

# Student Project Report

Magnus Dierking

September 2024

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>3</b>
2.1	Group Theory . . . . .	3
2.1.1	Group Action and Representation . . . . .	4
2.1.2	Left Action and Group Conjugacy . . . . .	4
2.2	Riemannian Manifolds . . . . .	7
2.2.1	Lie Derivative and Pushforward . . . . .	8
2.2.2	Geodesics and Exponential Map . . . . .	8
2.3	Diffusion Models . . . . .	9
2.3.1	Concept . . . . .	10
2.3.2	Training . . . . .	11
<b>3</b>	<b>Lie Theory for <math>\text{SO}(3)</math> and <math>\text{SE}(3)</math></b>	<b>13</b>
3.1	Left-Invariant Vector Fields . . . . .	13
3.2	The Exponential and Log Map of Lie Groups . . . . .	14
3.3	Tangent Spaces and Lie Algebra . . . . .	14
3.3.1	$\text{SO}(3)$ . . . . .	15
3.3.2	$\text{SE}(3)$ . . . . .	16
3.4	Adjoint Representation and Commutator . . . . .	17
3.4.1	$\text{SO}(3)$ . . . . .	18
3.4.2	$\text{SE}(3)$ . . . . .	19
3.5	Metrics on $\text{SO}(3)$ . . . . .	19
<b>4</b>	<b>Brownian Motion on <math>\text{SO}(3)</math></b>	<b>21</b>
4.1	The Heat Kernel and Gaussian Densities . . . . .	21
4.2	The Haar Measure and Inverse Transform Sampling . . . . .	22
<b>5</b>	<b>Diffusion on <math>\text{SO}(3)</math></b>	<b>24</b>
5.1	Toy Example on $\text{SO}(3)$ . . . . .	24
5.2	Implementation Remarks . . . . .	26
<b>6</b>	<b>Topology Bits</b>	<b>27</b>
6.1	Introduction . . . . .	27
6.2	The case $\text{SO}(3)$ and Relationship with other Rotation Representations . . . . .	28
6.3	Practical Implications . . . . .	29

# 1 Introduction

This document represents the final report of my student project undertaken during the summer term of 2024 at TU Darmstadt. The primary objective of this project is to explore the mathematical structures of rotations and rigid body transformations as Lie Groups, with the intention of laying a foundation for a potential Master's thesis on these concepts.

Rotations and rigid body transformations are central in both theoretical mathematics and various applied fields, such as robotics, computer vision, and mechanics. However, these fields often introduce these concepts from distinct perspectives. On one hand, pure mathematics tends to emphasize abstract, rigorous definitions, and the derivation of properties, frequently without prioritizing practical applications or intuitive understanding. On the other hand, in applied disciplines, such as robotics or computer vision, the theoretical foundations are often overlooked or simplified. In these fields, rotations and rigid body transformations are typically treated as practical tools, with focus placed on building intuitive understanding and achieving functional results.

In this report, we aim to bridge this gap to some extent by connecting the abstract mathematical foundations of rotations and rigid body transformations to a more application-oriented viewpoint. Our approach focuses on developing intuitive insights into the central ideas of Lie Groups, while maintaining a strong connection to their formal structure. By doing so, we hope to make some of the more abstract concepts that are central to Lie Groups and Lie algebras feel more accessible and intuitive.

The results presented in this report are based on a combination of theoretical studies in differential geometry, Riemannian geometry, and group theory, complemented by lectures attended at TU Darmstadt. Additionally, research papers that focus on practical applications of these concepts in areas such as robotics and computer vision have provided valuable insights and have influenced the direction of this work.

We begin by examining how rotations in three-dimensional space can be described using linear transformations, specifically the set of orthogonal matrices with determinant 1, denoted by the special orthogonal group:

$$\text{SO}(3) := \{\mathbf{R} \in \mathbb{R}^{3 \times 3} : \mathbf{R}\mathbf{R}^T = \mathbf{I}, \det(\mathbf{R}) = 1\} \subset \text{O}(3), \quad (1)$$

where  $\text{O}(n)$  represents the broader group of orthogonal matrices that also allow  $\det(\mathbf{R}) = -1$ , which includes reflections in addition to rotations. Thus,  $\text{SO}(n)$  forms a subset of  $\text{O}(n)$ , capturing only the proper rotations, excluding reflections.

To describe the full range of rigid body motions, which combine both rotational and translational components, we extend this concept by defining the special Euclidean group, denoted as  $\text{SE}(3)$ :

$$\text{SE}(3) := \{g = (\mathbf{R}, \mathbf{t}) | \mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3\}. \quad (2)$$

Here, each element of  $\text{SE}(3)$  consists of a rotation  $\mathbf{R}$  from  $\mathbf{R} \in \text{SO}(3)$  and a translation vector  $\mathbf{t} \in \mathbb{R}^3$ , forming a group that can describe all rigid body transformations in three-dimensional space. Both  $\text{SO}(3)$  and  $\text{SE}(3)$  are examples of Lie Groups, a special class of mathematical objects that possess a particularly rich structure. A Lie Group combines the properties of a group (an algebraic structure equipped with an operation that satisfies certain axioms such as closure, associativity, and the existence of an identity element and inverses) with the smooth, geometric structure of a differentiable manifold. In particular, the Lie Groups  $\text{SO}(3)$  and  $\text{SE}(3)$  exhibit additional structure due to their relationship with Riemannian geometry, which allows for the study of these groups using techniques from both algebra and differential geometry.

In the following sections, we will delve deeper into the mathematical structure of these Lie Groups, analyzing their properties and understanding how they are used to model physical phenomena in both theoretical and applied settings. Specifically, we will examine the Lie algebras associated with  $\text{SO}(3)$  and  $\text{SE}(3)$ , which provide powerful tools for studying the local behavior of these groups and their corresponding manifolds.

## 2 Preliminaries

A valuable general perspective in many branches of mathematics is the idea of applying structure to a set. A set, in its most basic form, is simply a collection of abstract elements with no inherent relationships or additional properties. However, by endowing a set with a specific structure, we obtain a "space," which we can then analyze according to the rules of the structure we've introduced. The type of space is determined by the nature of the structure applied: for instance, adding a topology creates a topological space, while introducing a metric defines a metric space, and so forth. This abstraction allows mathematicians to move from purely set-theoretic thinking to studying more nuanced properties that arise when additional structure is introduced. The significance of this approach is that it reveals how certain properties of the elements of a set are influenced and constrained by the imposed structure. For example, in a metric space, we gain the notion of distance, which leads to concepts like convergence and continuity that have no meaning in the set itself without a metric.

In this work, we focus on the application of the theoretical framework of Lie groups, which are mathematical structures that emerge from combining two well-established concepts: groups and manifolds.

This dual structure means that Lie groups are not only algebraically rich, providing a group operation that satisfies familiar properties, but they also carry a geometric and smooth structure, allowing for differential calculus to be applied. The interaction between these two structures leads to deep and far-reaching mathematical properties, as well as a wide range of applications in fields like geometry, physics, and engineering.

In particular, the study of Lie groups provides a natural framework for understanding continuous symmetries, which are key to understanding physical transformations in three-dimensional space. For instance, the special orthogonal group  $SO(3)$  describes the symmetries of rotations in 3D space, while the special Euclidean group  $SE(3)$  extends this by including translations, describing the full range of rigid body transformations.

### 2.1 Group Theory

A group is a set  $G$  together with any binary operation  $(\cdot)$  on the elements of that set, which satisfies the following four fundamental properties [10]:

- **Closure:** combining any two elements in  $G$  via  $\cdot$  yields an element of  $G$  again
- **Associativity:**  $(g \cdot h) \cdot i = g \cdot (h \cdot i)$ ,  $\forall g, h, i \in G$
- **Existence of an identity element:**  $e \in G$ , such that

$$g \cdot e = e \cdot g = e, \quad \forall g \in G$$

- **Existence of an inverse element:**  $g^{-1} \in G$ , such that

$$g \cdot g^{-1} = g^{-1} \cdot g = e, \quad \forall g \in G$$

In many formulations, the binary operation  $(\cdot)$  of the tuple  $(G, \cdot)$  is often left implicit, especially when it is clear from context.

For the special orthogonal group  $SO(3)$ , we define the binary operation as matrix multiplication. Specifically, combining two rotation matrices via matrix-matrix multiplication always yields another rotation matrix, meaning the group is closed under this operation. Associativity is inherently a property of matrix multiplication, so it holds automatically. The identity element is the identity matrix, denoted  $II$ , and the inverse of a rotation matrix is its transpose, which also belongs to  $SO(3)$ . Thus, all group properties are satisfied by  $SO(3)$ .

Similarly, for the special Euclidean group  $SE(3)$ , we use homogeneous coordinates to combine rotations and translations. An element of  $SE(3)$  can be represented by a transformation matrix of the form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^\top & 1 \end{pmatrix}. \quad (3)$$

The group operation for  $SE(3)$  is matrix multiplication of these homogeneous matrices, which combines both the rotational and translational components. Just as with  $SO(3)$ , matrix multiplication ensures closure and associativity. The identity element is the matrix

$$\mathbf{I}_{SE(3)} = \begin{pmatrix} \mathbf{I} & \mathbf{0}_3 \\ \mathbf{0}_3^\top & 1 \end{pmatrix} \quad (4)$$

and the inverse of any transformation matrix  $\mathbf{T}$  can be computed, ensuring the existence of inverse elements.

With this framework, we can consistently combine elements of  $SO(3)$  and  $SE(3)$  and reliably predict the properties of the resulting elements. These group structures provide a powerful foundation for describing and understanding rotations and rigid body motions in three-dimensional space.

### 2.1.1 Group Action and Representation

Mathematically, a group is considered as an abstract algebraic structure defined solely by its elements and the operation that combines them. However, in both theory and application it is of great interest to consider the effect that a group element  $g \in G$  has on other elements within the group or on sets of other mathematical objects. This is done via so-called group actions.

A particularly important case arises when the set on which the group acts on is a vector space, such as  $\mathbb{R}^n$ . In this case, the action of the group is referred to as a group representation. Formally, a representation of a group  $G$  on a vector space  $V$  is a map

$$\rho: G \rightarrow \text{GL}(V), \quad (5)$$

where  $\text{GL}(V)$  is the general linear group of all invertible linear transformations on  $V$ . This map is required to be a group homomorphism, meaning for any  $g, h \in G$  it preserves the group structure

$$\rho(g \cdot h) = \rho(g) \circ \rho(h). \quad (6)$$

For our examples, these functions are the linear maps encoded by the matrices (1) and (2), who act on  $\mathbb{R}^3$  and  $\mathbb{R}^4$  via standard matrix-vector multiplication. They preserve the group structure as required in (6), since we can combine both types of transformations in a consistent way via matrix-matrix multiplication to obtain the combined rotation.

