Reparameterizing Distributions on Lie Groups

Falorsi, L.; de Haan, P.; Davidson, T.R.; Forré, P.

Reparameterizing Distributions on Lie Groups

Luca Falorsi[1,2] **Pim de Haan**[1,3] **Tim R. Davidson**[1,2] **Patrick Forré**[1]

[1]University of Amsterdam [2]Aiconic [3]UC Berkeley

Abstract

Reparameterizable densities are an important way to learn probability distributions in a deep learning setting. For many distributions it is possible to create low-variance gradient estimators by utilizing a 'reparameterization trick'. Due to the absence of a general reparameterization trick, much research has recently been devoted to extend the number of reparameterizable distributional families. Unfortunately, this research has primarily focused on distributions defined in Euclidean space, ruling out the usage of one of the most influential class of spaces with non-trivial topologies: *Lie groups*. In this work we define a general framework to create reparameterizable densities on arbitrary Lie groups, and provide a detailed practitioners guide to further the ease of usage. We demonstrate how to create complex and multimodal distributions on the well known oriented group of 3D rotations, SO(3), using normalizing flows. Our experiments on applying such distributions in a Bayesian setting for pose estimation on objects with discrete and continuous symmetries, showcase their necessity in achieving realistic uncertainty estimates.

1 INTRODUCTION

Formulating observed data points as the outcomes of probabilistic processes, has proven to provide a useful framework to design successful machine learning models. Thus far, the research community has drawn almost exclusively from results in probability theory limited to Euclidean and discrete space. Yet, expanding the set of possible spaces under consideration to those with a

non-trivial topology has had a longstanding tradition and giant impact on such fields as physics, mathematics, and various engineering disciplines. One significant class describing numerous spaces of fundamental interest is that of *Lie* groups, which are groups of symmetry transformations that are simultaneously differentiable manifolds. Lie groups include rotations, translations, scaling, and other geometric transformations, which play an important role in several application domains. Lie group elements are for example utilized to describe the rigid body rotations and movements central in robotics, and form a key ingredient in the formulation of the Standard Model of particle physics. They also provide the building blocks underlying ideas in a plethora of mathematical branches such as *holonomy* in Riemannian geometry, *root systems* in Combinatorics, and the *Langlands Program* connecting geometry and number theory.

Many of the most notable recent results in machine learning can be attributed to researchers' ability to combine probability theoretical concepts with the power of deep learning architectures, e.g. by devising optimization strategies able to directly optimize the parameters of probability distributions from samples through backpropagation. Perhaps the most successful instantiation of this combination, has come in the framework of *Variational Inference* (VI) (Jordan et al., 1999), a Bayesian method used to approximate intractable probability densities through optimization. Crucial for VI is the ability to posit a flexible family of densities, and a way to find the member closest to the true posterior by optimizing the parameters.

These variational parameters are typically optimized using the *evidence lower bound* (ELBO), a lower bound on the data likelihood. The two main approaches to obtaining estimates of the gradients of the ELBO are the *score function* (Paisley et al., 2012; Mnih and Gregor, 2014) also known as REINFORCE (Williams, 1992), and the *reparameterization trick* (Price, 1958; Bonnet, 1964; Salimans et al., 2013; Kingma and Welling, 2013; Rezende et al., 2014). While various works have shown the latter to provide lower variance estimates, its use is limited by the absence of a general formulation for

all variational families. Although in recent years much work has been done in extending this class of reparameterizable families (Ruiz et al., 2016; Naesseth et al., 2017; Figurnov et al., 2018), none of these methods explicitly investigate the case of distributions defined on non-trivial manifolds such as Lie groups.

The principal contribution of this paper is therefore to extend the reparameterization trick to Lie groups. We achieve this by providing a general framework to define reparameterizable densities on Lie groups, under which the well-known Gaussian case of Kingma and Welling (2013) is recovered as a special instantiation. This is done by pushing samples from the Lie algebra into the Lie group using the exponential map, and by observing that the corresponding density change can be analytically computed. We formally describe our approach using results from differential geometry and measure theory.

In the remainder of this work we first cover some preliminary concepts on Lie groups and the reparameterization trick. We then proceed to present the general idea underlying our reparameterization trick for Lie groups (ReLie[1]), followed by a formal proof. Additionally, we provide an implementation section[2] where we study three important examples of Lie groups, deriving the reparameterization details for the n-Torus, $\mathbb{T}^N$, the oriented group of 3D rotations, SO(3), and the group of 3D rotations and translations, SE(3). We conclude by creating complex and multimodal reparameterizable densities on SO(3) using a novel non-invertible normalizing flow, demonstrating applications of our work in both a supervised and unsupervised setting.

2 PRELIMINARIES

In this section we first cover a number of preliminary concepts that will be used in the rest of this paper.

2.1 Lie Groups and Lie Algebras

Lie Group, G: A *Lie group*, G is a group that is also a smooth manifold. This means that we can, at least in local regions, describe group elements continuously with parameters. The number of parameters equals the dimension of the group. We can see (connected) Lie groups as continuous symmetries where we can continuously traverse between group elements[3]. Many relevant Lie groups are matrix Lie groups, which can be expressed as a subgroup of the Lie group $\mathrm{GL}(n, \mathbb{R})$ of invertible square matrices with matrix multiplication

as product.

Lie Algebra, $\mathfrak{g}$: The *Lie algebra* $\mathfrak{g}$ of a N dimensional Lie group is its tangent space at the identity, which is a vector space of N dimensions. We can see the algebra elements as infinitesimal generators, from which all other elements in the group can be created. For matrix Lie groups we can represent vectors $\boldsymbol{v}$ in the tangent space as matrices $\mathbf{v}_\times$.

