Reparameterizing Distributions on Lie Groups

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

University of Amsterdam

Aiconic

U.C. Berkeley

A Probabilistic Model of The World

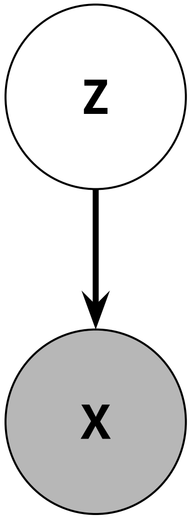
- Generally, in machine learning we aspire to model the world around us
- We attempt to infer latent structure responsible for observed events
- Unfortunately, our observations can be **incomplete** and **noisy**
- Hence, we utilize a probabilistic approach to allow for imperfect predictions

Variational Inference

- Given: Observations (**X**)   Assumed: Latent Structure (**Z**)
  - We would like to infer the posterior over latent variables given observations, p(**Z**|**X**)
- Deep Learning: Powerful to learn arbitrary functions

- Bayes Rule: intractable in practice
- Alternative: **Variational Inference (VI)**
  - transforms inference task into an optimization problem
  - find distribution q(**Z**), that best approximates true posterior

- **Reparameterization trick** crucial to enable VI in DL



Reparameterizable distributions

$$\nabla_\theta \mathbb{E}_{z\sim q(\mathbf{z};\theta)}[f(\mathbf{z})] = \boxed{\nabla_\theta} \mathbb{E}_{\epsilon\sim s(\epsilon)}[f(\mathcal{T}(\epsilon;\theta))]$$
$$= \mathbb{E}_{\epsilon\sim s(\epsilon)}[\boxed{\nabla_\theta} f(\mathcal{T}(\epsilon;\theta))]$$

where $\mathcal{T}(\epsilon;\theta) = z \sim q(z;\theta)$, if $\epsilon \sim s(\epsilon)$

- We want to:
  - Take unbiased, low variance gradients with respect to the distribution’s parameters
- We need to be able to:
  - Take samples of the distribution by first sampling a simple distribution independent on the parameters and then transforming it into the target distribution
  - Compute the density (often useful for computing KL using MC estimates)

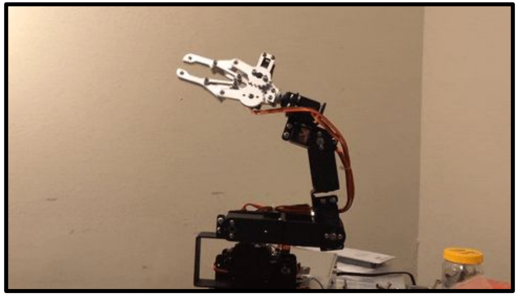
So, is reparameterizing distributions solved?

- **No general trick** to reparameterize arbitrary density

- In recent years much research has been done on extending the class of reprametrizable densities:
    - Normalizing Flows (Rezender & Mohammed 2014)
    - Acceptance Rejection Sampling (Naesseth et al. 2017)
    - Implicit Reparameterization (Figurnov et al. 2018)

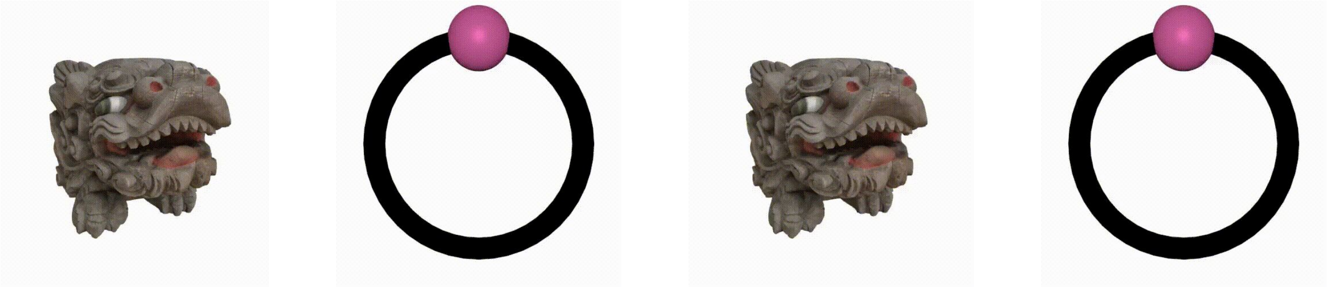
- However almost all past work assumes a **‘flat’ Euclidean space**

