}(n)$ is totally geodesic in $\mathcal{DS}(n)$ under $\breve{g}_{st}$, the restricted version of $\breve{g}$ on $\mathcal{DT}(n)$. Therefore, according to theorem 2, any $\breve{g}$-mean is both scale-compatible and rotation-compatible.

In summary, via the decomposition of the Riemannian manifold $(\mathcal{DS}(n), \breve{g})$, we have shown how to compute the mean of direct similarities in $\mathcal{DS}(n)$ induced by $\breve{g}$, and that it is guaranteed to be scale-compatible and rotation-compatible, but not translation-compatible, due to theorem 3.

6 SRT Divergences

We have seen that the Euclidean distance and its means are simple, and efficient to compute. However, the distance is not left-invariant and the means are biased in scale due to rotations. We have seen that the Lie divergence and all left-invariant distances induce means which are biased in scale due to translations, and also must be computed via an iterative process. In this section, we construct a new family of left-invariant divergences in $\mathcal{DS}(n)$, which are simpler and more efficient than the Lie divergence and Riemannian distances in computing divergences and means. We call them the *SRT divergences.*

6.1 Divergence Construction

Inspired by the natural Riemannian distance in §5.2, in which the rotation component can be separated from other components, we define an SRT divergence to be an $\ell 2$-norm of component-wise divergences, thus:

$$d_\alpha(\mathbf{X}, \mathbf{Y}) := \sqrt{\frac{d_s(\mathbf{X}, \mathbf{Y})^2}{\sigma_s^2} + \frac{d_r(\mathbf{X}, \mathbf{Y})^2}{\sigma_r^2} + \frac{d_{t;\alpha}(\mathbf{X}, \mathbf{Y})^2}{\sigma_t^2}}. \tag{52}$$

where d_s, d_r and $d_{t;\alpha}$ measure scale, rotation and translation divergences respectively, $\sigma_s, \sigma_r, \sigma_t > 0$ are bandwidth coefficients, and $\alpha \in \mathbb{R}$ is a divergence parameter. Varying $\sigma_s, \sigma_r, \sigma_t$ alters the relative sensitivity of the SRT divergence to different types of transformations, making SRT divergences more flexible than existing divergences.

The component-wise divergences are defined as follows:

$$d_s(\mathbf{X}, \mathbf{Y}) := |\ln(\mathbf{X}_s/\mathbf{Y}_s)|, \tag{53}$$

$$d_r(\mathbf{X}, \mathbf{Y}) := \|\mathbf{X}_r - \mathbf{Y}_r\|_{\mathrm{F}}, \tag{54}$$

$$d_{t;\alpha}(\mathbf{X}, \mathbf{Y}) := \frac{\|\mathbf{X}_t - \mathbf{Y}_t\|}{\sqrt{\mathbf{X}_s^{1+\alpha}\mathbf{Y}_s^{1-\alpha}}}. \tag{55}$$

It is clear that d_α in equation (52) can be evaluated in closed-form. SRT divergences do not satisfy sub-additivity[11]. The conjugate of a divergence d_α, denoted by $d_{\bar{\alpha}}$, is given by the following:

$$d_{\bar{\alpha}}(\mathbf{X},\mathbf{Y}) := d_\alpha(\mathbf{Y},\mathbf{X}) = d_{-\alpha}(\mathbf{X},\mathbf{Y}). \tag{56}$$

Therefore, every $(d_\alpha, d_{-\alpha})$ is a pair of divergences preserving conjugate symmetry, and d_0 is symmetric.

The first two distances, given by equations(53) and (54), are bi-invariant distances on $\mathcal{D}(n)$ and $\mathcal{R}(n)$ respectively. The quotient $\sqrt{\mathbf{X}_s^{1+\alpha}\mathbf{Y}_s^{1-\alpha}}$ is introduced in equation (55) to make the translation distance $\|\mathbf{X}_t - \mathbf{Y}_t\|$ invariant to scaling. As a result, SRT divergences are proved to be left-invariant by the following theorem.

Theorem 4 (Proof in §A.7) *For every $\alpha \in \mathbb{R}$ and $\sigma_s, \sigma_r, \sigma_t \in \mathbb{R}^+$, d_α is left-invariant.*

6.1.1 Discussion of Rotation Distances

Even though in equation (54) the Euclidean distance of rotation matrices (Downs, 1972), also known as the extrinsic rotation distance (Moakher, 2002) (**extR**), is used as the distance of rotations to simplify the theoretical explanation, it is possible to use *any* left-invariant rotation distance at all. For rotation distances in 3D, we implemented the intrinsic rotation distance (Park and Ravani, 1997) (**intR**) given in (46) and a more efficient extrinsic quaternion distance (**quat**) of Ravani and Roth (1983), given by:

$$d_{\mathrm{Q}}(\mathbf{X}_r,\mathbf{Y}_r) = \sqrt{1 - |\mathbf{q}(\mathbf{X}_r)^{\mathrm{T}}\mathbf{q}(\mathbf{Y}_r)|}, \tag{57}$$

where $\mathbf{q}(\mathbf{X}_r)$ is the quaternion representation of the rotation component $\mathbf{X}_r$, and $|\cdot|$ is needed to account for the fact that $\mathbf{q}(\mathbf{X}_r)$ and $-\mathbf{q}(\mathbf{X}_r)$ represent the same rotation. Readers interested in existing rotation distances and means are referred to (Hartley et al, 2013).

6.2 Mean Computation

Intrinsic means, such as those induced by the Lie divergence or left-invariant distances, are slow to compute, requiring an iterative optimization. In contrast, means induced by SRT divergences, SRT means, are efficient to compute, as shown by the following lemma.

[11] For example, consider $\sigma_s = \sigma_r = \sigma_t = 1$, $\mathbf{A} := m\left(e^{-10}\boldsymbol{I}_3, \hat{\boldsymbol{e}}_1\right)$, $\mathbf{B} := m\left(e^{10}\boldsymbol{I}_3, \hat{\boldsymbol{e}}_1\right)$, and $\mathbf{C} := m\left(\boldsymbol{I}_3, \mathbf{0}\right)$ in $\mathcal{DS}(3)$. For any $\alpha \geq 0$: $d_\alpha(\mathbf{A},\mathbf{B}) = 20$, $d_\alpha(\mathbf{B},\mathbf{C}) = \sqrt{10^2 + e^{-10(1+\alpha)}}$ and $d_\alpha(\mathbf{A},\mathbf{C}) = \sqrt{10^2 + e^{10(1+\alpha)}}$, proving that $d_\alpha(\mathbf{A},\mathbf{C}) > d_\alpha(\mathbf{A},\mathbf{B}) + d_\alpha(\mathbf{B},\mathbf{C})$.

Lemma 6 (Proof in §A.8) *Let $\overline{\mathbf{X}}$ be the SRT mean of a set of direct similarities $\{\boldsymbol{X}_i\}_{i=1}^N$ with positive weights $\{\mathsf{w}_i > 0\}_{i=1}^N$. The components of $\overline{\mathbf{X}}$ are given by:*

$$\ln\overline{\mathbf{X}}_s = \operatorname*{argmin}_{z\in\mathbb{R}} \mathsf{w}_{\underline{i}}(z - \ln\boldsymbol{X}_{\underline{i};s})^2 + V_t \mathrm{e}^{(\alpha-1)z}, \tag{58}$$

$$\overline{\mathbf{X}}_r = \mathrm{sop}(\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i};r}), \tag{59}$$

$$\overline{\mathbf{X}}_t = \bar{\mathbf{t}}, \tag{60}$$

where $\bar{\mathbf{t}} = \mathsf{v}_{\underline{i}}\boldsymbol{X}_{\underline{i};t}/\mathsf{v}_{\underline{j}}$ is a weighted mean of translations with weights $\mathsf{v}_i = \frac{\mathsf{w}_i}{\boldsymbol{X}_{i;s}^{1+\alpha}}$, $V_t = \frac{\sigma_s^2}{\sigma_t^2}\mathsf{v}_{\underline{i}}\left\|\bar{\mathbf{t}} - \boldsymbol{X}_{\underline{i};t}\right\|^2$ is a multiple of their weighted variance, and the function to minimize in equation (58) *is convex. In addition, when* $\alpha = 1$,

$$\overline{\mathbf{X}}_s = \mathrm{e}^{\mathsf{w}_{\underline{i}}\ln\boldsymbol{X}_{\underline{i};s}/\mathsf{w}_{\underline{j}}}. \tag{61}$$

If quaternions are used to represent 3D rotations instead, the extrinsic mean of rotations in quaternion space (Ravani and Roth, 1983) is efficiently approximated as the length-normalized version of

$$\bar{\mathbf{q}} = \mathrm{sign}(\mathbf{q}(\boldsymbol{X}_{\underline{i}})^{\mathrm{T}}\hat{\mathbf{q}})\mathsf{w}_{\underline{i}}\mathbf{q}(\boldsymbol{X}_{\underline{i}}), \tag{62}$$

where $\hat{\mathbf{q}}$ is an estimate[12] of the mean.

We note that any SRT mean is equivariant since its underlying divergence is left-invariant. It follows immediately from equations (59) and (60) that every SRT mean is unbiased in rotation and translation, since $\overline{\mathbf{X}}_r$ is an *extrinsic mean* of rotations $\boldsymbol{X}_{i;r}$ and $\overline{\mathbf{X}}_t$ is an *arithmetic mean* of translations $\boldsymbol{X}_{i;t}$. It is scale-compatible because equations (59) and (60) involve only rotations $\boldsymbol{X}_{i;r}$ and translations $\boldsymbol{X}_{i;t}$ respectively. It is rotation-compatible because when all translations $\boldsymbol{X}_{i;t}$ are the same, $V_t = 0$, and equations (58) and (60) become a scale mean and a translation mean respectively. SRT means are in general not translation-compatible. However, when $\alpha = 1$, the second term of (58) vanishes and $\overline{\mathbf{X}}_s$ becomes a *geometric mean* of scales $\boldsymbol{X}_{i;s}$, making SRT means *unbiased* in scale and translation-compatible. When $\alpha = 0$, the biasedness of the scale component is illustrated in figure 4.

7 Evaluation

Table 1 summarizes the properties of SRT divergences and means, along with those of existing divergences,

[12] We set $\hat{\mathbf{q}}$ to one of the input quaternions (that with the highest weight, if computing a weighted mean), or in the case of mean shift, the mean from the previous iteration of mean shift. The evaluation of equation (62) could be repeated, replacing $\hat{\mathbf{q}}$ with $\bar{\mathbf{q}}$ each iteration, until convergence, which is rapid as $\hat{\mathbf{q}}$ affects only the sign applied to each input quaternion, but in practice we used just one iteration.

stated or derived in the previous sections. Through experiments, we aim to demonstrate two further points. Firstly, using synthetic direct similarities, we will empirically evaluate the divergences and means introduced earlier, in terms of bias and also computational efficiency, demonstrating the desirable properties of our new SRT divergences. Secondly, using publicly available data for a 3D object recognition and registration problem, we will show that SRT divergences can lead to improved performance in a real-world application.

7.1 Synthetic evaluation

Table 2 visualizes pseudo-geodesics induced by different divergences and transformations. A pseudo-geodesic between two direct similarities $\mathbf{X}$ and $\mathbf{Y}$ is defined as a sequence of means of the set $\{(\mathbf{X}, a), (\mathbf{Y}, 1-a)\}$ where a runs from 0 to 1. In the figures intermediate frames illustrate weighted means of the two given direct similarities. They form the geodesic between the two direct similarities if the underlying divergence is a metric (hence the name pseudo-geodesic). The table clearly shows how existing means are biased in scale.
Figure 4 plots scale-bias as a function of translation variance, excluding the unbiased SRT divergence. It shows that, the scale-biasedness increases as the translation variance increases. However, the scale-biasedness is negligible for all divergences when the translation variance is less than 10^{-3}.

Table 3 reports the average computational speed for each divergence in $\mathcal{DS}(n)$. The two SRT divergences present the fastest computational time, evaluating one million divergences in just 86 milliseconds, while the Lie divergence requires 6.4 seconds on the same task.