Exponential Map, $\exp(\cdot)$: The structure of the algebra creates a map from an element of the algebra to a vector field on the group manifold. This gives rise to the *exponential map*, $\exp : \mathfrak{g} \to G$ which maps an algebra element to the group element at unit length from the identity along the flow of the vector field. The zero vector is thus mapped to the identity. For compact connected Lie groups, such as SO(3), the exponential map is surjective. Often, the map is not injective, so the inverse, the log map, is multi-valued. The exponential map of matrix Lie groups is the matrix exponential.

Adjoint Representation, $\mathrm{ad_x}$: The Lie algebra is equipped with with a bracket $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, which is bilinear. The bracket relates the structure of the group to structure on the algebra. For example, $\log(\exp(x)\exp(y))$ can be expressed in terms of the bracket. The bracket of matrix Lie groups is the commutator of the algebra matrices. The adjoint representation of $x \in \mathfrak{g}$ is the matrix representation of the linear map $\mathrm{ad_x} : \mathfrak{g} \to \mathfrak{g} : y \mapsto [x, y]$.

2.2 Reparameterization Trick

The reparameterization trick (Price, 1958; Bonnet, 1964; Salimans et al., 2013; Kingma and Welling, 2013; Rezende et al., 2014) is a technique to simulate samples $z \sim q(z, \theta)$ as $z = \mathcal{T}(\epsilon; \theta)$, where $\epsilon \sim s(\epsilon)$ is independent from θ [4], and the transformation $\mathcal{T}(\epsilon; \theta)$ should be differentiable w.r.t. θ. It has been shown that this generally results in lower variance estimates than score function variants, thus leading to more efficient and better convergence results (Titsias and Lázaro-Gredilla, 2014; Fan et al., 2015). This reparameterization of samples z, allows expectations w.r.t. $q(z; \theta)$ to be rewritten as $\mathbb{E}_{q(z; \theta)}[f(z)] = \mathbb{E}_{s(\epsilon)}[f(\mathcal{T}(\epsilon; \theta))]$, thus making it possible to directly optimize the parameters of a probability distribution through backpropagation.

Unfortunately, there exists no general approach to defining a reparameterization scheme for arbitrary distributions. Although there has been a significant amount of research into finding ways to extend or generalize the reparameterization trick (Ruiz et al., 2016; Naesseth et al., 2017; Figurnov et al., 2018), to the best

---

[1] Pronounced 'really'.

[2] Our implementation is available at
https://github.com/pimdh/relie

[3] We refer the interested reader to (Hall, 2003).

[4] At most weakly dependent (Ruiz et al., 2016).

Luca Falorsi[1,2], Pim de Haan[1,3], Tim R. Davidson[1,2], Patrick Forré[1]

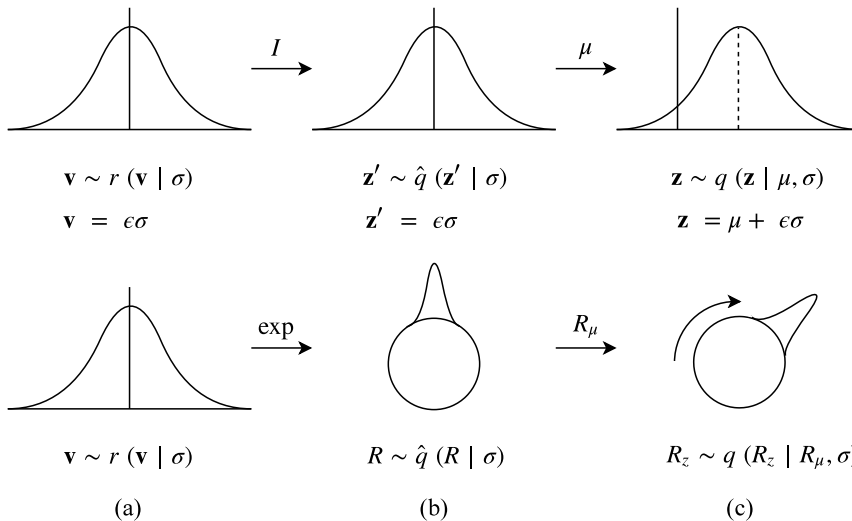Figure 2.1: Illustration of reparameterization trick of a Lie group v. the classic reparameterization trick.

of our knowledge no such trick exists for spaces with non-trivial topologies such as Lie groups.

3 REPARAMETERIZING DISTRIBUTIONS ON LIE GROUPS

In this section we will first explain our reparameterization trick for distributions on Lie groups (ReLie), by analogy to the classic Gaussian example described in (Kingma and Welling, 2013), as we can consider $\mathbb{R}^N$ under addition as a Lie group with Lie algebra $\mathbb{R}^N$ itself. In the remainder we build an intuition for our general theory drawing both from geometrical as well as measure theoretical concepts, concluded by stating our formal theorem.

3.1 Reparameterization Steps

The following reparameterization steps (a), (b), (c) are illustrated in Figure 2.1.

(a) We first sample from a reparameterizable distribution $r(\mathbf{v}|\sigma)$ on $\mathfrak{g}$. Since the Lie algebra is a real vector space, if we fix a basis this is equivalent to sampling a reparameterizable distribution from $\mathbb{R}^N$. In fact, the basis induces an isomorphism between the Lie algebra and $\mathbb{R}^N$ (see Appendix G).

