Gaussian Process Dynamical Models over Dual Quaternions

Muriel Lang, Martin Kleinsteuber, Oliver Dunkley and Sandra Hirche

Abstract— **This paper presents a method for learning nonlinear rigid body dynamics in the special Euclidean group $SE(3)$. The method is based on the Gaussian process dynamical model (GPDM), which combines two Gaussian processes (GPs), one for representing unknown dynamics in a space $\mathbb{R}^d$ with reduced dimensionality and the other for transforming the reduced space back to the state space of the high dimensional measurements $\mathbb{R}^D$. We introduce in this paper an enhanced GPDM, which extends the dynamics modeling space from Euclidean space to the special Euclidean group $SE(3)$. This allows for accurate modeling of unknown dynamics incorporating rotation and translation. Therefore, the unknown dynamics are described by a GP over dual quaternions, denoted by $\mathrm{GP}_{\mathbb{H}_D}$, which extends the state of the art GP to a non-Euclidean input space $SE(3)$. Further, we provide a proof that the squared exponential kernel used in the $\mathrm{GP}_{\mathbb{H}_D}$ defines a valid covariance function. In conclusion we illustrate how the GPDM over dual quaternions outperforms the traditional GPDM depending on the amount of training data and rotation magnitude.**

I. INTRODUCTION

In this work, we consider nonlinear dynamics on the special Euclidean group $SE(3)$, the group of relative distance and orientation preserving transformations in 3D space. An application domain is human robot interaction, where human motion modeling is required. A main challenge for such a task lies in finding a model that can capture the nonlinearities of the dynamics. Well known models, such as switching linear dynamical systems (SLDS) [1] or nonlinear dynamical systems (NLDS) [2] can account only to a certain extend for nonlinearities in the dynamics and require a large training data set for accurate maximum a posteriori estimations. Classical system identification methods assume that the structure of the differential equations is known [3], which in general can't be provided in applications of human motion modeling tasks. Therefore, we propose a Bayesian approach for modeling an unknown nonlinear dynamic, namely a Gaussian process (GP) based model. Its main advantage, compared to the aforementioned models or neural networks, lies in accounting for the uncertainty in the model. GP based models nominate a best estimator for the predictions and provide the uncertainty of each prediction.

However, a drawback of all GP based models is that they are only defined in Euclidean space. The GP model is based on the inverse of a Gram matrix [4], which only exists if a Euclidean vector space structure is available. In an application where orientations are required, Euler angles are typically used as a rotation representation, see [5], as this representation is approximately Euclidean for rotations

All authors are with the Department of Electrical and Computer Engineering, Technische Universität München, D-80290 Munich
`muriel.lang at tum.de`

close to zero. This approach is appropriate, if one assumes sufficiently small rotation angles. However, when the rate of orientation change becomes significantly large due to high angular speed, low sampling frequency or temporary loss of input data, another rotation parametrization is required. We propose to use unit quaternions S_3 to parametrize rotations and dual quaternions $\mathbb{H}_D$ to parametrize transformations consisting of rotations and translations, i.e. rigid motions. The dual quaternion representation is advantageous, as it represents rigid motions in a unified manner, provides smooth motions [6] and is close to a minimal representation. In our previous work [7] we introduced a GP in the special Euclidean group $SE(3)$ parametrized by dual quaternions.

The contribution of this paper is the enhancement of the Gaussian process dynamical model (GPDM) to a GPDM over dual quaternions by introducing a GP over dual quaternions for modeling rotational and translational dynamics in the reduced dimensional space of hidden variables. We introduce a modified distance measure defined on the special Euclidean group $SE(3)$. It is used in the covariance function of the introduced GP. We prove that this covariance function over dual quaternions is valid. The differences between the GP over dual quaternions and the traditional GP incorporating Euler angles are empirically evaluated inside the GPDM.

The paper is organized as follows. In Section II the problem is formulated. Section III briefly explains the GP on the special Euclidean group $SE(3)$. Section IV introduces the enhanced GPDM over dual quaternions. Finally, in Section V the estimation quality of the enhanced GPDM for nonlinear dynamics is examined.

Notation. In the remainder of the paper we denote the set of vectors that are perpendicular to vector $\mathbf{q}$ by $\mathbf{q}^{\perp}$; $\sim$ stands for distributed as; $\simeq$ denotes an isomorphism; $\approx$ means approximately equal and $\rtimes$ is the semi-direct product.

II. PROBLEM FORMULATION

The objective of this paper is to develop a data-driven modeling approach for nonlinear dynamics in the special Euclidean group $SE(3)$. As human robot interaction is a main application area, we require the model to be suitable for human behavior modeling incorporating significant rotations. Our approach to this problem is to enhance the GPDM to a GPDM over dual quaternions, which can handle these rotations.