### 2.1.2 Left Action and Group Conjugacy

As mentioned above, group elements can also act on each other. The most straightforward example of this is the left action, which combines two group elements via the binary group operation. Specifically, for  $g \in G$ , the left action on  $h \in G$  is given by:

$$L_{g*}: G \rightarrow G, \quad h \mapsto g \cdot h. \quad (7)$$

This simply means that  $g$  acts on  $h$  by applying the group operation, producing another element of the group. This form of action is fundamental in many areas of group theory and is easy to interpret as the composition of transformations.

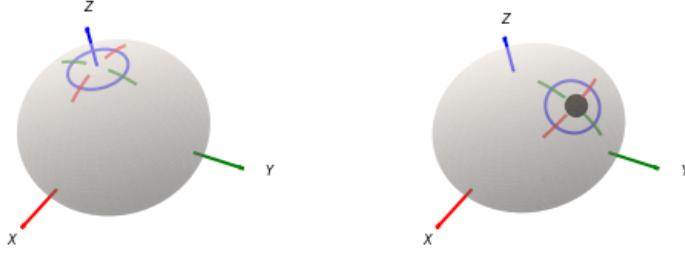


Figure 1: The matrix  $\mathbf{G}$  that is used to change perspective realizes a rotation by  $-45^\circ$  around the  $x$ -axis. We consider the effect of a rotation on the point  $(0, 0, 1)^\top$  and mark the reached point.

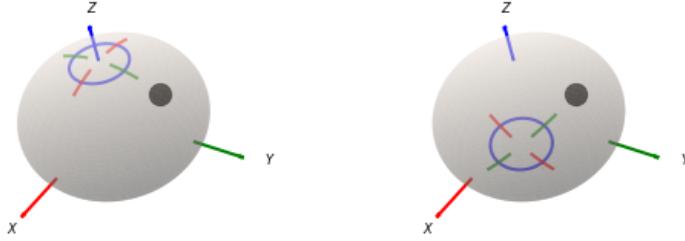


Figure 2: We mark the point computed above before applying any rotation. The conjugate  $\mathbf{R}'$  yields a new group element, which is a  $90^\circ$  around the new  $z$ -axis. It thereby realizes the same operation, but after changing our perspective to that of the point.

Another important way to do so is via group conjugacy. An element  $g \in G$  is conjugate to  $x \in G$ , if there exists a  $g \in G$ , such that

$$h = gxg^{-1}. \quad (8)$$

There is a very intuitive interpretation of this action as a change of perspective, which will become essential later on. In this case,  $h$  expresses the element  $x$  from the perspective of  $g$ .

In order to gain visual intuition for  $\text{SO}(3)$ , we consider e.g. a  $90^\circ$  rotation around the  $z$ -axis encoded via a rotation matrix  $\mathbf{R}$ . Based on another matrix  $\mathbf{G}$  encoding a  $-45^\circ$  rotation around the  $x$ -axis, we can compute

$$\mathbf{R}' = \mathbf{GRG}^{-1}, \quad (9)$$

which is a rotation of  $90^\circ$  around a new axis, obtained by rotating the  $z$ -axis  $45^\circ$  around the  $x$ -axis. More intuitively, we obtained a new rotation  $\mathbf{R}'$ , which realizes the same  $90^\circ$  rotation, just "seen" from the perspective / the new reference frame of  $G$ . Please refer to Figure ?? for an illustration. Adding translations, we can also use the group conjugacy with  $\text{SE}(3)$  to describe a rigid body transformation from a different reference frame. Albeit based on the same idea, gaining visual intuition for this case is slightly more involved, since rotational and translational part interact with

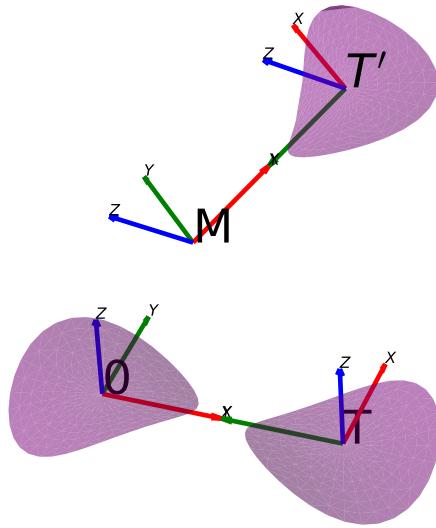


Figure 3: Visualization of the change-of-perspective intuition for  $SE(3)$ . The conjugate action of  $\mathbf{M}$  on  $\mathbf{T}$  results in a new transformation  $\mathbf{T}'$  that -if applied to  $\mathbf{M}$ - realizes the same relative transformation that  $\mathbf{T}$  does on the world frame.

each other and we have to consider a slightly different interpretation in terms of changing our perspective.

As rigid body transformations are usually defined relative to some reference frame, we have to apply the result of a conjugate action to the frame that performs the action. Please refer to Figure 3 for clarification<sup>1</sup>. Another interesting example that is worth mentioning is the general linear group  $GL(n)$  of invertible matrices, for which the definition we made above directly corresponds to that of matrix similarity from linear algebra. The change of perspective is also quite intuitive in this case: we consider the same linear transformation, only expressed in a different basis. Note that the inverses in both definitions are switched: Two matrices  $\mathbf{A}, \mathbf{B} \in GL(n)$  are similar if there exists a  $\mathbf{P} \in GL(n)$ , such that

$$\mathbf{B} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}. \quad (10)$$

This is a mere consequence of different conventions in group theory and linear algebra.

---

<sup>1</sup>Actually, we could use the same procedure for  $SO(3)$ . This would result in the coloured "crosshair" in Figure 2 being centered around the point of perspective, while still indicating the same rotation.

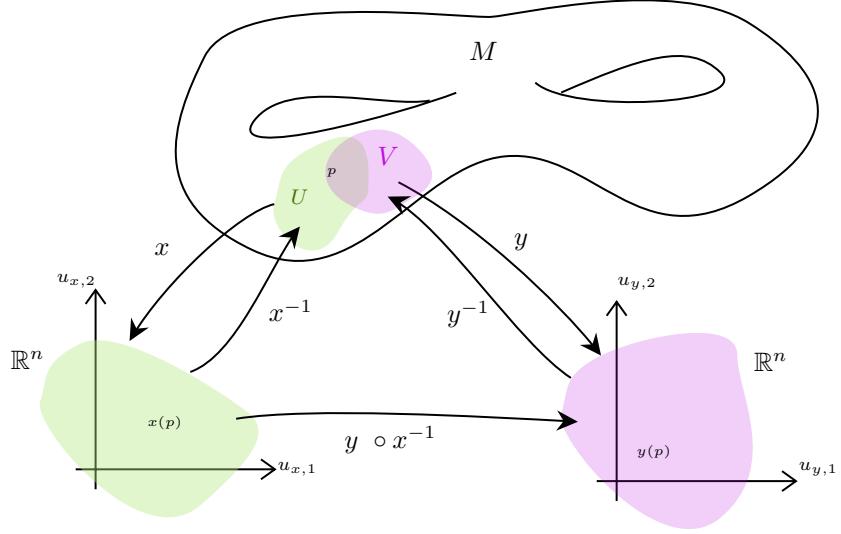


Figure 4: Manifold  $M$  with charts  $x$  and  $y$ , mapping open subsets  $U, V$  onto open subsets of  $\mathbb{R}^n$ . In the general setting, the local basis does not have to be the standard basis and usually varies between charts. Remember that while being an open subset of  $\mathbb{R}^n$ , the metric is given by  $g$  and not the standard inner product.

## 2.2 Riemannian Manifolds

The fundamental idea of an  $n$ -dimensional manifold  $M$  is that of a space that, while potentially complex in its global structure, locally resembles Euclidean  $\mathbb{R}^n$ . More precisely, each point on the manifold has a neighborhood that is homeomorphic to an open subset of  $\mathbb{R}^n$ , meaning it shares the same local topological structure as Euclidean space. This allows us to work with manifolds as if they were Euclidean spaces in a local sense, even though their global properties may be much more complicated.

Much like sets, manifolds can be equipped with additional structures that define various special types of manifolds. One such structure, which is of particular interest to us, is the structure of a Riemannian manifold. A Riemannian manifold is differentiable and endowed with a so-called Riemannian metric  $g$ . This metric allows us to generalize familiar geometric notions from Euclidean space, such as lengths of curves, angles between vectors, and distances between points.

The local homeomorphisms that relate the manifold to Euclidean space are called local coordinates or charts. A chart  $x$  is a smooth map that assigns each point on the manifold to a point in an open subset of  $\mathbb{R}^n$ , consider e.g. Figure 4. In a general setting, no single chart suffices to describe the entire manifold. Therefore, a collection of charts, called an atlas, is used to cover the manifold. Importantly, these charts must be compatible with one another, meaning that where two charts overlap, the transition map between them is smooth. This smooth compatibility ensures that we can transition between different local descriptions of the manifold seamlessly, preserving the differentiable structure.

Linearizing the local coordinates around a point  $p \in M$  yields a smoothly-varying set of vectors that span a vector space at each point on the manifold, known as the tangent space  $T_p M$ . This space provides a linear approximation of the manifold in the neighborhood of  $p$ , allowing us to perform local operations using the tools of linear algebra, even though the manifold itself might have a more complex global structure. Formally, a tangent vector at a point  $p \in M$  is typically written in terms

of a local basis  $\{e_i\}$  via  $v = \sum_i \xi^i e_i$ , where  $\xi^i$  are local components of the vector in the given chart. The metric  $g$  can be thought of as a point-wise inner product on these local vector spaces. Mathematically, it is a point-wise symmetric bilinear form that is non-degenerate and positive definite

$$g_p: T_p M \times T_p M \rightarrow \mathbb{R}_{>0}, \quad (11)$$

giving us the tools to measure geometric quantities in a manner that extends beyond flat spaces. Given two tangent vectors  $v = \sum_i \xi^i e_i$  and  $w = \sum_j \eta^j e_j$  in local coordinates, we can express the metric via

$$g(X, Y) = \sum_{ij} \xi^i \eta^j g_{ij}. \quad (12)$$

In the flat euclidean case of  $\mathbb{R}^n$ , this would simply be the regular inner product ( $g_{ij} = 1$ ), each tangent space would be a local copy of  $\mathbb{R}^n$  and tangent vectors are regular euclidean vectors. Using the generalization above, more complex geometries can be encoded. A Riemannian manifold is referred to by the tuple  $(M, g)$ , although multiple metrics can encode the same geometric structure.

### 2.2.1 Lie Derivative and Pushforward

To perform more advanced mathematical operations on a manifold, we need to express them in terms of the tangent spaces, since these allow us to handle differential structures locally. Two key tools for this are the Lie derivative, which measures how a function or tensor field changes along a vector field, and the pushforward, which provides a way to map tangent vectors between manifolds or within the same manifold via smooth maps.

