



**HAL**  
open science

# A Bi-Invariant Statistical Model Parametrized by Mean and Covariance on Rigid Motions

Emmanuel Chevallier, Nicolas Guigui

► **To cite this version:**

Emmanuel Chevallier, Nicolas Guigui. A Bi-Invariant Statistical Model Parametrized by Mean and Covariance on Rigid Motions. *Entropy*, 2020, 22 (4), pp.432. 10.3390/e22040432 . hal-02568245

**HAL Id: hal-02568245**

**<https://hal.science/hal-02568245v1>**

Submitted on 8 May 2020

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

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

# A Bi-Invariant Statistical Model Parametrized by Mean and Covariance on Rigid Motions

Emmanuel Chevallier<sup>1</sup> and Nicolas Guigui<sup>2</sup>

<sup>1</sup>Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille 13013, France

<sup>2</sup>Université Côte d'Azur, Inria Epione, Sophia Antipolis 06902, France

## Abstract

This paper aims to describe a statistical model of wrapped densities for bi-invariant statistics on the group of rigid motions of a Euclidean space. Probability distributions on the group are constructed from distributions on tangent spaces and pushed to the group by the exponential map. We provide an expression of the Jacobian determinant of the exponential map of  $SE(n)$  which enables the obtaining of explicit expressions of the densities on the group. Besides having explicit expressions, the strengths of this statistical model are that densities are parametrized by their moments and are easy to sample from. Unfortunately, we are not able to provide convergence rates for density estimation. We provide instead a numerical comparison between the moment-matching estimators on  $SE(2)$  and  $\mathbb{R}^3$ , which shows similar behaviors.

**keywords:** wrapped distributions; rigid motions; Euclidean groups; differential of the exponential; moment-matching estimator; density estimation; sampling

## 1 Introduction

This work is an extended version of the conference paper [1], focused on  $SE(2)$ . We provide here a formula for  $SE(n)$  with arbitrary  $n \geq 2$ , and a numerical evaluation of the convergence of the moment-matching density estimator on  $SE(2)$ .

Probability density estimation problems generally fall in one of two categories: estimating a density on a Euclidean vector space or estimating a density on a non-Euclidean manifold. In turn, estimation problems on non-Euclidean manifolds can be divided in different categories depending on the nature of the manifold. The two main classes of non-Euclidean manifold encountered in statistics are Riemannian manifolds and Lie groups. On Riemannian manifolds, the objects studied in statistics should be consistent with the Riemannian distance. For instance, means of distributions are defined as points minimizing the average square Riemannian distances. On a Lie group, the objects should be consistent with the

group law. Direct products of compact Lie groups and vector spaces for examples belong to both categories, they admit a Riemannian metric invariant by left and right multiplications. However, in full generality, Lie groups do not admit such nice metrics, hence the need for statistical tools based solely on the group law and not on the Riemannian distance.

The definition of a statistical mean on Lie groups was addressed by Pennec and Arsigny in [2] where authors define bi-invariant means on arbitrary Lie groups as exponential barycenters [3]. Once the bi-invariant mean is defined, higher order bi-invariant centered moments can be defined in the tangent space at the mean. We build on this notion of moments to address the problem of constructing statistical models on  $SE(n)$ , the group of direct isometries of  $\mathbb{R}^n$ . The wrapped distributions model we propose has several advantages. First, it is stable under left and right multiplications. Second, densities have explicit expressions and are parameterized by their mean and covariance rather than their concentration matrix as for normal distributions defined in [4]. Third, the densities are easy to sample from. To do so, we construct wrapped densities on  $SE(n)$  similar to the densities defined in [11, 5, 6, 7, 9, 10, 8] on Riemannian manifolds. Similar types of probability distributions have already been considered for robotics applications on  $SE(3)$  to model uncertainty in motion estimation, see for instance [12].

Harmonic analysis is another well-known approach to density estimation, see [13] for  $SE(2)$  and [14, 16, 15] on other manifolds. Beside the technicalities and numerical difficulties introduced by harmonic analysis on non-abelian and non-compact groups, the main motivation for using wrapped distributions over harmonic analysis techniques, is that it enables the definition of parametric models.

This work is based on two facts. First, the exponential map can be translated from the identity element to any point of the group regardless of the choice of left or right multiplication. This property was already of primary importance in the construction of the bi-invariant mean [2] and enables the definition of bi-invariant estimation procedures. The second important fact is that the Jacobian of the exponential map on  $SE(n)$  admits a closed form expression which we compute in Section 3.2. This Jacobian provides an easy way to define probability densities with explicit expressions on the group by pushing densities from tangent spaces using the exponential map.

Unfortunately, the literature on convergence of bi-invariant moments on Lie group is still very limited. Therefore, we were not able to characterize the convergence of estimators using the proposed model. Instead, we compared numerically the convergence of the moment-matching estimator on  $SE(2)$  and on  $\mathbb{R}^3$ .

The paper is organized as follows. Section 2 describes the group of direct isometries of the Euclidean space. Section 3 includes relevant properties of the exponential mapping and the computation of the Jacobian determinant. Section 4 recalls the definitions of the first and second centered moments on a Lie group. A statistical model together with a sampling and an estimation procedure is introduced in Section 5. Section 6 concludes the paper.

## 2 Euclidean Groups

For a condensed introduction to Lie group theory for robotics, see [17], and for several relevant calculations on low-dimensional rigid motions, see the series of notes [18, 19, 20].

$SE(n)$  is the set of all direct isometries of the Euclidean space  $\mathbb{R}^n$ . The composition law of maps makes  $SE(n)$  a group. For each element  $g$  of  $SE(n)$  there are a unique rotation  $R$  and a unique vector  $t$  such that

$$g(u) = Ru + t,$$

hence the isometry  $g$  can be represented by the couple  $(R, t)$ . The group structure of  $SE(n)$  is not a direct product between the special orthogonal group and the group of translations, but a semi direct product with translations as the normal subgroup:

$$\begin{aligned} SE(n) &= SO(n) \ltimes_{\phi} \mathbb{R}^n \\ (R, t)(R', t') &= (RR', \phi_R(t') + t) \end{aligned}$$

where we simply have  $\phi_R = R$ . Let  $\Psi_{(R,t)}$  denote the conjugation by  $(R, t)$ . A short calculation gives

$$\Psi_{(R,t)}(R', t') = (R, t)(R', t')(R, t)^{-1} = (RR'R^t, -RR'R^t t + Rt' + t).$$

Recall that  $Ad_{R,t} = d(\Psi_{(R,t)})_e$ . Hence, after unfolding the elements of the Lie algebra  $\mathfrak{se}(n)$  into column vectors, the matrix representation of  $Ad_{R,t}$  is given by

$$Ad_{(R,t)} : \begin{pmatrix} Ad_R & 0 \\ C & R \end{pmatrix} \tag{1}$$

where  $C$  is a  $n$  by  $\frac{n(n-1)}{2}$  matrix,  $Ad_R$  is the adjoint representation of rotations. The structure of this adjoint matrix implies first that  $SE(n)$  is unimodular, i.e., admits a bi-invariant measure and the derivative of the exponential admits an explicit expression as we will see in Section 3.2. To see that  $SE(n)$  is unimodular, consider a left-invariant volume form  $\omega$ . The volume form is bi-invariant if and only if

$$dL_g \circ dR_{g^{-1}}(\omega_e) = \omega_e,$$

or equivalently  $\det(dL_g \circ dR_{g^{-1}}) = \det(Ad_g) = 1$ . Since  $SO(n)$  is compact, it admits a bi-invariant measure. Hence  $\det(Ad_R) = 1$ , and we have

$$\det(Ad_{(R,t)}) = \det(Ad_R) \cdot \det(R) = 1.$$

We note  $\mu_G$  the bi-invariant measure associated with  $\omega$ . The fact that  $SE(n)$  is unimodular has a significant impact on the definition of statistical tools: it is possible to manipulate densities of probability distributions with respect to a canonical measure.

