A EXPERIMENTS: ADDITIONAL DETAILS

A.1 A Small Note on Alternative Proxies

The main focus of the experiments in this work is to show how our framework enables the usage of reparameterizable distributions on arbitrary Lie groups in a probabilistic deep learning setting, which to the best of our knowledge is not possible with current alternatives. The experiments therefore represent typical prototypes of applications, which can now be tackled using a general approach. To avoid confusion, it might very well be possible to design specialized one-off solutions for learning distributions on specific Lie groups, however, in this paper we aim at providing a *general* framework for doing this task.

A.2 Supplementary Details on VI Experiment

Setup In this proto-typical Variational Inference experiment we provide an intuitive example of the need for complex distributions in the difficult task of estimating which group actions of SO(3) leave a symmetrical object invariant. For didactic purposes we take two ordered points, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}3$, and perform LI-Flow VI to learn the approximate posterior over rotations. We evaluate the learned distribution by comparing its samples to those of the true posterior obtained using the Metropolis-Hastings algorithm.

Results Results are shown in Fig. 5.1. As expected, the discovered distribution over SO(3) group actions is a rotational subgroup, $\mathcal{S}^1$. Clearly, the learned approximate posterior almost perfectly matches the true posterior. Instead, using a simple centered distribution such as the pushforward of a Gaussian as the variational family, would make learning the observed topology problematic, as all probability mass would focus around a single rotation.

A.3 Supplementary Details on MLE Experiment

Setup We generate a random vector x_0 that has a linear Lie group action. Then we create a random variable $g \in G$ uniformly distributed representing the pose and a noisy version $g' = \exp(\epsilon)g$ with $\epsilon \sim (0, 0.01)$. We observe $x' = g'(x_0)$ and need to predict g. This corresponds to having noisy observations of an object x from different poses and needing to estimate the pose $p(g|x')$. When the object is symmetrical, that is a subgroup $D \subset G$ exists such that $d(x_0) = x_0$ for all $d \in D$, $p(g|x')$ should have modes corresponding to the values in D.

Results This is evaluated on SO(3). The object x_0 is taken to an element of the representation space of SO(3), as in (Falorsi et al., 2018). It is made symmetric by taking the average of $\{d(x_0)|d \in D\}$. D is taken to be the cyclic group of order 3 corresponding to rotations of $^{2\pi}/_{3}$ along one axis. The results show in Figure 5.2 reveal that the LI Flow successfully learns complicated conditional distributions.

B DISTRIBUTIONS ON THE CIRCLE

As an example of how the reparametrizable distribution on Lie groups behaves in practice, we illustrate in Figure B.1 the distribution that arises when a univariate Normal distribution is pushed forward to the Lie group $SO(2)$, homeomorphic to the circle, with the exponential map.

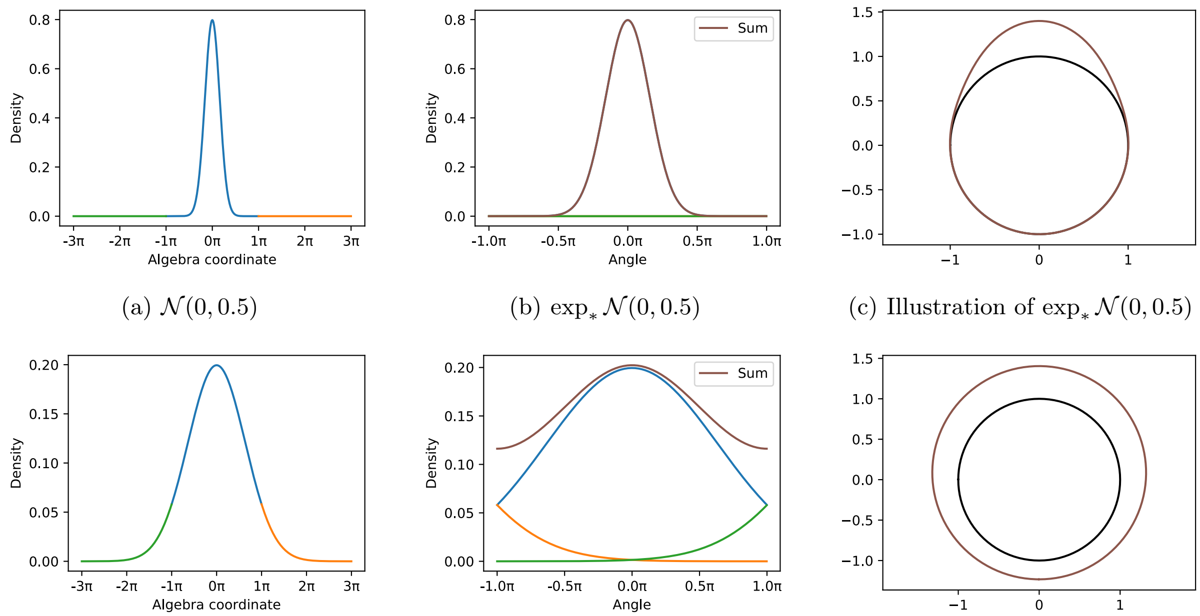
C PREREQUISITES

Definition 1 (Absolutely continuous measures, see Klenke (2014))**.** *Let* $(X, \mathcal{A})$ *be a measurable space, and* $\nu, m : \mathcal{A} \to [0, \infty]$ *two measures on* $(X, \mathcal{A})$*. Then* ν *is said to be* absolutely continuous *with respect to* m*, written as* $\nu \ll m$*, iff for all* $A \in \mathcal{A}$ *we have that*

$$m(A) = 0 \implies \nu(A) = 0. \tag{7}$$

Definition 2 (Density between two measures)**.** *Let* $(X, \mathcal{A})$ *be a measurable space, and* $\nu, m : \mathcal{A} \to [0, \infty]$ *two measures on* $(X, \mathcal{A})$*. One says that* ν *has a* density *w.r.t.* m *iff there is a measurable function* $f : X \to \mathbb{R}_{\geq 0}$ *such that for all* $A \in \mathcal{A}$ *we have:*

$$\nu(A) = \int_A f \, dm.$$



(a) $\mathcal{N}(0, 0.5)$ (b) $\exp_* \mathcal{N}(0, 0.5)$ (c) Illustration of $\exp_* \mathcal{N}(0, 0.5)$

