Equivariant Functional Shape Analysis in SO(3) with Applications to Gait Analysis

Dissertation zur Erlangung
des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August Universität Göttingen
im Promotionsprogramm *Mathematical Sciences*
der Georg-August University School of Science (GAUSS)

vorgelegt von
Fabian J.E. Telschow
aus Kiel

Göttingen, 2016

Betreuungsausschuss:

Prof. Dr. Stephan F. Huckemann
Institut für Mathematische Stochastik, Georg-August Universität Göttingen

Prof. Dr. Axel Munk
Institut für Mathematische Stochastik, Georg-August Universität Göttingen

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Stephan F. Huckemann

Koreferent: Prof. Dr. Axel Munk

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Thomas Schick
Mathematisches Institut, Georg-August Universität Göttingen

Prof. Dr. David Russell Luke
Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen

PD Dr. Timo Aspelmeier
Institut für Mathematische Stochastik, Georg-August Universität Göttingen

Ph.D. Michael R. Pierrynowski
School of Rehabilitation Science, McMaster University, Canada

Tag der mündlichen Prüfung: 16.09.2016

Contents

Acknowledgments

"Difficult times disrupt your conventional ways of thinking and push you to forge better habits of thought, performance and being."

– Robin S. Sharma

This thesis would never have been finished, if there would not exist many amazing people around me who helped, supported and motivated me during the last years. I really want to thank all who have contributed making that happen.

First of all, I thank my first advisor Stephan Huckemann for his patience with me, for time for discussions, giving me the opportunity to travel and do research abroad and introducing me to the topic of non-Euclidean statistics with this exciting application of gait analysis. Further, I thank him for his careful proofreading of this thesis helping to improve its structure and eliminating a lot of bad habits in my writings. Secondly, I thank Michael Pierrynowski who recorded the data underlying this thesis, provided a lot of biomechanical insights and interesting articles. Especially, I want to thank him for the detailed explanations of the experimental setup and discussions about the data analysis. Real data analysis is always a great challenge and can be extremely frustrating from time to time, but nevertheless I am glad that I had the opportunity to study such a complex data set and had help from a biomechanics expert.

I owe many thanks to Max Sommerfeld. I am not only grateful for his suggestions where I could look for answers, but mostly for his faith in me, his interest in my work, and for encouraging me to not give up. Without his support – sometimes only by drinking an espresso together and talking during a work break – I am not sure whether I would have finished my dissertation. Additionally, I thank him for proofreading Chapter 1 and providing a lot of constructive suggestions on making the presentation clearer.

I also really want to thank my friends and colleagues Benjamin Heuer and Alexander Hartmann. They were often a great help in solving small mathematical puzzles and they resolved many of my confusions often by simply arguing that my question is simple. Among other useful inputs, I thank Benjamin Heuer, especially, for his discussions and explanations on uniform convergence of stochastic processes. However, I am also glad that our office hours did not had much of an overlap, since otherwise probably both of us would never have finished their dissertations, since there exist so many fun topics – other than mathematics – to talk about. To Alexander, I owe thanks for proofreading the introduction and providing very helpful Latex support. Last but not least, I thank both of them for really enjoyable lunch breaks, for making the office a fun place to work at and for entertaining board game evenings.

I would like to thank Florian Boßmann from the numerics department for his

help in the development of the algorithm to solve the minimizing problem, which arose in the time warping procedure for curves in Lie groups. My second advisor, Axel Munk, I thank for helping me a few times with discussions, inspiring me to learn about simultaneous confidence bands and providing many interesting articles. I also thank Christian Böhm, the system administrator of the FBMS, for his computer support, improving the working atmosphere in the FBMS and prioritization of some last simulations I needed for this thesis. Moreover, I thank Thomas Schick, Russel Luke and Timo Aspelmeier for taking the time to be members of my oral exam committee.

