# Local Gaussian Processes on Compact Lie Groups

**Bochuan Liu, Mingyang Zhao*, Xiaohong Jia**
State Key Laboratory of Mathematical Sciences
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
University of Chinese Academy of Sciences
{liubochuan,zhaomingyang,xhjia}@amss.ac.cn

## Abstract

Gaussian processes are a cornerstone of modern probabilistic machine learning, providing principled uncertainty quantification and flexible nonparametric modeling. While the Gaussian kernel is widely used in Euclidean spaces, many real-world problems involve data residing on non-Euclidean domains, particularly compact Lie groups that naturally encode *symmetries*. In this work, we construct a local exponential-coordinate kernel on arbitrary compact Lie groups by decomposing the group into a *maximal torus* and its orthogonal complement, which overcomes the limitation that no globally positive definite Gaussian kernel exists on compact Lie groups. For the prominent Lie groups $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$, we derive closed-form kernel expressions using the Rodrigues formula. Extensive experiments verify the positive definiteness of the proposed kernel and demonstrate its effectiveness in Gaussian process regression, achieving more accurate predictions across the entire manifold compared to prior methods, while offering significantly higher computational efficiency. Code is available at https://github.com/liubochuan/Local-Gaussian-Processes-on-Compact-Lie-groups.

## 1 Introduction

*Gaussian processes* (GPs) are powerful probabilistic models with widespread applications in machine learning [1, 2, 3], statistics [4, 5], and control engineering [6, 7]. Their key appeal lies in providing both *predictions* and principled *uncertainty quantification*, a capability that is indispensable for tasks such as Bayesian optimization [8], variational inference [9, 10], and reinforcement learning [11, 12].

A GP is fully specified by its mean function and kernel function, with the latter encoding prior assumptions about the smoothness and structure of the process. Among the most widely used kernels is the Gaussian (also known as the squared exponential or RBF) kernel, which is typically defined in a Euclidean vector space. However, motivated by applications in robotics [13, 14], computer vision [15, 16], and computational chemistry [17, 18], where data often reside on manifolds such as the *rotation group* $\mathbf{SO}(3)$ or the *Grassmann manifold* $\mathcal{G}_n^r$, there has been growing interest in generalizing GPs from Euclidean vector spaces to manifolds [19, 20].

The key challenge in this generalization lies in preserving the positive definiteness of the kernel function, a fundamental requirement for a valid kernel. A natural approach to generalizing the Gaussian kernel to a Riemannian manifold is to replace the Euclidean norm with the *geodesic distance*. However, as shown in [21], this construction manner yields a positive semi-definite kernel for all bandwidths $\lambda > 0$ only when the underlying manifold is isometric to a Euclidean space. Consequently, on spaces with *non-zero* curvature or on *compact* Riemannian manifolds that cannot be isometric to a Euclidean space, the geodesic Gaussian kernel is provably indefinite. This limitation is particularly restrictive because the "for all $\lambda > 0$" condition is what allows the bandwidths can be learned freely from data.

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

This theoretical obstruction has motivated many innovative work. For instance, [22] proposed a positive definite Gaussian kernel on the space of positive definite matrices (a non-compact Riemannian manifold) via the log-Euclidean geodesic distance. For spaces of negative curvature, such as hyperbolic space, [23] developed a positive definite hyperbolic RBF kernel by embedding the hyperbolic space into the Euclidean space. For general compact Riemannian manifolds, [24] defined a Matérn kernel on a flat torus via periodic summation and further generalized the construction to arbitrary compact Riemannian manifolds through the Laplace-Beltrami operator. [25] proposed the use of reproducing kernel Krein space-based methods, which relax the positive definiteness requirement by only requiring kernels that admit a positive decomposition. Rather than requiring positive definiteness for all bandwidths, [26] instead sought bandwidth intervals on which geodesic exponential kernels are positive definite with high probability over finite datasets. Furthermore, [27] identified spaces where Gaussian kernel is not positive definite for any parameter value. However, there is still no general method to construct a positive definite kernel on compact Riemannian manifolds for all parameters.

As a representative Riemannian manifold, the *compact Lie group* naturally describes *symmetries* encountered in various applications, particularly in the physical sciences and engineering [28, 29, 30, 31]. The invariance of a Gaussian process kernel to such symmetries gives rise to the most natural generalization of the concept of stationarity to such spaces [32]. For instance, [32] developed techniques to derive algebraic expressions for the heat kernel and Matérn kernel on compact Lie groups and their homogeneous spaces by combining results from differential geometry and representation theory. In contrast, in this work, we present a novel method and show how to construct a positive definite exponential-coordinate kernel on compact Lie groups via the Lie algebra decomposition, which has many applications for a large class of non-Euclidean spaces arising from symmetries.

We bridge the gap between the geodesic Gaussian kernel and the practical need for such a kernel on compact Lie groups through a *constructive* framework. Our key insight is to leverage the local Euclidean property of these groups: instead of seeking a global kernel, we construct a local exponential-coordinate kernel that is provably positive definite within a neighborhood. To aid understanding, we first present a construction of the local geodesic Gaussian kernel on *compact connected abelian Lie groups*, which are isomorphic to a flat torus $\mathbb{T}^n$. Although [24] provided the intuition that the naive geodesic Gaussian kernel is *locally correct* on flat torus, it does not construct an explicit neighborhood admitting a positive definite Gaussian kernel for all bandwidths. Instead, it constructs a global kernel via the periodic summation over the lattice using Euclidean Gaussian kernel, which is no longer a geodesic Gaussian kernel. In contrast, we show that for any point $p \in \mathbb{T}^n$, the *injectivity radius* defines a geodesically convex neighborhood isometric to a Euclidean domain, on which the geodesic Gaussian kernel naturally remains positive definite. Crucially, owing to the homogeneity of compact Lie groups, this neighborhood is uniform in both size and geometry across the entire manifold. Moreover, the geodesic convexity of the neighborhood guarantees that the form of the constructed geodesic kernel is consistent across the manifold.

Building on the above theoretical foundation, we further construct the local exponential-coordinate kernel on an *arbitrary compact Lie group*. Using the tool of *Cartan subalgebra* [33], we decompose such a group into a *maximal torus* and its orthogonal complement. This allows us to factorize the local exponential-coordinate kernel as a product of two components: the local Gaussian kernel on the torus (from our preceding result) and a standard Gaussian kernel on the Euclidean space. We present a complete algorithm for this framework. As concrete examples, we derive closed-form expressions for two groups of central importance in robotics and computer vision: the rotation group $\mathbf{SO}(3)$ and the special unitary group $\mathbf{SU}(2)$. The Rodrigues formula is employed for efficient kernel computation. Extensive experiments on $\mathbf{SO}(3)$ verify the positive definiteness of the proposed kernel across a wide range of configurations. We further prove, via a *quaternion-based analysis*, that the kernel is positive definite almost everywhere on the entire group. Finally, we benchmark our kernel against state-of-the-art Matérn kernels on compact Riemannian manifolds and Lie groups, demonstrating competitive predictive accuracy and substantially higher computational efficiency. Moreover, a robotics denoising experiment on simulated manipulator trajectories further demonstrates the practical advantage of our geometry-aware kernel over the Euclidean RBF baseline.

The main technical contributions of this work are fourfold:

1. We develop a constructive framework for building local Gaussian kernels on compact Lie groups that are provably positive definite.

2. We characterize a family of local Gaussian kernels on the flat torus $\mathbb{T}^n$ that take a consistent form defined via the geodesic distance.
3. We derive closed-form expressions for the local Gaussian kernel on the rotation group $\mathbf{SO}(3)$ and the special unitary group $\mathbf{SU}(2)$. Moreover, we prove that the constructed kernel is positive definite almost everywhere on $\mathbf{SO}(3)$.
4. We conduct extensive Gaussian process regression experiments on $\mathbf{SO}(3)$ comparing the proposed kernel against state-of-the-art Matérn kernels, demonstrating superior accuracy and computational efficiency. A robotics denoising experiment further shows that our exponential-coordinate kernel outperforms the Euclidean RBF baseline across all noise levels.

## 2 Preliminaries

**Definition 2.1** (Gaussian processes)**.** Let $\mathcal{X}$ be a data set. A stochastic process $f : \mathcal{X} \to \mathbb{R}$ is called a Gaussian process, *i.e.*, $f(\mathbf{x}) \sim \mathcal{GP}(m, k)$, if for any finite collection of points $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathcal{X}$, the random vector $f(\mathbf{x}) = [f(\mathbf{x}_1), ..., f(\mathbf{x}_n)]$ follows a multivariate Gaussian distribution. The Gaussian process is completely determined by its mean function $m(\mathbf{x})$ and covariance or kernel function $k(\mathbf{x}, \mathbf{x}')$:

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \end{aligned} \tag{1}$$

Here $\mathbb{E}$ is the expectation operator.

The requirement for a bivariate function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ to be a kernel is *positive semi-definite* (also called *positive definite* in the literature), in the sense that for any $n$ and any $\mathbf{x} \in \mathcal{X}^n$, the kernel matrix $\mathbf{K_{xx}}$ is positive semi-definite. One of the most popular kernel functions on $\mathbb{R}^n$ is the Gaussian kernel (or RBF kernel):

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2l^2}\right), \tag{2}$$

where $\sigma$ and $l \in \mathbb{R}$.

### 2.1 A Naive Generalization of Kernels on Lie Groups

We are interested in generalizing the Gaussian kernel Eq. (2) from the Euclidean setting to general compact Lie groups. In our context, let $\mathcal{G}$ be a compact Lie group equipped with a bi-invariant metric $Q$, *i.e.*, a Riemannian metric on $\mathcal{G}$ for which left and right translations are both isometries. A naive generalization is to replace the Euclidean norm in Eq. (2) with a geodesic distance:

$$k(x, y) = \exp\left(-\lambda d(x, y)^2\right), \forall x, y \in \mathcal{G}, \lambda > 0, \tag{3}$$

where $d$ is the geodesic distance on $\mathcal{G}$ with respect to $Q$. However, $k$ may not be a well-defined kernel on $\mathcal{G}$ by the following theorem.

**Theorem 2.2.** *(Theorem 2 in [21]) Let $\mathcal{M}$ be a complete Riemannian manifold with associated geodesic distance $d$. The geodesic Gaussian kernel $k(x, y)$ in Eq. (3) is positive semi-definite for all $\lambda > 0$ if and only if $\mathcal{M}$ is isometric to a Euclidean space.*

Since a Euclidean space is not compact, the kernel $k(\mathbf{x}, \mathbf{x}')$ in Eq. (3) cannot be positive definite for any $\lambda > 0$ on a compact Lie group. Instead, we aim to define a local Gaussian kernel motivated by the local Euclidean property of compact Lie groups. Let us first consider the simplest case: the compact connected abelian Lie group.

## 3 Local Geodesic Kernel on Compact Connected Abelian Lie Groups

Before constructing the local exponential-coordinate kernel on a general compact Lie group, we first study the simplest and most fundamental case: the compact connected abelian Lie group, *i.e.*, the flat torus. Owing to its flatness, this case is particularly tractable and serves as a building block for the maximal torus used in the general construction. In this section, we detail the construction of a local Gaussian kernel on such groups.

**Lemma 3.1.** *(Theorem 7.3 in [34]) A Riemannian manifold $\mathcal{M}$ is flat if and only if it is locally isometric to a Euclidean space.*

**Proposition 3.2.** *Let $\mathcal{G}$ be a compact connected abelian Lie group endowed with a bi-invariant metric $Q$. Then, $\mathcal{G}$ is flat.*

*Proof.* All theoretical proofs are presented in the Appendix. □

From Lemma 3.1, for any point $p \in \mathcal{G}$, there exists a neighborhood $U_p \subset \mathcal{G}$ that is locally Euclidean, allowing us to define a Gaussian kernel on it unconditionally (positive definite for all bandwidths). The key challenge lies in constructing such a neighborhood $U_p$. To address this, we introduce the concept of the *injectivity radius* in the following.

**Definition 3.3** (Injectivity radius)**.** For any point $p \in \mathcal{M}$, let $T_p^{\mathcal{M}}$ be the tangent space at $p$. The injectivity radius $\mathrm{Inj}(p)$ at $p$ is defined as the biggest $R > 0$ such that $\exp_p : B_R(0) \subset T_p^{\mathcal{M}} \to \mathcal{M}$ is a diffeomorphism onto its image, where $B_R(0) = \{v \in T_p^{\mathcal{M}} | \|v\|_p < R\}$. The injectivity radius of $\mathcal{M}$ is defined as the infimum of the injectivity radius at all points.

**Lemma 3.4.** *The injectivity radius $Inj(p)$ is a constant on any homogeneous Riemannian manifold.*

In our context, the Lie group $\mathcal{G}$ is automatically homogeneous under a bi-invariant metric, and thus the injectivity radius is constant on $\mathcal{G}$. Throughout this paper, we denote this constant by $r$. Let us now concentrate on the flat torus, which serves as the most classical model for a compact connected abelian Lie group.

**Flat torus.** The flat torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ is a compact connected abelian Lie group with the product manifold structure, where the multiplication in $\mathbb{S}^1$ is given by complex multiplication, and multiplication in $\mathbb{T}^n$ is defined componentwise. Equivalently, let $\Gamma$ be a discrete subgroup of $\mathbb{R}^n$, then $\mathbb{T}^n$ can be defined as a quotient space $\mathbb{T}^n \cong \mathbb{R}^n/\Gamma$. The projection map $\pi : \mathbb{R}^n \to \mathbb{T}^n$ given by $\pi(x) = \bar{x} = x + \Gamma$ is a local isometric mapping (isometry), and $\mathbb{R}^n$ is a *universal cover* of $\mathbb{T}^n$.

As a fundamental classification result in Lie theory, we have the following conclusion.

**Lemma 3.5.** *(Corollary 2.20 in [35]) All compact connected abelian Lie groups are tori, up to isomorphism.*

Therefore, determining $\mathrm{Inj}(p)$ for a compact connected abelian Lie group is essentially equivalent to finding $\mathrm{Inj}(p)$ for a torus $\mathbb{T}^n$.

**Lemma 3.6.** *[36] For a Riemannian manifold with curvature $K \leq 0$, the injectivity radius satisfies*

$$\mathrm{Inj}(p) = \frac{1}{2} L(\gamma_p), \tag{4}$$

*where $\gamma_p$ is the shortest geodesic loop based at $p$ (i.e., $p$ is both the start and end point) and $L(\gamma_p)$ denotes its length.*

From this lemma, we can determine the injectivity radius for $\mathbb{T}^n$.

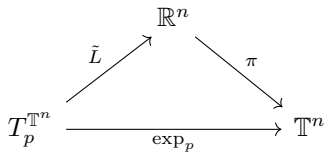
**Theorem 3.7.** *For any $p \in \mathbb{T}^n$, its injectivity radius satisfies:*

$$\mathrm{Inj}(p) = \frac{1}{2} \min_{v \in \Gamma \setminus \{0\}} ||v|| \tag{5}$$

*Remark* 3.8. If we take $\Gamma = \mathbb{Z}^n$, then $r = \frac{1}{2}$.

Now we are ready to construct the local neighborhood that admits a positive semi-definite Gaussian kernel. For any point $p \in \mathbb{T}^n$, let $B_r(0) \subset T_p^{\mathbb{T}^n}$ be the tangent ball defined by the injectivity radius, and $U_p = \exp_p(B_r(0))$ be its diffeomorphic image under the exponential map. We claim that $U_p$ is isometric to an open ball in $\mathbb{R}^n$.

**Lemma 3.9.** *Let $\tilde{p}$ be a lifting of $p$ in $\mathbb{R}^n$: $\pi(\tilde{p}) = p$. Since $\pi$ is a local isometry, $d\pi_{\tilde{p}} : \mathbb{R}^n \to T_p^{\mathbb{T}^n}$ is an isomorphism. Let $L$ be the inverse of the differential $d\pi_{\tilde{p}}^{-1}$ and $\tilde{L} : T_p^{\mathbb{T}^n} \to \mathbb{R}^n$ be the translation of $\tilde{p}$ under $L(v)$: $\tilde{L}(v) = \tilde{p} + L(v)$. Then the following diagram commutes:*




*Specifically,* $\pi \circ \tilde{L} = \exp_p$.

**Lemma 3.10.** *Let* $B_r(\tilde{p})$ *be the open ball in* $\mathbb{R}^n$ *determined by the injectivity radius* $r$*. Then,* $\tilde{L}(B_r(0)) = B_r(\tilde{p})$.

The commutativity of the diagram (Lemma 3.9) and Lemma 3.10 imply that

$$U_p = \exp_p(B_r(0)) = \pi(\tilde{L}(B_r(0))) = \pi(B_r(\tilde{p})). \tag{6}$$

Consequently, $\pi$ maps $B_r(0)$ onto $U_p$. Moreover, thanks to the injectivity radius, its restriction on $B_r(\tilde{p})$ is injective and hence a diffeomorphism.

**Theorem 3.11.** $\pi|_{\pi^{-1}(U_p)} : B_r(\tilde{p}) \to U_p$ *is an isometry.*

Theorem 3.11 implies that for any $p \in \mathbb{T}^n$, there exists a neighborhood of $p$, defined by the injectivity radius, that is Euclidean. Consequently, we can define a Gaussian kernel on this neighborhood unconditionally:

$$k(x, y) = \exp(-\lambda d_U(x, y)^2), \forall x, y \in U_p, \tag{7}$$

where $d_U$ denotes the geodesic distance in $U_p$. By Theorem 2.2, this kernel is positive definite for all $\lambda > 0$. Moreover, the geodesic distance $d_U$ is independent of the choice of neighborhood $U_p$. We demonstrate this property using the concept of *geodesic convexity*.

**Definition 3.12.** Let $\mathcal{M}$ be a Riemannian manifold. A subset $C \subset \mathcal{M}$ is called geodesic convex if given any two points in $C$, there exists a unique minimizing geodesic contained within $C$ joining these two points.

**Proposition 3.13.** $U_p$ *is geodesic convex.*

By Proposition 3.13 and Definition 3.12, for any two points $x, y \in U_p$, the geodesic distance in $U_p$ coincides with that in $\mathbb{T}^n$: $d_U(x, y) = d(x, y)$. We have thus obtained the final result.

**Theorem 3.14.** *For any* $p \in \mathbb{T}^n$*, there exists a neighborhood* $U_p$ *given by Eq. 6 in which the Gaussian kernel*

$$k_p(x, y) = \exp(-\lambda d(x, y)^2), \forall x, y \in U_p, \lambda > 0 \tag{8}$$

*is well-defined, where* $d(x, y)$ *is the geodesic distance in* $\mathbb{T}^n$*. This local Gaussian kernel only depends on point* $p$*, and is consistent in the sense of injectivity radius and geodesic distance.*

## 4 Local Exponential-Coordinate Kernel on Compact Lie Groups

In the previous section, we exploit the flatness of compact connected abelian Lie groups. Nevertheless, a general compact Lie group is not flat. This raises a fundamental difficulty: how can we find a local neighborhood in a compact Lie group $\mathcal{G}$ that admits a positive definite kernel for all parameters? We solve this problem using the key concept of *maximal torus*.

**Definition 4.1** (Maximal torus)**.** A torus in a compact Lie group $\mathcal{G}$ is a compact connected abelian Lie subgroup of $\mathcal{G}$. A maximal torus $\mathcal{T}$ is one which is the maximal among all tori.

Note that a maximal torus $\mathcal{T}$ always exists in a compact Lie group. Another important concept associated with a maximal torus is the *Cartan subalgebra*.

**Definition 4.2.** Let $\mathfrak{g}$ be the Lie algebra of a compact Lie group $\mathcal{G}$. A maximal abelian subalgebra of $\mathfrak{g}$ is called a Cartan subalgebra.

As a main result in Lie theory, the following conclusion holds.

**Proposition 4.3** ( [37])**.** *Let* $\mathcal{G}$ *be a compact Lie group. There is a one to one correspondence between the maximal tori of* $\mathcal{G}$ *and the Cartan subalgebras of* $\mathfrak{g}$.

We now return to the problem at hand. Let $\mathcal{G}$ be a compact Lie group endowed with a bi-invariant metric $Q$, and let $r$ be its injectivity radius. Suppose that $\mathcal{T} \subset \mathcal{G}$ is a fixed maximal torus in $\mathcal{G}$ and $\mathfrak{t}$ is its corresponding Cartan subalgebra. Using the Cartan subalgebra, we can decompose the Lie algebra as $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{t}^\perp$, where $\mathfrak{t}^\perp = \{X \in \mathfrak{g} \mid \langle X, Y \rangle = 0,\ \forall Y \in \mathfrak{t}\}$ is the orthogonal complement of $\mathfrak{t}$ in $\mathfrak{g}$. This decomposition inspires us to define the Gaussian kernel on $\mathfrak{t}$ and $\mathfrak{t}^\perp$ separately, and then derive the desired Gaussian kernel on $\mathcal{G}$. First, we have the following conclusion.