(b) Next we apply the exponential map to $\mathbf{v}$, to obtain an element, $g \sim \hat{q}(g|\sigma)$ of the group. If the distribution $r(\mathbf{v}|\sigma)$ is concentrated around the origin, then the distribution of $\hat{q}(g|\sigma)$ will be concentrated around the group identity. In the Gaussian example on $\mathbb{R}^N$, this step corresponds to the identity operation, and $r = \hat{q}$. As this transformation is in general not the identity

operation, we have to account for the possible change in volume using the change of variable formula[5]. Additionally the exponential map is not necessarily injective, such that multiple points in the algebra can map to the same element in the group. We will have a more in depth discussion of both complications in the following subsection.

(c) Finally, to change the location of the distribution $\hat{q}$, we left multiply g by another group element g_μ, applying the group specific operation. In the classic case this corresponds to a translation by μ. If the exponential map is surjective (like in all compact and connected Lie groups), then g_μ can also be parameterized by the exponential map[6].

3.2 Theory

Geometrical Concepts When trying to visualize the change in volume, moving from the Lie algebra space to that of the group manifold, we quickly reach the limits of our geometrical intuition. As concepts like volume and distance are no longer intuitively defined, naturally our treatment of integrals and probability densities should be reinspected as well. In mathematics these concepts are formally treated in the fields of differential and Riemannian geometry. To gain insight into building quantitative models of the above-mentioned concepts, these fields start from the local space behavior instead. This is done through the notion of the Riemannian metric, which formally corresponds to "at-

---

[5]In a sense, this is similar to the idea underlying *normalizing flows* (Rezende and Mohamed, 2015)

[6]Care must be taken however when g_μ is predicted by a neural network to avoid homeomorphism conflicts as explored in (Falorsi et al., 2018; de Haan and Falorsi, 2018)
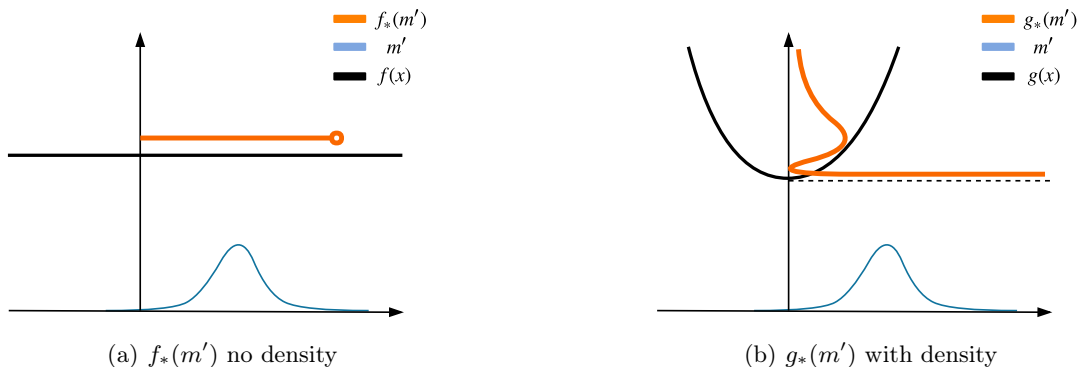
(a) $f_*(m')$ no density

(b) $g_*(m')$ with density

Figure 3.1: Example of the two non-injective mappings, $f(x) = 1$ and $g(x) = x^2 + c$, where the blue line denotes the initial Gaussian density of m', and the orange line the transformed density. In (a) the pushforward measure of m' by f collapses to δ_1, while for (b) the pushforward measure by g has a density.

taching" to the tangent space T_pG at every point p a scalar product $\langle\ ,\ \rangle_p$. This allows to measure quantities like length and angles, and to define a local volume element, in small infinitesimal scales. Extrapolating from this approach we are now equipped to measure sets and integrate functions, which corresponds to having a measure on the space[7]. Notice that this measure arose directly from the geometric properties defined by the Riemannian metric. By carefully choosing this metric, we can endow our space with some desirable properties. A standard choice for Lie groups is to use a left invariant metric, which automatically induces a left invariant measure ν, called the *Haar measure* (unique up to a constant):

$$\nu(gE) = \nu(E), \quad \forall g \in G, E \in \mathcal{B}[G],$$

where gE is the set obtained by applying the group element to each element in the set E. More intuitively, this implies that left multiplication doesn't change volume.

Measure Theoretical Concepts Perhaps a more natural way to view this problem comes from measure theory, as we're trying to push a measure on $\mathfrak{g}$, to a space G with a possibly different topology. Whenever discussing densities such as r in $\mathbb{R}^N$, it is implicitly stated that we consider a density w.r.t. the Lebesgue measure λ. What this really means is that we are considering a measure m, absolutely continuous (a.c.) w.r.t. λ, written as m $\ll \lambda$[8]. Critically, this is equivalent to stating there exists a density r, such that

$$m(E) = \int_E r \, d\lambda, \quad \forall E \in \mathcal{B}(\mathbb{R}^N),$$

where $\mathcal{B}(\mathbb{R}^N)$ is the Borel σ-algebra, i.e. the collection of all measurable sets. When applying the exponential map, we define a new measure on G[9], technically called the *pushforward* measure, $\exp_*(m)$[10]. However, G already comes equipped with another measure ν, not necessarily equal to $\exp_*(m)$. Hence, if we consider a prior distribution that has a density on ν, in order to compute quantities such as the Kullback-Leibler divergence we also need $\exp_*(m) \ll \nu$, meaning it has a density $\hat{q}$ w.r.t. ν.