(d) $\mathcal{N}(0, 2)$ (e) $\exp_* \mathcal{N}(0, 2)$ (f) Illustration of $\exp_* \mathcal{N}(0, 2)$

Figure B.1: Density of pushforward of Normal distributions with zero mean and scale 0.5 and 2 to the Lie group $SO(2)$. Following Equation 3, the density on the group in (b) and (e) at angle θ is simply the sum of the algebra density of the pre-images of θ. The circular representation in (c) and (f) illustrate the density q on the group by drawing a loop with radius $1 + q(\theta)$, for angle θ.

It is knows (see Klenke (2014)) that a density (if existent) is unique up to a m-zero measure and it is often denoted as: $f(x) = \frac{d\nu}{dm}(x)$.

Theorem C.1 (Radon-Nikodým, see Klenke (2014) Cor. 7.34)**.** *Let* $(X, \mathcal{A})$ *be a measurable space, and* $\nu, m : \mathcal{A} \to [0, \infty]$ *two* σ*-finite measures on* $(X, \mathcal{A})$*. Then one has the equivalence:*

$$\nu \text{ has a density w.r.t. } m \iff \nu \ll m.$$

Definition 3 (Pushforward measure)**.** *Let* $(X, \mathcal{A}, m)$ *be a measure space,* $(Y, \mathcal{B})$ *a measurable space, and let* $f : X \to Y$ *be a measurable map. Then the* pushforward measure *of* m *along* f*, in symbols* $f_* m$*, is defined as follows*

$$(f_* m)(B) := m(f^{-1}(B)), \quad \text{for } B \in \mathcal{B}. \tag{8}$$

Definition 4 (The standard measure on (pseudo-)Riemannian manifolds, see (Schreiber and Bartels, 2018))**.** *Let* (M, g) *be a (pseudo-)Riemannian manifold with metric tensor* g*. The standard measure* m_g *on* M *w.r.t.* g *is in local (oriented) coordinates per definition given by the density* $\sqrt{|\det(g)|}$ *w.r.t. the Lebesgue measure, where* $|\det(g)|(x)$ *is the absolute value of the determinant of the matrix of* g *in the local coordinates at point* x*. Note that the standard measure w.r.t.* g *always exists.*

We are mainly interested in probability distributions on (pseudo-)Riemannian manifolds (M, g) that have a density w.r.t. the standard measure m_g (i.e. that are absolute continuous w.r.t. m_g).

D CHANGE OF VARIABLES

Consider a n dimensional Lie group G and its Lie algebra $\mathfrak{g}$. Then a scalar product $\langle\ ,\ \rangle$ on $\mathfrak{g}$ induces left invariant Riemannian metric on G in the following way:

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \langle \mathrm{d}(L_{a^{-1}})_a \mathbf{x}, \mathrm{d}(L_{a^{-1}})_a \mathbf{y} \rangle \quad \forall a \in G, \quad \mathbf{x}, \mathbf{y} \in T_a G \tag{9}$$

Where $\mathrm{d}(L_{a^{-1}})_a : T_aG \to T_eG \cong \mathfrak{g}$ is the differential of the Left action by a^{-1}. Since we have now given to the Lie group a Riemannian manifold structure, we can endow G with a regular Borel measure ν. Notice that from the construction of the metric ν is a left-invariant measure, this also called left Haar measure. The left Haar measure is unique up to a scaling constant, determined by the choice of scalar product. Also the scalar product in the Lie algebra induces a measure λ in $\mathfrak{g}$[15] that is invariant with respect to vector addition and unique up to a constant. The following Proposition gives a general formula for the change of variables in Riemannian manifolds:

Lemma D.1. *(Proposition 1.3 Howe (1989)) Let M and N be Riemannian manifolds and Φ a diffeomorphism of M onto N. For $p \in M$ let $|\det(d\Phi_p)|$ denote the absolute value of the determinant of the linear isomorphism $d\Phi_p : T_pM \to T_{\Phi(p)}N$ when expressed in terms of any orthonormal bases. Then given a function F:*

$$\int_N F(q)\ dq = \int_M F(\Phi(p))|\det(d\Phi_p)|dp \tag{10}$$

if dp and dq denote the Riemannian measures on M and N, respectively

In order to change variables we therefore need an orthonormal basis for the tangent space T_aG at each one of the group elements $a \in G$.

Similarly as we built the Riemannian metric, this is given by the differential of the Left group action.

In fact given $\mathbb{B} = (\mathbf{e}_i)_{i\in[n]}$ a basis of the Lie algebra, then a basis $\mathbb{B}_a$ for T_aG is given by $(\mathrm{d}(L_a)_e(e_i))_{i\in[n]}$. If $(\mathbf{e}_i)_{i\in[n]}$ is orthonormal then $((dL_a)_e(e_i))_{i\in[n]}$ is an orthonormal basis for T_aG considering G endowed with the Riemannian metric defined in Equation 9:

$$\langle \mathrm{d}(L_a)_e\mathbf{e}_i, \mathrm{d}(L_a)_e\mathbf{e}_j\rangle_a = \langle \mathbf{e}_i, \mathbf{e}_j\rangle = \delta_{ij} \quad \forall a \in G, \quad i,j \in [n] \tag{11}$$