We consider a time series of states $\mathbf{x}_t$ in the special Euclidean group $SE(3)$ with time index $t \geq 0$. Each state $\mathbf{x}_t$ consists of a relative distance and orientation preserving transformation parametrized by a dual quaternion. We assume a system on $SE(3)$ to be evolving according to some

general nonlinear but unknown dynamics $f : SE(3) \to SE(3)$,

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t) + \mathbf{n}_{\mathbf{x}_t}, \qquad (1)$$

where $\mathbf{n}_{\mathbf{x}_t}, t \geq 0$ is a noise process in $SE(3)$. Our objective is to find an approximation of the dynamics based on observed data, employing a Gaussian process (GP) based modeling approach. It should be noted, however, that a Euclidean vector space structure is required for defining a Gaussian distribution, the output of a GP model. As on the considered special Euclidean group $SE(3)$, this structure isn't provided, the dynamics (1) is not directly amenable to a GP based modeling approach. In consequence, we propose to reformulate the problem in the following way: Let the mapping $\tilde{f}$ be a vector field on $SE(3)$ with dynamic

$$\mathbf{v}_t = \tilde{f}(\mathbf{x}_t) + \tilde{\mathbf{n}}_{\mathbf{x}_t}, \qquad (2)$$

where $\mathbf{v}_t \in T_{\mathbf{x}_t} SE(3)$ is a velocity vector in the tangent space of $SE(3)$ at state $\mathbf{x}_t$. We assume the noise to be some white Gaussian noise $\tilde{\mathbf{n}}_{\mathbf{x}} \sim \mathcal{N}(0, \Sigma_{\mathbf{x}})$ in the tangent bundle $TSE(3)$. The function $\tilde{f}$ can now be modeled by a GP over dual quaternions, abbreviated in the following by $\mathrm{GP}_{\mathbb{H}_D}$. In addition, we define $h : TSE(3) \to SE(3)$ to map from the tangent bundle $TSE(3)$ back to the initial group $SE(3)$ such that $f = h \circ \tilde{f}$ holds. The mapping h is determined analytically. The pair $(\mathbf{v}_t, \mathbf{x}_t) \in TSE(3)$ defines a vector $\mathbf{v}_t \in T_{\mathbf{x}_t} SE(3)$ inducing $h_{\mathbf{x}_t} : T_{\mathbf{x}_t} SE(3) \to SE(3)$ such that,

$$\mathbf{x}_{t+1} = h_{\mathbf{x}_t}(\mathbf{v}_t). \qquad (3)$$

A mapping from the special Euclidean group $SE(3)$ to the observation space $\mathbb{R}^D$ is required additionally, to complete the Gaussian process dynamical model (GPDM). It is given by a function $g : SE(3) \to \mathbb{R}^D$ with

$$\mathbf{y}_t = g(\mathbf{x}_t) + \mathbf{n}_{\mathbf{y}_t}, \qquad (4)$$

where $\mathbf{y}_t \in \mathbb{R}^D, D \in \mathbb{N}$ a high dimensional vector-valued observation. We denote by $\mathbf{n}_{\mathbf{y}_t} \in \mathbb{R}^D, t \geq 0$ zero-mean Gaussian noise. Together, we obtain a GPDM over dual quaternions, which is suitable to model nonlinear rigid body dynamics in the special Euclidean group $SE(3)$.

The observation space $\mathbb{R}^D$ of the human motion is parametrized by translation vectors and rotation angles to match state of the art sensing devices, that mostly provide 3D rotation angles and 3D positions. Note, Euler angles close to zero are approximately Euclidean and with high sampling rates, the incremental orientation changes are sufficiently small. However, in the latent space with reduced dimensionality we require the dual quaternion parametrization. Therefore, the GPDM over dual quaternions is required for nonlinear dynamics in the special Euclidean group $SE(3)$.

III. GP OVER DUAL QUATERNIONS

This section illustrates the problem of including orientation into a GP. First, relevant background information about GP modeling is provided. Then, the $\mathrm{GP}_{\mathbb{H}}$ for pure rotations is recapitulated before the $\mathrm{GP}_{\mathbb{H}_D}$ defined in [7] is briefly explained. We provide a proof for the squared exponential covariance functions in $\mathrm{GP}_{\mathbb{H}}$ and $\mathrm{GP}_{\mathbb{H}_D}$ to be valid.

A. Background on GP Models

A GP is defined as a set of random variables, such that every finite subset is jointly Gaussian distributed [8]. The GP is fully specified by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$. Hence, a GP is given by

$$f(x) = \mathrm{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \qquad (5)$$

This covariance function $k(\mathbf{x}, \mathbf{x}')$ is required to generate a symmetric and positive semi-definite Gram matrix $\mathbf{K}$ [9], which is specified by $(\mathbf{K})_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ for $\mathbf{x}_i$ and $\mathbf{x}_j$ elements of any finite set of realizations $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The Gaussian Process dynamical model (GPDM) [10] is an unsupervised learning method for high dimensional data. It models the joint distribution of the observed data in the high dimensional *observation space* and the corresponding latent variables in a low dimensional *latent space*. Further, the GPDM comprises a dynamical model in the latent space. Principal component analysis (PCA) is interpreted as GP prior on the mapping from latent to observation space.

B. GP on the Hypersphere S_3

For better understanding, the GP on S_3 parametrized by unit quaternions is briefly explained in this section. Following our previous work [7], we model the mapping $\tilde{f}$ by a GP on unit quaternions. Hence, the state $\mathbf{x}_t$ in eq. (2) is a rotation quaternion $\mathbf{q}_{\mathrm{rot}_t} = q_w + q_x i_x + q_y i_y + q_z i_z \in \mathbb{H}$, with parameter entries $q_w, q_x, q_y, q_z \in \mathbb{R}$ such that $\|\mathbf{q}_{\mathrm{rot}_t}\| = 1$. The output velocity $\mathbf{v}_t$ is denoted by $\mathbf{v}_{TS_{\mathbf{q}_{\mathrm{rot}_t}}}$ for $TS_{\mathbf{q}_{\mathrm{rot}_t}}$ the tangent space of the corresponding orientation quaternion $\mathbf{q}_{\mathrm{rot}_t}$. In GP modeling a Gaussian distribution is provided for any output $\mathbf{v}_t$. In the following we explain how to transfer this Gaussian to a distribution over states as in (3). A visualization of the projected Gaussian is provided in Fig. 1. **Central Projection:** We model the prediction uncertainty on the unit sphere S_3 by projecting the Gaussian distribution from the tangent space to the sphere.
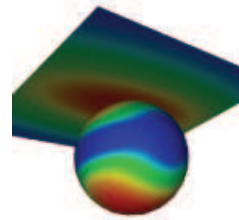

Fig. 1. The Gaussian distribution in the tangent space at the north pole is projected to the sphere S_2 via central projection.