Among many other friends who supported me each in their own way I want to thank especially René Schulz who taught me in his homebrewed fitness studio that weight lifting – with some intermezzi of serious dialogues – can be a lot of fun, Dominic Simm for just being around and being always available, when I need someone to talk to, and Sina Mayer – my pen pal and good friend – for many lovely walks and stimulating conversations at the Leine and in the forests around Göttingen. I also thank my good friend Artur Wachtel for being so brave to read and discuss with me many parts of *Foundations Of Mechanics* by Abraham and Marsden and *Science And The Modern World* by Whitehead, which was often very confusing and challenging, but nevertheless inspiring. Furthermore, I am grateful to share a deep and long friendship with Matthias Lienert which is one of the most stable relationships in my life, although we are living quite far apart since many years.

Last but not least, I owe special thanks to my family – my parents, my two sisters, my grandfather and my two grandmothers who unfortunately both passed away during my work on this thesis – for their tremendous support, patience, advice and love through all these years. Without all of them, life would be much harder than it actually is and they are the invisible helpful and necessary hands in the background of such a thesis. I will also always remember the brother of one of my grandmothers, who lived in Göttingen and with whom I spent quite a few evenings drinking good wine and eating good food while having fascinating conversations; rest in peace, uncle Manfred.

The research in this dissertation was funded by *Deutsche Forschungsgesellschaft* (Grant DFG HU 1575/4).

Introduction

> *"Grenander [1993] has often emphasized that when we want to model some collections of patterns, it is very important to consider the symmetries of the situation – whether there is an underlying group."*
>
> – Mumford and Desolneux [2010]

What this is all about

In this thesis, we model data stemming from biomechanical gait analysis of the knee joint and analyze these models in order to answer questions about human gait. Our aim is, that all developed methodologies and procedures are tailored to the specific challenges arising from the analysis of real gait data and involve only canonical choices. Interestingly, taking this approach seriously, naturally requires methods from different branches of mathematics, for example, functional data analysis and Lie groups.

Indeed, knee joint data are curves in the group of 3×3 rotations. Therefore, we introduce *Gaussian perturbation of a center curve models*, which are functional models for random curves of rotations exhibiting perturbation consistency (i.e., their center curve can be consistently estimated). The latter is not, eo ipso, true for all perturbation models. For example, perturbation models for shapes in Euclidean spaces introduced in Goodall [1991] do not always have this property (see Lele [1993], Kent and Mardia [1997], Le [1998] and Huckemann et al. [2011]).

In the spirit of Grenander [1993] and his pattern theory marker placements and self-selected walking speeds are identified as the underlying symmetries of center curves of the data at hand. These symmetries can be modeled as a specific Lie group $\mathcal{S}$. Therefore, in order to do inference on two center curves, we remove the effect of $\mathcal{S}$ between these center curves using Procrustes methods, which give strongly consistent estimators. Here *removing* means estimating a group element from $\mathcal{S}$ making the center curves similar.

In order to do inference, we provide different statistical tools to analyze biomechanical gait data and visualize differences between center curves, for example, permutation tests and under the *concentrated error assumption* (i.e., the perturbations are with high probability close to the center curve) simultaneous confidence sets for the center curve.

These methods are applied to real gait data from an experiment studying reliable identification of volunteers and the effect of short kneeling. An additional challenge due to the design of such biomechanical experiments are *small sample sizes*.

Since our setup, as mentioned above, naturally arises from our application, we first have to understand the experiment in more detail.

Motivation from Biomechanics

This thesis is primarily motivated by a specific data set collected in an experiment, designed and carried out by Michael Pierrynowski and Jodi Gallant, McMaster University, Canada. In this experiment, data consisting of relative rotations between the lower and upper leg of different volunteers were collected by tracking the 3D motion of markers placed on specific skin locations of the lower extremities.