**Theorem 4.4.** *For any point $p \in \mathcal{G}$, there exists a $r_p > 0$ and $\epsilon_p > 0$ such that the map*

$$\Phi_p : B_{\mathfrak{t}}(0, r_p) \times B_{\mathfrak{t}^\perp}(0, \epsilon_p) \to \mathcal{G}, \Phi_p(X, Y) = p \exp(X) \exp(Y) \tag{9}$$

*is a diffeomorphism onto its image $U_p = \Phi_p(B_{\mathfrak{t}}(0, r_p) \times B_{\mathfrak{t}^\perp}(0, \epsilon_p))$, where $B_{\mathfrak{t}}(0, r_p) = \{X \in \mathfrak{t} \mid \|X\| < r_p\}$ and $B_{\mathfrak{t}^\perp}(0, \epsilon_p) = \{Y \in \mathfrak{t}^\perp \mid \|Y\| < \epsilon_p\}$ denote the open balls in $\mathfrak{t}$ and $\mathfrak{t}^\perp$, respectively.*

We now define the Gaussian kernel on $U_p$. By virtue of the diffeomorphism, this construction factorizes into two components: a Gaussian kernel on $B_{\mathfrak{t}}(0, r_p)$ lying in the *flat torus*, and a Gaussian kernel on $B_{\mathfrak{t}^\perp}(0, \epsilon_p)$ lying in a *Euclidean space*.

**Proposition 4.5.** *Let $r_p \leq r$, and let $U_e \subset \mathcal{T}$ be a neighborhood of the identity element $e$ in $\mathcal{T}$. Then the exponential map $\exp_e : B_{\mathfrak{t}}(0, r_p) \to \exp_e\big(B_{\mathfrak{t}}(0, r_p)\big) \subset U_e$ is an isometry.*

Since $U_e$ is Euclidean, $B_{\mathfrak{t}}(0, r_p)$ is also Euclidean via the exponential map, and therefore admits a well-defined Gaussian kernel. Moreover, the orthogonal complement $\mathfrak{t}^\perp$ equipped with the restricted inner product is naturally isometric to a Euclidean space by choosing an orthonormal basis of $\mathfrak{t}^\perp$. Consequently, $B_{\mathfrak{t}^\perp}(0, \epsilon_p)$ can be identified with an open ball in $\mathbb{R}^d$ on which a Gaussian kernel can be defined. Combining these results, we obtain the positive definite exponential-coordinate kernel on $U_p$ for all bandwidths.

**Theorem 4.6.** *For any $q, q' \in U_p$, let $\Phi_p^{-1}(q) = (X_p(q), Y_p(q))$ and $\Phi_p^{-1}(q') = (X_p(q'), Y_p(q'))$ be the preimage under the diffeomorphism. Then the exponential-coordinate kernel $k_p(q, q') = \exp(-\lambda_1 ||X_p(q) - X_p(q')||^2 - \lambda_2 ||Y_p(q) - Y_p(q')||^2)$ is positive semi-definite for all $\lambda_1, \lambda_2 > 0$.*

We summarize the construction of local exponential-coordinate kernel on compact Lie groups in Algorithm 1 (Appendix C).

## 5 Examples

In this section, we present two concrete examples on the representative Lie groups $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$ to illustrate the construction of the local exponential-coordinate kernel.

### 5.1 Local Exponential-Coordinate Kernel on $\mathbf{SO}(3)$

The special orthogonal group $\mathbf{SO}(3) = \{Q \in \mathbf{GL}_3(\mathbb{R}) | QQ^\top = Q^\top Q = \mathbf{I}, \det(Q) = 1\}$ is a compact, connected, and semisimple Lie group of dimension 3. Its Lie algebra is the set of all skew symmetric matrices $\mathfrak{so}(3) = \{X \in \mathfrak{gl}_3(\mathbb{R}) | X + X^\top = 0\}$. Its Killing form is given by $B(X, Y) = \operatorname{tr}(XY)$.

**Proposition 5.1** ( [38]). *Let $\mathcal{G}$ be a $n$-dimensional semisimple connected Lie group. Then $\mathcal{G}$ is compact if and only if its Killing form $B$ is negative definite.*

By Proposition 5.1, the negative Killing form $-B$ defines an inner product on Lie algebra $\mathfrak{so}(3)$. In fact, $-B$ is the Frobenius inner product on $\mathfrak{so}(3)$: $-B(X, Y) = \operatorname{tr}(X^\top Y) = \langle X, Y \rangle_F$. Since the adjoint representation $\mathrm{Ad}_g$ is a Lie algebra automorphism, the Killing form is also adjoint invariant by Proposition A.13. Moreover, we have the following conclusion.

**Theorem 5.2.** *The negative Killing form $-B$ can be extended to a bi-invariant metric on $\mathbf{SO}(3)$.*

To construct the local exponential-coordinate kernel via Algorithm 1 , we first identify a maximal torus in $\mathbf{SO}(3)$ and its Cartan subalgebra, which together decompose the Lie algebra and provide the local coordinates for $\mathbf{SO}(3)$.

**Theorem 5.3.** *The subgroup*

$$\mathcal{T} = \left\{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \middle| \theta \in [0, 2\pi) \right\} \tag{10}$$

*consisting of all rotations about the z-axis in $\mathbb{R}^3$ is a maximal torus in $\mathbf{SO}(3)$, its Cartan subalgebra is given by*

$$\mathfrak{t} = \left\{ \begin{pmatrix} 0 & -a & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \middle| a \in \mathbb{R} \right\}. \tag{11}$$

*Remark* 5.4. $\mathcal{T}$ in Theorem 5.3 is the set of all rotations about the $z$-axis. In fact, for any fixed axis in $\mathbb{R}^3$, the set of all rotations about this axis forms a maximal torus of $\mathbf{SO}(3)$.

**Proposition 5.5.** *The orthogonal complement of $\mathfrak{t}$ in $\mathfrak{so}(3)$ is given by*

$$\mathfrak{t}^\perp = \left\{ \begin{pmatrix} 0 & 0 & b \\ 0 & 0 & -c \\ -b & c & 0 \end{pmatrix} \middle| b, c \in \mathbb{R} \right\}. \tag{12}$$

To determine the radius of the open ball in Theorem 4.4, we need the injectivity raidus $r$ of $\mathbf{SO}(3)$. This in turn requires the sectional curvature, which we compute next.

**Proposition 5.6.** *$(\mathbf{SO}(3), -B)$ has constant sectional curvature equal to $\frac{1}{8}$.*

The following lemma provides a formula for the injectivity radius of a manifold with positive sectional curvature.

**Lemma 5.7.** *[36] Let $(\mathcal{M}, g)$ be a compact Riemannian manifold with constant section curvature $K > 0$. Then*

$$\mathrm{Inj}(\mathcal{M}) = \frac{1}{2}(\textit{length of the shortest closed geodesic}). \tag{13}$$

**Theorem 5.8.** *The injectivity radius $r$ of $\mathbf{SO}(3)$ is $\sqrt{2}\pi$.*

Now we are ready to compute the neighborhood $U_p$ in $\mathbf{SO}(3)$ and define the Gaussian kernel $k_p$ on it. From Theorem 4.4, it suffices to compute $U_e$ and $k_e$ without loss of generality. Since for each point $p$, $U_p = pU_e$ and $\Phi_p^{-1} = \Phi_e^{-1}p^{-1}$. To compute the neighborhood $U_e$ explicitly, we employ the Rodrigues formula (see Appendix B.3.1 for detailed derivation) which gives the explicit formula of exponential map from $\mathfrak{so}(3)$ to $\mathbf{SO}(3)$. In our context, for simplicity, we take $r_p = \epsilon_p = r$ for all $p \in \mathbf{SO}(3)$. Then we deduce the explicit form of $B_{\mathfrak{t}}(0, r)$ and $B_{\mathfrak{t}^\perp}(0, r)$ in the following.

**Lemma 5.9.** $B_{\mathfrak{t}}(0, r) = \{X = aL_1 | |a| < \pi, a \in \mathbb{R}\}$, $B_{\mathfrak{t}^\perp}(0, r) = \{Y = bL_2 + cL_3 | b^2 + c^2 < \pi^2, b, c \in \mathbb{R}\}$.

The notations $L_1, L_2, L_3$ is a basis for $\mathfrak{so}(3)$ explained in Appendix B.14. Finally, we have the neighborhood $U_e$ as follows.

**Theorem 5.10.** *The neighborhood $U_e \subset \mathbf{SO}(3)$ in Theorem 4.4 is given by the composition of two rotations determined by three angles $a, \phi, \psi$:*

$$U_e = \{R_{\mathbf{z}}(a)R_{\mathbf{n}}(\psi) | |a| < \pi, 0 < \psi < \pi\}, \tag{14}$$

*where $\mathbf{z} = [0, 0, 1]^\perp$ represents the z-axis and $\mathbf{n} = [0, \sin(\phi), \cos(\phi)]^\perp$,$0 < \phi < 2\pi$.*

To determine the local Gaussian kernel on $U_e$, we need the explicit form of the inverse map $\Phi_e^{-1}$ which enables us to compute the Gaussian kernels on two components.

**Theorem 5.11.** *For any $g = [g_{ij}] \in U_e$, $\Phi_e^{-1}(g) = (X, Y)$ is given by*

$$Y = \psi \sin(\phi) L_2 + \psi \cos(\phi) L_3, \tag{15}$$

*where $\psi = \arccos(g_{33}), \phi = \arctan(-\frac{g_{31}}{g_{32}})$. Let $R = g \exp(Y)^\top = [R_{ij}]$, then $X = aL_1, a = \arctan(\frac{R_{21}}{R_{11}})$.*

Using the inverse of $\Phi_e$ in Theorem 5.11, the exponential-coordinate kernel on $U_e$ is provided by Theorem 4.6.

In fact, by the explicit choice of $r_p$ and $\epsilon_p$, the local chart $U_p$ is sufficiently large to cover $\mathbf{SO}(3)$ almost everywhere. Consequently, $k_p$ serves as a virtually global positive definite kernel on $\mathbf{SO}(3)$ (see Appendix C.4 for a complete analysis).

### 5.2 Local Exponential-Coordinate Kernel on $\mathbf{SU}(2)$

The special unitary group $\mathbf{SU}(2) = \{Q \in \mathbf{GL}_2(\mathbb{C}) | QQ^* = Q^*Q = \mathbf{I}, \det(Q) = 1\}$ is a compact, connected, simple (hence semisimple) real Lie group of dimension 3. Here, $Q^* = \bar{Q}^\top$ is the conjugate transpose of $Q$. From the definition of $\mathbf{SU}(2)$, we can obtain the concrete form of matrix in $\mathbf{SU}(2)$:

$$Q = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix}, |\alpha|^2 + |\beta|^2 = 1. \tag{16}$$

The Lie algebra of $\mathbf{SU}(2)$ is the set of all skew-Hermitian matrices with trace zero: $\mathfrak{su}(2) = \{X \in \mathfrak{gl}_2(\mathbb{C}) | X + X^* = 0, \mathrm{tr}(X) = 0\}$. Its Killing form is given by $B(X, Y) = 4\mathrm{tr}(XY)$. A direct computation yields the explicit form of matrix in $\mathfrak{su}(2)$: $X = \begin{pmatrix} ia & -\bar{z} \\ z & -ia \end{pmatrix}, a \in \mathbb{R}, z \in \mathbb{C}$.

By Theorem 5.1, $-\frac{1}{4}B$ is an inner product on Lie algebra $\mathfrak{su}(2)$ and can be extended to a bi-invariant metric on Lie group $\mathbf{SU}(2)$ by its adjoint invariant property. The construction for $\mathbf{SU}(2)$ follows the same procedure as for $\mathbf{SO}(3)$: the local chart is provided by the Cartan subalgebra and its orthogonal complement, and the local Gaussian kernel is determined once the inverse map $\Phi_e^{-1}$ is derived. We therefore only present the key intermediate results and the final closed-form expressions.

**Theorem 5.12.** *The set of all diagonal elements* $T = \{\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} | \theta \in [0, 2\pi)\}$ *in* $\mathbf{SU}(2)$ *forms a maximal torus.*

**Theorem 5.13.** *The Cartan subalgebra of* $T$ *in Theorem 5.12 is the set of all diagonal elements in* $\mathfrak{su}(2)$*:*

$$\mathfrak{t} = \{\begin{pmatrix} ia & 0 \\ 0 & -ia \end{pmatrix} | a \in \mathbb{R}\}. \tag{17}$$

**Theorem 5.14.** *The orthogonal complement of* $\mathfrak{t}$ *in Theorem 5.13 is given by*

$$\mathfrak{t}^\perp = \{\begin{pmatrix} 0 & -b + ic \\ b + ic & 0 \end{pmatrix}, b, c \in \mathbb{R}\}. \tag{18}$$

**Proposition 5.15.** $(\mathbf{SU}(2), -\frac{1}{4}B)$ *has a constant sectional curvature equal to* $\frac{1}{2}$.

**Theorem 5.16.** *The injectivity radius* $r$ *of* $\mathbf{SU}(2)$ *is* $\sqrt{2}\pi$.

Following the same approach as for $\mathbf{SO}(3)$, we now construct the neighborhood $U_e$ of identity element in Lie group $\mathbf{SU}(2)$ (Theorem 4.4). For this purpose, we employ the Rodrigues formula for $\mathbf{SU}(2)$ (please refer to Appendix B.4.1 for detailed derivation), which provides the explicit formula for the exponential map from $\mathfrak{su}(2)$ to $\mathbf{SU}(2)$. Let $r_p = \epsilon_p = r = \sqrt{2}\pi$, then the next lemma gives the explicit form of open balls $B_{\mathfrak{t}}(0, r)$ and $B_{\mathfrak{t}^\perp}(0, r)$.

**Lemma 5.17.** $B_{\mathfrak{t}}(0, r) = \{X = aX_1 | |a| < \pi, a \in \mathbb{R}\}$, $B_{\mathfrak{t}^\perp}(0, r) = \{Y = bX_2 + cX_3 | b^2 + c^2 < \pi^2, b, c \in \mathbb{R}\}$.

The notations $X_1, X_2, X_3$ is a basis for $\mathfrak{su}(2)$ explained in Appendix B.21.

**Theorem 5.18.** *The neighborhood* $U_e \subset \mathbf{SU}(2)$ *in Theorem 4.4 is given by the following form:*

$$U_e = \left\{ \begin{pmatrix} \cos(\varphi)e^{ia} & -\sin(\varphi)e^{i(a-\theta)} \\ \sin(\varphi)e^{-i(a-\theta)} & \cos(\varphi)e^{-ia} \end{pmatrix} \middle| \begin{matrix} |a|<\pi,\ 0<\varphi<\pi, \\ 0<\theta<2\pi \end{matrix} \right\}. \tag{19}$$

**Theorem 5.19.** *For any* $g = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} \in U_e$*, the inverse* $\Phi_e^{-1}(g) = (X, Y)$ *is given by*

$$X = aX_1, a = \arg(\alpha), \text{ and } Y = \varphi\cos(\theta)X_2 + \varphi\sin(\theta)X_3, \tag{20}$$

*where* $\varphi = \arccos(|\alpha|)$, $\theta = \arg(\alpha\beta)$. *Here* $\arg \in [-\pi, \pi)$ *denotes the principal argument of a complex number* $z$.

The similar global coverage result of $U_p$ also holds for the Lie group $\mathbf{SU}(2)$. Consequently, $k_p$ also serves as a virtually global positive-definite kernel on $\mathbf{SU}(2)$ (see Appendix C.5 for details).

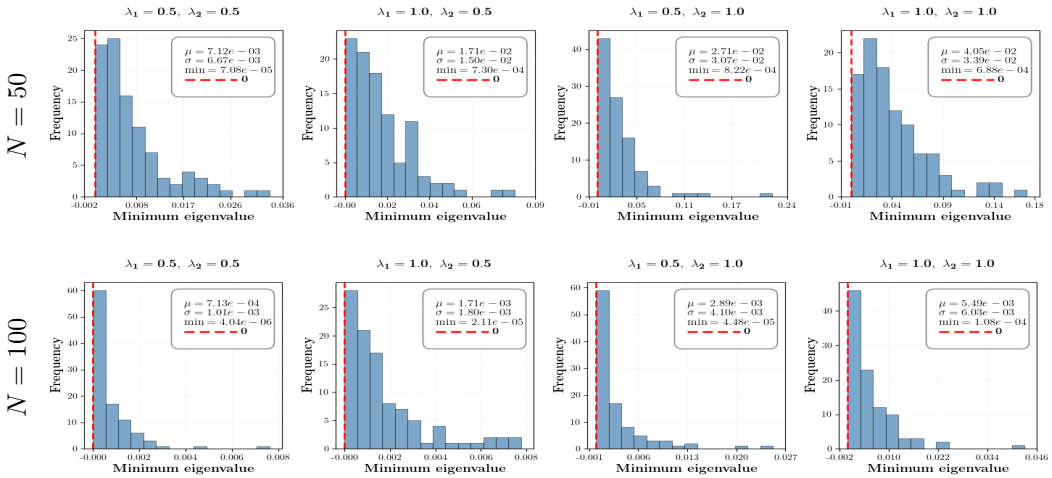



Figure 1: Minimum eigenvalue distribution on $\mathbf{SO}(3)$ with sample sizes $N = 50$ (top) and $N = 100$ (bottom). Each histogram shows 100 independent trials. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

## 6 Experimental Results and Discussion

In this section, we present numerical experiments to demonstrate the positive definiteness of the constructed exponential-coordinate kernel, and then apply it to Gaussian process regression tasks. More experimental details, results, and comparison are provided in Appendix C. All experiments are conducted on a server equipped with an Intel Xeon Gold 6326 CPU @ 2.90GHz and an NVIDIA GeForce RTX 3090 GPU (24 GB).

### 6.1 Positive Definiteness Verification on $\mathbf{SO}(3)$

We empirically verify the positive semi-definiteness of our local exponential-coordinate kernel on $\mathbf{SO}(3)$. For four representative weight pairs $(\lambda_1, \lambda_2) = (0.5, 0.5), (1.0, 0.5), (0.5, 1.0), (1.0, 1.0)$, we uniformly sample $N$ points from the neighborhood $U_e$, construct the kernel matrix, and compute its minimum eigenvalue. This procedure is repeated for 100 independent trials to obtain statistical significance. Fig. 1 shows the distribution of minimum eigenvalues for $N = 50$ (top) and $N = 100$ (bottom). Each subfigure displays the mean $\mu$, standard deviation $\sigma$, and the minimum $\min$ of the minimum eigenvalues over 100 independent trials. The red dashed line marks zero. As observed, all minimum eigenvalues are strictly positive across all parameter combinations and sample sizes, providing strong empirical evidence that our kernel function is positive definite within $U_e$, consistent with Theorem 4.6. Additional experiments, including the influence of data size, hyperparameter pairs, and verification on neighborhoods of arbitrary points, are provided in Appendix C.2.

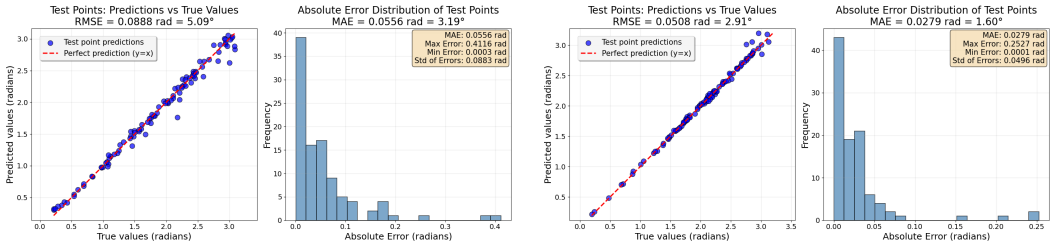
### 6.2 Gaussian Process Regression on $\mathbf{SO}(3)$

Subsequently, we perform Gaussian process regression on the neighborhood $U_e \subset \mathbf{SO}(3)$ using our local exponential-coordinate kernel. The target function is chosen as the rotation angle:

$$f : \mathbf{SO}(3) \to \mathbb{R}, \quad f(\mathbf{R}) = \arccos\left(\frac{\mathrm{tr}(\mathbf{R}) - 1}{2}\right). \tag{21}$$

We sample $n$ rotation matrices $\mathbf{R}_1, \ldots, \mathbf{R}_n$ from $U_e$ as training inputs, with corresponding noisy observations $y_i = f(\mathbf{R}_i) + \epsilon_i$, where $\epsilon_i$ denotes independent and identically distributed Gaussian noise. An additional set of $m$ matrices $\mathbf{R}_1^*, \ldots, \mathbf{R}_m^*$ are sampled from $U_e$ as test inputs. The Gaussian process model is trained on the dataset $\{(\mathbf{R}_i, y_i)\}_{i=1}^n$ and then used to predict the function values at the test points. The prediction errors are evaluated by comparing the predicted values $y_j$ with the ground truth $y_j^* = f(\mathbf{R}_j^*)$ using two metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

In our experiment configuration, we set the noise $\epsilon_i = 0.02$, hyperparameter pair $\lambda_1 = \lambda_2 = 0.1$, and the size of test points $m = 100$. The results for $n = 100$ and $n = 200$ are shown in Fig. 2a




(a) $n = 100$ (b) $n = 200$

Figure 2: Gaussian process regression on $U_e \subset \mathbf{SO}(3)$ with $n = 100$ and $n = 200$ training points, respectively. The strong alignment between predicted and true values demonstrates the effectiveness of our constructed local Gaussian kernel.

and Fig. 2b, respectively. In both cases, the predicted values cluster tightly around the diagonal, confirming that the kernel faithfully recovers the target function. As expected, the prediction accuracy improves with larger training sets: the MAE decreases from $3.19°$ for $n = 100$ to $1.60°$ for $n = 200$. Moreover, we benchmark our kernel against Matérn kernels on compact Riemannian manifolds [24] and on compact Lie groups [32]. Our kernel achieves both higher predictive accuracy and significantly higher computational efficiency. For more experiment details (including regression on $\mathbf{SU}(2)$), please refer to Appendix C.3 and C.4.

