

# Gaussian Process for 6-DoF Rigid Motions

Muriel Lang\*

Martin Kleinsteuber\*

Sandra Hirche\*

November 20, 2017

## Abstract

Data-driven modeling approaches receive significant attention in robotics as they are capable of representing system behavior to which first-order principles cannot be employed. Modeling of human motions, based on observations is one of the many application areas. So far, however, the available probabilistic approaches cannot handle dynamics evolving in the space of rigid motions, as rotations are not appropriately considered. In this article, we present a mathematical framework for Gaussian process modeling, where the valid input domain is generalized to full rigid motions, namely the special Euclidean group  $SE(3)$ . The kernel functions inside the Gaussian process are modified to exploit properties of the input data representation by dual quaternions. We further prove that the presented covariance functions maintain the Gaussian process properties. The correctness and accuracy of our approach is validated on simulated and real human motion data. We analyze the estimation performance of the novel Gaussian process framework in comparison to state of the art techniques, and show significantly improved model behavior of rigid motions.

**Keywords:** Gaussian process • Data-driven modelling • Nonlinear dynamics • Dual quaternions • Covariance function • Special Euclidean group • Rigid motion dynamics • Manifold modelling

## 1 Introduction

Data-driven approaches gain in popularity due to their capability to represent complex dynamics, which are difficult to model from first order physical principles. Application domains include human motion prediction for human-robot interaction (HRI), modeling of complex dynamical systems such as tendon-driven robots and unknown object estimation, e.g.

---

\*all authors are with the Department of Electrical and Computer Engineering, Technical University of Munich, D-80333 Munich, Germany  
muriel.lang@tum.de

objects containing liquid. Under such conditions rigid body dynamics, i.e. dynamics evolving on the special Euclidean group  $SE(3)$  need to be learned and predicted. The learning approach we will present in this article applies to all rigid-body motions. As our primary application domain, however, involves human motion estimation, we will motivate this research by choosing examples from this domain. Human action recognition and human movement prediction are of great importance in numerous applications. According to Sebanz and Knoblich (2009), it is instrumental to predict the behavior of the interaction partner for intuitive interaction. In complex task settings, however, no analytical model is available to describe human motor behavior (Medina et al, 2011). Although simple movement prediction can be carried out based on analytical models, such as the minimum jerk profile for point-to-point movements (Corteville et al, 2007; Jarrasse et al, 2008), learning approaches are considered superior for more complex motions (Miossec and Kheddar, 2008).

Widely studied data-driven models for nonlinear dynamics of human behavior are Gaussian processes (GPs), Gaussian mixture models (GMMs) and Hidden Markov models (HMMs); applied for instance by Wang et al (2008), Khansari-Zadeh and Billard (2011) and Kulić et al (2007), respectively. These approaches convince owing to their robustness in describing system dynamics from mere observations. While HMMs are particularly popular for describing motion primitives, they are limited to discrete state space. GP and GMM are defined on continuous space and therefore suitable for modeling dynamics in continuous state space. Neither the GP nor the GMM allows for mathematically correct modeling of orientations though. Both models are defined in Euclidean space, as Gaussian measures require a real underlying vector space to be correctly normalized. Hence, they are suitable solely for modeling data that exhibit an underlying vector space structure, such as translation, linear velocity and forces. However, for rotations no representation exists in any space owing the required vector space structure. The common procedure for rotations is to use an approximately Euclidean representation, e.g. the Euler angles (Ko and Fox (2008), Kim and Billard (2012)). If rotation angles are small, this approach may suffice as the space of Euler orientations close to zero is nearly Euclidean. However, a model generalization to non-Euclidean space is required, when the orientation between successive input data is large, for example at high angular speed, low sampling frequency or partial input data loss.

In this article, we focus on the GP for its property to account for the uncertainty in each prediction besides providing a best estimator. The main contributions concerns the introduction of a rigorous mathematical framework for modeling rigid motion dynamics by GPs and the development of a recursive reformulation of the GP framework that allows for online application. Even though the fundamentals of GP modeling are maintained in our approach, major changes in the kernel function are required to achieve a mathematically correct generalization of the input domain to the special Euclidean group  $SE(3)$ . In addition, a Gaussian-like bell curve is defined on dual quaternions; the parametrization we employ for the rigid motions. This probability distribution serves as a best estimator and uncertainty estimate in the dynamics state space.

## 1.1 Related Work

This section gives an overview of GP models and kernel methods for estimating nonlinear rigid motion dynamics in continuous time with special focus on model generalizations to non-Euclidean space.

A GP is a Bayesian model, in which the inputs are projected to some high dimensional feature space using the kernel trick before regression (respectively classification) is performed (Rasmussen and Williams, 2006). Since its first application to dynamical systems by Ko and Fox (2008), the GP gains popularity in this domain. Kang and Park (2015) and Wang et al (2008), for instance, use Gaussian process dynamical models for motion estimation. Their approaches employ additional lower dimensional GP layers for improved prediction results. These lower dimensional spaces are assumed to possess a Euclidean vector space structure, what makes the approaches unsuitable in our setting.

Kernel functions are the core of a GP model and in general non-trivial to extend beyond the Euclidean space. In Fukumizu et al (2009), the subclass of characteristic kernels are investigated, i.e. where the mapping is injective. An example for a characteristic kernel is given on the special orthogonal group  $SO(3)$  which represents rotations in  $\mathbb{R}^3$ . However, the correlation of orientations, described by the characteristic kernel, is not related to a distance function based on the rotation magnitude between orientations. The manifold metrics in a Hilbert space embedding, investigated in Harandi et al (2014), do not necessarily result in valid kernel functions. One counter example is a kernel resulting from the Stein-divergence. Jayasumana et al (2013) introduce a theorem that states sufficient and necessary conditions to obtain valid Gaussian kernels on metric spaces. We will make use of their results in the development of our kernel function. None of the kernel approaches listed so far was applied to GP models to the best of our knowledge.

Another interesting line of research considers manifold-based learning methods. The approaches of Tenenbaum et al (2000), Roweis and Saul (2000) and Belkin and Niyogi (2003) employ nonlinear dimensionality reduction methods and find application in machine learning as non-Gaussian manifold kernel methods. In contrast to these approaches, the manifold kernel dimension reduction (Nilsson et al, 2007) is a supervised setting for dimensionality reduction. The authors combine the approach with manifold regression through Laplacian eigenmaps. The kernel formulation of those methods admits an interpretation of the algorithms as warping of the input space into a feature space where the manifold is flat (Ham et al, 2004).

As our interest lies on GP modeling, we survey approaches that incorporate GPs over non-Euclidean spaces in the following. An embedding of latent variables into non-Euclidean space is presented for GP latent variable models by Urtasun et al (2008). However, as only locally linear embeddings are considered, the standard GP structure is maintained. Another approach (Del Castillo et al, 2015) introduces a geodesic GP. The GP model is used to train a surface embedding in Euclidean space. That is, the GP is defined in Euclidean space but maps geodesic distances to a Euclidean equivalent. Lang and Hirche (2017) describe a GP generalization using a pose representation by axis-angle pseudo-vector and translation vector, which can be used for modeling rigid motion dynamics. This approach, however, represents uncertainty on a locally flat neighborhood around the estimation only, but does not allow for representing uncertainty on  $SE(3)$  globally. Further, Calandra et al (2014) introduce a manifold GP that jointly models 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. 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. As we focus our investigations to generalized GP models over rigid motions, this allows us to describe the manifold  $M$  explicitly and to define an explicit mapping from the generalized GP output space to the manifold.

## 1.2 Innovations and Article Structure

In the following, the innovations of the work and the article structure is detailed. Our approach and contribution concerns the development of a complete mathematical framework for GP modeling of rigid motion dynamics. Beyond that, we introduce a recursive version of the GP over rigid motions for online applications. As an example we then parametrize the rigid motions in the non-Euclidean space  $SE(3)$  by dual quaternions and perform the necessary computations for mean and kernel of the GP model generalization using this representation. The general approach, however, is not limited to a specific rigid motion parametrization but can equally be conducted using homogeneous transformation matrices, for instance.

We start by providing a formal GP definition and then introduce GP kernel functions on unit and dual quaternions. All functions are proven to fulfill the kernel characteristics. Thereafter, a novel recursive generalized GP model for rigid motion dynamics is introduced, which allows for online GP regression. By default, the model uncertainty of the generalized GP is encoded in a Gaussian defined in the gradient field to  $SE(3)$ . Thus, we complete the framework by developing a proper uncertainty prediction method in  $SE(3)$  directly. Therefore, we define a probability distribution on  $SE(3)$ , which possesses the properties of the Gaussian bell curve. Preliminary results to this work have been published in the conference papers (Lang et al, 2014, 2015). Here we provide a more general framework on GP modeling, introduce an online GP version and add experimental results.

Our generalized GP model is validated on simulated and real experimental data. In a regression performance analysis, synthetically generated dynamics are modeled by the GP over dual quaternions, and are evaluated in comparison to the traditional GP. A second experiment assesses the prediction behavior of GP models using different metrics on the unit sphere  $S_3$  as similarity measures on real human gaze motion. In a third experiment, the recursive GP model is used for online motion estimation in an object handover task using real data. The evaluations confirm that the introduced framework for GP modeling over rigid motions significantly outperforms the traditional way of including orientation into the GP via Euler angles.

## 2 Concept of GP Generalization to $SE(3)$

In this section we explain how we generalize the GP to the non-Euclidean space of rigid motions. Our goal is to enhance the GP model to appropriately describe nonlinear rigid-body dynamics in the special Euclidean group  $SE(3)$ . Note that the underlying principle is independent of the rigid motion representation. Then, we discuss most common parametrizations

and our decision to use dual quaternions.

Let us start by considering the tangent bundle

$$TSE(3) = \{(\mathbf{g}, \mathbf{n}_{\mathbf{g}}) \mid \mathbf{g} \in SE(3), \mathbf{n}_{\mathbf{g}} \in T_{\mathbf{g}}SE(3)\}, \quad (1)$$

where the elements are pairs consisting of a pose  $\mathbf{g} \in SE(3)$  and a term  $\mathbf{n}_{\mathbf{g}} \in T_{\mathbf{g}}SE(3)$ . A rigid-body motion dynamics is defined by a smooth mapping

$$f: SE(3) \times T_{\mathbf{g}}SE(3) \rightarrow TSE(3). \quad (2)$$

The dynamics maps the rigid motions  $\mathbf{g} \in SE(3)$  including a noise process  $\mathbf{n}_{\mathbf{g}} \in T_{\mathbf{g}}SE(3)$  to the tangent bundle  $TSE(3)$ , representing the velocity space. One of the key elements of the GP generalization is to define (2) in such a way that the output, a predictive Gaussian probability distribution, lies in the tangent bundle. GP output is inherently not defined on  $SE(3)$ , as  $SE(3)$  does not provide the vector space structure required for the GP output. As the codomain of dynamics (2) is the velocity space, it exhibits the required vector space structure, and the GP output is thus defined. To take maximum advantage of GP property to account for the uncertainty, it is additionally desirable to provide an uncertainty estimate also on the rigid motions directly. Given the dynamics input space is a Riemannian manifold  $M$  (Subbarao, 2008) we can project the probability distribution from the velocity space canonically to  $SE(3)$ . Hence, we are able to define a Gaussian pendant on the rigid motions.

In the following, we discuss various representations of  $\mathbf{g}$  in  $SE(3)$  to obtain a more concrete description of the dynamical system (2): the most popular pose representation for  $\mathbf{g} \in SE(3)$  is a vector of Euler angles concatenated with translation vector  $\mathbf{t}$ , i.e.  $\mathbf{g} = (\phi, \psi, \vartheta, t_x, t_y, t_z)^{\top}$ . However, the rotation angles are not Euclidean by definition. Therefore, the Euclidean distance between Euler angles, required for stationary kernel functions  $k(\mathbf{g}, \mathbf{g}')$ , is erroneous. Another possible parametrization of the rotation part of  $\mathbf{g}$  is the axis-angle orientation representation, which can be derived from Euler's rotation theorem. However, this representation lacks an important property required for motion generation in the GP prediction: no efficient composition algorithm is available, as the vector sum is not employable. A further alternative is the representation of  $SE(3)$  in terms of  $4 \times 4$  homogeneous transformation matrices  $\mathbf{g} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix}$ . Using this parametrization, the tangent space representation is obtained by left transition  $\mathfrak{g}$  of the Lie algebra,  $T_{\mathbf{g}}SE(3) = \mathfrak{g}\mathfrak{se}(3)$ , where the Lie algebra is represented by the set of twists,

$$\mathfrak{se}(3) = \left\{ \begin{pmatrix} \Omega & \tau \\ 0 & 0 \end{pmatrix} \mid \Omega \in \mathbb{R}^{3 \times 3}, \Omega^{\top} = -\Omega, \tau \in \mathbb{R}^3 \right\}. \quad (3)$$

The linear velocities are represented by vector  $\tau$  and angular velocities  $\omega_{\phi}$ ,  $\omega_{\psi}$ ,  $\omega_{\vartheta}$  can be encoded in the skew symmetric

matrix  $\Omega$  via

$$\Omega = \begin{pmatrix} 0 & \omega_\phi & \omega_\psi \\ -\omega_\phi & 0 & \omega_\vartheta \\ -\omega_\psi & -\omega_\vartheta & 0 \end{pmatrix}. \quad (4)$$

A use of this representation is theoretically plausible. However, instead of handling bulky  $4 \times 4$  matrices we decide in favor of the more elegant dual quaternion representation  $\mathfrak{g} \in \mathbb{H}_D$  for 6-DoF rigid body motions in this work. Unit quaternions produce smooth rotations and can be efficiently composed via quaternion multiplication. Further, they extend naturally to dual quaternions, while preserving their desired properties. Hence, they permit encapsulating rotation and translation in a unified representation (Thomas, 2014) in a closed to minimal representation with just 7 free parameters. In addition, this representation allows for particularly simple global projection from the velocity space onto the manifold of rigid motions  $SE(3)$ .

