Computationally Efficient Rigid-body Gaussian Process for Motion Dynamics

Muriel Lang and Sandra Hirche*

March 2, 2017

Abstract

In this paper we address the modeling and learning of complex non-linear rigid-body motions employing Gaussian processes. As the common procedure of using Euler angles in the Gaussian process results in inaccurate predictions for large rotations, we represent the input data by axis-angle pseudo-vectors for rotations and Euclidean vectors for translation. Our decision in favor of this representation of the special Euclidean group $SE(3)$ is due to its computational efficiency. To allow Gaussian process estimation on a non-Euclidean input domain, such as the space of rigid motions, we generalize the model by introducing novel mean and covariance functions on $SE(3)$. We prove that those functions fulfill the requirements of Gaussian processes. The proposed approach is validated on simulated and on real human motion data. Our results demonstrate significant benefits of the proposed rigid-body Gaussian process with respect to alternative variants in terms of regression performance and computational efficiency.

Keywords: Probability and Statistical Methods; Learning and Adaptive Systems; Behaviour-Based Systems;

1 INTRODUCTION

Data-driven modeling approaches become more and more popular in robotics, in particular in domains where models employing first order principles fail. This includes the modeling of human motion for human-robot interaction [1], programming by demonstration [2], and many other. Gaussian process (GP) regression, Gaussian mixture models (GMM) and neural networks (NNs), especially recurrent neural networks (RNNs), have been employed for data-driven modeling in continuous-time space. GP regression is a powerful nonparametric Bayesian regression method that has become increasingly popular for modeling system dynamics [3]. The objective of this paper is to model and learn complex rigid motion dynamics from mere observations. We propose to employ GP regression for its suitability to model non-linear mappings and for its natural ability to model predictive conditional probabilities including a best estimator and a prediction confidence. This uncertainty estimate depending on the distance to the training data set is highly valuable in safety-critical applications such as human-robot interaction [4]. Nevertheless, a major drawback of this model is that it is defined in Euclidean space. Even though a formulation with non-Euclidean input space is possible in principle, the traditional formulation

*Both authors are with the Chair of Information-oriented Control, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany {muriel.lang, hirche}@tum.de

requires a Euclidean vector space structure on the input as well as the output space. As for rotations there exists no representation in Euclidean space, they cannot be modeled accurately. The common procedure, see e.g. [5], is to use the Euler angles representation for orientations, as close to zero their space is almost Euclidean. However, when the rotation is large, for example at high angular speed, low sampling frequency or if training data is sparse, this approach leads to inaccurate predictions. Therefore, we generalize in the present paper the GP to incorporate input domains within the non-Euclidean space of rigid motions $SE(3)$. As a shortcoming of the GP model in generals is its computational expense, we seek for a computationally efficient representation.

So far, only few works consider the problem of generalizing the GP model to non-Euclidean input space, even though it is crucial for achieving accurate estimation results. As in GP modeling an essential part concerns the definition of kernel functions over spaces different to the Euclidean space, we start by reviewing related work on this topic. Gaussian kernels for a Hilbert space embedding are investigated in [6]. However, a manifold metric does not necessarily result in a valid kernel function, e.g. the kernel from the Stein-divergence. Sufficient and necessary conditions to obtain valid Gaussian kernels on metric spaces is introduced in [7], which we will use to develop the kernel function in our rigid-body GP. Also few works are available on GPs variants on non-Euclidean spaces. An embedding of latent variables into non-Euclidean space is presented in [8], considering topologically-constrained latent variable models. However, as only locally linear embeddings are considered, the standard GP structure is maintained. Another related approach, presented by [9], introduces a geodesic GP. The GP model is used to train a surface embedding in Euclidean space. That is to say, the GP is defined in Euclidean space but maps geodesic distances to a Euclidean equivalent. A manifold GP [10] models jointly two composed functions $m \circ g$, where m is a deterministic mapping from an abstract manifold M into a Euclidean feature space and g the GP regression task. The approach aims to model complex and non-differentiable functions, for which the smoothness assumptions on standard kernel functions are unsuitable. In contrast to this approach where both mappings m and g are abstract, we propose an alternative approach in which a concrete manifold is given. In our preliminary work [11] and [3] another valid GP enhancement to the special Euclidean group $SE(3)$ is introduced, namely the GP over dual quaternions. However, the use of quaternions harbors disadvantages, such as increased computational complexity and training failures due to the topology of quaternions that induces sensitive reaction to the hyperparameter choice.

The contribution of this paper is a novel computationally efficient and robust rigid-body GP suitable to model motion dynamics in the special Euclidean group $SE(3)$ in a rigorous way. In contrast to previous work, we propose the representation of rigid motions by axis-angle and Euclidean vectors to develop a GP model, that generalizes to non-Euclidean input space. We present novel mean and kernel functions and provide a proof for the important generalized squared exponential kernel to be valid. The proposed approach outperforms comparable GP variants at large in computational efficiency and accuracy as we demonstrate experimentally.

The paper is structured as follows. The general modeling approach is presented in Sec. 2, before the rigid-body GP is developed in Sec. 3, which involves the introduction of the rigid motion representation, the mean and kernel functions and a comparison with an alternative GP generalization, the GP over dual quaternions. An experimental evaluation is provided in Sec. 4.

2 MODELING APPROACH

In the following, we briefly describe the challenges of modeling rigid-body motion dynamics by Gaussian processes (GPs) and our approach using axis-angle pseudo-vector and Euclidean vector for rotation and translation representation, respectively. Compactly stated, *a Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution* [12] and a precise, more explicit formulation is given in the following.

Definition 1. *Let $\mathcal{X}$ be a (multidimensional) index set, and denote by $\{\varphi(\mathbf{x})\}_{\mathbf{x}\in\mathcal{X}}$ a real-valued stochastic process over $\mathcal{X}$. Such a process is called Gaussian, if and only if any finite collection of random variables $\{\varphi(\mathbf{x}_1), \ldots, \varphi(\mathbf{x}_\nu)\}$, where $\nu \in \mathbb{N}$, is ν-dimensional multivariate Gaussian.*

We call $\varphi(\mathbf{x})$ distributed according to a GP,

$$\varphi(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}),\, k(\mathbf{x},\, \mathbf{x}')), \tag{1}$$

for a mean function $m(\mathbf{x})$ and a kernel function $k(\mathbf{x},\, \mathbf{x}')$ such that

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}\left[\varphi(\mathbf{x})\right], \\ k(\mathbf{x},\, \mathbf{x}') &= \mathbb{E}\left[(\varphi(\mathbf{x}) - m(\mathbf{x}))\,(\varphi(\mathbf{x}') - m(\mathbf{x}')\right]. \end{aligned} \tag{2}$$

GP characteristics A GP model returns for any input value $\mathbf{x}^* \in \mathcal{X}$ an estimate for $\varphi(\mathbf{x}^*)$ in form of a Gaussian distribution $\mathcal{N}(\mu_{\mathbf{x}^*}, \sigma_{\mathbf{x}^*})$ and Gaussian distributions require an underlying Euclidean vector space structure to be correctly normalized. Intuitively, the terms distance, angle, vector, addition and inner product need to be defined (inducing a flat space) for that the normalization factor of a Gaussian correctly scales the probability measure. As in non-Euclidean curved space density is not uniform, the normalization is corrupted and thus, no Gaussian probability is defined.