### 6.3 Robotics Denoising Experiment

We further validate our method on a robotics task using the PyBullet simulator, which is widely employed for trajectory generation and control. We generate 1,000 rotation matrices from the end-effector trajectory of a simulated Franka Panda manipulator. The task is to recover the clean rotation angle $f(R) = \arccos((\mathrm{tr}(R) - 1)/2)$ from a noisy rotation matrix, where additive isotropic Gaussian noise is applied to the rotation matrix to simulate realistic sensor noise. We use 500 points for training and 100 for testing, and report RMSE and MAE against the Euclidean RBF baseline. As shown in Table 1, our exponential-coordinate kernel consistently outperforms the RBF baseline across all noise levels. This improvement stems from the fact that our kernel respects the intrinsic geometry of $\mathbf{SO}(3)$ by operating in the Lie algebra coordinates, whereas the RBF baseline flattens the rotation matrix into a 9-dimensional Euclidean vector, which ignores the underlying manifold structure.

| Noise | Method | MAE (°) | RMSE (°) |
|---|---|---|---|
| 0.02 | RBF | 0.74 | 0.95 |
| | Ours | **0.66** | **0.86** |
| 0.05 | RBF | 1.62 | 1.78 |
| | Ours | **1.25** | **1.39** |
| 0.08 | RBF | 2.02 | 2.58 |
| | Ours | **1.88** | **2.48** |
| 0.10 | RBF | 2.39 | 3.17 |
| | Ours | **2.31** | **3.07** |

Table 1: Robotics denoising results on $\mathbf{SO}(3)$.

## 7 Conclusions

We presented a novel and fully constructive method for defining a local exponential-coordinate kernel on compact Lie groups, which enables the description of symmetries naturally. Our key contribution is to exploit the decomposition of the Lie algebra, factorizing the kernel into a local Gaussian kernel on the flat torus, defined via the geodesic distance within a neighborhood determined by the injectivity radius, and a standard Gaussian kernel on the Euclidean component. The product of these two parts is provably positive definite. For two practically important Lie groups $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$, we derive closed-form solutions for the kernel using the Rodrigues formula. Extensive experiments confirm the positive definiteness of the proposed kernels under various settings. Gaussian process regression using these kernels achieves accurate predictions both on local charts and across the entire manifold, demonstrating their practical effectiveness for probabilistic modeling on compact Lie groups.

## Acknowledgments

The author would like to deeply thank the anonymous referees who provided helpful and constructive comments. This work was supported by the National Natural Science Foundation for Young Scholars (Type A) (12525114), the National Natural Science Foundation of China (62602661), and the Key Project of the National Natural Science Foundation of China (12494550 and 12494551).

## References

[1] Matthias Seeger. Gaussian processes for machine learning. *International Journal of Neural Systems*, 14(02):69–106, 2004. 1

[2] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Machine learning of linear differential equations using Gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017. 1

[3] Victor Alves, Vitor Gazzaneo, and Fernando V Lima. A machine learning-based process operability framework using Gaussian processes. *Computers & Chemical Engineering*, 163:107835, 2022. 1

[4] Alan E Gelfand and Erin M Schliep. Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, 18:86–104, 2016. 1

[5] James Hensman, Nicolo Fusi, and Neil D Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013. 1

[6] Juš Kocijan and Alexandra Grancharova. Application of Gaussian processes to the modelling and control in process engineering. *Innovations in Intelligent Machines-5: Computational Intelligence in Control Systems Engineering*, pages 155–190, 2014. 1

[7] Adrian Dudek and Jerzy Baranowski. Gaussian processes for signal processing and representation in control engineering. *Applied Sciences*, 12(10):4946, 2022. 1

[8] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012. 1

[9] Hugh Salimbeni and Marc Deisenroth. Doubly stochastic variational inference for deep Gaussian processes. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 1

[10] Silvia Sellán and Alec Jacobson. Stochastic poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 41(6):1–12, 2022. 1

[11] Marc Peter Deisenroth, Dieter Fox, and Carl Edward Rasmussen. Gaussian processes for data-efficient learning in robotics and control. *IEEE transactions on pattern analysis and machine intelligence*, 37(2):408–423, 2013. 1

[12] Marc Deisenroth and Carl E Rasmussen. Pilco: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on machine learning (ICML-11)*, pages 465–472, 2011. 1

[13] Sylvain Calinon and Noémie Jaquier. Gaussians on Riemannian manifolds for robot learning and adaptive control. *arXiv preprint arXiv:1909.05946*, 2019. 1

[14] Noémie Jaquier, Leonel Rozo, Sylvain Calinon, and Mathias Bürger. Bayesian optimization meets Riemannian manifolds in robot learning. In *Conference on Robot Learning*, pages 233–246. PMLR, 2020. 1

[15] Sadeep Jayasumana, Richard Hartley, and Mathieu Salzmann. Kernels on Riemannian manifolds. In *Riemannian computing in computer vision*, pages 45–67. Springer, 2016. 1

[16] Pavan Turaga, Ashok Veeraraghavan, and Rama Chellappa. Statistical analysis on stiefel and grassmann manifolds with applications in computer vision. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008. 1

[17] Alexander Smith, Benjamin Laubach, Ivan Castillo, and Victor M Zavala. Data analysis using Riemannian geometry and applications to chemical engineering. *Computers & Chemical Engineering*, 168:108023, 2022. 1

[18] Dirk Lebiedz, Volkmar Reinhardt, and Jochen Siehr. Minimal curvature trajectories: Riemannian geometry concepts for slow manifold computation in chemical kinetics. *Journal of Computational Physics*, 229(18):6512–6533, 2010. 1

[19] Lizhen Lin, Niu Mu, Pokman Cheung, and David Dunson. Extrinsic Gaussian processes for regression and classification on manifolds. *Bayesian Analysis*, 14(3):887–906, 2019. 1

[20] Mu Niu, Zhenwen Dai, Pokman Cheung, and Yizhu Wang. Intrinsic Gaussian process on unknown manifolds with probabilistic metrics. *Journal of Machine Learning Research*, 24(104):1–42, 2023. 1

[21] Aasa Feragen, Francois Lauze, and Soren Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3032–3042, 2015. 1, 3

[22] Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–80, 2013. 2

[23] Pengfei Fang, Mehrtash Harandi, and Lars Petersson. Kernel methods in hyperbolic spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10665–10674, October 2021. 2

[24] Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Deisenroth (he/him). Matérn Gaussian processes on Riemannian manifolds. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12426–12437. Curran Associates, Inc., 2020. 2, 10, 33, 34, 36, 37

[25] Nathael Da Costa, Cyrus Mostajeran, Juan-Pablo Ortega, and Salem Said. Geometric learning with positively decomposable kernels. *Journal of Machine Learning Research*, 25(326):1–42, 2024. 2

[26] Aasa Feragen and Søren Hauberg. Open problem: Kernel methods on manifolds and metric spaces. what is the probability of a positive definite geodesic exponential kernel? In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1647–1650, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR. 2

[27] Nathaël Da Costa, Cyrus Mostajeran, and Juan-Pablo Ortega. The Gaussian kernel on the circle and spaces that admit isometric embeddings of the circle. In *International Conference on Geometric Science of Information*, pages 426–435. Springer, 2023. 2

[28] Annika Lang and Christoph Schwab. Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *The Annals of Applied Probability*, pages 3047–3094, 2015. 2

[29] Jaehong Jeong, Mikyoung Jun, and Marc G Genton. Spherical process models for global spatial statistics. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 32(4):501, 2017. 2

[30] Michael Hutchinson, Alexander Terenin, Viacheslav Borovitskiy, So Takao, Yee Teh, and Marc Deisenroth. Vector-valued Gaussian processes on Riemannian manifolds via gauge independent projected kernels. *Advances in Neural Information Processing Systems*, 34:17160–17169, 2021. 2

[31] Noémie Jaquier, Viacheslav Borovitskiy, Andrei Smolensky, Alexander Terenin, Tamim Asfour, and Leonel Rozo. Geometry-aware bayesian optimization in robotics using Riemannian matérn kernels. In *Conference on Robot Learning*, pages 794–805, 2022. 2

[32] Iskander Azangulov, Andrei Smolensky, Alexander Terenin, and Viacheslav Borovitskiy. Stationary kernels and Gaussian processes on Lie groups and their homogeneous spaces i: the compact case. *Journal of Machine Learning Research*, 25(280):1–52, 2024. 2, 10, 33, 35, 36, 37

[33] James E Humphreys. *Introduction to Lie algebras and representation theory*. Springer Science & Business Media, 2012. 2

[34] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006. 4, 14

[35] John Frank Adams. *Lectures on Lie groups*. University of Chicago Press, 1982. 4

[36] P Petersen. Riemannian geometry. *Graduate Texts in Mathematics/Springer-Verlarg*, 2006. 4, 7

[37] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, volume 222 of *Graduate Texts in Mathematics*. Springer, Cham, 2015. 5, 19

[38] Marcos M Alexandrino and Renato G Bettiol. Lie groups with bi-invariant metrics. In *Lie groups and geometric aspects of isometric actions*, pages 27–47. Springer, 2015. 6, 15, 18

[39] J Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911. 17

[40] John Willard Milnor and David W Weaver. *Topology from the differentiable viewpoint*, volume 21. Princeton university press, 1997. 31

# A Background on Differential Geometry and Lie Theory

## A.1 Riemannian Manifolds and Curvature

**Definition A.1** (Riemannian manifold)**.** A Riemannian manifold $\mathcal{M}$ is a smooth manifold endowed with a Riemannian metric $g$ which assigns each point $p \in \mathcal{M}$ an inner product in a smooth way.

A Riemannian metric on manifold provides a way to measure the length of a curve on manifold. Specifically, let $\gamma : [0, 1] \to \mathcal{M}$ be a curve on manifold, its length is given by $L(\gamma) = \int_0^1 \langle \gamma'(t), \gamma'(t) \rangle^{\frac{1}{2}} dt$. This gives rise to a distance on manifold called the *geodesic distance*.

**Definition A.2** (Geodesic distance)**.** Let $(\mathcal{M}, g)$ be a Riemannian manifold. Then $d_g(p, q) = \inf\{L(\gamma) \mid \gamma : [0, 1] \to \mathcal{M} \text{ is a piecewise smooth curve joining } p \text{ and } q\}$ is a metric on $\mathcal{M}$ called the geodesic distance associated with the Riemannian metric $g$.

**Definition A.3** (Isometry)**.** Let $\mathcal{M}$ and $\mathcal{M}'$ be two Riemannian manifolds and let $f : \mathcal{M} \to \mathcal{M}'$ be a diffeomorphism. Then $f$ is called an isometry if $\langle df_p(X), df_p(Y) \rangle_{f(p)} = \langle X, Y \rangle_p, \forall X, Y \in T_p^{\mathcal{M}}$.

All isometries from $\mathcal{M}$ to itself forms a group under composition called its *isometry group* denoted by $\mathrm{Iso}(\mathcal{M})$. There is a natural group action from $\mathrm{Iso}(\mathcal{M})$ to manifold $\mathcal{M}$. A Riemannian manifold is said to be *homogeneous* if its isometry group acts *transitively*, i.e. for each pair of points $p, q \in \mathcal{M}$, there is an isometry $F \in \mathrm{Iso}(\mathcal{M})$ such that $F(p) = q$.

An essential concept in Riemannian geometry is *curvature*. Let $(\mathcal{M}, g)$ be an $n$-dimensional Riemannian manifold equipped with its *Levi-Civita connection* $\nabla$. A *Riemannian curvature tensor* is defined as the $(1, 3)$-tensor field given by the following expression, for all vector fields $X, Y, Z, W \in \mathfrak{X}(\mathcal{M})$:

$$R(X, Y)Z = \nabla_{[X,Y]}Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z \tag{22}$$

We can further define a $(0, 4)$-tensor using the metric:

$$R(X, Y, Z, W) = g(R(X, Y)Z, W). \tag{23}$$

Using the curvature tensor, we can define the *sectional curvature* of the plane spanned by two linearly independent tangent vectors $X$ and $Y$ as:

$$\sec(X,Y) = \frac{R(X,Y,X,Y)}{g(X,X)g(Y,Y) - g(X,Y)^2}. \tag{24}$$

It's worth noting that the sectional curvature only depends on the plane, not specific on the choice of the basis $X$ and $Y$. An important property of Riemannian curvature tensor is its invariance under isometry.

**Lemma A.4.** *[34] Let $F : \mathcal{M} \to \mathcal{N}$ be an isometry between two Riemannian manifolds and $F_*$ its pushforward operator. Let $R_\mathcal{M}$ and $R_\mathcal{N}$ be two Riemannian curvature tensors on $\mathcal{M}$ and $\mathcal{N}$, respectively. Then the following holds:*

$$F_*(R_\mathcal{M}(X,Y)Z) = R_\mathcal{N}(F_*X, F_*Y)F_*Z, \forall X, Y, Z \in \mathfrak{X}(M). \tag{25}$$

**Definition A.5.** A Riemannian manifold is said to be flat if its Riemannian curvature tensor is everywhere zero.

For each point $p \in \mathcal{M}$, the Riemannian curvature $R$ gives rise to a trilinear map: $R_p : T_p^\mathcal{M} \times T_p^\mathcal{M} \times T_p^\mathcal{M} \to T_p^\mathcal{M}$. Then we define the map: $\mathrm{Ric}_p : T_p^\mathcal{M} \times T_p^\mathcal{M} \to \mathbb{R}$ by

$$\mathrm{Ric}_p(Y,Z) = \mathrm{tr}(X \to R_p(X,Y)Z). \tag{26}$$

As $p$ varies over $\mathcal{M}$, this defines a $(0,2)$-tensor field Ric, known as the *Ricci tensor*.

**Definition A.6.** A Riemannian manifold $(\mathcal{M}, g)$ is called an Einstein manifold if its Ricci tensor is proportional to its metric tensor.

**Lemma A.7.** *[34] Every 3-dimensional Einstein manifold has constant sectional curvature.*

### A.2 Lie Groups and Lie Algebras

**Definition A.8** (Lie group)**.** A real Lie group $\mathcal{G}$ is a group that is also a real smooth manifold, in which the group operations of multiplication and inversion are smooth maps

For any $g \in \mathcal{G}$, we define the left (resp. right) translation by $\mathrm{L}_g : \mathcal{G} \to \mathcal{G}$, $\mathrm{L}_g(h) = gh$ and $\mathrm{R}_g : \mathcal{G} \to \mathcal{G}$, $\mathrm{R}_g(h) = hg$. They are obviously both diffeomorphisms. Given a Riemannian metric on $\mathcal{G}$, if $\mathrm{L}_g$ (resp. $\mathrm{R}_g$) is an isometry for all $g \in \mathcal{G}$, then it is called *left (resp. right)-invariant*. If a Riemannian metric is both left-invariant and right-invariant, then it's called *bi-invariant*.

**Proposition A.9.** *Every compact Lie group admits a bi-invariant metric $Q$.*

**Definition A.10** (Lie algebra)**.** A Lie algebra $\mathfrak{g}$ is a real vector space endowed with a bilinear operator (called a Lie bracket):

$$[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g},$$

such that for all $x, y, z \in \mathfrak{g}$,

(a) $[x,y] = -[y,x]$, (anti-commutativity)

(b) $[[x,y],z] + [[y,z],x] + [[z,x],y] = 0.$ (Jacobi identity)

Let $\mathcal{G}$ be a Lie group and $\mathfrak{g}$ its Lie algebra. An inner product $\langle .,. \rangle$ on $\mathfrak{g}$ can be naturally extended to a left-invariant metric on its Lie group $\mathcal{G}$. Since $\mathfrak{g}$ is identified with the tangent space of $\mathcal{G}$ at its identity $T_e^\mathcal{G}$, for any $g \in \mathcal{G}$, we define the inner product on $T_g^\mathcal{G}$ by pulling back the elements via the differential of the left translation:

$$\langle X, Y \rangle_g := \langle d(L_g)_e^{-1}(X), d(L_g)_e^{-1}(Y) \rangle. \tag{27}$$

By construction, this metric is actually left-invariant.

### A.3 Killing Form and Bi-invariant Metric

**Definition A.11.** Let $\mathfrak{g}$ be a Lie algebra. For any $x \in \mathfrak{g}$, the *adjoint endomorphism* $\mathrm{ad}_x : \mathfrak{g} \to \mathfrak{g}$ is defined as $\mathrm{ad}_x(y) = [x,y]$

**Definition A.12** (Killing form)**.** Suppose $\mathfrak{g}$ is finite dimensional. $B : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$ defined by $B(x, y) = \mathrm{tr}(\mathrm{ad}_x \circ \mathrm{ad}_y)$ is a symmetric bilinear form called the Killing form on $\mathfrak{g}$. The Killing form of a Lie group is defined as the Killing form of its Lie algebra.

**Proposition A.13.** *The Killing form is automorphism invariant:* $B(s(x), s(y)) = B(x, y)$, $\forall s \in \mathrm{Aut}(\mathfrak{g})$.

Let $\mathcal{G}$ be a Lie group and $\mathfrak{g}$ its Lie algebra. For each $g \in \mathcal{G}$, let $\Phi_g : \mathcal{G} \to \mathcal{G}$ be the conjugation: $\Phi_g(h) = ghg^{-1}$. We define $\mathrm{Ad}_g$ to be the derivative of $\Phi_g$ at the identity: $\mathrm{Ad}_g := d(\Phi_g)_e : \mathfrak{g} \to \mathfrak{g}$. $\mathrm{Ad}_g$ is a Lie algebra automorphism. The map $\mathrm{Ad} : \mathcal{G} \to \mathrm{Aut}(\mathfrak{g}), g \to \mathrm{Ad}_g$ is a group homomorphism called the *adjoint representation* of $\mathcal{G}$. An inner product $\langle , \rangle$ on $\mathfrak{g}$ is called *adjoint invariant* if $\langle \mathrm{Ad}_g(x), \mathrm{Ad}_g(y) \rangle = \langle x, y \rangle$, for all $g \in \mathcal{G}$, $x, y \in \mathfrak{g}$.

**Proposition A.14.** *Let* $\langle ., . \rangle$ *be an adjoint invariant inner product on Lie algebra* $\mathfrak{g}$*, then the metric given by Eq.* (27) *is bi-invariant.*

*Proof.* It suffices to verify that the defined metric is right-invariant. Without loss of generality, the right-invariance can be summarized to the identity element:

$$\langle X, Y \rangle = \langle d(R_g)_e(X), d(R_g)_e(Y) \rangle_g, \forall g \in \mathcal{G}, X, Y \in \mathfrak{g}. \tag{28}$$

Observe that the conjugation map can be written as $\Phi_g = R_{g^{-1}} \circ L_g$, differentiating at the identity and applying the chain rule yields

$$d(\Phi_g)_e = d(R_{g^{-1}})_g \circ d(L_g)_e. \tag{29}$$

Since $d(L_{g^{-1}})_g$ is the inverse of $d(L_g)_e$, this identity is equivalent to $\mathrm{Ad}_g \circ d(L_{g^{-1}})_g = d(R_{g^{-1}})_g$. By the adjoint invariance and left-invariant property of the metric, for each $X, Y \in T_g^{\mathcal{G}}$ we have:

$$\begin{aligned} \langle d(R_{g^{-1}})_g(X), d(R_{g^{-1}})_g(Y) \rangle &= \langle d(L_{g^{-1}})_g(X), d(L_{g^{-1}})_g(Y) \rangle \\ &= \langle X, Y \rangle_g \end{aligned} \tag{30}$$

Replacing $X, Y$ in Eq. (30) by $d(R_g)_e(X)$ and $d(R_g)_e(Y)$ in Eq. (28), we finish the proof. □

### A.4 Totally Geodesic Submanifolds

**Definition A.15.** A submanifold $\mathcal{N} \subset (\mathcal{M}, g)$ is called totally geodesic if for each $p \in \mathcal{N}$ a neighborhood of $0 \in T_p^{\mathcal{N}} \subset T_p^{\mathcal{M}}$ is mapped into $\mathcal{N}$ via the exponential map $\exp_p$ for $\mathcal{M}$.

The following classical result characterizes totally geodesic subgroups in Lie groups with bi-invariant metrics.

**Lemma A.16.** *Let* $\mathcal{G}$ *be a compact Lie group endowed with a bi-invariant metric. Then the subgroup* $\mathcal{H} \subset \mathcal{G}$ *is totally geodesic if and only if* $H$ *is closed in* $\mathcal{G}$*.*

As a direct result of Lemma A.16, the maximal torus $T \in \mathcal{G}$ is totally geodesic since it's a compact subgroup. The following lemma relates the injectivity radius of a manifold and its totally geodesic submanifold.