### 3 GP over Rigid Motions

In this section we introduce the GP generalization for modeling 6-DoF rigid motions over dual quaternions. We start by briefly recapitulating the pose representation by a dual quaternion, before the GP model is defined and characteristic model properties are highlighted. For better understanding, the generalized GP framework is firstly developed for pure rotations, and then extended to whole rigid motions. Finally, we introduce a novel recursive generalized GP algorithm that is suitable for online applications.

#### 3.1 Dual Quaternion Representation

The set  $\mathbb{H}$  denotes the skew field of quaternions. A 3D rotation is represented by a unit quaternion  $\mathbf{q}_{\text{rot}} = \mathbf{q}_w + \mathbf{q}_x i_x + \mathbf{q}_y i_y + \mathbf{q}_z i_z \in \mathbb{H}$ , with parameters  $\mathbf{q}_w, \mathbf{q}_x, \mathbf{q}_y, \mathbf{q}_z \in \mathbb{R}$  such that  $\|\mathbf{q}_{\text{rot}}\| = 1$ . A 3D translation is represented by an imaginary quaternion  $\mathbf{q}_{\text{tra}} = \mathbf{q}'_x i_x + \mathbf{q}'_y i_y + \mathbf{q}'_z i_z$ , where  $\mathbf{q}'_x, \mathbf{q}'_y, \mathbf{q}'_z \in \mathbb{R}$ . Together, rotation and translation define a rigid motion. Thus, the representations  $\{(\mathbf{q}_{\text{rot}}, \mathbf{q}_{\text{tra}})\} \simeq S_3 \times \mathbb{R}^3$  are isomorphic. We divide the sphere  $S_3$  into the equivalence classes  $\{\pm 1\}$ , because the unit quaternions  $S_3$  are a double coverage of the rotations  $SO(3)$ .

**Proposition 1.** (Lang et al, 2015) Using the semi-direct product  $\rtimes$  to connect the spaces  $S_3$  and  $\mathbb{R}^3$ , the dual quaternions representing rigid motions are isomorphic to the special Euclidean group,

$$SE(3) \simeq S_3 / \{\pm 1\} \rtimes \mathbb{R}^3. \quad (5)$$

The dual extension of the quaternions  $\mathbb{H}$ , namely the ring of the dual quaternions, is given by

$$\mathbb{H}_D = \{ \mathbf{g} \mid \mathbf{g} = \mathbf{q}_{\text{re}} + \varepsilon \mathbf{q}_{\text{du}} \ \& \ \mathbf{q}_{\text{re}}, \mathbf{q}_{\text{du}} \in \mathbb{H} \}, \quad (6)$$

where  $\mathbf{q}_{\text{re}}$  and  $\mathbf{q}_{\text{du}}$  denote quaternions in the real and the dual space, respectively, and  $\varepsilon$  is a dual unit which holds  $\varepsilon^2 = 0$  (for more details see Ata, 2008). The dual quaternion  $\mathbf{g}$  representing the pose  $(\mathbf{q}_{\text{rot}}, \mathbf{q}_{\text{tra}})$ , is composed by

$$\mathbf{g} := \underbrace{\mathbf{q}_{\text{rot}}}_{\mathbf{q}_{\text{re}}} + \varepsilon \underbrace{\frac{1}{2} \mathbf{q}_{\text{tra}} \mathbf{q}_{\text{rot}}}_{\mathbf{q}_{\text{du}}}. \quad (7)$$

The dual quaternion multiplication of rigid motions  $\mathbf{g}$  and  $\mathbf{g}'$ , defined by  $\mathbf{g} * \mathbf{g}' = \mathbf{q}_{\text{re}} \mathbf{q}'_{\text{re}} + \varepsilon (\mathbf{q}_{\text{re}} \mathbf{q}'_{\text{du}} + \mathbf{q}_{\text{du}} \mathbf{q}'_{\text{re}})$ , acts as successive execution of the motions. Further, the dual quaternion conjugate of  $\mathbf{g}$  is defined by  $\bar{\mathbf{g}} = \bar{\mathbf{q}}_{\text{re}} + \varepsilon \bar{\mathbf{q}}_{\text{du}}$ , where  $\bar{\mathbf{q}}$  is the quaternion conjugate of  $\mathbf{q}$ .

### 3.2 Gaussian Process Characteristics

Let us disambiguate the stochastic process by providing a formal definition.

**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  $\{\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_v)\}$  is  $v$ -dimensional multivariate Gaussian.

Rasmussen and Williams (2006) show that in function space, a Gaussian process is fully specified by a mean function  $m(\mathbf{x})$  and a kernel function  $k(\mathbf{x}, \mathbf{x}')$

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[\varphi(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(\varphi(\mathbf{x}) - m(\mathbf{x}))(\varphi(\mathbf{x}') - m(\mathbf{x}'))], \end{aligned} \quad (8)$$

and thus, we can write

$$\varphi(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (9)$$

In this work, we consider vector-valued stochastic processes  $f = (\varphi_1, \dots, \varphi_n)^\top$ , where each component  $\varphi_i$  is a Gaussian process  $\varphi_i(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  with same classes of mean and kernel functions for all  $i$ . The predictive Gaussian probability distributions provided by the GP for each component of  $\dot{\mathbf{x}}_{v+1}$  are joint into one multivariate Gaussian distribution

$$\dot{\mathbf{x}}_{v+1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (10)$$

where  $\boldsymbol{\mu} := (\mu_1, \dots, \mu_n)^\top$  and  $\sigma_i$  the diagonal entries of the otherwise empty covariance matrix  $\boldsymbol{\Sigma}$ . In practice, the mean function  $m(\mathbf{x})$ , is often set to zero as this simplifies calculation without limiting the expressiveness of the process (see

Bishop, 2006, Chap. 6.4.1). Therefore, and as their handling is straight forward, we will no further investigate mean functions in this article, but set  $m(\cdot) \equiv 0$  in the GP generalization. The essential part concerns the modeling of the kernel  $k(\mathbf{x}, \mathbf{x}')$ . We focus on generalization of the squared exponential kernel function

$$k_{\text{eu}}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{[d_{\text{eu}}(\mathbf{x}, \mathbf{x}')]^2}{2\lambda^2}\right), \quad (11)$$

where  $d_{\text{eu}}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|$  is the Euclidean distance and  $H = (\lambda, \sigma_f)$  the hyperparameters length-scale and signal noise, as this kernel is part of the most widely used kernel functions in machine learning.

In real world applications it is common procedure to encode rigid motions in a 6D state vector  $\mathbf{g} = (\phi, \psi, \theta, t_x, t_y, t_z)^\top$ . Employing (9) induces the traditional 6-DoF GP

$$\check{f}(\mathbf{g}) = \text{GP}_{\mathbb{R}^6}(m(\mathbf{g}), k(\mathbf{g}, \mathbf{g}')). \quad (12)$$

To distinguish the GP over dual quaternions in  $SE(3)$  from the traditional 6-DoF GP, we denote them by  $\text{GP}_{\mathbb{H}_D}$  and  $\text{GP}_{\mathbb{R}^6}$ , respectively. An application of  $\text{GP}_{\mathbb{R}^6}$  is reasonable solely if it can be guaranteed that all occurring angles are in close neighborhood to zero, where Euler angles are approximately Euclidean. Elsewise, the preferred GP model is the  $\text{GP}_{\mathbb{H}_D}$ .

### 3.3 Gaussian Process over Unit Quaternions

To support an intuitive understanding, this section explains a GP generalization over pure rotations parametrized by unit quaternions. For the sake of clarity, the index “rot” for quaternions  $\mathbf{q}_{\text{rot}} \in S_3$  is omitted in the remainder of this section. We contrast three distance measures on the sphere and show that each of them induces a squared exponential kernel.

#### 3.3.1 Unit Quaternion Metrics

Figure 1 visualizes three distance measures that define metrics on the unit sphere  $S_3$ , i.e. they are non-negative, symmetric and fulfil the triangle inequality. All three metrics are equally applicable in a GP generalization, but differ essentially in the underlying concept. The quaternion norm

$$d_{\text{norm}}(\mathbf{q}, \mathbf{q}') = \sqrt{\mathbf{q}\mathbf{q}'} = \|\mathbf{q} - \mathbf{q}'\|, \quad (13)$$

where  $\mathbf{q}, \mathbf{q}' \in S_3$  (depicted by the red dashed line in Fig. 1) provides identical results to the regular vector norm of points in the space  $\mathbb{R}^4$ . Its pros concern the low computational cost and the direct applicability of theoretical findings that are valid in Euclidean space. Contrary to intuition, however, the quaternion norm metric yields

$$2d_{\text{norm}}(\mathbf{q}, \mathbf{q}_\Delta \mathbf{q}) > d_{\text{norm}}(\mathbf{q}, (\mathbf{q}_\Delta \mathbf{q}_\Delta) \mathbf{q}), \quad (14)$$

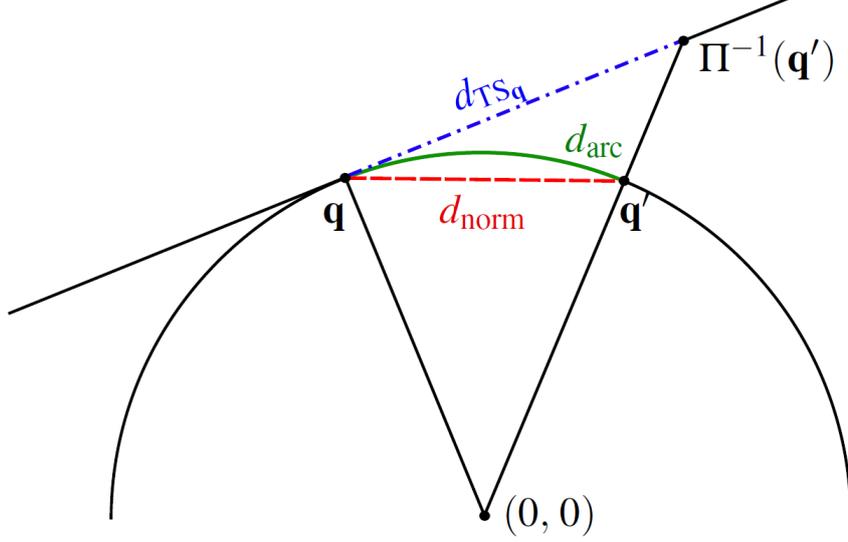


Figure 1: The distance measures consisting of quaternion norm, arc length and projection to the tangent space are depicted by a red dashed, green solid and blue dashed-dotted line, respectively.

when applying twice a non-trivial rotation  $\mathbf{q}_\Delta \neq (1, 0, 0, 0)^\top$ . The arc metric (depicted by the green solid line) between unit quaternions  $\mathbf{q}, \mathbf{q}' \in \mathcal{S}_3$  is calculated via

$$d_{\text{arc}}(\mathbf{q}, \mathbf{q}') = \min |\arccos(\pm \mathbf{q}_{\Delta w})|, \quad (15)$$

where  $\mathbf{q}_{\Delta w}$  is the real part of the transforming unit quaternion  $\mathbf{q}_\Delta = \mathbf{q}'\bar{\mathbf{q}}$  from  $\mathbf{q}$  to  $\mathbf{q}'$ . It is an alternative metric, which satisfies equality in (14) for all  $\mathbf{q}, \mathbf{q}' \in \mathbb{H}$  with  $\langle \mathbf{q}, \mathbf{q}' \rangle \geq 0$ . Both distance measures introduced so far are bounded, in fact  $\max(d_{\text{norm}}) = \sqrt{2}$  and  $\max(d_{\text{arc}}) = \frac{\pi}{2}$ . An unbounded metric can be obtained when one quaternion, for instance  $\mathbf{q}'$  is projected to the other's tangent space  $TS_{\mathbf{q}}$ . In Section 4.1 we will present the central projection  $\Pi$  as it was applied by Lang (2011), generating infinite distance for  $\pi$ -rotations. The origin of  $TS_{\mathbf{q}}$  is identical to quaternion  $\mathbf{q}$  and thus, the tangent space distance  $d_{TS_{\mathbf{q}}}(\mathbf{q}, \mathbf{q}')$  between projected quaternions is computed using the Euclidean norm in the space  $TS_{\mathbf{q}} \simeq \mathbb{R}^3$ ,

$$d_{TS_{\mathbf{q}}}(\mathbf{q}, \mathbf{q}') = \|\Pi_{\mathbf{q}}^{-1}(\mathbf{q}')\| = \|\mathbf{q}'_{TS_{\mathbf{q}}}\| = \|\Pi_{\mathbf{q}'}^{-1}(\mathbf{q})\|. \quad (16)$$

The appropriate choice of the metric is application dependent. While the tangent space norm  $d_{TS_{\mathbf{q}}}$  assumes identical costs for a  $\pi$ -rotation and infinite translation, the quaternion norm  $d_{\text{norm}}$  produces only little inequality in (14) for small rotations. We suggest the use of the arc metric  $d_{\text{arc}}$  in all other applications. The proof for (13) and (16) to define metrics is trivial, as both distance measures are vector norms in the Euclidean space. The arc distance (15) is non-negative by definition of the arccos function, it is symmetric, as  $\Delta \mathbf{q}^{-1} = -\Delta \mathbf{q}^{-1}$  and a proof for the triangle inequality to be valid is provided in Matsuoka et al (2011, pp. 195-196).