While Def. 1 implies no requirement to the domain $\mathcal{X}$ of function $\varphi(\cdot)$, despite the need to allow the definition of a valid mean function $m(\cdot)$ and kernel function $k(\cdot,\cdot)$, the target set $\varphi(\mathcal{X})$ is required to possess a Euclidean vector space structure and be scalar. Thus, set $\varphi(\mathcal{X})$ is isomorphic to $\mathbb{R}$. To consider vector-valued stochastic processes $f = (\varphi_1, \ldots, \varphi_n)$, the common procedure is to model each component φ_l, $l = 1, \ldots, n$, by a Gaussian process $\varphi_l(\mathbf{x}) \sim \mathrm{GP}(m_l(\mathbf{x}),\, k_l(\mathbf{x},\, \mathbf{x}'))$ with the same classes of mean and kernel functions for all l. In many applications the mean function is set to zero, $m_l(\mathbf{x}) \equiv 0_{\varphi(\mathcal{X})}$ as this reduces the computational expense without limiting the expressive power of the process [13, Chap. 6.4.1]. Hence, the heart piece of GP modeling concerns the definition of the kernel function. Apart from few exceptions, the kernel functions can be divided in two classes: stationary and dot-product kernels. So far, in both kernel function classes the vast majority assumes also domain $\mathcal{X}$ to equal $\mathbb{R}^n$. We will provide one basic kernel for each class of kernel functions as a baseline from which more kernels can be constructed.

Modeling rigid-body dynamics Here we employ GP regression to model non-linear rigid-body motion dynamics. By considering rigid bodies only, the possible movements consist of rotation and translation, which together generate the special Euclidean group $SE(3)$, the space of rigid motions $\mathfrak{g}$. We consider the specification in continuous time via a mapping from state space to the space of its time

derivatives. Thus, the rigid-body dynamics are given by

$$f\colon SE(3)) \to TSE(3), \tag{3}$$

a mapping from the rigid motions $\mathfrak{g} \in SE(3)$ to the according velocity space $TSE(3)$, which is described by a tangent bundle

$$TSE(3) = \{(\mathfrak{g}, \dot{\mathfrak{g}}) |\, \mathfrak{g} \in SE(3),\ \dot{\mathfrak{g}} \in T_{\mathfrak{g}}SE(3)\}\,, \tag{4}$$

consisting of pairs of a rigid motion $\mathfrak{g}$ and corresponding time derivative $\dot{\mathfrak{g}}$ in the tangent space $T_{\mathfrak{g}}SE(3)$. We further assume that the observation of the velocity is disturbed by a noise process in the tangent space, i.e. $\mathbf{n}_{\mathfrak{g}} \in T_{\mathfrak{g}}SE(3)$, and we obtain noisy measurements of the velocity through the mapping

$$\tilde{f}\colon SE(3) \times T_{\mathfrak{g}}SE(3) \to TSE(3). \tag{5}$$

Our objective is to employ GP regression for finding an approximation of the unknown dynamics (3). Advantageously, in this formulation the output space of the GP is a tangent bundle $TSE(3)$, which provides a Euclidean vector space structure. In consequence, if we find valid mean and kernel functions over the rigid motion space $SE(3)$, the GP output in form of a Gaussian distribution is well-defined, and therefore also the rigid-body GP. For the proposed representation of rigid-body motions a discrete-time integration is available, allowing the representation in form of a discrete-time dynamics by composing dynamics (3) and the integration.

3 RIGID-BODY GAUSSIAN PROCESS

In this section, we introduce our approach to generalize GPs to modeling rigid-body motion dynamics. After introducing the rigid motion representation by axis-angle and Euclidean vector, we review GP model learning and prediction to point out the requirements for the aspired GP generalization. We see it is sufficient to define valid mean functions and kernel functions to obtain the generalization to 6-DoF motion dynamics. Thereafter, we introduce a set of fundamental functions containing the zero mean and a dot-product kernel as well as a stationary kernel function.

3.1 Rigid Motion Representation

From Euler's fixed point theorem (1776) [14] we know that in 3D space any rigid-body displacement with one fixed point, i.e. any composition of rotations, can be equivalently described by a single rotation about some non-trivial axis of rotation through the fixed point. This axis is also called Euler axis. Thus, the set of unit length Euler axes $\mathbf{u}$ together with a rotation angle θ parametrizes the rotation group, called special Orthogonal group $SO(3)$

$$SO(3) \subset \left\{\theta\mathbf{u} \in \mathbb{R}^3 \middle|\, \|\mathbf{u}\| = 1 \wedge \theta \in [0, \pi]\right\}. \tag{6}$$

The set given in (6) defines the solid ball $B_\pi(0)$ in $\mathbb{R}^3$ with radius $0 \leq r \leq \pi$ centered around the origin and is thus closed, dense and compact. Ambiguity in the representation occurs for $\theta = \pi$, as $\pi\mathbf{u} = -\pi\mathbf{u}$ define the same rotation. To obtain an isomorphism between the rotation group $SO(3)$ and the axis-angle

representation, we additionally fix the Euler axis representation for $\theta = \pi$ and obtain

$$\begin{aligned}\tilde{B}_\pi(0) := B_\pi(0) \setminus \big\{\pi\mathbf{u}\big|\, \mathbf{u}_z < 0 \vee (\,\mathbf{u}_z = 0 \wedge\, \mathbf{u}_y < 0) \\ \vee (\,\mathbf{u}_z = \,\mathbf{u}_y = 0 \wedge\, \mathbf{u}_x < 0)\big\}.\end{aligned} \tag{7}$$

This parametrization of the rotation group by a Euler axis and a rotation angle is a minimal and unique, $SO(3) \simeq \tilde{B}_\pi(0)$.