To propagate the 3D prediction uncertainty of velocity $\mathbf{v}_{TS_t}$ to the state space S_3, function h is defined to be the central projection

$$\Pi_{\mathbf{q}_{\mathrm{rot}_t}} : TS_{\mathbf{q}_{\mathrm{rot}_t}} \to S_3$$
$$\Pi_{\mathbf{q}_{\mathrm{rot}_t}}(\mathbf{v}_{TS_t}) = \{-\tfrac{\mathbf{v}}{\|\mathbf{v}\|}, \tfrac{\mathbf{v}}{\|\mathbf{v}\|}\}, \qquad (6)$$

where $\mathbf{v} = \mathbf{q}_{\mathrm{rot}_t} + \mathbf{B}\mathbf{v}_{TS_t}$. The basis $\mathbf{B}$ is the canonical representation of the 3D tangent space $TS_{\mathbf{q}_{\mathrm{rot}_t}}$ in the 4D space $\mathbb{R}^4$ [11]. As claimed in eq. (3), we obtain a unit quaternion $\mathbf{q}_{\mathrm{rot}_{t+1}} = \Pi_{\mathbf{q}_{\mathrm{rot}_t}}(\mathbf{v}_{TS_t})$ from the velocity $\mathbf{v}_{TS_t}$ such that $f = h \circ \tilde{f}$ holds. A probability density on the sphere

is obtained by projecting the set closure $\{\infty\}$ of the tangent space $TS_{\mathbf{q}_{\mathrm{rot}_t}}$ to the set of orthogonal vector-valued points $\mathbf{q}_{\mathrm{rot}_t}^{\perp}$ on the sphere,

$$\Pi_{\mathbf{q}_{\mathrm{rot}_t}}(\infty) = \mathbf{q}_{\mathrm{rot}_t}^{\perp} := \{\mathbf{q} \in S_3 | \mathbf{q} \perp \mathbf{q}_{\mathrm{rot}_t}\}. \tag{7}$$

As the set closure $\{\infty\}$ has probability 0, we obtain a continuous completion of the probability distribution projected to the sphere.

Remark 3.1: The unit quaternions are a double coverage of the 3D orientations, i.e. opposing points $\pm\mathbf{q}$ on the sphere S_3 represent the same orientation. Thus, the special Orthogonal group $SO(3)$ is isomorphic to the group $S_3/\{\pm 1\}$ with quaternion multiplication as a group operation. Informally speaking, the unit sphere S_3 is reduced about the equivalence class $\{\pm 1\}$ of opposing unit quaternions.

Squared Exponential Covariance Function: It is common to set the mean function to zero in GP applications, as the essential part in GP model learning is modeling the covariance function $k(\mathbf{x}, \mathbf{x}')$. The squared exponential covariance function

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{d^2(\mathbf{x}, \mathbf{x}')}{2\lambda^2}\right), \tag{8}$$

where $d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|$ the Euclidean $\mathcal{L}^2$ distance and the hyperparameters $\sigma_f, \lambda > 0$ signal variance and length scale, is widely used in state of the art GP modeling [4]. In the following we provide a squared exponential kernel function for unit quaternions, based on the length of the arc section between two quaternions $\mathbf{q}$ and $\mathbf{q}'$

$$d_{\mathrm{arc}}(\mathbf{q}, \mathbf{q}') = \min \arccos(\langle \mathbf{q}, \pm\mathbf{q}' \rangle), \tag{9}$$

where $\langle \mathbf{q}, \mathbf{q}' \rangle = \frac{1}{2}(\mathbf{q}\overline{\mathbf{q}'} + \mathbf{q}'\overline{\mathbf{q}})$ is the scalar product over quaternions and $\overline{\mathbf{q}}$ denotes the quaternion conjugate. The arc length defines a metric on the hypersphere S_3 [12]. To be able to proof Theorem 3.3, we need a conditionally positive definite covariance function [13], which is defined as follows.

Definition 3.2: Let $\mathcal{X}$ be a nonempty set. A real-valued symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a conditionally positive definite (cpd) covariance function, if the Gram matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ satisfies $\mathbf{c}^\top \mathbf{K} \mathbf{c} \geq 0$ for any vector $\mathbf{c} \in \mathbb{R}^n$ with $0 = \mathbf{c}^\top \mathbf{1} = \sum_{i=1}^n c_i$.

Theorem 3.3: The function $k : S_3 \times S_3 \to \mathbb{R}_0^+$ over unit quaternions $\mathbf{q}, \mathbf{q}' \in S_3$, given by

$$k_{\mathrm{arc}}(\mathbf{q}, \mathbf{q}') = \sigma_f^2 \exp\left(-\frac{d_{\mathrm{arc}}(\mathbf{q}, \mathbf{q}')^2}{2\lambda^2}\right), \tag{10}$$

where d_{arc} as in (9) and the hyperparameters $\sigma_f, \lambda > 0$, defines a covariance function, i.e. is symmetric and positive definite.