Figure 5 summarizes the average speed of computing means induced by each divergence. The take-home message here is that the evaluation time varies by orders of magnitude between extrinsic and intrinsic means. For example, the mean of 100 direct similarities induced by the SRT divergences and the Euclidean distance can be computed in 10-20 microseconds, while that induced by the Lie divergence or the natural Riemannian distance takes tenths of a second to compute.

7.2 3D object recognition and registration

We evaluate the divergences in the context of a real-world application—3D object recognition and registration—using a publicly-available dataset: the Toshiba CAD model point clouds dataset (Pham et al, 2012), which consists of direct similarity votes for shapes within 3D point clouds.

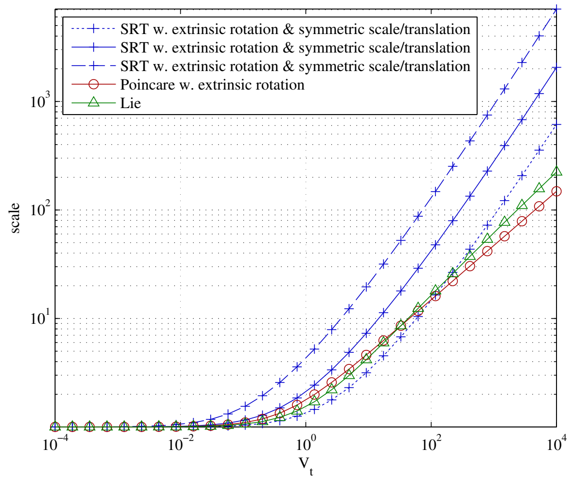


Fig. 4 Scale-biasedness due to translations of the SRT divergence with $\alpha = 0$. All input scales are set to 1. Based on (58), the scale component of the mean depends only on V_t, a weighted variance of translations. Different SRT curves (blue) correspond to different values of coeefficient $\frac{\sigma_s^2}{\sigma_t^2}$ of V_t, which are from left to right 0.5, 1, and 2 respectively. When V_t is large, the logarithm of the output scale becomes linear with the logarithm of V_t.

Table 3 Time for computing one million divergences**.**

Divergence type	Time to compute 10^6 divergences
Euclidean	0.246s
Lie	6.369s
intR + Poincaré	0.608s
intR + asymST	0.504s
intR + symST	0.462s
quat + Poincaré	0.231s
quat + asymST	**0.086s**
quat + symST	**0.086s**

These votes, computed from real data (Pham et al, 2011), can be used to infer the position of known shapes within the data, and compared with the ground truth, which is also provided. We use the *intrinsic Hough transform* (Woodford et al, 2013) as the inference framework; this method finds modes in a kernel density estimate of the posterior distribution of object poses and identities, and has been shown to perform well on this dataset. Within this framework we use each of the divergences discussed here in a Gaussian kernel on each vote, and quantitatively compare the recognition and registration scores of the resulting methods.

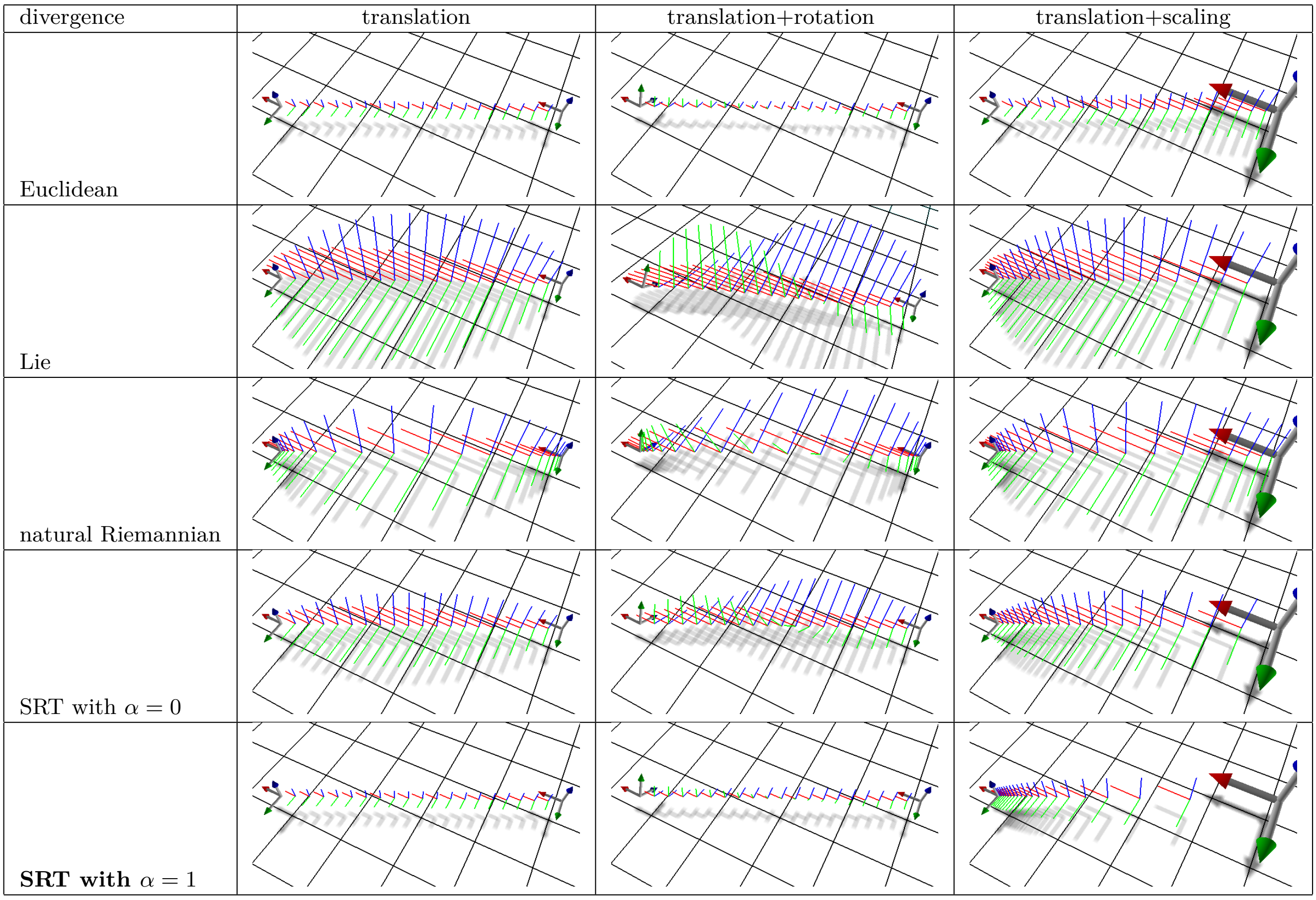
7.2.1 Test framework

The dataset (Pham et al, 2012) consists of 1000 test sets of votes, each computed from a point cloud containing a single, rigid object, one of ten test objects. Each vote consists of a direct similarity representing a putative 7D

Table 1 **Properties** of divergences and associated means in $\mathcal{DS}(n)$.

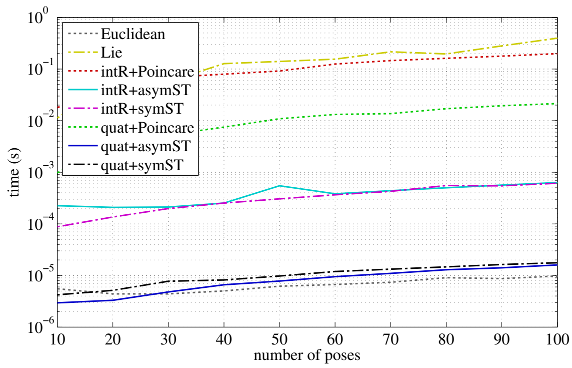
Properties	Euclidean	Lie	natural Riemannian	symmetric SRT $\alpha = 0$	unbiased SRT $\alpha = 1$
Divergence:					
Closed-form	✓	✓	✓	✓	✓
Left-invariant	✗	✓	✓	✓	✓
Symmetric	✓	✓	✓	✓	✗
Sub-additive	✓	✗	✓	✗	✗
Mean:					
Closed-form	✓	✗	✗	✗	✓
Equivariant	✓	✓	✓	✓	✓
Scale-compatible	✓	✓	✓	✓	✓
Rotation-compatible	✗	✓	✓	✓	✓
Translation-compatible	✓	✗	✗	✗	✓

Table 2 **(Pseudo-)geodesics of direct similarities.** Each intermediate direct similarity is a weighted mean of the two end direct similarities. The (pseudo-)geodesics induced by all except the symmetric SRT divergence are independent of the bandwidth σ.

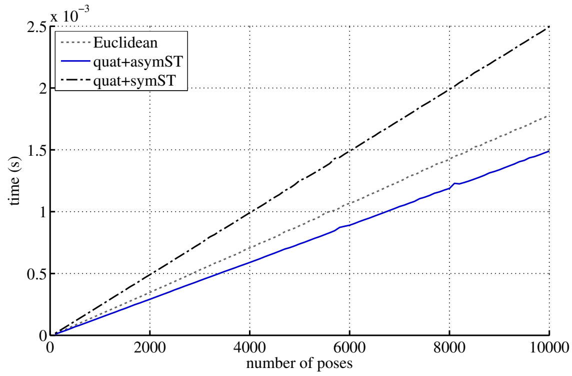


pose (scale, rotation and translation) of the object, a corresponding object identity number proposing which object is present, and a weight indicating the relative strength of the vote. Finally, the ground truth pose and object identity for each test set are also given.

The intrinsic Hough transform inference framework (Woodford et al, 2013) returns a list of putative objects' identity and pose, along with their relative likelihoods. Recognition and registration scores are then computed as per (Pham et al, 2011); to recap, the recogntion score is computed by comparing the identity of the most likely object with the ground truth identity, for all 1000 tests; the registration score is similarly computed, by checking that each of the following measures



(a) Less than 100 direct similarities.



(b) More than 100 direct similarities.

Fig. 5 Average mean computation times. The Euclidean distance and SRT divergences with quaternions yield fastest times since they have closed-form formulæ for the means. In (b), only divergences with closed-form means are considered.

of pose accuracy, on scale, rotation and translation respectively, is less than 1:

$$\tau_s(\mathbf{X},\mathbf{Y}) = 20\,|\ln \mathbf{X}_s - \ln \mathbf{Y}_s| \tag{63}$$

$$\tau_r(\mathbf{X},\mathbf{Y}) = \frac{12}{\pi}\arccos\left(\frac{\operatorname{trace}(\mathbf{X}_r^{\mathrm{T}}\mathbf{Y}_r)-1}{2}\right), \tag{64}$$

$$\tau_t(\mathbf{X},\mathbf{Y}) = 10\frac{\|\mathbf{X}_t - \mathbf{Y}_t\|}{\sqrt{\mathbf{X}_s\mathbf{Y}_s}}, \tag{65}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the pose of the most likely object of the ground truth identity and the ground truth pose respectively. In the case of an object having symmetries, there are multiple $\mathbf{Y}$s, and divergence to the closest is used.

7.2.2 Learning bandwidth parameters

Each divergence tested has a set of bandwidth parameters. Therefore, to compare these divergences fairly, we need to use suitable bandwidth parameters. To do this we learn a good set of bandwidth parameters for each divergence, using 40 training sets of votes (separate from the test sets) which are also provided (Pham et al, 2012), and find parameters which maximize the registration score on this training data. However, since the registration score is a discrete measure[13], we use the following real-valued, robust registration measure to allow standard gradient-based optimizers to be used:

$$E = \sum_i \rho\left(\left\|\begin{bmatrix}\tau_s(\boldsymbol{X}_i,\boldsymbol{Y}_i)\\ \tau_r(\boldsymbol{X}_i,\boldsymbol{Y}_i)\\ \tau_t(\boldsymbol{X}_i,\boldsymbol{Y}_i)\end{bmatrix}\right\|_\infty\right), \qquad \rho(x) = \begin{cases} x^2 & \text{if } |x| \le 1 \\ 1 + 2\ln x & \text{otherwise}\end{cases} \tag{66}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ are an estimate of the most likely pose for the ground truth object class and the ground truth pose for the i^{th} training set respectively. The particular robust kernel $\rho(\cdot)$ chosen, which we call the *log tail kernel*, is continuously differentiable and has a quadratic region for the training sets which meet registration criteria of equations (63)–(65), encouraging these to be modelled well, and a strictly increasing, strongly sublinear region for the other training sets, encouraging the resulting bandwidths to better model those training sets which are closer to meeting the registration criteria than those further away. The ensures broad convergence of the bandwidth training algorithm described next.