Then with respect of this basis the matrix representation U of the differential of the exponential $\mathrm{d}\exp_{\mathbf{x}}$ has entries:

$$U_{ij} = \langle (\mathrm{d}(L_{\exp(\mathbf{x})})_e)(\mathbf{e}_i), \mathrm{d}\exp_{\mathbf{x}}(\mathbf{e}_j)\rangle_{\exp(\mathbf{x})} = \langle \mathbf{e}_i, d(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})} \circ d\exp_{\mathbf{x}}(\mathbf{e}_j)\rangle$$

Where the equality follows from (9) [16] . From this equality it is clear that U is equal to the matrix representation of the endomorphism $\mathrm{d}(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})} \circ \mathrm{d}\exp_{\mathbf{x}} : \mathfrak{g} \to \mathfrak{g}$ with repect to the basis $\mathbb{B}$. Since the determinant an endomorphism is a quantity defined independently of the choice of the basis. The volume change term is independent on the choice of scalar product and metric and it is given by the determinant of the endomorphism $\mathrm{d}(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})} \circ \mathrm{d}\exp_{\mathbf{x}}$ that can be computed with respect of any basis of $\mathfrak{g}$. [17]

Then Theorem 1.7 of Hermann (1980) gives a general expression of this endomorphism for every Lie group:

Theorem D.2. *(Theorem 1.7 of Hermann (1980)) Let G be a Lie group with Lie algebra $\mathfrak{g}$. The exponential mapping of the manifold $\mathfrak{g}$ into G has the differential:*

$$d\exp_{\mathbf{x}} = d(L_{\exp(\mathbf{x})})_e \circ \frac{1-\exp(-\operatorname{ad}_{\mathbf{x}})}{\operatorname{ad}_{\mathbf{x}}}, \tag{12}$$

where $\dfrac{1-\exp(-\operatorname{ad}_{\mathbf{x}})}{\operatorname{ad}_{\mathbf{x}}}$ is a formal expression to indicate the infinite power series $\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!}(\operatorname{ad}_{\mathbf{x}})^k$.

Now simply by composing on the left each side of (12) with $d(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})}$ we have that:

$$\mathrm{d}(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})} \circ \mathrm{d}\exp_{\mathbf{x}} = \frac{1-\exp(-\operatorname{ad}_{\mathbf{x}})}{\operatorname{ad}_{\mathbf{x}}} := \sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!}(\operatorname{ad}_{\mathbf{x}})^k \tag{13}$$

[15]It is sufficient to consider $\mathfrak{g}$ with the Riemannian metric given by "copying" the scalar product at each point. This could be formalized considering $\mathfrak{g}$ itself a Lie group with respect to vector addition and repeating the same argument used for G

[16]Notice that here in the following derivations we identify the tangent space at a point $\mathbf{x}$ of the Lie algebra with the Lie algebra itself.

[17]Notice that even if the formal construction uses an explicit choice of scalar product and basis the induced measures ν and λ are independent of this choice up to a scalar multiplicative constant. Moreover since the choice of the constant for λ automatically the constant for ν the change of volume term is completely independent from the choice of scalar product and basis, as showed above. Regardless of these considerations the density of the pushforward measure will in general dependent of the choice of basis and scalar product, an in depth discussion of this behaviour is given in Appendix G

Combining this expression with Proposition D.1 we have the general expression for the change of variables in Lie groups:

Lemma D.3. *Let λ and ν defined as above. Let $U \subseteq \mathfrak{g}$ an open set in which $\exp_{|U} : U \to \exp(U) \subseteq G$ is a diffeomorphism. Let f measurable function in $\mathfrak{g}$ and h a measurable function in G. Then we have:*

$$\int_U f \; d\lambda = \int_{\exp(U)} f(\exp^{-1}(a))|J(\exp^{-1}(a))|^{-1} d\nu \tag{14}$$

$$\int_U h(\exp(x))|J(\mathbf{x})| d\lambda = \int_{\exp(U)} h \; d\nu, \tag{15}$$

where:

$$J(\mathbf{x}) = \det\left(\frac{1 - \exp(-\operatorname{ad}_{\mathbf{x}})}{\operatorname{ad}_{\mathbf{x}}}\right) \tag{16}$$

When we can find all eigenvalues of $\operatorname{ad}_{\mathbf{x}}$ the following theorem gives a closed form for $J(\mathbf{x})$.

Theorem D.4. *Let G be a Lie Group and $\mathfrak{g}$ its Lie algebra, then the expression*

$$J^{-1}(\mathbf{x}) := \det\left(\frac{1 - \exp(-\operatorname{ad}_x)}{\operatorname{ad}_x}\right) = \prod_{\substack{\lambda \in Sp(ad_x) \\ \lambda \neq 0}} \frac{1 - e^{-\lambda}}{\lambda}, \tag{17}$$

where $Sp(\cdot)$ is the spectrum of the operator, i.e. the set of its (complex) eigenvalues, i.e. the multiset of roots of the characteristic polynomial of the operator (in complex field), in which each element is repeated as many times as its algebraic multiplicity.

Proof. Let P a matrix representation on a given basis of the endomorphism ad_x. Then we have:

$$J^{-1}(x) = \det\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} P^k\right) = \det_{\mathbb{C}}\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} P^k\right), \tag{18}$$

where $\det_{\mathbb{C}}(\cdot)$ is the determinant in complex field. Formally this is the determinant applied to the complexification of the endomorphism. Now let $Q \in \mathrm{GL}(n, \mathbb{C})$ such that $P = Q^{-1}(D+N)Q$ where $(D+N)$ is the Jordan normal form of P where D is the diagonal matrix that has as entries elements of the spectrum of P and N is a nilpotent matrix. Then we have:

$$\det_{\mathbb{C}}\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} P^k\right) = \det_{\mathbb{C}}\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} (D+N)^k\right) = \det_{\mathbb{C}}\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} (D)^k\right), \tag{19}$$

where the last equality follows from the fact that $(D+N)^k = D^k + N'$ where N' is an another nilpotent matrix, and from the fact that the determinant of a triangular matrix depends only on the diagonal entries. Using the definition of D we can then write:

$$J^{-1}(\mathbf{x}) = \det_{\mathbb{C}}\left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} (D)^k\right) = \prod_{\lambda \in Sp(\operatorname{ad}_{\mathbf{x}})} \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!} (\lambda)^k\right) \tag{20}$$