A convenient way to represent elements of  $SE(n)$  is to identify the isometry  $(R, t)$  with the matrix

$$\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \in GL_{n+1}(\mathbb{R}).$$

It is easy to check that the composition of isometries corresponds to the matrix multiplication.  $SE(n)$  is thus seen as a Lie subgroup of  $GL_{n+1}(\mathbb{R})$ . Our density modelling framework is intrinsic and does not depend on a specific choice of coordinates. However, it is useful for some computations to set a reference basis. The tangent space at the identity element, noted  $T_eSE(n)$ , is spanned by the matrices of the form

$$A_{i,j} = \begin{pmatrix} E_{i,j} - E_{j,i} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad T_i = \begin{pmatrix} 0 & e_i \\ 0 & 0 \end{pmatrix}$$

where  $E_{i,j}$  is the  $n \times n$  matrix with a 1 at index  $(i, j)$  and zeros elsewhere and  $e_i$  is the  $i$ -th basis vector of  $\mathbb{R}^n$ . Let  $\mathcal{B}_e = (A_{i,j}) \cup (T_i)$  be the reference basis of  $T_eSE(n)$ .  $\mathcal{B}_e$  can be translated by left multiplication to make a left-invariant field of basis  $\mathcal{B}$ . Depending on context  $A$  will denote an  $n \times n$  skew-symmetric matrix or its embedding in the Lie algebra of  $GL_{n+1}$ , and tangent vectors will be noted with the letter  $u$ :  $u = (A, T)$ .

Recall that a skew-symmetric matrix can be block-diagonalized with 2 by 2 rotations on the diagonal, followed by a 0 when the dimension is odd. For each  $n$  by  $n$  skew-symmetric matrix  $A$ , we note  $\theta_1, \dots, \theta_{\lfloor \frac{n}{2} \rfloor}$  the set of angles of the 2 by 2 rotations.

The identification of  $SE(n)$  with a Lie subgroup of  $GL_{n+1}(\mathbb{R})$  makes the computation of the exponential map easy: the group exponential is simply the matrix exponential. Let  $U$  be the subset of  $T_eSE(n)$  defined by

$$U = \{u = (A, T) \mid \forall i, \theta_i \in [-\pi, \pi]\}.$$

It can be checked that the exponential map on  $U$  is a bijection. Therefore, we can define the logarithm on  $SE(n)$  as the inverse of the exponential on  $U$ .

### 3 Bi-Invariant Local Linearizations

Moments and densities are defined using local linearizations of the group. Hence, to obtain bi-invariant statistics, the linearization must be compatible with left and right multiplications. This section describes why the exponential map provides such linearizations from arbitrary elements.

Though we do not use this formalism, the construction of the exponential at  $g$  can be viewed in the general setting of Cartan connections on Lie groups. The exponential at  $g$  is then the exponential of a bi-invariant connection, see [21, 23, 22].

### 3.1 The Exponential at Point $G$

Since the exponential maps the lines of the tangent space at  $e$  to the one parameter subgroups of  $SE(n)$ , it is a natural candidate to linearize the group near the identity. To linearize the group around an arbitrary element  $g$ , it is possible to move  $g$  to the identity by a multiplication by  $g^{-1}$ , use the linearization at identity to obtain a tangent vector in  $T_e SE(n)$ , and map the resulting tangent vector to  $T_g SE(n)$  with a multiplication by  $g$ . Fortunately, we can check that this procedure does not depend on a choice of left or right multiplication. Recall that on a Lie group,

$$g \exp(u) g^{-1} = \exp(\text{Ad}_g(u)) = \exp(dL_g(dR_{g^{-1}}(u))) = \exp(dR_{g^{-1}}(dL_g(u))),$$

where  $dL_g$  and  $dR_g$  are the differentials of the left and right multiplication. This property enables the transport of the exponential application to any element of the group without ambiguity on the choice of left or right multiplication,

$$\begin{aligned} \exp_g : T_g SE(n) &\rightarrow SE(n) \\ u &\mapsto \exp_g(u) = g \cdot \exp(dL_{g^{-1}}u) = \exp(dR_{g^{-1}}u) \cdot g, \end{aligned}$$

see Figure 1 for a visual illustration.

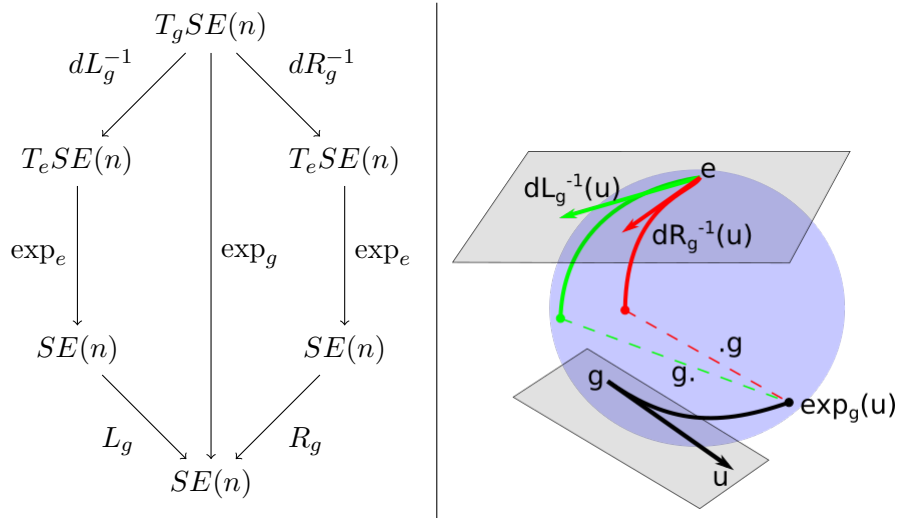


Figure 1: Commutation of the Adjoint/conjugation and the exponential.

Note  $U_g \subset T_g SE(n) = dL_g(U)$  the injectivity domain of  $\exp_g$ . The logarithm  $\log_g : SE(n) \rightarrow U_g$  becomes

$$\log_{g_0}(g) = dL_{g_0}(\log(g_0^{-1}g)) \tag{2}$$

$$= dR_{g_0}(\log(gg_0^{-1})). \tag{3}$$

We now have a linearization of the group around an arbitrary  $g \in SE(n)$ . The bi-invariant nature of the linearization is summarized in Figure 2. Independence from the choice of left or right multiplication in the definition of the exponential at an arbitrary point was the key ingredient of the definition of the bi-invariant mean in [2]. It is again a key property in our statistical model.