For a given set of bandwidth parameters, $\boldsymbol{X}_i$ can be computed using the test inference framework. E is then computed and locally minimized w.r.t. the bandwidth parameters, first using a coarse local search grid which moves towards the lowest value, then refined using Levenberg-Marquadt minimization. However, running the full inference framework at each iteration makes this process very slow. Instead, we make use of the fact that we know the ground truth pose of the training data, and approximate $\boldsymbol{X}_i$ by initializing it at $\boldsymbol{Y}_i$ and use one step of mean shift (Cheng, 1995) to move it to a more likely location under the probability density given by equation (58). In fact, since a single step of mean shift has been shown to be a step in the direction of steepest gradient, with a magnitude proportional to that of the steepest gradient (Cheng, 1995), minimizing E in this way is equivalent to minimizing the magnitude of the steepest gradient of the probability density function at the ground truth solution, i.e. encouraging the ground truth solution to be a local mode. This novel parameter learning strategy is reminiscent of contrastive divergence (Hinton, 2002).

The initial bandwidth parameters for all methods are set to 0.1; the final computed bandwidths are given in Table 4.

[13] Given 40 training sets, only 41 different registration scores are achievable.

Table 4 Learned values of kernel bandwidth.

Divergence type	Bandwidth		
	σ_s	σ_r	$\sigma/\sigma_t/\sigma_{st}$
Euclidean (σ)	–	–	0.226
Lie (σ)	–	–	0.273
extR + Poincaré (σ_r, σ_{st})	–	0.100	0.058
quat + Poincaré (σ_r, σ_{st})	–	0.146	0.173
intR + Poincaré (σ_r, σ_{st})	–	0.619	0.173
extR + asymST (σ_s, σ_r, σ_t)	0.100	0.360	0.100
quat + asymST (σ_s, σ_r, σ_t)	0.103	0.100	0.104
intR + asymST (σ_s, σ_r, σ_t)	0.101	0.360	0.100
extR + symST (σ_s, σ_r, σ_t)	0.101	0.358	0.118
quat + symST (σ_s, σ_r, σ_t)	0.100	0.144	0.144
intR + symST (σ_s, σ_r, σ_t)	0.103	0.613	0.144

Table 5 3D object recognition and registration results.

Divergence type	Scores (%)		Av. comp. time (µs/vote)	
	Recog.	Regis.	Divergence	Mean
Euclidean	44.8	62.0	1.92	**4.66**
Lie	53.3	67.0	18.4	6530
extR + Poincaré	41.4	55.1	2.11	75.7
quat + Poincaré	37.1	57.2	**1.28**	87.7
intR + Poincaré	36.9	51.3	4.57	159
extR + asymST	67.7	75.6	2.65	7.13
quat + asymST	69.0	**75.8**	1.50	9.28
intR + asymST	67.7	75.1	5.15	116
extR + symST	67.3	75.4	2.52	45.8
quat + symST	72.4	75.2	1.48	50.4
intR + symST	**73.4**	74.6	4.67	137

7.2.3 Results

The recognition and registration rates for each divergence on the 3D object recognition and registration task are summarized in Table 5. The results show that all six SRT divergences introduced in this paper perform significantly better at both the recognition and registration tasks, demonstrating the efficacy of the SRT divergence framework in a real application. Furthermore, the SRT divergence computation times are comparable with the fastest methods, while the mean computation times of the extR + asymST and quat + asymST SRT divergences are also comparable with the fastest methods, demonstrating that this improvement in performance need not come with a speed penalty.

8 Conclusions

In this article we have reviewed three families of divergences on direct similarities, ranging from extrinsic to intrinsic, and introduced a new family, SRT divergences. We have further proven a number of desirable properties of these divergences and their induced means, and have discovered closed forms for some of them.

The Euclidean distance and Euclidean means are simple, and fast to compute, but the distance is not left-invariant and the means are biased in scale due to rotations. The Lie divergence is left-invariant, but not a metric. We have developed closed forms for the Lie divergence in 2D and 3D; however, Lie means are estimated iteratively, making them very slow to compute. Any mean induced by any left-invariant distance is likewise slow to compute. In addition, means induced by the Lie divergence and all left-invariant distances are biased in scale due to translations.

The proposed SRT divergences, though not metric, are fast to compute, as are the means induced by them. In contrast to all the existing divergences evaluated, the asymmetric SRT divergence produces means which are completely unbiased. The fact that the asymmetric SRT divergence is not symmetric does not create any difficulty; one just has to take care as to the direction of the divergence.

Evaluating all these divergences on a challenging real-world application, 3D object recognition and registration, we demonstrate that SRT divergences provide a significant boost in performance, some of which do so with insignificant computation time penalty over existing methods.

A Proofs

A.1 Rotation Group $\mathcal{SO}(n)$

The rotation group $\mathcal{SO}(n)$ is the group of n-dimensional rotation matrices:

$$\mathcal{SO}(n) = \left\{ \mathbf{R} \in GL(n) : \mathbf{R}^{\mathrm{T}}\mathbf{R} = \boldsymbol{I}_n \wedge \det(\mathbf{R}) = 1 \right\}. \tag{67}$$

A number of known facts related to $\mathcal{SO}(n)$ are required in the proofs. They are summarized here.

Because rotation preserves the Euclidean norm, the eigenvalues of a rotation matrix $\mathbf{R}$ are unit complex numbers $e^{i\theta_k}$, for $\theta_k \in \mathbb{R}$ and $k = 1, \ldots, n$. Since $\mathbf{R}$ is a real matrix, both $e^{i\theta_k}$ and $e^{-i\theta_k}$ are eigenvalues of $\mathbf{R}$. The Lie algebra of $\mathcal{SO}(n)$, denoted by $\mathfrak{so}(n)$, contains skew-symmetric matrices

$$\mathfrak{so}(n) = \left\{ \mathbf{W} \in GL(n) : \mathbf{W}^{\mathrm{T}} = -\mathbf{W} \right\}. \tag{68}$$

The complex eigenvalues of $\mathbf{W}$ are 0 and complex conjugate pairs $\pm i\theta_k$ (i.e. logarithms of the eigenvalues of $\mathbf{R}$). The matrix exponential series in (4) and its inverse in (16) send points back and forth between $\mathcal{SO}(n)$ and $\mathfrak{so}(n)$. However, (16) converges only when $|\theta_k| < \pi$ for all k.

The space of skew-symmetric matrices is isomorphic to the space of bivectors in geometric algebra, which is a vector space. If we define a basis for the space, any skew-symmetric matrix can be represented compactly as a vector. Denote by $\hat{\boldsymbol{e}}_i$ a *single-entry* unit vector having a 1 at row i and 0 elsewhere, with size matching the context. Let single-entry matrices $\hat{\boldsymbol{E}}_{i,j}$ be defined as $\hat{\boldsymbol{E}}_{i,j} := \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_j^{\mathrm{T}}$. Consider $n_r := \frac{n(n-1)}{2}$ n-by-n matrices $(\boldsymbol{B}_k)_{k=1}^{n_r}$:

$$\boldsymbol{B}_k := (-1)^{i+j}(\hat{\boldsymbol{E}}_{i,j} - \hat{\boldsymbol{E}}_{j,i}), \tag{69}$$

where variables k and i, j are related by $k = n_r + 2 - j - (n-1-i/2)(i-1)$ with $1 \le i < j \le n$. Then, any skew-symmetric matrix $\mathbf{W} \in \mathfrak{so}(n)$ is uniquely represented as $\mathbf{W} = \mathbf{x}_{\underline{k}} \boldsymbol{B}_{\underline{k}}$ for some $\mathbf{x} \in \mathbb{R}^{n_r}$. Hereinafter, $\mathbf{W}_\times := \mathbf{x}$ denotes the vector representation of skew-symmetric matrix $\mathbf{W}$ via this map. Conversely, given a vector $\mathbf{x} \in \mathbb{R}^{n_r}$, $\mathbf{x}^\times$ denotes the skew-symmetric matrix $\mathbf{x}_{\underline{k}} \boldsymbol{B}_{\underline{k}}$. In fact, when $n = 3$, $\mathbf{x}^\times$ corresponds to the matrix representation of the cross product, i.e. for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, $\mathbf{x}^\times \mathbf{y} = \mathbf{x} \times \mathbf{y}$.

Consider the exponential map $\mathbf{x} \to \mathrm{e}^{\mathbf{x}^\times}$ that maps points in $\mathbb{R}^{n_r}$ to $\mathcal{SO}(n)$. Since $\frac{\partial \mathbf{x}^\times}{\partial \mathbf{x}_k} = \boldsymbol{B}_k$, the partial derivatives of the function $\mathrm{e}^{\mathbf{x}^\times}$ are given by:

$$\frac{\partial \mathrm{e}^{\mathbf{x}^\times}}{\partial \mathbf{x}_k} = \mathrm{e}^{\mathbf{x}^\times} \Psi_{\mathbf{x}}(\boldsymbol{B}_k), \tag{70}$$

where $\Psi_{\mathbf{x}} : \mathfrak{so}(n) \to \mathfrak{so}(n)$ satisfies (Hall, 2003, theorem 3.5):

$$\Psi_{\mathbf{x}}(\mathbf{V}) := \sum_{k=0}^{\infty} \frac{\mathrm{ad}^k_{-\mathbf{x}^\times}(\mathbf{V})}{(k+1)!}, \tag{71}$$

where $\mathrm{ad} : \mathfrak{so}(n) \times \mathfrak{so}(n) \to \mathfrak{so}(n)$ is a Lie algebra automorphism given by $\mathrm{ad}_{\mathbf{U}}(\mathbf{V}) = [\mathbf{U}, \mathbf{V}] := \mathbf{UV} - \mathbf{VU}$ and $[\cdot, \cdot]$ is called the commutator. We further denote by $\tilde{\Psi}_{\mathbf{x}}$ the matrix that represents $\Psi_{\mathbf{x}}(\cdot)$ in local coordinates, i.e. for all $\mathbf{v} \in \mathbb{R}^{n_r}$,

$$\tilde{\Psi}_{\mathbf{x}} \mathbf{v} = \Psi_{\mathbf{x}}(\mathbf{v}^\times)_\times. \tag{72}$$

Deriving an efficient form $\Psi_{\mathbf{x}}$ (and $\tilde{\Psi}_{\mathbf{x}}$ equivalently) in the general case $\mathcal{SO}(n)$ is not straightforward, which is to be addressed elsewhere. However, in case of $\mathcal{SO}(2)$ and $\mathcal{SO}(3)$, we can derive it by analyzing the commutator in what follows.

A.1.1 Rotation Group $\mathcal{SO}(2)$

The 2D rotation group is one-dimensional. Every $\mathbf{W} \in \mathfrak{so}(2)$ is written uniquely as $\theta \boldsymbol{B}_1$ for some $\theta \in \mathbb{R}$. Thus, every rotation $\mathbf{R} \in \mathcal{SO}(2)$ equals $\mathrm{e}^{\theta \boldsymbol{B}_1}$, and the commutator $[\cdot, \cdot]$ always vanishes. As a result, $\Psi_{\mathbf{x}}(\mathbf{V}) = \mathbf{V}$, leading to $\tilde{\Psi}_{\mathbf{x}} = \boldsymbol{I}_{n_r}$.

A.1.2 Rotation Group $\mathcal{SO}(3)$

The 3D rotation group is apparently more complicated. It has three dimensions. The eigenvalues of a rotation matrix $\mathbf{R} \in \mathcal{SO}(3)$ are $\{1, e^{i\theta}, e^{-i\theta}\}$, where $\mathrm{trace}\mathbf{R} = 2\cos\theta + 1$. Like in $\mathcal{SO}(2)$, $\ln\mathbf{R}$ converges when $|\theta| < \pi$, in which case $\ln \mathbf{R}$ is reduced to $\frac{0.5\theta}{\sin\theta}\left(\mathbf{R} - \mathbf{R}^{\mathrm{T}}\right)$. If instead $\mathbf{W} \in \mathfrak{so}(3)$ is given, the eigenvalues of $\mathbf{W}$ are $\{0, i\theta, -i\theta\}$ where $\theta = \|\mathbf{W}\|_{\mathrm{F}} / \sqrt{2}$, and $\mathrm{e}^{\mathbf{W}}$ is derived by Rodrigues' formula in (31).