The intention of this study was mainly to answer the following two questions. *Firstly, can we reliably identify and distinguish volunteers?* Intuitively, this holds true, since often it is even possible to identify or distinguish individuals just by the sound of their footsteps. *Secondly, is it possible to detect and maybe locate changes in the relative rotations of the lower and upper leg due to degenerative processes (e.g., early onset of osteoarthritis) or specific tasks carried out directly before recording data?*

To this end, healthy volunteers of this study were asked to kneel for 15 minutes prior to recording some parts of the data. This specific posture was chosen, since epidemiological studies identified repetitive, prolonged and deep kneeling as an occupational risk factor for tibiofemoral (knee) osteoarthritis (see Baker et al. [2002], Knoop et al. [2011], Gaudreault et al. [2013] and Kajaks and Costigan [2015]). This association is thought to be caused by overly lax knee ligaments, which are unable to stabilize the motion of the knee joint during activity (see Solomonow [2004]).

The challenges here are that changes in joint motion are usually small compared to the much larger functional joint motion and at the same time the sample sizes are small ($N < 15$). However, small sample sizes are natural from a biomechanical perspective, since it is likely that the influence of the kneeling task fades away after a while.

Within the scope of the analyzed study the answer to the two raised questions will turn out to be: yes, we can. The statistical tools developed in this thesis are able to reliably distinguish and identify volunteers and, moreover, detect changes in the gait pattern of many volunteers after kneeling.

Data and its Acquisition

Unless invasive and expensive methods are used (e.g., X-ray imaging), the typical practice in biomechanical gait analysis of the knee joint is to collect data of locomotion using photogrammetry (i.e., obtaining data from (moving) images of a camera system). In the experiment underlying this thesis, this was done as follows: first anthropometric data of a volunteer (e.g., leg length, knee circumference) was collected. Afterwards reflective markers were placed by an experienced technician on anatomically defined and identifiable skin locations of the right and left lower extremities according to a standard protocol. While the volunteer walked a straight line across the laboratory, the spatial positions of these markers were tracked by a camera system. From the spatial positions together with the anthropometric data a proprietary software (a black box for this thesis) constructed a link-segment model of the lower extremities. It consisted of estimated

knee joint centers and orthonormal coordinate frames (ONFs) of the lower and upper leg, which approximated ONFs of the underlying bones – tibia (main bone in the shank) and femur (bone in the thigh). The relative movements of the bones with

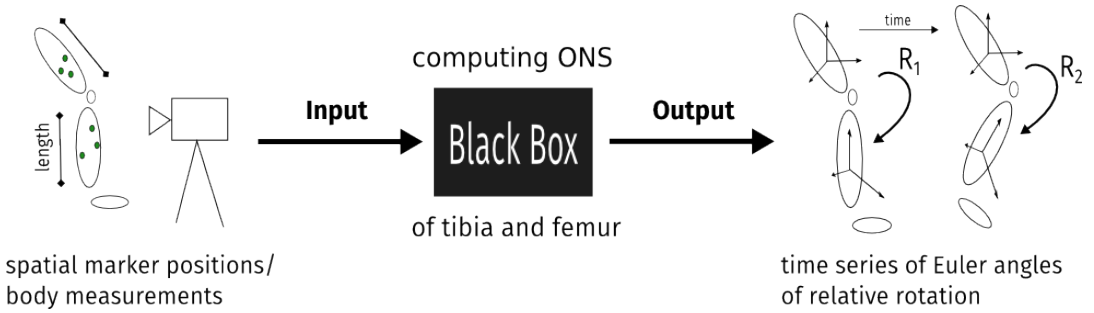

Figure 1: *Schematically illustrated data collection process using a photogrammatic measurement device.*

respect to the laboratory coordinate system were represented by the relative rotations between these frames. The proprietary software of the measurement device provided these relative rotations as a time series of Euler angles (see Appendix A.2), which are specific charts of the group of rotations

$$SO(3) = \left\{ R \in \mathbb{R}^{3\times 3} \mid R^T R = I_{3\times 3} \text{ and } \det(R) = 1 \right\}. \tag{1}$$

In the biomechanical context the three Euler angles describe flexion-extension, abduction-adduction and internal-external rotation of the knee joint.

Schematically, this data collection process is shown in Figure 1. Data recorded using such a protocol exhibit different sources of errors:

I. natural variations in the kinematics of human motor tasks, if specific motion patterns are repeated several times by a volunteer.