### 3.3.2 Squared Exponential Kernel Functions

Over the metrics from Section 3.3.1, kernel functions on  $S_3$  can be defined. We focus on the kernel function resulting from the arc length metric, as the underlying metric preserves the unit quaternion topology.

**Theorem 1.** A valid kernel function  $k_{\text{arc}} : S_3 \times S_3 \rightarrow \mathbb{R}_0^+$  over unit quaternions  $\mathbf{q}, \mathbf{q}' \in S_3$ , is defined by

$$k_{\text{arc}}(\mathbf{q}, \mathbf{q}') = \sigma_f^2 \exp\left(-\frac{[d_{\text{arc}}(\mathbf{q}, \mathbf{q}')]^2}{2\lambda^2}\right), \quad (17)$$

where  $d_{\text{arc}}$  as in (15) and the hyperparameters  $\sigma_f, \lambda > 0$ .

To prove Theorem 1 we need to show that  $k_{\text{arc}}(\mathbf{q}, \mathbf{q}')$  is symmetric and positive definite  $\forall \mathbf{q}, \mathbf{q}' \in S_3$ . Alternatively, we can show Lemma 1 following below, to prove Theorem 1. Therefore, we need to define the notion of conditional positive definiteness (cpd).

**Definition 2** ((Zhang et al, 2005)). Let  $\mathcal{X}$  be a nonempty set. A real-valued symmetric function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is called a conditionally positive definite (cpd) kernel function, if and only if the Gram matrix  $\mathbf{K} \in \mathbb{R}^{v \times v}$  satisfies  $\mathbf{c}^\top \mathbf{K} \mathbf{c} \geq 0$  for any vector  $\mathbf{c} \in \mathbb{R}^v$  with  $\mathbf{c}^\top \mathbf{1} = 0$ , where  $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^v$ .

**Lemma 1.** The squared negative distance function

$$k_{\text{negd}}^2(\mathbf{q}, \mathbf{q}') := -([d_{\text{arc}}(\mathbf{q}, \mathbf{q}')]^2) \quad (18)$$

defines a cpd kernel function.

*Lemma 1.* Let the functions  $h_1 : [-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$  given by  $h_1(x) = \pi/2 - \arccos(x)$  and  $h_2 : [-1, 1] \rightarrow [\frac{3}{2}\pi, \frac{\pi}{2}]$  given by  $h_2(x) = \pi/2 + \arccos(x)$ . Let us assume for now that the functions  $h_1$  and  $h_2$  are kernels. Given the finite product of kernels defines a kernel (Bishop, 2006), then also the function  $h : [-1, 1] \rightarrow [-\frac{3}{4}\pi^2, \frac{1}{2}\pi^2]$  defined by  $h(x) = h_1(x)h_2(x)$  is also a valid kernel. Hence, we can rewrite (18) to

$$k_{\text{negd}}^2(\mathbf{q}, \mathbf{q}') = -\frac{\pi^2}{4} + h(\min(\langle \mathbf{q}, \pm \mathbf{q}' \rangle)). \quad (19)$$

From Zhang et al (2005) we know that a constant  $c \in \mathbb{R}$  is cpd, a valid kernel is also cpd and the finite sum of cpd kernels is cpd. This induces, that (19) is cpd. Thus, the proof is complete, if we can prove that the functions  $h_1$  and  $h_2$  are valid kernels. According to Zhang et al (2005, Theorem 4) we know that the Taylor series expansion of a function defined on the unit sphere, such as  $\arcsin(x)$ , defines a valid covariance function, iff all coefficients are non-negative. Reformulating  $h_1(x) = \arcsin(x)$ , the function can be expressed as infinite polynomial with non-negative coefficients via Taylor expansion. Likewise, the function  $h_2$  can be reformulated in terms of the arcsin function,  $h_2(x) = \pi + \arcsin(-x)$ . As the sum of kernels defines a kernel function according to Bishop (2006), we obtain that  $h_1$  and  $h_2$  both define valid kernels. □ □

*Theorem 1.* If a kernel  $k$  is cpd, it follows that  $\exp(ck)$  is a valid kernel function for all constants  $c > 0$  according to Jayasumana et al (2013). Further, we know from Bishop (2006) that any finite constant  $\tilde{c} > 0$  is a positive definite kernel function and the product of valid kernels defines a positive definite kernel function. Together, this proves that the function  $k_{\text{arc}}(\mathbf{q}, \mathbf{q}')$  defines a positive definite kernel function for all  $\sigma_f, \lambda > 0$ .  $\square$   $\square$

By introducing the distance measures (13) resp. (16) into the formula for the squared exponential kernel (11), we obtain kernels  $k_{\text{norm}}$  and  $k_{TS}$ , respectively, in the Euclidean space. They, too, allow for a simple GP generalization to 3-DoF rotations.

### 3.4 Gaussian Process over Dual Quaternions

In this section, we introduce the  $\text{GP}_{\mathbb{H}_D}$  over rigid motions, parametrized by dual quaternions  $\mathbb{H}_D$ . First, we present a distance measure over dual quaternions, and then introduce a generalized squared exponential kernel function on  $SE(3)$ .

#### 3.4.1 Dual Quaternion Distance Measure

Motivated by the fact that the magnitude of the transforming unit quaternion results in the topologically correct distance measure of rotations, we propose to base the dual quaternion distance on a magnitude measure between rigid motions  $\mathfrak{g}, \mathfrak{g}' \in \mathbb{H}_D$ ,

$$\mathfrak{g}_\Delta = \bar{\mathfrak{g}} * \mathfrak{g}'. \quad (20)$$

The transformation dual quaternion  $\mathfrak{g}_\Delta$  has to be applied to  $\mathfrak{g}$ , to arrive in  $\mathfrak{g}'$ . As the dual quaternion norm equals 1 for all rigid motions, we decompose  $\mathfrak{g}_\Delta = \mathbf{q}_{\text{re}_\Delta} + \varepsilon \mathbf{q}_{\text{du}_\Delta}$  into rotation  $\mathbf{q}_{\text{rot}_\Delta}$  and translation  $\mathbf{q}_{\text{tra}_\Delta}$  using (7) and define the transformation magnitude measure via

$$d_{\text{mag}}(\mathfrak{g}, \mathfrak{g}') = \sqrt{[d_{\text{arc}}(\mathbf{q}_0, \mathbf{q}_{\text{rot}_\Delta})]^2 + \|\mathbf{q}_{\text{tra}_\Delta}\|^2}, \quad (21)$$

where  $\mathbf{q}_0 = (1, 0, 0, 0)^\top$  denotes the zero rotation. Instead of  $d_{\text{arc}}$ , any of the quaternion metrics defined in Section 3.3.1 could be used and (21) would hold true defining a distance measure. We focus on distance (15) as it describes topologically correct quaternion divergence. The square root is introduced in the definition, as this allows for inserting  $d_{\text{mag}}^2$  into the kernel function without getting combined terms of rotational and translational parts.

#### 3.4.2 Squared Exponential Kernel Function

Analogously to the GP model generalization over pure rotations, a squared exponential kernel function for rigid motions in  $SE(3)$  is obtained from the proposed dual quaternion distance measure.

**Theorem 2.** A valid kernel function  $k_{\text{mag}} : (S_3 \times \mathbb{R}^3) \times (S_3 \times \mathbb{R}^3) \rightarrow \mathbb{R}_0^+$  over dual quaternions  $\mathfrak{g}, \mathfrak{g}' \in S_3 \times \mathbb{R}^3$ , is defined

by

$$k_{\text{mag}}(\mathbf{g}, \mathbf{g}') = \sigma_f^2 \exp\left(-\frac{[d_{\text{mag}}(\mathbf{g}, \mathbf{g}')]^2}{2\lambda^2}\right), \quad (22)$$

where  $d_{\text{mag}}$  as in (21) and the hyperparameters  $\sigma_f, \lambda > 0$ .

*Theorem 2.* We reformulate the function (22) using the property  $\exp(a+b) = \exp(a)\exp(b)$ , as for each factor separately, it is simple to show, that the factor defines a kernel function. Thus, we obtain

$$k_{\text{mag}}(\mathbf{g}, \mathbf{g}') = \underbrace{\sigma_f^2 \exp\left(-\frac{[d_{\text{arc}}]^2}{2\lambda^2}\right)}_{\star} \underbrace{\exp\left(-\frac{\|\mathbf{q}_{\text{tra}}\|^2}{2\lambda^2}\right)}_{\diamond}. \quad (23)$$

The factor  $\star$  defines a valid kernel function according to Theorem 1 and the factor  $\diamond$  describes a valid kernel of standard form (11). To finalize the proof, we apply the kernel construction rule via multiplication. This rule states that any finite product of kernels defines a kernel, see Bishop (2006, Chap. 6.2).  $\square$   $\square$

### 3.5 Online GP over Rigid Motions

So far, we discussed GP models that are suitable for offline model learning. Those models are convenient for applications with static environment. In real world applications, however, it often is impractical or even impossible to a-priori train the model as for instance the data capturing might be extensive and time-consuming. Further, under changing circumstances the training data can be meaningless for the test case. Therefore, we introduce in the following a recursive algorithm that allows for online model training in the generalized GP framework. It is inspired by the Kernel Recursive Least-Squares (KRLS) algorithm by Van Vaerenbergh et al (2012) for standard GP models in Euclidean space.

Let us consider a dynamical system of form (2) over dual quaternions and a training data set  $\mathcal{D}_v := \{(\mathbf{g}_i, \dot{\mathbf{g}}_i)\}_{i=1}^v$  consisting of  $v$  input-output pairs from the dynamics  $f = (\varphi_1, \dots, \varphi_n)^\top$ . As each output component is processed separately and joined into one output distribution afterwards, see Sec. 3.2, we denote by  $\dot{\mathbf{g}}_i^{(l)}$  the  $l$ -th component of the output velocity for  $l = 1, \dots, n$  and employ per GP the according training data set  $\mathcal{D}_v^{(l)} := \{(\mathbf{g}_i, \dot{\mathbf{g}}_i^{(l)})\}_{i=1}^v$ . Then, the online GP algorithms comprises two main steps. First, it provides an  $n$ -dimensional predictive distribution for any new input  $\mathbf{g}_{v+1}$  given training data  $\mathcal{D}_v$  and second, per component  $l$  it updates the joint Gaussian distribution

$$p(\varphi_l(\mathbf{g}_1), \dots, \varphi_l(\mathbf{g}_v) | \mathcal{D}_v^{(l)}) = \mathcal{N}(\varphi_l(\mathbf{g}_1), \dots, \varphi_l(\mathbf{g}_v) | \boldsymbol{\mu}_v^{(l)}, \boldsymbol{\Sigma}_v^{(l)}) \quad (24)$$

to a new conditional posterior  $p(\varphi_l(\mathbf{g}_1), \dots, \varphi_l(\mathbf{g}_{v+1}) | \mathcal{D}_{v+1}^{(l)})$  including the observation  $(\mathbf{g}_{v+1}, \dot{\mathbf{g}}_{v+1}^{(l)})$ . The mean vector  $\boldsymbol{\mu}_v^{(l)} \in \mathbb{R}^v$  and the covariance matrix  $\boldsymbol{\Sigma}_v^{(l)} \in \mathbb{R}^{v \times v}$  in (24) describe mean and covariance, respectively, of the prediction vector  $(\varphi_l(\mathbf{g}_1), \dots, \varphi_l(\mathbf{g}_v))^\top$  in the  $l$ -th GP component.