**Lemma A.17.** *Let* $\mathcal{N} \subset (\mathcal{M}, g)$ *be a totally geodesic submanifold. Then for each point* $p \in \mathcal{N}$*,* $\mathrm{Inj}_p(\mathcal{N}) \geq \mathrm{Inj}_p(\mathcal{M})$*, where* $\mathrm{Inj}_p(\mathcal{N})$ *and* $\mathrm{Inj}_p(\mathcal{M})$ *denote the injectivity radii of* $\mathcal{N}$ *and* $\mathcal{M}$ *at* $p$*, respectively.*

## B Technical Lemmas and Auxiliary Proofs

### B.1 Proofs of Section 3

**Proposition B.1.** *[38] Let* $\mathcal{G}$ *be a Lie group endowed with a bi-invariant metric. Then the following hold:*

$$R(X, Y)Z = \frac{1}{4}[[X, Y], Z], \forall X, Y, Z \in \mathfrak{g}, \tag{31}$$

$$R(X, Y, X, Y) = \frac{1}{4}||[X, Y]||^2. \tag{32}$$

**Corollary B.2.** *Let $\mathcal{G}$ be a compact connected abelian Lie group with a bi-invariant metric. Then $\mathcal{G}$ is flat.*

*Proof.* Since $\mathcal{G}$ is abelian, its Lie algebra $\mathfrak{g}$ is abelian, and thus by Proposition B.1, the Riemannian curvature tensor is zero on $\mathfrak{g}$. Because the metric is bi-invariant, the left translation $L_g$ is an isometry for each point $g \in \mathcal{G}$. Applying Lemma A.4, we conclude that the Riemannain curvature tensor vanishes identically on $\mathcal{G}$, hence $\mathcal{G}$ is flat (Definition A.5). □

**Lemma B.3.** *The injectivity radius is a constant on any homogeneous Riemannian manifold.*

*Proof.* For any $p, q \in \mathcal{M}$, there exists an isometry $\varphi : \mathcal{M} \to \mathcal{M}$ such that $\varphi(p) = q$ since $\mathcal{M}$ is homogeneous. Because $\varphi$ is an isometry, it satisfies $\exp_q \circ d\varphi_p = \varphi \circ \exp_p$. Hence, $\exp_p$ is injective if and only if $\exp_q$ is injective. By the definition of injectivity radius, it follows that $\mathrm{Inj}(p) = \mathrm{Inj}(q)$. This finishes the proof. □

**Theorem B.4.** *For any $p \in \mathbb{T}^n$, the injectivity radius satisfies:*

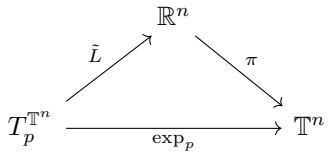
$$\mathrm{Inj}(p) = \frac{1}{2} \min_{v \in \Gamma \setminus \{0\}} ||v|| \tag{33}$$

*Proof.* Since $\pi$ is a local isometry, the geodesics in $\mathbb{T}^n$ are the projection of straight lines in $\mathbb{R}^n$. Consequently, a geodesic loop at $p$ takes the form: $\gamma(t) = \pi(x + ty)$, where $x, y \in \mathbb{R}^n$. $\gamma(0) = p$ means that $x \in \pi^{-1}(p)$ and $\gamma(1) = p$ is equivalent to $y \in \Gamma$.

$$\begin{aligned} L(\gamma) &= \int_0^1 \langle \gamma'(t), \gamma'(t) \rangle^{\frac{1}{2}} dt \\ &= \int_0^1 \langle d\pi(y), d\pi(y) \rangle^{\frac{1}{2}} dt \\ &= \int_0^1 ||y|| dt = ||y||. \end{aligned} \tag{34}$$

By Lemma 3.6, we derive the final result. □

**Lemma B.5.** *Let $\tilde{p}$ be a lifting of $p$ in $\mathbb{R}^n$: $\pi(\tilde{p}) = p$. Since $\pi$ is a local isometry, $d\pi_{\tilde{p}} : \mathbb{R}^n \to T_p^{\mathbb{T}^n}$ is an isomorphism. Let $L$ be the inverse of the differential $d\pi_{\tilde{p}}^{-1}$ and $\tilde{L} : T_p^{\mathbb{T}^n} \to \mathbb{R}^n$ be the translation of $\tilde{p}$ under $L(v)$: $\tilde{L}(v) = \tilde{p} + L(v)$. Then the following diagram commutes:*




*Specifically, $\pi \circ \tilde{L} = \exp_p$.*

*Proof.* This can be checked easily. We consider the straight line $l(t) = \tilde{p} + tL(v)$ in $\mathbb{R}^n$. Since $\pi$ is a covering map, $\gamma(t) = \pi(l(t))$ is a geodesic in $\mathbb{T}^n$. It's easily to verify that $\gamma(0) = p$ and $\gamma'(0) = d\pi_{\tilde{p}}(L(v)) = v$ by the definition of $L$. By the uniqueness of geodesic, we can deduce the commutativity: $\exp_p(v) = \gamma(1) = \pi(\tilde{p} + L(v)) = \pi(\tilde{L}(v))$. □

**Lemma B.6.** $\tilde{L}(B_r(0)) = B_r(\tilde{p})$.

*Proof.* For any $v \in B_r(0)$, $\tilde{L}(v) = \tilde{p} + L(v)$, then $||\tilde{L}(v) - \tilde{p}|| = ||L(v)|| = ||v|| < r$, so that $\tilde{L}(v) \in B_r(\tilde{p})$. For any $x \in B_r(\tilde{p})$, there exists a unique $v \in T_p^{\mathbb{T}^n}$, such that $\tilde{L}(v) = x$. It follows that $\tilde{p} + L(v) = x$, so that $||v|| = ||L(v)|| = ||x - \tilde{p}|| < r$, which means that $x \in B_r(0)$. This finshes the proof. □

**Theorem B.7.** $\pi|_{\pi^{-1}(U_p)} : B_r(\tilde{p}) \to U_p$ *is an isometry.*

*Proof.* It suffices to show that $\pi$ is injective on $B_r(\tilde{p})$. For any $x, y \in B_r(\tilde{p})$ satisfying $\pi(x) = \pi(y)$, then we have $x - y \in \Gamma$. If $x - y \neq 0$, then Theorem 3.7 implies that $||x - y|| \geq 2r$. However, $||x - y|| \leq ||x - \tilde{p}|| + ||y - \tilde{p}|| < 2r$. This is a contradiction. Hence, $x = y$. This finishes the proof. □

**Proposition B.8.** *$U_p$ is geodesic convex.*

*Proof.* For any two points $x, y \in U_p$, since $\pi : B_r(\tilde{p}) \to U_p$ is an isometry, there exist unique $x', y' \in B_r(\tilde{p})$ such that $\pi(x') = x, \pi(y') = y$. Since $B_r(\tilde{p})$ is convex, the straight line $l$ connecting $x'$ and $y'$ is contained in $B_r(\tilde{p})$. Since $\pi$ is a covering map, the image of $l$ under $\pi$ is a geodesic in $\mathbb{T}^n$ which is within $U_p$ by Eq. 6. □

### B.2 Proofs of Section 4

**Theorem B.9.** *For any point $p \in \mathcal{G}$. There exists $r_p > 0$ and $\epsilon_p > 0$ such that the map*

$$\begin{aligned} \Phi_p : B_{\mathfrak{t}}(0, r_p) \times B_{\mathfrak{t}^\perp}(0, \epsilon_p) &\to \mathcal{G} \\ \Phi_p(X, Y) &= p \exp(X) \exp(Y) \end{aligned} \tag{35}$$

*is a diffeomorphism onto its image $U_p = \Phi_p(B_{\mathfrak{t}}(0, r_p) \times B_{\mathfrak{t}^\perp}(0, \epsilon_p))$.*

*Proof.* Let's consider the smooth map: $\Phi_p : \mathfrak{t} \times \mathfrak{t}^\perp \to \mathcal{G}$ defined by $\Phi_p(X, Y) = p \exp(X) \exp(Y)$. For any $(X, Y) \in \mathfrak{t} \times \mathfrak{t}^\perp$, $\gamma(t) = (tX, tY)$ is a curve satisfying $\gamma(0) = (0, 0)$, $\gamma'(0) = (X, Y)$. Then

$$\begin{aligned} d(\Phi_p)|_{(0,0)}(X, Y) &= \Phi_p^{'}(\gamma(t))(0) \\ &= L_p^{'}(\exp(tX) \exp(tY))(0) \\ &= d(L_p)_e((\exp(tX) \exp(tY))'(0)) \\ &= d(L_p)_e(X + Y). \end{aligned} \tag{36}$$

For any $g \in T_p^{\mathcal{G}}$, given the diffeomorphism $L_p$ and the direct sum of $\mathfrak{g}$, there exist unique $X \in \mathfrak{t}$ and $Y \in \mathfrak{t}^\perp$ such that $d(L_p)_e(X + Y) = g$. Hence $d(\Phi_p)|_{(0,0)}$ is bijective. By the inverse function theorem, there is a neighborhood of $(0, 0)$ in $\mathfrak{t} \times \mathfrak{t}^\perp$ and a neighborhood $U_p$ of $p$ in $\mathcal{G}$ such that $\Phi_p$ is a diffeomorphism. □

**Proposition B.10.** *Take $r_p \leq r$. Let $U_e$ be the neighborhood of identity element $e$ in $T$ defined in Section 3. Then $\exp_e : B_{\mathfrak{t}}(0, r_p) \to \exp_e(B_{\mathfrak{t}}(0, r_p)) \subset U_e$ is an isometry.*

*Proof.* Since $\exp_e$ is a local isometry under the bi-invariant metric, it's an isometry from $B_{\mathfrak{t}}(0, r')$ to $U_e$, where $r'$ is the injectivity radius of $T$. Because $T$ is a totally geodesic submanifold, by Lemma A.17, we conclude that $r' \geq r$. Consequently, $B_{\mathfrak{t}}(0, r_p) \subset B_{\mathfrak{t}}(0, r')$, and therefore $\exp_e$ is an isometry from $B_{\mathfrak{t}}(0, r_p)$ onto its image in $U_e$. □

**Lemma B.11** (Schur Product Theorem [39])**.** *The Hadamard product of two positive semi-definite matrices is also positive semi-definite.*

**Theorem B.12.** *For any $q, q' \in U_p$, let $\Phi_p^{-1}(q) = (X_p(q), Y_p(q))$ and $\Phi_p^{-1}(q') = (X_p(q'), Y_p(q'))$ be the preimage under the diffeomorphism. Then the Gaussian kernel $k_p(q, q') = \exp(-\lambda_1 ||X_p(q) - X_p(q')||^2 - \lambda_2 ||Y_p(q) - Y_p(q')||^2)$ is positive semi-definite for all $\lambda_1, \lambda_2 > 0$.*

*Proof.* Since $B_{\mathfrak{t}}(0, r_p)$ and $B_{\mathfrak{t}^\perp}(0, r_p)$ are Euclidean, $\exp(-\lambda_1 ||X_p(q) - X_p(q')||^2)$ and $\exp(-\lambda_2 ||Y_p(q) - Y_p(q')||^2)$ are both positive semi-definite. By the Schur Product theorem, their product is also positive semi-definite. □

### B.3 Proofs and Derivations for Lie Group $\mathbf{SO}(3)$

#### B.3.1 Rodrigues formula for $\mathbf{SO}(3)$

Let $\mathbf{k} \in \mathfrak{so}(3)$, $\mathbf{u} \in \mathbb{R}^3$ is the vector generating $\mathbf{k}$: $\mathbf{k} = [\mathbf{u}]_\times$. Specifically, if $\mathbf{u} = [k_x, k_y, k_z]^\top$, then $\mathbf{k} = \begin{pmatrix} 0 & -k_z & k_y \\ k_z & 0 & -k_x \\ -k_y & k_x & 0 \end{pmatrix}$. Let $\mathbf{u} = \theta\mathbf{n}$, $\theta = ||\mathbf{u}||$ is the rotation angle, $\mathbf{n}$ is the unit vector, and $\mathbf{K} = \frac{\mathbf{k}}{\theta}$ is the unit skew-symmetric matrix. Then

$$\exp(\mathbf{k}) = \mathbf{I} + \sin(\theta)\mathbf{K} + (1 - \cos(\theta))\mathbf{K}^2, \tag{37}$$

is the rotation matrix throuth the angle $\theta$ counterclockwise about the unit axis $\mathbf{n}$, denoted by $R_{\mathbf{n}}(\theta)$.

#### B.3.2 Proofs of Section 5.1

**Theorem B.13.** *The subgroup*

$$\mathcal{T} = \left\{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \middle| \theta \in [0, 2\pi) \right\} \tag{38}$$

*consisting of all rotations about the $z$-axis in $\mathbb{R}^3$ is a maximal torus in $\mathbf{SO}(3)$, and its Cartan subalgebra is given by*

$$\mathfrak{t} = \left\{ \begin{pmatrix} 0 & -a & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \middle| a \in \mathbb{R} \right\}. \tag{39}$$

*Proof.* $\mathcal{T}$ is homeomorphic and isomorphic to $\mathbb{S}^1$, thus $\mathcal{T}$ is a compact and connected abelian subgroup. Suppose there is an abelian subgroup $\mathcal{S} \supsetneq \mathcal{T}$, then there exists a rotation $R \in S \setminus \mathcal{T}$ which commutes with all rotations about the $z$-axis. This is impossible. Hence $\mathcal{T}$ is maximal. Its Cartan subalgebra $\mathfrak{t}$ is the Lie algbera of $\mathcal{T}$. Let $\gamma(t) = \left\{ \begin{pmatrix} \cos(\theta(t)) & -\sin(\theta(t)) & 0 \\ \sin(\theta(t)) & \cos(\theta(t)) & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$ be a curve in $T$ with $\theta(0) = 0$. Then $\gamma'(0) = \left\{ \begin{pmatrix} 0 & -\theta'(0) & 0 \\ \theta'(0) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}$. This finishes the proof. □

**Proposition B.14.** *The orthogonal complement of $\mathfrak{t}$ in $\mathfrak{so}(3)$ is given by*

$$\mathfrak{t}^\perp = \left\{ \begin{pmatrix} 0 & 0 & b \\ 0 & 0 & -c \\ -b & c & 0 \end{pmatrix} \middle| b, c \in \mathbb{R} \right\}. \tag{40}$$

*Proof.* $L_1 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $L_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$, $L_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ is an orthogonal basis of $\mathfrak{so}(3)$. For any $X \in \mathfrak{t}^\perp$, $X = aL_1 + bL_2 + cL_3$ and satisfies $\langle X, Y \rangle = 0, \forall Y \in \mathfrak{t}$. We can write $Y = xL_1$. This gives $2ax = 0$ for all $x \in \mathbb{R}$, so that $a = 0$. This finishes the proof. □

**Lemma B.15.** *[38] Let $\mathcal{G}$ be a semisimple compact and connected Lie group with Killing form $B$. Then $(\mathcal{G}, -B)$ is an Einstein manifold.*

**Proposition B.16.** *$(\mathbf{SO}(3), -B)$ has constant sectional curvature equal to $\frac{1}{8}$.*

*Proof.* By Lemma B.15 and A.7, $\mathbf{SO}(3)$ has constant sectional curvature. Therefore, it suffices to examine its Lie algebra $\mathfrak{so}(3)$. Following the definition of sectional curvature (Eq. 24) and Proposition B.1, for an orthonormal basis $X, Y$ in a plane of $\mathfrak{so}(3)$, the sectional curvature takes the form: $\sec(X, Y) = \frac{1}{4}||[X, Y]||^2$. Following the proof of Theorem 5.5, we may choose $X = \frac{1}{\sqrt{2}}L_1$, $Y = \frac{1}{\sqrt{2}}L_2$. Since $[X, Y] = -\frac{L_3}{2}$, we obtain $\sec(X, Y) = \frac{1}{8}$. □

**Theorem B.17.** *The injectivity radius $r$ of $\mathbf{SO}(3)$ is $\sqrt{2}\pi$.*

*Proof.* Topologically, $\mathbf{SO}(3) \cong \mathbb{RP}^3 \cong \mathbb{S}^3/\{p, -p\}$ [37]. Therefore, the closed geodesic in $\mathbf{SO}(3)$ is the image of a great semicircular arc in $\mathbb{S}^3$. Since the total length of great circle in $\mathbb{S}^3$ is $4\sqrt{2}\pi$, the length of closed geodesic in $\mathbf{SO}(3)$ is $2\sqrt{2}\pi$. It then follows from the Lemma 5.7 that the injectivity radius is $\text{Inj}(\mathbf{SO}(3)) = \frac{1}{2} \cdot 2\sqrt{2}\pi = \sqrt{2}\pi$. □

**Lemma B.18.** $B_{\mathfrak{t}}(0, r) = \{X = aL_1 || a| < \pi, a \in \mathbb{R}\}$, $B_{\mathfrak{t}^\perp}(0, r) = \{Y = bL_2 + cL_3 | b^2 + c^2 < \pi^2, b, c \in \mathbb{R}\}$.

*Proof.* By Theorem 5.3, any $X \in \mathfrak{t}$ can be written as $X = aL_1$. Its norm is then $||X|| = |a|||L_1|| = \sqrt{2}|a|$. The condition $||X|| < \sqrt{2}\pi$ therefore implies $|a| < \pi$. Similarly, Proposition 5.5 implies that any $Y \in \mathfrak{t}^\perp$ can be expressed as $Y = bL_2 + cL_3$. From the orthogonality of $L_2$ and $L_3$, its norm is $||Y|| = \sqrt{2}\sqrt{b^2 + c^2}$. The requirement $||Y|| < \sqrt{2}\pi$ thus yields $b^2 + c^2 < \pi^2$. □

**Theorem B.19.** *The neighborhood $U_e \subset \mathbf{SO}(3)$ in Theorem 4.4 is given by the composition of two rotations determined by three angles $a, \phi, \psi$:*

$$U_e = \{R_{\mathbf{z}}(a)R_{\mathbf{n}}(\psi) || a| < \pi, 0 < \psi < \pi\}, \tag{41}$$

*where $\mathbf{z} = [0, 0, 1]^\perp$ represents the z-axis and $\mathbf{n} = [0, \sin(\phi), \cos(\phi)]^\perp$,$0 < \phi < 2\pi$.*

*Proof.* For any $X \in B_{\mathfrak{t}}(0, r)$, by Lemma B.18, $X = aL_1$, where $|a| < \pi$. The vector corresponds to $X$ is $\mathbf{u} = [0, 0, a]^\top$, its norm is $\theta = |a|$, the unit skew-symmetric matrix $\mathbf{K} = \pm L_1$. Therefore, by Rodrigues formula we can deduce that

$$\begin{aligned} \exp(X) &= \mathbf{I} \pm \sin(\theta)L_1 + (1 - \cos(\theta))L_1^2 \\ &= \begin{pmatrix} \cos(\theta) & -\sin(\pm\theta) & 0 \\ \sin(\pm\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= R_{\mathbf{z}}(a). \end{aligned} \tag{42}$$

Similarly, for any $Y \in B_{\mathfrak{t}^\perp}(0, r)$, by Lemma B.18, $Y = bL_2 + cL_3$, where $b^2 + c^2 < \pi^2$. The vector corresponds to $Y$ is $\mathbf{v} = [0, b, c]^\top$, its norm $\psi = \sqrt{b^2 + c^2}$ and the unit skew-symmetric matrix $\mathbf{M} = \frac{bL_2 + cL_3}{\sqrt{b^2+c^2}} = \sin(\phi)L_2 + \cos(\phi)L_3$. By Rodrigues formula, it follows that

$$\begin{aligned} \exp(Y) &= \mathbf{I} + \sin(\psi)\mathbf{M} + (1 - \cos(\psi))\mathbf{M}^2 \\ &= \begin{pmatrix} 1-(1-\cos\psi)\sin^2\phi & (1-\cos\psi)\sin\phi\cos\phi & \sin\psi\sin\phi \\ (1-\cos\psi)\sin\phi\cos\phi & 1-(1-\cos\psi)\cos^2\phi & -\sin\psi\cos\phi \\ -\sin\psi\sin\phi & \sin\psi\cos\phi & \cos\psi \end{pmatrix} \\ &= R_{\mathbf{n}}(\psi), \end{aligned} \tag{43}$$

where $\mathbf{n} = \frac{\mathbf{v}}{\psi} = [0, \sin(\phi), \cos(\phi)]^\top$. Then we can obtain the desirable result: $U_e = \{\exp(X)\exp(Y) | X \in B_{\mathfrak{t}}(0, r), Y \in B_{\mathfrak{t}^\perp}(0, r)\} = \{R_{\mathbf{z}}(a)R_{\mathbf{n}}(\psi) || a| < \pi, 0 < \psi < \pi\}$. □

**Theorem B.20.** *For any $g = [g_{ij}] \in U_e$, $\Phi_e^{-1}(g) = (X, Y)$ is given by*

$$Y = \psi \sin(\phi)L_2 + \psi \cos(\phi)L_3, \tag{44}$$

*where $\psi = \arccos(g_{33})$, $\phi = \arctan(-\frac{g_{31}}{g_{32}})$. Let $R = g\exp(Y)^\top = [R_{ij}]$, then*

$$X = aL_1, a = \arctan(\frac{R_{21}}{R_{11}}). \tag{45}$$

*Proof.* Following the proof of Theorem (B.19), The last row of $\exp(X)\exp(Y)$ and coincides with the last row of $\exp(Y)$ given by Eq. (43). Consequently, we obtain

$$-\sin(\psi)\sin(\phi) = g_{31}, \sin(\psi)\cos(\phi) = g_{32}, \cos(\psi) = g_{33}. \tag{46}$$

From these relations, it follows that

$$\phi = \tan 2(-\frac{g_{31}}{g_{32}}), \psi = \arccos(g_{33}). \tag{47}$$

Moreover, the form of $\exp(X)$ yields $\cos(a) = R_{11}, \sin(a) = R_{21}$, so that $a = \tan 2(\frac{R_{21}}{R_{11}})$. □