In the case the exponential map is injective, it can easily be shown that the pushforward measure has a density on ν[11]. However, that these requirements are not *necessarily* fulfilled can be best explained through a simple example: Consider $f : \mathbb{R} \to \mathbb{R}$, s.t. $f(x) = 1, \forall x \in \mathbb{R}$, this function is clearly differentiable (see Fig. 1(a)). If we take a measure m', with a Gaussian density, the pushforward of m' by f is a Dirac delta, δ_1, for which it no longer holds that $f_*(m') \ll \lambda$. Intuitively, this happens because f is not injective since all points $x \in \mathbb{R}$ are mapped to 1, such that all the mass of the pushforward measure is concentrated on a single point.

Yet, this does not mean that all non-injective mappings can not be used. Instead, consider $g : \mathbb{R} \to \mathbb{R}$, s.t. $g(x) = x^2 + c, \forall x \in \mathbb{R}$ with $c \in \mathbb{R}$ a constant, and m' as before (see Fig. 1(b)). Although g is clearly not injective, for the pushforward measure by g we still have $g_*(m') \ll \lambda$. The key property here, is that it's possible to partition the domain of g into the sets $(-\infty, 0) \cup (0, \infty) \cup \{0\}$. For the first two, we can now apply the change of variable method on each, as g is injective when restricted to either. The zero set can be ignored,

---

[7]We refer the interested reader to (Lee, 2012).
[8]See definition 1, Appendix C

[9]We can do this since the exponential map is differentiable, thus continuous, thus measurable.
[10]See definition 3, Appendix C
[11]In fact, as discussed before we can always reduce to this case by defining a measure with limited support.

Luca Falorsi[1,2]**, Pim de Haan**[1,3]**, Tim R. Davidson**[1,2]**, Patrick Forré**[1]

since it has Lebesgue measure 0. This partition-idea can be generally extended for Lie groups, by proving that the Lie algebra domain can be partitioned in a set of measure zero and a countable union of open sets in which the exponential map is a diffeomorphism. This insight proven in Lemma E.5, allows us to prove the general theorem:

Theorem 3.1. *Let G, $\mathfrak{g}$, m, λ, ν be defined as above, then $\exp_*(m) \ll \nu$ with density:*

$$p(a) = \sum_{\mathbf{x} \in \mathfrak{g}: \exp(\mathbf{x}) = a} r(\mathbf{x}) |J(\mathbf{x})|^{-1}, \quad a \in G, \quad (1)$$

where $J(\mathbf{x}) := \det \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_{\mathbf{x}})^k \right)$

Proof. See Appendix E.6 □

Location Transformation Having verified the pushforward measure has a density, $\hat{q}$, the final step is to recenter the location of the resulting distribution. In practice, this is done by left multiplying the samples by another group element. Technically, this corresponds to applying the left multiplication map

$$L_{g_\mu} : G \to G, \quad g \mapsto g_\mu g$$

Since this map is a diffeomorphism, we can again apply the change of variable method. Moreover, if the measure on the group ν is chosen to be the Haar measure, as noted before applying the left multiplication map leaves the Haar measure unchanged. In this case the final sample thus has density

$$g_z \sim q(g_z | g_\mu, \sigma) = \hat{q}(g_\mu^{-1} g_z | \sigma)$$

Additionally, the entropy of the distribution is invariant w.r.t. left multiplication.

4 IMPLEMENTATION

In this section we present general implementation details, as well as worked out examples for three interesting and often used groups: the n-Torus, $\mathbb{T}^N$, the oriented group of 3D rotations, SO(3), and the 3D rotation-translation group, SE(3). The worst case reparameterization computational complexity for a matrix Lie group of dimension n can be shown[12] to be $O(n^3)$. However for many Lie groups closed form expressions can be derived, drastically reducing the complexity.

4.1 Computing $J(\mathbf{x})$

The term $J(\mathbf{x})$ as appearing in the general reparameterization theorem 3.1, is crucial to compute the change

---
[12]See Appendix D.4.

of volume when pushing a density from the algebra to the group. Here, we given an intuitive explanation for Matrix Lie groups in d dimensions with matrices of size $n \times n$. For a formal general explanation, we refer to Appendix D.

The image of the exp map is the d dimensional manifold, Lie group G, embedded in $\mathbb{R}^{n \times n}$. An infinitesimal variation in the input around point $\mathbf{x} \in \mathfrak{g}$, creates an infinitesimal variation of the output, which is restricted to the d dimensional manifold G. Infinitesimally this gives rise to a linear map between the tangent spaces at input and output. This is the Jacobian.

The change of volume is the determinant of the Jacobian. To compute it, we express the tangent space at the output in terms of the chosen basis of the Lie algebra. This is possible, since a basis for a Lie algebra provides a unique basis for the tangent space throughout G. This can be computed analytically for any $\mathbf{x}$, since the exp map of matrix Lie groups is the matrix exponential, for which derivatives are computable. Nevertheless a general expression of $J(\mathbf{x})$ exists for any Lie Group and is given in terms of the complex eigenvalue spectrum $Sp(\cdot)$ of the adjoint representation of $\mathbf{x}$, which is a linear map:

Theorem 4.1. *Let G be a Lie Group and $\mathfrak{g}$ its Lie algebra, then it can be shown that $J(\mathbf{x})$ can be computed using the following expression*

$$J(\mathbf{x}) := \prod_{\substack{\lambda \in Sp(\mathrm{ad}_{\mathbf{x}}) \\ \lambda \neq 0}} \frac{\lambda}{1 - e^{-\lambda}} \quad (2)$$