II. instrumental errors affecting photogrammetric measurements, resulting in an error on observed marker coordinates.

In Chiari et al. [2005] the instrumental errors are further divided into two general types

IIa) instrumental systematic error (ISE), for example, camera calibration.

IIb) instrumental random error (IRE), for example, electronic noise.

Moreover, Chiari et al. [2005] provide a survey of solutions to these errors, which are still considered state of the art. Usually, ISEs are correctable either by applying an appropriate preprocessing to recorded marker positions before analyzing them or by optimization of the measurement devices. In contrast natural variability in human kinematics and IREs should be subject to statistical modeling.

However, this is not the complete story. There is another type of stochastic errors one has to take into account. We will call them

III. experimental effects.

They are systematic differences in the data stemming from unwanted variability between two experiments, since they are not carried out under perfectly identical conditions. An example can be found in Della Croce et al. [2005]. They discuss and survey the findings of the biomechanics community on precision of marker placements and its effect on variability and reliability of photogrammetric measurements. In the literature, numerous methods have been proposed to reduce this error in the *pre-data recording phase* (e.g., Noehren et al. [2010] for a review and an example). In contrast the literature on corrections during the *post-data recording phase* is sparse; to the best of our knowledge these methods are covered by Woltring [1994], Rivest [2005], Ball and Greiner [2012] and Baudet et al. [2014] and all of them have limitations as we will discuss in Section 3.2. Therefore, we will present in Chapter 3 a new post-data recording solution to this effect. A second example of an experimental effect stems from the common practice in the biomechanical community to record gait data at self-selected walking speeds and not for example on a treadmill (for a comparison of these procedures see Sloot et al. [2014]). This introduces additional variability of the data and is known in the statistics community as *time warping effects*. Statistical methods reducing this effect (e.g., Kneip and Gasser [1992], Wang et al. [1997] and Ramsay and Li [1998]) have been also applied in the biomechanical literature (e.g., Sadeghi et al. [2003]). Since both effects are visible in the data set analyzed in this thesis, our main focus is to establish a statistical sound procedure dealing with them, which thereafter allows for inferential methods. Exemplarily, Figure 2 shows two samples from our data set belonging to the same volunteer. Between recording of the two samples the markers were replaced by an experienced technician, resulting in offsets visible especially in the x- and z-angle.

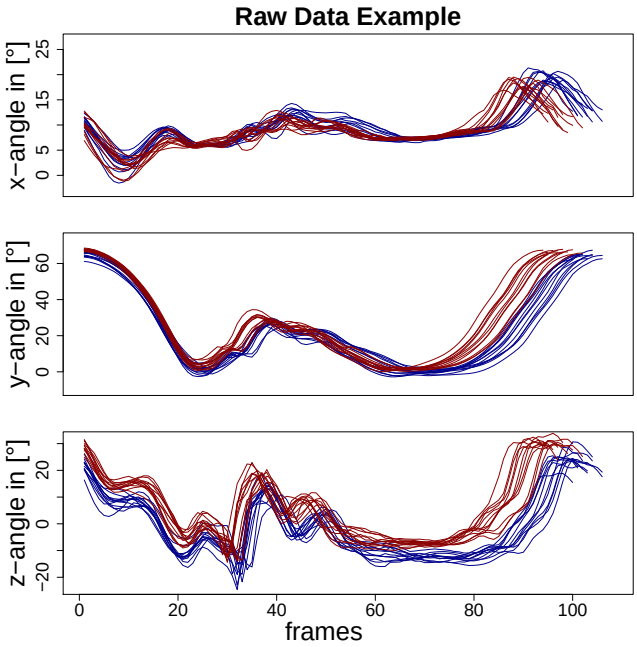

Figure 2: *Example of Euler angle curves of gait cycles (MF-MF) of the same volunteer before (red) and after (blue) marker replacement. In the blue session the volunteer in average walks slower.*

Gait cycles and sessions. A *gait cycle* for a specific leg in biomechanics is usually defined as the time period during locomotion from the first contact with the ground of one foot – called *heel contact* (HC) – to the next heel contact of the same foot. It is divided into the *stance phase*, where the foot has contact with the ground, and the *swing phase* where the foot is in the air. Special events during

a gait cycle while walking include *mid stance* (MS), which roughly corresponds to the minimum flexion (i.e., a local minimum of the y-Euler angle) within the stance phase, *toe off* (TO) the transition between stance and swing phase and *maximal flexion* (MF) corresponding to the global maximal flexion. In Figure 3, we show an example of a gait cycle together with its special events.