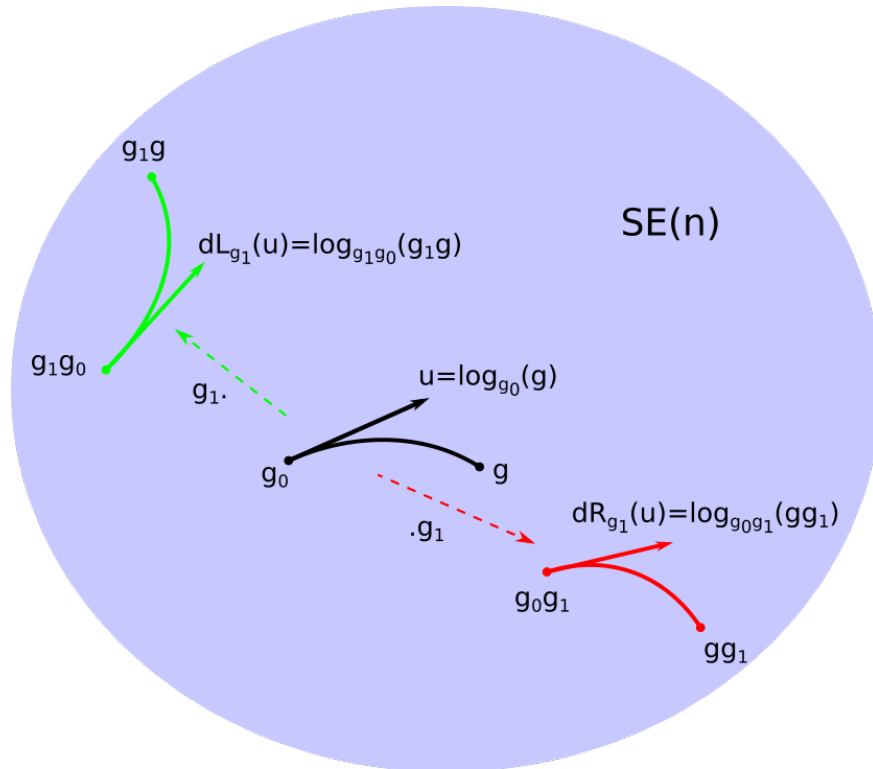


Figure 2: Bi-invariant linearization.

The strength of the exponential map is that it turns some general algebra problems into linear algebra. Once the space has been lifted to a tangent space, the problem of left and right invariances is reduced to the study of the commutation with the differentials of left and right multiplications. Since the tangent spaces do not have a canonical basis or scalar product, the manipulations we perform such as computing a mean, a covariance or estimating a density should not depend on the choice of a particular coordinate system. Hence if these manipulations commute with all the linear invertible transformations, in particular with the left and right differentials, they induce bi-invariant operations.

### 3.2 Jacobian Determinant of the Exponential

A measure  $\mu$  on  $T_g SE(n)$  can be pushed forward to the group using the exponential at  $g$ . This push-forward measure is noted  $\exp_{g*}(\mu)$ . Since  $\exp_g$  commutes with the right and left actions, so does the push-forward of measures. To obtain expressions of the densities on the group, it is necessary to compute the Jacobian determinant of the exponential, see Figure 3.

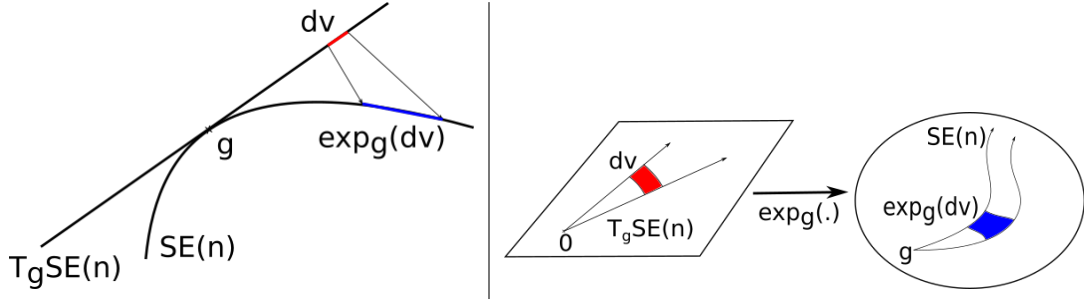


Figure 3: To push a density from a tangent space to the group, it is necessary to know the ratios between red and blue areas.

Assume  $\mu$  has a density  $f$  with respect to a Lebesgue measure of  $T_g SE(n)$  and that its support is contained in an injectivity domain  $U_g$  of  $\exp_g$ . The density  $f_{SE(n)}$  of the measure pushed on the group is given by

$$f_{SE(n)}(\exp_g u) = \frac{d \exp_{g*}(\mu)}{d\mu_G}(\exp_g(u)) = \frac{1}{|\det(d(\exp_g)_u)|} f(u)$$

where  $d(\exp_g)_u$  is the differential of  $\exp_g$  at the vector  $u$  expressed in the left-invariant reference field of basis. Since  $SE(n)$  is unimodular, i.e.,  $\mu_G$  is bi-invariant, the density of the pushed forward measure also commutes with the left and right translations of  $SE(n)$ .

We now compute this Jacobian determinant at the identity element. For the sake of notation, we drop the index  $e$  and let  $d \exp_u$  be the differential of the group exponential at the tangent vector  $u$  expressed in the bases  $\mathcal{B}_e$  and  $\mathcal{B}_{\exp(u)}$ .  $d \exp_u$  has the following expression (see [24] and [20]):

$$d \exp_u = dL_{\exp_u} \circ \left( \sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} ad_u^k \right).$$

Since  $\det(dL_{\exp_u}) = 1$ , the Jacobian determinant of the exponential is given by the determinant of the series. Fortunately, the adjoint action can be diagonalized and the determinant can be computed explicitly. Recall that  $ad_{u=(A,T)} = d(Ad_{(R,t)})_{(R,t)=e}(A,T)$ . Using Equation(1) we have that the matrix of  $ad_u$  has the following form



$$ad_u : \begin{pmatrix} ad_A & 0 \\ D & A \end{pmatrix}, \quad (4)$$

where on the left side  $A$  is an  $n \times n$  skew-symmetric matrix,  $ad_A$  is the adjoint map in the Lie algebra of skew-symmetric matrices, and  $D$  is an  $\frac{n(n-1)}{2}$  by  $n$  matrix. Since the matrix of  $ad_{(A,T)}$  is block triangular,

$$\det(d \exp_u) = \det \left( \sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} ad_A^k \right) \cdot \det \left( \sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} A^k \right).$$

Both determinants are obtained by diagonalizing  $A$  and  $ad_A$ . Take a  $d \times d$  real skew-symmetric matrix  $M$ . There is an unitary matrix  $P$  such that  $M = PD\bar{P}^t$ , where  $D$  is diagonal with eigenvalues  $\mathbf{i}\lambda_1, -\mathbf{i}\lambda_1, \dots, \mathbf{i}\lambda_{\lfloor \frac{d}{2} \rfloor}, -\mathbf{i}\lambda_{\lfloor \frac{d}{2} \rfloor}$  with  $\lambda_i \in \mathbb{R}$  followed by a 0 when  $d$  is odd. For  $\lambda \neq 0$  we have,

$$\sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} \lambda^k = \frac{1 - e^{-\lambda}}{\lambda},$$

and when  $\lambda = 0$ , the left term equals 1 and the right term can be extended by continuity. The right terms are the eigenvalues of the series in  $M$ . Hence using the fact that the determinant of a diagonalizable matrix is the product of its eigenvalues we have

$$\begin{aligned} \det \left( \sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} M^k \right) &= \prod_i \frac{1 - e^{-\mathbf{i}\lambda_i}}{\mathbf{i}\lambda_i} \frac{1 - e^{\mathbf{i}\lambda_i}}{-\mathbf{i}\lambda_i} \\ &= \prod_i 2 \frac{1 - \cos(\lambda_i)}{\lambda_i^2}. \end{aligned} \quad (5)$$