The special orthogonal group $SO(3)$ obtains its group structure from the operation composition of rotations. As the axis-angle representation only defines a pseudo-vector, none of the vector operations addition and multiplication is suitable for rotation composition. Instead, we define the rotation composition of two rotations $\theta_1\mathbf{u}_1$ and $\theta_2\mathbf{u}_2$ in $\tilde{B}_\pi(0)$, inspired by unit quaternion multiplication,

$$\theta_2\mathbf{u}_2 \circ \theta_1\mathbf{u}_1 = \Bigg(\underbrace{2\,\mathrm{acos}(\big|a - bc\big|)}_{\theta_3}\ \underbrace{\frac{\sqrt{1-a^2} + b(\,\mathbf{u}_2 \times\, \mathbf{u}_1)}{\sqrt{1-(a^2-bc)^2}}}_{\mathbf{u}_3}\Bigg), \tag{8}$$

where we substitute $a = \cos\left(\frac{\theta_2+\theta_1}{2}\right)$, $b = \sin\left(\frac{\theta_2}{2}\right)\sin\left(\frac{\theta_1}{2}\right)$ and $c = 1 + \,\mathbf{u}_2\,\mathbf{u}_1$.

In the proposed GP generalization, we seek to model a dynamics as given in (5), mapping from the rigid motions into the velocity space. In many control applications, however, it is of interest to generate a prediction of the motion including uncertainty certificates in $SE(3)$. This requires the integration of the velocity signal and the corresponding transformation of uncertainty, which is not trivial in general, as the space of rigid motions $SE(3)$ is non-Euclidean as soon as rotation is present. Consequently, the Gaussian describing the uncertainty in velocity space, has to be transformed to a curved space, where probability distributions inherently are not defined.

In the following, we discuss how the exponential function is employed to integrate predictive angular velocities to the system states. $SO(3)$ is a compact Lie group with corresponding Lie algebra $\mathfrak{so}(3)$, the set of skew-symmetric matrices

$$\mathfrak{so}(3) = \left\{\Omega \in \mathbb{R}^{3\times 3}\middle|\Omega^\top = -\Omega\right\}. \tag{9}$$

It is well-known, see [15, Chap. 3.7], that for Lie groups the exponential map $\exp : \mathfrak{so}(3) \to SO(3)$ pulls back elements to the special Orthogonal group. Given any angular velocity $\omega \in \mathbb{R}^3$, where $\theta_\omega = \|\omega\|$ and $\mathbf{u}_\omega = \frac{\omega}{\|\omega\|}$, the corresponding skew-symmetric matrix

$$[\omega]_\times = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \tag{10}$$

is pulled back to an element $\boldsymbol{R}$ in the special Orthogonal group $SO(3)$ using

$$\boldsymbol{R} = \exp([\theta_\omega \mathbf{u}_\omega]_\times). \tag{11}$$

From [16] we know that (11) can be calculated using Rodrigues' rotation formula

$$\boldsymbol{R} = \cos\theta_\omega \boldsymbol{I} + \sin\theta_\omega [\,\mathbf{u}_\omega]_\times + (1 - \cos\theta_\omega)\,\mathbf{u}_\omega\,\mathbf{u}_\omega^\top, \tag{12}$$

where $\boldsymbol{I}$ denotes the identity matrix. This induces that in axis-angle space the exponential function acts locally as identity on the angular velocities $\omega \in \mathfrak{so}(3)$, what formally means $\exp(\omega) = \exp(\theta_\omega \mathbf{u}_\omega) = \theta_\omega \mathbf{u}_\omega \in SO(3)$. Consequently, a Gaussian, representing the velocity prediction uncertainty, which is defined in the tangent space, describes equally the uncertainty over rotations in $SO(3)$ for a small neighborhood U around $\theta_\omega \mathbf{u}_\omega$. Outside the neighborhood the variance is still approximated by the Gaussian, but with increasing Euclidean distance, the approximation quality decreases rapidly. Hence, we can use the exponential function to integrate in discrete-time from the tangent bundle $TSE(3)$ to the special Euclidean group $SE(3)$. Thus, we obtain discrete-time dynamics by composing dynamics (3) with the integration, $\exp \circ f$.

Since required for the generalized GP kernel definition, we additionally introduce the dot product and a distance measure between rotations for the axis-angle representation. The dot product of two rotations in axis-angle is defined as the regular scalar product in vector spaces,

$$\langle \theta_1 \mathbf{u}_1, \theta_2 \mathbf{u}_2 \rangle = |\theta_1| \, |\theta_2| \cos \sphericalangle \, \mathbf{u}_1 \mathbf{u}_2. \tag{13}$$

As distance function we propose to define the length of the geodesic between rotations. For rotation matrices and unit quaternions such arc distance functions are used by [11] and [17]. Moreover, [16] and [3] prove that for both representations the arc distances define metrics. The arc distance for axis-angle can be derived from unit quaternions considering that the angle between unit quaternions is half the angle between rotations, because unit quaternions are a double coverage of the rotation group $SO(3)$. Hence, we obtain as metric

$$d_{\text{arc}}(\theta_1 \mathbf{u}_1, \theta_2 \mathbf{u}_2) = 2 \operatorname{acos} \left| \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} + \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \, \mathbf{u}_1^\top \mathbf{u}_2 \right|. \tag{14}$$

In the following, we present our approach to combine rotation with translation, which we represent by Euclidean vectors $\mathbf{v} \in \mathbb{R}^3$. As $\tilde{B}_\pi(0)$ is a unique rotation representation, the link with translations can be realized using the standard Cartesian set product. The spaces of rotation $\tilde{B}_\pi(0)$ and translation $\mathbb{R}^3$ jointly define isomorphically the special Euclidean group

$$SE(3) \simeq \tilde{B}_\pi(0) \times \mathbb{R}^3 \tag{15}$$

and any rigid motion $\mathfrak{g} \in SE(3)$ can be equivalently represented by $\mathfrak{g} \equiv (\theta \mathbf{u}, \, \mathbf{v}) \in \tilde{B}_\pi(0) \times \mathbb{R}^3$.

The group structure on $SE(3)$ is inherited from the group structure on the subspaces, namely composition. As translation composition in $\mathbb{R}^3$ is defined via vector addition, we obtain as composition operation of rigid motions in $SE(3)$

$$(\theta_3 \mathbf{u}_3, \, \mathbf{v}_3) = \big(\theta_2 \mathbf{u}_2 \circ \theta_1 \mathbf{u}_1, \theta_2 \mathbf{u}_2 (\, \mathbf{v}_1) + \, \mathbf{v}_2 \big), \tag{16}$$

where $\theta_2 \mathbf{u}_2 \circ \theta_1 \mathbf{u}_1$ as defined in (8) and $\theta_2 \mathbf{u}_2 (\, \mathbf{v}_1)$ is obtained using Rodrigues rotation formula

$$\theta_2 \mathbf{u}_2 (\mathbf{v}_1) = \mathbf{v}_1 \cos \theta_2 + \sin \theta_2 (\mathbf{u}_2 \times \, \mathbf{v}_1) + \, \mathbf{u}_2 (\mathbf{u}_2^\top \mathbf{v}_1)(1 - \cos \theta_2) \tag{17}$$

for translation vectors $\mathbf{v}_i \in \mathbb{R}^3$, $i = 1, 2$.