An attentive reader may recognize that Figure 2 tacitly includes a different definition of gait cycles. For robustness reasons explained in Chapter 6 a gait cycle in this thesis is the time period between two successive maximal flexion events.

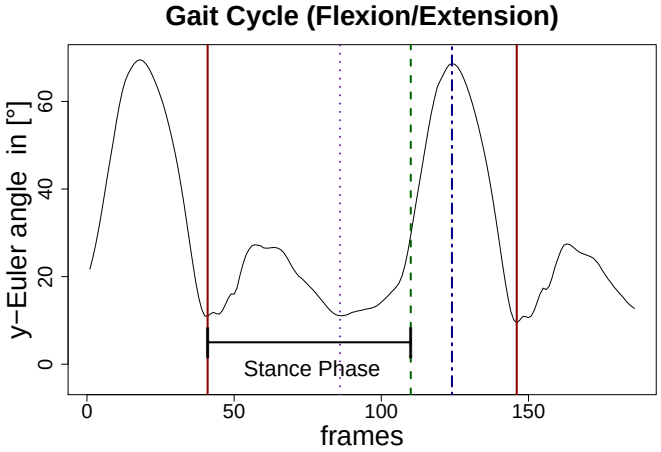

Figure 3: *from left to right, heel contact (red, solid), mid stance (violet, dotted), toe off (green, dash), maximal flexion (blue, twodash)*

Furthermore, a *session* (e.g., Marin et al. [2003]) within a biomechanical experiment is a random sample of walking trials of a volunteer collected under the same experimental conditions, in particular the same marker placements.

Short Overview, Mathematical Challenges and Related Work

Chapter 1

Data space. The first step in the development of a statistical methodology, is to define the data space. We will argue in Chapter 1 that the correct model space of the data at hand is given by the space $\mathcal{X}$ of continuously differentiable curves with values in the Lie group $SO(3)$ of 3×3 rotation matrices. To the best of our knowledge, this is a novel approach in the analysis of biomechanical gait data. Indeed, the commonly used approach is performing statistical analysis directly in Euler angle charts. Here, since gait cycles are approximately periodic, it is often assumed that the curves can be represented by finite Fourier series. Based on this assumption, simultaneous confidence bands of the mean or the difference of means computed using the bootstrap are proposed in Olshen et al. [1989], Johansen and Johnstone [1990] and Lenhoff et al. [1999]. Herein the coordinates are often assumed to be independent. Until now simultaneous confidence bands of Euler angle curves and tests for scalar observables (e.g. step length, force at heel

contact) seem to be the standard methods used in the biomechanics community (among others, see Duhamel et al. [2004], McGinley et al. [2009], Pierrynowski et al. [2010], and Cutti et al. [2014]). Recently, Pataky [2010] and Pataky et al. [2013, 2015] point out that ignoring the correlation between coordinates may result in an improper statistical analysis. However, they still apply multivariate test statistics directly to the Euler angles, for example, pointwise Hotelling T^2 tests with uniform critical values computed using the Gaussian kinematic formula (Cao et al. [1999]).

Considering $SO(3)$ as the data space in biomechanics rather than functional data analysis using the Euler angles is done in Rancourt et al. [2000]. However, they only consider point clouds in $SO(3)$ and develop tests for equality of means of different groups. Rivest [2001] analyzes a $SO(3)$-valued model with fixed rotation axis and applies it to elbow motions with the conclusion that this model does not describe elbow motions well, since the rotation axis varies in time. The articles Haddou et al. [2010] and Rivest et al. [2012] propose $SO(3)$ methods to estimate rotation axis of ankle joints, but they do not model it as functional data.