### B.4 Proofs and Derivations for Lie Group $\mathbf{SU}(2)$

#### B.4.1 Rrodrigues formula for $\mathbf{SU}(2)$

Let $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ be the Pauli matrices: $\sigma^2 = \mathbf{I}$. For any $X = \begin{pmatrix} ia & -b+ic \\ b+ic & -ia \end{pmatrix} \in \mathfrak{su}(2)$, let $\theta = 2\sqrt{a^2+b^2+c^2}$ be the angle, $\mathbf{n} = [n_x, n_y, n_z]^\top = \frac{2[-c,b,-a]^\top}{\theta}$ be the unit vector. Then $X$ can be written in the form:

$$X = -\frac{i\theta}{2}(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z) = -\frac{i\theta}{2}\mathbf{n}\cdot\boldsymbol{\sigma}, \tag{48}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ denotes the triple of involution matrices $\sigma_x, \sigma_y, \sigma_z$. The exponential of $X$ is given by:

$$\exp(X) = \cos(\frac{\theta}{2})\mathbf{I} - i\sin(\frac{\theta}{2})(\mathbf{n}\cdot\boldsymbol{\sigma}). \tag{49}$$

#### B.4.2 Proofs of Section 5.2

**Theorem B.21.** *The orthogonal complement of* $\mathfrak{t}$ *in Theorem 5.13 is given by*

$$\mathfrak{t}^\perp = \{\begin{pmatrix} 0 & -b+ic \\ b+ic & 0 \end{pmatrix}, b, c \in \mathbb{R}\}. \tag{50}$$

*Proof.* Let $X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, X_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$. This is an orthogonal basis of $\mathfrak{su}(2)$ with $\langle X_i, X_i\rangle = 2$. For any $Y \in \mathfrak{t}^\perp$, we can write $Y = xX_1 + bX_2 + cX_3$. The condition that $Y$ is orthogonal to all elements in $\mathfrak{t}$ implies that $2ax = 0, \forall a \in \mathbb{R}$. Hence $x = 0$, and $Y = bX_2 + cX_3$. □

**Proposition B.22.** $(\mathbf{SU}(2), -\frac{1}{4}B)$ *has constant sectional curvature equal to* $\frac{1}{2}$.

*Proof.* By Lemma B.15 and A.7, $\mathbf{SU}(2)$ has constant sectional curvature. The sectional curvature is of the form $\sec(X,Y) = \frac{1}{4}||[X,Y]||^2$ under an orthonormal basis of $\mathfrak{su}(2)$. From the proof in Theorem B.21, we choose $X = \frac{1}{\sqrt{2}}X_1, Y = \frac{1}{\sqrt{2}}X_2$. By a direct computation, we have $[X,Y] = -X_3$, thus $\sec(X,Y) = \frac{1}{2}$. □

**Theorem B.23.** *The injectivity radius* $r$ *of* $\mathbf{SU}(2)$ *is* $\sqrt{2}\pi$.

*Proof.* From the explicit form of matrix $Q \in \mathbf{SU}(2)$ given in Eq. (16), it follows that $\mathbf{SU}(2)$ is isometric to the unit sphere $\mathbb{S}^3 \subset \mathbb{R}^4$. A closed geodesic in $\mathbb{S}^3$ is the great circle, whose length is $2\pi R$, where $R$ is the radius of the sphere. Using the relation between radius and curvature, we have $\frac{1}{R^2} = \frac{1}{2}$, thus $R = \sqrt{2}$. By Lemma 5.7, the injectivity radius of $\mathbf{SU}(2)$ is therefore given by $\mathrm{Inj}(\mathbf{SU}(2)) = \sqrt{2}\pi$. □

**Lemma B.24.** $B_{\mathfrak{t}}(0,r) = \{X = aX_1 | |a| < \pi, a \in \mathbb{R}\}$, $B_{\mathfrak{t}^\perp}(0,r) = \{Y = bX_2 + cX_3 | b^2+c^2 < \pi^2, b, c \in \mathbb{R}\}$.

**Theorem B.25.** *The neighborhood* $U_e \subset \mathbf{SU}(2)$ *in Theorem 4.4 is given by the following form:*

$$U_e = \left\{ \begin{pmatrix} \cos(\varphi)e^{ia} & -\sin(\varphi)e^{i(a-\theta)} \\ \sin(\varphi)e^{-i(a-\theta)} & \cos(\varphi)e^{-ia} \end{pmatrix} \middle| \begin{matrix} |a|<\pi,\ 0<\varphi<\pi, \\ 0<\theta<2\pi \end{matrix} \right\}. \tag{51}$$

*Proof.* For any $X \in B_{\mathfrak{t}}(0,r)$, Lemma B.24 gives $X = aX_1$, with $|a| < \pi$.Assuming $a > 0$, then the angle $\theta = 2a$, unit vector $\mathbf{n} = [0,0,-1]^\top$. Applying the Rodrigues formula yields $\exp X = \cos(a)\mathbf{I} + i\sin(a)\sigma_z = \begin{pmatrix} e^{ia} & 0 \\ 0 & e^{-ia} \end{pmatrix}$. One can get the same result for the case of $a < 0$. Next, for any $Y \in B_{\mathfrak{t}^\perp}(0,r)$, by Lemma B.24, we have $Y = bX_2 + cX_3$, $b^2+c^2 < \pi^2$. Then the rotation angle and axis are given by $\theta = 2\sqrt{b^2+c^2}$, $\mathbf{n} = \frac{[-c,b,0]^\top}{\sqrt{b^2+c^2}}$. Define $\varphi = \sqrt{b^2+c^2}$ and let

**Algorithm 1** Local Exponential-Coordinate Kernel on a Compact Lie group

**Ensure:** Compact Lie group $\mathcal{G}$ and a point $p \in \mathcal{G}$
**Require:** Neighborhood $U_p$ and Local exponential-coordinate kernel defined on it
1: Find a bi-invariant metric on $\mathcal{G}$
2: Compute its Lie algebra $\mathfrak{g}$
3: Compute a maximal torus $\mathcal{T}$ of $\mathcal{G}$ and its corresponding Cartan subalgebra $\mathfrak{t}$
4: Decompose the Lie algebra with its Cartan subalgebra $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{t}^{\perp}$
5: Compute the injectivity radius $r$ of $\mathcal{G}$
6: Take $r_p \leq r$ and compute $U_p$ and $\Phi_p^{-1}$ by Eq. (35)
7: The local Exponential-Coordinate kernel is given by Theorem B.12

$\theta$ be such that $\cos(\theta) = \frac{b}{\sqrt{b^2+c^2}}, \sin(\theta) = \frac{c}{\sqrt{b^2+c^2}}$. Applying the Rodrigues formula again, we can deduce that

$$\begin{aligned} \exp(Y) &= \cos(\varphi)\mathbf{I} - i\sin(\varphi)(-\sin(\theta)\sigma_x + \cos(\theta)\sigma_y) \\ &= \begin{pmatrix} \cos(\varphi) & -\sin(\varphi)e^{-i\theta} \\ \sin(\varphi)e^{i\theta} & \cos(\varphi) \end{pmatrix}. \end{aligned} \tag{52}$$

By substituting these expressions into the definition of $U_e = \{\exp(X)\exp(Y) | X \in B_{\mathfrak{t}}(0,r), Y \in B_{\mathfrak{t}^{\perp}}(0,r)\}$, we can get the desired result. □

**Theorem B.26.** *For any* $g = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} \in U_e$, *the inverse* $\Phi_e^{-1}(g) = (X, Y)$ *is given by*

$$X = aX_1, a = \arg(\alpha), \tag{53}$$

*and*

$$Y = \varphi\cos(\theta)X_2 + \varphi\sin(\theta)X_3, \tag{54}$$

*where* $\varphi = \arccos(|\alpha|)$, $\theta = \arg(\alpha\beta)$. *Here* $\arg \in [-\pi, \pi)$ *denotes the principal argument of a complex number* $z$.

*Proof.* By Theorem B.25, we have

$$\alpha = \cos(\varphi)e^{ia}, \beta = \sin(\varphi)e^{-i(a-\theta)}. \tag{55}$$

From the first relation, we obtain

$$a = \arg(\alpha), \varphi = \arccos(|\alpha|). \tag{56}$$

Multiplying the two equations gives $\alpha\beta = \sin(\varphi)\cos(\varphi)e^{i\theta}$, which yields $\theta = \arg(\alpha\beta)$. □

# C Additional Experimental Results

## C.1 Pseudo Code

We present the pseudo-code in Algorithm 1 for better understanding of how to construct the local exponential-coordinate kernel on compact Lie groups.

## C.2 Positive Definiteness Verification

We first verify the positive definiteness of our proposed local Gaussian kernel on Lie group $\mathbf{SO}(3)$ from three perspectives: sample size, hyperparameter configuration, and base point location, which enables us to establish the kernel's theoretical soundness across all relevant dimensions. Since $\mathbf{SU}(2)$ and $\mathbf{SO}(3)$ are locally isometric, the local behavior of kernel on $\mathbf{SU}(2)$ is mathematically equivalent to that on $\mathbf{SO}(3)$, we omit the empirical validation on $\mathbf{SU}(2)$ for brevity.

**Sample Size**

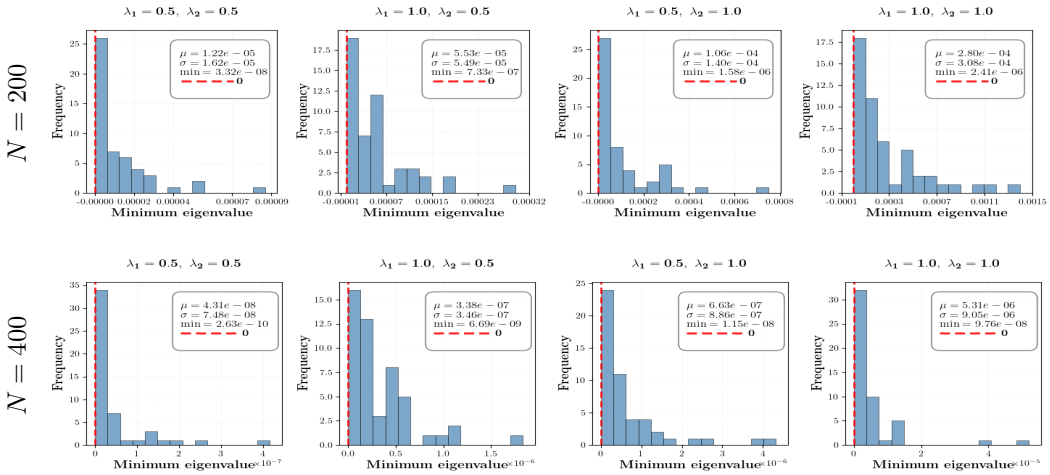



Figure 3: Minimum eigenvalue distribution on $\mathbf{SO}(3)$ with sample sizes $N = 200$ (top) and $N = 400$ (bottom). Each histogram shows 50 independent trials. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

To examine the robustness of the positive definiteness of our kernel on $\mathbf{SO}(3)$ with respect to sample size, we repeat the verification experiment in Section 6.1 with a larger sample size $N$. Figure 3 shows the distribution of minimum eigenvalues for $N = 200$ (top row) alongside the result for $N = 400$ (bottom row). All eigenvalues remain strictly positive, confirming that the positive definiteness holds regardless of sample size. One can observe that the minimum eigenvalue tends to zero as the data size increases, which is consistent with the theoretical expectation that the kernel matrix becomes increasingly ill-conditioned but remains positive definite.

**Hyperparameter Pair**

In this subsection, we further explore the influence of the hyperparameter pair $(\lambda_1, \lambda_2)$ on the positive definiteness of our local Gaussian kernel on $\mathbf{SO}(3)$. For each experimental configuration, we sample $N = 100$ points and conduct 100 independent trials. Figure 4 shows the distribution of minimum eigenvalues for different isotropic hyperparameter pairs. The results indicate that larger hyperparameter values lead to larger minimum eigenvalues, suggesting that the kernel becomes better conditioned as $(\lambda_1, \lambda_2)$ increase.

We then investigate anisotropic hyperparameter pairs. Figure 5 shows results with fixed $\lambda_1 = 0.5$ and varying $\lambda_2$, while Figure 6 shows the symmetric case with fixed $\lambda_2 = 0.5$ and varying $\lambda_1$. In both scenarios, when one hyperparameter is held constant, the minimum eigenvalues increase as the other hyperparameter increases. Crucially, all minimum eigenvalues remain strictly positive across all configurations, confirming that our kernel remains positive definite for both isotropic and anisotropic parameter choices.

**Arbitrary Points**

To verify that positive definiteness holds on the neighborhood $U_p$ of arbitrary point $p \in \mathbf{SO}(3)$, we extend the experiment for a set of base points that form a uniform covering of $\mathbf{SO}(3)$. The uniform sampling strategy leverages $\mathbb{S}^3$ as the universal covering of $\mathbf{SO}(3)$. Specifically, we can identify $\mathbb{S}^3$ with $\mathbf{SU}(2)$ by the following homeomorphism:

$$\begin{aligned} \psi : \mathbb{S}^3 &\longrightarrow \mathbf{SU}(2) \\ (a, b, c, d) &\longmapsto \begin{pmatrix} a + ib & -c + id \\ c + id & a - ib \end{pmatrix}. \end{aligned} \tag{57}$$

For each matrix $U \in \mathbf{SU}(2)$, $U$ can be uniquely determined by a unit axis $\mathbf{n} = [n_x, n_y, n_z]^\top$ and an angle $\theta$:

$$U = \cos(\frac{\theta}{2}) - i \sin(\frac{\theta}{2})(\mathbf{n} \cdot \boldsymbol{\sigma}), \tag{58}$$

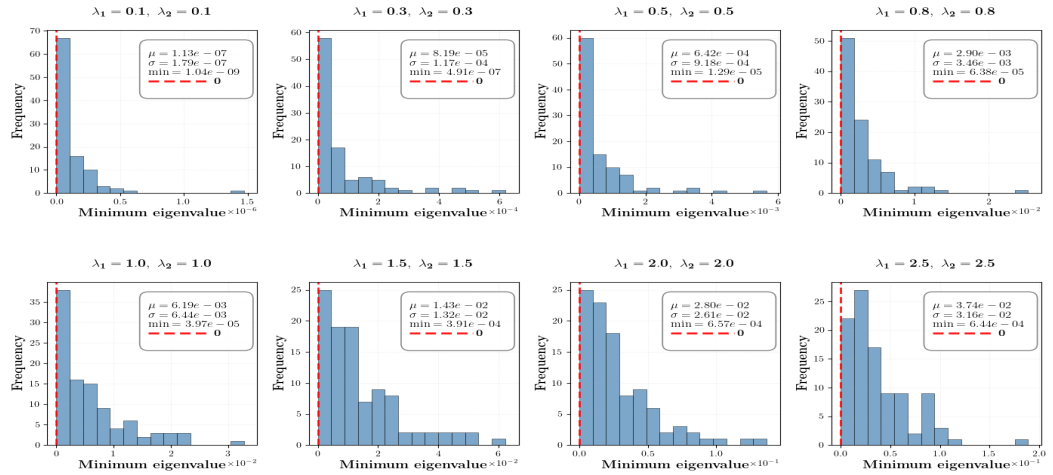



Figure 4: Minimum eigenvalue distribution on $\mathbf{SO}(3)$ with different isotropic weight pairs $(\lambda_1, \lambda_2)$. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

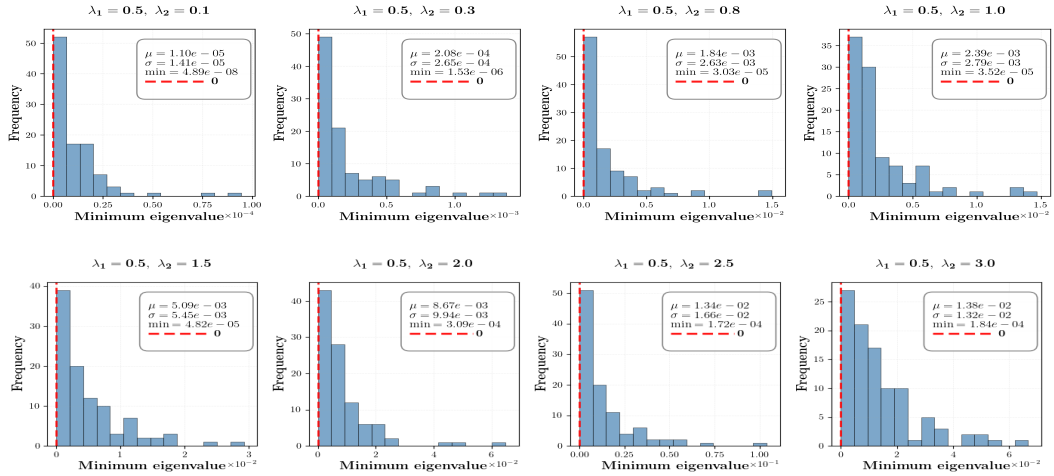



Figure 5: Minimum eigenvalue distribution on $\mathbf{SO}(3)$ with different anisotropic hyperparameter pairs with fixed $\lambda_1 = 0.5$ and varying $\lambda_2$. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

where $\boldsymbol{\sigma}$ is the Pauli matrices B.4.1. This gives an rotation matrix in $\mathbf{SO}(3)$ by Rodrigues formula B.3.1:

$$R = \mathbf{I} + \sin(\theta)K + (1 - \cos(\theta))K^2, \tag{59}$$

where $K = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix}$ is the skew-symmetric matrix corresponding to $\mathbf{n}$. Then we can define the universal covering map: $\phi : \mathbf{SU}(2) \to \mathbf{SO}(3)$ by $\phi(U) = K$. One can verify that $\phi(-U) = R$, so that the degree of $\phi$ is 2. If we write $U$ as $\begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix}$, then we can derive the explicit formula for $R$:

$$R = \begin{pmatrix} \mathrm{Re}(\alpha^2 - \beta^2) & \mathrm{Im}(\alpha^2 + \beta^2) & -2\mathrm{Re}(\alpha\beta) \\ -\mathrm{Im}(\alpha^2 - \beta^2) & \mathrm{Re}(\alpha^2 + \beta^2) & 2\mathrm{Im}(\alpha\beta) \\ 2\mathrm{Re}(\alpha\bar{\beta}) & 2\mathrm{Im}(\alpha\bar{\beta}) & |\alpha|^2 - |\beta|^2 \end{pmatrix}. \tag{60}$$

The composition $f = \phi \circ \psi$ defines a universal covering map $f : \mathbb{S}^3 \to \mathbf{SO}(3)$, where each pair of antipodal points on $\mathbb{S}^3$ is identified with a single element in $\mathbf{SO}(3)$. To obtain a uniform sampling of

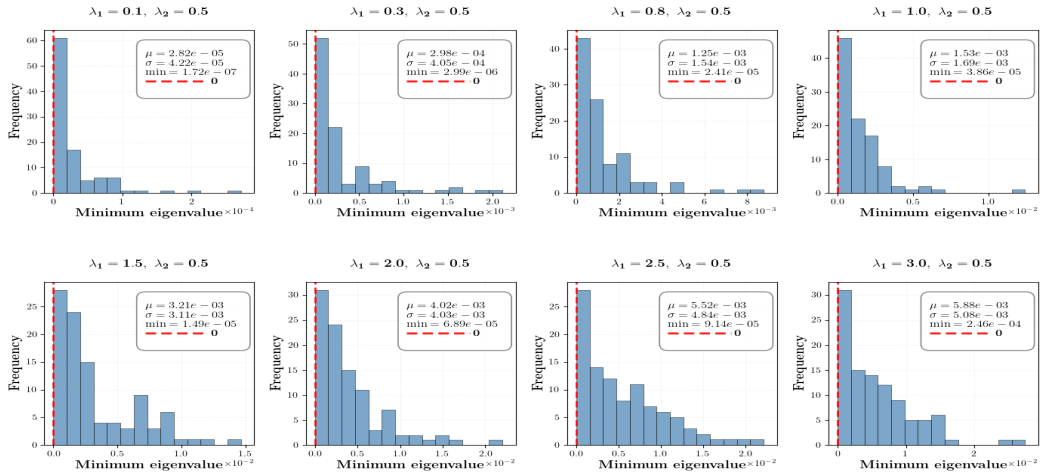



Figure 6: Minimum eigenvalue distribution on $\mathbf{SO}(3)$ with different anisotropic hyperparameter pairs with fixed $\lambda_2 = 0.5$ and varying $\lambda_1$. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

$\mathbf{SO}(3)$, we uniformly sample 16 points from the upper hemisphere of $\mathbb{S}^3$ and map them via $f$. This yields a set of rotations that uniformly cover $\mathbf{SO}(3)$ without redundancy.

For each base point $p$, we uniformly sample $N = 100$ points from its neighborhood $U_p = pU_e$, and compute the minimum eigenvalue of the corresponding kernel matrix. This procedure is repeated for 100 independent trials to obtain statistical significance, with the hyperparameter pair fixed to be $(\lambda_1, \lambda_2) = (1, 1)$. Figure 7 shows the distribution of minimum eigenvalues for sixteen uniformally sampled base points $p_1, \ldots, p_{16} \in \mathbf{SO}(3)$, where each subfigure displays the histogram of the 100 minimum eigenvalues for one base point. As observed, all minimum eigenvalues are strictly positive across all sixteen base points, confirming that our kernel remains positive definite on $U_p$ for any $p \in \mathbf{SO}(3)$.