To measure rates of change of e.g. a scalar function  $f: M \rightarrow \mathbb{R}$  along a flow of a vector field on the manifold, we can use the so-called Lie derivative. Given a direction  $v = \sum_i \xi^i e_i$  in the tangent space  $T_p M$ , we define

$$\partial_v f(p) = \sum_{i=1}^n \xi^i \partial_i (f \circ x^{-1}) \Big|_{x(p)}, \quad (13)$$

which incorporates information about the manifold structure via the inverse of the coordinate mapping  $x^{-1}$  and the partial derivatives with respect to the local coordinates  $\partial_i$ . This expression captures how the function changes as we move in the direction of the vector  $v$  in the coordinate chart.

Similarly, we can define a linearization of a map  $\phi$  between points of the same manifold or between two different ones, resulting in mappings between tangent spaces. These are formalized via the pushforward operator  $\phi_*(v)$ . Here, we consider a map  $\phi: M \rightarrow M$  and assume that both points can be described by the same chart. Given a tangent vector  $v \in T_p M$  for  $p \in M$ , this map then sends  $v$  to a tangent vector  $\phi_*(v) \in T_{\phi(p)} M$  in the tangent space at the image point  $\phi(p)$ . Using again local components  $\xi^i$  provided by a chart  $x$ , we compute the image vector  $\eta$  at  $p$  via

$$\eta := \sum_i \xi^i \partial_i (x \circ \phi \circ x^{-1}) \Big|_{x(p)}. \quad (14)$$

This essentially allows us to transport tangent vectors along a map by translating the abstract transformation on the manifold into a linear one on local vector spaces.

### 2.2.2 Geodesics and Exponential Map

When performing calculations in tangent spaces, we are working based on local first order approximations. In order to map these back onto the manifold, we need a generalization of straight lines

or shortest paths that respects the non-trivial geometry. This is achieved by an acceleration-free motion using the tangent vector in question as a starting velocity, a so-called geodesic curve. This is not straight-forward, since each point has a potentially different basis. However, since these are required to be smoothly varying for Riemannian manifolds, we can express this change in closed form via the so-called Christoffel-Symbols  $\Gamma_{ij}^k$ . Each symbol captures the  $k$ -th component of the change of the basis vector  $e_j$  when moving infinitesimally into the direction of  $e_i$ . A geodesic curve  $c(t)$  with  $t \in [0, 1]$  on the manifold with tangent vector  $\gamma'(t)$  fulfills the non-linear system of ODE's given by

$$\gamma''^k(t) + \sum_{ij} \gamma'^i(t) \gamma'^j(t) \Gamma_{ij}^k \circ c(t) = 0, \quad k = 1, \dots, n. \quad (15)$$

Intuitively, this equation ensures that the acceleration vector of the path is adjusted to cancel out the effect of the non-trivial geometry. However, the resulting system of equations cannot be solved in closed form for every metric.

For notational convenience, we can define a mapping between an initial tangent vector  $v = \gamma'(0) \in T_p M$  and the end-point of the associated geodesic, the so-called exponential map

$$\exp_p(v) := c_{p,v}(1), \quad (16)$$

which is computed by solving the system of ODE's up until time 1. Consider Figure 5 for an illustration.

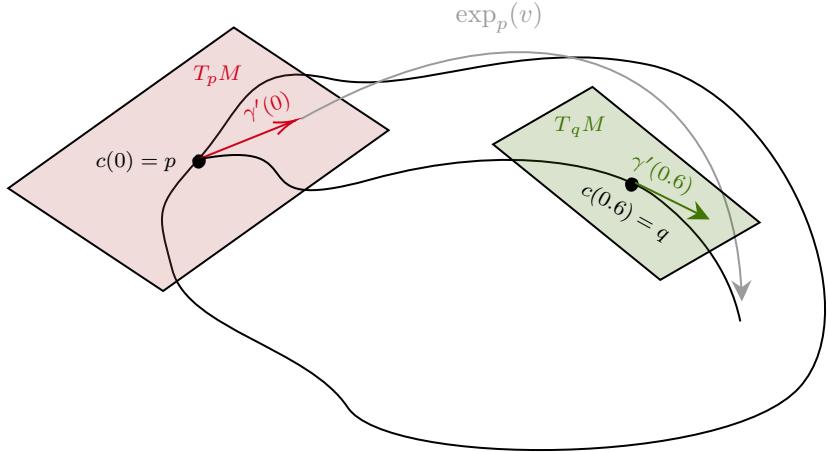


Figure 5: Conceptual sketch of the exponential map  $\exp_p(v)$  of a Riemannian manifold. Starting from an initial point  $c(0) = p$ , we generate a curve on the manifold by following the tangent vector  $\gamma'(t)$  in every point. If the tangent vectors fulfill (15), the motion is acceleration free and generalizes a straight line in euclidean space to the manifold.

### 2.3 Diffusion Models

As a first application of Lie Theory, we want to consider adapting a diffusion model to the  $\text{SO}(3)$  manifold. As a short introduction, we follow the excellent blog post [11] and add intuition via visualizations and additional explanations.

### 2.3.1 Concept

Diffusion models are generative models inspired by the physical process of diffusion in thermodynamics. This describes how e.g. heat spreads from areas of high temperature to its surroundings over time, eventually reaching an equilibrium state.

The main idea is to model this process using a Markov chain that repeatedly adds noise to a given data distribution and then learn to invert it. The hope is that after training, we are able to generate new samples of the data distribution by simulating the reverse process starting from an arbitrary (mostly Gaussian) distribution.

We assume data that is distributed with some "real" data distribution  $\mathbf{x} \sim q(\mathbf{x})$ . In order to provide visual intuition, we train the 3D toy dataset shown in Figure 6 based on the introduced concepts. In order to obtain closed form updates, we define a process based on Gaussian distributions to increasingly corrupt the data artificially and in a controlled manner.

In the most basic formulation, we begin by defining a pre-determined noise or variance schedule  $\beta = (\beta_1, \dots, \beta_t)$  with  $\beta_i \in (0, 1)$ . Based on the usual Markovian assumption that the immediate past completely determines the present, we can express the noising via the so-called forward process

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) \sim \mathcal{N}(\sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}). \quad (17)$$

With this, any trajectory probability can be computed via

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{i=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}). \quad (18)$$

Since we exclusively use Gaussians, we can use the reparametrization trick to express any intermediate forward step in closed form. For an arbitrary update, we can write

$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1}. \quad (19)$$

For brevity, we employ the notation

$$\alpha_t = 1 - \beta_t \quad \bar{\alpha}_t = \prod_{i=1}^t \alpha_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (20)$$

Since  $1 - \beta_t < 1$ , this can be understood as repeatedly scaling down the vector representing the current iterate and then adding back to it a small noise vector. Recursively inserting this update, we obtain

$$\begin{aligned} \mathbf{x}_t &= \sqrt{\alpha_t} \mathbf{x}_{t-1} + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1} \\ &= \sqrt{\alpha_t} (\sqrt{\alpha_{t-1}} \mathbf{x}_{t-2} + \sqrt{1 - \alpha_{t-1}} \boldsymbol{\epsilon}_{t-2}) + \sqrt{1 - \alpha_t} \boldsymbol{\epsilon}_{t-1} \\ &= \dots \\ &= \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon} \end{aligned}$$

Altogether, we obtain the closed form, single-shot sampling procedure defined by

$$\begin{aligned} q(\mathbf{x}_t | \mathbf{x}_0) &\sim \mathcal{N}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \\ \mathbf{x}_t &= \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \end{aligned} \quad (21)$$

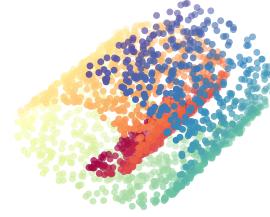


Figure 6: Swiss roll toy dataset.

Usually, we choose larger variances the noisier the data gets  $\beta_1 < \beta_2 < \dots < \beta_T$ , which results in  $\bar{\alpha}_1 > \dots > \bar{\alpha}_T$ . Therefore, as  $\bar{\alpha}$  gets smaller for later timesteps (leading to  $\sqrt{\bar{\alpha}_t} \rightarrow 0$  and  $\sqrt{1 - \bar{\alpha}_t} \rightarrow 1$ ), the distribution of the samples converges to a standard Gaussian distribution, as can be retraced in Figure 17. Our objective is to learn the reversed process by using a neural network



Figure 7: Exemplary forward diffusion process for the swiss role example. We show the datapoints at timesteps 10, 25, 50, 100, 200 out of 500. Notice how the mixing of colors increases between the fourth and fifth image, bringing the data distribution closer and closer to a unit Gaussian.

with parameters  $\theta$  to encode

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \sim \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}(\mathbf{x}_t, t)), \quad (22)$$

where based on the noise schedule  $\boldsymbol{\Sigma}(\mathbf{x}_t, t)$  can be derived in closed form.

### 2.3.2 Training

The formal objective is a maximum likelihood approach, where we basically demand the model to produce our initial data with high probability. For this, we have to minimize the negative log-likelihood

$$L = -\log \left( \int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T} \right), \quad (23)$$

which in this form is intractable as it would require marginalization over space and time dimensions. As it is done in other generative frameworks, we can use the Evidence Lower Bound to obtain a tractable approximative objective

$$\begin{aligned} -\log p_\theta(\mathbf{x}_0) &\leq -\log(p_\theta)(\mathbf{x}_0) + D_{KL}(q(\mathbf{x}_{1:T}|\mathbf{x}_0) || p_\theta(\mathbf{x}_{1:T}|\mathbf{x}_0)) \\ &= -\log(p_\theta)(\mathbf{x}_0) + \mathbb{E}_{\mathbf{x}_{1:T} \sim q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T}) p_\theta(\mathbf{x}_0)} \right] \\ &= \mathbb{E}_{\mathbf{x}_{1:T} \sim q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} \right]. \end{aligned} \quad (24)$$

Intuitively, we can think of this objective this way: The negative log-likelihood of a datapoint  $\mathbf{x}_0$  is high, if even for unlikely "noising paths"  $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$  we are able to generate a de-noising path with a high probability. Mathematically, this yields very small ratios in the expectation, of which the negative logarithm is a high number.

Inversely, if even for highly likely noising-paths (e.g. only very small noise at every step), we are not able to create a high-probability path, the output of  $\mathbf{x}_0$  is unlikely.