Proof of Theorem 3.3: From [13] we know that $\exp(ck)$ is a positive definite covariance function for all constants $c > 0$ iff k is cpd. Further, [14] proves that any finite constant $\tilde{c} > 0$ is a positive definite covariance function and the product of valid kernels defines a positive definite covariance function. By choosing $\tilde{c} = \sigma_f^2$ and $c = \frac{1}{2\lambda^2}$ in kernel $k_{\mathrm{arc}}(\mathbf{q}, \mathbf{q}')$ defined as in (10), we have the proof completed, if we can show that the function $k_{\mathrm{negd}}^2(\mathbf{q}, \mathbf{q}') := -(d_{\mathrm{arc}}(\mathbf{q}, \mathbf{q}')^2)$ defines a cpd covariance function. Therefore, let $\phi : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$,

given by $\phi(x) = \pi/2 - \arccos(x)$ and $\psi : [-1, 1] \to [\frac{3}{2}\pi, \frac{\pi}{2}]$, given by $\psi(x) = \pi/2 + \arccos(x)$ be two auxiliary functions. The procedure is to prove that both auxiliary functions define valid covariance functions and then, to show that $k_{\mathrm{negd}}^2(\mathbf{q}, \mathbf{q}')$ can be constructed from these auxiliary functions.

The arccos function can be expressed in terms of the arcsin

$$\arccos(x) = \frac{\pi}{2} - \arcsin(x). \tag{11}$$

Further, it is well known, that arcsin can be represented as infinite series via Taylor expansion

$$\arcsin(x) = \sum_{j=0}^{\infty} c_j x^{2j+1}, \text{ where } c_j = \frac{2\Gamma\left(j + \frac{1}{2}\right)}{\sqrt{\pi}(2j+1)j!} \tag{12}$$

are constants depending on the index j and $\Gamma(\tilde{x})$ the gamma function. The constants $c_j > 0 \ \forall j = 0, 1, 2, \dots$ because $\Gamma(\tilde{x}) > 0 \ \forall \tilde{x} > 0$. Together, we obtain that the function $\phi(x)$ reduces to $\arcsin(x)$, an infinite polynomial with non-negative coefficients and the function $\psi(x)$ to

$$\psi(x) = \frac{\pi}{2} + \frac{\pi}{2} - \arcsin(x) = \pi + \arcsin(-x). \tag{13}$$

On introducing $\langle \mathbf{q}, \mathbf{q}' \rangle$ for $x \in [-1, 1]$ into the auxiliary functions ϕ and ψ we obtain $\phi(\langle \mathbf{q}, \mathbf{q}' \rangle) = \arcsin(\langle \mathbf{q}, \mathbf{q}' \rangle)$ and $\psi(\langle \mathbf{q}, \mathbf{q}' \rangle) = \pi + \arcsin(\langle \mathbf{q}, -\mathbf{q}' \rangle)$, because $-\langle \mathbf{q}, \mathbf{q}' \rangle = \langle \mathbf{q}, -\mathbf{q}' \rangle$. According to Theorem 4 in [13] we know that the Taylor series expansion of a function defined on the unit sphere, such as (12) defines a valid covariance function, iff all coefficients are non-negative. Further, we know from [14] that the sum of valid covariance functions defines a positive definite covariance function. Hence, we obtain that both auxiliary functions define valid covariance functions.

Now, we define a new auxiliary function, depending on the former two, $\theta : [-1, 1] \to [-\frac{3}{4}\pi^2, \frac{1}{2}\pi^2]$, $\theta(x) = \phi(x)\psi(x)$. Introducing $\langle \mathbf{q}, \mathbf{q}' \rangle$ provides

$$\theta(\langle \mathbf{q}, \mathbf{q}' \rangle) = \frac{\pi^2}{4} - \arccos^2(\langle \mathbf{q}, \mathbf{q}' \rangle), \tag{14}$$

what defines a valid covariance function according to [14].

The squared negative distance kernel $k_{\mathrm{negd}}^2 = -d_{\mathrm{arc}}^2$ can be rewritten as

$$k_{\mathrm{negd}}^2 = \underbrace{-\frac{\pi^2}{4}}_{\star} + \underbrace{\underbrace{\theta(\langle \mathbf{q}, \tilde{\mathbf{q}}' \rangle)}_{\triangle}}_{\diamond}, \tag{15}$$

where $\tilde{\mathbf{q}}' \in \{\pm\mathbf{q}'\}$ chosen such that $\arccos(\langle \mathbf{q}, \tilde{\mathbf{q}}' \rangle)$ minimal. From [13] we know that

- $\star$ any constant $c \in \mathbb{R}$ is cpd
- $\triangle$ any valid covariance function is also cpd
- $\diamond$ the finite sum of cpd covariance functions is a cpd covariance function

and hence, obtain that $k_{\mathrm{negd}}^2(\mathbf{q}, \mathbf{q}')$ is a cpd covariance function as defined in Definition 3.2.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

C. GP on the Special Euclidean Group $SE(3)$

In this section, we introduce the $\mathrm{GP}_{\mathbb{H}_D}$ over the special Euclidean group $SE(3)$, parametrized by dual quaternions $\mathbb{H}_D$

$$\mathbb{H}_D = \{\mathbf{dq} \mid \mathbf{dq} = \mathbf{q}_{\mathrm{re}} + \varepsilon\mathbf{q}_{\mathrm{du}} \ \& \ \mathbf{q}_{\mathrm{re}}, \mathbf{q}_{\mathrm{du}} \in \mathbb{H}\}, \tag{16}$$

where $\varepsilon \neq 0$ is a dual unit, i.e. it satisfies $\varepsilon^2 = 0$ [15]. A state configuration with 6 degrees of freedom (DoF) can be represented by a dual quaternion

$$\mathbf{dq} := \mathbf{q}_{\text{rot}} + \frac{\varepsilon}{2} \mathbf{q}_{\text{tra}} \mathbf{q}_{\text{rot}}, \qquad (17)$$

where a unit quaternion $\mathbf{q}_{\text{rot}}$ describes the orientation and an imaginary quaternion $\mathbf{q}_{\text{tra}} = q'_x i_x + q'_y i_y + q'_z i_z$, with parameters $q'_x, q'_y, q'_z \in \mathbb{R}$ the position.