### C.3 Gaussian Process on $\mathbf{SO}(3)$

In this subsection, we extend the experiment for Gaussian process regression on $\mathbf{SO}(3)$ in terms of training dataset, hyperparameter pair, noise level and base points. Finally, we extend our local Gaussian process to the global manifold using the Gaussian mixture model. The target function remains the rotation angle $f(R) = \arccos((\mathrm{tr}(R) - 1)/2)$ for all experiments.

#### C.3.1 Gaussian Process Regression on $U_e$

We first extend the experiment described in Section 6.2 across three dimensions: training data size, hyperparameter pair and noise level.

**Training Dataset**

We test the performance of Gaussian process regression using our proposed Gaussian kernel under larger training dataset sizes: $n = 200, 300, 600, 1000$. In all experiments, we fix the hyperparameters $\lambda_1 = \lambda_2 = 1$, the number of test points $m = 100$, and the noise level $\epsilon_i \sim \mathcal{N}(0, 0.02^2)$. The prediction results for each training size are shown in Figures 2b through 10. As expected, the prediction errors (RMSE and MAE) decrease consistently as the training dataset size increases, demonstrating the improved accuracy of the Gaussian process model with more training data.

**Hyperparameter Pair**

We investigate the influence of the kernel hyperparameters $(\lambda_1, \lambda_2)$ on the performance of Gaussian process regression on $\mathbf{SO}(3)$. For all experiments, we fix the training dataset size to $n = 500$, the noise level to $\sigma = 0.02$, and the number of test points to $m = 100$. Both isotropic ($\lambda_1 = \lambda_2$) and anisotropic ($\lambda_1 \neq \lambda_2$) cases are tested. As shown in Figure 11 for the isotropic case and Figure 12 for the anisotropic case, all tested hyperparameter configurations yield accurate predictions. A clear

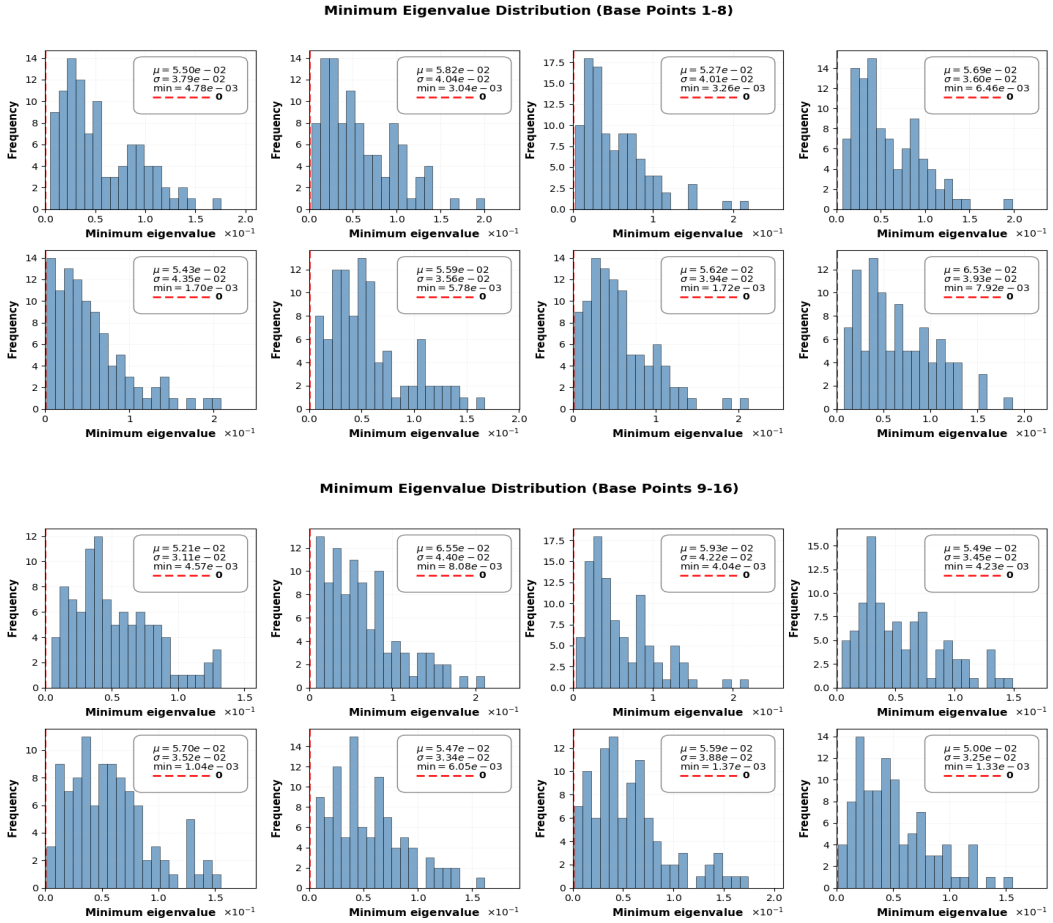



Figure 7: Minimum eigenvalue distribution on neighborhood $U_p$ for 16 uniformally sampled base points $p \in \mathbf{SO}(3)$ with fixed hyperparameter pair $(\lambda_1, \lambda_2) = (1, 1)$. All eigenvalues are positive, confirming the positive definiteness of our kernel function.

trend is observed: smaller $\lambda$ values produce smaller prediction errors. This holds for both isotropic and anisotropic cases. When $\lambda$ increases, the kernel becomes more localized, leading to reduced smoothness and increased prediction errors.

### Noise Level

We investigate the robustness of our local Gaussian kernel to observation noise with fixed training size $n = 500$ and test size $m = 100$. The performance is evaluated using Mean Absolute Error (MAE) across increasing noise levels $\sigma = [0.02, 0.05, 0.08, 0.12, 0.15, 0.18, 0.2]$. Figure 13 presents results for both isotropic ($\lambda_1 = \lambda_2 = \lambda$) and anisotropic hyperparameter configurations. As expected, MAE increases monotonically with noise level across all settings, from approximately $0.02$ rad at $\sigma = 0.02$ to $0.10$ rad at $\sigma = 0.2$. In the isotropic case, $\lambda = 0.5$ performs best overall, while $\lambda = 1.0$ achieves the smallest error under low noise ($\sigma = 0.02, 0.05$). Globally, MAE increases with hyperparameter $\lambda$, despite some fluctuations at individual noise levels. In the anisotropic case, all curves exhibit similar performance, except that configurations with the largest hyperparameter ($\lambda_1 = 2.5$ or $\lambda_2 = 2.5$) show the highest errors under high noise ($\sigma = 0.18, 0.2$). Symmetric hyperparameter pairs $(\lambda_1, \lambda_2)$ and $(\lambda_2, \lambda_1)$ yield comparable results. Overall, the configuration $(\lambda_1, \lambda_2) = (0.5, 1.0)$ achieves the best performance across most noise levels, combining the strengths of both as observed in the isotropic case: $\lambda = 0.5$ performs best globally while $\lambda = 1.0$ excels under low noise.

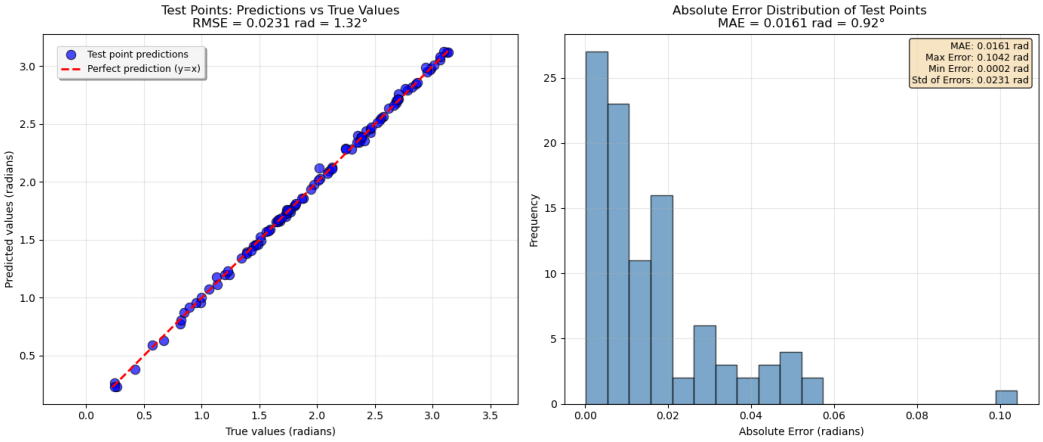



Figure 8: Training dataset size $n = 300$

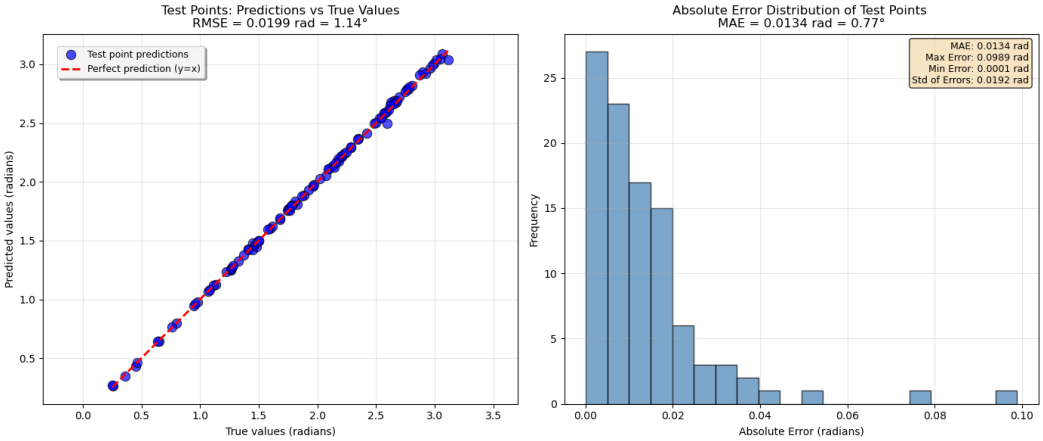



Figure 9: Training dataset size $n = 600$

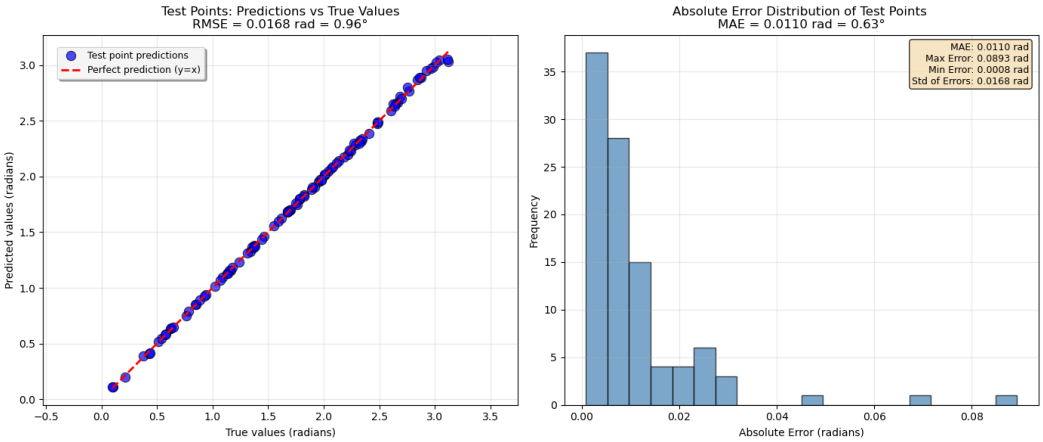
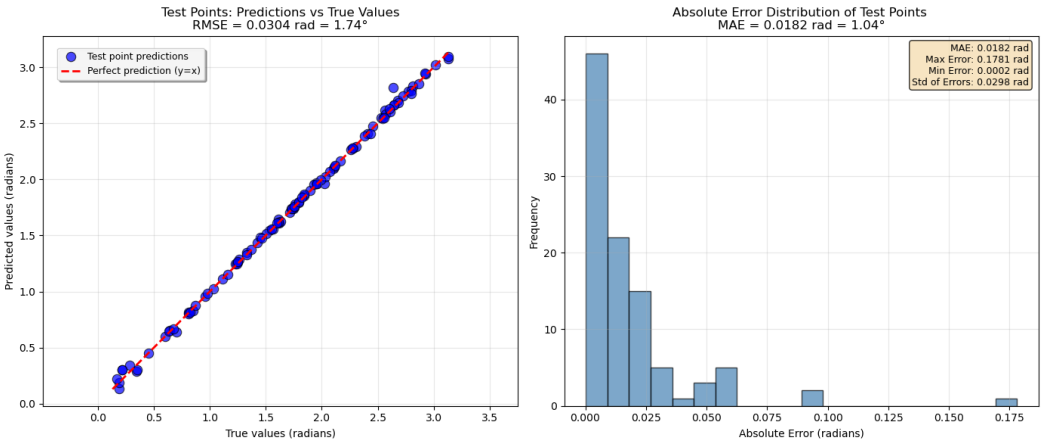
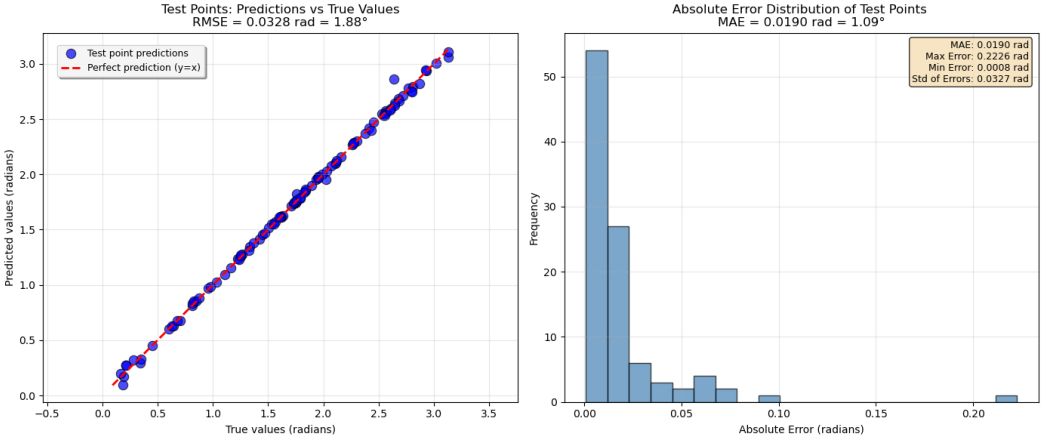


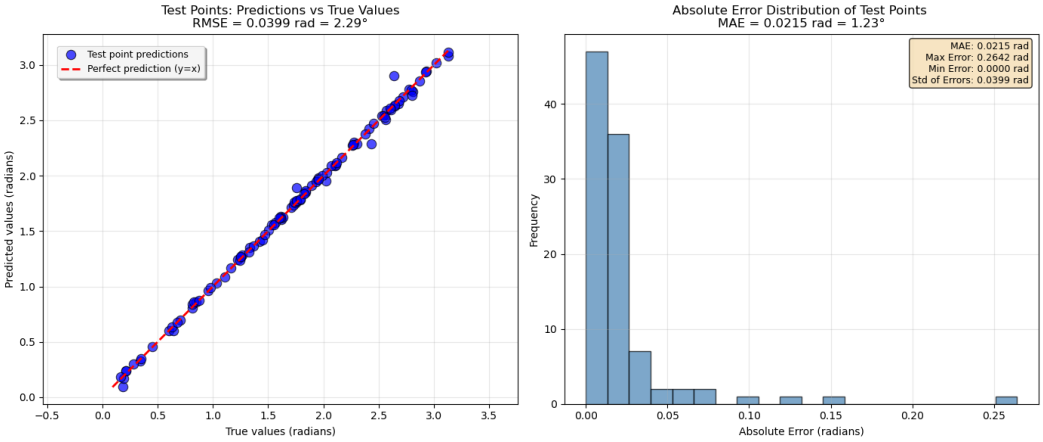

Figure 10: Training dataset size $n = 1000$




(a) $\lambda_1 = \lambda_2 = 0.5$

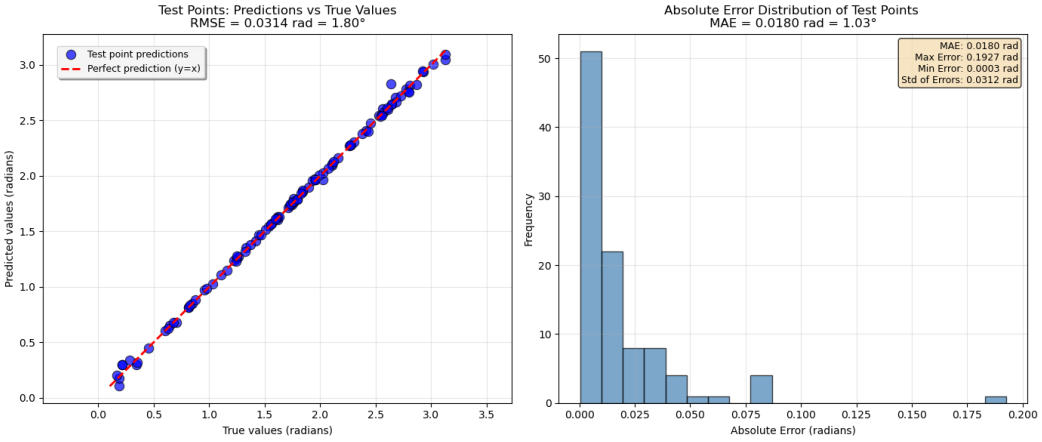



(b) $\lambda_1 = \lambda_2 = 1$

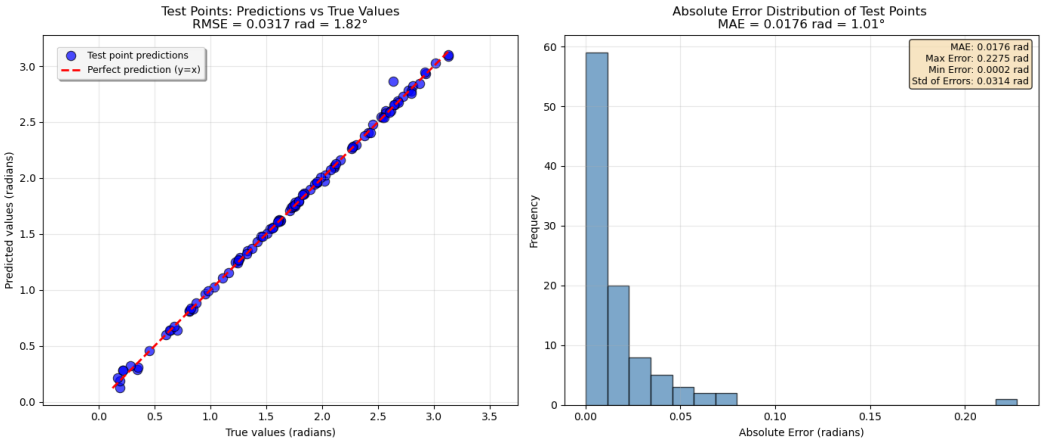



(c) $\lambda_1 = \lambda_2 = 2$

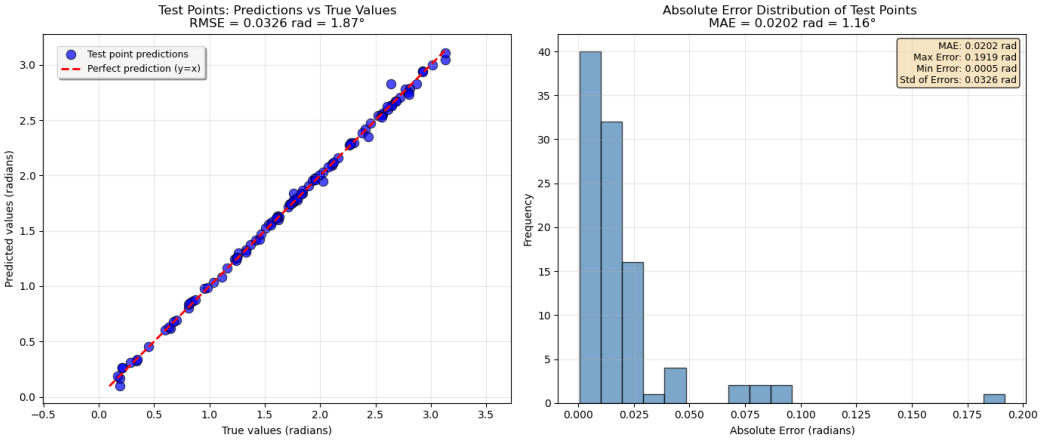
Figure 11: Gaussian Process Regression on $U_e \subset \mathbf{SO}(3)$ with isotropic hyperparameter pair. As observed, the error becomes larger with increase of hyperparameter.