Now if $\lambda = 0$ then $\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!}(\lambda)^k = 1$. Else, if $\lambda \neq 0$ then $\sum_{k=0}^{+\infty} \frac{(-1)^k}{(k+1)!}(\lambda)^k = \frac{1 - e^{-\lambda}}{\lambda}$ □

D.1 Matrix Lie Groups

In the case of a matrix Lie group $G < \mathrm{GL}(n, \mathbb{R}) \subseteq \mathrm{M}(n, \mathbb{R})$ we can exploit the fact our group is embedded in $\mathrm{M}(n, \mathbb{R}) \simeq \mathbb{R}^{n \times n}$ to give an alternative way to compute a matrix representation of $\mathrm{d}(L_{\exp(\mathbf{x})^{-1}})_{\exp(\mathbf{x})} \circ \mathrm{d}\exp_{\mathbf{x}}$. This corresponds to what in the literature is known as the Left Jacobian J_l

Here we show how we can derive the expression of J_l from the formal framework described in the previous Sections, using the additional information given by the fact that we are in a matrix Lie group. This is done using the fact that at each point $a \in G$ the tangent space $T_aG < T_a(\mathrm{GL}(n,\mathbb{R})) \tilde{=} \mathrm{M}(n,\mathbb{R})$ can be identified with a subspace of the real $n \times n$ matrices.

In fact let $p \in \mathrm{GL}(n,\mathbb{R})$, considering $\mathrm{GL}(n,\mathbb{R})$ as an open subset of $\mathrm{M}(n,\mathbb{R})$ then the canonical basis $(E_{ij})_{ij}$ of $\mathrm{M}(n,\mathbb{R})$ induces the isomorphism $\psi_p : \mathrm{M}(n,\mathbb{R}) \to T_p(\mathrm{GL}(n,\mathbb{R})),\ E_{ij} \mapsto \partial_{E_{ij}}|_p$. With this identification the diffe rential of the exp is a map from $\mathrm{M}(n,\mathbb{R})$ to $\mathrm{M}(n,\mathbb{R})$ and can be directly computed taking derivatives. The same holds for the differential of the left group action. Moreover the following Lemma shows that it corresponds to a matrix left multiplication. With this isomorphism we can see that the differential of left multiplication corresponds exactly to left matrix multiplication:

Lemma D.5. *Let $P, Q \in \mathrm{GL}(n,\mathbb{R})$ and let L_P the left action of P then $d(L_P)_Q$ identifying both the tangent spaces with $\mathrm{M}(n,\mathbb{R})$ using the isomorphisms ψ_P, ψ_{PQ} is the following function:*

$$d(L_P)_Q : \mathrm{M}(n,\mathbb{R}) \to \mathrm{M}(n,\mathbb{R}) \tag{21}$$
$$X \mapsto PX \tag{22}$$

Proof. Let $X \in \mathrm{M}(n,\mathbb{R})$ then $\forall f \in C^\infty(\mathrm{GL}(n,\mathbb{R}))$

$$\left[d(L_P)_Q\left(\partial_{X|_Q}\right)\right](f) = \partial_{X|_Q}(f \circ L_P) = \frac{d}{dt}_{|t=0}(f \circ L_P(Q + tX)) = \tag{23}$$
$$= \frac{d}{dt}_{|t=0} f\left(L_P(Q + tX)\right) = \frac{d}{dt}_{|t=0} f\left(PQ + tPX\right) = \left(\partial_{PX|_{PQ}}\right)(f) \tag{24}$$

□

These considerations lead to the following result:

Theorem D.6. *Now let $G < \mathrm{GL}(n,\mathbb{R})$ be a matrix Lie group, $\mathbb{B} := (\mathbf{v}_i)_i$ a basis of the Lie algebra. Then the Lie algebra endomorphism $d(L_{\exp(X)^{-1}})_{\exp(X)} \circ d\exp_X$ has matrix representation with respect to $\mathbb{B}$:*

$$J_l(X) = \left[\left(\exp(X)^{-1}\frac{\partial \exp}{\partial \mathbf{v}_1}(X)\right)^\vee \middle| \quad \cdots \quad \middle| \left(\exp(X)^{-1}\frac{\partial \exp}{\partial \mathbf{v}_n}(X)\right)^\vee\right] \in \mathrm{M}(n,\mathbb{R}) \tag{25}$$

Which is called the left-Jacobian. Where $(\cdot)^\vee := \varphi_{\mathbb{B}} : \mathfrak{g} \to \mathbb{R}^n$ is the ismomorphism given by the basis $\mathbb{B}$.

Proof. Considering G as embedded in $\mathrm{GL}(n,\mathbb{R})$ Then the tangent space at each point can be identified with a vector subspace of $\mathrm{M}(n,\mathbb{R})$.