### 3.5.1 Initialization

Per dynamics output component  $l = 1, \dots, n$  the generalized online GP regression is fully described by the variables mean  $\mu_v^{(l)}$ , covariance  $\Sigma_v^{(l)}$  and the inverse of the Gram matrix  $\mathbf{S}_v^{(l)} := \mathbf{K}_v^{(l)-1}$ , which is introduced to save the computational effort of matrix inversion. Given a first observation pair  $\{(\mathbf{g}_1, \hat{\mathbf{g}}_1^{(l)})\}$ , the variables are initialized by setting

$$\begin{aligned}\mu_1^{(l)} &= \frac{\hat{\mathbf{g}}_1^{(l)} k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1)}{\sigma_{\mathbf{n}}^2 + k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1)} \\ \Sigma_1^{(l)} &= k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1) - \frac{k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1)^2}{\sigma_{\mathbf{n}}^2 + k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1)} \\ \mathbf{S}_1^{(l)} &= \frac{1}{k_{\text{mag}}(\mathbf{g}_1, \mathbf{g}_1)},\end{aligned}\tag{25}$$

where  $k_{\text{mag}}$  as defined in Theorem 2 and  $\sigma_{\mathbf{n}}^2$  denotes the variance of the noise process  $\mathbf{n}_{\mathbf{g}}$  defined below (2).

### 3.5.2 Update Step

Given the recursive variables  $\mu_v^{(l)}$ ,  $\Sigma_v^{(l)}$  and  $\mathbf{S}_v^{(l)}$  and a new observation  $\mathbf{g}_{v+1}$  we seek to infer a scalar predictive mean value  $\hat{\mathbf{g}}_{v+1}^{(l)}$  and variance  $\hat{\sigma}_{\hat{\mathbf{g}}^{(l)}v+1}^2$  for each of the unknown output components  $\hat{\mathbf{g}}_{v+1}^{(l)}$ , before we update the posterior Gaussian distribution.

**Prediction** In order to keep the final equations simple and readable, four auxiliary variables are introduced

$$\begin{aligned}\mathbf{s}_{v+1} &= \mathbf{S}_v^{(l)} \mathbf{k}_{v+1} \\ \gamma_{v+1}^2 &= k_{\text{mag}}(\mathbf{g}_{v+1}, \mathbf{g}_{v+1}) - \mathbf{k}_{v+1}^\top \mathbf{S}_v^{(l)} \mathbf{k}_{v+1} \\ \mathbf{h}_{v+1} &= \Sigma_v^{(l)} \mathbf{s}_{v+1} \\ \hat{\sigma}_{\hat{\mathbf{f}}v+1}^2 &= \gamma_{v+1}^2 + \mathbf{s}_{v+1}^\top \mathbf{h}_{v+1},\end{aligned}\tag{26}$$

where  $\mathbf{k}_{v+1}$  is a column vector with  $(\mathbf{k}_{v+1})_i = k_{\text{mag}}(\mathbf{g}_i, \mathbf{g}_{v+1})$ . The  $n$  components of the predictive mean and variance are obtained by

$$\begin{aligned}\hat{\mathbf{g}}_{v+1}^{(l)} &= \mathbf{s}_{v+1}^\top \mu_v^{(l)} \\ \hat{\sigma}_{\hat{\mathbf{g}}^{(l)}v+1}^2 &= \sigma_{\mathbf{n}}^2 + \hat{\sigma}_{\hat{\mathbf{f}}v+1}^2\end{aligned}\tag{27}$$

for  $l = 1, \dots, n$ . They are joint into a single multidimensional predictive distribution for the input  $\mathbf{g}_{v+1}$  via

$$\mathcal{N}\left(\left(\hat{\mathbf{g}}_{v+1}^{(1)}, \dots, \hat{\mathbf{g}}_{v+1}^{(n)}\right)^\top, \text{diag}(\hat{\sigma}_{\hat{\mathbf{g}}^{(1)}v+1}^2, \dots, \hat{\sigma}_{\hat{\mathbf{g}}^{(n)}v+1}^2)\right).\tag{28}$$

**Model Update** After observing the real output components  $\hat{\mathbf{g}}_{v+1}^{(l)}$  the recursive variables  $\mu_v^{(l)}$ ,  $\Sigma_v^{(l)}$  and  $\mathbf{S}_v^{(l)}$  are updated to the posterior distribution for  $(v+1)$ -th observation via estimation error reduction

$$\begin{aligned}
\mu_{v+1}^{(l)} &= \begin{pmatrix} \mu_v^{(l)} \\ \hat{\mathbf{g}}_{v+1}^{(l)} \end{pmatrix} + \frac{\hat{\mathbf{g}}_{v+1}^{(l)} - \hat{\mathbf{g}}_{v+1}^{(l)}}{\hat{\sigma}_{\mathbf{f}_{v+1}}^2} \begin{pmatrix} \mathbf{h}_{v+1} \\ \hat{\sigma}_{\mathbf{f}_{v+1}}^2 \end{pmatrix} \\
\Sigma_{v+1}^{(l)} &= \begin{pmatrix} \Sigma_v^{(l)} & \mathbf{h}_{v+1} \\ \mathbf{h}_{v+1}^\top & \hat{\sigma}_{\mathbf{f}_{v+1}}^2 \end{pmatrix} - \frac{1}{\hat{\sigma}_{\hat{\mathbf{g}}^{(l)}v+1}^2} \begin{pmatrix} \mathbf{h}_{v+1} \\ \hat{\sigma}_{\mathbf{f}_{v+1}}^2 \end{pmatrix} \begin{pmatrix} \mathbf{h}_{v+1} \\ \hat{\sigma}_{\mathbf{f}_{v+1}}^2 \end{pmatrix}^\top \\
\mathbf{S}_{v+1}^{(l)} &= \begin{pmatrix} \mathbf{S}_v^{(l)} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} + \frac{1}{\gamma_{v+1}^2} \begin{pmatrix} \mathbf{s}_{v+1} \\ -1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_{v+1} \\ -1 \end{pmatrix}^\top.
\end{aligned} \tag{29}$$

### 3.5.3 Discussion and Application

The generalized online GP allows for the same precision as its offline pendant. Given a set of training data  $\mathcal{D}_v$ , a set of hyperparameters  $H$  in kernel (22) and a new input value  $\mathbf{g}_{v+1}$ , the generalized GP model yields the same predictive distribution as obtained by mean and variance defined in (27), which can be verified by simple calculations. Further, while the standard GP has complexity  $\mathcal{O}(h^3)$ , as it is governed by matrix inversion, the recursive GP algorithm reduces the computational complexity to  $\mathcal{O}(h^2)$ . As observations are added consecutively to the training data, the algorithm is suitable for learning the GP model online. For application in practice, however, the recursive GP requires the following adaptations:

**Fixed-Budget Online GP** We denote by dictionary the set of training data on which the next prediction is made. So far, the recursive GP has an ever-growing dictionary. In order to limit the requirements of computation and memory we introduce the fixed-budget online GP, possessing a limited dictionary size. In case a predefined budget  $\beta$  is exhausted, i.e. on obtaining  $(\beta+1)$ -th observation, we select and remove the least relevant pair  $(\mathbf{g}_j, \hat{\mathbf{g}}_j) \in \mathcal{D}_{\beta+1}$ . The index  $j$  is determined by a cost function measuring the information loss caused by removal of the pair

$$j = \arg \min_i \left( \frac{\left( \mathbf{S}_{\beta+1}^{(l)} \boldsymbol{\mu}_{\beta+1}^{(l)} \right)_i}{\left( \mathbf{S}^{(l)} \right)_{i,i}} \right)^2 \tag{30}$$

Then, the operators  $(\cdot)_{-i}$  and  $(\cdot)_{-i,-i}$ , which remove the  $i^{\text{th}}$  row in a vector and both the  $i^{\text{th}}$  row and column in a matrix, respectively, are applied to remove the corresponding information from the regression variables. Hence, by removal of

$j$ -th observation

$$\begin{aligned}
\boldsymbol{\mu}_{\beta+1}^{(l)} &= \left( \boldsymbol{\mu}_{\beta+1}^{(l)} \right)_{-j} \\
\Sigma_{\beta+1}^{(l)} &= \left( \Sigma_{\beta+1}^{(l)} \right)_{-j,-j} \\
\mathbf{S}_{\beta+1}^{(l)} &= \left( \mathbf{S}_{\beta+1}^{(l)} \right)_{-j,-j} - \frac{\left( \mathbf{S}_{\beta+1}^{(l)} \right)_{-j,j} \left( \mathbf{S}_{\beta+1}^{(l)} \right)_{-j,j}^\top}{\left( \mathbf{S}_{\beta+1}^{(l)} \right)_{j,j}}
\end{aligned} \tag{31}$$

we re-obtain in the  $(\beta + 1)$ -th recursion step a dictionary of size  $\beta$ .

**Forgetting Strategy** Under non-stationary working circumstances the generalized recursive GP additionally needs a strategy to “forget” past samples to be able to adapt to a varying underlying mapping (Van Vaerenbergh et al, 2012). The basic idea of the forgetting strategy is to slightly reset the main vector and the covariance matrix to their initial values after each kernel updates. A forgetting factor  $\delta \in (0, 1]$  is applied to the mean and covariance of the generalized online GP by

$$\begin{aligned}
\Sigma_v^{(l)} &= \delta \Sigma_v^{(l)} + (1 - \delta) \mathbf{K}_v^{(l)} \\
\boldsymbol{\mu}_v^{(l)} &= \sqrt{\delta} \boldsymbol{\mu}_v^{(l)}.
\end{aligned} \tag{32}$$

The smaller the forgetting factor  $\delta$ , the faster the algorithm is able to track changes, meanwhile it learns less. When choosing  $\delta = 0$ , the regression is with complete forgetting, whereas  $\delta = 1$  corresponds to no forgetting.

## 4 Distribution over Dual Quaternions

This section discusses a completion of the GP framework for modeling rigid motion dynamics. We present a projection from Euclidean space to the sphere  $S_3$ , which is used to define a probability distribution on the rigid motions and as integrator from velocity space  $TSE(3)$  to  $SE(3)$ .

### 4.1 Projected Gaussian Distribution on $SE(3)$

Due to defining input/output spaces in form of dynamics (2), the generalized GP model provides the uncertainty prediction as a Gaussian distribution in velocity space  $TSE(3)$ . The standard dynamics formulation, in contrast, maps from state space to state space,

$$\tilde{f} : SE(3) \times T_g SE(3) \rightarrow SE(3). \tag{33}$$

Therefore, we provide here a method to transform the predictive Gaussian uncertainty estimate from the velocity space  $TSE(3)$  into  $SE(3)$ .

We start with providing a probability distribution on pure rotation quaternions, visualized in Fig. 2, by employing a projec-

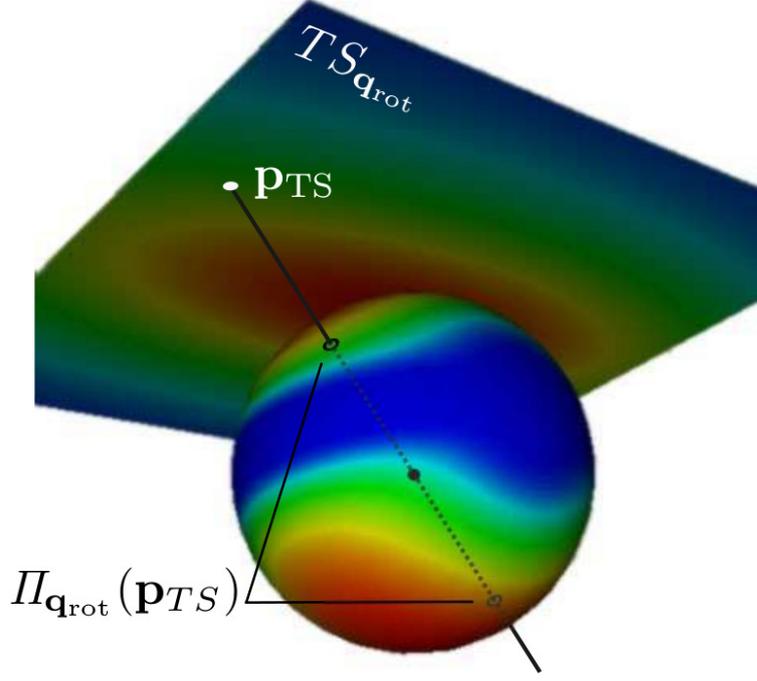


Figure 2: The Gaussian distribution in the tangent plane to the north pole is projected to the sphere  $S_2$  via central projection. Hence, each point in the plane is projected to both opposite intersections of the sphere with a straight line through the center of the sphere and the point in the tangent plane.