Even the statistical literature considering curves or functional data with values in manifolds is sparse. Brillinger and Stewart [1998] model elephant-seal movements using diffusions on S^2. Ball et al. [2008] uses Brownian motion and Ornstein-Uhlenbeck processes on the planar shape space with applications to cell shape modeling. In Su et al. [2014] a time warping procedure for curves in manifolds is introduced and used to reduce the temporal variation in bird migration and hurricane track data, which are described as curves in the unit sphere S^2. They also apply it to classification of vehicle trajectories. Kendall [2015] seeks to employ barycenter techniques developed in Kendall et al. [2011], to investigate temporal association between successive hurricane tracks.

Experimental effects. The second step is to model the experimental effects: marker placement and self-selected walking speeds. It turns out that these effects can be described as a Lie group action on the data space $\mathcal{X}$ of the Lie group $\mathcal{S} = \mathcal{I}_0\big(SO(3)\big) \times \mathrm{Diff}^+[0,1]$ the Cartesian product of a certain subgroup $\mathcal{I}_0\big(SO(3)\big)$ of the isometry group of $SO(3)$ (see Definition 1.1.4) with the group of monotone increasing diffeomorphisms $\mathrm{Diff}^+[0,1]$ (see Formula (1.6)). Although time warping is currently used in biomechanics and the marker placement effect is known as correctable by multiplying rotation matrices from each side to the observed curves (e.g., Rivest [2005] and Ball and Greiner [2012]) there exists no clear exposure, and the connection to isometries of $SO(3)$ (see Theorem 1.1.7) was not known. In consequence, the reliability of gait analysis data is still discussed in the literature. Further informations on this topic will be given in Section 3.2.

The fact that the experimental effects are realized by a Lie group action on $\mathcal{X}$, naturally implies that we are within the setting of shape analysis (see among many others Huckemann et al. [2010, 2011] and Celledoni and Eslitzbichler [2015]). Therefore we aim to incorporate the effects of marker placements and self-selected walking speeds into the statistical analysis, using approaches from shape analysis.

Gaussian pertubation models. In Section 1.3 we define *right Gaussian perturbations* (rGP) models describing random curves γ in $SO(3)$ by

$$\gamma(t) = \gamma_0(t)\mathrm{Exp}(A_t)\,,$$

where $\gamma_0 \in \mathcal{X}$ is a deterministic *center curve*, $\mathrm{Exp} : \mathfrak{so}(3) \to SO(3)$ the Lie exponential and a Gaussian process A with almost surely differential sample paths having values in the Lie algebra $\mathfrak{so}(3) \cong \mathbb{R}^3$ of $SO(3)$ (see Definition 1.3.3). Here we call A the *generating Gaussian process* of the rGP. These models are generalizations to functional data of the models given in Rancourt et al. [2000], who himself builds on Downs [1972]. Similar models are also used in Fletcher [2013] for linear regression on manifolds. The important property of these models for biomechanics is that rGPs form a statistical group model i.e., they are $\mathcal{S}$-invariant, meaning that, if $g \in \mathcal{S}$ and γ follows an rGP model, then also $g.\gamma$ (the action of g on γ) follows an rGP model (see Theorem 1.3.5). This especially implies the independence of the coordinate system demanded in Pataky et al. [2013]. In contrast, the Euler angle curves of $g.\gamma$ and γ cannot be assumed to be both stemming from Gaussian processes as implicitly assumed in Pataky et al. [2013, 2015], since the combination of Euler charts and the action of g is a nonlinear transformation.

Since $SO(3)$ is non-commutative, thus in general $\mathrm{Exp}(X)\mathrm{Exp}(Y) \neq \mathrm{Exp}(X + Y)$ for $X, Y \in \mathfrak{so}(3)$, multiplying the error process from the right to the center curve is not a canonical choice. Other reasonable and canonical models are

$$\gamma(t) = \mathrm{Exp}(B_t)\gamma_0(t) \quad \text{and} \quad \gamma(t) = \mathrm{Exp}(C_t)\gamma_0(t)\mathrm{Exp}(D_t)$$

with Gaussian processes B, C, D having almost surely differential sample paths. In Theorem 1.3.7 and 1.3.8 we show that it suffices for our data to consider rGPs only.