Then given this identification, taking $X \in \mathfrak{g} \subseteq \mathrm{M}(n,\mathbb{R})$ the quantities $d\exp_X(\mathbf{v}_i) = \frac{\partial \exp}{\partial \mathbf{v}_i}(X) \in \mathrm{M}(n,\mathbb{R})$ are real valued matrices and can be simply obtained deriving the expression of the exponential in each entry. Moreover we have $\left[d(L_{\exp(X)^{-1}})_{\exp(X)} \circ d\exp_X\right](\mathbf{v}_i) = \exp(X)^{-1}\frac{\partial \exp}{\partial \mathbf{v}_i}(X) \in \mathfrak{g} \subseteq \mathrm{M}(n,\mathbb{R})$ where the equality is given by considering the left group action as the restriction of $L_{exp(X)} : \mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R})$ to $\mathfrak{g}$ and applying the Lemma D.5.This gives an explicit description on how the endomorphism acts on each vector of the basis. From this we can build its matrix representation $J_l(X)$. This gives us the thesis. □

E PUSHFORWARD DENSITY

E.1 Preliminary Lemmata

Lemma E.1 (See (Duistermaat and Kolk, 2000) Cor. 1.5.4)**.** *For a Lie Group G with algebra $\mathfrak{g}$ and exponential map $\exp : \mathfrak{g} \to G$, the set of singular points Σ is the set:*

$$\begin{aligned}\Sigma &= \{\mathbf{x} \in \mathfrak{g} | \det(T_{\mathbf{x}} \exp) = 0\} \\ &= \bigcup_{k \in \mathbb{Z}\setminus\{0\}} k\Sigma_1,\end{aligned}$$

with

$$\Sigma_1 := \{\mathbf{x} \in \mathfrak{g} | \det((\mathrm{ad}_{\mathbf{x}})_{\mathbb{C}} - 2\pi i I) = 0\},$$

where $(\mathrm{ad}\, X)_{\mathbb{C}}$ *denotes the adjoint representation of the real Lie algebra* $\mathfrak{g}$ *as a linear operator on the complex vector space.*

Lemma E.2. *Let* $f \in \mathbb{C}[X_1, \dots, X_n]$ *be a complex polynomial viewed as a function on the real vector space* $\mathbb{R}^n$*:*

$$f : \mathbb{R}^n \to \mathbb{C}, \quad x \mapsto f(x)$$

Then either f *is identically zero or the set of roots* $\{x \in \mathbb{R}^n \mid f(x) = 0\}$ *has Lebesgue measure zero in* $\mathbb{R}^n$*.*

Proof. The problem is reduced to the real polynomial $g \in \mathbb{R}[X_1, \dots, X_n]$ defined by

$$g := \mathrm{Re}(f)^2 + \mathrm{Im}(f)^2$$

It has the same set of (real) roots as f and g is identically zero if and only if f is. The statement then follows from the theorem of Okamoto. A simple proof can be found in (Caron and Traynor, 2005). □

Lemma E.3. *For a Lie Group* G *with algebra* $\mathfrak{g}$ *and exponential map* $\exp : \mathfrak{g} \to G$*, the set of singular points* Σ *is closed and has Lebesgue measure 0.*

Proof. Σ is closed because it is the preimage of the closed set $\{0\} \subset \mathbb{R}$ of the continuous function $\det(T_X \exp)$.

Let $f(X) = \det((\mathrm{ad}\, X)_{\mathbb{C}} - 2\pi i I)$. f is a polynomial in X, because ad is linear and det polynomial. f can not be identically zero, as $\{0\} \notin \Sigma$, because exp is a diffeomorphism in a neighbourhood of $0 \in \mathfrak{g}$ (see (Duistermaat and Kolk, 2000) 1.3.4). Thus, the set of roots of f, namely Σ_1, has Lebesgue measure zero. It follows that also Σ has Lebesgue measure zero. □

Definition 5. *(Sets of Lebesgue measure 0 on a Manifold) If* M *is a smooth n-manifold we say that a subset* $A \subseteq M$ *has measure zero in* M *if for every smooth chart* (U, φ) *the subset* $\varphi(A \cap U) \subseteq \mathbb{R}^n$ *has n-dimensional measure zero.*

Lemma E.4. *Let* M *a smooth manifold. Then* $\forall p \in M$ *and* U *open neighbourhood of* p *there exists* $U' \subseteq U$ *open neighbourhood of* p *such that* $\partial U'$ *has Lebesgue measure* 0.

Proof. Take a smooth chart (V, φ) such that $p \in V$. Let $V' := V \cap U$ open set. Then $\varphi(V')$ is an open set in $\mathbb{R}^n$ such that $\varphi(p) \in \varphi(V')$. Take then an open ball $B(\varphi(p), r)$ with $r > 0$ such that $\subseteq \varphi(V')$. If define $U' := \varphi^{-1}(B(\varphi(p), r))$ we have that U' is an open neighborhood of p and that $\varphi(\partial U') = \partial B(\varphi(p), r)$ has measure 0 in $\mathbb{R}^n$. Then Lemma 6.6 of Lee (2012) implies that $\partial U'$ has measure 0 in M □

Lemma E.5. *Let* N *and* M *smooth manifolds of the same dimension and* $F : M \to N$ *a smooth map. Let* $D := \{p \in M : F \text{ is a local diffeomorphism at } p\} \subseteq M$*. Then* D *can be partitioned in* $D = B \bigcup \left(\cup_{k=1}^{+\infty} A_k\right)$ *such that* B *has Lebesgue measure 0 and for every* k A_k *is an open set such that* $F|_{A_k} : A_k \to F(A_k)$ *is a diffeomorphism.*

Proof. We first show that D is open: $\forall p \in D$ since F is a local diffeomorphism at p there exists a neighbourhood $U_p \ni p$ such that $F|_{U_p}$ is a diffeomorphism. Then $U_p \subseteq D$. This shows that $\mathring{D} = D$ thus D is open. Therefore D inherits a manifold structure from M as a sub-manifold, meaning that D is second countable, implying D is Lindelöf (see (Lee, 2010), Thm. 2.50). This means that every open cover has a countable subcover.