tion from the tangent space onto the unit sphere  $S_3$ . Let  $TS_{\mathbf{q}_{\text{rot}}}$  be the tangent space of  $\mathbf{q}_{\text{rot}}$  to the unit hypersphere  $S_3$ . The central projection  $\Pi_{\mathbf{q}_{\text{rot}}}$  propagates a 3D Gaussian distribution  $\mathcal{N}(\mu_{\text{rot}}, \Sigma_{\text{rot}})$  given in tangent space  $TS_{\mathbf{q}_{\text{rot}}}$  to the sphere  $S_3$  via

$$\begin{aligned} \Pi_{\mathbf{q}_{\text{rot}}} : TS_{\mathbf{q}_{\text{rot}}} &\rightarrow S_3/\{\pm 1\} \\ \Pi_{\mathbf{q}_{\text{rot}}}(\mathbf{p}_{TS}) &= \pm \frac{\mathbf{p}_{\mathbf{B}}}{\|\mathbf{p}_{\mathbf{B}}\|}, \end{aligned} \quad (34)$$

where  $\mathbf{p}_{\mathbf{B}} = \mathbf{q}_{\text{rot}} + \mathbf{B}\mathbf{p}_{TS} \in \mathbb{R}^4$  and  $\mathbf{p}_{TS} \in \mathbb{R}^3$  any point in the tangent space  $TS_{\mathbf{q}_{\text{rot}}}$ . The basis  $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\} \in \mathbb{R}^{4 \times 3}$  is the canonical representation of the 3D tangent space  $TS_{\mathbf{q}_{\text{rot}}}$  in the space  $\mathbb{R}^4$ . When projecting the Gaussian density  $\varphi_{TS}$  of the distribution  $\mathcal{N}(\mu_{\text{rot}}, \Sigma_{\text{rot}})$  to the sphere  $S_3/\{\pm 1\}$  using (34), its properties are inherited, as (34) is a bijective mapping. To obtain a probability density, we normalize over the new domain  $S_3/\{\pm 1\}$ ,

$$\varphi_{S_3} = \frac{1}{C} \varphi_{TS}, \quad \text{where } C = \int_{S_3/\{\pm 1\}} \varphi_{TS}(\Pi_{\mathbf{q}_{\text{rot}}}^{-1}(\mathbf{q})) d\mathbf{q}, \quad (35)$$

following previous investigations in Lang and Feiten (2012) and project the set closure  $\{\infty\}$  of  $TS_{\mathbf{q}}$  to the set of orthogonal vector-valued points  $\mathbf{q}_{\text{rot}}^\perp$  on the sphere,

$$\Pi_{\mathbf{q}_{\text{rot}}}(\infty) = \mathbf{q}_{\text{rot}}^\perp := \{\mathbf{q}' \in S_3 | \mathbf{q}' \perp \mathbf{q}_{\text{rot}}\}. \quad (36)$$

As in the tangent space the set closure  $\{\infty\}$  has probability 0, we obtain a continuous completion of the projected probability distribution on the sphere. The projected probability distribution is denoted by  $\mathcal{N}(\mathbf{q}_{\text{rot}}, \mu_{\text{rot}}, \Sigma_{\text{rot}})$ .

**Remark:** *The central projection satisfies the desired property of mapping the Gaussian distribution on each hemisphere: opposing quaternions that represent the same orientation receive the same probability value. This fits exactly the topology of unit quaternions. Consequently, the positive probability on the sphere of distribution  $\mathcal{N}(\mathbf{q}_{\text{rot}}, \mu_{\text{rot}}, \Sigma_{\text{rot}})$  does not reach a  $\pi$  rotation from the tangent point  $\mathbf{q}_{\text{rot}}$ . In applications, where rotational velocities and uncertainties are so large, that an orientation estimate could be off by  $\pi$ , another mapping should be considered to transform the uncertainty to the orientations, e.g. the stereographic projection. Further, the projection's inverse  $\Pi_{\mathbf{q}_{\text{rot}}}^{-1}(\mathbf{q}')$  is calculated analytically using simple trigonometric calculus and the intercept theorem  $\forall \mathbf{q}' \in S_3, \mathbf{q}' \neq \mathbf{q}_{\text{rot}}$ .*

Second, we extend the projection (34) to full 6-DoF rigid motions. As Prop. 1 holds, defining a Gaussian counterpart on  $S_3/\{\pm 1\} \times \mathbb{R}^3$  is equivalent to defining one on  $SE(3)$ . Further, an isomorphism  $T_{\mathfrak{g}}SE(3) \simeq TS_{\mathbf{q}_{\text{rot}}} \times \mathbb{R}^3$  is induced by (5), for  $\mathfrak{g}$  representing  $(\mathbf{q}_{\text{rot}}, \mathbf{q}_{\text{tra}})$ . Thus, the full projection is defined by

$$\begin{aligned} \Pi_{\mathfrak{g}} : T_{\mathfrak{g}}SE(3) &\rightarrow S_3 \times \mathbb{R}^3 \\ \Pi_{\mathfrak{g}}((\mathbf{p}_{TS}, \mathbf{p}_{\mathbb{R}^3})) &= \begin{cases} \Pi_{\mathbf{q}_{\text{rot}}}(\mathbf{p}_{TS}), & \text{for } \Pi_{\mathfrak{g}}|_{TS_{\mathbf{q}_{\text{rot}}}} \\ \mathbf{q}_{\text{tra}} + \mathbf{p}_{\mathbb{R}^3}, & \text{for } \Pi_{\mathfrak{g}}|_{\mathbb{R}^3}. \end{cases} \end{aligned} \quad (37)$$

A distribution on the rigid motion space  $S_3/\{\pm 1\} \times \mathbb{R}^3$  is obtained by a regular concatenation of the projected Gaussian  $\mathcal{N}(\mathbf{q}_{\text{rot}}, \mu_{\text{rot}}, \Sigma_{\text{rot}})$  with a standard Gaussian  $\mathcal{N}(\mu_{\text{tra}}, \Sigma_{\text{tra}})$  in the Euclidean space of translations  $\mathbb{R}^3$ ,

$$\mu = \begin{pmatrix} \mu_{\text{rot}} \\ \mu_{\text{tra}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{\text{rot}} & 0 \\ 0 & \Sigma_{\text{tra}} \end{pmatrix}. \quad (38)$$

An analogous approach has been performed previously by the author (Lang and Feiten, 2012).

## 4.2 Integration from Tangent Bundle to Rigid Motions

In this section we discuss the use of the full projection (37) as integrator from the velocity space to the special Euclidean group  $SE(3)$ . We show that projection (37) composed with our dynamics formulation (2) provides the standard dynamics (33),  $\tilde{f} = \Pi_{\mathfrak{g}} \circ f$ .

It is well-known that rotational velocities  $\omega$  are obtained from quaternions  $\mathbf{q}_{\text{rot}}$  by

$$\omega(t) = 2 \frac{d\mathbf{q}_{\text{rot}}(t)}{dt} \mathbf{q}_{\text{rot}}(t). \quad (39)$$

However, in many practical applications the explicit form of  $\mathbf{q}_{\text{rot}}(t)$  is unknown. Thus, we propose to approximate the rotational velocity  $\omega$  by  $\omega_{\Delta} = \omega \Delta t$ , obtained by zero-order hold discretization with constant time intervals  $\Delta t > 0$ .

After discretizing dynamics (2), the output at a certain time instance  $t > 0$  becomes  $(\mathbf{g}(t), \mathbf{v}_{\mathbf{g}}(t)) \in TSE(3)$ . Employing isomorphism  $T_{\mathbf{g}}SE(3) \simeq TS_{\mathbf{q}_{\text{rot}}} \times \mathbb{R}^3$ , the tangent space velocity  $\mathbf{v}_{\mathbf{g}} = (\mathbf{v}_{TS}, \mathbf{v}_{\mathbb{R}^3})$  is decomposed into a rotation and a translation part and the relation between rotational tangent space velocity and the discrete-time rotational velocity becomes apparent,

$$\boldsymbol{\omega}_{\Delta} = 2 \mathbf{v}_{TS}. \quad (40)$$

The transforming unit quaternion  $\mathbf{q}_{\Delta}$ , rotating from  $t$ -th to  $(t + \Delta t)$ -th orientation, can be calculated using the exponential function as integrator of the discrete time rotational velocity  $\boldsymbol{\omega}_{\Delta}$ ,

$$\mathbf{q}_{\Delta} = \exp\left(\frac{\boldsymbol{\omega}_{\Delta}}{2}\right). \quad (41)$$

For integrating directly to  $\mathbf{g}(t + 1)$  however, we propose to employ the full projection (37)

$$\Pi_{\mathbf{g}}((\mathbf{v}_{TS}, \mathbf{v}_{\mathbb{R}^3})) = (\mathbf{q}_{\text{rot}}(t + \Delta t), \mathbf{q}_{\text{tra}}(t + \Delta t)), \quad (42)$$

where  $\mathbf{q}_{\text{rot}}(t + \Delta t) = \mathbf{q}_{\Delta} \mathbf{q}_{\text{rot}}$  and  $\mathbf{q}_{\text{tra}}(t + \Delta t) = \mathbf{q}_{\text{tra}} + \mathbf{v}_{\mathbb{R}^3}$ . Thus, we spare the extra quaternion multiplication and after dual quaternion composition (7) we end up with the output of the standard dynamics (33).

## 5 Experimental Evaluation

In this section we validate the introduced framework for generalized GP modeling over rigid motions ( $\text{GP}_{\mathbb{H}_D}$ ) in comparison to the state of the art  $\text{GP}_{\mathbb{R}^6}$ . Our evaluations confirm that our approach significantly outperforms the traditional way of including orientation into the GP via Euler angles. We carry out a numerical performance analysis on simulated data, in which we calculate the prediction error as a function of the rotation magnitude and of the number of training samples. In a second experiment we compare the behavior of GP regression across the three unit quaternion distance metrics introduced in Sec. 3.3.1. For convenience, we additionally include the Euclidean distance using the Euler angle representation in the comparison. The experiment uses real human gaze motion data. A third experiment is based on real data of a human object handover trajectory. We perform online GP modeling over 6-DoF rigid motions using the recursive GP over dual quaternions and investigate the ability of the introduced framework of a generalized GP to model usual human motion behavior.

### 5.1 Numerical Evaluation

In this section we evaluate the approximation quality of the  $\text{GP}_{\mathbb{H}_D}$  over dual quaternions in comparison to the state of the art  $\text{GP}_{\mathbb{R}^6}$ . We evaluate the prediction accuracy of  $\text{GP}_{\mathbb{H}_D}$  and  $\text{GP}_{\mathbb{R}^6}$  as a function of the sample size of the training data and of the rotation range in the dynamics drift term of ten synthetically generated ground truth dynamics of form (2). The

evaluation shows the necessity to introduce the  $\text{GP}_{\mathbb{H}_D}$  over dual quaternions for modeling nonlinear dynamics of rigid motions, as it clearly outperforms the traditional  $\text{GP}_{\mathbb{R}^6}$ .

**Experimental Conditions** We randomly generate ten dynamics drift terms  $f$  of form (2) based on the trigonometric functions  $\{\sin, \cos, \tan\}$ , with constants in the range  $[\frac{1}{8}, 5]$  and basal operations  $\{\pm, \times, \div, \circ\}$ . Each 6-DoF dynamics component comprises three to seven operations. As an example, one of the dynamics is given by

$$f(\mathbf{g}) = \begin{pmatrix} c \left( \frac{\pi}{3} - \cos(\mathbf{q}_{\text{re}_x}) + \sin(\mathbf{q}_{\text{re}_y} + \mathbf{q}_{\text{du}_x}) \right) \\ c \left( 1 - \mathbf{q}_{\text{re}_x} \sin(\mathbf{q}_{\text{du}_z} + \mathbf{q}_{\text{du}_x} + \mathbf{q}_{\text{re}_z}) \right) \\ c \left( \frac{\pi}{8} + \sin(\mathbf{q}_{\text{du}_y} \mathbf{q}_{\text{re}_z} - 1) \right) \\ 2 \cos(\mathbf{q}_{\text{re}_x} \mathbf{q}_{\text{re}_y}) \sin(\mathbf{q}_{\text{du}_x} + \mathbf{q}_{\text{re}_x}) + \mathbf{q}_{\text{du}_x} \\ \sin\left(\tan\left(\frac{\mathbf{q}_{\text{re}_z}}{3} + \frac{1}{2}\right) - 2 \mathbf{q}_{\text{re}_y}\right) \\ 3 \cos(\mathbf{q}_{\text{du}_y} \mathbf{q}_{\text{du}_z}) - \frac{\mathbf{q}_{\text{re}_y}}{4} \end{pmatrix}, \quad (43)$$

where  $\mathbf{g} = (\mathbf{q}_{\text{re}}, \mathbf{q}_{\text{du}})$  and  $c > 0$  a constant. For each of the dynamics, we generate randomly sampled training data sets and a ground truth data set for evaluating the estimation accuracy. For reference, each data set is converted to the traditional representation  $\mathbf{g} = (\phi, \psi, \vartheta, t_x, t_y, t_z)^\top$  invoking (5). To assure numerical stability we add an artificial white Gaussian noise with arbitrarily selected standard deviation  $\sigma = 3 \cdot 10^{-3}$  to the training data sets. We use the squared exponential kernels (11) respectively (22) in the GP models.