Remark 3.4: The representations

$$SE(3) \simeq S_3/\{\pm 1\} \rtimes \mathbb{R}^3 \qquad (18)$$

are isomorphic with dual quaternion multiplication as group operation on $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$. The special Euclidean group is homeomorpic to $SO(3) \times \mathbb{R}^3$, where $\times$ denotes the direct product. As the space of rotations with quaternion representation is $S_3/\{\pm 1\}$, we require the semi-direct product $\rtimes$ to link rotation and translation space.

The multiplication of dual quaternions $\mathbf{dq}$ and $\mathbf{dq}'$ is defined as

$$\mathbf{dq} * \mathbf{dq}' = \mathbf{q}_{\text{re}} \mathbf{q}'_{\text{re}} + \varepsilon \left(\mathbf{q}_{\text{re}} \mathbf{q}'_{\text{du}} + \mathbf{q}_{\text{du}} \mathbf{q}'_{\text{re}} \right). \qquad (19)$$

The $\text{GP}_{\mathbb{H}_D}$ over dual quaternions models the vector field $\tilde{f}$ on $SE(3)$ defined in (2). The input is parametrized by dual quaternions, the output is a vector consisting of rotational velocities $\mathbf{v}_{TS_t}$ in the tangent space concatenated with translational velocities $\mathbf{v}_{\mathbf{p}_t}$. To obtain a dual quaternion as in eq. (3), we define the mapping h to be the central projection for the rotation and the identity function for the translation,

$$h(\mathbf{v}_t) = \begin{cases} \Pi_{\mathbf{q}_{\text{rot}_t}}(\mathbf{v}_{TS_t}) & \text{for rotational velocites} \\ \text{id}\,(\mathbf{v}_{\mathbf{p}_t}) & \text{for translational velocities.} \end{cases} \qquad (20)$$

To define a covariance function for $\text{GP}_{\mathbb{H}_D}$ we require a magnitude measure between input states, which is calculated from the transforming dual quaternion: To arrive in $\mathbf{dq}'$ we apply to the configuration $\mathbf{dq}$ the transformation $\vec{\mathbf{dq}} = \overline{\mathbf{dq}} * \mathbf{dq}'$, where $\overline{\mathbf{dq}}$ denotes the dual quaternion conjugate. From the transforming dual quaternion $\vec{\mathbf{dq}} = \vec{\mathbf{q}}_{\text{re}} + \varepsilon \vec{\mathbf{q}}_{\text{du}}$, rotation $\mathbf{q}_{\text{rot}}$ and translation $\mathbf{q}_{\text{tra}}$ are obtained by dual quaternion decomposition using (17). We define the transformation magnitude measure as

$$d_{\text{mag}}(\mathbf{dq}, \mathbf{dq}') = \sqrt{d_{\text{arc}}(\mathbf{q}_0, \vec{\mathbf{q}}_{\text{rot}})^2 + \|\vec{\mathbf{q}}_{\text{tra}}\|^2}, \qquad (21)$$

where $\mathbf{q}_0 = (1,0,0,0)^\top$ denotes the zero rotation. Similar to the case, where the GP is restricted to the sphere S_3, a valid covariance function is defined based on this magnitude measure in the following theorem.

Theorem 3.5: The function $k : (S_3 \times \mathbb{R}^3) \times (S_3 \times \mathbb{R}^3) \to \mathbb{R}_0^+$ over dual quaternions $\mathbf{dq}, \mathbf{dq}' \in S_3 \times \mathbb{R}^3$, given by

$$k_{\text{mag}}(\mathbf{dq}, \mathbf{dq}') = \sigma_f^2 \, \exp\left(-\frac{d_{\text{mag}}(\mathbf{dq}, \mathbf{dq}')^2}{2\lambda^2} \right). \qquad (22)$$

where d_{mag} as in (21) and the hyperparameters $\sigma_f, \lambda > 0$, defines a valid covariance function.

Proof: The function k_{mag} can be rewritten as

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

According to Theorem 3.3 $\star$ is a covariance function. The proof for $\diamond$ to be a covariance function is identical the one for the standard squared exponential kernel provided in [14] Section 6.2 for eq. (6.23). Hence, it follows that $k_{\text{mag}}(\mathbf{dq}, \mathbf{dq}')$ defines a positive definite covariance function.

$\square$

IV. GPDM OVER DUAL QUATERNIONS

In this section we introduce a GPDM, in which the low dimensional space of latent dynamics is $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$ or a subset of it. This latent space is parametrized by dual quaternions $\mathbb{H}_D$. The $\text{GP}_{\mathbb{H}_D}$ over dual quaternions is used for learning the dynamical model in the latent space, as well as for the mapping from the latent space $\mathbb{H}_D$ to a high dimensional observation space. In applications of full body human motion modeling the dimension of the observation space generally exceeds the number of dimensions where calculations are reasonably fast such that dimension reduction is required to obtain a GP prior for the dynamics in the latent space $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$.