Proof. See Appendix D.4 □

4.2 Three Lie Group Examples

The n-Torus, $\mathbb{T}^N$: The n-Torus is the cross-product of n times $\mathcal{S}^1$. It is an abelian (commutative) group, which is interesting to consider as it forms an important building block in the theory of Lie groups. The n-Torus has the following matrix representation:

$$T(\boldsymbol{\alpha}) := \begin{bmatrix} B_{\alpha_1} & & \\ & \ddots & \\ & & B_{\alpha_n} \end{bmatrix}, B_\alpha := \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix},$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n) \in \mathbb{R}^n$. The basis of the Lie algebra is composed of $2n \times 2n$ block-diagonal matrices with 2×2 blocks s.t. all blocks are 0 except one that is equal to L:

$$L(\boldsymbol{\alpha}) = \begin{bmatrix} \alpha_1 L & & \\ & \ddots & \\ & & \alpha_n L \end{bmatrix}, \quad L := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

The exponential map is s.t. the pre-image can be defined from the following relationship $L(\boldsymbol{\alpha}) \mapsto T(\boldsymbol{\alpha})$:

$$\exp(L(\boldsymbol{\alpha} + 2\pi\boldsymbol{k})) = \exp(L(\boldsymbol{\alpha})), \quad \boldsymbol{k} \in \mathbb{Z}^n$$

The pushforward density is defined as

$$J(L(\boldsymbol{\alpha})) = 1$$

$$\hat{q}(T(\boldsymbol{\alpha})|\sigma) = \sum_{\boldsymbol{k} \in \mathbb{Z}^n} r\left(\boldsymbol{\alpha} + 2\boldsymbol{k}\pi|\sigma\right) \tag{3}$$

It can be observed that there is no change in volume. The resulting distribution on the circle or 1-Torus, which is also the Lie group SO(2), is illustrated in Appendix B.

The Special Orthogonal Group, SO(3)**:** The Lie group of orientation preserving three dimensional rotations has its matrix representation defined as

$$\mathrm{SO}(3) := \{R \in \mathrm{GL}(3, \mathbb{R}) : R^\top R = I \wedge \det(R) = 1\}$$

The elements of its Lie algebra $\mathfrak{so}(3)$, are represented by the 3D vector space of skew-symmetric 3×3 matrices. We choose a basis for the Lie algebra:

$$L_{1,2,3} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This provides a vector space isomorphism between $\mathbb{R}^3$ and $\mathfrak{so}(3)$, written as $[\,\cdot\,]_\times : \mathbb{R}^3 \to \mathfrak{so}(3)$. Assuming the decomposition $\mathbf{v}_\times = \theta \mathbf{u}_\times$, s.t. $\theta \in \mathbb{R}_{\geq 0}$, $\|\mathbf{u}\| = 1$, the exponential map is given by the Rodrigues rotation formula (Rodrigues, 1840)

$$\exp(\mathbf{v}_\times) = \mathbf{I} + \sin(\theta)\mathbf{u}_\times + (1 - \cos(\theta))\mathbf{u}_\times^2 \tag{4}$$

Since SO(3) is a compact and connected Lie group this map is surjective, however it is not injective. The complete preimage of an arbitrary group element can be defined by first using the principle branch $\log(\cdot)$ operator to find the unique Lie algebra element next to the origin, and then observing the following relation

$$\exp(\theta\mathbf{u}_\times) = \exp((\theta + 2k\pi)\mathbf{u}_\times) \quad k \in \mathbb{Z}$$
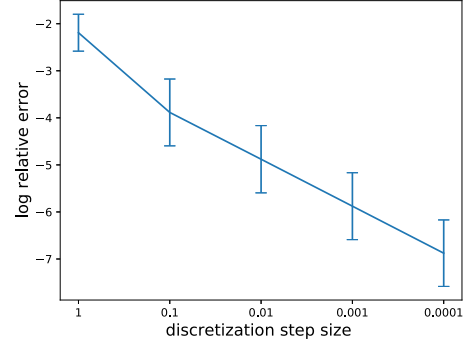
In practice, we will already have access to such an element of the Lie algebra due to the sampling approach. The pushforward density defined almost everywhere as

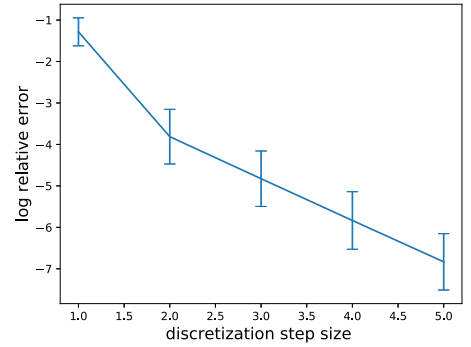$$J(\mathbf{v}) = \frac{\|\mathbf{v}\|^2}{2 - 2\cos\|\mathbf{v}\|} \tag{5}$$

$$\hat{q}(R|\sigma) = \sum_{k \in \mathbb{Z}} r\left(\left.\frac{\log(R)}{\theta(R)}(\theta(R) + 2k\pi)\right|\sigma\right) \frac{(\theta(R) + 2k\pi)^2}{3 - \mathrm{tr}(R)},$$

where $R \in \mathrm{SO}(3)$ and

$$\theta(R) = \|\log(R)\| = \cos^{-1}\left(\frac{\mathrm{tr}(R) - 1}{2}\right)$$

(a) SO(3)

(b) SE(3)

Figure 4.1: Plotted the log relative error of analytical and numerical estimations of $J(\mathbf{x})$ for the groups SO(3) and SE(3), equations (5), (6). Numerical estimation of Jacobian performed by taking small discrete steps of decreasing size (x-axis) in each Lie algebra direction. The error is evaluated at 1000 randomly sampled points.