Lie Groups in the Wild



- Yet **many problems** are naturally defined on **non-trivial spaces**:
    - Example 1: estimate position and orientation of a Drone
    - Example 2: infer camera position and orientation
    - Example 3: maximum entropy reinforcement learning on a Robotic Arm
    - Example 4: inference of an orthogonal matrix in a Bayesian Neural Network
- All of the above can be described in terms of ‘Lie Groups’ !

What is a Lie Group



- Manifold: **locally** Euclidean, but **globally** non-trivial
- Lie Groups, manifolds with group structure:
  - The elements of the space can operate on the space itself by group multiplication
- Very useful construct to describe changes of objects over time!

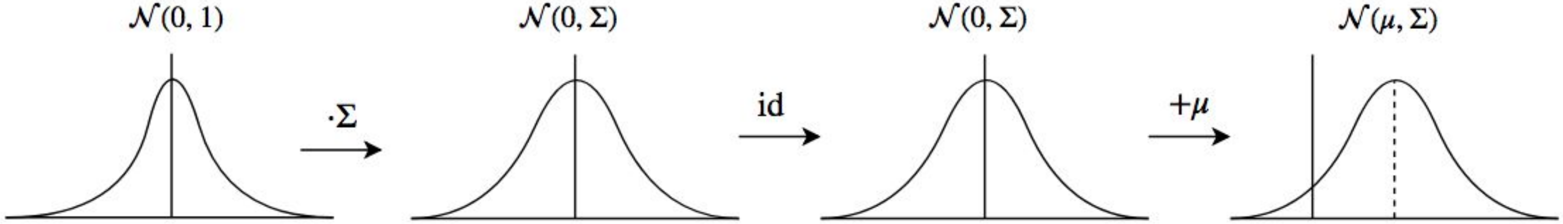
Lie Group Examples

SO(2)

SO(3)

SE(3)