The dot product on $SE(3)$ with the representation by axis-angle pseudo-vector and translation vector

is defined as in regular vector spaces,

$$\langle\, \mathfrak{g}_1,\, \mathfrak{g}_2\rangle = \langle \theta_1 \mathbf{u}_1, \theta_2 \mathbf{u}_2\rangle + \langle\, \mathbf{v}_1,\, \mathbf{v}_2\rangle. \tag{18}$$

As distance function on $SE(3)$ we define the root over a sum of squares, inspired by [11], but include a convex combination of weights, $\sum_i \rho_i = 1$, where $\rho_i \geq 0$, to allow an application dependent scaling between rotation and translation, as we know from [17] that any distance metric in $SE(3)$ will ultimately depend on a choice of length scale. Together, we define as distance function

$$d(\, \mathfrak{g}_1,\, \mathfrak{g}_2) = \sqrt{\rho_1 \left[d_{\text{arc}}(\theta_1 \mathbf{u}_1, \theta_2 \mathbf{u}_2)\right]^2 + \rho_2 \|\, \mathbf{v}_1 - \, \mathbf{v}_2\|^2}. \tag{19}$$

The distance function (19) allows for incorporation of domain knowledge. To give an intuitive example, one can imagine the divergence in motion dynamics of a spinning top and a large truck. While the cost for a 2π rotation of a spinning top may be in the range of millimeters in translation, a truck may cover large distances easier than a full rotation. This dynamics difference should be reflected in the similarity measure between poses, as application specific weighting of rotation and translation increases the regression performance of the GP.

3.2 GP Model Learning and Prediction

In the following we provide an intuitive understanding of the principle of GP modeling and discuss the requirements for a GP generalization to estimate rigid-body dynamics.

Model training Model training ultimately refers to optimizing the parameters of the mean and the kernel function. We consider a smooth mapping $\check{f} : \mathcal{X} \to \mathcal{V}$, with $\check{f}(\, \mathbf{x}) = \dot{\mathbf{x}}$ and a training data set $\mathcal{D}_\nu = \{(\, \mathbf{x}_i,\, \dot{\mathbf{x}}_i)\}_{i=1}^{\nu}$ consisting of n-dimensional input-output pairs. Per output dimension $l = 1, \ldots, n$, a set of GP hyperparameters h_l, configuring the pre-identified classes of mean function $m_l(\cdot)$ and kernel function $k_l(\cdot, \cdot)$, is adjusted to approximate the rigid-body dynamics $\check{f}$. Optimal accuracy, given the input values $\mathbf{x}_i$, is obtained for the dimension-wise globally maximum likelihoods of the observations $\dot{\mathbf{x}}_i$. As both functions, mean and kernel, solely depend on the GP domain, see (2), it consequently suffices to provide valid mean functions and kernel functions per dimension l, defined on that domain $\mathcal{X}$ to generalize GP model training to a non-Euclidean domain.

Prediction For any new value $\mathbf{x}_{\nu+1}$ the GP returns a predictive Gaussian probability distribution defined in the target set $\mathcal{V}_l$ for each component of $\dot{\mathbf{x}}_{\nu+1}$, namely $\mathcal{N}(\mu_l,\, \sigma_l)$ for index $l = 1,\, \ldots,\, n$. Mean μ_l and variance σ_l of the l-th component is calculated analytically,

$$\begin{aligned} \mu_l &= m(\, \mathbf{x}_{\nu+1}) + \mathbf{K}_* \mathbf{K}^{-1} \begin{pmatrix} (\, \dot{\mathbf{x}}_1)_l - m(\, \mathbf{x}_1) \\ \vdots \\ (\, \dot{\mathbf{x}}_\nu)_l - m(\, \mathbf{x}_\nu) \end{pmatrix} \\ \sigma_l &= k(\, \mathbf{x}_{\nu+1},\, \mathbf{x}_{\nu+1}) - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^\top, \end{aligned} \tag{20}$$

where the Gram matrix $\mathbf{K}$ is obtained by $(\mathbf{K})_{ij} = k(\, \mathbf{x}_i,\, \mathbf{x}_j)$ for $i, j \in \{1, \ldots, \nu\}$ and each entry of the row vector $(\mathbf{K}_*)_j = k(\, \mathbf{x}_{\nu+1},\, \mathbf{x}_j)$. Thus, the necessary and sufficient requirement for the output space $\mathcal{V}$ is to allow the definition of a predictive Gaussian distribution, which in turn requires a real

underlying vector space structure to be correctly normalized. Due to the structure of dynamics (3), the requirement is fulfilled and the predictive output is defined.

3.3 Mean and Kernel Functions

In the following we introduce a set of valid mean and kernel functions defined on the special Euclidean group $SE(3)$. For modeling dynamics of form (3) we know from (2) that each mean and kernel function has the form

$$\begin{aligned} m\colon SE(3) &\to TSE(3) \\ k\colon SE(3) \times SE(3) &\to \mathbb{R}. \end{aligned} \tag{21}$$

For the mean function $m(\cdot)$ no further requirements need to be satisfied. In the special Euclidean group, the zero mean maps any input $\mathfrak{g} \in SE(3)$ to the zero element of the tangent bundle $TSE(3)$, i.e. $m(\mathfrak{g}) \equiv 0_{T_{\mathfrak{g}}SE(3)}$. Kernel functions encode the correlation between the elements of the dynamics input domain $SE(3)$ and constitute the essential part of GP modeling. A kernel is required to be symmetric and positive semi-definite to define a valid kernel function. As in literature various strategies are available to construct more elaborate kernel functions by composition from elementary kernels, e.g. see [18], we focus on introducing one representative per kernel class.

Dot-product kernel From [12] we know that the dot-product kernel defines a valid kernel function for elements $\mathbf{x}_i,\ \mathbf{x}_j \in \mathbb{R}^n, n \in \mathbb{N},\ \forall i, j$. In our setting, we seek to model rigid-body motion dynamics, where the rigid motions $\mathfrak{g} \in SE(3)$ are represented by axis-angle pseudo-vectors concatenated with Euclidean translation vectors. That means, $\mathfrak{g} \equiv (\theta\mathbf{u},\ \mathbf{v}) \in \tilde{B}_\pi(0) \times \mathbb{R}^3$. In this space, the dot product (18) describes the same operation as the vector space dot product in $\mathbb{R}^6$. Thus it follows,

$$k_{\text{dot}}(\,\mathfrak{g}_i,\,\mathfrak{g}_j) = \langle\,\mathfrak{g}_i,\,\mathfrak{g}_j\rangle + \sigma_0^2\delta_{ij}, \tag{22}$$

with hyperparameter $\sigma_0^2 \geq 0$, defines a valid kernel function on the rigid motions $\forall i, j$. Throughout the paper, we use the standard formulation, see [12], of introducing a noise term in the kernel function as required for (5), where the δ-function is defined by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{23}$$

Stationary kernel One of the most important stationary kernels is the squared exponential kernel. In the following we show that the kernel based on the distance function (19) is indeed a valid kernel.