A. GPDM Dynamics

We consider a mapping $g : \mathbb{H}_D \times \mathbb{R}^{r \times s} \to \mathbb{R}^D$, with $r, s \in \mathbb{N}$ from the latent to observation space

$$\mathbf{y}_{t+1} = g(\mathbf{dq}_{t+1}, \mathbf{B}) + \mathbf{n}_{\mathbf{y}_t}, \qquad (24)$$

where $\mathbf{y}_{t+1} \in \mathbb{R}^D$ the observations at the subsequent time step and $\mathbf{n}_{\mathbf{y}_t}$, zero-mean white Gaussian noise The dual quaternion $\mathbf{dq}_{t+1}$, is obtained using function h as defined in eq. (20),

$$\mathbf{dq}_{t+1} = h(\mathbf{v}_{\mathbf{dq}_t}). \qquad (25)$$

By slight abuse of denotation, we continue to refer to the dynamics in the latent space with $\tilde{f}$ even though we restrict the input to a time series of dual quaternions $\mathbf{dq}_t, t \geq 0$ and include unknown function parameters $\mathbf{A} \in \mathbb{R}^{r' \times s'}$ (respectively $\mathbf{B} \in \mathbb{R}^{r \times s}$ for mapping g) written in matrix form into the parametrization.

$$\mathbf{v}_{\mathbf{dq}_t} = \tilde{f}(\mathbf{dq}_t, \mathbf{A}) + \tilde{\mathbf{n}}_{\mathbf{dq}_t} \qquad (26)$$

where $\mathbf{v}_{\mathbf{dq}_t}$ the latent velocities and $\tilde{\mathbf{n}}_{\mathbf{dq}_t}$ zero-mean white Gaussian noise. Following [16], the mappings $\tilde{f}$ and g are considered to consist of linear combinations of possibly nonlinear basis functions. Using this dynamics representation, we are able to "marginalize" over the parameter matrices $\mathbf{A}$ and $\mathbf{B}$ specifying the unknown weights of the basis functions. Marginalizing means to reformulate the GP densities conditioned to the parameters. This is required to insert the GP describing the latent dynamics into the GP modeling the latent mapping.

Marginalizing the parameter $\mathbf{B}$ over g yields a Gaussian density for the mapping from latent to observation space [10], which can be expressed as a product of l GPs

$$p(\mathbf{Y}|\mathbf{DQ}, \beta) = \frac{1}{\sqrt{(2\pi)^{nl}|\mathbf{K}_{\mathbf{Y}}|^l}} \exp\left(-\frac{1}{2} \sum_i \sum_j (\mathbf{K}_{\mathbf{Y}}^{-1})_{ij} \mathbf{y}_j \mathbf{y}_i^\top \right), \qquad (27)$$

where $\mathbf{DQ} = \{\mathbf{dq}_1, \ldots, \mathbf{dq}_n\}$ the n latent states, corresponding to the n D-dimensional observations $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, further $\mathbf{K_Y}$ a covariance matrix determined by $(\mathbf{K_Y})_{ij} = k_{\mathrm{mag}}(\mathbf{dq}_i, \mathbf{dq}_j)$ for $i, j \in \{1, \ldots, n\}$ and k_{mag} the covariance function defined as in (22). The parameter $\beta = (\sigma_f, \lambda)$ combines the hyperparameters of the covariance function k_{mag}.

The density of the joint dynamics mapping in the latent states $\mathbf{DQ}$ is obtained by marginalizing over the unknown parameter matrix $\mathbf{A}$, while assuming an isotropic Gaussian prior on the columns of $\mathbf{A}$. Denoting the set of dual quaternion velocities by $\mathbf{V_{DQ}}$ we calculate the probability density

$$p(\mathbf{V_{DQ}}|\mathbf{DQ}, \alpha) = \frac{p(\mathbf{dq}_1)}{\sqrt{(2\pi)^{(n-1)m}|\mathbf{K_{DQ}}|^m}}$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{i=2}^n \sum_{j=2}^n (\mathbf{K_{DQ}^{-1}})_{ij} |\langle \mathbf{dq}_j, \mathbf{dq}_i\rangle|\right), \quad (28)$$

where we assume $\mathbf{dq}_1$ to have an isotropic projected Gaussian prior. This prior is obtained by projecting separately defined Gaussian priors for rotational and translational part to $SE(3)$ via h, see (20). The rotational Gaussian prior is required to be defined in the tangent space of the rotational part $\mathbf{q}_{\mathrm{re}_1}$. Further, $\mathbf{K_{DQ}}$ denotes the covariance matrix over the latent states $\mathbf{dq}_1, \ldots, \mathbf{dq}_{n-1}$ with matrix entries $(\mathbf{K_{DQ}})_{ij} = k_{\mathrm{mag}}(\mathbf{dq}_j, \mathbf{dq}_j)$, where the covariance function defined as in (22) and α is a vector containing the hyperparameters of covariance function k_{mag}. Note, the matrices $\mathbf{K_Y}$ and $\mathbf{K_{DQ}}$ differ solely in the values of the hyperparameters α and β inside the covariance function k_{mag}. The exponent $m \in \{1, \ldots, 6\}$ is the maximum number of nonzero imaginary entries of any dual quaternion in the latent space. The absolute value of $\langle \mathbf{dq}_j, \mathbf{dq}_i\rangle$ is obtained on exploiting the linearity of the bilinear form in both arguments and the dual unit properties $|\varepsilon| = 1$ and $\varepsilon^2 = 0$.

Remark 4.1: The density function (28) of the dynamics in $S_3 \times \mathbb{R}^3$ is not Gaussian, because the latent state $\mathbf{dq}_1$ appears outside the exponential covariance function.