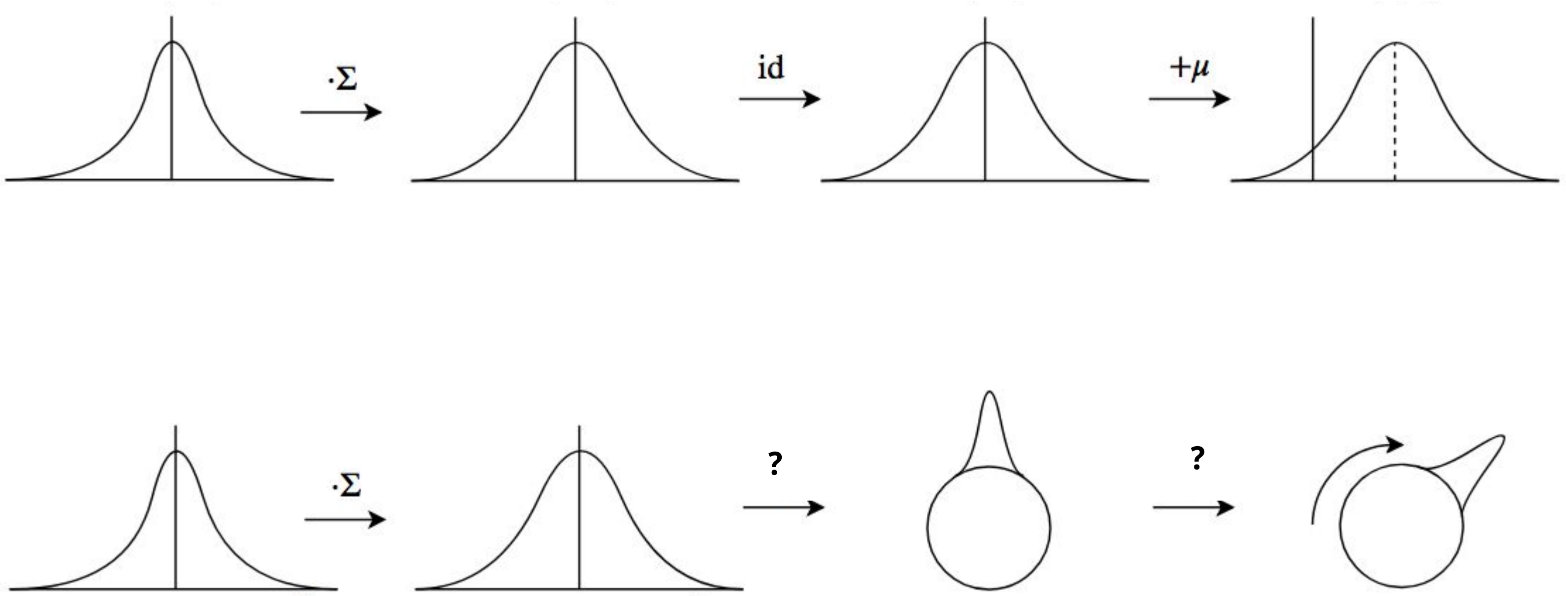
Gaussian Reparametrization



Location-scale reparameterization trick

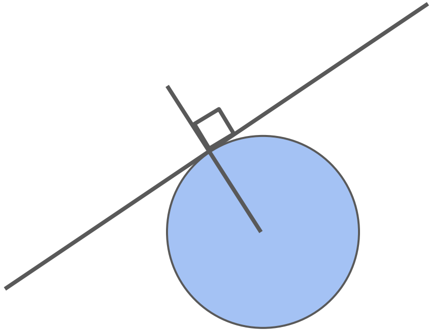
a. Sample from standard Gaussian
b. Change scale
c. Change location by using translation operation

Reparameterize Lie Groups



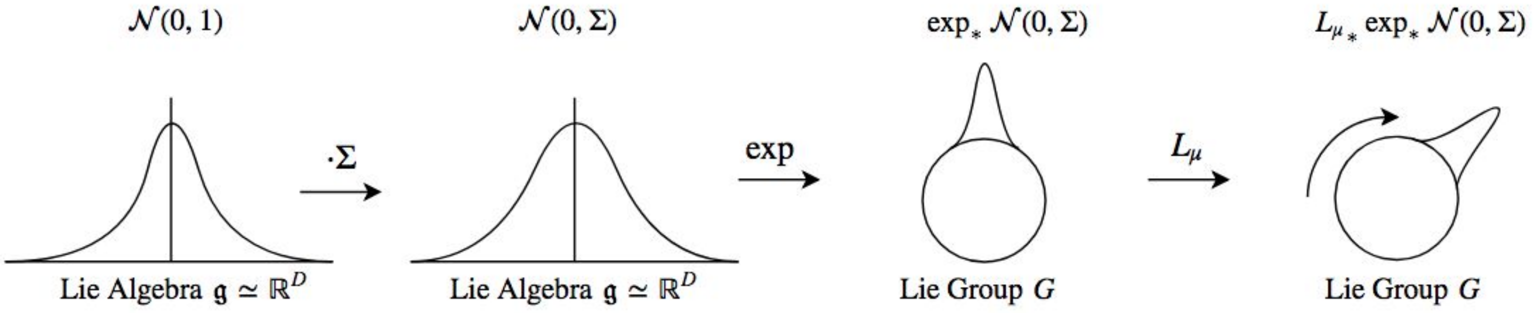
Reparameterize Lie Group

- Reparametrize a density in Euclidean space:
  - Tangent space at group identity: Lie algebra $\mathfrak{g} \simeq \mathbb{R}^D$ [SO(2): $\mathbb{R}$]
- Wrapping the algebra to the Group:
  - Exponential map: $\exp : \mathbb{R}^D \simeq \mathfrak{g} \longrightarrow \mathcal{G}$ [SO(2): mod 2π]
  - For matrix Lie groups this is the matrix exponential
- Moving the distribution on the lie group on a target position:
  - Action of group on itself by left group multiplication $L_\mu(g) = \mu \cdot g$ [SO(2): +θ mod 2π]



Reparameterize General Lie Group:

- Reparameterize density in $\mathbb{R}^N$
- Exponential map
- Act by left multiplication



Where does the difficulty 'Lie'?