Theorem 1. *The function* $k_{se} : SE(3) \times SE(3) \to \mathbb{R}_0^+$,

$$k_{se}(\,\mathfrak{g}_i,\,\mathfrak{g}_j) := \sigma_f^2\ \exp\left(-\frac{[d(\,\mathfrak{g}_i,\,\mathfrak{g}_j)]^2}{2\lambda^2}\right) + \sigma_n^2\delta_{ij}, \tag{24}$$

where we consider a distance function d *defined by (19) and positive hyperparameters* $h = (\lambda, \sigma_f, \sigma_n)$*, namely length-scale* $\lambda > 0$*, signal variance* $\sigma_f > 0$ *and signal noise* $\sigma_n > 0$*, defines a valid kernel function.*

Proof. We know that the finite sum of kernels provides a kernel [13, Chap. 6.2] and it follows from definition that $\sigma_n^2 \delta_{ij} \geq 0$ defines a valid kernel. Hence, we can focus on the first summand in the reminder of this proof. Introducing (19) in the noise-free squared exponential kernel, we can rewrite (24) neglecting the noise term $\sigma_n^2 \delta_{ij}$ as

$$\underbrace{\sigma_f^2 \exp\left(-\frac{\rho_1 [d_{\text{arc}}(\theta_i \mathbf{u}_i, \theta_j \mathbf{u}_j)]^2}{2\lambda^2}\right)}_{\bigstar} \underbrace{\exp\left(-\frac{\rho_2 \| \mathbf{v}_i - \mathbf{v}_j \|^2}{2\lambda^2}\right)}_{\clubsuit}. \tag{25}$$

A proof for the standard squared exponential kernel $\clubsuit$ to be valid is provided in [13, Chap. 6.2]. As any kernel scaled by a positive constant and any finite product of kernels yields each a new valid kernel [18], it suffices to show that also $\bigstar$ defines a valid kernel, to complete the present proof. To formally show that $\bigstar$ is a kernel, some mathematical formalism is required and we therefore need a new term, namely cpd:

Definition 2 ([19])**.** *Let $\mathcal{X}$ be a non-empty set. A real-valued symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called conditional positive definite (cpd) kernel function, if and only if the Gram matrix $\mathbf{K} \in \mathbb{R}^{\nu \times \nu}$ satisfies $\mathbf{c}^\top \mathbf{K} \mathbf{c} \geq 0$ for any vector $\mathbf{c} \in \mathbb{R}^\nu$ with $\mathbf{c}^\top \mathbf{1} = 0$, where $\mathbf{1} = (1, \ldots, 1)^\top$.*

We utilize the cpd property in $\bigstar$. A kernel k being cpd induces that $\exp(ck)$ defines a valid kernel function $\forall c > 0$ [19]. Hence, it suffices to show that the squared negative distance function

$$k_{\text{snd}}(\theta\, \mathbf{u}, \theta' \mathbf{u}') := -\left([d_{\text{arc}}(\theta\, \mathbf{u}, \theta' \mathbf{u}')]^2\right) \tag{26}$$

defines a cpd kernel. To derive that, we transform the orientation representation to unit quaternions, obtained from the axis-angle representation by $\mathbf{q} = \pm(\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\, \mathbf{u})$. Then, introducing (14) into (26), we obtain

$$k_{\text{snd}}(\mathbf{q}, \mathbf{q}') = -4\left(\operatorname{acos}\left|\mathfrak{Re}(\mathbf{q}\overline{\mathbf{q}}')\right|\right)^2 \tag{27}$$

where $\mathfrak{Re}$ denotes the real part of a quaternion and $\overline{\mathbf{q}}$ the quaternion conjugate. Let us substitute $\mathfrak{z} := \left|\mathfrak{Re}(\mathbf{q}\overline{\mathbf{q}}')\right|$. It is well-known, that $\operatorname{acos}(\mathfrak{z}) = \frac{\pi}{2} - \operatorname{asin}(\mathfrak{z})$. Hence, we can rewrite

$$k_{\text{snd}}(\mathbf{q}, \mathbf{q}') = -\pi^2 + 4 \operatorname{asin}(\mathfrak{z})\,(\pi - \operatorname{asin}(\mathfrak{z}))\,. \tag{28}$$

From [19] we know that any constant $c \in \mathbb{R}$ is cpd, a valid kernel always is also cpd and the finite sum of cpd kernels is cpd. Thus, if we can prove that both auxiliary functions $g_1(\mathfrak{z}) = \pi - \operatorname{asin}(\mathfrak{z})$ and $g_2(\mathfrak{z}) = \operatorname{asin}(\mathfrak{z})$ for $\mathfrak{z} \in [-1, 1]$, define valid kernels, it follows that (28) is cpd, and therewith (27) and (26), which would conclude the proof. In the remaining part of the proof, we analyze the auxiliary functions $g_1(\mathfrak{z})$ and $g_2(\mathfrak{z})$. Our aim is to apply [19, Theorem 4] stating that a function $f(\langle \mathbf{q}, \mathbf{q}' \rangle)$ defined on the unit sphere, $|\mathbf{q}| = |\mathbf{q}'| = 1$, in an infinite dimensional Hilbert space is a valid kernel if and only if its Taylor series expansion has only non-negative coefficients. It is well known, that the coefficients of asin in infinite series writing are non-negative and that $-\operatorname{asin}(\mathfrak{z}) = \operatorname{asin}(-\mathfrak{z})$ holds. To cope with the minus in function $g_1(\mathfrak{z})$, we re-substitute $-\mathfrak{z}$. The real part of a quaternion product is calculated

$$\mathfrak{Re}(\mathbf{q}\overline{\mathbf{q}}') = \langle \pm\mathbf{q}, \overline{\mathbf{q}}' \rangle. \tag{29}$$

Taking advantage of the fact that opposing unit quaternions $\pm\mathbf{q}$, represent the same orientation, $\pm\mathbf{q}$

and $\overline{\mathbf{q}}'$ can be chosen to lie on the same hemisphere of S_3, such that we obtain a positive value for $\langle \pm\, \mathbf{q}, \overline{\mathbf{q}}' \rangle$. Hence, we are allowed to apply [19, Theorem 4] and the proof is complete. □

3.4 Comparison of GP Variants

Besides the rigid-body GP over axis-angle and translation vector, which we investigated so far, another GP generalization suitable for modeling rigid motions in $SE(3)$ is available, namely the GP over dual quaternions [3, 11]. The underlying idea of the GP over dual quaternions is similar in a sense that dynamics are also represented from the space of rigid motions to its time derivatives. However, the topological structure of the rotation representation using unit quaternions on S_3 is different: In contrast to our rotation representation on a solid ball $\tilde{B}_\pi(0)$, defined in (7), the unit quaternions lie on the unit hypersphere S_3. They are combined with a translation vector to a dual quaternion, representing a full rigid motion at once. While the representation via dual quaternions does not allow for weighting, direct access to the weighting factors ρ_1, ρ_2 is provided by the distance function (19).