For every $p \in D$ consider $U_p \in D$, neighbourhood of p such that $F|_{U_p}$ is a diffeomorphism. Then by Lemma E.4 there exists $U'_p \subseteq U_p$ open neighbourhood of p such that $\partial U'_p$ has Lebesgue measure 0. Consider then the cover $\{U'_p : p \in D\}$, by Lindelöf property it has a countable subcover $\{A'_n\}_{n=1}^{+\infty}$. We then iteratively build the sets $A_1 := A'_1$, $A_n := A'_n \setminus \overline{\left(\cup_{k=1}^{n-1} A'_k\right)}$. Then by construction the sets A_n are open and $F|_{A_n}$ is a diffeomorphism. Moreover defining $B := D \setminus \left(\cup_{k=1}^{+\infty} A_k\right)$ we are are left to show that B has Lebesgue measure 0. This simply follows from the fact that by construction $B \subseteq \cup_{k=1}^{+\infty} \partial A'_k$ and that the sets A'_k were defined to have boundary of Lebesgue measure 0. To see that $B \subseteq \cup_{k=1}^{+\infty} \partial A'_k$ consider $b \in B$ and define the set $N_b = \{n \in \mathbb{N}^+ : b \in A'_n\}$, the set is not empty since the sets A'_k form a cover of D. Let then $m \in N_b$ be the smallest element in N_b. Since $b \in B$ then $b \notin A_1 = A'_1$ therefore $m > 1$. Moreover $b \notin A_m$ and since $b \in A'_m$ we have that $b \in \overline{\left(\cup_{k=1}^{m-1} A'_k\right)} = \left(\cup_{k=1}^{m-1} A'_k\right) \cup \partial\left(\cup_{k=1}^{m-1} A'_k\right)$. By definition of m, $b \notin \left(\cup_{k=1}^{m-1} A'_k\right)$, then $b \in \partial\left(\cup_{k=1}^{m-1} A'_k\right) \subseteq \cup_{k=1}^{m-1} \partial A'_k \subset \cup_{k=1}^{+\infty} \partial A'_k$ □

E.2 Main Theorem

Now suppose we have samples from a measure $\mathrm{m} \ll \lambda$ with density r. We can then "push" the samples to elements in G through the exp map. The resulting samples will be distributed according to the pushforward measure $\exp_*(\mathrm{m})$ on G. The following theorem ensures that $\exp_*(\mathrm{m})$ is a.c. with respect to the left Haar measure ν and gives an expression for the density

Theorem E.6. *Let G, $\mathfrak{g}$, m, λ, ν defined as above. Then $\exp_*(m) \ll \nu$ with density:*

$$p(a) = \sum_{\{\mathbf{x}\in\mathfrak{g}:\exp(\mathbf{x})=a\}} r(\mathbf{x})|J(\mathbf{x})|^{-1}, \tag{26}$$

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

Proof. Using Lemma E.3, we partition $\mathfrak{g}$ in the open set A such that A is the set of points in which exp is a local diffeomorphism and $\Sigma := \mathfrak{g} \setminus A$.

Using Lemma E.5, we further partition A in countably many open sets $\{A_k\}_{k\in I}$, for some index set I, and a set B, such that $\exp|_{A_k}$ is a diffeomorphism for all k and B is of Lebesgue measure 0. Define for $S \in \mathcal{B}[\mathfrak{g}]$:

$$\mathrm{m}_S : \mathcal{B}[\mathfrak{g}] \to [0,1] : E \mapsto \mathrm{m}(E \cap S)$$

Then we have, since $\mathrm{m} \ll \lambda$ and $\lambda(\Sigma) = \lambda(B) = 0$:

$$\mathrm{m} = \mathrm{m}_\Sigma + \mathrm{m}_B + \sum_{k\in I} \mathrm{m}_{A_k} = \sum_{k\in I} \mathrm{m}_{A_k}$$

Consider the pushforward measure $\exp_*(\mathrm{m})$, we have for all $D \in \mathcal{B}[G]$:

$$\begin{aligned}
(\exp_*(\mathrm{m}))(D) &= \mathrm{m}(\exp^{-1}(D)) \\
&= \sum_{k\in I} \mathrm{m}_{A_k}(\exp^{-1}(D)) \\
&= \sum_{k\in I} (\exp_*(\mathrm{m}_{A_k}))(D) \\
&= \sum_{k\in I} ((\exp_{|A_k})_*(\mathrm{m}))(D) \\
\implies \exp_*(\mathrm{m}) &= \sum_{k\in I} (\exp_{|A_k})_*(\mathrm{m}),
\end{aligned}$$

where we define:

$$((\exp_{|A_k})_*(\mathrm{m}))(D) = \mathrm{m}(\exp^{-1}(D) \cap A_k)$$

Notice that $\exp_{|A_k} : A_k \to \exp(A_k)$ is now a diffeomorphism, so the change of variable formula in (14) can be applied:

$$\begin{aligned}
((\exp_{|A_k})_*(\mathrm{m}))(D) &= \int_{\exp_{|A_k}^{-1}(D)\cap A_k} r \; d\lambda \\
&= \int_{D\cap\exp(A_k)} (r \circ \exp_{|A_k}^{-1}) \cdot (|J|^{-1} \circ \exp_{|A_k}^{-1}) \; d\nu
\end{aligned} \tag{27}$$

Then $(\exp_{|A_k})_*(\mathrm{m}) \ll \nu$ and since $\exp_*(\mathrm{m}) = \sum_{k\in I}((\exp_{|A_k})_*(\mathrm{m}))$ then $\exp_*(\mathrm{m}) \ll \nu$.