Based on reformulations using Bayes Theorem, the Markov property and logarithm rules, this expression can be rewritten as

$$\mathbb{E}_q \left[ -\log \left( \frac{p(\mathbf{x}_T)}{q(\mathbf{x}_T|\mathbf{x}_0)} \right) - \sum_{i=2}^T \log \left( \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right) - \log(p_\theta(\mathbf{x}_0|\mathbf{x}_1)) \right], \quad (25)$$

where the term in red can be derived in closed form

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I}) \quad (26)$$

$$\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_t}{1 - \bar{\alpha}_t} \mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t \quad (27)$$

$$\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t, \quad (28)$$

which enables us to learn a parametrized mean  $\tilde{\boldsymbol{\mu}}_\theta(\mathbf{x}_t, \mathbf{x}_0)$ . In practice, we go one step further by using interesting reformulation of Markovian dynamical systems that considers the trajectory of the noise instead of the states, e.g. derived in [4]. For simple settings such as ours, this can easily be applied by simply rearranging

$$\mathbf{x}_0 = \frac{1}{\bar{\alpha}_t} (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}) \quad \rightarrow \quad \tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_t \right) \quad (29)$$

Since  $\mathbf{x}_t$  is available at training time based on our one-shot sampling, we can now instead use the neural network to predict the noise vector

$$\tilde{\boldsymbol{\mu}}_\theta(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) \quad (30)$$

Inserting this idea back into the loss term of the means yields

$$L = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}} \left[ \frac{(1 - \alpha_t)^2}{2\alpha_t(1 - \bar{\alpha}_t) \|\Sigma_\theta\|_2^2} \|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_t, t)\|^2 \right] \quad (31)$$

Empirical studies showed that leaving out the weighting term in this objective can actually help to stabilize the training without affecting the results significantly. We thereby reduced the maximum likelihood objective to a mean squared error of the noise vectors. After training, we can then sample from  $\mathcal{N}(0, \mathbf{I})$  and iterate

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left( \mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \tilde{\beta}_t \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \quad (32)$$

to generate new samples of  $q(\mathbf{x})$ . For the swiss role toy dataset in Figure 6, the result is shown in Figure 17.

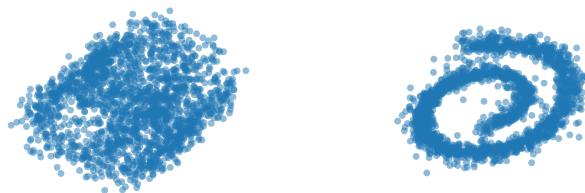


Figure 8: Samples generated from a multivariate unit Gaussian using the trained diffusion model shown from two different perspectives.

### 3 Lie Theory for SO(3) and SE(3)

As introduced in Section 2.1, we encounter two important groups:

- SO(3), the group of rotations around the origin in  $\mathbb{R}^3$ , defined via

$$\text{SO}(3) := \{\mathbf{R} \in \mathbb{R}^{3 \times 3} : \mathbf{R}\mathbf{R}^T = \mathbf{I}, \det(\mathbf{R}) = 1\} \subset \text{O}(3), \quad (33)$$

which excluded reflections as improper rotations.

- SE(3), the group of rigid body transformations, which combines SO(3) with translations

$$\text{SE}(3) = \{g = (\mathbf{R}, \mathbf{t}) | \mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3\}. \quad (34)$$

These groups not only fulfill the algebraic group axioms (closure, associativity, identity, and inverses) but also exhibit a continuous structure. The constraints that define SO(3) and SE(3) (such as orthogonality and determinant conditions) naturally endow them with the structure of smooth manifolds, giving rise to the concept of Lie groups. Our introduction mainly uses [9] and [8] as reference.

A Lie group  $\mathcal{G}$  is a mathematical object that is simultaneously a group and a smooth manifold. The manifold structure allows for the smooth differentiation of functions, while the group structure enables the algebraic operations of multiplication and inversion. This additional group structure can be used to greatly simplify some of the general definitions on Riemannian manifolds we gave above. The aim will again be to translate operations on the manifold into linear algebra formulations on the tangent space. The resulting theory is particularly rich and wide-ranging, which is why we will directly focus on the two groups we use in applications. To distinguish tangent vectors on Lie groups from those on general manifolds, we adopt the following notation:

- For any general Lie group, we use  $x^\wedge$
- For SO(3), the elements of the tangent space at any point  $\mathbf{R}$  are denoted by  $\omega^\wedge$ .
- For SE(3), the elements of the tangent space at any point  $\mathbf{T}$  are denoted by  $\xi^\wedge$ .

#### 3.1 Left-Invariant Vector Fields

Given any tangent vector  $x_h^\wedge \in T_h\mathcal{G}$ , the group structure enables us to extend this vector to a vector field on the entire group  $\mathcal{G}$ . This means that we can assign a tangent vector  $x_g^\wedge$  to every group element  $g \in \mathcal{G}$ , where the extension is smooth and compatible with the group operations.

To achieve this, we can leverage the pushforward operator  $L_{g*}$  associated with the left group action, which was introduced in Section 2.2.1. It allows us to map tangent vectors from  $T_h\mathcal{G}$  to tangent vectors at  $T_{gh}\mathcal{G}$  via

$$x_{gh}^\wedge) L_{g*}(x_h^\wedge). \quad (35)$$

In the case of our matrix Lie groups SO(3) and SE(3), it can be shown that the left action and its linearization (the pushforward) are represented simply by matrix multiplication.

The ability to extend a tangent vector to the whole group highlights a key feature of Lie groups. Unlike general manifolds, where transitioning between tangent spaces may require complicated coordinate transformations, Lie groups allow for a global, consistent transformation via their group operations.

## 3.2 The Exponential and Log Map of Lie Groups

Reconsidering the definition of the exponential map from the perspective of left-invariant vector fields significantly simplifies the expressions. By leveraging the group structure, we can express geodesics and their differential equations in a more tractable way. In Section 2.2.2, we defined geodesics as curves  $c(t)$  on a manifold driven by a tangent vector  $v = \gamma'$ . In the case of our matrix Lie groups, we know that the tangent vector can be extended to the entire group via the left-invariant vector fields using matrix multiplication.

For any Lie group, let  $x^\wedge \in T_e \mathcal{G}$  be a tangent vector at the identity element  $e$ . The curve  $c(t)$ , which represents a geodesic starting at the identity, satisfies the differential equation

$$c'(t) = x^\wedge = c(t) \cdot x^\wedge \quad (36)$$

This expression reflects the fact that at each point along the geodesic, the tangent vector is related to the group element itself by the left multiplication (hence the left-invariance of the vector field). Solving this differential equation is straightforward and yields the solution:

$$c(t) = \exp(c^\wedge t), \quad (37)$$

which is much simpler than the general expression in equation (15).

To generalize this notion to arbitrary group elements, we can use the left action again to transform any geodesic to an arbitrary starting point. However, working with the identity element is often simpler and more convenient, as the notation becomes more concise, and the identity element plays a central role in Lie group theory anyways.

With this, we can derive expressions to map our matrix groups back and forth between tangent space and the respective manifold. For any matrix  $\mathbf{A}$  of either group, the exponential map is the series

$$\mathbf{A}(x^\wedge) = \exp(x^\wedge) = \sum_{n=0}^{\infty} \frac{1}{n!} (x^\wedge)^n. \quad (38)$$

Inversely, the map from the manifold to tangent vector at the identity is given by the so-called log-map

$$\log(\mathbf{A}) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} (\mathbf{A} - \mathbf{1})^n. \quad (39)$$

## 3.3 Tangent Spaces and Lie Algebra

The core idea of using Lie Theory for rotations and rigid body transformations is to use the underlying Lie Theory to shift all required computations on the complex group manifold to the tangent spaces instead. This shift allows us to operate using linear algebra tools for computation, while still respecting the underlying geometric structure.

While this makes the algebraic operations more complicated, the structure of  $\text{SO}(3)$  and  $\text{SE}(3)$  allows us to not only obtain closed-form expressions for many of them, they also allow us to perform them all in a single tangent- and thereby vector space. A Lie group theoretically allows us to choose an arbitrary element to construct this common space, but historically the identity element  $e$  is used to define the so-called Lie Algebra

$$\mathfrak{g} = T_e \mathcal{G}.$$

With this and the knowledge of the preceding sections, we can now determine what the tangent vectors and operations on the Lie Algebra of our matrix groups actually look like.

### 3.3.1 SO(3)

For elements  $\mathbf{R}$  of the rotation group  $\text{SO}(3)$ , the associated manifold structure results from the constraint  $\mathbf{R}^\top \mathbf{R} = \mathbf{1}$ . Taking the time derivative yields

$$\mathbf{R}^\top \dot{\mathbf{R}} = -(\mathbf{R}^\top \dot{\mathbf{R}})^\top,$$

which means this combined expression is a skew-symmetric matrix. To get to the Lie Algebra, we look at the identity element of rotations (identity matrix) and realize that then the expression says

$$\mathbf{1}^\top \dot{\mathbf{R}} = (\mathbf{1}^\top \dot{\mathbf{R}})^\top,$$

showing that the set of these skew-symmetric matrices

$$\mathfrak{so}(3) = \left\{ \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \right\}, \quad (40)$$

form the tangent space at the identity and therefore our Lie Algebra  $\mathfrak{so}(3)$ . A commonly used basis is

$$E_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad E_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (41)$$

which we denote infinitesimal generators, since they essentially encode infinitesimal rotations. Because these matrices are the basis of a vector space structure, we can identify them with euclidean vectors

$$\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^\top \in \mathbb{R}^3. \quad (42)$$

This directly corresponds to the well-known axis-angle representation with

$$\boldsymbol{\omega} = \theta \hat{\boldsymbol{\omega}}, \quad \theta = \|\boldsymbol{\omega}\|. \quad (43)$$

To switch between both representations of tangent vectors, we use the notation

$$(\boldsymbol{\omega})^\wedge = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}^\vee = \boldsymbol{\omega}. \quad (44)$$

For  $\text{SO}(3)$  with tangent vectors  $\boldsymbol{\omega}^\wedge$ , the exponential map can be obtained in closed form based on the infinite series in Equation (38). This yields the well-known Rodriguez formula

$$\mathbf{R}(\boldsymbol{\omega}) = \exp(\boldsymbol{\omega}^\wedge) = \sum_{n=0}^{\infty} \frac{1}{n!} (\boldsymbol{\omega}^\wedge)^n \quad (45)$$

$$= \mathbf{1} + \hat{\boldsymbol{\omega}}^\wedge \frac{\sin(\theta)}{\theta} + (\hat{\boldsymbol{\omega}}^\wedge)^2 \frac{(1 - \cos(\theta))}{\theta^2}. \quad (46)$$

Inversely, the log-map can be derived to yield

$$\theta = \begin{cases} \arccos\left(\frac{\text{tr}(\mathbf{R}) - 1}{2}\right), & \mathbf{R} \neq \mathbf{I}, \\ 2\pi k & \text{else,} \end{cases} \quad \hat{\boldsymbol{\omega}} = \begin{cases} \frac{\theta}{2\sin(\theta)}(\mathbf{R} - \mathbf{R}^\top) & \mathbf{R} \neq \mathbf{I}, \\ \mathbf{0} & \text{else.} \end{cases} \quad (47)$$