Another difference is that the topological structure of quaternions requires a more complex integration than our novel GP generalization: In the GP over dual quaternions, the integration is handled by an operation that projects values from the velocity space to the manifold of rigid motions $SE(3)$. In addition, a computationally demanding tangent space change operation is necessary, to disambiguate the tangent bundle representative, which becomes redundant in the novel rigid-body GP over axis-angle and Euclidean vector, as our integration via exponential function locally acts as identity, see (11) and below. Further, even though both GP variants use geodesic distances, differences in the definition of the metrics result in different GP training behavior, as the distance function affects the correlation between rigid motions. In the experimental evaluations we demonstrate that the GP over axis-angle and translation vector exhibits in the training phase a more robust behavior towards the initial choice of hyperparameters.

The model training complexity of both GP variants is equal to the standard GP complexity of $O(n^3)$, where n is the number of training samples, which is due to inversion of an $n \times n$ covariance matrix. With a fixed training data size, the GP prediction complexity for all 3 variants is $O(n)$, where n is the number of predicted points. In addition, we analyze the computational demand by counting floating-point operations (flops). Given a training set of n samples, the required number of flops for calculating the covariance matrix using the squared exponential kernel function is $18n^2 + 36n + 16$ for the standard GP, $24n^2 + 85n + 18$ for the GP over dual quaternions and $22n^2 + 26n + 15$ for the novel rigid-body GP, respectively. This results in significant runtime improvement, in particular as the number of training samples increases, which is demonstrated in the next section.

4 EXPERIMENTAL RESULTS

In this section we evaluate the proposed approach of a rigid-body GP in three experiments; first regarding runtime of the GP prediction, second regarding accuracy in the rotation estimation and third regarding its suitability in a real world application. We compare our results with the GP over dual quaternions and with the often applied standard GP in Euclidean space, where rotations are represented inaccurately as Euler angles. All experiments were implemented in MATLAB using the GPML toolbox [20] on a commercially available PC with intel core i5-6360U processor and 8 GB RAM. The GPML toolbox was extended to additionally comprise the generalized GP over axis-angle and translation vector and the GP over dual quaternions. We focus on the generalized squared exponential kernel (24), as for this kernel

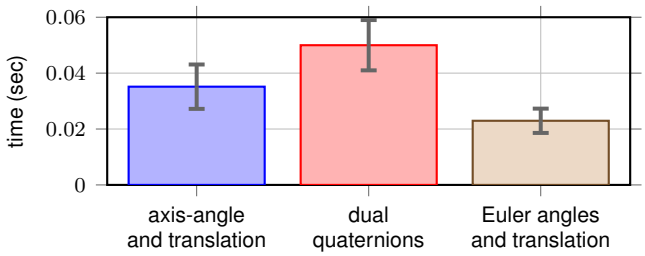

Figure 1: Average GP velocity prediction runtime for 10001 unknown rigid-body motions $\mathfrak{g}$. Per model the box height shows the average required time and the gray error bar the standard deviation over 1000 repetitions.

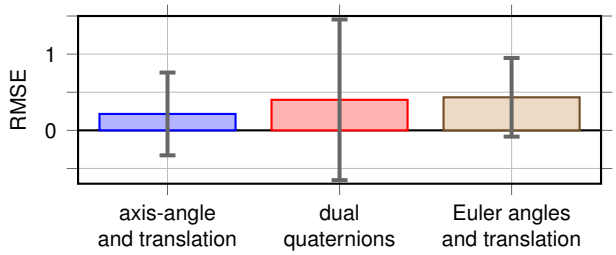

Figure 2: Rooted mean square error in the GP prediction of 1000 test points for 1000 synthetically generated dynamics. Per model the box height shows the mean RMSE and the gray error bar the standard deviation over the 1000 experiment repetitions.

a pendant is available in quaternion space and as for the dot-product kernel no significant difference is expected, as calculations are equal to those in Euclidean space.

4.1 Runtime Comparison

We generate 80 training pairs $\mathcal{D}_\nu = \{(\mathfrak{g}_i,\ \dot{\mathfrak{g}}_{i,l})\}_{i=1}^{\nu}$ in this simulation, consisting of randomly drawn rigid-body motions $\mathfrak{g} \in SE(3)$ as GP input and corresponding velocity output $\dot{\mathfrak{g}}_l$ in one dimension $l \in \{1, \ldots, 6\}$, which is obtained by the pseudo-random output generator algorithm provided in the GPML toolbox. Each training pair is once represented as axis-angle and translation vector, once as dual quaternion and once as Euler angles with translation vector. After the model training phase we compare the runtime required for GP velocity prediction at 10001 new randomly drawn rigid-body motions for the three representations. To obtain reliable and reproducible results, we repeat the experiment 1000 times. Figure 1 visualizes the resulting average prediction times of the 10001 test points and the standard deviation of the average required time for each representation. An independent-samples t-test is conducted to compare the runtime of the newly introduced rigid-body GP with the GP over dual quaternions. The superiority of the GP parametrized by axis-angle and Euclidean vector is verified by a t-test, $t(1998) = 37.8$, $p = 0$, showing a significant improvement in runtime.

4.2 Rotation Prediction Accuracy

In a second simulation we generate random rotation dynamics, mapping from the special Orthogonal group $SO(3)$ to the rotational velocities $TSO(3)$. We focus on rotations, as this is the well-known challenging issue in GP modeling. From every artificially generated dynamics f as in (3) we obtain 1000 ground truth testing pairs consisting each of a rotation with corresponding rotational velocity. From this data we draw 100 i.i.d. random training samples. After the GP training, we compare the predicted GP

values for the testing rotations with the ground truth velocities to evaluate the prediction accuracy for three rotational GP variants; the GP over axis-angle, the GP over unit quaternions (which is a special case of the GP over dual quaternions with purely rotational input) and the GP over Euler angles.