$A$  is by definition skew-symmetric. Since the adjoint representation of  $SO(n)$  is compact, there is a basis of matrices such that  $ad_A$  are skew-symmetric. Hence, Equation (5) enables the computation of  $\det(d \exp_u)$  from the eigenvalues of  $A$  and  $ad_A$ . The eigenvalues of  $ad_A$  are usually obtained by computing the roots of the complexified Lie algebra of the group  $SO(n)$ , see [25, chap. 3, sec. 8]. We provide a direct computation in appendix. If  $n$  is even, we have then

$$\det(d \exp_u) = \prod_i 2 \frac{1 - \cos(\theta_i)}{\theta_i^2} \cdot \prod_{i < j} \left( 4 \cdot \frac{1 + \cos(\theta_i + \theta_j)}{(\theta_i + \theta_j)^2} \cdot \frac{1 + \cos(\theta_i - \theta_j)}{(\theta_i - \theta_j)^2} \right) \quad (6)$$

and for  $n$  odd,

$$\det(d \exp_u) = \left( \prod_i 2 \frac{1 - \cos(\theta_i)}{\theta_i^2} \right)^2 \cdot \prod_{i < j} \left( 4 \cdot \frac{1 + \cos(\theta_i + \theta_j)}{(\theta_i + \theta_j)^2} \cdot \frac{1 + \cos(\theta_i - \theta_j)}{(\theta_i - \theta_j)^2} \right). \quad (7)$$

Let  $J_g(u) = |\det(d\exp_{g,u})|$ . Since  $\exp_g(u) = g \cdot \exp_e(dL_{g^{-1}}u)$ ,

$$d\exp_{g,u} = dL_g \circ d\exp_{e,dL_{g^{-1}}(u)} \circ dL_{g^{-1}}.$$

Furthermore,

$$dL_{g^{-1}}(\mathcal{B}_g) = \mathcal{B}_e \text{ and } \mathcal{B}_{\exp_g(u)} = dL_g \left( \mathcal{B}_{\exp_e(dL_{g^{-1}}(u))} \right).$$

Hence expressed in the basis  $\mathcal{B}_g$  and  $\mathcal{B}_{\exp_g(u)}$ , the determinant of  $d\exp_{g,u}$  is given by

$$J_g(u) = J_e(dL_{g^{-1}}(u)).$$

When all tangent vectors are expressed in the left-invariant basis, it is possible to drop the subscripts and write

$$J(A, T) = J(\theta_1, \dots, \theta_{\lfloor \frac{n}{2} \rfloor}, T) = \left| \det(d\exp_{(A, T)}) \right|. \quad (8)$$

On  $SE(2)$  we simply have

$$J(\theta, T) = \left| 2 \frac{1 - \cos(\theta)}{\theta^2} \right|. \quad (9)$$

## 4 First and Second Moments of a Distribution on a Lie Group

### 4.1 Bi-Invariant Means

Bi-invariant means on Lie groups have been introduced by Pennec and Arsigny, see [2]. An element  $\bar{g}$  in a Lie group  $G$  is said to be a bi-invariant mean of  $g_1, \dots, g_k \in G$  or of probability distribution  $\mu$  on  $G$ , if

$$\sum_i \log_{\bar{g}}(g_i) = 0 \text{ or } \int_G \log_{\bar{g}}(g) d\mu(g) = 0.$$

Observe that  $\bar{g}$  is not necessarily unique, see [2, 26, 27] for more details. Using Equation (2), it is straightforward to check that the mean is compatible with left and right multiplications:

$$dL_{g'} \left( \sum_i \log_g(g_i) \right) = \sum_i \log_{g'g}(g'g_i) \quad \text{and} \quad dR_{g'} \left( \sum_i \log_g(g_i) \right) = \sum_i \log_{gg'}(g_i g'),$$

Hence if  $\sum_i \log_{\bar{g}}(g_i) = 0$  we also have  $\sum_i \log_{g'\bar{g}}(g'g_i) = 0$  and  $\sum_i \log_{\bar{g}g'}(g_i g') = 0$ .

## 4.2 Covariance in a Vector Space

In this section, the bold letter  $\mathbf{u}$  represents a vector and the letter  $u$  its coordinate in a basis. Let us recall the definition of the covariance of a distribution on a vector space in a coordinate system. Let  $\mathbf{e}_1, \dots, \mathbf{e}_n$  be a basis of the vector space  $V$  and  $\mu$  a distribution on  $V$ . The covariance of  $\mu$  in  $V$  is defined by

$$\Sigma = E_\mu((u - \bar{\mu})(u - \bar{\mu})^t) = \int_V (u - \bar{\mu})(u - \bar{\mu})^t d\mu(u),$$

where  $u$  and  $\bar{\mu}$  are the coordinate expressions of the vector  $\mathbf{u}$  and the average of  $\mu$  and  $E_\mu(\cdot)$  is the expectation with respect to  $\mu$ .

Let  $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be such that  $K(\|x\|)$  is a probability density on  $\mathbb{R}^n$  whose covariance matrix in the canonical basis is the identity matrix, and  $\mu$  be the distribution on  $V$  whose density is

$$\frac{d\mu}{d\lambda_{\mathbf{e}}}(\mathbf{u}) = \frac{1}{\det(\sqrt{\Sigma})} K\left(\sqrt{u^t \Sigma^{-1} u}\right). \quad (10)$$

where  $\lambda_{\mathbf{e}}$  is the Lebesgue measure induced by  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . It is easy to check that the covariance of  $\mu$  is  $\Sigma$ .

Since the tangent space of a Lie group does not have a canonical basis, it is sometimes useful to define objects independently of coordinates. The coordinate free definition of the covariance becomes

$$\Sigma = \int_V (\mathbf{u} - \bar{\mu}) \otimes (\mathbf{u} - \bar{\mu}) d\mu(\mathbf{u}).$$

Recall that  $V \otimes V$  is naturally identified with the space of bilinear forms on  $V^*$ . Let  $B^*$  be the bilinear form on  $V^*$  associated with  $\Sigma$ . If  $B^*$  is positive definite, it induces an isomorphism between  $V^*$  and  $V$  and is then naturally identified with a bilinear form  $B$  on  $V$ . The definition  $\mu$  in Equation (10) becomes

$$\frac{d\mu}{d\lambda_B}(\mathbf{u}) = K\left(\sqrt{B(\mathbf{u}, \mathbf{u})}\right),$$

where  $\lambda_B$  is the Lebesgue measure on  $V$  induced by  $B$ . In this formulation it clearly appears that  $\mu$  does not depend on a basis.