Not that by the formula above, the exponential map is surjective, but not injective. In some situations, it can be useful to introduce the capitalized version both maps for notational convenience

$$\mathbf{R} = \text{Exp}(\boldsymbol{\omega}) = \exp((\theta \hat{\boldsymbol{\omega}})^\wedge) \quad (48)$$

$$\boldsymbol{\omega}^\wedge = \text{Log}(\mathbf{R}) = (\log(\mathbf{R}))^\vee \quad (49)$$

This simply takes the vector representation of the tangent vector instead of the matrix representation we used above. The notation is clarified in the following chart

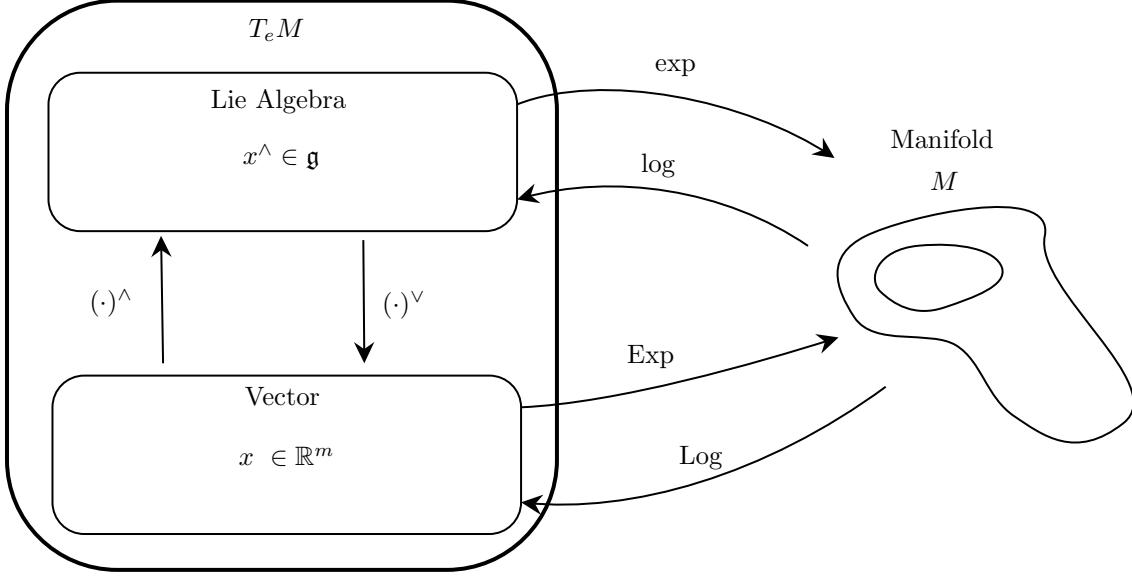


Figure 9: Relationship between different representations of the Lie Algebra inspired by [8].

### 3.3.2 SE(3)

To obtain similar notions for SE(3), we can follow the same arguments as above to obtain the Lie Algebra

$$\mathfrak{se}(3) = \left\{ \begin{pmatrix} (\boldsymbol{\phi})^\wedge & \boldsymbol{\rho} \\ \mathbf{0}_3^\top & 0 \end{pmatrix} : \boldsymbol{\rho}, \boldsymbol{\phi} \in \mathbb{R}^3 \right\} \quad (50)$$

with  $(\boldsymbol{\phi})^\wedge \in \text{SO}(3)$  and parametrization

$$(\boldsymbol{\xi})^\wedge = \begin{pmatrix} 0 & -\omega_3 & \omega_2 & \rho_1 \\ \omega_3 & 0 & -\omega_1 & \rho_2 \\ -\omega_2 & \omega_1 & 0 & \rho_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (51)$$

Similarly to the SO(3) case, we can define a basis consisting of 6 matrices by setting one of the parameters to one and the remaining to zero. These can again be identified with euclidean vectors, this time in  $\mathbb{R}^6$  because of the additional translational parameters

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\rho} \\ \boldsymbol{\phi} \end{pmatrix}. \quad (52)$$

The exponential map  $\exp: \mathfrak{se}(3) \rightarrow \text{SE}(3)$  can also be obtained via an infinite series and expressed via homogeneous coordinates yields

$$\mathbf{T} = \exp(\boldsymbol{\xi}^\wedge) = \begin{pmatrix} \text{exp}_{\text{SO}(3)}(\boldsymbol{\phi}^\wedge) & \mathbf{V}(\boldsymbol{\phi}^\wedge)\boldsymbol{\rho} \\ \mathbf{0} & 1 \end{pmatrix}, \quad (53)$$

with

$$\mathbf{V}(\boldsymbol{\phi}^\wedge) = \mathbf{I} + \boldsymbol{\xi}^\wedge + \left(\frac{1 - \cos(\theta)}{\theta^2}\right)(\boldsymbol{\xi}^\wedge) + \left(\frac{\theta - \sin(\theta)}{\theta^3}\right)(\boldsymbol{\xi}^\wedge)^2. \quad (54)$$

Inversely, we can compute the log-map by first using the log map of  $\text{SO}(3)$  to compute  $\boldsymbol{\phi}^\wedge$  from  $\mathbf{R}$ . This can then be used to compute the translational part via, yielding

$$\boldsymbol{\xi}(\mathbf{T}) = \begin{pmatrix} \mathbf{V}^{-1}(\boldsymbol{\phi}^\wedge)\mathbf{t} \\ \text{log}_{\text{SO}(3)}(\mathbf{R}) \end{pmatrix}. \quad (55)$$

It is important to note that  $\text{SE}(3)$  performs translation and rotation as a continuum, it thereby differs from  $\text{T}(3) \times \text{SO}(3)$  in how it deals with the translational part and its interaction with the shared  $\text{SO}(3)$  part.

### 3.4 Adjoint Representation and Commutator

So far, we can translate between vectors of the tangent space of the identity, the so-called Lie Algebra and points on the actual manifold via the exponential and log map. For being able to perform all computations within the Lie Algebra vector space as promised above, we need to account for operations that take place at arbitrary points points on the manifold.

This brings us back to the concept of group conjugacy and perspective change introduced in Section 2.1.2. Just as group conjugacy can be seen as expressing an element from a different "perspective," we can use the group structure to change the tangent space from the identity element to any other element on the manifold.

We can obtain a linear operator called the Adjoint  $\text{Ad}$  by linearizing the conjugate group action at the identity element via

$$\text{Ad}_g: T_e \mathcal{G} \rightarrow T_e \mathcal{G}, \quad \text{Ad}_g(x^\wedge) = \frac{d}{dt}(g \exp(tx^\wedge)g^{-1}) \Big|_{t=0}. \quad (56)$$

This new action can be used to change the perspective from which we see an infinitesimal rotation to another point on the manifold. To explicitly compute this, we can use the fact that any element  $h$  close to the identity can be expanded as

$$h = e + \epsilon x^\wedge + \mathcal{O}(\epsilon^2), \quad (57)$$

since, as an element of the Lie algebra,  $x^\wedge$  is defined as the linear approximation of  $h$  in  $e$ . Writing out the conjugate action of any  $g \in \mathcal{G}$  based on this and using the fact that  $geg^{-1} = e$ , we obtain

$$ghg^{-1} = e + \epsilon(gx^\wedge g^{-1}) + \mathcal{O}(\epsilon^2). \quad (58)$$

Since this has the exact same form as the equation before, the term  $gx^\wedge g^{-1}$  is the linearization of the conjugate action we are seeking in (56). The adjoint is a linear operator, which is why we can represent it as a matrix operating on the vector-representations of our Lie groups.

Lastly, we have to consider how infinitesimal elements interact with each other by inspecting another derivative of the adjoint representation. For a  $y^\wedge \in \mathfrak{g}$ , we consider the conjugate action by another infinitesimal element  $x^\wedge \in \mathfrak{g}$  via

$$\begin{aligned} \frac{d}{dt} \left( \exp(ty^\wedge) x^\wedge \exp(-ty^\wedge) \right) \Big|_{t=0} &= \left( \frac{d}{dt} \exp(ty^\wedge) \right) \Big|_{t=0} x^\wedge \left( \exp(-ty^\wedge) \right) \Big|_{t=0} \\ &+ \left( \exp(ty^\wedge) \right) \Big|_{t=0} x^\wedge \left( \frac{d}{dt} \exp(-ty^\wedge) \right) \Big|_{t=0}. \end{aligned}$$

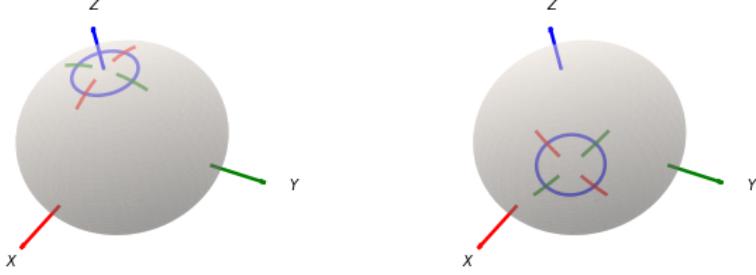


Figure 10: Rotation used in Figure 2, this time realized using the Lie Algebra. We first compute the vector representation of the rotation matrix  $\mathbf{R}$  via  $\text{Log}(\mathbf{R})$ . Then, we can simply apply the rotation matrix  $\mathbf{G}$  to that vector to realize the adjoint action  $\text{Ad}_{\mathbf{R}}$  and map the result back to the manifold using  $\text{Exp}$ .

The exponential map of a zero tangent vector is simply the identity, while the derivative of the exponential map at the starting point brings us back to the respective element of the tangent space. This yields

$$\frac{d}{dt} \left( \exp(ty^\wedge) x^\wedge \exp(-ty^\wedge) \right) \Big|_{t=0} = y^\wedge x^\wedge - x^\wedge y^\wedge, \quad (59)$$

which is the adjoint representation  $\text{ad}_{y^\wedge}(x^\wedge)$  of the Lie Algebra. The operator that realizes this linear map is the infamous commutator or Lie bracket, denoted by  $[y^\wedge, x^\wedge]$ . We say that the Lie Algebra acts on itself via commutators.