- Exponential map:
    - In general non-injective
    - In some groups can be constant under local variations

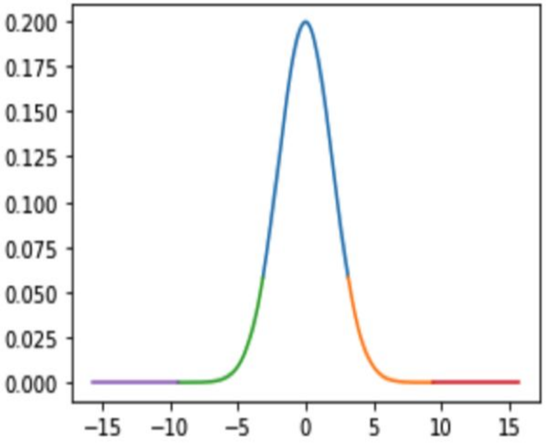
Where does the difficulty ‘Lie’?

- Exponential map:
  - In general **non-injective**
  - In some groups can be constant under local variations (in a set of measure zero)
- We prove it has a density

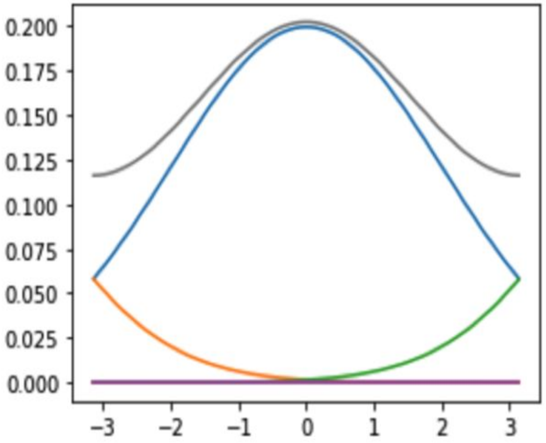
$$p(a) = \sum_{\mathbf{x} \in \mathfrak{g} : \exp(\mathbf{x}) = a} r(\mathbf{x}) |J(\mathbf{x})|^{-1}$$

- Contributions from all pre-image algebra elements
- Change of volume
  - From auto-diff
  - From Lie Algebra representation

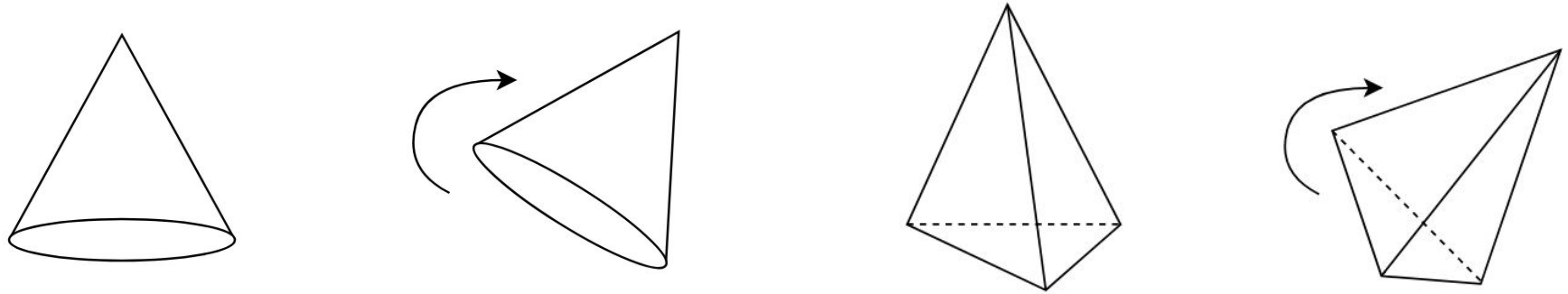
Non-injectivity



Exponential Map



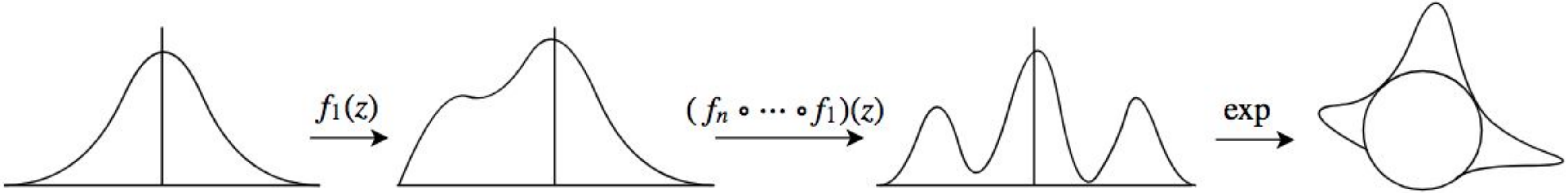
Synthetic example: SO(3)