In order to analyse the prediction quality of the  $\text{GP}_{\mathbb{H}_D}$  and the  $\text{GP}_{\mathbb{R}^6}$ , we employ the RMSE between the predictive mean  $\hat{\mathbf{g}}$  and the true function value  $f(\mathbf{g})$ . For the translational output  $f(\mathbf{g})|_{\text{tra}}$  we find no significant difference in the prediction quality of  $\text{GP}_{\mathbb{H}_D}$  and the traditional GP in our experiments. For growing training data sets, the average RMSE in translational GP prediction converges around  $10^{-3}$  for both models, the size of which corresponds to the magnitude of the noise with the difference in the RMSEs ranging in the magnitude of  $10^{-4}$ . Therefore, we concentrate in the remainder of this evaluation on the quality of the rotational part of the predictive mean  $\hat{\mathbf{g}}_{\text{rot}}$ . In order to obtain reliable evaluation results, the RMSE is calculated for a test set  $\{(\mathbf{g}_i, f(\mathbf{g}_i))\}_{i=1}^{10000}$  of iid pairs. The experiment is implemented in Matlab using the GPML toolbox available on [gaussianprocess.org](http://gaussianprocess.org) (Rasmussen and Nickisch, 2010). For the hyperparameter optimization, we use the algorithm provided in the toolbox with a Gaussian likelihood. We initialize the training phase each time with the same set of default hyperparameters  $H_0 = (\lambda_0, \sigma_{f0}, \sigma_{n0})$  for  $\text{GP}_{\mathbb{H}_D}$ . As the signal range in  $\text{GP}_{\mathbb{R}^6}$  is  $\pi$  times the signal range in  $\text{GP}_{\mathbb{H}_D}$ , we set the default starting hyperparameters in  $\text{GP}_{\mathbb{R}^6}$  to  $(\lambda_0, \pi\sigma_{f0}, \pi\sigma_{n0})$ .

**RMSE Depending on Rotation Magnitude** In the first evaluation, we compare the prediction accuracy of the GP models, depending on the rotation magnitude in (43). We double the constant  $c$  in the dynamics  $f$  from 0.125 to 8 in stepwise. The number of training samples is fixed to 2000 for all 70 (seven steps times ten dynamics) repetitions, and each time, the learning is performed per dynamics with the same set of uniformly drawn random samples. The average prediction RMSEs with standard deviation of both GP models are contrasted. In all tested conditions the  $\text{GP}_{\mathbb{H}_D}$  clearly

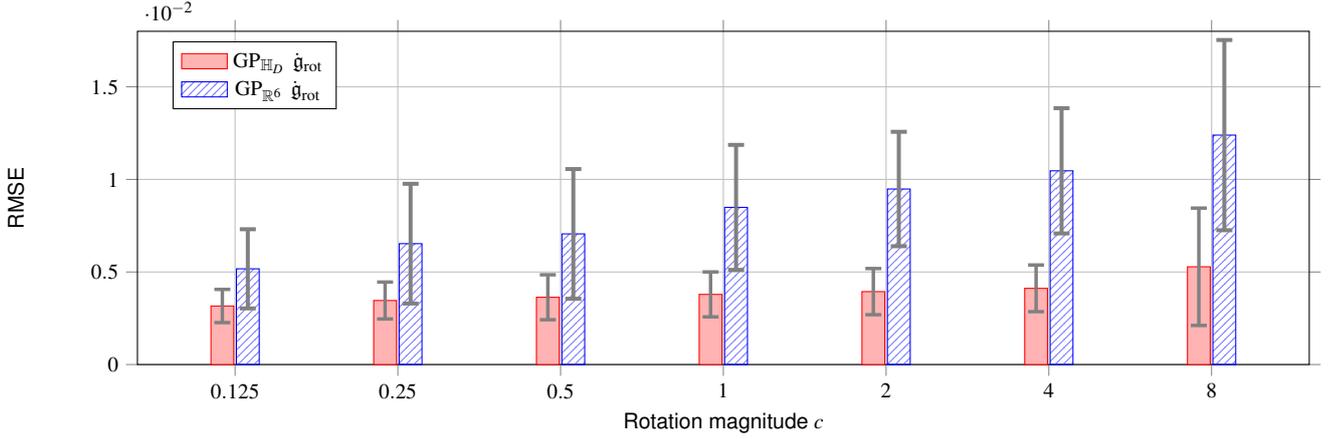


Figure 3: The prediction quality is visualized for a fixed sample size of the training data, but increasing rotation magnitude in the dynamics  $f$ . The bars depict the average RMSEs in the rotational velocity prediction, where solid red colors show the GP $_{\mathbb{H}_D}$  and striped blue the traditional GP. The standard deviation among the ten different dynamics is shown by the error bars.

outperforms the GP $_{\mathbb{R}^6}$  prediction accuracy as visualized in Fig. 3. For small rotational velocities, the error difference between the models is less significant and the superiority of GP $_{\mathbb{H}_D}$  may not counterbalance the additional computational complexity. For increased rotation magnitude the performance of the GP $_{\mathbb{R}^6}$  is drastically decreased while the accuracy of the GP $_{\mathbb{H}_D}$  is significantly less affected. The largest condition tested,  $c = 8$ , incorporated up to three full turns in the randomly generated dynamics. The corresponding rightmost result in Fig. 3 indicates that the modeling abilities of GP $_{\mathbb{H}_D}$  can be exhausted for even larger rotation magnitudes. At a certain magnitude the dynamics gets too edged for appropriate application of the squared exponential kernel.

**RMSE Depending on Number of Samples** In the second numerical evaluation, we illustrate the GP prediction quality for GP $_{\mathbb{H}_D}$  and the traditional GP $_{\mathbb{R}^6}$  for different numbers of training samples. The sample set size is relevant for the ability of a GP to capture the underlying function, as the stepwise orientation change increases for smaller sample sets. In this setup the rotation magnitude is fixed to  $c = 1$ . The model learning is performed on increasing sizes of training data sets, doubled in each step from 125 to 8000 samples. Figure 4 shows that the GP $_{\mathbb{H}_D}$  approximates the underlying ground truth dynamics significantly better than the GP $_{\mathbb{R}^6}$  except for the condition with 125 training samples. In the first tested condition, no significant difference could be determined by a independent-sampled t-test due to the large variance in the RMSE,  $t(9) = 0.74, p = 0.47$ . On the one hand, the results indicate that the GP $_{\mathbb{H}_D}$  has difficulties finding optimal hyperparameters for very sparse sample sets. On the other hand, this evaluation proves the superiority of GP $_{\mathbb{H}_D}$  in modeling rigid motion dynamics involving significant rotation over GP $_{\mathbb{R}^6}$  for sufficiently large sample sets.

## 5.2 GP Behavior depending on Unit Quaternion Distance

In this section a real data experiment is conducted to compare the estimation behavior of the generalized GP depending on the underlying distance metric over unit quaternions in the squared exponential kernel. The results are compared to the

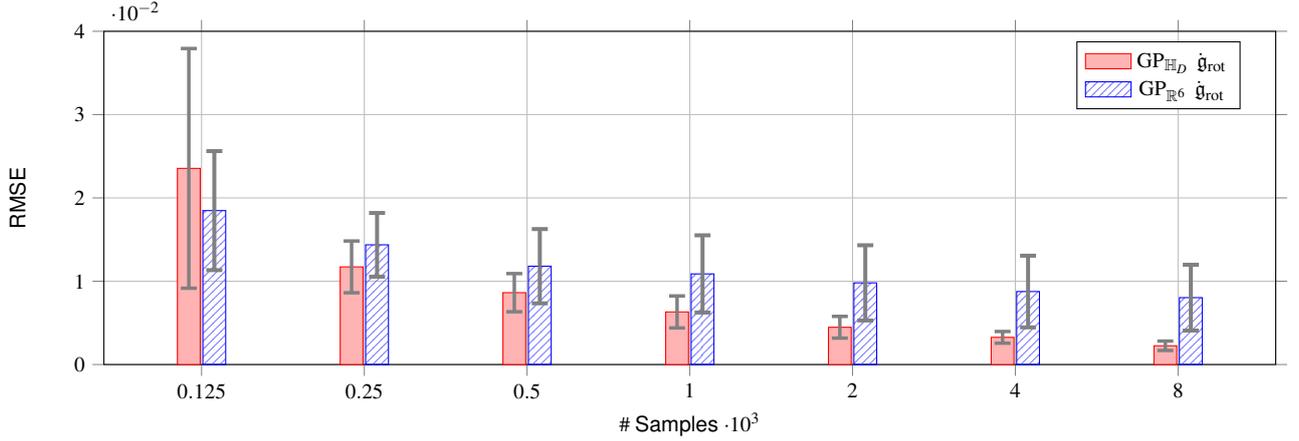


Figure 4: For increasing numbers of samples used in GP training the prediction quality is visualized. The solid red bars depict the RMSE in the rotation prediction of GP $_{\mathbb{H}_D}$  and the striped blue bars of traditional GP.

traditional GP model incorporating Euler angles as rotation representation and the standard Euclidean distance.

**Experimental Setup** A human participant draws repetitively 40 large “Z” letters on a flip chart. Even though not instructed to do so, the human naturally follows the pen with the eye. Using a combined system of motion capturing by Qualisys motion tracker (Sweden) and Dikablis eye tracking (Germany), the human gaze motion was captured at 100 Hz in the workspace coordinates. Figure 5 illustrates one experiment trial. From the measurements, we obtain approximate derivatives of the gaze motion using zero-order hold. We consider solely the head and gaze rotation, as we aim to compare *pure rotation modeling behavior* of the generalized GP describing human gaze dynamics of form (2). The captured data includes rotation up to 1.4 rad per time step, as gaze jumps in saccades, and involves a high noise level as the eye tracking system in our setup is nonlinearly delayed against the motion capture system. We intentionally select a challenging experimental setting to exhaust the GP modeling abilities as only borderline model abilities bring up the behavioral differences. We train the GP models on the first 34 out of the 40 trials and test its performance on the remaining six repetitions. As in the numerical evaluation, we fix the initial hyperparameter values  $H_0 = (\lambda_0, \sigma_{f0}, \sigma_{n0})$  for all GPs over unit quaternions and scale signal variance and signal noise about  $\pi$  for the reference GP over Euler angles.

**Visualization Interpretation** Figure 6 illustrates the prediction behavior of the four compared GP models, namely Arc, Norm,  $TS_{\mathbf{q}}$  and  $(\phi, \psi, \vartheta)$ . The GP models we compare are the following: in first row a generalized GP using the arc metric (15) is shown, in second row the underlying metric is the quaternion norm (13), in third row the generalized GP is based on the tangent space distance (16). Fourth row illustrates the prediction behavior of the standard GP model over Euler angles using the Euclidean distance in the kernel function (11) for reference. More precisely, the first three rows visualize each the rotational velocity dimensions,  $\{\dot{\mathbf{g}}_{TS_x}, \dot{\mathbf{g}}_{TS_y}, \dot{\mathbf{g}}_{TS_z}\}$  of the tangent space to the sphere  $S_3$  for the conditions Arc, Norm and  $TS_{\mathbf{q}}$ , while the fourth row visualizes the synchronized rotational velocities  $\{\dot{\phi}, \dot{\psi}, \dot{\vartheta}\}$  for the Euler angles  $(\phi, \psi, \vartheta)$ . For better visibility of the differences in the estimation we do not show all test trials of the “Z”

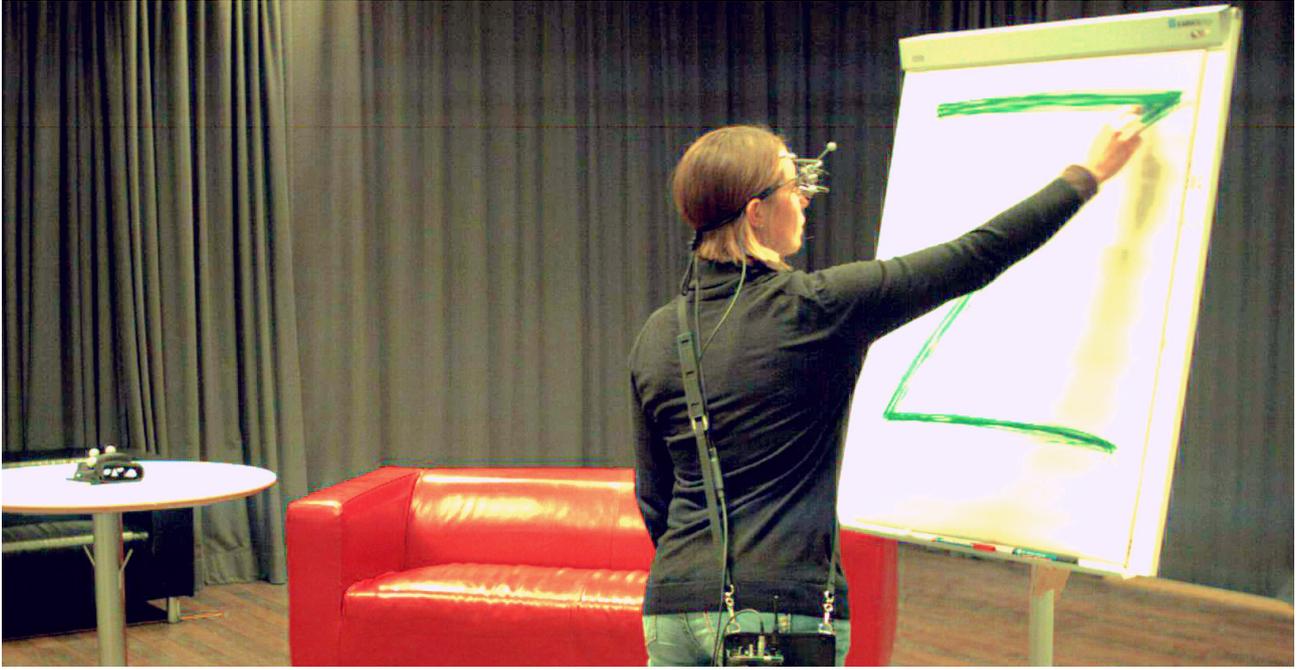


Figure 5: A human participant repetitively draws large “Z” letters on a flip chart, while naturally tracking the pen with the eye. The human gaze motion is captured in workspace coordinates using a combination of eye tracking and marker based head motion tracking.