In order to find the expression for the density we observe that $(\exp_{|A_k})_*(\mathrm{m})$ has density $r(\exp_{|A_k}^{-1}(a))J^{-1}(\exp_{|A_k}^{-1}(a))\mathbb{I}_{\exp(A_k)}(a)$ where $a \in G$ and $\mathbb{I}$ is the indicator function. Then we have that the density of $\exp_*(\mathrm{m})$ with respect to ν is

$$\sum_{k \in I} r(\exp_{|A_k}^{-1}(a))|J^{-1}(\exp_{|A_k}^{-1}(a))|\mathbb{I}_{\exp(A_k)}(a) = \sum_{\{\mathbf{x} \in \mathfrak{g} : \exp(\mathbf{x}) = a\}} r(\mathbf{x})|J(\mathbf{x})|^{-1}, \tag{28}$$

where the last equality is true almost everywhere in G. This can be seen if we define the set $N \subseteq G$ as all the points $p \in G$ in which $\{\exp_{|A_k}^{-1}(p) : k \in I\} \neq \{x \in \mathfrak{g} : \exp(x) = p\}$. Then N has Lebesgue measure 0. In fact $N \subseteq \exp(B \cup \Sigma)$ and since $B \cup \Sigma$ has measure zero in $\mathfrak{g}$ and exp is smooth then by Theorem 6.9 in Lee (2012) $\exp(B \cup \Sigma)$ has measure 0.

□

F COMPUTATIONAL COMPLEXITY

F.1 Complexity of the Reparameterization Trick

In this appendix we will analyze the complexity of performing the reparameterization trick when working with a Lie group G of dimension n. For simplicity we will assume in the following considerations that G is a matrix Lie group. The complexity is given by the cost of computing the exp map and its differential. The exp map for a matrix lie group is given by the matrix exponential

$$\exp(X) = \sum_{k=0}^{\infty} \frac{X^k}{k!} \quad X \in \mathrm{M}(n, \mathbb{R}), \tag{29}$$

which involves an infinite summation. In general the worst case complexity for computing a good approximation of the above expression is $O(n^3)$ [18].

For the differential of the exp map, the computation via the left-Jacobian (25) is generally also cubic in n, as it involves a matrix inversion. The alternative is to use equation (17) in Theorem D.4, in which case the complexity is cubic in n as well. In fact because the Lie algebra $\mathfrak{g}$ is a vector space of dimension n, then since $\mathrm{ad}_{\mathbf{x}} \in \mathrm{End}(\mathfrak{g})$ fixed a basis for $\mathfrak{g}$, $\mathrm{ad}_{\mathbf{x}}$ has a matrix representation as an element of $\mathrm{GL}(n, \mathbb{R})$. One can either compute the exp of this matrix, or find its eigenvalues, both operations are cubic in n.

Despite the above considerations, for specific Lie Group there might exist specific analytic calculations to derive closed form expressions for the exponential map and for the eigenvalues of the adjoint, using group specific properties. This can in practice lead to a significant reduction in computational complexity, as it is shown in the specific examples of Section 4.2.

F.2 Approximation of Infinite Summations

In Appendix E we have proven that the the pushforward measure of a probability measure in the Lie algebra is a well defined measure on the Lie Group, with a density with respect to the Haar measure on the group. However the expression of the density at a point depends on a potentially infinite summation. In general since $1 = [\exp_*(\mathrm{m})](G) = \int_G p d\nu$, the density is finite almost everywhere in G. This means that the infinite series can be truncated at the N-th term, still retaining an arbitrarily good approximation (that depends on N). In practice we have observed that when using an exponentially decaying distribution on the lie algebra, only an handful of terms are sufficient to get a good approximation. However it is difficult to derive general bounds and to determine a priori a good value for N, as this will greatly depend on the choice of base distribution, on the specific Lie Group and on the way we decide to enumerate the points in $\exp^{-1}(a)$.

A possible alternative to avoid infinite summations is to use a compactly supported distribution, this reduces the infinite series to a finite summation, since the terms become definitely 0. Notice that since compactly supported functions are dense in $L^1(\mathbb{R}^n)$ and that $r \in L^1(\mathbb{R}^n)$, there is always a compactly supported function that approximates r arbitrarily well.

[18] The interested reader is referred to (Moler and Van Loan, 2003) for a survey on the possible ways to compute matrix exponentials with a detailed explanation for the complexity of each method.

Connected to this approach, it is possible to choose a density supported in the injectivity radius of the exponential map. The summation then reduces to one term. Moreover if the base density r is smooth then the density of the pushforward will also be smooth.

G CHOICE OF BASIS AND SCALAR PRODUCT IN THE LIE ALGEBRA

In the previous Sections the starting point for obtaining a reparameterizable density on the Lie Group G was using a reparameterizable density on the corresponding Lie algebra $\mathfrak{g}$.

Since computations usually can only be done on real values we need a concrete representation of the abstractly defined Lie algebra $\mathfrak{g}$ as some real vector space $\mathbb{R}^n$.

This amounts to say that we need to choose a concrete basis $b = (b_1, \ldots, b_n)$ with the $b_i \in \mathfrak{g}$ and identify linear combinations $v = \sum_{i=1}^n x_i \cdot b_i \in \mathfrak{g}$ with the corresponding vector $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Every such a choice of basis gives us a linear isomorphism:

$$\psi_b : \mathbb{R}^n \cong \mathfrak{g}, \qquad x = (x_1, \ldots, x_n)^T \mapsto \sum_{i=1}^n x_i \cdot b_i$$

Furthermore, the standard scalar product on $\mathbb{R}^n$ induces a scalar product $< .,. >$ on $\mathfrak{g}$ via aboves isomorphism.

We can then proceed in two ways:

1. In case we can directly and intrinsically define a (probability) measure m on $\mathfrak{g}$ then we can take any basis b and push m via $\varphi_b := \psi_b^{-1}$ to $\mathbb{R}^n$ (to get $\varphi_{b,*}\mathrm{m}$). We then can use the real valued representation there to reparameterize the corresponding density. All results can then be pulled back to $\mathfrak{g}$ with ψ_b.

2. The second way is to start directly from a reparameterizable measure m' on $\mathbb{R}^n$ and then define: $\mathrm{m}_b := \psi_{b,*}\mathrm{m}'$.