According to the above-mentioned basis in $\mathfrak{so}(n)$, the three basis tangent vectors in $\mathfrak{so}(3)$ are $\boldsymbol{B}_1 = \hat{\boldsymbol{E}}_{3,2} - \hat{\boldsymbol{E}}_{2,3}$, $\boldsymbol{B}_2 = \hat{\boldsymbol{E}}_{1,3} - \hat{\boldsymbol{E}}_{3,1}$ and $\boldsymbol{B}_3 = \hat{\boldsymbol{E}}_{2,1} - \hat{\boldsymbol{E}}_{1,2}$. They satisfy: $[\boldsymbol{B}_1, \boldsymbol{B}_2] = \boldsymbol{B}_3$, $[\boldsymbol{B}_2, \boldsymbol{B}_3] = \boldsymbol{B}_1$ and $[\boldsymbol{B}_3, \boldsymbol{B}_1] = \boldsymbol{B}_2$. Based on these equations, we derive a closed form for computing the commutator in local coordinates:

$$[\mathbf{x}^\times, \mathbf{y}^\times]_\times = \mathbf{x}^\times \mathbf{y} = \mathbf{x} \times \mathbf{y}. \tag{73}$$

Since $\mathrm{ad}_{-\mathbf{x}^\times}(\mathbf{V}) = [-\mathbf{x}^\times, \mathbf{V}] = [\mathbf{V}, \mathbf{x}^\times]$ for all skew-symmetric matrices $\mathbf{V}$, and $(\mathbf{x}^\times)^3 = -\|\mathbf{x}\|^2 \mathbf{x}^\times$, applying (73) to (71), we obtain for all $\mathbf{v} \in \mathbb{R}^3$:

$$\Psi_{\mathbf{x}}(\mathbf{v}^\times)_\times = \mathbf{v} - \mathbf{x}^\times \mathbf{v} \frac{1 - \cos\theta}{\theta^2} + (\mathbf{x}^\times)^2 \mathbf{v} \frac{\theta - \sin\theta}{\theta^3}. \tag{74}$$

Since by definition, $\tilde{\Psi}_{\mathbf{x}} \mathbf{v} = \Psi_{\mathbf{x}}(\mathbf{v}^\times)_\times$, this leads to:

$$\tilde{\Psi}_{\mathbf{x}} = \boldsymbol{I}_3 - \mathbf{x}^\times \frac{1 - \cos\theta}{\theta^2} + (\mathbf{x}^\times)^2 \frac{\theta - \sin\theta}{\theta^3}. \tag{75}$$

We note that $\tilde{\Psi}_{\mathbf{x}}^{\mathrm{T}} = \tilde{\Psi}_{-\mathbf{x}}$ and:

$$\tilde{\Psi}_{\mathbf{x}}^{\mathrm{T}} \tilde{\Psi}_{\mathbf{x}} = \boldsymbol{I}_3 + (\mathbf{x}^\times)^2 \frac{\theta^2 + 2\cos\theta - 2}{\theta^4}. \tag{76}$$

A.2 Proofs for Lemma 1

The objective function to minimize in (6) is rewritten as:

$$\begin{aligned} \mathcal{E}(\mathbf{X}) &= \mathsf{w}_{\underline{i}} d_{\mathrm{E}}(\boldsymbol{X}_{\underline{i}}, \mathbf{X})^2 \\ &= \mathsf{w}_{\underline{i}} \left\| \boldsymbol{X}_{\underline{i};s} \boldsymbol{X}_{\underline{i};r} - \mathbf{X}_s \mathbf{X}_r \right\|_{\mathrm{F}}^2 + \mathsf{w}_{\underline{j}} \left\| \boldsymbol{X}_{\underline{j};t} - \mathbf{X}_t \right\|^2. \end{aligned} \tag{77}$$

Minimizing the second term with respect to $\mathbf{X}_t$ gives (12). Hence, the problem becomes finding $\mathbf{X}_s$ and $\mathbf{X}_r$ that minimize the first term. To do this, we define $\mathsf{v}_i := \mathsf{w}_i \mathbf{X}_s^2$ and $\boldsymbol{V}_i := \boldsymbol{X}_{i;s} \boldsymbol{X}_{i;r} / \mathbf{X}_s$ for all $i = 1, \ldots, N$, and rewrite the first term as:

$$E'(\mathbf{X}) := \mathsf{v}_{\underline{i}} \left\| \boldsymbol{V}_{\underline{i}} - \mathbf{X}_r \right\|_{\mathrm{F}}^2. \tag{78}$$

The idea is to find $\mathbf{X}_r$ that minimizes $E'(\mathbf{X})$ given $\mathbf{X}_s$ first, and then to use the resultant formula to find $\mathbf{X}_s$. Since $\mathbf{X}_r$ is a rotation matrix, minimizing $E'(\mathbf{X})$ with respect to $\mathbf{X}_r$ has been solved in (Downs, 1972; Sibson, 1979). It is analogous to the classical orthogonal Procrustes problem which seeks the orthogonal matrix that most closely transforms a given matrix to a second one. The minimizer for $\mathsf{v}_{\underline{i}} \left\| \boldsymbol{V}_{\underline{i}} - \mathbf{X}_r \right\|_{\mathrm{F}}^2$ is given by:

$$\operatorname*{argmin}_{\mathbf{X}_r \in \mathcal{SO}(n)} \mathsf{v}_{\underline{i}} \left\| \boldsymbol{V}_{\underline{i}} - \mathbf{X}_r \right\|_{\mathrm{F}}^2 = \mathrm{sop}(\mathsf{v}_{\underline{j}} \boldsymbol{V}_{\underline{j}}). \tag{79}$$

Function $\mathrm{sop}(\cdot)$ is invariant to direct dilation, i.e. $\mathrm{sop}(\mathbf{X}) = \mathrm{sop}(a\mathbf{X})$ for any $a > 0$ (Downs, 1972; Sibson, 1979). This gives us a formula for finding $\overline{\mathbf{X}}_r$ independently from $\overline{\mathbf{X}}_s$:

$$\begin{aligned} \overline{\mathbf{X}}_r &= \mathrm{sop}(\mathsf{v}_{\underline{i}} \boldsymbol{V}_{\underline{i}}) = \mathrm{sop}(\mathsf{w}_{\underline{i}} \mathbf{X}_s^2 \boldsymbol{X}_{\underline{i};s} \boldsymbol{X}_{\underline{i};r} / \mathbf{X}_s) \\ &= \mathrm{sop}(\mathsf{w}_{\underline{i}} \boldsymbol{X}_{\underline{i};s} \boldsymbol{X}_{\underline{i};r}), \end{aligned} \tag{80}$$

which is (11).

Given that we have found $\overline{\mathbf{X}}_r$ without knowing $\overline{\mathbf{X}}_s$, we substitute $\overline{\mathbf{X}}_r$ back to (78), further reducing the problem to minimizing $E'(\mathbf{X})$ with respect to $\mathbf{X}_s$, which is a quadratic minimization problem, to which the unique solution is given in (10).

A.3 Matrix Exponential and Logarithm in $\mathcal{DS}(n)$

For a short hand notation, we write $\xi(a\boldsymbol{I}_n + \mathbf{W})$ as $\boldsymbol{E}_n$. Additionally, the quotient operator is overloaded to denote that $\frac{\mathbf{A}}{\mathbf{B}} := \mathbf{B}^{-1}\mathbf{A} = \mathbf{A}\mathbf{B}^{-1}$ if square matrices $\mathbf{A}$ and $\mathbf{B}$ commute and $\mathbf{B}$ is invertible. To derive a closed form for $\mathrm{e}^{\mathbf{Y}}$, we derive a closed form for $\boldsymbol{E}_n$ and use basic facts about $\mathcal{SO}(n)$ are summarized in §A.1. One way to find a closed form for $\boldsymbol{E}_n$ is to notice that $\mathbf{Z}\xi(\mathbf{Z}) = \xi(\mathbf{Z})\mathbf{Z} = e^{\mathbf{Z}} - \boldsymbol{I}_n$, and obtain

$$\boldsymbol{E}_n = \frac{e^a e^{\mathbf{W}} - \boldsymbol{I}_n}{a\boldsymbol{I}_n + \mathbf{W}} \tag{81}$$

if $a\boldsymbol{I}_n + \mathbf{W}$ is invertible. Since the eigenvalues of $\mathbf{W}$ are 0 and complex conjugate pairs $\pm i\theta_k$ (see §A.1), it occurs when $a \neq 0$.

Conversely, since $(\boldsymbol{I}_n - \mathbf{Z})\eta(\mathbf{Z}) = \eta(\mathbf{Z})(\boldsymbol{I}_n - \mathbf{Z}) = -\ln \mathbf{Z}$,

$$\eta(s\mathbf{R}) = \frac{(\ln s)\boldsymbol{I}_n + \ln \mathbf{R}}{s\mathbf{R} - \boldsymbol{I}_n} \tag{82}$$

if $s\mathbf{R} - \boldsymbol{I}_n$ is invertible. Similarly, since the eigenvalues of $\mathbf{R}$ are 1 and $e^{\pm i\theta_k}$ (see §A.1), the eigenvalues of $s\mathbf{R} - \boldsymbol{I}_n$ are $s-1$ and $se^{\pm i\theta_k} - 1$. Hence, $s\mathbf{R} - \boldsymbol{I}_n$ is invertible unless $s = 1$. We write $\eta(s\mathbf{R})$ as $\boldsymbol{L}_n$ hereinafter.

In theory, we can use equations (81) and (82) in computing matrix exponential and logarithm. However, this approach involves a matrix inversion operation which is costly to compute and it only works when the numerator matrix is invertible. More importantly, the forms become numerically unstable when one of the eigenvalues of $s\mathbf{R} - \boldsymbol{I}_n$ is close to zero.

In what follows, we further simplify (81) and (82) to forms which do not involve matrix inversion when $n = 2, 3$, leading to closed forms for $e^{\mathbf{Y}}$ and $\ln\mathbf{X}$ in $\mathcal{DS}(2)$ and $\mathcal{DS}(3)$. It is possible to generalize the work of Gallier and Xu (2002) to find an efficient method for computing $e^{\mathbf{Y}}$ and $\ln\mathbf{X}$ in $\mathcal{DS}(n)$ with $n \geq 4$, but the work is much more difficult, requiring solving an inverse problem for each computation, thereby is out of this paper's scope.

Consider a real diagonalizable d-by-d matrix $\mathbf{Z}$ for some integer d. There is an efficient approach to compute $\mathbf{Z}^k$. Using eigen-value decomposition, we diagonalize $\mathbf{Z} = \mathbf{Q}\text{diag}(\mathsf{w}_1, \ldots, \mathsf{w}_d)\mathbf{Q}^{\text{H}}$ where $(\mathsf{w}_i)_{i=1}^d$ are complex eigenvalues, $\mathbf{Q}^{\text{H}}$ is the conjugate transpose of $\mathbf{Q}$, and $\mathbf{Q}$ is a unitary matrix, with each column $\mathbf{Q}_{:,i}$ being an eigenvector corresponding to the eigenvalue w_i. Then $\mathbf{Z}^k = \mathbf{Q}\text{diag}(\mathsf{w}_1^k, \ldots, \mathsf{w}_d^k)\mathbf{Q}^{\text{H}}$.

The approach can be generalized to computing a matrix series. Let $f(z) = \sum_{k=0}^{\infty} \mathsf{a}_k z^k$ be a series over a complex number z. Let $\mathbf{F}(\mathbf{Z}) = \sum_{k=0}^{\infty} \mathsf{a}_k \mathbf{Z}^k$ be its matrix version. If $\mathbf{Z}$ is diagonalizable then $\mathbf{F}(\mathbf{Z}) = \mathbf{Q}\text{diag}(f(\mathsf{w}_1), \ldots, f(\mathsf{w}_d))\mathbf{Q}^{\text{H}}$.