The Special Euclidean Group, SE(3)**:** This Lie group extends SO(3) by also adding translations. Its matrix representation is given by

$$\begin{bmatrix} R & \mathbf{u} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{u} \in \mathbb{R}^3, R \in \mathrm{SO}(3)$$

The Lie algebra, $\mathfrak{se}(3)$ is similarly built concatenating a skew-symmetric matrix and a $\mathbb{R}^3$ vector

$$S(\boldsymbol{\omega}, \mathbf{u}) = \begin{bmatrix} \boldsymbol{\omega}_\times & \mathbf{u} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{u}, \boldsymbol{\omega} \in \mathbb{R}^3,$$

A basis can easily be found combining the basis elements for $\mathfrak{so}(3)$ and the canonical basis of $\mathbb{R}^3$. The exponential map from algebra to group is defined as

$$\begin{bmatrix} \boldsymbol{\omega}_\times & \mathbf{u} \\ 0 & 0 \end{bmatrix} \mapsto \begin{bmatrix} \exp(\boldsymbol{\omega}_\times) & V\mathbf{u} \\ 0 & 1 \end{bmatrix},$$

where $\exp(\cdot)$ is defined as in equation (4), and

$$V = I + \left(\frac{1 - \cos(\|\boldsymbol{\omega}\|)}{\|\boldsymbol{\omega}\|^2}\right)\boldsymbol{\omega}_\times + \left(\frac{\|\boldsymbol{\omega}\| - \sin(\|\boldsymbol{\omega}\|)}{\|\boldsymbol{\omega}\|^3}\right)\boldsymbol{\omega}_\times^2$$

Luca Falorsi[1,2], Pim de Haan[1,3], Tim R. Davidson[1,2], Patrick Forré[1]

From the expression of the exponential map it is clear that the preimage can be described similar to SO(3). Finally the pushforward density is defined almost everywhere as

$$J(S(\boldsymbol{\omega}, \mathbf{u})) = \left[\frac{\|\boldsymbol{\omega}\|^2}{2 - 2\cos\|\boldsymbol{\omega}\|}\right]^2 \qquad (6)$$

$$\hat{q}(M|\sigma) =$$

$$\sum_{k\in\mathbb{Z}} r\left(S\left(\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}(\|\boldsymbol{\omega}\| + 2k\pi), \mathbf{u}\right)\Big|\sigma\right)\left[\frac{\|\boldsymbol{\omega} + 2k\pi\|^2}{2 - 2\cos\|\boldsymbol{\omega}\|}\right]^2,$$

where ω and u such that $S(\boldsymbol{\omega}, \mathbf{u}) = \log(M)$. The log can be easily defined from the log in SO(3).

5   RELATED WORK

Various work has been done in extending the reparameterization trick to an ever growing amount of variational families. Figurnov et al. (2018) provide a detailed overview, classifying existing approaches into (1) finding *surrogate distributions*, which in the absence of a reparameterization trick for the desired distribution, attempts to use an acceptable alternative distribution that *can* be reparameterized instead (Nalisnick and Smyth, 2017). (2) *Implicit reparameterization gradients*, or *pathwise* gradients, introduced in machine learning by Salimans et al. (2013), extended by Graves (2016), and later generalized by Figurnov et al. (2018) using implicit differentiation. (3) *Generalized reparameterizations* finally try to generalize the standard approach as described in the preliminaries section. Notable are (Ruiz et al., 2016), which relies on defining a suitable invertible standardization function to allow a weak dependence between the noise distribution and the parameters, and the closely related (Naesseth et al., 2017), focusing on rejection sampling.

All of the techniques above can be used orthogonal to our approach, by defining different distributions over the Lie algebra. While some even allow for reparameterizable densities on spaces with non-trivial topologies[13], none of them provide the tools to correctly take into account the volume change resulting from pushing densities defined on $\mathbb{R}^N$ to arbitrary Lie groups. In that regard the ideas underlying *normalizing flows* (NF) (Rezende and Mohamed, 2015) are the closest to our approach, in which probability densities become increasingly complex through the use of injective maps. Two crucial differences however with our problem domain, are that the change of variable computation now needs to take into account a transformation of the underlying space, as well as the fact that the exponential

---

[13]For example Davidson et al. (2018) reparameterize the von Mises-Fisher distribution which is defined on $\mathcal{S}^M$, with $\mathcal{S}^1$ isomorphic to the Lie group SO(2).
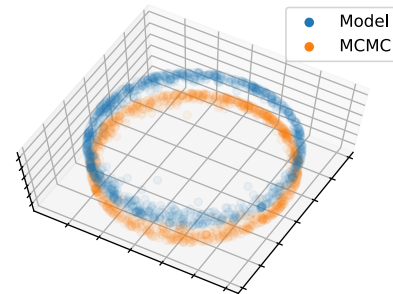
(a) Line symmetry
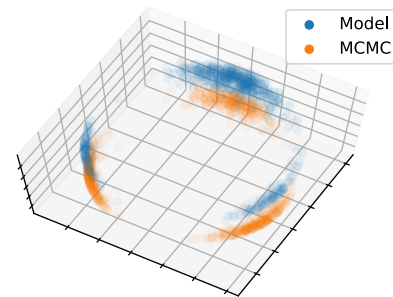
(b) Triangular symmetry

Figure 5.1: Samples of the Variational Inference model and Markov Chain Monte Carlo of Experiment 6.1. Outputs are shifted in the z-dimension for clarity.

map is generally not injective. NF can be combined with our work to create complex distributions on Lie groups, as is demonstrated in the next section.