(a) $\lambda_1 = 1, \lambda_2 = 0.5$

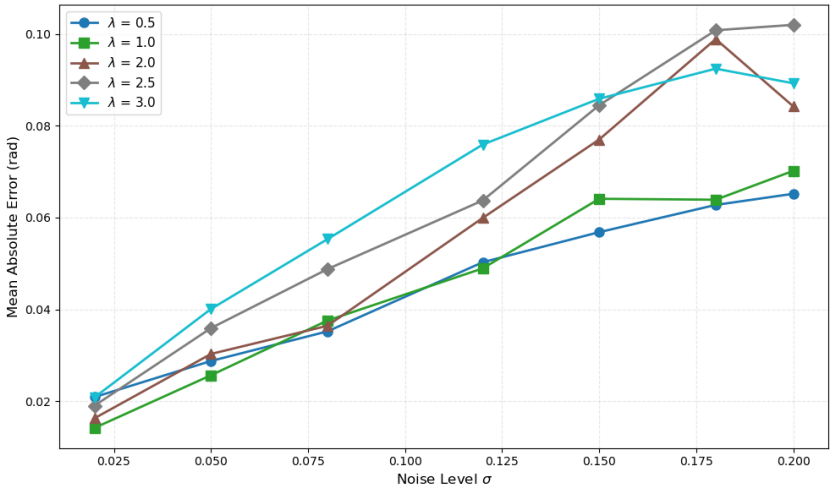



(b) $\lambda_1 = 0.5, \lambda_2 = 1$

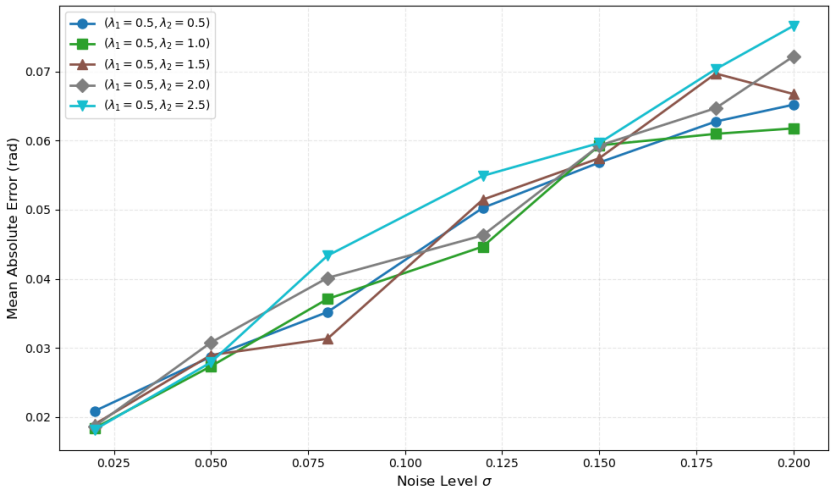



(c) $\lambda_1 = 2, \lambda_2 = 1$

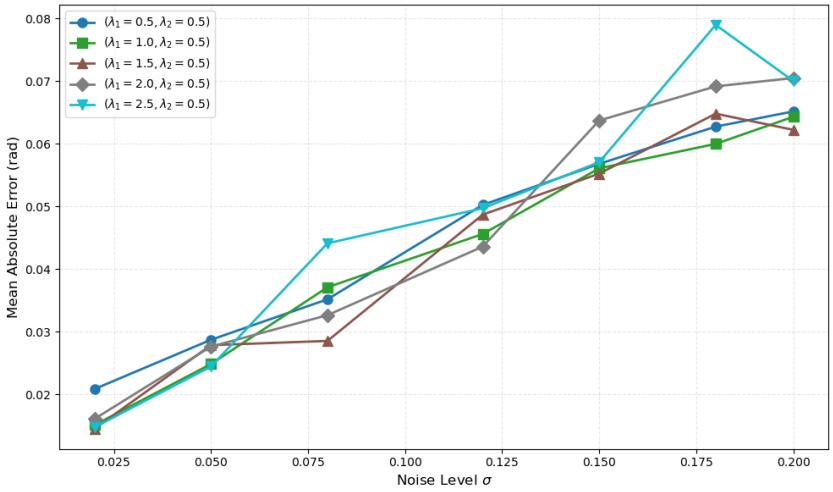
Figure 12: Gaussian Process Regression on $U_e \subset \mathbf{SO}(3)$ with anisotropic hyperparameter pair.




(a) Isotropic case $\lambda_1 = \lambda_2 = \lambda$




(b) Anisotropic case with fixed $\lambda_1 = 0.5$ and varying $\lambda_2$




(c) Anisotropic case with fixed $\lambda_2 = 0.5$ and varying $\lambda_1$

Figure 13: Noise test for gaussian process regression on $U_e \subset \mathbf{SO}(3)$ with different hyperparameter pairs. All experiments use $n = 500$ training data and $m = 100$ test data. The results demonstrate that our kernel function is robust to noise under different hyperparameter pairs.

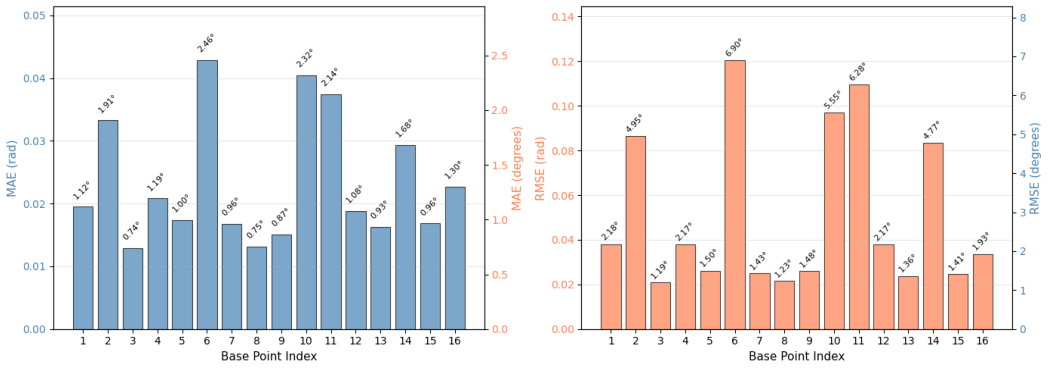



Figure 14: Gaussian process regression result on 16 base points covering $\mathbf{SO}(3)$. Left: MAE with left axis in radians and right axis in degrees; Right: RMSE with left axis in radians and right axis in degrees. All experiments use $n = 500$ training points and $m = 100$ test points with noise level $\sigma = 2\%$ and kernel hyperparameter pair $(\lambda_1, \lambda_2) = (0.5, 1.0)$. The corresponding numerical results are listed in Table 2.

### C.3.2 Gaussian Process Regression on Arbitrary Charts

To further validate the applicability of our local Gaussian kernel across $\mathbf{SO}(3)$, we conduct Gaussian process regression on neighborhoods $U_p$ of other non-identity elements $p \in \mathbf{SO}(3)$. Leveraging the compactness of $\mathbf{SO}(3)$, it can be covered by a finite number of neighborhoods $\{U_{p_i}\}_{i=1}^{N}$. We uniformly sample $N = 16$ base points $\{p_i\}_{i=1}^{16}$ on $\mathbf{SO}(3)$ and perform Gaussian process regression independently on each neighborhood $U_{p_i} = p_i U_e$. All experiments use $n = 500$ training points and $m = 100$ test points with noise level $\sigma = 2\%$ and hyperparameter pair $(\lambda_1, \lambda_2) = (0.5, 1.0)$.

The results are visualized in Figure 14, which displays bar plots of MAE and RMSE across the 16 base points, with left axes in radians and right axes in degrees. The corresponding numerical values are summarized in Table 2. The mean MAE across all base points is $0.0234$ rad ($1.34°$) with a small standard deviation of $0.0097$ rad ($0.55°$), indicating consistent performance across different regions of $\mathbf{SO}(3)$. Similarly, the mean RMSE is $0.0507$ rad ($2.91°$) with standard deviation $0.0341$ rad ($1.95°$). These results demonstrate that our local Gaussian kernel maintains excellent and stable predictive performance across the entire manifold, confirming its suitability for global Gaussian process regression on $\mathbf{SO}(3)$.

## C.4 Global Coverage Analysis and Baseline Comparison for $\mathbf{SO}(3)$

In this subsection, we provide a complete analysis of the global coverage property of our local chart $U_e$ in $\mathbf{SO}(3)$. We first prove that $U_e$ covers the entire group almost everywhere, then compare our Gaussian kernel with other global baselines.

### C.4.1 Rotations and Quaternions

To establish the global coverage of $U_e$, we utilize the quaternion representation of rotations, a formalism that significantly simplifies the composition and algebraic manipulation of elements in $\mathbf{SO}(3)$.

**Definition C.1.** A quaternion $\mathbf{q} \in \mathbb{H}$ is represented in the form:

$$\mathbf{q} = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{61}$$

where $w, x, y, z$ are real numbers, and $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$ are the basis vectors.

A quaternion is said to be unit if its norm $\|q\| = \sqrt{w^2 + x^2 + y^2 + z^2} = 1$. A multiplicative group structure called the *Hamilton product* can be defined on the quaternions in the following way. The product is first given for the basis elements, and then extended to all quaternions by using the associative and distributive laws. The multiplication for basis elements are defined by:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \quad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}, \tag{62}$$

| Base Point Index | MAE | | RMSE | |
|---|---|---|---|---|
| | (rad) | (deg) | (rad) | (deg) |
| 1 | 0.0195 | 1.12° | 0.0380 | 2.18° |
| 2 | 0.0333 | 1.91° | 0.0865 | 4.95° |
| 3 | 0.0129 | 0.74° | 0.0208 | 1.19° |
| 4 | 0.0208 | 1.19° | 0.0379 | 2.17° |
| 5 | 0.0174 | 1.00° | 0.0262 | 1.50° |
| 6 | 0.0429 | 2.46° | 0.1205 | 6.90° |
| 7 | 0.0168 | 0.96° | 0.0250 | 1.43° |
| 8 | 0.0131 | 0.75° | 0.0215 | 1.23° |
| 9 | 0.0151 | 0.87° | 0.0259 | 1.48° |
| 10 | 0.0405 | 2.32° | 0.0969 | 5.55° |
| 11 | 0.0374 | 2.14° | 0.1096 | 6.28° |
| 12 | 0.0189 | 1.08° | 0.0378 | 2.17° |
| 13 | 0.0163 | 0.93° | 0.0238 | 1.36° |
| 14 | 0.0293 | 1.68° | 0.0833 | 4.77° |
| 15 | 0.0168 | 0.96° | 0.0246 | 1.41° |
| 16 | 0.0227 | 1.30° | 0.0337 | 1.93° |
| **Mean** | 0.0234 | 1.34° | 0.0507 | 2.91° |
| **Std** | 0.0097 | 0.55° | 0.0341 | 1.95° |

Table 2: The numerical results of Gaussian process regression on neighborhoods of 16 uniformly sampled base points on $\mathbf{SO}(3)$.

the basis element 1 is defined as the multiplicative identity. Combining these rules, $\mathbf{ijk} = -1$.

The relationship between a rotation and a quaternion is given by the following Proposition.

**Proposition C.2.** *Let $R \in \mathbf{SO}(3)$ be a rotation defined by the unit axis $\mathbf{u} = [u_x, u_y, u_z]^\top$ and angle $\theta$. Then $R$ can be represented by a unit quaternion $\mathbf{q}$:*

$$\mathbf{q} = \cos(\frac{\theta}{2}) + \sin(\frac{\theta}{2})(u_x\mathbf{i} + u_y\mathbf{j} + u_z\mathbf{k}). \tag{63}$$

A key advantage of the quaternion formulation is that the composition of two rotations corresponds directly to quaternion multiplication:

**Proposition C.3.** *If $\mathbf{q}_A$ and $\mathbf{q}_B$ are the unit quaternions representing rotations $R_A$ and $R_B$, then the composite rotation $R_C = R_B R_A$ is represented by the quaternion product $\mathbf{q}_C = \mathbf{q}_B \mathbf{q}_A$.*

### C.4.2 Theoretical Proof

**Lemma C.4.** *Let $f : \mathcal{M} \to \mathcal{N}$ be a smooth map between two manifolds with dim$(\mathcal{M})$ < dim$(\mathcal{N})$. Then the image $f(\mathcal{M})$ has Lebesgue measure 0 in $\mathcal{N}$.*

*Proof.* This is a direct result of Sard's Theorem [40]. □

**Theorem C.5** (Global Coverage of $U_e$)**.** *Let $U_e = \{R_{\mathbf{z}}(a)R_{\mathbf{n}(\phi)}(\psi) | |a| < \pi, 0 < \psi < \pi, 0 < \phi < 2\pi\}$ be the chart given in Theorem B.19, where $\mathbf{n}(\phi) = [0, \sin(\phi), \cos(\phi)]^\top$. Then $U_e$ covers $\mathbf{SO}(3)$ almost everywhere, i.e., its complement $\mathbf{SO}(3) \setminus U_e$ has Lebesgue measure 0.*

*Proof.* Consider the parameterization map $\Phi : \mathcal{D} = [-\pi, \pi] \times [0, 2\pi] \times [0, \pi] \to \mathbf{SO}(3)$ defined by

$$\Phi(a, \phi, \psi) = R_{\mathbf{z}}(a)R_{\mathbf{n}(\phi)}(\psi).$$

Then $U_e = \Phi(\mathcal{D}^\circ)$, where $\mathcal{D}^\circ$ is the interior of $\mathcal{D}$.

**Step 1: $\Phi$ is surjective.** To prove that $\Phi$ is surjective, we will show that for any $R \in \mathbf{SO}(3)$, there exists $a, \phi, \varphi \in \mathcal{D}$ such that

$$R_{\mathbf{z}}(a)R_{\mathbf{n}(\phi)}(\psi) = R. \tag{64}$$

Let $\mathbf{q} = w_x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ be the quaternion representing $R$. By Proposition C.2, the quaternion corresponding to $R_{\mathbf{z}}(a)$ is $\mathbf{q}(a) = \cos(\frac{a}{2}) + \sin(\frac{a}{2})\mathbf{k}$, and the quaternion corresponding to $R_{\mathbf{n}(\phi)}(\psi)$ is $\mathbf{q}(\phi, \psi) = \cos(\frac{\psi}{2}) + \sin(\frac{\psi}{2})(\sin(\phi)\mathbf{j} + \cos(\phi)\mathbf{k})$. Applying Proposition C.3, the quaternion representing the composite rotation $R_{\mathbf{z}}(a)R_{\mathbf{n}(\phi)}(\psi)$ is

$$\begin{aligned}\mathbf{q}(a, \phi, \psi) &= \mathbf{q}(a)\mathbf{q}(\phi, \psi) \\ &= \cos(\frac{a}{2})\cos(\frac{\psi}{2}) - \sin(\frac{a}{2})\sin(\frac{\psi}{2})\cos(\phi) - \sin(\frac{a}{2})\sin(\frac{\psi}{2})\sin(\phi)\mathbf{i} \\ &+ \cos(\frac{a}{2})\sin(\frac{\psi}{2})\sin(\phi)\mathbf{j} + (\cos(\frac{\psi}{2})\sin(\frac{a}{2}) + \cos(\frac{a}{2})\sin(\frac{\psi}{2})\cos(\phi))\mathbf{k}. \end{aligned} \tag{65}$$

Thus the problem is equivalent to the following system of equations has a solution:

$$\begin{cases} w = \cos(\frac{a}{2})\cos(\frac{\psi}{2}) - \sin(\frac{a}{2})\sin(\frac{\psi}{2})\cos(\phi), & \text{(66a)} \\ x = -\sin(\frac{a}{2})\sin(\frac{\psi}{2})\sin(\phi), & \text{(66b)} \\ y = \cos(\frac{a}{2})\sin(\frac{\psi}{2})\sin(\phi), & \text{(66c)} \\ z = \cos(\frac{\psi}{2})\sin(\frac{a}{2}) + \cos(\frac{a}{2})\sin(\frac{\psi}{2})\cos(\phi). & \text{(66d)} \end{cases}$$

**Case 1:** $y \neq 0$**.** By Eq. (66b) and (66c), we have $\frac{x}{y} = -\tan(\frac{a}{2})$, which implies that $a = 2\arctan(-\frac{x}{y}) \in (-\pi, \pi)$. Combining Eq. 66a and 66d, we can derive a system of equations for $\phi$ and $\psi$:

$$A\begin{pmatrix}\cos(\frac{\psi}{2}) \\ \sin(\frac{\psi}{2})\cos(\phi)\end{pmatrix} = \begin{pmatrix} w \\ z \end{pmatrix}, \tag{67}$$

where $A = \begin{pmatrix}\cos(\frac{a}{2}) & -\sin(\frac{a}{2}) \\ \sin(\frac{a}{2}) & \cos(\frac{a}{2})\end{pmatrix}$. Since $\det(A) = 1$, the system of equations is solvable. Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = A^{-1}\begin{pmatrix} w \\ z \end{pmatrix}$, we deduce that

$$\cos(\frac{\psi}{2}) = v_1, \quad \sin(\frac{\psi}{2})\cos(\phi) = v_2. \tag{68}$$

From the first equation in Eq. 68, we get the solution for $\psi$: $\psi = 2\arccos(v_1) \in [0, 2\pi]$. If $v_2 = 0$, then $\phi = \frac{\pi}{2}$. If $v_2 \neq 0$, then combining the second equation in Eq 68 and 66c, it follows that $\phi = \arctan(\frac{y}{\cos(\frac{a}{2})v_2})$. If $\psi \in (\pi, 2\pi]$, then we take $\tilde{\psi} = 2\pi - \psi \in [0, \pi)$, $\tilde{\phi} = \phi + \pi \pmod{2\pi}$, $\tilde{a} = a$. This gives the quaternion $\mathbf{q}(\tilde{a}, \tilde{\phi}, \tilde{\psi}) = -\mathbf{q}(a, \phi, \psi)$ which defines the same rotation as the quaternion $\mathbf{q}(a, \phi, \psi)$, thus also solves the Eq. 64.

**Case 2:** $y = 0$**.**

- **Subcase 2.1:** $\cos(\frac{a}{2}) = 0$**.** This implies that $a = \pi$. The other three equations in Eq. 66 reduces to:

$$\begin{cases} w = -\sin(\frac{\psi}{2})\cos(\phi), & \text{(69a)} \\ x = -\sin(\frac{\psi}{2})\sin(\phi), & \text{(69b)} \\ z = \cos(\frac{\psi}{2}). & \text{(69c)} \end{cases}$$

  if $\sin(\frac{\psi}{2}) = 0$, then $\psi = 0$ and $\phi$ can be an arbitrary angle. If $\sin(\frac{\psi}{2}) \neq 0$, then $\psi = 2\arccos(z)$, $\phi = \arctan\frac{x}{w}$.

- **Subcase 2.2:** $\sin(\frac{\psi}{2}) = 0$**.** Then $\psi = 0$, the equation system 66 reduces to:

$$\begin{cases} w = \cos(\frac{a}{2}), & \text{(70a)} \\ z = \sin(\frac{a}{2}). & \text{(70b)} \end{cases}$$

  If $\cos(\frac{a}{2}) = 0$, then $a = \pi$. If $\cos(\frac{a}{2}) \neq 0$, then $a = 2\arctan(\frac{z}{w})$. $\phi$ is an arbitrary value.

- **Subcase 2.3:** $\sin(\phi) = 0$ This implies that $\phi = 0, \pi, 2\pi$. If $\phi = 0, 2\pi$, the equation system reduces to:

$$\begin{cases} w = \cos(\frac{a}{2})\cos(\frac{\psi}{2}) - \sin(\frac{a}{2})\sin(\frac{\psi}{2}), & \text{(71a)} \\ z = \sin(\frac{a}{2})\cos(\frac{\psi}{2}) + \cos(\frac{a}{2})\sin(\frac{\psi}{2}). & \text{(71b)} \end{cases}$$

  This is similar to the Case 1. The case for $\phi = \pi$ is similar to the case $\phi = 0$ or $2\pi$, and we omit the proof for brevity.

**Step 2:** $\mathbf{SO}(3) \setminus U_e \subset \Phi(\partial\mathcal{D})$**.** For any $R \in \mathbf{SO}(3) \setminus U_e$, by the surjectivity of $\Phi$, there exists $d \in \mathcal{D}$ such that $\Phi(d) = R$. Since $R \notin U_e$, $d \in \partial D$, so that $R \in \Phi(\partial D)$. Since $\mathcal{D}$ is a 3-manifold with boundary, its boundary $\partial D$ is a manifold without boundary of dimension 2. By Lemma C.4, its image $\Phi(\partial D)$ in $\mathbf{SO}(3)$ has Lebesgue measure 0. Consequently, $\mathbf{SO}(3) \setminus U_e$ has Lebesgue measure 0, i.e. $U_e$ covers $\mathbf{SO}(3)$ almost everywhere.

□

**Corollary C.6** (Global Coverage of $U_p$)**.** *Let $U_p \subset \mathbf{SO}(3)$ be the local chart defined in Theorem 4.4. Then $U_p$ covers $\mathbf{SO}(3)$ almost everywhere.*

*Proof.* This is direct result of Theorem C.8 and the construction of $U_p = pU_e$. □

**Corollary C.7** (Global Positive Definiteness)**.** *The local Gaussian kernel $k_p$ defined in Theorem B.12 is positive definite on $\mathbf{SO}(3)$ except on a null set (Lebesgue measure 0). Consequently, it can be used as a global covariance function for Gaussian process regression on the entire group.*

*Proof.* This is a direct result of Corollary C.9. □

### C.4.3 Comparison with Global Baselines

Owing to the global coverage analysis in the previous subsection C.4.2, we conduct the Gaussian process regression on Lie group $\mathbf{SO}(3)$ using our proposed exponential coordinate kernel $k_e$ (which has been treated as a global kernel) and compare with the state-of-the-art methods including Matérn kernel on compact Riemannian manifolds [24] and Matérn kernel on compact Lie groups [32]. Both baseline methods are based on the spectral theory of the Laplace-Beltrami operator: the former utilizies eigenfunction expansions, while the latter employs explicit summation over irreducible representations via character formulas derived from the *Peter-Weyl* theorem. In practice, both baselines approximate the kernel by truncating the infinite series to a finite number of terms.