## 4.3 Covariance of a Distribution on $SE(N)$

Let  $\mu$  be a distribution on  $SE(n)$  such that its bi-invariant mean  $\bar{g}$  is uniquely defined. The covariance tensor of  $\mu$  is defined as

$$\Sigma = E_\mu(\log_{\bar{g}}(\mathbf{g}) \otimes \log_{\bar{g}}(\mathbf{g})) = \int_{SE(n)} \log_{\bar{g}}(\mathbf{g}) \otimes \log_{\bar{g}}(\mathbf{g}) d\mu(g) \in T_{\bar{g}}SE(n) \otimes T_{\bar{g}}SE(n),$$

see Figure 4 for a visual illustration.

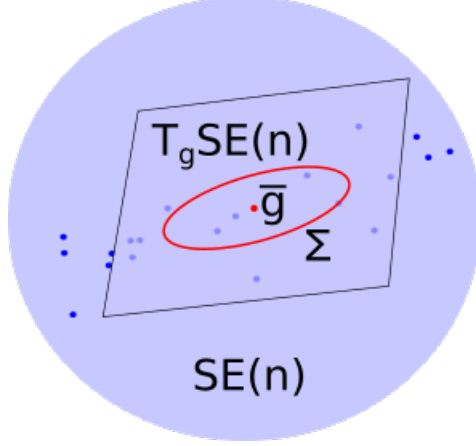


Figure 4: Covariance of an empirical measure.

Again, using Equation (2) and the bi-invariance of the mean, the compatibility of the covariance with left and right multiplication is straightforward. Note  $g \cdot \Sigma$  and  $\Sigma \cdot g$  the pushforwards by left and right multiplication by  $g$  of the tensor  $\Sigma$ . We have then

$$\begin{aligned}
 g' \cdot \Sigma &= E_{\mu} (dL_{g'}(\log_{\bar{g}}(\mathbf{g})) \otimes dL_{g'}(\log_{\bar{g}}(\mathbf{g}))) \\
 &= E_{\mu} (\log_{\mathbf{g}'\bar{g}}(\mathbf{g}'\mathbf{g}) \otimes \log_{\mathbf{g}'\bar{g}}(\mathbf{g}'\mathbf{g})) \\
 &= \Sigma'
 \end{aligned}$$

where  $\Sigma'$  is the covariance of the distribution  $g' \cdot \mu$ , the push-forward of  $\mu$  by  $L_{g'}$ . The same goes for right multiplications. However, it is important to note that for a covariance  $\Sigma$  defined on  $T_g SE(n)$ , pushing the covariance to the tangent space at identity using left and right multiplication usually gives different results:

$$g^{-1} \cdot \Sigma = Ad_{g^{-1}}(\Sigma \cdot g^{-1}) \neq \Sigma \cdot g^{-1},$$

where  $Ad_g(\cdot)$  is interpreted as the map on tensors induced by the adjoint representation.

For two distributions  $\mu_1$  and  $\mu_2$  with different means, their covariance tensors are objects defined in different tangent spaces. The collection of all these spaces form the tangent bundle  $TSE(n)$ , and covariances are identified to points in the tensor bundle  $TSE(n) \otimes TSE(n)$ .

In the reference field of basis  $\mathcal{B}$ , the covariance  $\Sigma$  has a matrix  $\Sigma$  given by

$$\Sigma = \int_{SE(n)} \log_{\bar{g}}(g) \log_{\bar{g}}(g)^t d\mu(g).$$

In principal geodesic analysis, the matrix  $\Sigma$  is sometimes referred to as a linearized quantity in contrast to the *exact* principal geodesic analysis, see [28].

## 5 Statistical Models for Bi-Invariant Statistics

### 5.1 The Model

Let  $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be such that

$$(i) \int_{\mathbb{R}^n} K(\|u\|) du = 1$$

$$(ii) \int_{\mathbb{R}^n} uu^t K(\|u\|) du = I, \quad I \text{ being the } n \times n \text{ identity matrix}$$

$$(iii) K(x > a) = 0 \text{ for some } a \in \mathbb{R}.$$

Condition (i) imposes that  $K(\|u\|)$  is a probability density on  $\mathbb{R}^n$ , condition (ii) that the covariance matrix is the identity matrix and condition (iii) that it has a bounded support.

The statistical model is defined by pushing densities of the form  $K(\|u\|)$  from tangent spaces to the group via the exponential, where the Euclidean norms on tangent spaces are parameters of the distributions. To avoid summing densities over the multiple inverse images of the exponential map, it is convenient to deal with densities  $K(\|u\|)$  whose support are included in injectivity domains, hence the (iii) requirement. Let  $C_g$  be the set of covariance matrices compatible with the injectivity domain  $U_g$ ,

$$C_g = \{ \Sigma | \forall u \notin U_g, u^t \Sigma^{-1} u > a \},$$

see Figure 5. When covariance matrices are expressed in the left-invariant reference basis, the set  $C_g$  is the same for all  $g$  and the subscript can be dropped.

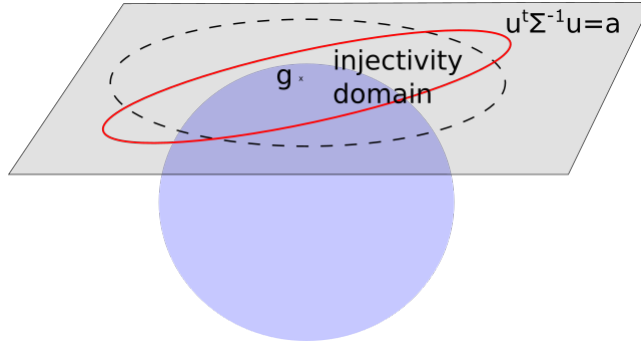


Figure 5:  $\Sigma \notin C_g$ .

When  $\Sigma \in C_g$ , the support of the probability distribution  $\mu$  on  $T_g SE(n)$  defined by

$$\frac{d\tilde{\mu}}{d\lambda_g}(u) = \frac{1}{\sqrt{\det(\Sigma)}} K\left(\sqrt{u^t \Sigma^{-1} u}\right),$$

where  $\tilde{\mu}$  is the lift of  $\mu$  by  $\log_g$ , is contained in  $U_g$ . Here  $\lambda_g$  denotes the Lebesgue measure of  $T_g SE(n)$ . The density of the push-forward of  $\mu$  is then

$$f(\exp_g(u)) = \frac{1}{J(u)\sqrt{\det(\Sigma)}} K\left(\sqrt{u^t \Sigma^{-1} u}\right), \quad (11)$$

or, expressed at  $g' \in SE(n)$ ,

$$f(g') = \frac{1}{J(\log_g(g'))\sqrt{\det(\Sigma)}} K\left(\sqrt{\log_g(g')^t \Sigma^{-1} \log_g(g')}\right), \quad (12)$$

where  $J$  is given in Equation (8). The set of such probability densities when  $g$  and  $\Sigma$  vary form a natural parametric statistical model:

$$\mathcal{M} = \{f_{g,\Sigma} : g \in SE(n) \text{ and } \Sigma \in C_g\}.$$

The commutation relations of Section 3.1 imply that  $\mathcal{M}$  is closed under left and right multiplications. The fact that  $g$  and  $\Sigma$  are the moments of  $f_{g,\Sigma}$  plays a major role in the relevance of the model  $\mathcal{M}$ . This fact holds when  $\Sigma$  is small enough, a more precise result should follow in a future work.