Even though both view points seem to be equivalent, only the first one is independent of the representation as the "true" measure m on $\mathfrak{g}$ was already given. The second method will highly depend on the choice of the basis b and the measure m'. Therefore, if possible, the first approach is preferred. However in practice specifying measures or densities directly in $\mathbb{R}^n$ is easier as the abstract definition of $\mathfrak{g}$ is not directly accessible. We will discuss this further in the following.

As mentioned before, the standard scalar product on $\mathbb{R}^n$ induces a scalar product $< .,. >$ on $\mathfrak{g}$ via aboves isomorphism and thus a left-invariant Riemannian metric on G.

So the whole Riemannian geometric structure of the Lie group G is sensitive to the choice of the basis on $\mathfrak{g}$. Also, if we would now sample from a skewed distribution $p(x)$ on $\mathbb{R}^n$ and push the samples to G via the maps:

$$\mathbb{R}^n \overset{\psi_b}{\cong} \mathfrak{g} \xrightarrow{\exp} G,$$

then these would in general not be the same as when using another basis for the isomorphism. To summarize, we need to choose the basis carefully and keep the dependence on it in mind.

Now let us assume that we already have a specified scalar product $< .,. >$ on $\mathfrak{g}$. Then a natural choice would be to take a orthonormal basis $b = (b_1, \ldots, b_n)$ w.r.t. the given scalar product, i.e. we have: $< b_i, b_j >= \delta_{i,j}$. Then still skewed distributions $p(x)$ on $\mathbb{R}^n$ would be mapped to different distributions under a different choice of orthonormal basis. In case $p(x)$ is invariant under orthonormal transformations (i.e. $p(g.x) = p(x)$ for all $g \in O(n)$ and $x \in \mathbb{R}^n$) like Normal distributions of form $p(x) = \mathcal{N}(x|0, \sigma^2 \cdot I)$ then the pushforward of $p(x)$ onto G would not depend on the choice of orthonormal basis.

But note that the notion of orthonormality strongly depends on the chosen scalar product $< .,. >$ on $\mathfrak{g}$ and the number of choices one can make are i.g. infinite. Different scalar products lead to different orthonormal basises.

So it is left to discuss how to choose a scalar product on $\mathfrak{g}$ or a Riemannian metric g on G, resp.. To reduce the number of Riemannian metrics we can impose additional desirable properties onto them, like bi-invariance.

Theorem G.1 (See (Milnor, 1976; Alexandrino and Bettiol, 2015)). *1. Any Lie group G that is isomorphic to the direct product of a compact Lie group K and $\mathbb{R}^n$, $n \geq 0$, admits a bi-invariant (i.e. left- and right-invariant) Riemannian metric g.*

2. If G is connected then also the reverse statement holds.

3. If G admits a bi-invariant Riemannian metric then the Lie exponential map and the Riemannian exponential map at the identity agree.

4. If G is a compact and simple Lie group then the bi-invariant Riemannian metric is unique up to a positive constant $c > 0$.

It turns out that for certain types of Lie groups there is even a natural choice of scalar product, the so called *negative Killing form.*

Theorem G.2 (See (Milnor, 1976; Alexandrino and Bettiol, 2015)). *Let G be a Lie group and for $x, y \in \mathfrak{g}$ define the* negative Killing form *as:*

$$< x, y >:= -\mathrm{tr}\left(\mathrm{ad}_x \circ \mathrm{ad}_y\right)$$

We then have the following results:

1. G is semisimple iff and $< .,. >$ is non-degenerate.

2. If G is semisimple and compact then $< .,. >$ induces a bi-invariant Riemannian metric on G.

G.1 Summary

1. Consider the case that we have a simple and compact Lie groups G (e.g. $SO(2)$ or $SO(3)$).
2. Then take the negative Killing form (up to scale $c > 0$) as scalar product on $\mathfrak{g}$:

$$< x, y >:= -\mathrm{tr}\left(\mathrm{ad}_x \circ \mathrm{ad}_y\right).$$

3. The left multiplications L_a of $< .,. >$, for $a \in G$, then induces a bi-invariant Riemannian metric g on G.
4. g induces the bi-invariant Haar measure m_g on G, which on arbitrary local charts is given by the density $\sqrt{|\det(g)|}$ w.r.t. Lebesgue measure.
5. In case we can compute $\mathrm{m}_g(G)$, re-scaling the scalar product by multiplying it with the factor $c := \frac{1}{\sqrt[n]{\mathrm{m}_g(G)^2}}$ with $n = \dim(G)$ makes the then induced bi-invariant Haar measure normalized (i.e. $\mathrm{m}_g(G) = 1$).
6. In any case, choose a orthonormal basis $b_1, \ldots, b_n$ of $\mathfrak{g}$ w.r.t. $< .,. >$ and fix the isomorphism:

$$\varphi_b: \mathbb{R}^n \cong \mathfrak{g}, \qquad x = (x_1, \ldots, x_n)^T \mapsto \sum_{i=1}^n x_i \cdot b_i$$

7. Then the pushforward (via exp) onto G of probability distributions $p(x)$ on $\mathbb{R}^n$ that are invariant under $O(n)$ w.r.t. $< .,. >$ are independent of the chosen basis and independent of the chosen bi-invariant metric up to scale.
8. For example for the Normal distribution $p(x) = \mathcal{N}(x|0, \sigma^2 \cdot I)$ this basically just reduces to the choice of variance σ^2 (even when not normalized, since multiplication with $c > 0$ only changes the variance).

Remark 1. *If G is only a semisimple Lie group then the negative Killing form $< .,. >$ on $\mathfrak{g}$ can still be used to induce a bi-invariant pseudo-Riemannian metric on G and thus a bi-invariant Haar measure m_g, which still on local oriented coordinates is given by the density $\sqrt{|\det(g)|}$ w.r.t. the Lebesgue measure.*