We notice that the complex version of $\xi(\cdot)$ in (19) leads to a closed form,

$$\xi(z) = \sum_{k=0}^{\infty} \frac{z^k}{(k+1)!} = \frac{e^z - 1}{z}, \tag{83}$$

which leads to obtaining $\boldsymbol{E}_n = \xi(a\boldsymbol{I}_n + \mathbf{W})$ via diagonalizing $a\boldsymbol{I}_n + \mathbf{W}$. Similarly, the complex version of $\eta(\cdot)$ in (21) leads to another closed form,

$$\eta(z) = \sum_{k=0}^{\infty} \frac{(1-z)^k}{k+1} = \frac{\ln z}{z-1}, \tag{84}$$

also suggesting us to obtain $\boldsymbol{L}_n = \eta(s\mathbf{R})$ (if it converges) via diagonalizing $s\mathbf{R}$.

We derive closed forms for $e^{\mathbf{Y}}$ and $\ln\mathbf{X}$ in $\mathcal{DS}(2)$ using this approach. The same idea is used with some extra work to derive closed forms for $e^{\mathbf{Y}}$ and $\ln\mathbf{X}$ in $\mathcal{DS}(3)$. The convergence condition for $\ln\mathbf{X}$ turns out to be the same as that for $\ln\mathbf{R}$ (with $\mathbf{R} \in \mathcal{SO}(n)$ and $n = 2, 3$): the rotation angle does not exceed 180°.

A.3.1 Matrix Exponential and Logarithm in $\mathcal{DS}(2)$

Let $\mathbf{Y} = M(a\boldsymbol{I}_2 + \mathbf{W}, \mathbf{u}) \in \mathfrak{ds}(2)$ be the matrix whose matrix exponential is to be computed. Since $\mathbf{W}$ is a 2-by-2 skew-symmetric matrix, $\mathbf{W} = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$ for some $\theta \in \mathbb{R}$. We rewrite $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{Q}\text{diag}(i\theta, -i\theta)\mathbf{Q}^{\text{H}}, \tag{85}$$

where $\mathbf{Q} = (\mathbf{v}, \bar{\mathbf{v}})$, $\mathbf{v} = \frac{1}{\sqrt{2}}(1, -i)^{\text{T}}$, and $\bar{\mathbf{v}} = \frac{1}{\sqrt{2}}(1, i)^{\text{T}}$ is the complex conjugate of $\mathbf{v}$, leading to $a\boldsymbol{I}_2 + \mathbf{W} = \mathbf{Q}\text{diag}(a + i\theta, a - i\theta)\mathbf{Q}^{\text{H}}$. Thus, substituting this equation to (19) and using (83) and to simplify the series, we get:

$$\begin{aligned} \boldsymbol{E}_2 &= \mathbf{Q}\text{diag}(\xi(a+i\theta), \xi(a-i\theta))\mathbf{Q}^{\text{H}} \\ &= \xi(a+i\theta)\mathbf{v}\mathbf{v}^{\text{H}} + \xi(a-i\theta)\bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}. \end{aligned} \tag{86}$$

Directly calculating the real part and the imaginary part of $\xi(a+i\theta) = \frac{e^{a+i\theta}-1}{a+i\theta}$ gives $\xi(a+i\theta) =: \xi_{\mathfrak{r}} + i\xi_{\mathfrak{i}}$, where $\xi_{\mathfrak{r}}$ and $\xi_{\mathfrak{i}}$ are given in (24) and (25) respectively. Since $\xi(a+i\theta)$ is the complex conjugate of $\xi(a-i\theta)$, it follows that

$$\boldsymbol{E}_2 = \xi_{\mathfrak{r}}\left(\mathbf{v}\mathbf{v}^{\text{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}\right) + i\xi_{\mathfrak{i}}\left(\mathbf{v}\mathbf{v}^{\text{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}\right). \tag{87}$$

By the definition of $\mathbf{v}$ and $\bar{\mathbf{v}}$, $\mathbf{v}\mathbf{v}^{\text{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}} = \hat{\boldsymbol{E}}_{1,1} + \hat{\boldsymbol{E}}_{2,2}$ and $\mathbf{v}\mathbf{v}^{\text{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}} = i(\hat{\boldsymbol{E}}_{1,2} - \hat{\boldsymbol{E}}_{2,1})$ ($\hat{\boldsymbol{E}}_{i,j}$ are defined in §A.1). This gives a closed form for $\mathrm{e}^{\mathbf{Y}}$, as shown in (23).

Similarly, given $\mathbf{X} = m(s\mathbf{R}, \mathbf{t}) \in \mathcal{DS}(2)$, to derive a closed form for $\ln\mathbf{X}$, we need a closed form for $\boldsymbol{L}_2 = \eta(s\mathbf{R})$. Let $\mathbf{W} := \ln\mathbf{R}$ be the principal matrix logarithm of rotation matrix $\mathbf{R}$. Suppose we have diagonalized $\mathbf{W}$ as above. Taking the matrix exponential of $\mathbf{W}$ via the expanded version in (85), we obtain $\mathbf{R} = \mathbf{Q}\text{diag}(e^{i\theta}, e^{-i\theta})\mathbf{Q}^{\text{H}}$. This leads to a closed form for computing θ from $\mathbf{R}$: $\theta = \arctan(\mathbf{R}_{2,1}/\mathbf{R}_{1,1})$. More importantly, as we substitute this diagonalized form of $\mathbf{R}$ to (21) using (84) to simply the series, we get:

$$\begin{aligned} \boldsymbol{L}_2 &= \mathbf{Q}\text{diag}(\eta(se^{i\theta}), \eta(se^{-i\theta}))\mathbf{Q}^{\text{H}} \\ &= \eta(se^{i\theta})\mathbf{v}\mathbf{v}^{\text{H}} + \eta(se^{-i\theta})\bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}. \end{aligned} \tag{88}$$

By directly calculating the real part and the imaginary part of $\eta(se^{i\theta}) = \frac{\ln s + i\theta}{se^{i\theta}-1} =: \eta_{\mathfrak{r}} + i\eta_{\mathfrak{i}}$, we get closed forms for $\eta_{\mathfrak{r}}$ and $\eta_{\mathfrak{i}}$ as shown in (28) and (29) respectively. Separating the real part of $\boldsymbol{L}_2$ from the imaginary part of $\boldsymbol{L}_2$, we also have

$$\boldsymbol{L}_2 = \eta_{\mathfrak{r}}(\mathbf{v}\mathbf{v}^{\text{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}) + i\eta_{\mathfrak{i}}(\mathbf{v}\mathbf{v}^{\text{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}), \tag{89}$$

using the same argument as above. Therefore, we have obtained a closed form for $\boldsymbol{L}_2$ as shown in (27), and then a closed form for $\ln\mathbf{X}$ as shown in (26).

A.3.2 Matrix Exponential and Logarithm in $\mathcal{DS}(3)$

Finding closed forms for matrix exponential and logarithm in $\mathcal{DS}(3)$ requires more work. Let $\mathbf{Y} = M(a\boldsymbol{I}_3 + \mathbf{W}, \mathbf{u}) \in \mathfrak{ds}(3)$. Suppose $\mathbf{W}$ is decomposed into $\mathbf{W} = \mathbf{Q}\text{diag}(0, i\theta, -i\theta)\mathbf{Q}^{\text{H}}$ (see §A.1) where $\mathbf{Q} = (\mathbf{n}, \mathbf{v}, \bar{\mathbf{v}})$, $\mathbf{n}$ is the normal vector representing the axis of rotation, and $\mathbf{v}$ and $\bar{\mathbf{v}}$ are a pair complex conjugate unit vectors. We can rewrite $\mathbf{W}$ as:

$$\mathbf{W} = i\theta\mathbf{v}\mathbf{v}^{\text{H}} - i\theta\bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}}. \tag{90}$$

Dividing both sides of (90) by $i\theta$ and squaring up the result (noticing that $\mathbf{v}$ and $\bar{\mathbf{v}}$ are orthonormal to each other) yields:

$$\mathbf{v}\mathbf{v}^{\text{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}} = \frac{\mathbf{W}}{i\theta}, \tag{91}$$

$$\mathbf{v}\mathbf{v}^{\text{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}} = \frac{\mathbf{W}^2}{-\theta^2}. \tag{92}$$

Since $(\mathbf{n}, \mathbf{v}, \bar{\mathbf{v}})$ form an orthonormal basis, we have:

$$\mathbf{n}\mathbf{n}^{\text{T}} + \mathbf{v}\mathbf{v}^{\text{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\text{H}} = \boldsymbol{I}_3, \tag{93}$$

which leads to

$$\mathbf{n}\mathbf{n}^{\text{T}} = \boldsymbol{I}_3 + \frac{\mathbf{W}^2}{\theta^2}, \tag{94}$$

by substituting (92) to it.

Diagonalizing $\boldsymbol{E}_3 = \xi(a\boldsymbol{I}_3 + \mathbf{W})$, we obtain

$$\boldsymbol{E}_3 = \xi(a)\mathbf{n}\mathbf{n}^{\mathrm{T}} + \xi(a+i\theta)\mathbf{v}\mathbf{v}^{\mathrm{H}} + \xi(a-i\theta)\bar{\mathbf{v}}\bar{\mathbf{v}}^{\mathrm{H}}. \tag{95}$$

Using the same argument as in the case of $\mathcal{DS}(2)$ for deriving $\boldsymbol{E}_2$, it follows that

$$\boldsymbol{E}_3 = \frac{e^a - 1}{a}\mathbf{n}\mathbf{n}^{\mathrm{T}} + \xi_{\mathfrak{r}}(\mathbf{v}\mathbf{v}^{\mathrm{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\mathrm{H}}) + i\xi_{\mathfrak{i}}(\mathbf{v}\mathbf{v}^{\mathrm{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\mathrm{H}}), \tag{96}$$

where $\xi_{\mathfrak{r}}$ and $\xi_{\mathfrak{i}}$ are defined in (24) and (25) respectively. Substituting (90), (91) and (92) to (96), we obtain a closed form for $\boldsymbol{E}_3$, i.e. (32), leading to a closed form for $\mathrm{e}^{\mathbf{Y}}$, i.e. (30).

Given $\mathbf{X} = m(s\mathbf{R}, \mathbf{t}) \in \mathcal{DS}(3)$, finding a closed form for $\ln \mathbf{X}$ is done similarly. The function $\boldsymbol{L}_3 = \eta(s\mathbf{R})$ gives

$$\boldsymbol{L}_3 = \eta(s)\mathbf{n}\mathbf{n}^{\mathrm{T}} + \eta_{\mathfrak{r}}(\mathbf{v}\mathbf{v}^{\mathrm{H}} + \bar{\mathbf{v}}\bar{\mathbf{v}}^{\mathrm{H}}) + i\eta_{\mathfrak{i}}(\mathbf{v}\mathbf{v}^{\mathrm{H}} - \bar{\mathbf{v}}\bar{\mathbf{v}}^{\mathrm{H}}), \tag{97}$$

where $\eta_{\mathfrak{r}}$ and $\eta_{\mathfrak{i}}$ are defined in (28) and (29) respectively. Substituting (94), (91) and (92) to (97), taking into account that $\eta(s) = \frac{\ln s}{s-1}$, we obtain a closed form for $\boldsymbol{L}_3$, (35). A closed form for $\ln\mathbf{X}$ follows, (33).

It can be verified that our closed forms for $e^{\mathbf{Y}}$ and $\ln \mathbf{X}$ are generalizations of those derived for $\mathcal{SE}(3)$ presented in (Agrawal, 2006), i.e. when $s = 1$.

A.4 Proof for Theorem 1

Note that in this theorem we only consider input sets $\{(\boldsymbol{X}_i, \mathsf{w}_i) \in \mathcal{DS}(n) \times \mathbb{R}^+\}_{i=1}^N$ for which the mean is unique. A recent study from Arnaudon and Miclo (2014) shows that in a complete manifold including $\mathcal{DS}(n)$, the mean induced by the Riemannian distance d_{R} via (7) is almost surely unique.