## 5.2 Sampling Distributions of $\mathcal{M}$

An advantage of constructing distributions from tangent spaces is that they are easy to sample: it suffices to be able to sample from the probability density  $p$  on  $\mathbb{R}$  proportional to  $K(x)$ ,  $p \propto K$ . Recall that the dimension of tangent spaces is  $d = \frac{n(n+1)}{2}$ . Let  $v_i = (x_{1,i}, \dots, x_{d,i})^t$  be random column vectors with  $x_{k,l}$  i.i.d. reals distributed according to  $p$ . Then the vectors

$$u_i = \Sigma^{\frac{1}{2}} v_i$$

are i.i.d. of density  $\frac{1}{\sqrt{\det(\Sigma)}} K\left(\sqrt{u^t \Sigma^{-1} u}\right)$  on  $\mathbb{R}^d$ , and the points

$$g_i = \exp_g(u_i)$$

are i.i.d. according to the density  $f_{g,\Sigma}$  on  $SE(n)$ .

### 5.3 Evaluation of the Convergence of the Moment-Matching Estimator

All the experiments in this section were performed using the Python package `geomstats`, see [29], available at <http://geomstats.ai>. Let  $g_1, \dots, g_k$  be points in  $SE(n)$  with a unique bi-invariant mean  $\hat{g}$  and such that the empirical covariance

$$\hat{\Sigma}(g_1, \dots, g_k) = \frac{1}{k} \sum_i \log_{\hat{g}}(g_i) \log_{\hat{g}}(g_i)^t$$

is contained in  $C_{\hat{g}}$  and that the moments of  $f_{\hat{g}, \hat{\Sigma}} \mu_G$  are  $(\hat{g}, \hat{\Sigma})$ . The compatibilities with left multiplications,

$$f_{g'g, g'\Sigma} = g' \cdot f_{g, \Sigma} \quad \text{and} \quad \hat{\Sigma}(g' \cdot g_1, \dots, g' \cdot g_k) = g' \cdot \hat{\Sigma}(g_1, \dots, g_k),$$

and right multiplications, implies that the maximum likelihood and the moment-matching estimators are bi-invariant.

On the one hand, finding the maximum likelihood estimation when  $g_1, \dots, g_k$  are i.i.d. requires an optimization procedure. On the other hand, matching moments is straightforward, provided that the moments of  $f_{\bar{g}, \Sigma}$  are  $(\bar{g}, \Sigma)$ . In most cases, this moment-matching estimator is expected to have reasonable convergence properties; however there are currently no theoretical results on the convergence of bi-invariant means and covariance on Lie groups. Hence for now it is only possible to provide empirical convergence on specific examples. Let

$$K(x) = \frac{3}{4\pi 5^{3/2}} \mathbf{1}_{[0, \sqrt{5}]}(x), \quad \Sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (13)$$

The function  $K$  verifies i),ii) and ii) of Section 5.1. Since  $\sqrt{5} < \frac{\pi}{2}$ ,  $\Sigma_1$  and  $\Sigma_2$  are admissible covariances,  $\Sigma_{1,2} \in C$ .  $\Sigma_2$  is chosen such that it correlates the rotation and translation coordinates.

Given a set of i.i.d. samples  $g_1, \dots, g_k$  of the density  $f_{e, \Sigma}$ , the estimated density of the moment-matching estimator is  $f_{\hat{g}, \hat{\Sigma}}$ . For the sake of notations, we drop subscripts and simply write  $f$  and  $\hat{f}$ . To characterize the convergence of the estimator, we compare the convergence of  $\hat{f}$  on  $SE(2)$  with the analogous moment-matching estimator on  $T_e SE(2) \sim \mathbb{R}^3$  using the samples  $\log(g_1), \dots, \log(g_k)$ .

Any  $L^p$  distance between densities provides a way to evaluate the convergence in a bi-invariant way. The  $L^1$  is particularly meaningful in the context of probabilities and presents the advantage of being independent from a reference measure. Therefore, we evaluated the expectation of the  $L^1$  distance to  $f$ :

$$e_k = E_f \left( \int_{SE(2)} |f(g) - \hat{f}(g)| d\mu_G(g) \right),$$

and the Euclidean analogous, where  $k$  is the number of samples of  $f$ . The integrals over  $SE(2)$  can be estimated using a Monte-Carlo sampling adapted to the distributions. Indeed,

$$\int |f - \hat{f}| d\mu_G = \int_S \left| 1 - \frac{\hat{f}}{f} \right| f d\mu_G + \int_{S^c} \hat{f} d\mu_G \quad (14)$$

$$= E_f \left( \left| 1 - \frac{\hat{f}}{f} \right| \right) + 1 - E_f \left( \frac{\hat{f}}{f} \right) \quad (15)$$

$$\approx 1 + \sum_i \left( \left| 1 - \frac{\hat{f}}{f}(u_i) \right| - \frac{\hat{f}}{f}(u_i) \right) \quad (16)$$

where  $S$  is the support of  $f$  and the  $(u_i)$  are i.i.d. samples of  $f$ . The  $L^1$  distances between  $f$  and  $\hat{f}$  are estimated using 5000 Monte-Carlo samples, and the expectation of the  $L^1$  distance is estimated using 200 estimates  $\hat{f}$ . Figure 6 depicts the decay of the expected  $L^1$  distance with the number of samples for the  $SE(2)$  and  $\mathbb{R}^3$  cases using the covariance  $\Sigma_1$  and Figure 7 using the covariance  $\Sigma_2$ . For a given covariance  $\Sigma$ , the error decay on  $SE(2)$  and  $\mathbb{R}^3$  seem to be asymptotically related by a multiplicative factor close to 1. Future work should focus on gaining insights into the phenomena underlying the error decay in the general case.