(a) (b)

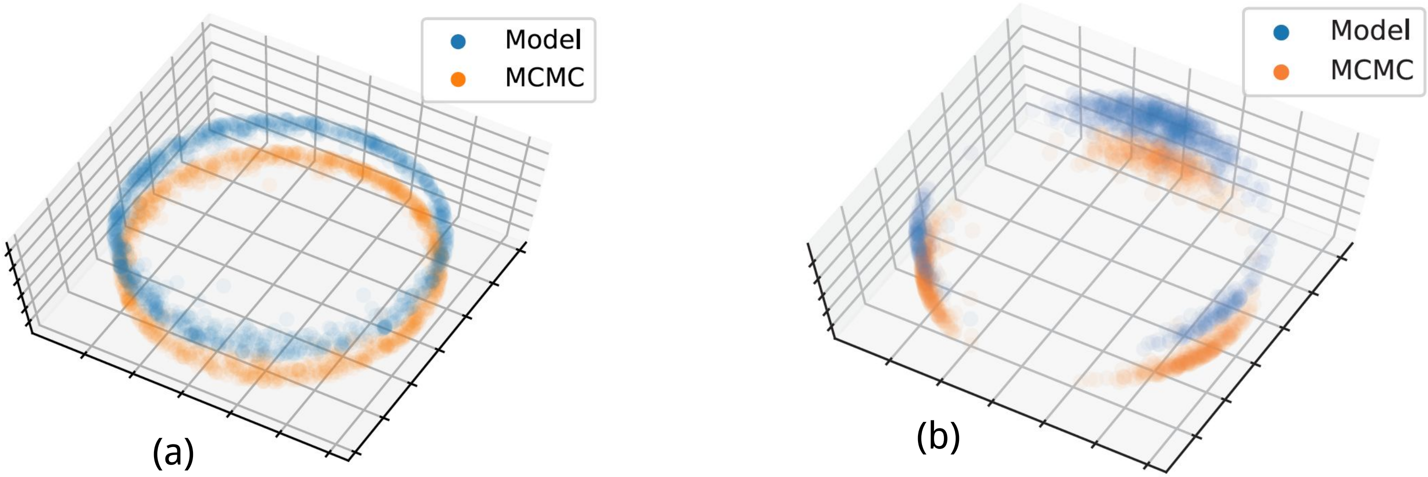
- Group of 3D oriented rotations, describe rotations of static object
- Task: given two views of an object, find which rotation(s) relate(s) them
- Variational Inference on SO(3)-valued latent
    - Object (a) has rotational symmetry around one axis
    - Object (b) has threefold discrete symmetry around one axis

Normalizing Flow



- True posterior: circular subgroup / discrete subgroup (order 3)
- **Multi-modal** distributions needed:
  - Recall we can use **arbitrary base distribution** in the Lie Algebra
  - Why not use Normalizing Flows (NF)?
- Basic NF: use repeated invertible transformations to create complex density
  - Parameterize very complex distribution in Lie Algebra and push to Lie Group

Results



- Markov Chain Monte Carlo true posterior v. approximate model posterior
- Modeled results are lifted in z-axis for clarity
- Almost perfectly capture latent dynamics!

Takeaway

- Go beyond Euclidean spaces!
- Easy to use framework for distributions on all Lie Groups
- Can express powerful multimodal distributions
- Usable for:
    - Variational Inference
    - Maximum Likelihood Estimation

(For technical details + lively discussion: visit our Poster **#We116** later today!)

Thank You for your Time!

Contact: **luca.falorsi@aiconic.com** **pimdehaan@gmail.com** **tim.davidson@aiconic.com**