The sum of squared divergences in (6) under d_{L} is expressed as:

$$\mathcal{E}(\mathbf{X}) = \mathsf{w}_{\underline{i}} \left\|\ln(\boldsymbol{X}_{\underline{i}}^{-1}\mathbf{X})\right\|_{\mathrm{F}}^2. \tag{98}$$

We rely on the minimal representation ϕ in §2.1 to derive the proof. Let $l_{\mathbf{Z}}(\mathbf{x}) := \phi^{-1} \circ L_{\mathbf{Z}} \circ \phi(\mathbf{x})$ be the equivalence of the left translation operator $L_{\mathbf{Z}}$ for some $\mathbf{Z} \in \mathcal{DS}(n)$ under ϕ. By inspection, we establish the following equations:

$$l_{\mathbf{Z}}(\mathbf{x})_s = \ln\mathbf{Z}_s + \mathbf{x}_s, \tag{99}$$

$$l_{\mathbf{Z}}(\mathbf{x})_r = \ln(\mathbf{Z}_r \mathrm{e}^{\mathbf{x}_r^{\times}})_{\times}, \tag{100}$$

$$l_{\mathbf{Z}}(\mathbf{x})_t = \mathbf{Z}_s\mathbf{Z}_r\mathbf{x}_t + \mathbf{Z}_t. \tag{101}$$

In this section only, we work on the *inverses* $\boldsymbol{X}_i^{-1}$ instead of $\boldsymbol{X}_i$ themselves so that we can expand $\ln(\boldsymbol{X}_i^{-1}\mathbf{X})$ easily. Let $m(\mathsf{s}_i\boldsymbol{R}_i, \boldsymbol{t}_i) := \boldsymbol{X}_i^{-1}$ for all $i = 1, \ldots, N$. Via ϕ, using (20), (99) to (101) and the fact $\left\langle \boldsymbol{I}_n, \ln(\boldsymbol{R}_i\mathrm{e}^{\mathbf{x}_r^{\times}}) \right\rangle = 0$ since the diagonal part of any skew-symmetric matrix is zero, (98) expands to:

$$\begin{aligned}\mathcal{E}(\phi(\mathbf{x})) = {} & n\mathsf{w}_{\underline{i}}(\ln\mathsf{s}_{\underline{i}} + \mathbf{x}_s)^2 + \mathsf{w}_{\underline{i}}\left\|\ln(\boldsymbol{R}_{\underline{i}}\mathrm{e}^{\mathbf{x}_r^{\times}})\right\|_{\mathrm{F}}^2 \\ & + \mathsf{w}_{\underline{i}}\left\|\eta(\boldsymbol{Z}_{\underline{i}}(\mathbf{x}))\boldsymbol{z}_{\underline{i}}(\mathbf{x})\right\|^2,\end{aligned} \tag{102}$$

where $\boldsymbol{Z}_i(\mathbf{x}) := \mathsf{s}_i\mathrm{e}^{\mathbf{x}_s}\boldsymbol{R}_i\mathrm{e}^{\mathbf{x}_r^{\times}}$ and $\boldsymbol{z}_i(\mathbf{x}) := \mathsf{s}_i\boldsymbol{R}_i\mathbf{x}_t + \boldsymbol{t}_i$ for all $i = 1, \ldots, N$, and $\eta(\cdot)$ is defined in (21).

A.4.1 Non-translation-compatibility

Without loss of generality, we set $(\mathsf{s}_i, \boldsymbol{R}_i) = (\bar{s}, \bar{\mathbf{R}})$ for all $i = 1, \ldots, N$ and some constant $\bar{s} > 0$ and $\bar{\mathbf{R}} \in \mathcal{SO}(n)$. Instead of finding the mean $\overline{\mathbf{X}}$, we prove that for any $\mathbf{t} \in \mathbb{R}^n$, $m\left(\bar{s}^{-1}\bar{\mathbf{R}}^{\mathrm{T}}, \mathbf{t}\right)$ cannot be a mean (note that we are working with $\boldsymbol{X}_i^{-1}$). If this is the case, the actual mean(s) cannot be translation-compatible.

Differentiating (102) with respect to a variable $\mathbf{x}_s$ yields:

$$\frac{\partial E \circ \phi}{\partial \mathbf{x}_s} = 2n\mathsf{w}_{\underline{i}}(\mathbf{x}_s + \ln\mathsf{s}_{\underline{i}}) + \mathsf{w}_{\underline{i}}\boldsymbol{z}_{\underline{i}}(\mathbf{x})^{\mathrm{T}}\eta_s'(\boldsymbol{Z}_{\underline{i}}(\mathbf{x}))\boldsymbol{z}_{\underline{i}}(\mathbf{x}), \tag{103}$$

where $\eta_s'(\mathbf{Z}_i(\mathbf{x}))$ is given by (omitting the variable $\mathbf{x}$):

$$\eta_s'(\boldsymbol{Z}_i) := \frac{\partial \eta}{\partial \mathbf{x}_s}(\boldsymbol{Z}_i)^{\mathrm{T}}\eta(\boldsymbol{Z}_i) + \eta(\boldsymbol{Z}_i)^{\mathrm{T}}\frac{\partial \eta}{\partial \mathbf{x}_s}(\boldsymbol{Z}_i). \tag{104}$$

Since $\frac{\partial \boldsymbol{Z}_i}{\partial \mathbf{x}_s}(\mathbf{x}) = \boldsymbol{Z}_i(\mathbf{x})$ for all $i = 1, \ldots, N$, we get:

$$\begin{aligned}\frac{\partial \eta \circ \boldsymbol{Z}_i}{\partial \mathbf{x}_s} &= \frac{\partial}{\partial \mathbf{x}_s}\sum_{k=1}^{\infty}\frac{(\boldsymbol{I}_n - \boldsymbol{Z}_i)^k}{k+1} \\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{k}\frac{-1}{k+1}(\boldsymbol{I}_n - \boldsymbol{Z}_i)^{l-1}\boldsymbol{Z}_i(\boldsymbol{I}_n - \boldsymbol{Z}_i)^{k-l}. \\ &= \sum_{k=1}^{\infty}\frac{-k}{k+1}(\boldsymbol{I}_n - \boldsymbol{Z}_i)^{k-1}\boldsymbol{Z}_i.\end{aligned} \tag{105}$$

where the last equation holds because $\boldsymbol{Z}_i$ commutes with $(\boldsymbol{I}_n - \boldsymbol{Z}_i)^l$ for all integer l.

Let $\tilde{\mathbf{x}} := \phi^{-1}\left(m\left(\bar{s}^{-1}\bar{\mathbf{R}}^{\mathrm{T}}, \mathbf{t}\right)\right)$ for an arbitrary translation $\mathbf{t} \in \mathbb{R}^n$. By definition, $\boldsymbol{Z}_i(\tilde{\mathbf{x}}) = \boldsymbol{I}_n$ for all $i = 1, \ldots, N$. At point $\mathbf{x} = \tilde{\mathbf{x}}$, we apply (105) to find the derivative of $\eta(\boldsymbol{Z}_i)$, and (21) to evaluate $\eta(\boldsymbol{Z}_i)$ itself, we get:

$$\eta \circ \boldsymbol{Z}_i(\tilde{\mathbf{x}}) = \boldsymbol{I}_n, \tag{106}$$

$$\frac{\partial \eta \circ \boldsymbol{Z}_i}{\partial \mathbf{x}_s}(\tilde{\mathbf{x}}) = -0.5\boldsymbol{Z}_i = -0.5\boldsymbol{I}_n. \tag{107}$$

It follows from (104) that:

$$\eta_s'(\boldsymbol{Z}_i(\tilde{\mathbf{x}})) = -\boldsymbol{I}_n. \tag{108}$$

Substituting the equation back to (103), we obtain:

$$\frac{\partial E \circ \phi}{\partial \mathbf{x}_s}(\tilde{\mathbf{x}}) = -\mathsf{w}_{\underline{i}}\boldsymbol{z}_{\underline{i}}(\tilde{\mathbf{x}})^{\mathrm{T}}\boldsymbol{z}_{\underline{i}}(\tilde{\mathbf{x}}) = -\mathsf{w}_{\underline{i}}\left\|\boldsymbol{z}_{\underline{i}}(\tilde{\mathbf{x}})\right\|^2. \tag{109}$$

Clearly, the right-hand side of the above equation is always negative. Because the partial derivative $\frac{\partial E \circ \phi}{\partial \mathbf{x}_s}$ at $\tilde{\mathbf{x}}$ does not vanish, no direct similarity $m\left(\bar{s}^{-1}\bar{\mathbf{R}}^{\mathrm{T}}, \mathbf{t}\right)$ can be a mean, proving the actual mean is not translation-compatible.

A.4.2 Scale-compatibility and rotation-compatibility

To verify scale-compatibility and rotation-compatibility, we set $\boldsymbol{t}_i := -\mathsf{s}_i^{-1}\boldsymbol{R}_i^{\mathrm{T}}\bar{\mathbf{t}}$ for all $i = 1, \ldots, N$ and some constant $\bar{\mathbf{t}} \in \mathbb{R}^n$ (so that all $\boldsymbol{X}_i$'s translation components equal $\bar{\mathbf{t}}$).

Differentiating $E \circ \phi(\mathbf{x})$ with respect to $\mathbf{x}_t$ yields:

$$\frac{\partial E \circ \phi}{\partial \mathbf{x}_t} = 2\mathsf{w}_{\underline{i}}\boldsymbol{z}_{\underline{i}}^{\mathrm{T}}\eta(\boldsymbol{Z}_{\underline{i}})^{\mathrm{T}}\eta(\boldsymbol{Z}_{\underline{i}})\mathsf{s}_{\underline{i}}\boldsymbol{R}_{\underline{i}}. \tag{110}$$

It immediately follows that the derivative $\frac{\partial E \circ \phi}{\partial \mathbf{x}_t}$ only vanishes when $\mathbf{x}_t = \bar{\mathbf{t}}$, at which point the third term of (102) also vanishes. When this occurs, $\overline{\mathbf{X}}_s$ becomes a geometric mean, and $\overline{\mathbf{X}}_r$ becomes the mean of rotations under the intrinsic Riemannian distance (Park and Ravani, 1997). Hence, the scale-compatibility and rotation-compatibility properties are verified.

A.5 Proof for Theorem 2

We will prove the third statement, i.e. for translation-compatibility. The other two statements follow analogously.

Suppose all direct similarities are written as $\boldsymbol{x}_i := \phi^{-1}(\boldsymbol{X}_i)$, and the mean of them $\bar{\mathbf{X}}$ is written as $\bar{\mathbf{x}} := \phi^{-1}(\bar{\mathbf{X}})$ under the map ϕ. Denote by $\mathbf{x}_{sr}$ the first $n_r + 1$ components of $\mathbf{x}$. Translation compatibility means that if for all $i = 1, \ldots, N$, $\boldsymbol{x}_{i;sr} = \mathbf{t}'$ for some constant $\mathbf{t}' \in \mathbb{R}^{n_r+1}$, then $\bar{\mathbf{x}}_{sr} = \mathbf{t}'$. We will prove that the following statements are equivalent to each other.

– **A**: every unique mean induced by g is translation-compatible,
– **B**: $\mathcal{T}(n)$ is totally geodesic in $\mathcal{DS}(n)$.

A.5.1 From A to B

Choose any two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n_{ds}}$ such that $\mathbf{x}_{sr} = \mathbf{y}_{sr} = \mathbf{0}$ and that $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$ are within the injectivity radius of each other. Suppose $\gamma(u)$ is the g-geodesic with the arc-length parameterization going from $\gamma(0) = \phi(\mathbf{x})$ to $\gamma(a) = \phi(\mathbf{y})$ for some constant $a > 0$. It suffices to show that $\phi \circ \gamma(u)_i = 0$ for all $u \in [0, a]$ and all $i \in \mathcal{J}_s \cup \mathcal{J}_r$.