Procedure As we are interested in nonlinear motion behavior, we generate each rotation dynamics f randomly from the trigonometric functions $\mathcal{F} = \{\sin, \cos, \operatorname{atan}, \operatorname{acot}\}$. For simplicity of the description we consider vector-valued dynamics drift terms $f(\theta\mathbf{u}) = \theta'\mathbf{u}'$ of a specific structure only: For $i = 1, \ldots, 3$ each function component has the form

$$f_i(\theta\mathbf{u}) = g\left(\frac{d}{c}\right) \diamond \tilde{d} \star \tilde{c}, \tag{30}$$

where function $g \in \mathcal{F}$. The operators are defined by $\diamond := \pm$ and $\star := \times/\div$, the constants $c, \tilde{c} \in \{\pm 1, \pm\frac{2}{3}, \pm\frac{1}{3}\}$ influence the rotation magnitude in the dynamics and the variables $d, \tilde{d} \in \{\theta\mathbf{u}_1, \theta\mathbf{u}_2, \theta\mathbf{u}_3\}$ denote entries in the axis-angle pseudo-vector. We randomly generate 1000 rotation dynamics of the described form, transform the axis-angle representation to unit quaternions and Euler angles for both, dynamics input and output and unwrap the Euler angles to avoid the jump at $\pm\pi$. Per simulated dynamics we train the models using the standard procedure of minimizing the negative log marginal likelihood, where the hyperparameter start values $\lambda,\ \sigma_f$ are i.i.d. samples from the interval $(0, 3)$. The signal noise σ_n is set to 1 to assure numerical stability during the training phase. As evaluation, the single-step prediction accuracy of each test value in terms of the root-mean-square error (RMSE) is compared to the ground truth.

Results It is clearly visible in Fig. 2 that the average RMSE over 1000 repetitions of each time 1000 predictions of the proposed generalized GP over axis-angle outperforms the other GP variants, as it's estimation error is only about half the error of the other models. Independent-sampled t-tests are conducted to compare the axis-angle generalization with each of the other GPs. A significant improvement (with significance level 1%) in the scores for both comparisons is obtained: $t(5998) = 8.6$, $p = 0$ for condition axis-angle versus dual quaternions and $t(5998) = 16$, $p = 0$ for condition axis-angle versus Euler angles. These results demonstrate the robustness of the proposed rigid-body GP over axis-angle, as it succeeds best in approximating the dynamics under such generalized and suboptimal learning circumstances. Due to the projection required in the GP over unit and dual quaternions, respectively, this GP variant is highly sensitive to the starting values of the GP model learning and thus, performs inferior in learning a suitable GP model for many of the 1000 random dynamics. Hence, the results show large variance in the prediction error for the GP over unit quaternions, even though theoretically comparable estimation quality should be achievable.

4.3 Real World Experiment

In the last experiment we evaluate the proposed rigid-body GP in a real world application. Our goal is to learn human behavior from demonstration. Applications for reproduction and prediction of human-like movements are widespread in robotics, e.g. in programming by demonstration and assistive robotics. Consider a human-robot interaction scenario where a robot should help putting on a shoe to an elderly. In future work, we aim to design a human-model-based control for the robot, to assure intuitive assistance. Therefore, we require a model for human motion behavior in the specific task of passing and slipping

on a shoe to someone else's foot. We learn the human provider motion using the novel rigid-body GP parametrized by axis-angle pseudo-vector and Euclidean vector.

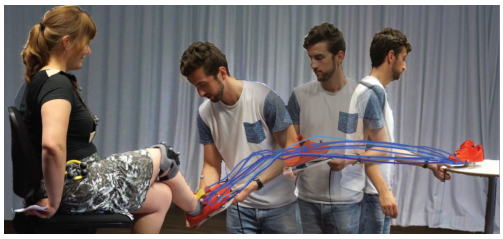

Figure 3: Human-human shoe providing experiment. One human puts on a shoe to another as preliminary study for elderly assistance in service robotics. The blue lines depict the translation part of the providers' 6D motion trajectories.

Experiment Procedure The setup for capturing GP training data is visualized in Fig. 3. The translation part of the trajectory bundle is depicted in blue. The GP is trained on 9 trials of shoe passing and slipping on, demonstrated by different human providers. The data is captured with a motion tracking system (Qualisys, Sweden) at 179 Hz and consists of 6D poses (rotation and translation of the shoe while grasped by the human). The corresponding 6D rotational and translational velocities are obtained by finite differences. The motion trajectories are synchronized on the onset and cut when the contact force between foot and shoe exceeds a predefined threshold, as we suggest to change the controller at that moment. To assure fast runtime of the GP prediction, the GP training is performed on 10% random samples of the captured data. We set the weights in (19) to $\rho_1 = 0.7$ and $\rho_2 = 0.3$, to compensate for the magnitude difference in the captured rotations and translations. The GP training with the total number of $\nu = 1752$ training data pairs requires 8.05 seconds.

Results and Accuracy After the training phase, we calculate 100 new motion trajectories, starting from previously unseen initial 6D poses. The trajectories are generated using single-step ahead mean prediction, which is integrated to pose by the exponential function. In Fig. 4 and Fig. 5 one predicted trajectory is visualized exemplary dimension wise by a black dashed line. The training trajectories are depicted by dotted lines for rotation and translation dimensions. The mean of the training trajectories is visualized by a red solid line and the dark blue dash-dotted line shows a trajectory, which starts at the same initial pose and is predicted using the standard GP over Euler angles and translation vector. All of the trajectories estimated using the novel rigid-body GP starting within the region of human starting poses, stayed throughout the movement within the variance of the human training data. The full trajectory prediction requires 0.84 seconds on average.

For accuracy evaluation we resample the training all to the same length of 3.5 seconds. Then, we determine per time step mean and variance of the training trajectories and use the so obtained time series distribution as our ground truth. For both GP variants we calculate per time step the Bhattacharyya distance to our reference distribution. On average, the distance to the proposed rigid-body GP is 2.9 and to the standard GP using Euler angles it is 4.4. In addition, a t-test is conducted to compare the Bhattacharyya distances per GP condition. We consider per GP variant the distances per time step as random samples from a distribution describing the similarity of the GP predictions and our ground truth distribution. The t-test shows a significant higher similarity for the proposed rigid-body GP than for the standard GP over Euler angles and translation vector, $t(68) = 4.1$, $p = 0$. For translation, Fig. 5, no

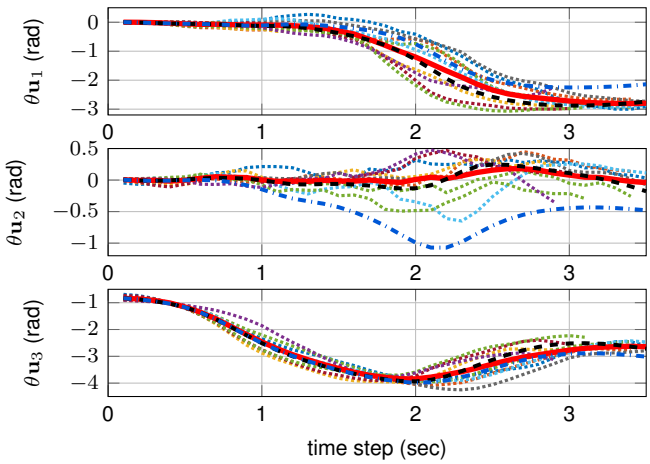

Figure 4: Rotation dimensions of the training data (colored dotted), the newly generated motion trajectory by GP prediction (dashed black), the mean of human demonstrations (red solid) and the standard GP prediction (blue dash-dotted).