Following the same experimental protocol as in the previous subsections, we set the training size to $n = 300$, test size to $m = 100$, and noise level to $\sigma = 2\%$. For our local exponential coordinate kernel, we fix the hyperparameter pair to $\lambda_1 = \lambda_2 = 0.5$ based on the ablation study in Appendix C.3.1. For the baseline Matérn kernels, we adopt the parameter $\nu = \frac{3}{2}, \frac{5}{2}, \infty$ and set the lengthscale to 0.5, following the same configuration as their original implementations.

Figs. 15 and 16 present the prediction results of our local exponential coordinate kernel against the two baseline methods, evaluating both predictive accuracy and computational efficiency. As observed, the Matérn kernels in both baselines achieve similar predictive performance, while our proposed local Gaussian kernel yields better accuracy in terms of both RMSE and MAE. The most notable advantage of our method, however, lies in its computational efficiency: training our local Gaussian kernel takes only 0.261 second, compared to 6.933 second for the Riemannian Matérn kernel [24] and 53 second for the Lie stationary Matérn kernel [32] on the same training data. This dramatic speedup

is attributed to the closed-form expression of our local exponential coordinate kernel, whereas the implementation of Matérn kernel requires truncating the infinite series, which inevitably introduces truncation error and incurs high computational cost.

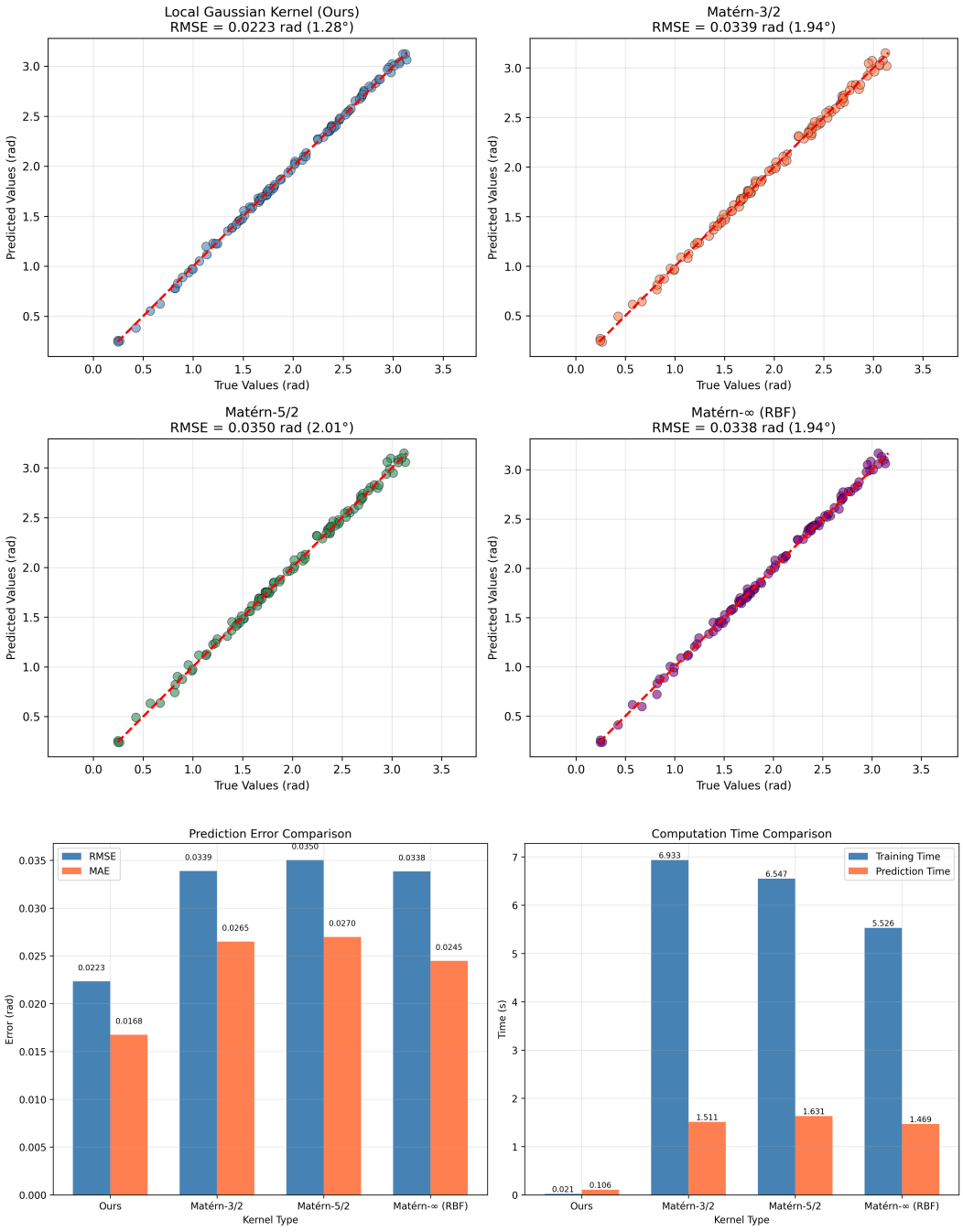



Figure 15: Comparison of our local exponential coordinate kernel and Matérn kernels [24] for Gaussian Process Regression on Lie group $\mathbf{SO}(3)$. The Matérn parameter $\nu$ is set to be $\frac{3}{2}$, $\frac{5}{2}$ and $\infty$, with training size $n = 300$. Our local exponential coordinate kernel achieves better accuracy and significant higher computational efficiency than the Matérn kernels.

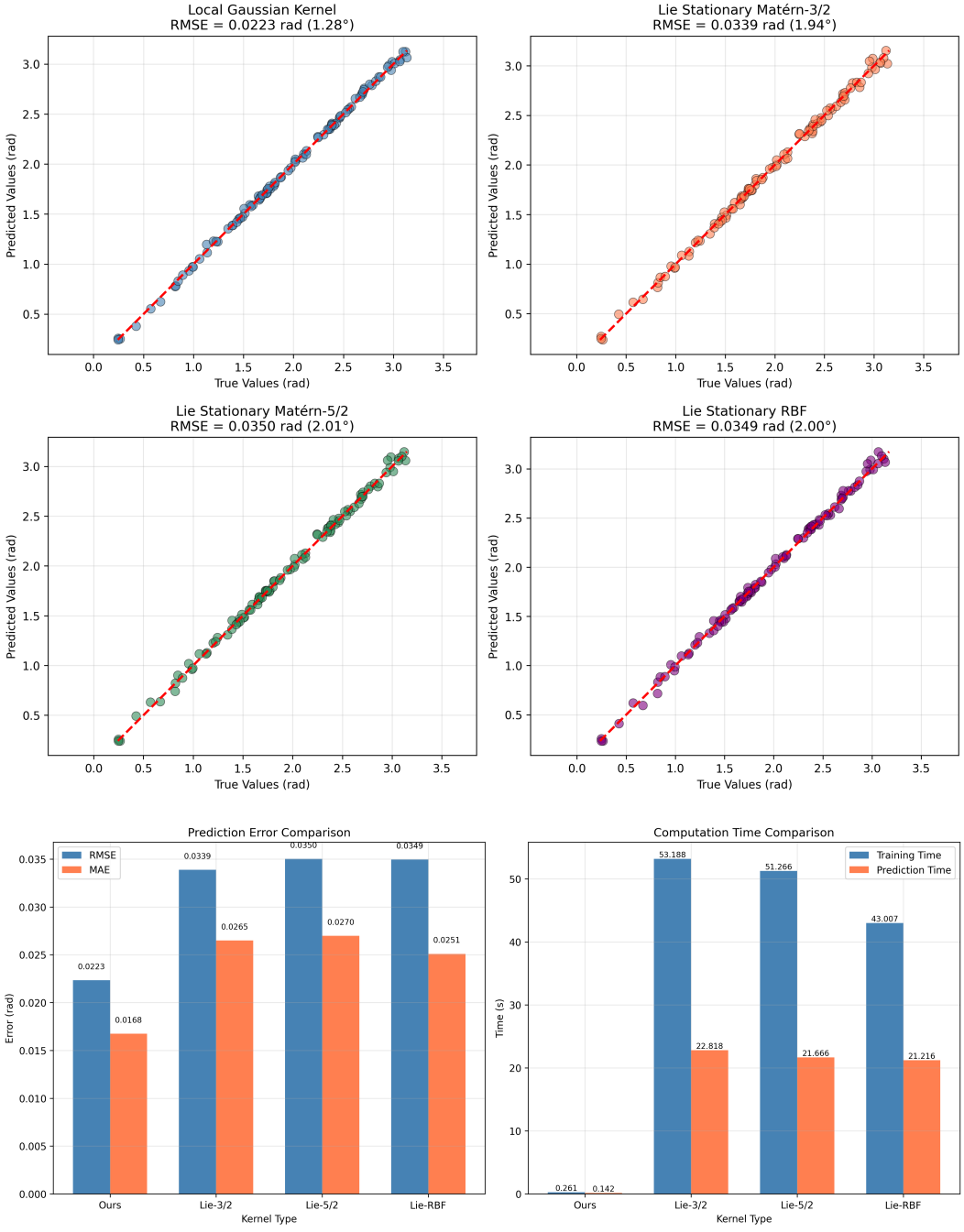



Figure 16: Comparison of our local exponential coordinate kernel and Lie stationary Matérn kernels [32] for Gaussian Process Regression on Lie group $\mathbf{SO}(3)$ with training size $n = 300$. While Lie stationary kernels rely on spectral series summation, our method employs a closed-form Gaussian kernel via torus decomposition. Results show that our approach yields both better accuracy and significantly higher computational efficiency across different test values $\nu = \frac{3}{2}, \frac{5}{2}, \infty$.

### C.5 Global Coverage Analysis and Baseline Comparison for $\mathbf{SU}(2)$

In this subsection, we first show that $U_e$ (in Theorem B.25) covers Lie group $\mathbf{SU}(2)$ almost everywhere, i.e. the local kernel $k_e$ can be treated as a global positive definite kernel on $\mathbf{SU}(2)$. We then compare our kernel with global baselines on Gaussian process regression tasks.

#### C.5.1 Theoretical Proof

**Theorem C.8** (Global Coverage of $U_e$)**.** *Let $U_e \subset \mathbf{SU}(2)$ be the chart given in Theorem B.25. Then $U_e$ covers $\mathbf{SU}(2)$ almost everywhere, i.e., its complement $\mathbf{SU}(2) \setminus U_e$ has Lebesgue measure 0.*

*Proof.* Consider the parametrization map

$$\Phi : \mathcal{D} = [-\pi, \pi] \times [0, \pi] \times [0, 2\pi] \longrightarrow \mathbf{SU}(2)$$

defined by

$$\Phi(a, \varphi, \theta) = \begin{pmatrix} \cos(\varphi)e^{ia} & -\sin(\varphi)e^{i(a-\theta)} \\ \sin(\varphi)e^{-i(a-\theta)} & \cos(\varphi)e^{-ia} \end{pmatrix}.$$

By construction, $U_e = \Phi(\mathcal{D}^\circ)$, where $\mathcal{D}^\circ$ denotes the interior of $\mathcal{D}$. We first prove that $\Phi$ is surjective. For any $U \in \mathbf{SU}(2)$, write

$$U = \begin{pmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1.$$

Let $\alpha = re^{ia}$ with $r = |\alpha| \in [0, 1]$ and $a = \arg(\alpha) \in [-\pi, \pi)$. Define $\varphi = \arccos(r) \in [0, \pi]$. If $\varphi = 0$, then $\beta = 0$ and $U = \begin{pmatrix} e^{ia} & 0 \\ 0 & e^{-ia} \end{pmatrix}$, which is obtained by taking any $\theta \in [0, 2\pi)$. If $\varphi > 0$, then from $\beta = \sin\varphi\, e^{-i(a-\theta)}$, we have $\theta = a + \arg(\beta) \pmod{2\pi}$. Thus we can find $(a, \varphi, \theta) \in \mathcal{D}$ satisfies $\Phi(a, \varphi, \theta) = U$, which proves surjectivity. For any $R \in \mathbf{SU}(2) \setminus U_e$, by the surjectivity of $\Phi$, there exists $d \in \mathcal{D}$ such that $\Phi(d) = R$. Since $R \notin U_e$, $d \in \partial D$, so that $R \in \Phi(\partial D)$. Hence $\mathbf{SU}(2) \setminus U_e \subset \Phi(\partial\mathcal{D})$. Since $D$ is a 3-dimensional manifold with boundary, its boundary $\partial D$ is a 2-dimensional smooth manifold without boundary. By Lemma C.4, $\Phi(\partial\mathcal{D})$ has Lebesgue measure zero in $\mathbf{SU}(2)$. Therefore, $\mathbf{SU}(2) \setminus U_e$ has Lebesgue measure 0, i.e. $U_e$ covers $\mathbf{SU}(2)$ almost everywhere. This completes the proof. □

**Corollary C.9** (Global Coverage of $U_p$)**.** *Let $U_p \subset \mathbf{SU}(2)$ be the local chart defined in Theorem 4.4. Then $U_p$ covers $\mathbf{SU}(2)$ almost everywhere.*

*Proof.* For any $R \in \mathbf{SU}(2) \setminus U_p$, we can express $R$ as $R = pR'$ for some $R' \in \mathbf{SU}(2)$. Since $U_p = pU_e$, $R' \notin U_e$. Hence the complement of $U_p$ in $\mathbf{SU}(2)$ is exactly the left translation image of the complement of $U_e$ in $\mathbf{SU}(2)$: $\mathbf{SU}(2) \setminus U_p = L_p(\mathbf{SU}(2) \setminus U_e)$. Since diffeomorphism preserves the null set of Lebesgue measure zero and $L_p$ is a diffeomorphism, $\mathbf{SU}(2) \setminus U_p$ has Lebesgue measure zero by Theorem C.8. □

**Corollary C.10** (Global Positive Definiteness)**.** *The local Gaussian kernel $k_p$ defined in Theorem B.12 is positive definite on $\mathbf{SU}(2)$ except on a null set of Lebesgue measure 0. Consequently, it can be used as a global covariance function for Gaussian process regression on the entire group.*

*Proof.* This is a direct result of Corollary C.9. □

#### C.5.2 Comparison with Global Baselines

Owing to the global coverage property in the previous subsection, we conduct the Gaussian process regression on Lie group $\mathbf{SU}(2)$ using our proposed exponential coordinate kernel $k_e$ (which has been treated as a global kernel) and compare with the state-of-the-art methods including Matérn kernel on compact Riemannian manifolds [24] and Matérn kernel on compact Lie groups [32] as in C.4.3. In practice, both baselines approximate the kernel by truncating the infinite series to a finite number of terms.

| Method | RMSE (°) | MAE (°) | Fit time (s) | Prediction time (s) |
|---|---|---|---|---|
| Ours | 1.15 | 0.78 | **2.758** | **0.009** |
| Matérn-$\frac{3}{2}$ | 1.14 | 0.85 | 10.225 | 0.929 |
| Matérn-$\frac{5}{2}$ | 1.24 | 0.90 | 10.211 | 1.307 |
| Matérn-$\infty$ | 0.57 | 0.46 | 9.950 | 0.882 |

Table 3: Comparison results of our exponential coordinate kernel with Matérn kernel [24] for Gaussian process regression on $\mathbf{SU}(2)$.

| Method | RMSE (°) | MAE (°) | Fit time (s) | Prediction time (s) |
|---|---|---|---|---|
| Ours | 1.15 | 0.86 | **0.014** | **0.073** |
| Matérn-$\frac{3}{2}$ | 2.61 | 1.85 | 1.457 | 1.161 |
| Matérn-$\frac{5}{2}$ | 2.22 | 1.60 | 1.325 | 1.391 |
| Matérn-$\infty$ | 0.75 | 0.59 | 1.350 | 1.292 |

Table 4: Comparison results of our exponential coordinate kernel with Lie stationary kernel [32] for Gaussian process regression on $\mathbf{SU}(2)$.

For this synthetic test, we take the rotation angle function on $\mathbf{SU}(2)$ as the target function:

$$f : \mathbf{SU}(2) \to \mathbb{R}, \quad f(R) = \arccos(\frac{\mathrm{tr}(R)}{2}). \tag{72}$$

We sample $n$ rotation matrices $\mathbf{R}_1, \ldots, \mathbf{R}_n$ from $\mathbf{SU}(2)$ as training inputs, with corresponding noisy observations $y_i = f(\mathbf{R}_i) + \epsilon_i$, where $\epsilon_i$ denotes independent and identically distributed Gaussian noise. An additional set of $m$ matrices $\mathbf{R}_1^*, \ldots, \mathbf{R}_m^*$ are sample as test inputs. The Gaussian process model is trained on the dataset $\{(\mathbf{R}_i, y_i)\}_{i=1}^n$ and then used to predict the function values at the test points. The prediction errors are evaluated by comparing the predicted values $y_j$ with the ground truth $y_j^* = f(\mathbf{R}_j^*)$ using two metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). In our experimental configuration, we set training size $n = 500$, test size $m = 100$, and noise level $\sigma = 2\%$.

Tables. 3 and 4 present the prediction results of our exponential coordinate kernel against the two baseline methods, evaluating both predictive accuracy and computational efficiency.As observed, our kernel achieves higher predictive accuracy than the Matérn kernels across both baselines and approaches the performance of the RBF (Matérn-$\infty$) kernel. The most notable advantage of our method lies in its computational efficiency: training and prediction with our exponential coordinate kernel are significantly faster than with the Riemannian Matérn kernel [24] and the Lie stationary Matérn kernel [32] on the same training data. This dramatic speedup stems from the closed-form expression of our local exponential coordinate kernel, whereas the Matérn kernels require truncating an infinite spectral series, which inevitably introduces truncation error and incurs high computational cost. We note that the two baselines employ different sampling strategies in their respective open-source implementations, which contributes to the observed differences in absolute training and prediction times between the two comparison experiments.

## D Limitations

While the proposed constructive framework is conceptually applicable to any compact Lie group, the current derivation of closed-form expressions for the local chart $U_p$ and the inverse map $\Phi_p^{-1}$ relies on the Rodrigues formula, which provides an explicit closed form for the exponential map on $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$. For higher-dimensional groups such as $\mathbf{SO}(n)$ and $\mathbf{SU}(n)$ with $n > 3$, as well as other compact Lie groups (*e.g.* exceptional groups), the exponential map does not admit an analogous simple closed form. Consequently, the product $\exp(X)\exp(Y)$ for $X \in \mathfrak{t}, Y \in \mathfrak{t}^\perp$ cannot be expressed analytically, and the inverse map $\Phi_p^{-1}$ cannot be obtained in closed form through the current approach. Extending the framework to other compact Lie groups may require alternative parameterizations of the exponential map or numerical inversion techniques.

## E Broader Impact

We present a constructive framework to build positive definite local Gaussian kernel on compact Lie groups, with closed-form expressions on $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$. By showing that the algebraic decomposition of the Lie algebra enables the construction of geodesic Gaussian kernel, our work suggests a general principle for kernel design on manifolds admitting a group structure.

We expect the primary impact to be practical. In robotics, the state spaces of most manipulators and mobile platforms involve $\mathbf{SO}(3)$ for end-effector orientation; our kernel enables Gaussian process models to incorporate geometric prior information directly on the rotation group, potentially improving model-based reinforcement learning, Bayesian optimization for grasp planning, and trajectory prediction. In computer vision, tasks such as camera pose estimation, 3D object tracking, and structure from motion rely heavily on $\mathbf{SO}(3)$ and $\mathbf{SU}(2)$ representations; our kernel provides a principled way to model uncertainty in these settings.

Beyond these immediate applications, the constructive methodology of decomposing a Lie group via its maximal torus may inspire analogous approaches for other structured non-Euclidean domains. We encourage practitioners in the physical sciences and engineering to consider how geometry-aware kernel design can improve uncertainty quantification and data efficiency in their respective fields.

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We claim the contributions of how to construct the local geodesic Gaussian kernel in the abstract and introduction parts.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please refer to Section D of the Appendix.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide the full set of assumptions for each theorem or proposition. All of their theoretical proofs are presented in the Appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide pseudocode to reproduce the implementation. We will release the code upon acceptance.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We will release the code upon acceptance.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We provide the full details of the experimental setting in the appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report the standard deviation of the minimum eigenvalue across 100 independent trials in the positive definiteness experiments C.2 and provide GP posterior standard deviations as error bars in the regression experiments C.3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments were conducted on a server equipped with an Intel Xeon Gold 6326 CPU @ 2.90GHz and an NVIDIA GeForce RTX 3090 GPU (24 GB). Training and prediction times for all methods are reported in Appendix C.4.3.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: This is a purely theoretical paper. No human subjects, sensitive data, or harmful applications are involved. We have reviewed the NeurIPS Code of Ethics and comply fully.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the broader impact, including potential applications in robotics and computer vision in Appendix E. As a foundational mathematical tool, we do not foresee negative societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper does not release any pretrained models, image generators, scraped datasets, or other assets that pose a risk of misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: This paper does not use any existing datasets or pretrained models.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The current paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.