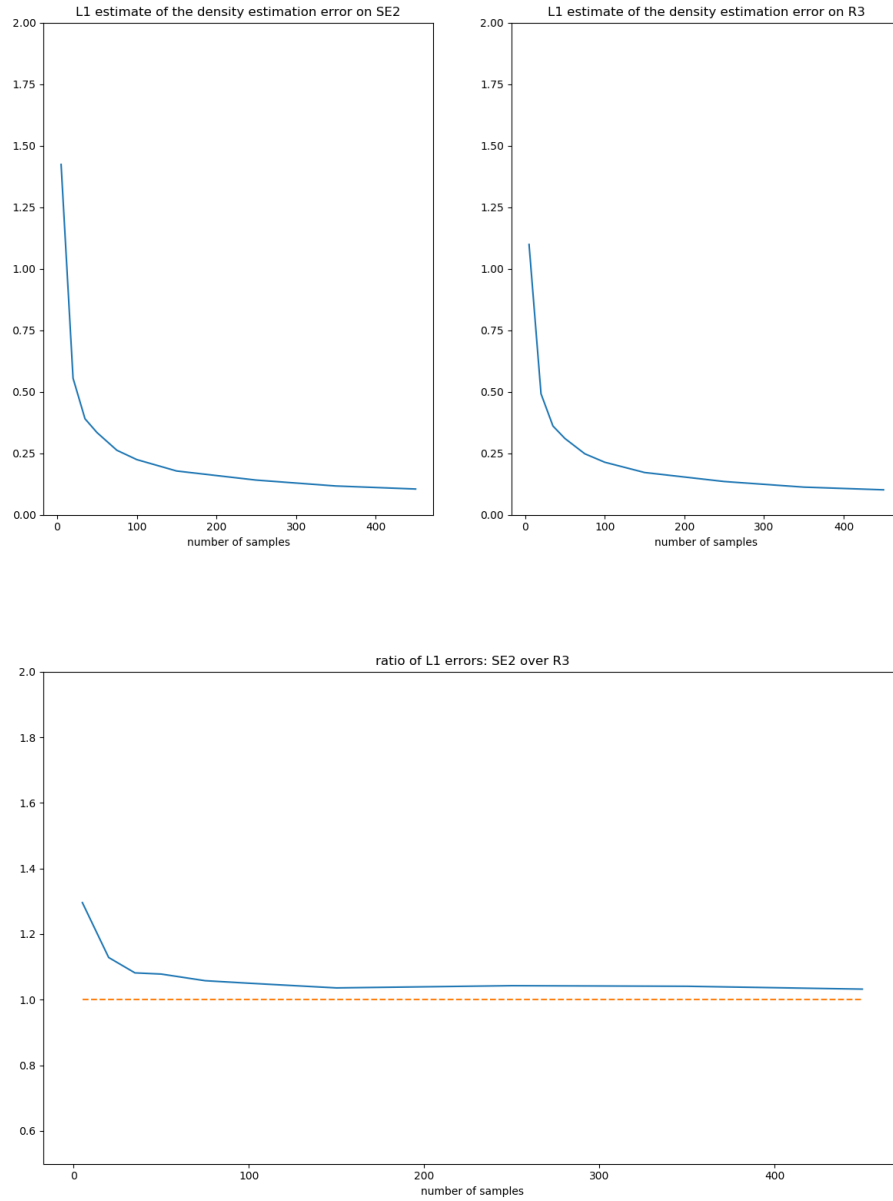


Figure 6:  $L^1$  errors and their ratios on  $SE(2)$  and  $T_e SE(2) \sim \mathbb{R}^3$  for the covariance  $\Sigma_1$ , see Equation (13).

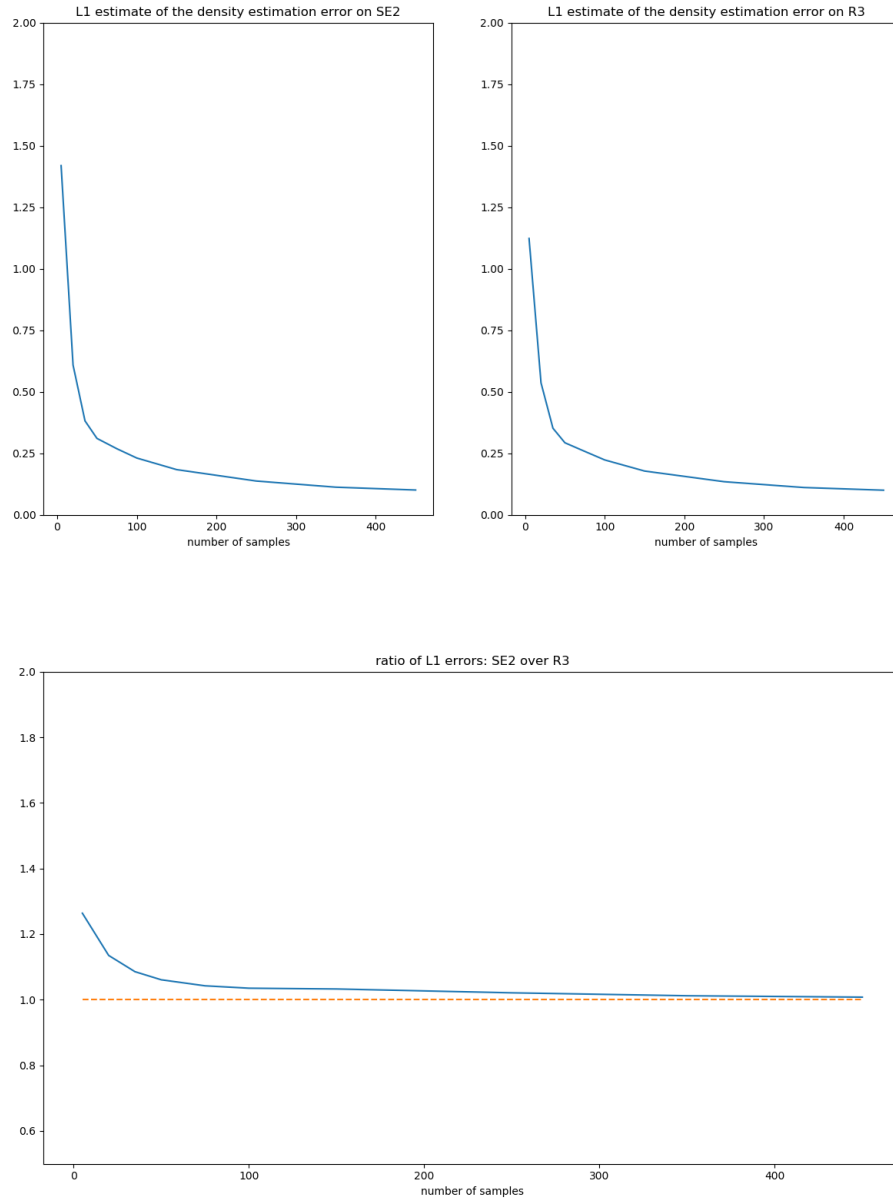


Figure 7:  $L^1$  errors and their ratios on  $SE(2)$  and  $T_e SE(2) \sim \mathbb{R}^3$  for the covariance  $\Sigma_2$ , see Equation (13).

## 6 Conclusion and Perspectives

In this paper, we have described a statistical model  $\mathcal{M}$  of densities for bi-invariant statistics on  $SE(n)$ . Even though we do not provide convergence rates, we showed experimentally on an example that the density estimation on  $SE(2)$  behaves similarly to the estimation on  $\mathbb{R}^3$ . Further works will focus on a deeper analysis of the performance of the moment-matching estimator, on proposing detailed algorithms to estimate densities in a mixture model, and on generalizing the construction to other Lie groups.

**Funding:** N. Guigui has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement G-Statistics No 786854).

**Acknowledgments:** Authors would like to thank the anonymous reviewers for their helpful contributions.

## A Eigenvalues of $ad_A$

Let  $\mathcal{A}$  be the set of skew-symmetric  $n \times n$  matrices. Let  $A \in \mathcal{A}$  and  $ad_A : X \in \mathcal{A} \mapsto AX - XA \in \mathcal{A}$ . We aim at computing the eigenvalues of  $ad_A$ . Since  $A$  is in  $\mathcal{A}$ , there exists an orthogonal matrix  $P$  such that  $D = P^{-1}AP$  is a matrix with vanishing entries outside of  $\lfloor \frac{n}{2} \rfloor$  2 by 2 matrices  $A_j = \begin{pmatrix} 0 & -a_j \\ a_j & 0 \end{pmatrix}$  along the diagonal. If  $n$  is even, the eigenvalues of  $ad_A$  are the numbers

$$\begin{aligned} & \mathbf{i}(\pm a_j \pm a_k), \quad 1 \leq j < k \leq \frac{n}{2} \\ & 0, \text{ with multiplicity } \frac{n}{2}. \end{aligned}$$

If  $n$  is odd the eigenvalues of  $ad_A$  are the numbers

$$\begin{aligned} & \mathbf{i}(\pm a_j \pm a_k), \quad 1 \leq j < k \leq \frac{n-1}{2} \\ & \pm a_j, \quad 1 \leq j \leq \frac{n-1}{2}, \\ & 0, \text{ with multiplicity } \frac{n-1}{2}. \end{aligned}$$