Suppose there is a number $\mathsf{u}_0 \in (0, a)$ and a dimension $i \in \mathcal{J}_s \cup \mathcal{J}_r$ such that $\phi \circ \gamma(\mathsf{u}_0)_i \neq 0$. Let $\mathsf{u}_1 < \mathsf{u}_0$ be the largest number such that $\phi \circ \gamma(\mathsf{u}_1)_i = 0$ and $\mathsf{u}_2 > \mathsf{u}_0$ be the smallest number such that $\phi \circ \gamma(\mathsf{u}_2)_i = 0$. Then, for any $u \in (\mathsf{u}_1, \mathsf{u}_2)$, we must have $\phi \circ \gamma(u)_i \neq 0$. However, $\gamma(\frac{\mathsf{u}_1+\mathsf{u}_2}{2})$ is the mean of $\gamma(\mathsf{u}_1)$ and $\gamma(\mathsf{u}_2)$. Hence, by our translation-compatibility definition, $\phi \circ \gamma(\frac{\mathsf{u}_1+\mathsf{u}_2}{2})_i = 0$, leading to a contradiction.

Therefore, $\mathcal{T}(n)$ is totally geodesic in $\mathcal{DS}(n)$.

A.5.2 From B to A

We only have to show that the mean is translation-compatible when all direct similarities $\boldsymbol{X}_i$ are in $\mathcal{T}(n)$. For other cases, the direct similarities must live in a coset $\mathbf{Z}\mathcal{T}(n) := \{\mathbf{ZX} : \mathbf{X} \in \mathcal{DS}(n)\}$ of $\mathcal{T}(n)$ for some $\mathbf{Z} \in \mathcal{DS}(n)$. In these cases, we left-translate the direct similarities by $\mathbf{Z}^{-1}$, compute the mean, left-translate it by $\mathbf{Z}$, and obtain a translation-compatible mean because the metric tensor g is left-invariant.

Suppose the induced metric tensor of g on $\mathcal{T}(n)$ is g'. Let d_{R} and d'_{R} denote the Riemannian distance in $\mathcal{DS}(n)$ and $\mathcal{T}(n)$ respectively. In fact, d'_{R} is just the restricted version of d_{R} on $\mathcal{T}(n)$. Let $\bar{\mathbf{X}}'$ be the mean induced by d'_{R} restricted to $\mathcal{T}(n)$, which is automatically translation-compatible. It suffices to prove that $\bar{\mathbf{X}}' = \bar{\mathbf{X}}$.

Let $f'_i(\mathbf{X}) := d'_{\mathrm{R}}(\mathbf{X}, \boldsymbol{X}_i)^2$ be the function that measures the restricted squared Riemannian distance between any translation $\mathbf{X} \in \mathcal{T}(n)$ and a given direct similarity $\boldsymbol{X}_i$. Let $f_i(\mathbf{X}) := d_{\mathrm{R}}(\mathbf{X}, \boldsymbol{X}_i)^2$ be the corresponding version in $\mathcal{DS}(n)$. By definition,

$$\bar{\mathbf{X}}' = \operatorname*{argmin}_{\mathbf{X} \in \mathcal{T}(n)} \mathsf{w}_{\underline{i}} f'_{\underline{i}}(\mathbf{X}), \tag{111}$$

$$\bar{\mathbf{X}} = \operatorname*{argmin}_{\mathbf{X} \in \mathcal{DS}(n)} \mathsf{w}_{\underline{i}} f_{\underline{i}}(\mathbf{X}). \tag{112}$$

According to (37), the gradient of $f'_i(\mathbf{X})$ is minus twice the velocity of the g'-geodesic γ that starts at $\gamma(0) = \mathbf{X}$ and ends at $\gamma(1) = \boldsymbol{X}_i$. Since this geodesic is also the g-geodesic connecting $\mathbf{X}$ with $\boldsymbol{X}_i$, we must have $\mathrm{grad} f'_i = \mathrm{grad} f_i$. Hence $\mathrm{grad}(\mathsf{w}_{\underline{i}} f'_{\underline{i}}) = \mathrm{grad}(\mathsf{w}_{\underline{i}} f_{\underline{i}})$. This implies that both gradients vanish concurrently. In other words, a g'-mean is also a g-mean.

This statement alone does not ensure that a g-mean is a g'-mean. However, because we restrict ourselves to the case that the g-mean is unique, if g'-mean exists, it must be the unique g-mean. Clearly g'-mean exists because the variance (6) is lower-bounded by 0 and $\mathcal{T}(n)$ is locally compact.

Therefore, $\bar{\mathbf{X}}' = \bar{\mathbf{X}}$.

A.6 Proof for Theorem 3

First, we determine the expression of the metric tensor at each element $\mathbf{X} \in \mathcal{DS}(n)$. Let $\mathbf{U}$ be a tangent vector at $\mathbf{X}$ with local coordinates $\mathbf{u} := \varphi_{\mathbf{X}}(\mathbf{U})$. Let $\mathbf{U}' := dL_{\mathbf{X}^{-1}}(\mathbf{U}) = \mathbf{X}^{-1}\mathbf{U}$ and let $\mathbf{u}' := \varphi_{\mathbf{0}}(\mathbf{U}')$ be its coordinates. By inspection, the linear relationship between $\mathbf{u}$ and $\mathbf{u}'$ is given by:

$$\mathbf{u}'_s = \mathbf{u}_s, \tag{113}$$

$$\mathbf{u}'_r = \tilde{\Psi}_{\mathbf{x}_r} \mathbf{u}_r, \tag{114}$$

$$\mathbf{u}'_t = \mathrm{e}^{-\mathbf{x}_s} \mathrm{e}^{-\mathbf{x}_r^\times} \mathbf{u}_t. \tag{115}$$

In other words, $\mathbf{u}' = \mathbf{D}_{\mathbf{x}}\mathbf{u}$, where $\mathbf{D}_{\mathbf{x}}$ is a block-diagonal matrix:

$$\mathbf{D}_{\mathbf{x}} := \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\Psi}_{\mathbf{x}_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathrm{e}^{-\mathbf{x}_s}\mathrm{e}^{-\mathbf{x}_r^\times} \end{bmatrix}. \tag{116}$$

With this, the metric tensor equation (39) at $\mathbf{X} = \phi(\mathbf{x})$ has a short form, $g_{\mathbf{X}}(\mathbf{U}, \mathbf{V}) = \varphi_{\mathbf{x}}(\mathbf{U})^{\mathrm{T}} \mathbf{G}(\mathbf{x}) \varphi_{\mathbf{x}}(\mathbf{V})$, where

$$\mathbf{G}(\mathbf{x}) := \mathbf{D}_{\mathbf{x}}^{\mathrm{T}} \tilde{\mathbf{G}} \mathbf{D}_{\mathbf{x}}. \tag{117}$$

Next, pick two arbitrary coordinates $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{n_{ds}}$ under the map ϕ such that $\mathbf{p} \neq \mathbf{q}$ and their first $n_r + 1$ components are zero: $\mathbf{p}_{sr} = \mathbf{q}_{sr} = \mathbf{0}$. Consider the geodesic γ going from $\gamma(0) = \phi(\mathbf{p})$ to $\gamma(1) = \phi(\mathbf{q})$. Let its coordinate functions be $\mathbf{x}(u) := \phi \circ \gamma(u)$. We will prove that $\mathcal{T}(n)$ is not totally geodesic in $\mathcal{DS}(n)$ by showing that no such γ exists that also satisfy $\mathbf{x}(u)_{sr} = \mathbf{0}$ for all $u \in (0, 1)$.

In Riemannian gemetry (e.g. see (Lee, 1997)), geodesics obey the following geodesic equations, for all $k = 1, \ldots, n_{ds}$:

$$\ddot{\mathbf{x}}_k(u) + \dot{\mathbf{x}}_{\underline{i}}(u)\dot{\mathbf{x}}_{\underline{j}}(u)\mathit{\Gamma}_{\underline{i},\underline{j};k}(\mathbf{x}(u)) = 0, \tag{118}$$

where $\ddot{\mathbf{x}}$ and $\dot{\mathbf{x}}$ are respectively first-order and second-order derivatives of $\mathbf{x}$, and $\mathit{\Gamma}_{i,\underline{j};k}$ are Christoffel symbols related to the local coordinates of the metric tensor via, for all $i, j, k = 1, \ldots, n_{ds}$:

$$\mathit{\Gamma}_{i,j;\underline{l}}(\mathbf{x})\mathbf{G}(\mathbf{x})_{\underline{l},k} = 0.5(\partial_i \mathbf{G}(\mathbf{x})_{j,k} + \partial_j \mathbf{G}(\mathbf{x})_{i,k} - \partial_k \mathbf{G}(\mathbf{x})_{i,j}), \tag{119}$$

where ∂_i for all $i = 1, \ldots, n_{ds}$ are partial derivative operators.

Under the extra condition that $\mathbf{x}(u)_{sr} = \mathbf{0}$, we get $\dot{\mathbf{x}}(u)_{sr} = \mathbf{0}$ and $\ddot{\mathbf{x}}(u)_{sr} = \mathbf{0}$, for all $u \in (0, 1)$. The first geodesic equation ($k = 1$) of (118) simplifies to:

$$\sum_{i \in \mathcal{J}_t} \sum_{j \in \mathcal{J}_t} \dot{\mathbf{x}}_i(u)\dot{\mathbf{x}}_j(u)\mathit{\Gamma}_{i,j;1}(\mathbf{x}(u)) = 0. \tag{120}$$

To find $\mathit{\Gamma}_{i,j;1}(\mathbf{x}(u))$ for all $i, j \in \mathcal{J}_t$, we substitute (117) to (119) and obtain $\mathit{\Gamma}_{i,j;1}(\mathbf{x}(u)) = \mathbf{H}(u)_{i,j}$, where:

$$\mathbf{H}(u) := 2\mathrm{e}^{-2\mathbf{x}_s}\mathrm{e}^{\mathbf{x}_r^\times} \tilde{\mathbf{G}}_{tt} \mathrm{e}^{-\mathbf{x}_r^\times}, \tag{121}$$

and $\tilde{\mathbf{G}}_{tt}$ is the n-by-n bottom-right submatrix of $\tilde{\mathbf{G}}$ (the matrix representing the metric tensor restricted to translation only). We rewrite (120):

$$\mathbf{v}_t(u)^{\mathrm{T}} \mathbf{H}(u) \mathbf{v}_t(u) = 0. \tag{122}$$

Analyzing $\mathbf{H}(u)$, we realize that this matrix is symmetric positive-definite since $\tilde{\mathbf{G}}_{tt}$ is symmetric positive-definite, $\mathrm{e}^{\mathbf{x}_r^\times}$ is a rotation matrix whose inverse is $\mathrm{e}^{-\mathbf{x}_r^\times}$, and $\mathrm{e}^{-2\mathbf{x}_s} > 0$. Thus, (122) holds if and only if $\mathbf{v}_t(u) = \mathbf{0}$ for all $u \in (0,1)$. Clearly, this is not possible because otherwise, we must have $\mathbf{x}_t(0) = \mathbf{x}_t(1)$ (since $\dot{\mathbf{x}}_t(u) = \mathbf{v}_t(u)$), leading to $\mathbf{p} = \mathbf{q}$.

Therefore, any geodesic from $\phi(\mathbf{p})$ to $\phi(\mathbf{q})$ must not lie entirely in $\mathcal{T}(n)$, proving $\mathcal{T}(n)$ is not totally geodesic in $\mathcal{DS}(n)$.