### 3.4.1 $\text{SO}(3)$

As with the actions before, the adjoint representation of  $\text{SO}(3)$  is simply given by the matrix-matrix product

$$\text{Ad}_{\mathbf{R}}(\omega^\wedge) = \mathbf{R} \omega^\wedge \mathbf{R}^{-1}, \quad \mathbf{R} \in \text{SO}(3), \omega^\wedge \in \mathfrak{so}(3). \quad (60)$$

To obtain the respective matrix operating on the vector representation, we can write out the adjoint representation and recollect the terms as a matrix-vector product, which yields

$$\text{Ad}_{\mathbf{R}} = \mathbf{R}, \quad (61)$$

we say that  $\text{SO}(3)$  is self-adjoint. To check whether this result agrees with our initial group conjugate example, we can compare the results of Figure 2 and 10, where we computed the same rotation using the adjoint. The commutator of  $\mathfrak{so}(3)$  is straight-forward and can be directly deduced from Equation (59), yielding

$$[\omega_1^\wedge, \omega_2^\wedge] = \omega_1^\wedge \omega_2^\wedge - \omega_2^\wedge \omega_1^\wedge = (\omega_1 \times \omega_2)^\vee \in \mathfrak{so}(3), \quad (62)$$

where  $\times$  is the regular cross-product. Albeit more abstract, we can extend our build-up intuition to this infinitesimal case. For example, we can consider an infinitesimal rotation around  $y$  given by  $E_y$

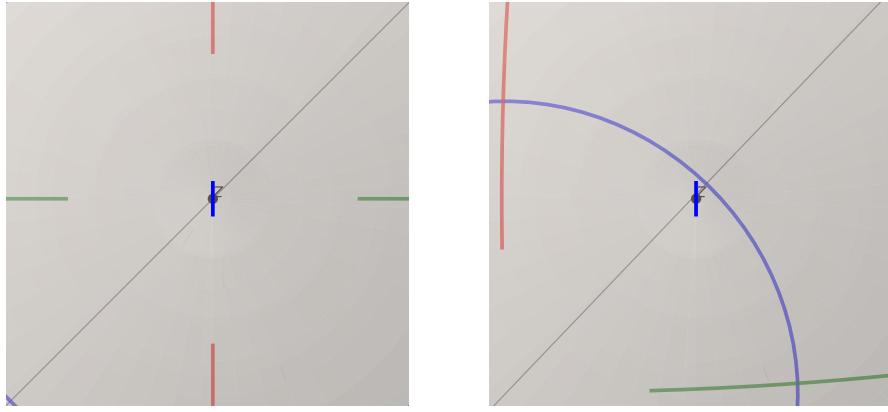


Figure 11: Zoomed-in view of a small two-step rotation around  $x$  followed  $y$ . By carefully observing the indicated diagonal, we can see that the  $z$ -axis is slightly shifted to the right. Thereby, from the perspective of the center of the circle, we performed a small positive rotation around our own  $z$ -axis.

seen from the perspective of  $E_x$ . Mathematically, this equates to an infinitesimal rotation around  $z$ :

$$[E_x, E_y] = E_x E_y - E_y E_x = E_z. \quad (63)$$

While this is a lot harder to grasp visually, we can consider Figure 11, where we see the result of two successive small rotations, one around  $x$  and one around  $y$ . As explained in the caption, the result includes a small rotation around  $z$  from the new perspective. Since both applied rotations should originally be infinitesimal, only this "tendency" to rotate around the  $z$  axis remains after taking the limit.

### 3.4.2 SE(3)

For rigid body transformations, we develop the adjoint based on  $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in \text{SE}(3)$  and  $\xi^\wedge = (\phi^\wedge, \rho) \in \mathfrak{se}(3)$  the same way as above, yielding

$$\text{Ad}_{\mathbf{T}}(\xi^\wedge) = (\mathbf{T}\xi^\wedge\mathbf{T}^{-1})^\vee = \begin{pmatrix} \mathbf{R} & \mathbf{t}^\wedge\mathbf{R} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \rho \\ \phi \end{pmatrix}. \quad (64)$$

Using the same reasoning as for  $\text{SO}(3)$ , the commutator of  $\mathfrak{se}(3)$  can be computed via

$$[\xi_1^\wedge, \xi_2^\wedge] = \left( \begin{pmatrix} \phi_1^\wedge & \rho_1^\wedge \\ \mathbf{0} & \phi_1^\wedge \end{pmatrix} \begin{pmatrix} \rho_2 \\ \phi_2 \end{pmatrix} \right)^\vee \in \mathfrak{se}(3) \quad (65)$$

## 3.5 Metrics on $\text{SO}(3)$

It remains to define a metric on  $\text{SO}(3)$  to connect this chapter back to our introduction about Riemannian manifolds in Section 2.2. A natural choice is provided by the Cartan-Killing form, a symmetric bilinear form defined on the Lie Algebra via

$$\langle x^\wedge, y^\wedge \rangle_{\text{Killing}} = \text{tr}(\text{ad}(x^\wedge) \circ \text{ad}(y^\wedge)) \stackrel{(1)}{=} -2\text{tr}(x^\wedge y^\wedge), \quad (66)$$

where (1) follows for compact Lie Groups such as  $\mathfrak{so}(3)$ . We know from Section 3.1 that any vector can be extended to a vector field on the manifold by left action, which is why the Killing form

extends to a metric on the whole manifold.

As a metric, the Cartan-Killing form tells us how close two rotations are and/or presents a way to measure some abstract notion of "length". Since these are abstract anyways, any scaling of the above is a reasonable choice for a metric, which is why you often see the definition

$$g_{SO(3)}(\omega_1^\wedge, \omega_2^\wedge) = \frac{1}{2} \text{tr}(\omega_1^\wedge (\omega_2^\wedge)^\top), \quad \omega_1^\wedge, \omega_2^\wedge \in \mathfrak{so}(3). \quad (67)$$

For  $\mathfrak{se}(3)$ , which is not compact because of the euclidean translation part, we can simply decompose the inner product to obtain

$$g_{SE(3)}(\xi_1^\wedge, \xi_2^\wedge) = -\text{tr}\left(\xi_1^\wedge \begin{pmatrix} \frac{1}{2}\mathbf{I} & \mathbf{0}_3 \\ \mathbf{0}_3^\top & 1 \end{pmatrix} (\xi_2^\wedge)^\top\right), \quad \xi_1^\wedge, \xi_2^\wedge \in \mathfrak{se}(3). \quad (68)$$

Mathematically, the trace is an obvious choice since it fulfills the properties of a metric (linear, symmetric, positive definite) automatically. For  $SO(3)$ , we could for example consider comparing the inner products of an infinitesimal rotation around  $x$  with others around  $z$ , a diagonal axis between  $x$  and  $z$  and  $x$  again. This results in

$$\begin{aligned} \frac{1}{2} \text{tr}(E_x, E_x) &= \frac{1}{2} \text{tr}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\right) = 1 \\ \frac{1}{2} \text{tr}(E_x, E_z) &= \frac{1}{2} \text{tr}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) = 0 \\ \frac{1}{2} \text{tr}(E_x, \frac{1}{\sqrt{2}}(E_x + E_z)) &= \frac{1}{2} \text{tr}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\right) = \frac{1}{\sqrt{2}}, \end{aligned}$$

where the mixed rotation was normalized by normalizing the respective axis of rotation  $(1, 0, 1)^\top$ . We see that the metric behaves as intended. Just like a straight line minimizes the distance between any points in euclidean space<sup>1</sup>, the Cartan Killing form can be used to derive the Geodesic distance

$$d(\mathbf{R}_1, \mathbf{R}_2) = \arccos\left(\frac{\text{tr}(\mathbf{R}_1 \cdot \mathbf{R}_2^\top) - 1}{2}\right). \quad (69)$$

With these concepts, we can generalize ideas from euclidean spaces to our manifolds. Since we say above that many formulas for  $SE(3)$  are based on notions of  $SO(3)$  and the latter tends to be easier to understand initially, we will focus the rest of this report on this case.

---

<sup>1</sup>The proof of this is surprisingly involved and requires us to minimize a functional involving the regular inner product. Replacing this with the Cartan Killing Form yields the geodesic distance.

## 4 Brownian Motion on $\text{SO}(3)$

We now want to consider the  $\text{SO}(3)$ -equivalent to the classical concepts of Brownian motion in euclidean spaces.

### 4.1 The Heat Kernel and Gaussian Densities

A fascinating connection exists between the partial differential equation that describes the diffusion of heat through a medium and the familiar Gaussian distribution. This relationship will provide a more intuitive understanding of the upcoming formulations. Consider the function  $u(\mathbf{x}, t)$  with  $\mathbf{x} \in \mathbb{R}^3$  and  $t \in \mathbb{R}_+$  describing the temperature at a position in space and time. In Euclidean space, this system is governed by

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = D\Delta u(\mathbf{x}, t), \quad (70)$$

where  $\Delta$  is the Laplace differential operator and  $D$  is a constant. All this equation says is that the rate of change (w.r.t. time) of the temperature is proportional to the average difference in temperature from the current point to neighboring points.

For a Dirac heat source at  $\mathbf{x} = \mathbf{0}$  at time  $t = 0$ , the solution to the above equation is given by

$$K(\mathbf{x}, t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right), \quad (71)$$

which has the same form as the density function of a Gaussian. The Laplacian operator can be generalized to arbitrary manifolds based on the metric, yielding the Laplace-Beltrami operator

$$\Delta_g f = \frac{1}{\sqrt{\det(g)}} \sum_k \partial_k \left( \sqrt{\det(g)} \sum_i g^{ij} \partial_i f \right) \quad (72)$$

Analogous to the euclidean case, the resulting partial differential equation can be solved at the identity of  $\text{SO}(3)$ . However, finding an explicit solution is more complex and requires the use of an infinite series expansion. According to derivations by Nikolayev et. al. [7], the heat kernel for  $\text{SO}(3)$  can be expressed as

$$f_\epsilon(\theta) = \sum_{\ell=0}^{\infty} (2\ell+1) \exp(-\ell(\ell+1)\epsilon^2) \frac{\sin((\ell+\frac{1}{2})\theta)}{\sin(\frac{\theta}{2})}, \quad \theta \in (-\pi, \pi]. \quad (73)$$

The parameter  $\epsilon$  acts as a concentration parameter that resembles the usual variance. According to the findings of Jagvaral et al. [3], this series converges quickly for  $\epsilon > 1$ , requiring only around 5 terms to reach sub-percentage accuracy. For more concentrated distributions, which are needed in e.g. the diffusion context, there exists a very good closed-form approximation, e.g. derived by Matthies et al. [6] of the form

$$f_\epsilon(\theta) \approx C(\theta) \sqrt{\pi} \epsilon^{-\frac{3}{2}} \exp\left(\frac{\epsilon}{4} - \frac{(\frac{\theta}{2})^2}{\epsilon}\right), \quad \epsilon < 1 \quad (74)$$

$$C(\theta) = \left( \frac{\theta - e^{-\frac{\pi^2}{\epsilon}} ((\theta - 2\pi)e^{\frac{\pi\theta}{\epsilon}} + (\theta + 2\pi)e^{-\frac{\pi\theta}{\epsilon}})}{2 \sin(\frac{\theta}{2})} \right) \quad (75)$$

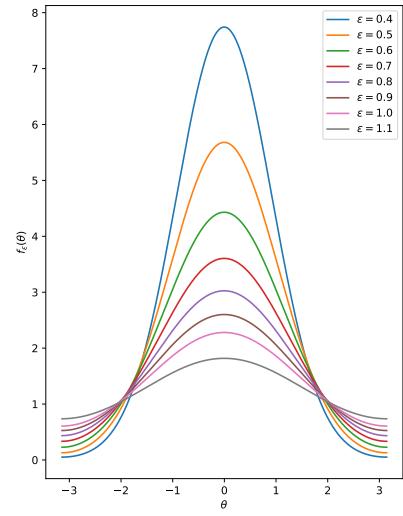


Figure 12: Heat kernel on  $\text{SO}(3)$  for different values of the concentration parameter. For  $\epsilon > 1$ , we computed 1000 summands of the infinite series (73), while for  $\epsilon < 1$  we used the approximation in Equation (75).