Following [17], on the hyperparameters uninformative priors $p(\alpha) \propto \prod \{\sigma_{f,\alpha}, \lambda_\alpha, \sigma_{l,\alpha}\}$ and $p(\beta) \propto \prod \{\sigma_{f,\beta}, \lambda_\beta\}$ are placed to avoid overfitting. Subsequently, we obtain the priors, the latent mapping and the dynamics over dual quaternions, which together define a generative mapping for high dimensional time series observations

$$p(\mathbf{DQ}, \mathbf{Y}, \alpha, \beta) = p(\mathbf{Y}|\mathbf{DQ}, \beta)\, p(\mathbf{V_{DQ}}|\alpha)\, p(\alpha)\, p(\beta). \quad (29)$$

This probability distribution comprises the composed information of both GPs, one for modeling the latent dynamics in the state space with reduced dimensions $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$ and one for describing the mapping from the latent space to the high dimensional observation space.

B. GPDM Learning

In this paragraph concerning hyperparameter learning, we focus on to the optimization method maximum a posteriori estimation. The learning is performed by minimizing the negative log-posterior of the hyperparameters $\mathcal{L} =$

$-\ln(p(\mathbf{DQ}, \alpha, \beta|\mathbf{Y}))$, which according to [16] is approximated to an additive constant by

$$\mathcal{L} = \mathcal{L}_\mathbf{Y} + \mathcal{L}_\mathbf{DQ} + \sum \ln\{\sigma_{f,\beta}, \lambda_\beta\} + \sum \ln\{\sigma_{f,\alpha}, \lambda_\alpha, \sigma_{l,\alpha}\}, \quad (30)$$

where

$$2\mathcal{L}_\mathbf{Y} = |\mathbf{K_Y}| + \sum_i \sum_j (\mathbf{K_Y^{-1}})_{ij}\, \mathbf{y}_j \mathbf{y}_i^\top$$

$$2\mathcal{L}_\mathbf{DQ} = m|\mathbf{K_{DQ}}| + \sum_{i=2}^n \sum_{j=2}^n (\mathbf{K_{DQ}^{-1}})_{ij} |\langle \mathbf{dq}_j, \mathbf{dq}_i\rangle| + \langle \mathbf{dq}_1, \mathbf{dq}_1\rangle.$$

The minimization is performed numerically using a gradient descent algorithm.

V. NUMERICAL EVALUATION

In this section we evaluate the approximation properties of an underlying ground truth dynamic $\tilde{f}$ incorporating translation and rotation for the $\mathrm{GP}_{\mathbb{H}_D}$ over dual quaternions and the state of the art GP, where rotation is parametrized using Euler angles. In [10] it is argued, that *mean-prediction* [18], a method ignoring process noise, is advantageous compared to a prediction within noise. However, on propagating a mean-prediction error of latent dynamics to the observation space, the GPDM output accuracy is limited to the error magnitude. Thus, we investigate the root-mean-square error (RMSE) in mean-prediction in the latent dynamics model. All evaluations are performed directly on GPs.

We evaluate the mean-prediction quality of $\mathrm{GP}_{\mathbb{H}_D}$ and the traditional GP depending on the sample size of the training data and the rotation range in the dynamic $\tilde{f}$. The training data set and the ground truth data set are generated using the arbitrary nonlinear function

$$\mathbf{v_x} = \tilde{f}(\mathbf{x}) = \begin{pmatrix} \cos(x) + \sin(y + R) \\ 1 - x\sin(Y + R + z) \\ \frac{\pi}{8} - \sin(Yz - 1) \\ 2c\cos(xy)\sin(R + x) + R \\ \sin(\tan(\frac{z}{3} + \frac{1}{2}) - 2cy) \\ 3c\cos(RY) - \frac{y}{4c} \end{pmatrix}, \quad (31)$$

where every input state $\mathbf{x} = (x, y, z, R, P, Y)$ can be isomorphically converted to a dual quaternion $\mathbf{dq}$ according to (18) and $c > 0$ a constant. An artificial white Gaussian noise with arbitrarily picked standard deviation $\sigma = 3 \cdot 10^{-2}$ is added to the training data set as recommended in [19], to assure numerical stability. To evaluate the quality of the $\mathrm{GP}_{\mathbb{H}_D}$ and the traditional GP, we use the RMSE between the GP predictions for $\mathbf{v_x}$ and the true function value $\tilde{f}(\mathbf{x})$. To obtain reliable testing results, the RMSE is calculated for 10000 iid testing states, where every entry of the input state $\mathbf{x}$ is within the interval $[-3, 3]$. We distinguish between the average RMSE of the GP prediction for translation $\mathbf{v_{x_{tra}}}$ respectively rotation $\mathbf{v_{x_{rot}}}$ for both GP models.

For the translational velocities $\mathbf{v_{x_{tra}}}$ we found in our experiments no significant difference in the prediction quality of $\mathrm{GP}_{\mathbb{H}_D}$ and the traditional GP. For growing training data sets, the average RMSE in translational GP prediction converges around 10^{-2} for both models and the difference in the RMSEs ranges in magnitude 10^{-4}. Therefore, we concentrate

in the remainder of the paper on the quality of the rotational part of GP predictions $\mathbf{v_{x_{rot}}}$.