repetitions in every column, but zoom in to the trials 1 and 2 in the first column, 3 and 4 in the second column and 5 and 6 in the third column. We visualize the motion velocities over time in seconds and reset the runtime to zero at the beginning of every trial. Every second trial is contrasted per column by yellow background and the vertical black line in every plot indicates the beginning of that next trial. Per dimension and test condition (Arc, Norm,  $TS_{\mathbf{q}}$  and  $(\phi, \psi, \vartheta)$ ) we depict by a red dashed line the ground truth velocities, by a blue solid line the mean prediction and by a gray shaded area the  $2\sigma$ -confidence region. Overall, the GP based on the arc metric (row one) outperforms the others, which can be seen most clearly in the third component of the velocity predictions. The squared exponential kernel over the quaternion norm overestimates the correlation of orientations slightly  $\forall \mathbf{q}, \mathbf{q}' \in S_3$ . The kernel used in third row holds  $k_{TS}(\mathbf{q}, \mathbf{q}') < k_{\text{arc}}(\mathbf{q}, \mathbf{q}')$   $\forall \mathbf{q} \neq \mathbf{q}'$  in contrast. Therefore, it shows a much faster decrease in correlation for growing dissimilarity of the quaternions. In our illustrative example it forces the optimization to significantly smaller length scale values than in the conditions Arc and Norm. This behavior can be valuable in applications where highly precise orientation is essential. The generalized GP models have in common that all of them significantly outperform the traditional GP model (condition  $(\phi, \psi, \vartheta)$ ) which has problems capturing the large rotational velocities.

### 5.3 Online Regression of Human Object Handover

The aim of this real data experiment is to generate a human motion model online. Hence, the recursive generalized GP over dual quaternions is employed to estimate dynamics of form (2) on  $SE(3)$ . We do not rely on task specific prior knowledge in our approach to guarantee maximal adequacy for arbitrary tasks. The example we choose in this experiment, visualized

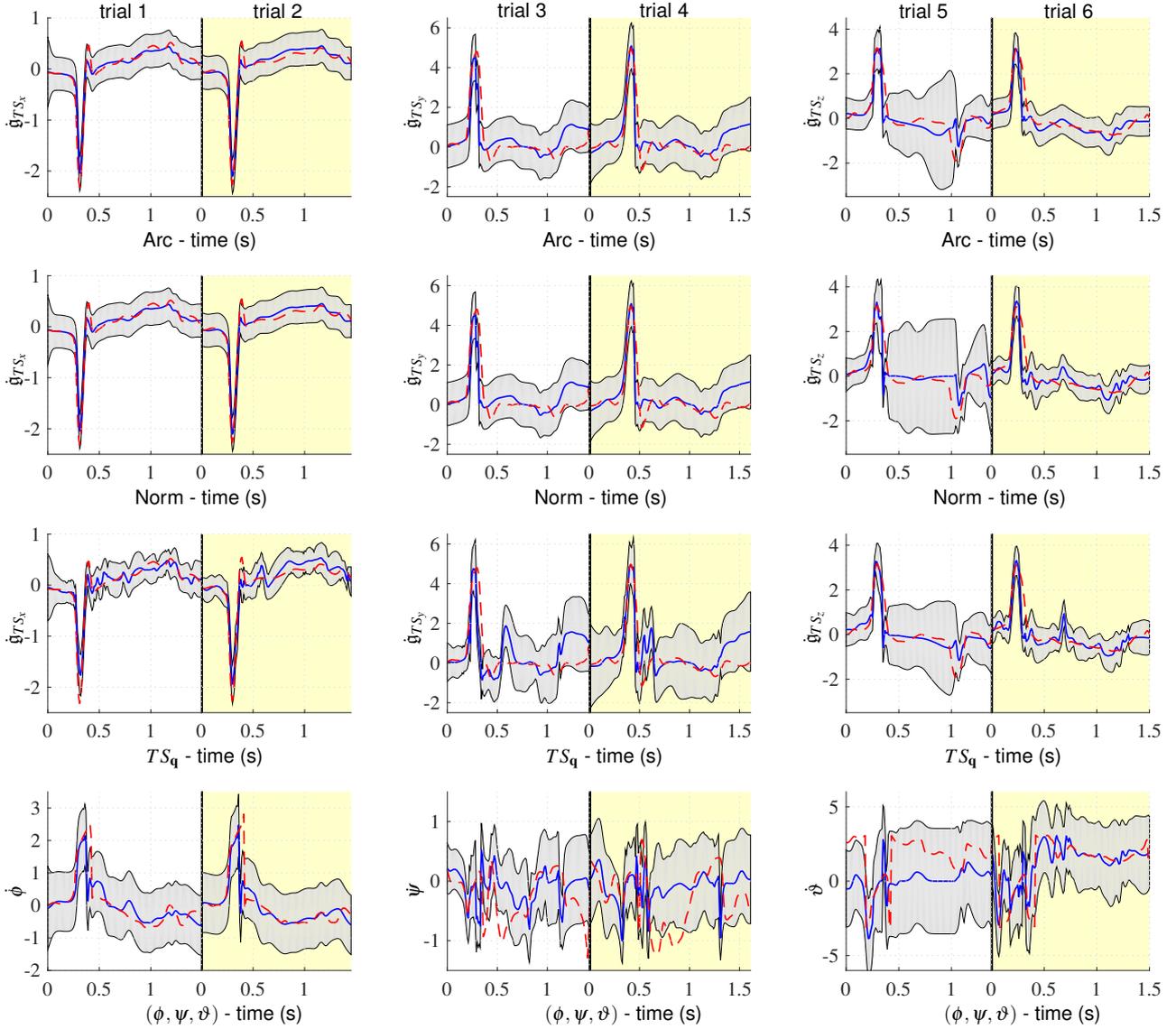


Figure 6: The prediction accuracy of the human gaze motion is illustrated. Blue solid lines show the mean prediction, red dashed lines the ground truth and gray shaded areas the prediction uncertainty. The columns correspond to the tangent bundle respectively rotational velocity dimensions, the rows to the different conditions. All columns are synchronized over time. In row one the GP model is based on the arc distance, in row two on the quaternion norm, in row three on the distance projected to the tangent space and in row four on the Euclidean distance in combination with the Euler angle representation. Column wise a second trajectory is highlighted by yellow background color.

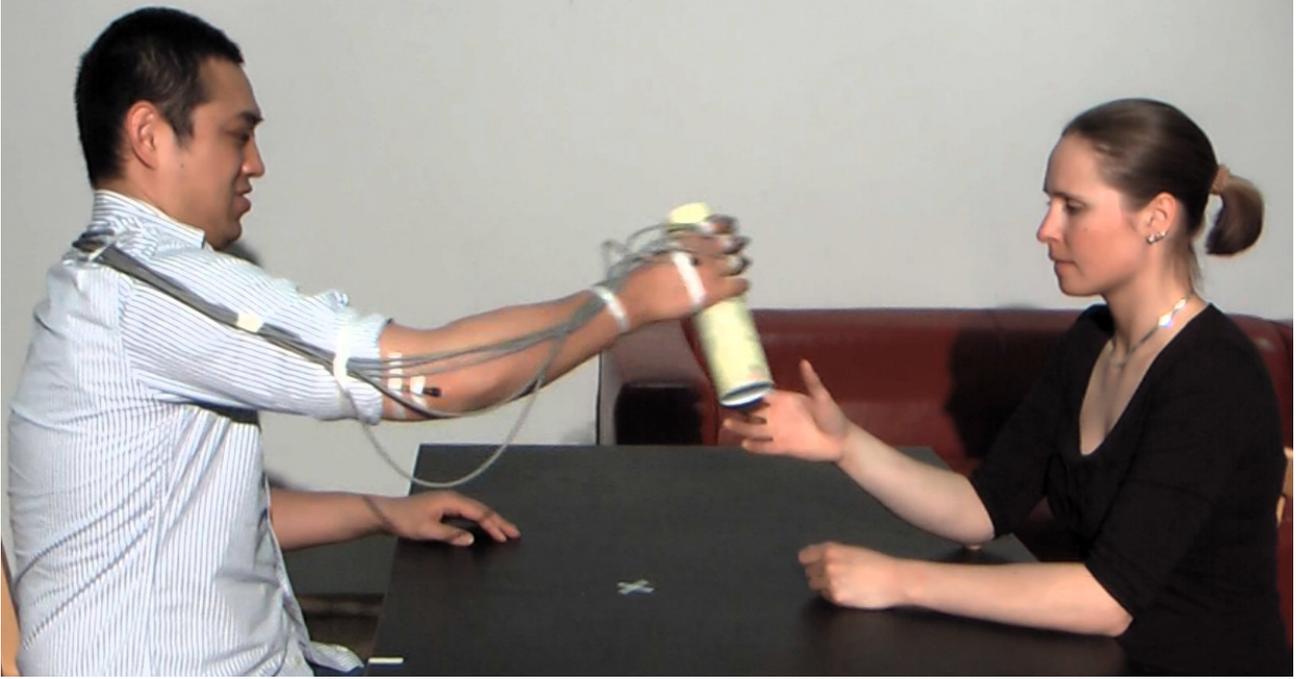


Figure 7: A natural 6D human handover is visualized. The object handover starts from an initial hand configuration, performs the object grasp and terminates at the final configuration after the object is handed over. The handover motion in this trial incorporates a half turn of the object.

in Figure 7, concerns a human object handover including significant rotational motion, namely turning the object around by a  $\pi$  rotation.

**Experimental Conditions** A pair of human participants performed an object handover where the passers hand motion was tracked using a magnetic motion tracking system (Polhemus Liberty) at 80 Hz. The subject handed over a cylinder (21 cm in length, 5.5 cm in diameter and weight 280 g). We model the human motion using the generalized online GP model and evaluate the results using the captured 6-DoF data as ground truth. The forgetting factor is fixed to 1 and the dictionary size limited to 100 observations, as a trade-off between regression performance and computational demand. The regression is performed online on a commercially available computer with Intel core i5 processor and 8 GB RAM. One iteration of the generalized GP model incorporates the calculation of a predictive mean and variance in all 6 output dimensions plus the update of the 6-DoF posterior distribution. On average, a recursion step takes 0.38 sec.

**Application of Generalized Online GP and Kernel Reset** From the captured data, we obtain the input set of dual quaternions  $\{\mathbf{g}_i\}_{i=1}^V$  representing the hand poses throughout the handover and the corresponding output velocity set  $\{\dot{\mathbf{g}}_i\}_{i=1}^V$ . Each 6D velocity consists of three translation velocities  $\dot{\mathbf{g}}_x$ ,  $\dot{\mathbf{g}}_y$ ,  $\dot{\mathbf{g}}_z$  and three rotational velocities represented in the tangent bundle  $\dot{\mathbf{g}}_{TS_x}$ ,  $\dot{\mathbf{g}}_{TS_y}$ ,  $\dot{\mathbf{g}}_{TS_z}$ . We model the dynamics using  $\text{GP}_{\mathbb{H}_D}$  with squared exponential kernel (22) and initialize the hyperparameters with  $H_0 = (1, 1, 1)$ . Per newly observed sample tuple  $(\mathbf{g}_{v+1}, \dot{\mathbf{g}}_{v+1})$  the training data set  $\mathcal{D}_v$  increases until the budget limit 100 is reached and the least relevant pair removal process is additionally activated.