Defining and working with distributions on homogeneous spaces, including Lie groups, was previously investigated in (Chirikjian and Kyatkin, 2000; Chirikjian, 2010; Wolfe and Mashner, 2011; Chirikjian, 2011; Chirikjian and Kyatkin, 2016; Ming, 2018). Barfoot and Furgale (2014) also discuss implicitly defining distributions on Lie groups, through a distribution on the algebra, focusing on the case of SE(3). However, these works only consider the neighbourhood of the identity, making the exponential map injective, but the distribution less expressive. In addition, generally only Gaussian distributions on the Lie Algebra are used in past work. Cohen and Welling (2015) devised harmonic exponential families which are a powerful family of distributions defined on homogeneous spaces. These works all did not concentrate on making the distributions reparameterizable. Mallasto and Feragen (2018) defined a wrapped Gaussian process on Riemannian manifolds through the pushforward of the exp map without providing an expression for the density.
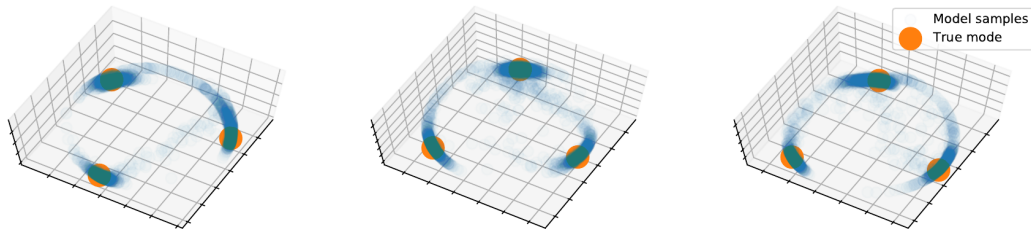
Figure 5.2: Samples from conditional SO(3) distribution $p(g|x)$ for different x, where x has the symmetry of rotations of $^{2\pi}/_3$ along one axis of Experiment 6.2. Shown are PCA embeddings of the matrices, learned using Locally Invertible Flow (LI-Flow) with Maximum Likelihood Estimation.

6   EXPERIMENTS

We conduct two experiments on SO(3) to highlight the potential of using complex and multimodal reparameterizable densities on Lie groups[14].

Normalizing Flow   To construct multimodal distributions Normalizing Flows are used (Dinh et al., 2014; Rezende and Mohamed, 2015):

$$\mathbb{R}^d \xrightarrow{f} \mathbb{R}^d \xrightarrow{r \cdot \tanh} \mathbb{R}^d \cong \mathfrak{g} \xrightarrow{\exp} G,$$

where f is an invertible Neural Network consisting of several coupling layers (Dinh et al., 2014), the $\tanh(\cdot)$ function is applied to the norm and a unit Gaussian is used as initial distribution. The hyperparameter r determines the non-injectivity of the exp map and thus of the flow. r must be chosen such that the image of $r \cdot \tanh$ is contained in the regular region of the exp map. For sufficiently small r, the entire flow is invertible, but may not be surjective, while for bigger r the flow is non-injective, with a finite inverse set at each $g \in G$, as exp is a local diffeomorphism and the image of $r \cdot \tanh$ has compact support. For details see Appendix E. For such a *Locally Invertible Flow* (LI-Flow), the likelihood evaluation requires us to branch at the non-injective function and traverse the flow backwards for each element in the preimage.

6.1   Variational Inference

In this experiment we estimate the SO(3) group actions that leave a symmetrical object invariant. This highlights how our method can be used in probabilistic generative models and unsupervised learning tasks. We have a generative model $p(x|g)$ and a uniform prior over the latent variable g. Using Variational Inference we optimize the Evidence Lower Bound to infer an approximate posterior $q(g|x)$ modeled with LI-Flow.

---
[14]See Appendix A for additional details.

Results are shown in Fig. 5.1 and compared to Markov Chain Monte Carlo samples. We observe the symmetries are correctly inferred.

6.2   Maximum Likelihood Estimation

To demonstrate the versatility of the reparameterizable Lie group distribution, we learn supervised pose estimation by learning a multimodal conditional distribution using MLE, as in (Dinh et al., 2017).

We created data set: $(x, g' = \exp(\epsilon)g)$ of objects x rotated to pose g and algebra noise samples ϵ. The object is symmetric for the subgroup corresponding to rotations of $^{2\pi}/_3$ along one axis. We train a LI-Flow model by maximizing: $\mathbb{E}_{x,g'} \log p(g'|x)$. The results in Fig. 5.2 reveal that the LI-Flow successfully learns a multimodal conditional distribution.

7   CONCLUSION

In this paper we have presented a general framework to reparameterize distributions on Lie groups (ReLie), that enables the extension of previous results in reparameterizable densities to arbitrary Lie groups. Furthermore, our method allows for the creation of complex and multimodal distributions through normalizing flows, for which we defined a novel *Locally Invertible Flow* (LI-Flow) example on the group SO(3). We empirically showed the necessity of LI-Flows in estimating uncertainty in problems containing discrete or continuous symmetries.

This work provides a bridge to leverage the advantages of using deep learning to estimate uncertainty for numerous application domains in which Lie groups play an important role. In future work we plan on further exploring the directions outlined in our experimental section to more challenging instantiations. Specifically, learning rigid body motions from raw point clouds or modeling environment dynamics for applications in optimal control present exciting possible extensions.

Luca Falorsi[1,2], Pim de Haan[1,3], Tim R. Davidson[1,2], Patrick Forré[1]

Acknowledgements

References

Alexandrino, M. M. and Bettiol, R. G. (2015). *Lie groups and geometric aspects of isometric actions.* Cham: Springer.