A. RMSE Depending on Rotation Magnitude

In a first experiment we compare the prediction quality of the GP models, depending on the rotation magnitude in (31). We increase stepwise the constant c in the dynamic $\tilde{f}$ from 0.25 to 4. The number of training samples is fixed to 10^3 for all repetitions and the learning is performed each time with the same set of uniformly drawn random samples. In case the rotational velocity is small, the RMSE of the $GP_{\mathbb{H}_D}$ and the traditional GP prediction are also small. When the rotation magnitude is increased, the performance of the traditional GP is hindered while accuracy of the $GP_{\mathbb{H}_D}$ is significantly less affected as visualized in Figure 2.
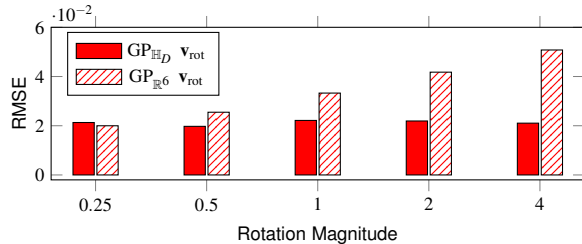
Fig. 2. The prediction quality is visualized for a fixed sample size of the training data, but increasing rotation magnitude in the dynamic $\tilde{f}$. The red bars depict the average RMSEs in the rotational velocity prediction, where solid colors show the $GP_{\mathbb{H}_D}$ and striped bars the traditional GP.

B. RMSE Depending on Number of Samples

In a second numerical evaluation we illustrate the GP prediction quality for $GP_{\mathbb{H}_D}$ and the traditional GP depending on the number of training samples. The model learning is performed on increasing training data sets from 500 to 8000 samples over the 6D space, while the constant c is fixed to 1. Figure 3 shows, that the $GP_{\mathbb{H}_D}$ approximates the underlying ground truth dynamics better than the traditional GP.
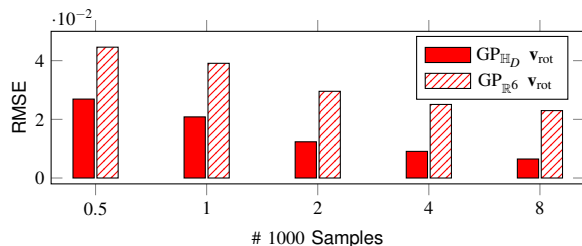
Fig. 3. For increasing numbers of samples used in GP training the prediction quality is visualized. The solid red bars depict the RMSE in the rotation prediction of $GP_{\mathbb{H}_D}$ and the striped bars of traditional GP.

VI. CONCLUSION

In this paper we introduce the GP over dual quaternions and the GPDM over dual quaternions. The extension of the GP input space to $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$ allows for modeling dynamics in the special Euclidean group $SE(3)$, which describes relative distances and orientation preserving transformations. We prove that the functions k, defined in Theorem (3.3) and Theorem (3.5), define valid covariance functions. Introducing

these covariance function into the GPDM density, enhances the GPDM to allow for modeling dynamics in the latent space $S_3/\{\pm 1\} \rtimes \mathbb{R}^3$ or a subspace of it. Through the rigid motion representation using dual quaternions, we demonstrate how the ground truth approximation accuracy of the $GP_{\mathbb{H}_D}$ outperforms the traditional GP on modeling strongly nonlinear dynamics.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] R. H. Shumway and D. S. Stoffer, "Dynamic linear models with switching," *Journal of the American Statistical Association*, vol. 86, no. 415, pp. pp. 763–769, 1991.
[2] D. Jordan and P. Smith, *Nonlinear Ordinary Differential Equations: An Introduction to Dynamical Systems*. Oxford applied and engineering mathematics, Oxford University Press, 1999.
[3] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall Information and System Sciences Series, Pearson Edu. Canada, 1987.
[4] C. Rasmussen and C. Williams, *Gaussian Processes for machine learning*. Adaptative computation and machine learning series, University Press Group Limited, 2006.
[5] 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.
[6] L. Kavan, S. Collins, J. Zara, and C. O'Sullivan, "Skinning with dual quaternions," in *2007 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 39–46, ACM Press, April/May 2007.
[7] 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.
[8] C. E. Rasmussen, "Gaussian processes for machine learning," MIT Press, 2006.
[9] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems 8*, pp. 514–520, MIT press, 1996.
[10] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models," in *In NIPS*, pp. 1441–1448, MIT Press, 2006.
[11] W. Feiten and M. Lang, "Mpg - a framework for reasoning on 6 dof pose uncertainty," (Corporate Technology, Intelligent Systems & Control, Siemens AG, D-80200 Munich), 2011. ICRA Workshop on Manipulation Under Uncertainty.
[12] D. Burago, Y. D. Burago, and S. Ivanov, *A Course in Metric Geometry*. Crm Proceedings, American Mathematical Society, 2001.
[13] D. Zhang, X. Chen, and W. S. Lee, "Text classification with kernels on the multinomial manifold," in *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference*, (New York, USA), pp. 266–273, ACM Press, 2005.
[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
[15] E. Ata and Y. Yayli, "Dual unitary matrices and unit dual quaternions," *Differential Geometry - Dynamical Systems*, vol. 10, pp. 1–12, 2008.
[16] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Trans. Pattern Anal. Machine Intell.*, 2007.
[17] N. Lawrence, "Probabilistic non-linear principal component analysis with gaussian process latent variable models," *J. Mach. Learn. Res.*, vol. 6, pp. 1783–1816, Dec. 2005.
[18] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. New York, NY, USA: Cambridge University Press, 2002.
[19] S. Van Vaerenbergh, M. Lazaro-Gredilla, and I. Santamaria, "Kernel recursive least-squares tracker for time-varying regression," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 23, pp. 1313–1326, Aug 2012.