*Proof.* Let  $g_P(X) = PXP^{-1}$ . Since  $g_P$  is invertible and that  $ad_A = g_P \circ ad_D \circ g_P^{-1}$ ,  $ad_A$  and  $ad_D$  have the same eigenvalues. Consider first the case  $n$  odd. Let  $X \in \mathcal{A}$ ,  $X$  can

be decomposed in  $\frac{n-1}{2} \times \frac{n-1}{2}$  2 by 2 sub-matrices  $B_{i,j}$ ,  $\frac{n-1}{2}$  1 by 2 sub-matrices  $u_j$  on the last line of  $X$ , and  $\frac{n-1}{2}$  2 by 1  $u_j^t$  on the last column, and a 1 by 1 sub-matrix  $x = X_{n,n}$ .  $Y = ad_A(X)$  can be decomposed the same way in sub-matrices  $c_{i,j}, v_{i,j}, v_{i,j}^t$  and  $y = Y_{n,n}$ . With the block products we obtain,

$$C_{i,j} = A_i B_{i,j} - B_{i,j} A_i, \quad 1 \leq j \leq k \leq \frac{n-1}{2}$$

$$v_j = -u_j A_j \quad 1 \leq j \leq \frac{n-1}{2},$$

$$y = 0.$$

it follows that each subspace  $A_{i,j,i \neq j}$  with vanishing entries outside the 2 by 2 blocks  $ij$  and  $ji$ , are  $ad_D$  stable. These spaces are four-dimensional and direct calculation shows that the eigenvalues of  $ad_D$  restricted to these spaces are  $\mathbf{i}(\pm a_i \pm a_j)$ . The subspace  $\mathcal{A}_i$  defined by the blocks  $u_i$  and  $u_i^t$  are stable as well, and the computation shows that the corresponding eigenvalues are  $\mathbf{i}(\pm a_j)$ .  $ad_D$  restricted to blocks  $\mathcal{A}_{i,i}$  vanishes, thus 0 is of multiplicity  $\frac{n-1}{2}$ . In the  $n$  even case, only the eigenvalues associated with blocks  $\mathcal{A}_{i,j}$  remain.  $\square$

## References

- [1] Chevallier, E. Towards Parametric Bi-Invariant Density Estimation on SE (2). In *Proceedings of International Conference on Geometric Science of Information*; Springer: Cham, Switzerland, 2019; pp. 695-702.
- [2] Pennec, X.; Arsigny, V. Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In *Matrix Information Geometry*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 123–166.
- [3] Émery, M.; Mokobodzki, G. Sur le barycentre d'une probabilité dans une variété. In *Séminaire de probabilités XXV*; Springer: Berlin/Heidelberg, Germany, 1991; pp. 220–233.
- [4] Pennec, X. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.* **2006**, *25*, 127.
- [5] Chevallier, E.; Forget, T.; Barbaresco, F.; Angulo, J. Kernel density estimation on the siegel space with an application to radar processing. *Entropy* **2016**, *18*, 396.
- [6] Chevallier, E. A family of anisotropic distributions on the hyperbolic plane. In *Proceedings of International Conference on Geometric Science of Information*; Nielsen, F., Barbaresco, F., Eds.; Springer: Cham, Switzerland, 2017; pp. 717–724.

- [7] Chevallier, E.; Kalunga, E.; Angulo, J. Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices. *SIAM J. Imaging Sci.* **2017**, *10*, 191–215.
- [8] Falorsi, L.; de Haan, P.; Davidson, T.R.; De Cao, N.; Weiler, M.; Forré, P.; Cohen, T.S. Explorations in homeomorphic variational auto-encoding. *arXiv preprint* **2018**, arXiv:1807.04689.
- [9] Forster, C.; Carlone, L.; Dellaert, F.; Scaramuzza, D. On-Manifold Preintegration for Real-Time Visual-Inertial Odometry. *IEEE Trans. Robot.* **2016**, *33*, 1-21.
- [10] Grattarola, D.; Livi, L.; Alippi, C. Adversarial autoencoders with constant-curvature latent manifolds. *Appl. Soft Comput.* **2019**, *81*, 105511.
- [11] Pelletier, B. Kernel density estimation on Riemannian manifolds. *Stat. Probab. Lett.* **2005**, *73*, 297–304.
- [12] Barfoot, T.D.; Furgale, P.T. Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Trans. Rob.* **2014**, *30*, 679–693.
- [13] Lesosky, M.; Kim, P.T.; Kribs, D.W. Regularized deconvolution on the 2D-Euclidean motion group. *Inverse Prob.* **2008**, *24*, 055017.
- [14] Hendriks, H. Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Stat.* **1990**, *18*, 832–849.
- [15] Huckemann, S.; Kim, P.; Koo, J.; Munk, A. Mobius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Stat.* **2010**, *38*, 2465–2498.
- [16] Kim, P.; Richards, D. Deconvolution density estimation on the space of positive definite symmetric matrices. In *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger*; World Scientific Publishing: Singapore, 2008; pp. 147–168.
- [17] Sola, J.; Deray, J.; Atchuthan, D. A micro Lie theory for state estimation in robotics. *arXiv preprint* **2018**, arXiv:1812.01537.
- [18] Eade, E. Lie groups for 2d and 3d transformations. Available online: <http://ethaneade.com/lie.pdf> (accessed on 10 April 2020).
- [19] Eade, E. *Lie Groups for Computer Vision*; Cambridge Univ. Press: Cambridge, UK, 2014.
- [20] Eade, E. Derivative of the Exponential Map. Available online: <http://ethaneade.com/lie.pdf> (accessed on 10 April 2020).

- [21] Cartan, É. On the geometry of the group-manifold of simple and semi-groups. *Proc. Akad. Wetensch.* **1926**, *29*, 803–815.
- [22] Lorenzi, M.; Pennec, X. Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration. *Int. J. Comput. Vision* **2013**, *105*, 111–127.
- [23] Postnikov, M.M. *Geometry VI: Riemannian Geometry*; Encyclopedia of mathematical science, Springer: Berlin/Heidelberg, Germany, 2001.
- [24] Rossmann, W. *Lie Groups: An Introduction Through Linear Groups*; Oxford University Press on Demand: Oxford, UK, 2006; Volume 5.
- [25] Helgason, S. *Differential Geometry, Lie Groups, and Symmetric Spaces*; Academic Press: Cambridge, MA, USA, 1979.
- [26] Pennec, X. Bi-invariant means on Lie groups with Cartan-Schouten connections. In *Proceedings of International Conference on Geometric Science of Information (GSI 2013)*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 59–67.
- [27] Arsigny, V.; Pennec, X.; ; Nicholas, Ayache, N. Bi-invariant Means in Lie Groups. Application to Left-invariant Polyaffine Transformations. Available online : <https://hal.inria.fr/inria-00071383/> (accessed on 9 April 2020).
- [28] Sommer, S.; Lauze, F.; Hauberg, S.; Nielsen, M. Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 43–56.
- [29] Nina Miolane, Alice Le Brigant, Johan Mathe, Benjamin Hou, Nicolas Guigui, et al.. Geomstats: A Python Package for Riemannian Geometry in Machine Learning. 2020. hal-02536154