Barfoot, T. D. and Furgale, P. T. (2014). Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Transactions on Robotics*, 30(3):679–693.

Bonnet, G. (1964). Transformations des signaux aléatoires a travers les systemes non linéaires sans mémoire. In *Annales des Télécommunications*, volume 19, pages 203–220. Springer.

Caron, R. and Traynor, T. (2005). The zero set of polynomials. *http://www1.uwindsor.ca/math/sites/uwindsor.ca.math/files/05-03.pdf*.

Chirikjian, G. S. (2010). Information-theoretic inequalities on unimodular lie groups. *Journal of geometric mechanics*, 2(2):119.

Chirikjian, G. S. (2011). *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, volume 2. Springer Science & Business Media.

Chirikjian, G. S. and Kyatkin, A. B. (2000). *Engineering applications of noncommutative harmonic analysis: with emphasis on rotation and motion groups.* CRC press.

Chirikjian, G. S. and Kyatkin, A. B. (2016). *Harmonic Analysis for Engineers and Applied Scientists: Updated and Expanded Edition.* Courier Dover Publications.

Cohen, T. S. and Welling, M. (2015). Harmonic exponential families on manifolds. *ICML*.

Davidson, T. R., Falorsi, L., Cao, N. D., Kipf, T., and Tomczak, J. M. (2018). Hyperspherical Variational Auto-Encoders. *UAI*.

de Haan, P. and Falorsi, L. (2018). Topological constraints on homeomorphic auto-encoding. *NIPS Workshop*.

Dinh, L., Krueger, D., and Bengio, Y. (2014). Nice: Non-linear independent components estimation. *ICLR*.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using real nvp. *ICLR*.

Duistermaat, J. J. and Kolk, J. A. C. (2000). *Lie groups.* Berlin: Springer.

Falorsi, L., de Haan, P., Davidson, T. R., De Cao, N., Weiler, M., Forré, P., and Cohen, T. S. (2018). Explorations in homeomorphic variational auto-encoding. *ICML Workshop*.

Fan, K., Wang, Z., Beck, J., Kwok, J., and Heller, K. A. (2015). Fast second order stochastic backpropagation for variational inference. In *NIPS*, pages 1387–1395.

Figurnov, M., Mohamed, S., and Mnih, A. (2018). Implicit reparameterization gradients. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 439–450. Curran Associates, Inc.

Graves, A. (2016). Stochastic backpropagation through mixture density distributions. *arXiv preprint arXiv:1607.05690*.

Hall, B. (2003). *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction.* Graduate Texts in Mathematics. Springer.

Hermann, R. (1980). Differential geometry, lie groups, and symmetric spaces (sigurdur helgason). *SIAM Review*, 22(4):524–526.

Howe, R. (1989). Review: Sigurdur helgason, groups and geometric analysis. integral geometry, invariant differential operators and spherical functions. *Bull. Amer. Math. Soc. (N.S.)*, 20(2):252–256.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *CoRR*.

Klenke, A. (2014). *Probability Theory - A Comprehensive Course.* Universitext. Springer, London, second edition.

Lee, J. (2010). *Introduction to topological manifolds*, volume 202. Springer Science & Business Media.

Lee, J. M. (2012). *Introduction to smooth manifolds*. Graduate texts in mathematics. Springer, New York, NY [u.a.], 2. ed. edition.

Mallasto, A. and Feragen, A. (2018). Wrapped gaussian process regression on riemannian manifolds. *CVPR*.

Milnor, J. (1976). Curvatures of left invariant metrics on lie groups. *Advances in Mathematics*, 21(3):293 – 329.

Ming, Y. (2018). Variational bayesian data analysis on manifold. *Control Theory and Technology*, 16(3):212–220.

Mnih, A. and Gregor, K. (2014). Neural variational inference and learning in belief networks. In *ICML*.

Moler, C. and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49.

Naesseth, C., Ruiz, F., Linderman, S., and Blei, D. (2017). Reparameterization Gradients through Acceptance-Rejection Sampling Algorithms. *AISTATS*, pages 489–498.

Nalisnick, E. and Smyth, P. (2017). Stick-breaking variational autoencoders. *ICLR*.

Paisley, J., Blei, D., and Jordan, M. (2012). Variational bayesian inference with stochastic search. *ICML*.

Price, R. (1958). A useful theorem for nonlinear devices having gaussian inputs. *IRE Transactions on Information Theory*, 4(2):69–72.

Rezende, D. and Mohamed, S. (2015). Variational inference with normalizing flows. *ICML*, 37:1530–1538.

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. *ICML*, pages 1278–1286.

Rodrigues, O. (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace: et de la variation des cordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J Mathematiques Pures Appliquees*, 5:380–440.

Ruiz, F. R., Titsias RC AUEB, M., and Blei, D. (2016). The generalized reparameterization gradient. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *NIPS*, pages 460–468. Curran Associates, Inc.

Salimans, T., Knowles, D. A., et al. (2013). Fixed-form variational posterior approximation through stochastic linear regression. *Bayesian Analysis*, 8(4):837–882.

Schreiber, U. and Bartels, T. (2018). volume form. `http://ncatlab.org/nlab/show/volume%20form`. Revision 10.

Titsias, M. and Lázaro-Gredilla, M. (2014). Doubly stochastic variational bayes for non-conjugate inference. In *ICML*, pages 1971–1979.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Wolfe, K. C. and Mashner, M. (2011). Bayesian fusion on lie groups. *Journal of Algebraic Statistics*, 2(1).