As it is our aim to demonstrate the wide applicability of the proposed approach besides the experimental validation, we

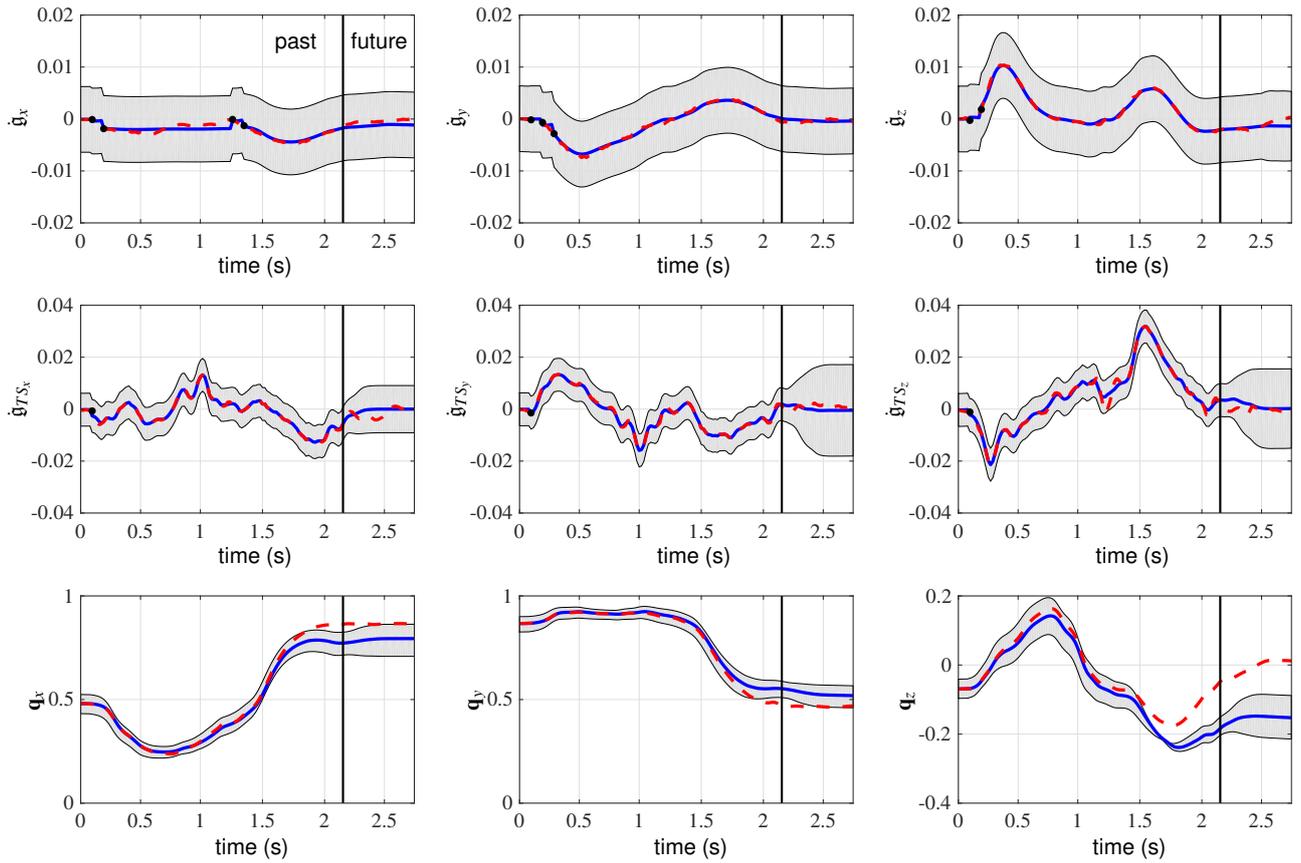


Figure 8: The estimation results of the generalized online GP algorithm are visualized. The vertical bar indicates the current time stamp. In the past, left of the bar, the Gaussian posterior is available, whereas on its right, in the future, the motion is predicted. The first row depicts the translational velocities, the second row the tangent space velocities and the third row shows the imaginary components of a quaternion, which is obtained by projecting the velocities of the row above to the sphere.

do not perform prior investigations to find optimal hyperparameters, despite the fact that this is the common routine in GP modeling. Instead, we train the hyperparameters on the four first observed pairs  $\{(\mathbf{g}_i, \hat{\mathbf{g}}_i)\}_{i=1}^4$  and perform the recursive GP over dual quaternions with this possibly suboptimal hyperparameter set. As these first poses might not be representative for the whole following motion, we accept less accurate prediction results in favor of robustness of the algorithm and a wide application range. In case the mean prediction deviates from the subsequently obtained ground truth value more than a tolerated prediction error (in this experiment we use 0.6 as threshold), we train new hyperparameters on the following four dual quaternions.

**Visualization** The estimation results are illustrated in Figure 8. For clarity the 6D output is shown per dimension over the runtime corresponding to the sample index  $i = 1, \dots, v$  instead of the GP input data  $\mathbf{g}$ . The vertical bar indicates the current time stamp, i.e. left of it is the past, where the Gaussian posterior is available and on its right is the future, where we estimate a predictive Gaussian. The ground truth is depicted by the red dashed line and the mean prediction by the blue solid line. The grey shaded area visualizes the  $2\sigma$  confidence interval for the estimation and captures thereby the noise involved in real data. In the first and second row we depict the three dimensions of the translational velocity and tangent bundle velocity, respectively. In the third row, the rotational velocity estimation is projected to the sphere  $S_3$  using (34). We visualize the three imaginary components of the resulting quaternions only. The velocity predictions, row one and two, demonstrate that in the past the mean prediction deviates little from the ground truth. The distance, how soon the mean prediction returns to zero in the future depends on the length-scale hyperparameter  $\lambda$  learned from four consecutive observed poses only. We indicate the kernel reset with black dots on the ground truth trajectory. In the third row, a slight asymmetry in the uncertainty estimate can be observed. It results from the projection to a curved space. The results show further that over time the prediction drifts from the ground truth as the velocity prediction error accumulates. Recall that we could significantly improve the estimation results through elaborate hyperparameter tuning, but as this contradicts the gist of online modeling, we refrain from any preliminary task specific tuning.

## 6 Conclusion

In this article we introduce a mathematical framework for GP modeling over 6-Dof rigid motions. The essential contribution is the development of a generalized GP in the special Euclidean group  $SE(3)$ , parametrized by dual quaternions. We explain the general concept behind the required GP generalization, develop all essential pre-requisites and finally introduce a mathematically firm solution for modeling rigid motions dynamics on  $SE(3)$  using the generalized GP. Therefore, diverse metrics on unit quaternions are introduced and the validity of resulting kernel functions is investigated. We extend the metrics from pure rigid rotations to full 6-DoF rigid motions and prove the mathematical correctness of the developed kernel functions. Moreover, we newly introduce an online GP over dual quaternions, that reduces the required computational demand from  $\mathcal{O}(h^3)$  to  $\mathcal{O}(h^2)$ . In an experimental evaluation, we analyze on simulated data how the ground truth approximation accuracy of the  $\text{GP}_{\mathbb{H}_D}$  outperforms the traditional  $\text{GP}_{\mathbb{R}^6}$  on modeling nonlinear dynamics. Additionally, we

contrast the characteristics of generalized GP kernels resulting from different metrics in 6D space in a real data experiment using human gaze motion and finally, we demonstrate in another real data experiment the estimation performance of the generalized online GP.

## Acknowledgements

The research leading to these results has received funding from the ERC Starting Grant “Control based on Human Models (con-humo)” under grant agreement n° 337654 and from 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

- Ata Y Erhan andf Yayli (2008) Dual unitary matrices and unit dual quaternions. *Differential Geometry - Dynamical Systems* 10:1–12
- Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15:1373–1396
- Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA
- Calandra R, Peters J, Rasmussen CE, P DM (2014) Manifold gaussian processes for regression. arXiv preprint arXiv:14025876
- Corteville B, Aertbelien E, Bruyninckx H, De Schutter J, Van Brussel H (2007) Human-inspired robot assistant for fast point-to-point movements. In: *IEEE International Conference on Robotics and Automation*, pp 3639–3644
- Del Castillo E, Colosimo BM, Tajbakhsh SD (2015) Geodesic gaussian processes for the parametric reconstruction of a free-form surface. *Technometrics* 57(1):87–99
- Fukumizu K, Gretton A, Schölkopf B, Sriperumbudur BK (2009) Characteristic kernels on groups and semigroups. In: Koller D, Schuurmans D, Bengio Y, Bottou L (eds) *Adv Neural Inf Process Syst.*, Curran Associates, Inc., pp 473–480
- Ham J, Lee DD, Mika S, Schölkopf B (2004) A kernel view of the dimensionality reduction of manifolds. In: *Proceedings of the Twenty-first International Conference on Machine Learning*, ACM, New York, NY, USA, ICML '04, pp 47–, DOI 10.1145/1015330.1015417
- Harandi MT, Salzmann M, Porikli F (2014) Bregman divergences for infinite dimensional covariance matrices. *CoRR* abs/1403.4334
- Jarrasse N, Paik J, Pasqui V, Morel G (2008) How can human motion prediction increase transparency? In: *IEEE International Conference on Robotics and Automation*, pp 2134–2139

- Jayasumana S, Hartley R, Salzmann M, Li H, Harandi M (2013) Kernel methods on the riemannian manifold of symmetric positive definite matrices. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp 73–80, DOI 10.1109/CVPR.2013.17
- Kang H, Park FC (2015) Motion optimization using gaussian process dynamical models. *Multibody System Dynamics* 34(4):307–325, DOI 10.1007/s11044-014-9441-8
- Khansari-Zadeh S, Billard A (2011) Learning stable nonlinear dynamical systems with gaussian mixture models. *IEEE Transactions on Robotics* 27(5):943–957
- Kim S, Billard A (2012) Estimating the non-linear dynamics of free-flying objects. *Rob Auton Sys* 60(9):1108–1122
- Ko J, Fox D (2008) Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. In: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pp 3471–3476, DOI 10.1109/IROS.2008.4651188
- Kulić D, Takano W, Nakamura Y (2007) Incremental learning of full body motions via adaptive factorial hidden markov models. In: 7th International Conference on Epigenetic Robotics
- Lang M (2011) Approximation of Probability Density Functions on the Euclidean Group Parametrized by Dual Quaternions. arxiv preprint, arxiv:1707.00532, Ludwig-Maximilians-Universität München, URL <https://arxiv.org/abs/1707.00532>
- Lang M, Feiten W (2012) Mpg - fast forward reasoning on 6 dof pose uncertainty. In: Robotics; Proceedings of ROBOTIK 2012; 7th German Conference on, pp 1–6
- Lang M, Hirche S (2017) Computationally Efficient Rigid-Body Gaussian Process for Motion Dynamics. *IEEE Robotics and Automation Letters* 2(3):1601–1608, DOI 10.1109/LRA.2017.2677469
- Lang M, Dunkley O, Hirche S (2014) Gaussian process kernels for rotations and 6d rigid body motions. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on, pp 5165–5170, DOI 10.1109/ICRA.2014.6907617
- Lang M, Kleinstueber M, Dunkley O, Hirche S (2015) Gaussian process dynamical models over dual quaternions. In: European Control Conference (ECC), Linz, Austria
- Matsuoka Y, Durrant-Whyte H, Neira J (2011) Robotics: Science and Systems VI. The MIT Press
- Medina JR, Lawitzky M, Mörtl A, Lee D, Hirche S (2011) An experience-driven robotic assistant acquiring human knowledge to improve haptic cooperation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp 2416–2422
- Miossec S, Kheddar A (2008) Human motion in cooperative tasks: Moving object case study. In: IEEE International Conference on Robotics and Biomimetics, ROBIO, pp 1509–1514, DOI 10.1109/ROBIO.2009.4913224

- Nilsson J, Sha F, Jordan MI (2007) Regression on manifolds using kernel dimension reduction. In: Proceedings of the 24th International Conference on Machine Learning, ACM, New York, NY, USA, ICML '07, pp 697–704, DOI 10.1145/1273496.1273584
- Rasmussen CE, Nickisch H (2010) Gaussian processes for machine learning (gpml) toolbox. *J Mach Learn Res* 11:3011–3015
- Rasmussen CE, Williams CK (2006) *Gaussian Processes for machine learning*. Adaptive computation and machine learning series, University Press Group Limited
- Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326, DOI 10.1126/science.290.5500.2323
- Sebanz N, Knoblich G (2009) Prediction in joint action: What, when, and where. *Topics in Cognitive Science* 1(2):353–367, DOI 10.1111/j.1756-8765.2009.01024.x
- Subbarao R (2008) Robust statistics over riemannian manifolds for computer vision. PhD thesis, Rutgers University; Graduate School - New Brunswick, DOI 10.7282/T3736R88
- Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500):2319
- Thomas F (2014) Approaching dual quaternions from matrix algebra. *IEEE Transactions on Robotics* 30(5):1037–1048, DOI 10.1109/TRO.2014.2341312
- Urtasun R, Fleet D, Geiger A, Popovic J, Darrell T, Lawrence N (2008) *Topologically-Constrained Latent Variable Models*. In: International Conference on Machine learning (ICML)
- Van Vaerenbergh S, Lazaro-Gredilla M, Santamaria I (2012) Kernel recursive least-squares tracker for time-varying regression. *IEEE Transactions on Neural Networks and Learning Systems* 23(8):1313–1326, DOI 10.1109/tnnls.2012.2200500
- Wang JM, Fleet DJ, Hertzmann A (2008) Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2):283–298
- Zhang D, Chen X, Lee WS (2005) Text classification with kernels on the multinomial manifold. In: SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM Press, New York, NY, USA, pp 266–273, DOI 10.1145/1076034.1076081