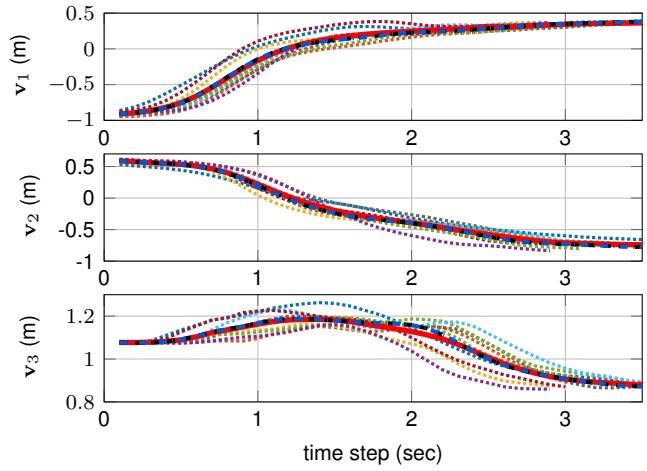

Figure 5: Translation dimensions of the training data (colored dotted), the newly generated motion trajectory by GP prediction (black dashed), the mean of human demonstrations (red solid) and the standard GP prediction (blue dash-dotted).

significant difference between the GP variants is visible, whereas the rotation, Fig. 4, is considerably better captured by the rigid-body GP over axis-angle and Euclidean vector.

5 CONCLUSIONS

In this paper we introduce the rigid-body GP, an approach for learning and predicting motion dynamics in the special Euclidean group $SE(3)$. We employ the axis-angle representation to obtain a computationally efficient GP variant, which generalizes the GP to input data domains in non-Euclidean space. The proposed rigid-body GP outperforms both alternative approaches, namely the GP over dual quaternions and the state of the art GP where rotations are represented by Euler angles.

We introduce mean and kernel functions and prove them to be valid. The generalized version of the most widely used kernel, the squared exponential kernel, is investigated under different aspects such as computational complexity, prediction accuracy, and suitability in real world experiments. Our

results suggest, that the rigid-body GP with the squared exponential kernel comprises outstanding model accuracy, robustness and computational efficiency.

ACKNOWLEDGMENT

This paper was recommended for publication by Editor Dr. Kevin Lynch upon evaluation of the Associate Editor and Reviewers' comments. The research leading to these results has received funding from the EU Horizon2020 project RAMCIP, under grant agreement n° 643433, the ERC Starting Grant "Control based on Human Models (con-humo)" under grant agreement n° 337654 and the European Union Seventh Framework Programme FP7/ 2007-2013 under grant agreement n° 601165 of the project "WEARHAP - WEARable HAPtics for humans and robots".

References

[1] S. Miossec and A. Kheddar, "Human motion in cooperative tasks: Moving object case study," in *IEEE International Conference on Robotics and Biomimetics, ROBIO 2008, February 22-25, 2009, Bangkok, Thailand*, 2008, pp. 1509–1514.

[2] S. Kim and A. Billard, "Estimating the non-linear dynamics of free-flying objects." *Robotics and Autonomous Systems*, vol. 60, no. 9, pp. 1108–1122, 2012.

[3] M. Lang, M. Kleinsteuber, O. Dunkley, and S. Hirche, "Gaussian Process Dynamical Models over Dual Quaternions," in *European Control Conference (ECC)*, Linz, Austria, Jul 2015.

[4] G. S. Aoude, B. D. Luders, J. M. Joseph, N. Roy, and J. P. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Auton. Robots*, vol. 35, no. 1, July 2013.

[5] J. Ko and D. Fox, "GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models," in *IROS 2008. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[6] M. T. Harandi, M. Salzmann, and F. Porikli, "Bregman Divergences for Infinite Dimensional Covariance Matrices," *CoRR*, 2014.

[7] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 73–80.

[8] R. Urtasun, D. Fleet, A. Geiger, J. Popovic, T. Darrell, and N. Lawrence, "*Topologically-Constrained Latent Variable Models*," in *International Conference on Machine Learning (ICML)*, 2008.

[9] E. del Castillo, B. M. Colosimo, and S. D. Tajbakhsh, "Geodesic Gaussian Processes for the Parametric Reconstruction of a Free-Form Surface," *Technometrics*, vol. 57, no. 1, pp. 87–99, 2015.

[10] R. Calandra, J. Peters, C. E. Rasmussen, and M. P. Deisenroth, "Manifold Gaussian Processes for regression," in *International Joint Conference on Neural Networks, IJCNN*, 2016, pp. 3338–3345.

[11] M. Lang, O. Dunkley, and S. Hirche, "Gaussian Process Kernels for Rotations and 6D Rigid Body Motions," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[12] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptative computation and machine learning series. University Press Group Limited, 2006.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[14] B. Palais and R. Palais, "Euler's fixed point theorem: The axis of a rotation," *Journal of Fixed Point Theory and Applications*, vol. 2, no. 2, pp. 215–220, 2007.

[15] B. Hall, *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, ser. Graduate Texts in Mathematics. Springer International Publishing, 2015.

[16] D. Q. Huynh, "Metrics for 3D Rotations: Comparison and Analysis," *J. Math. Imaging Vis.*, vol. 35, no. 2, pp. 155–164, Oct. 2009.

[17] F. C. Park, "Distance Metrics on the Rigid-Body Motions with Applications to Mechanism Design," *Journal of Mechanical Design*, vol. 117, no. 1, pp. 48–54, 1995.

[18] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Structure Discovery in Nonparametric Regression through Compositional Kernel Search," in *International Conference on Machine Learning*, 2013.

[19] D. Zhang, X. Chen, and W. S. Lee, "Text classification with kernels on the multinomial manifold," in *International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM Press, 2005, pp. 266–273.

[20] C. E. Rasmussen and H. Nickisch, "Gaussian Processes for Machine Learning (GPML) Toolbox," *J. Mach. Learn. Res.*, 2010.