In Figure 12, we can see that the resulting function behaves as intended and approaches a uniform distribution for  $\epsilon \rightarrow \infty$ . Because it is a solution to the diffusion equation,  $f_\epsilon$  can be used to define the manifold equivalent to the isotropic Gaussian distribution, which we denote  $\mathcal{IG}_{SO(3)}$ . Based on a mean rotation  $\mathbf{R}_\mu \in SO(3)$ , the probability density of a rotation  $\mathbf{R} \in SO(3)$  under an isotropic Gaussian on the manifold is given by

$$\mathcal{IG}_{SO(3)}(\mathbf{R}'; \mathbf{R}_\mu, \epsilon) = f_\epsilon(\arccos[\frac{1}{2}(\text{trace}(\mathbf{R}_\mu^\top \mathbf{R}') - 1)]), \quad (76)$$

Comparing the argument of the function  $f_\epsilon$  with the Geodesic distance in equation (69), we see that the density depends on the distance to the mean rotation, just as for the euclidean definition of a Gaussian.

Using these definition, we implicitly consider a uniform distribution over rotation axes and only use the angle to define the density. The main motivation for defining such a distribution is that it inherits many of the properties that are fundamental in many applications, e.g. certain combinations of Gaussians being Gaussian again. However, it is important to stress that we also limit ourselves to isometric Gaussians with a single scalar value encoding the variance, thereby assuming no correlation between dimensions.

## 4.2 The Haar Measure and Inverse Transform Sampling

In order to sample from the distribution we defined in the preceding section, we employ inverse transform sampling. This requires integrating over the pdf to obtain the cdf. Since we are no longer dealing with euclidean definitions, we have to be careful when performing integration, as we can no longer use the regular Lebesgue measure. It turns out that what we need is a measure  $\mu$  that is left and right invariant with respect to the group elements, i.e. for all Borel subsets  $S \subseteq G$  and  $g \in G$  it holds that

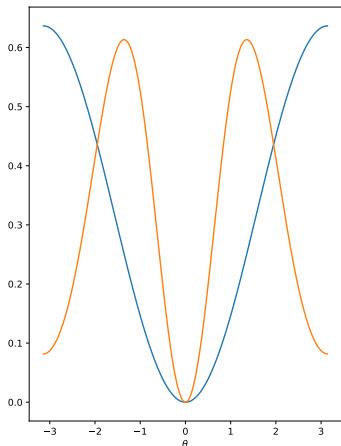


Figure 13: Probability density function (orange) of any angle for the isotropic Gaussian weighted by the Haar measure (blue) based on the heat kernel in Figure 12 with  $\epsilon = 0.5$ .

$$\mu(gS) = \mu(S) \quad \text{and} \quad \mu(Sg) = \mu(S). \quad (77)$$

It can be shown that such a measure can always be found and normalized for compact groups such as  $SO(3)$ , yielding a probability measure as intended. The resulting measure is called Haar measure and for  $SO(3)$  and our initial definitions above, it adds a weighting factor of the form

$$\frac{1 - \cos(\theta)}{\pi} \quad (78)$$

to integrals over  $\theta$ . To understand this formula, we have to remind ourselves of the euclidean pendant to the proposed sampling process. If we were to sample from a three-dimensional Gaussian distribution in an analogous way, we have to first sample a unit-norm translation vector uniformly and then scale it accordingly. It turns out that this scaling parameter is distributed as the so-called Rayleigh distribution, a scaled chi distribution with pdf

$$f(x; \sigma) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}. \quad x \geq 0. \quad (79)$$

Since the translation direction we sample along the scaling is distributed uniformly, the Rayleigh distribution has to account for the fact that at larger radii, there is more and more room for a point to be found at, which is not achieved by the usual Lebesgue measure alone. To gather intuition we

consider trying to sample a uniform point from the unit ball in 3D. If we imagine the ball to be divided into layers like an onion, the probability of the point being in a specific layer has to increase from the origin (where it is 0, since this requires hitting a single point in a continuum) to the surface, because there is more and more space available for the point to land in.

Mathematically, the volume element of spherical coordinates is proportional to  $r^2 dr$ , but since we are considering a density over radii, we get an extra factor of  $1/r$ , yielding the linear term in the pdf. In this case, the linear volume effect dominates at small radii, but at larger radii, the exponential decay in probability takes over, limiting the probability of very large distances.

For our rotation manifold, the Haar measure plays a similar role. Instead of translations, we are moving by combining rotations around arbitrary axes. The measure provides the necessary weighting to account for the structure of the manifold, ensuring that angles close to 0 (representing rotations close to the identity) receive almost no probability mass, while larger angles, corresponding to rotations further from the identity, account for the increased space on the manifold.

In Figure 14, the probability density function of the isotropic Gaussian (orange curve) is shown with the Haar measure (blue curve). For any given rotation axis, a rotation angle of 0 corresponds to the identity rotation, which—being a single point in a continuous manifold—must have no probability mass. As the rotation angle increases, more “space” becomes available on the manifold for the rotations, causing the pdf to grow accordingly. The fact that this growth is not linear as in the Euclidean case can be attributed to the structure of  $SO(3)$ . For instance, the geodesic distance, as discussed in Section 3.5, is related to the rotation angle by an inverse cosine function, which further differentiates it from the linear growth seen in flat Euclidean spaces.

It is important to note that the explanation provided here considers the Haar measure on its own, without the additional effects of the Gaussian exponential decay. The result holds for a uniform distribution on  $SO(3)$ , where the Haar measure dictates the probability density purely based on the geometric structure of the rotation space. In the case of the isotropic Gaussian, both the heat kernel and the Haar measure contribute to the overall pdf. The heat kernel, representing the Gaussian distribution, pulls probability mass closer to the identity rotation, while the Haar measure ensures that no mass is assigned to the identity itself, balancing the two effects. With this in mind, the cdf can be computed via

$$\int_{\theta} \frac{1 - \cos(\theta)}{\pi} f_{\epsilon}(\theta) d\theta, \quad \theta \in (-\pi, \pi]. \quad (80)$$

to perform inverse transform sampling. After this, we can randomly sample a normalized rotation axis uniformly over  $\mathbb{S}^2$  to obtain the axis-angle representation of the sampled group element.

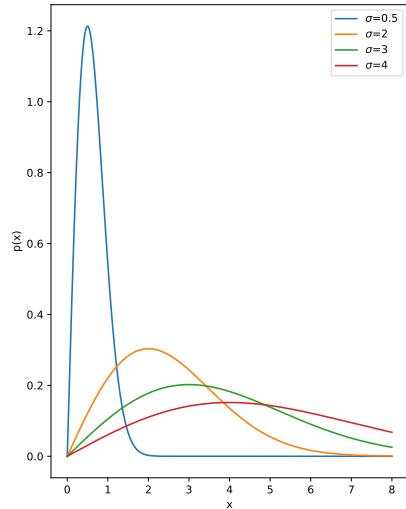


Figure 14: Probability density function of the Rayleigh distribution for different values of the parameter  $\sigma$ .

## 5 Diffusion on $\text{SO}(3)$

With the foundational concepts established in Section 3, we are now prepared to adapt the diffusion framework introduced in Section 2.3 specifically for the manifold of rotations,  $\text{SO}(3)$ . This adaptation is grounded in the workshop paper by Leach et al. [5].

In Euclidean space, we rely on operations such as vector addition and scalar multiplication. However, on the manifold we must define analogous operations that incorporate moving between the manifold and its associated Lie algebra, where computations can be more straightforwardly handled using linear algebra.

### 5.1 Toy Example on $\text{SO}(3)$

In order to add together scaled states as it is required in equations (32) and (21) of the diffusion framework, we can use the analogous

$$\lambda(\gamma, \mathbf{x}) = \exp(\gamma \log(\mathbf{x})), \quad (81)$$

where we pull the rotation back to the Lie Algebra, scale the resulting vector by  $\gamma$  and map it back to the manifold. Since rotations act on each other via matrix-matrix products as we learned in 3.1, we additionally have to replace the plus sign with the matrix product. For the forward process, we want to visualize how the new noising process affects the rotations of the toy dataset in Figure 15. We adapt the formulas in the following way:

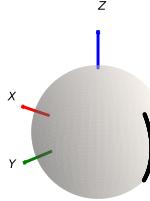


Figure 15: Rotation toy dataset depicted via the effect each rotation has on the point  $(0, 0, 1)^\top$ .

$$q(\mathbf{x}_t | \mathbf{x}_0) \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} \mathbf{x}_{t-1}, (1 - \bar{\alpha}_t) \mathbf{I}) \quad \downarrow \quad (82)$$

$$q(\mathbf{x}_t | \mathbf{x}_0) \sim \mathcal{IG}_{\text{SO}(3)}(\lambda(\sqrt{\bar{\alpha}_t}, \mathbf{x}_0), (1 - \bar{\alpha}_t)) \quad (83)$$

Instead of the euclidean reparametrization trick, we use the sampling procedure described in Section 4.2. The result is depicted in Figure 17, which fits our intuition of samples becoming more and more diverse. Because of the different topology in comparison with  $\mathbb{R}^n$ , the samples approach a uniform distribution  $\mathcal{U}_{\text{SO}(3)}$  instead of a unit Gaussian, which has to be considered when using the reverse process. For the reverse process, the formulas can be adapted via

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) \sim \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\boldsymbol{\beta}}_t \mathbf{I}) \quad (84)$$

$$\downarrow \quad (85)$$

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \sim \mathcal{IG}_{\text{SO}(3)}(\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\boldsymbol{\beta}}_t) \quad (86)$$

$$(87)$$

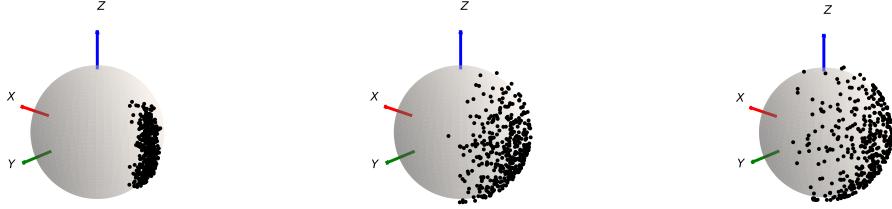


Figure 16: Effect of the forward diffusion on the rotation dataset.

with

$$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t \quad (88)$$

$$\downarrow \quad (89)$$

$$\tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0) = \lambda\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}, \mathbf{x}_0\right) \cdot \lambda\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}, \mathbf{x}_t\right). \quad (90)$$

The intuitive advantage of the equations above is that we now have a scheme to update rotation matrices without ever running into an issue because an intermediate step could yield a non-valid matrix. The objective now is to train a neural network to predict any representation of the sampled noise rotation. We follow our reference and try to predict the vector-representation of the noise  $\mathbf{R}$  by first computing the one-shot sample at time  $t$  via

$$\mathbf{x}_t \rightarrow \mathbf{R} \cdot \exp(\sqrt{\bar{\alpha}_t} \log(\mathbf{x}_0)), \quad \mathbf{x}_0 \sim q, \quad \mathbf{R} \sim \mathcal{IG}_{\text{SO}(3)}(\mathbf{I}, \sqrt{1-\bar{\alpha}_t}). \quad (91)$$

Since we cannot simply scale down samples from  $\mathcal{IG}_{\text{SO}(3)}(\mathbf{I}, \sqrt{1-\bar{\alpha}_t})$  to predict the noise, we perform learning on the vector representation of the Lie algebra, where we can simply scale the vector after applying the log-map.