A.7 Proof for Theorem 4

To prove that d_α is left-invariant, we show that d_α is related to a pseudo-seminorm by the formula $d_\alpha(\mathbf{X},\mathbf{Y}) = h_\alpha(\mathbf{X}^{-1}\mathbf{Y})$, where

$$h_\alpha(\mathbf{Z}) := \sqrt{\frac{(\ln\mathbf{Z}_s)^2}{\sigma_s^2} + \frac{\|\mathbf{Z}_r - \boldsymbol{I}_n\|_{\mathrm{F}}^2}{\sigma_r^2} + \frac{\|\mathbf{Z}_t\|^2}{\sigma_t^2\mathbf{Z}_s^{1-\alpha}}}, \quad (123)$$

$$\mathbf{X}^{-1}\mathbf{Y} = m\left(\frac{\mathbf{Y}_s}{\mathbf{X}_s}\mathbf{X}_r^{\mathrm{T}}\mathbf{Y}_r, \frac{\mathbf{X}_r^{\mathrm{T}}(\mathbf{Y}_t - \mathbf{X}_t)}{\mathbf{X}_s}\right). \quad (124)$$

Evaluating $h_\alpha(\mathbf{X}^{-1}\mathbf{Y})^2$ yields:

$$h_\alpha(\mathbf{X}^{-1}\mathbf{Y})^2 = \frac{(\ln\mathbf{Y}_s - \ln\mathbf{X}_s)^2}{\sigma_s^2} + \frac{\left\|\mathbf{X}_r^{\mathrm{T}}\mathbf{Y}_r - \boldsymbol{I}_n\right\|_{\mathrm{F}}^2}{\sigma_r^2} + \frac{\left\|\mathbf{X}_r^{\mathrm{T}}(\mathbf{Y}_t - \mathbf{X}_t)\right\|^2}{\sigma_t^2\mathbf{X}_s^{1+\alpha}\mathbf{Y}_s^{1-\alpha}} = d_\alpha(\mathbf{X},\mathbf{Y})^2, \quad (125)$$

where the last equation holds because the Frobenius norm and the vector norm are rotation-invariant. Since $\mathbf{X}^{-1}\mathbf{Y} = (\mathbf{ZX})^{-1}(\mathbf{ZY})$, it follows that $h_\alpha(\mathbf{X}^{-1}\mathbf{Y}) = h_\alpha((\mathbf{ZX})^{-1}(\mathbf{ZY}))$, proving d_α is left-invariant.

A.8 Proof for Lemma 6

The weighted sum of squared divergences in (6) can be rewritten as:

$$\mathsf{w}_{\underline{i}}d_\alpha(\boldsymbol{X}_{\underline{i}},\mathbf{X})^2 = \frac{E_s(\mathbf{X})}{\sigma_s^2} + \frac{E_r(\mathbf{X})}{\sigma_r^2} + \frac{E_{t;\alpha}(\mathbf{X})}{\sigma_t^2}, \quad (126)$$

where $E_s(\mathbf{X}) = \mathsf{w}_{\underline{i}}d_s(\boldsymbol{X}_{\underline{i}},\mathbf{X})^2$, $E_r(\mathbf{X}) = \mathsf{w}_{\underline{i}}d_r(\boldsymbol{X}_{\underline{i}},\mathbf{X})^2$, and $E_{t;\alpha}(\mathbf{X}) = \mathsf{w}_{\underline{i}}d_{t;\alpha}(\boldsymbol{X}_{\underline{i}},\mathbf{X})^2$. Since $\mathbf{X}_r$ only appears in $E_r(\mathbf{X})$, we obtain

$$\begin{aligned}\overline{\mathbf{X}}_r &= \operatorname*{argmin}_{\mathbf{R}\in\mathcal{SO}(n)} \mathsf{w}_{\underline{i}}\left\|\mathbf{R} - \boldsymbol{X}_{\underline{i};r}\right\|_{\mathrm{F}}^2. \\ &= \mathrm{sop}(\mathsf{w}_{\underline{i}}\boldsymbol{X}_{\underline{i};r}),\end{aligned} \quad (127)$$

where the last equation follows from (79). Likewise, since $\mathbf{X}_t$ only appears in $E_{t;\alpha}(\mathbf{X})$,

$$\overline{\mathbf{X}}_t = \operatorname*{argmin}_{\mathbf{t}\in\mathbb{R}^n} \frac{\mathsf{w}_{\underline{i}}}{\boldsymbol{X}_{\underline{i};s}^{1+\alpha}}\left\|\mathbf{t} - \boldsymbol{X}_{\underline{i};t}\right\|^2, \quad (128)$$

and (60) follows.

To find $\overline{\mathbf{X}}_s$, we substitute $\overline{\mathbf{X}}_t$ back to (6) and remove the rotation term $\frac{E_r(\mathbf{X})}{\sigma_r^2}$, we obtain an optimization problem:

$$\overline{\mathbf{X}}_s = \operatorname*{argmin}_{s\in\mathbb{R}^+} \frac{\mathsf{w}_{\underline{i}}}{\sigma_s^2}\ln\left(\frac{s}{\boldsymbol{X}_{\underline{i};s}}\right)^2 + \frac{\mathsf{w}_{\underline{i}}/\sigma_t^2}{\boldsymbol{X}_{\underline{i};s}^{1+\alpha}s^{1-\alpha}}\left\|\overline{\mathbf{X}}_t - \boldsymbol{X}_{\underline{i};t}\right\|^2. \quad (129)$$

Setting $z = \ln s$, we obtain the convex objective function (58) which is a sum of a quadratic term and an exponential term, the minimizer of which satisfies a transcendental equation $Az = \mathrm{e}^{Bz}$ for some constants A, B, therefore cannot be expressed as a closed form. However, any Newton-based approach would sufficiently find the minimizer.

When $\alpha = 1$ the exponential term vanishes, and we get (61).

Acknowledgment

The authors are thankful to Laurent Ducrohet, École Polytechnique and Peter Meer, Department of Electrical and Computer Engineering, Rudgers University for their valuable feedback on the analysis of Riemannian distances of the article.

References

Agrawal M (2006) A Lie algebraic approach for consistent pose registration for general euclidean motion. In: Proc. Int. Conf. on Intelligent Robot and Systems, pp 1891–1897

Arnaudon M, Miclo L (2014) Means in complete manifolds: uniqueness and approximation. ESAIM: Probability and Statistics 18:185–206

Arnold V, Vogtmann K, Weinstein A (1989) Mathematical Methods of Classical Mechanics. Graduate Texts in Mathematics, Springer

Arsigny V, Commowick O, Pennec X, Ayache N (2006a) A Log-Euclidean polyaffine framework for locally rigid or affine registration. In: Biomedical Image Registration, vol 4057, pp 120–127

Arsigny V, Pennec X, Ni (2006b) Bi-invariant means in lie groups. applications to left-invariant polyaffine transformations. Tech. rep., INRIA Technical Report No 5885

Begelfor E, Werman M (2006) Affine invariance revisited. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, vol 2, pp 2087–2094

Beltrami E (1868) Teoria fondamentale degli spazi di curvatura constante. Annali di Mat II(2):232–255

Bhattacharya R, Patrangenaru V (2003) Large sample theory of intrinsic and extrinsic sample means on manifolds. The Annals of Statistics 31(1):1–29

Bossa MN, Olmos S (2006) Statistical model of similarity transformations: building a multi-object pose model of brain structures. In: Workshop on Math. Methods in Biomedical Image Analysis

Carreira Perpinan M (2007) Gaussian mean-shift is an EM algorithm. IEEE Trans on Pattern Analysis and Machine Intelligence 29(5):767–776

Cetingul H, Vidal R (2009) Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp 1896–1902

Cheng SH, Higham NJ, Kenney CS, Laub AJ (2000) Approximating the logarithm of a matrix to specified accuracy. SIAM J Matrix Anal Appl 22:1112–1125

Cheng Y (1995) Mean shift, mode seeking, and clustering. IEEE Trans on Pattern Analysis and Machine Intelligence 17:790–799

Coxeter HSM (1961) Introduction to Geometry. John Wiley & Sons

Downs T (1972) Orientation statistics. Biometrika 59:665–676

Drost B, Ulrich M, Navab N, Ilic S (2010) Model globally, match locally: Efficient and robust 3D object recognition. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp 998–1005

Dubbelman G, Dorst L, Pijls H (2012) Manifold statistics for essential matrices. In: Proc. European Conf. on Computer Vision, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp 531–544

Eade E (2011) Lie groups for 2d and 3d transformations, URL `http://ethaneade.com/lie.pdf`, revised Dec. 2013

Fréchet M (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. Ann Inst H Poincaré 10:215–310

Gallier J, Xu D (2002) Computing exponentials of Skew-Symmetric matrices and logarithms of orthogonal matrices. Int J Robotics and Automation 17(4):10–20

Hall BC (2003) Lie Groups, Lie Algebras, and Representations: An Elementary Introduction. Springer-Verlag

Hartley R, Trumpf J, Dai Y, Li H (2013) Rotation averaging. Int J Computer Vision 103(3):267–305

Hartley RI, Zisserman A (2004) Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press

Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Computation 14:1771–1800

Karcher H (1977) Riemannian center of mass and mollifier smoothing. Comm Pure Appl Math 30(5):509–541

Khoshelham K (2007) Extending generalized Hough transform to detect 3D objects in laser range data. Workshop on Laser Scanning XXXVI:206–210

Knopp J, Prasad M, Willems G, Timofte R, Van Gool L (2010) Hough transform and 3D SURF for robust three dimensional classification. In: Proc. European Conf. on Computer Vision, pp 589–602

Lee J (1997) Riemannian Manifolds: An Introduction to Curvature. Graduate Texts in Mathematics, Springer

Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Computer Vision 77(1-3):259–289

Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. Mathematical Programming 45:1989

Moakher M (2002) Means and averaging in the group of rotations. SIAM J Matrix Anal Appl 24:1–16

O'Neill B (1983) Semi-Riemannian Geometry: With Applications to Relativity. No. v. 103 in Pure and Applied Mathematics, Academic Press

Opelt A, Pinz A, Zisserman A (2008) Learning an alphabet of shape and appearance for multi-class object detection. Int J Computer Vision 80(1)

Park FC (1995) Distance metrics on the rigid-body motions with applications to mechanism design. J Mech Des 117(1):48–54

Park FC, Ravani B (1997) Smooth invariant interpolation of rotations. ACM Trans Graph 16(3):277–295

Parzen E (1962) 1962. The Annals of Mathematical Statistics 33(3):1065–1076

Pelletier B (2005) Kernel density estimation on Riemannian manifolds. Statistics Probability Letters 73(3):297–304

Pennec X (1998) Computing the Mean of Geometric Features Application to the Mean Rotation. Tech. Rep. RR-3371, INRIA

Pennec X (2006) Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. JMIV 25(1):127–154

Pennec X, Ayache N (1998) Uniform distribution, distance and expectation problems for geometric features processing. J Math Imaging Vis 9:49–67

Pennec X, Thirion JP (1997) A Framework for Uncertainty and Validation of 3D Registration Methods based on Points and Frames. Int J Computer Vision 25(3):203–229

Pham MT, Woodford OJ, Perbet F, Maki A, Stenger B, Cipolla R (2011) A new distance for scale-invariant 3D shape recognition and registration. In: Proc. Int. Conf. on Computer Vision

Pham MT, Woodford OJ, Perbet F, Maki A, Stenger B (2012) Toshiba CAD model point clouds dataset. URL `http://www.toshiba.eu/eu/Cambridge-Research-Laboratory/Computer-Vision-Group/Stereo-Points/`

Poincaré H (1882) Théorie des groupes fuchsiens. Almqvist & Wiksells

Ravani B, Roth B (1983) Motion synthesis using kinematic mappings. J Mech Des 105(3):460–467

Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 27(3):832–837

Schönemann P (1966) A generalized solution of the orthogonal procrustes problem. Psychometrika 31(1):1–10

Schramm E, Schreck P (2003) Solving geometric constraints invariant modulo the similarity group. In: Int. Conf. on Computational Science and Applications, pp 356–365

Shotton J, Blake A, Cipolla R (2008) Multiscale categorical object recognition using contour fragments. IEEE Trans on Pattern Analysis and Machine Intelligence 30(7):1270–1281

Sibson R (1979) Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. Journal of the Royal Statistical Society Series B 41(2):217–229

Sternberg S (1999) Lectures on Differential Geometry. AMS Chelsea Publishing Series, Chelsea Publishing Company

Strasdat H, Montiel J, Davison AJ (2010) Scale drift-aware large scale monocular slam. Robotics: Science and Systems 2(3):5

Subbarao R, Meer P (2009) Nonlinear mean shift over Riemannian manifolds. Int J Computer Vision 84(1)

Tombari F, Di Stefano L (2010) Object recognition in 3D scenes with occlusions and clutter by Hough voting. In: Proc. Pacifc-Rim Symp. on Image and Video Technology, pp 349–355

Vaccaro C (2012) Heat kernel methods in finance: the SABR model. Quantitative finance papers, arXiv.org

Woodford OJ, Pham MT, Maki A, Perbet F, Stenger B (2013) Demisting the Hough transform for 3D shape recognition and registration. In: Int. J. Computer Vision

Zefran M, Kumar V (1998) Interpolation schemes for rigid body motions. Computer-Aided Design 30(3):179–189