After training, the network is used to guide the reverse diffusion process by mapping back the predicted Lie algebra elements to rotation matrices and applying them iteratively. The results for our toy dataset and 1000 diffusion steps can be seen in Figure 17. We can see that since our initial



Figure 17: Learned reverse diffusion process from random initial rotations (left) to new samples of the rotation dataset (right).

samples are not uniformly distributed, the resulting distribution is also distorted with more samples ending up on the upper half of the sphere.

## 5.2 Implementation Remarks

While the above adaptation was relatively straight forward to derive, the actual implementation can be tricky. For once, we see in Figure 14 that even for small  $\epsilon$  values, we tend to sample large rotation angles, because the Haar measure pushes the density mass away from the identity. To end up with small increments during the diffusion, a very small parameter is required, which can require double precision. The authors of [5], whose codebase we used for experiments artificially bias the sampling towards smaller angles.

Additionally, a loss function based on the metric in Section 3.5 can be used theoretically, but suffers from numerical instabilities. For the results above, a regular mean squared error between the Lie algebra vectors is used.

Added to this are the problems specific to diffusion models, which require careful tuning of the batch size and learning rate.

## 6 Topology Bits

The section above used a very constrained example to illustrate the theoretical ideas from the first sections. However, in practice the mentioned different topology of  $SO(3)$  can lead to problems in e.g. learning tasks. In this last chapter, we will quickly go over the most basic notions of topology and directly connect them to  $SO(3)$  and other rotation representations. For the sake of completeness, we remark that in the following, we are always referring to connected topological manifolds, while everything can also be introduced in a more general context.

### 6.1 Introduction

Topology is an infamously abstract branch of mathematics that is concerned with the classification of (topological) spaces based on their intrinsic geometric and structural properties without regard to precise distances or angles (aka metrics, which would yield metric spaces). It provides a formal framework for discussing continuity, connectedness, and boundaries in abstract spaces.

In essence, it studies properties that are preserved under continuous transformations like stretching and bending, but not by tearing or glueing. One important tool in topology are homotopic curves. Two curves on a manifold are homotopic, if they can be continuously transformed into one another. If the curves are assumed to be closed, we call them loops. These loops form a so-called equivalence relation, meaning that any two loops that can be transformed into each other are considered the same curve, including being shrunk to a single point. Studying the different types of curves that cannot be transformed into each other therefore provides a useful way to categorize topological spaces. A very common example is that of a torus, described in Figure 18. A manifold where every point can be shrunk to a point is denoted simply connected, think e.g. of the surface of a sphere. Non-trivial homotopy classes are an indication for additional structure of a manifold, which can complicate certain operations or proofs. This is why many advanced proofs in Topology and Riemannian geometry rely on the concept of a covering. A covering is a surjective map  $\pi: E \rightarrow M$  between two manifolds<sup>2</sup>  $E, M$ , where for every point  $p \in M$  there is a neighborhood  $U$  around it, whose preimage is a disjoint union

$$\pi^{-1}(U) = \bigcup_{\alpha \in A} U_\alpha, \quad \pi: U_\alpha \rightarrow U. \quad (92)$$

We can e.g. think of mapping the real number line to the unit circle via the complex exponential. Every small neighborhood on the circle is the image of countable infinite neighborhoods on  $\mathbb{R}$ . For the torus, we can construct a lattice as in Figure 19 by cutting the torus open and repeating it infinitely many times in each direction. Both of these examples have another property: both coverings are simply connected. These so-called universal coverings can be thought of as maximally disentangled versions of a manifold, compare e.g. the two visualizations for the torus: each red or green loop on the original torus becomes an infinite line on the covering. In many proofs these concepts are used to map an object to a covering, derive a property there and then map it back to the original manifold to avoid the more complicated structure that comes with the non-trivial homotopy classes.

---

<sup>2</sup>Technically, we differentiate between topological, differentiable and isometric covering depending on whether the manifolds are topological, differentiable or Riemannian. In these cases the map  $\pi$  has to be a homeomorphism (preserve topology) / diffeomorphism (bijective, differentiable) or an isometry (preserve metric).

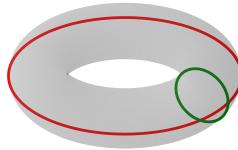


Figure 18: Homotopy classes of a torus. Neither the green nor the red loop can be shrunk to a point. Also, both types of curves cannot be transformed into each other. Since we can simply use the curves that loop around the marked lines multiple times, the homotopy class is  $\mathbb{Z} \times \mathbb{Z}$  (a positive integer number of loops for each of the two types).

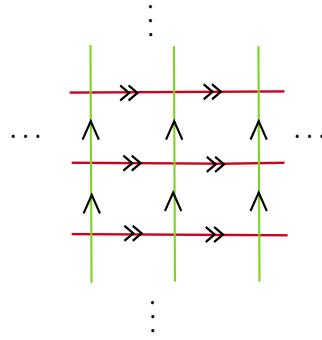


Figure 19: Covering of the torus constructed as a lattice. The coloured lines are the same ones as indicated in Figure 18, while the black symbols indicate identification.

## 6.2 The case $\text{SO}(3)$ and Relationship with other Rotation Representations

While  $\text{SO}(3)$  is compact and connected, it has a non-simply connected structure with homotopy class  $\mathbb{Z}_2$ . We can envision the axis-angle vectors as arrows attached to the origin pointing into space. The  $\text{SO}(3)$ -manifold is then a ball with radius  $\pi$ , where antipodal points are identified with each other. For reference and details on the somewhat unusual homotopy classes, we refer to the excellent visualizations provided by [1]. This property is somewhat surprising: it means that by not differentiating between rotations by  $\theta$  and  $2\pi + \theta$ , we end up with a rather complex manifold with additional non-trivial structure. Even more surprisingly, because the homotopy class is  $\mathbb{Z}_2$ , a double cover suffices to disentangle that structure and obtain a simply connected manifold, meaning that we do not differentiate between rotations that are different by  $4\pi$  instead. This manifold is the Lie group  $\text{SU}(2)$ , which is homeomorphic to the set of unit quaternions. Because they unwrap the complication caused by the antipodal identification, unit quaternions provide a simpler structure, which is why they have proven useful in practice. In addition to being a Lie group, unit quaternions offer a fascinating interpretation via Geometric Algebra, a field that is concerned with higher-dimensional objects called multivectors. This perspective can remove some of the difficulties quaternions pose to practitioners due to their abstract nature. However, Geometric Algebra itself requires a conceptual leap to get an intuitive understanding of multivectors that would require a report like this on its own.

What remains to be said is that the simpler structure of quaternions comes at the cost of introducing

ambiguities, as the double cover represents each rotation twice. This can have its own complications in practice.

### 6.3 Practical Implications

The practical complications that result from the more complex structure of the Lie algebra and the quaternion double cover can be summarized as follows:

- In the quaternion representation, where  $\pm \mathbf{q}$  represent the same rotation, tracking algorithms, such as for animation or motion capture, can suffer from sudden sign flips. This can happen during interpolation or tracking a continuous rotation, leading to disorienting behavior. Even though the rotation itself remains smooth, this can introduce discontinuities in the quaternion values, confusing algorithms that expect a smooth trajectory.
- When working with Lie algebra representations of rotations we used in this report, while the lower-dimensional representation simplifies certain operations (e.g., dealing with only three parameters for rotations), it also introduces singularities, particularly for large rotations. These singularities arise from the non-linear nature of the exponential map, stemming from the antipodal identification explained above. Near these singularities, small increments in the Lie algebra might correspond to large and unpredictable changes in the rotation group (to the antipodal point), making accurate computations difficult for large angles. However, for small rotations, these singularities are not problematic, which is why techniques like the diffusion framework in Section 5 tend to work well for local computations. For additional examples please refer to e.g. [2].

There are numerous other rotation representations that try to avoid the complications described here, but none of them offer a perfect solution and the choice of which one to use is highly dependent on the application.

## References

- [1] R. Behiel. *The Mystery of Spinors*. Accessed: 2024-09-30. Youtube. 2024. URL: <https://www.youtube.com/watch?v=b70IbMCIfs4&t=1143s>.
- [2] A. R. Geist, J. Frey, M. Zobro, A. Levina, and G. Martius. *Learning with 3D rotations, a hitchhiker’s guide to  $SO(3)$* . 2024.
- [3] Y. Jagvaral, F. Lanusse, and R. Mandelbaum. *Unified framework for diffusion generative models in  $SO(3)$ : applications in computer vision and astrophysics*. arXiv:2312.11707 [cs]. 2023.
- [4] K. J. H. Law, A. M. Stuart, and K. C. Zygalakis. *Data Assimilation: A Mathematical Introduction*. 2015.
- [5] A. Leach, S. M. Schmon, M. T. Degiacomi, and C. G. Willcocks. “Denoising Diffusion Probabilistic Models on  $SO(3)$  for Rotational Alignment”. en. In: (2022).
- [6] S. Matthies, J. Muller, and G. W. Vinel. “On the Normal Distribution in the Orientation Space”. en. In: *Texture, Stress, and Microstructure* 10.1 (1988), pp. 77–96.
- [7] D. I. Nikolayev and T. I. Savyolov. “Normal Distribution on the Rotation Group  $SO(3)$ ”. en. In: *Texture, Stress, and Microstructure* 29.3-4 (1997), pp. 201–233.
- [8] J. Solà, J. Deray, and D. Atchuthan. *A micro Lie theory for state estimation in robotics*. arXiv:1812.01537 [cs]. 2021.
- [9] THE GEOMETRY OF THE EXCEPTIONAL LIE GROUPS. <https://books.physics.oregonstate.edu/GELG/GELG.html>. Accessed: 2024-09-30.

- [10] J. L. Wellens. “A FRIENDLY INTRODUCTION TO GROUP THEORY”. In: 2015.
- [11] L. Weng. “What are diffusion models?” In: *lilianweng.github.io* (2021).