Chapter 2

Perturbation consistency. Section 2.1 analyzes some properties of rGP models. Especially, properties of the pointwise extrinsic (population) mean (PEM) and the pointwise extrinsic sample mean (PESM) are studied. These novel descriptors are a combination of the pointwise or cross-sectional population and sample mean from functional data analysis (see Ramsay [2006, Chapter 2.3]) and the extrinsic population and sample means from non-Euclidean statistics (see Bhattacharya and Patrangenaru [2003]).

In Theorem 2.1.7, we prove that for rGP models the PEM is identical with the center curve. Hence, pointwise application of the strong consistency result Theorem 3.4. in Bhattacharya and Patrangenaru [2003] yields that the PESM is pointwise a strongly consistent estimator of the center curve of rGP models. Theorem 2.1.10 extends this result to uniform convergence under additional mild regularity assumptions on the derivative of the generating Gaussian process which are reasonable for the data at hand. Among other ingredients, the proof is based on showing that a derived process is stochastically equicontinuous such that standard uniform stochastic limit theorems (e.g., Newey [1991], Andrews [1992]) can be applied (see also Appendix B).

This result has mainly two important consequences for the rest of the thesis. Firstly, it can be used to establish that the PESM of an rGP is with high probability a differentiable curve (see Corollary 2.1.11), which we need especially for the application of time warping procedures. Secondly, it ensures that our estimator, which can be used to remove the effect of different marker placements, is strongly consistent.

Simultaneous confidence sets. Under further reasonable assumptions on the generating Gaussian process A of an rGP model γ we provide in Sections 2.2 and 2.3 simultaneous confidence sets for its center curve. To this end a pointwise analogue $\{\tilde{H}_t^\gamma\}_{t\in[0,1]}$ of the one sample Hotelling T^2 statistic in $\mathbb{R}^3$ is defined and its maximum over $t \in I = [0,1]$ is considered. The idea is inspired by Rancourt et al. [2000], where an analogue of the Hotelling T^2 statistic and concentrated error assumptions are used to produce approximate confidence sets for a mean rotation. The main challenges in our approach is to estimate for given $\beta \in (0,1)$ the critical value

$$\tilde{h}_{\gamma,\beta} = \inf\left\{ h \in \mathbb{R}_{\geq 0} \;\middle|\; \mathbb{P}\left(\sup_{t\in I} \tilde{H}_t^\gamma \leq h \right) \geq 1-\beta \right\}.$$

Since Theorem 2.2.5 shows that $\tilde{H}^\gamma$ is approximately the Hotelling T^2 process of A under the assumption that the variance of the process A is small, we propose to use the Gaussian kinematic formula (see Cao et al. [1999], Taylor [2006] and Taylor and Worsley [2008]) to obtain an estimator for $\tilde{h}_{\gamma,\beta}$. Note that the use of the Gaussian kinematic formula is similar to its use in Pataky et al. [2013, 2015], but we take the geometry of $SO(3)$ into account. Especially, we emphasize that Theorem 2.2.4 shows that our simultaneous confidence sets are *equivariant* and hence compatible with marker replacements and different walking speeds.

Chapter 3

Removing self-selected walking speeds. In Su et al. [2014] a method for time warping of manifold valued curves is presented, which is a generalization of the time warping procedure introduced in Srivastava et al. [2011a] and Srivastava et al. [2011b] for $\mathbb{R}^D$-valued curves, $D \in \mathbb{N}$. Despite its elegancy, this method requires two (infinite dimensional) non-canonical choices, as will be discussed in more detail in Section 3.1. Therefore we introduce a time warping procedures applicable to curves with values in compact Lie groups, based on our novel *intrinsic length losses* (ILLs, see Definition 3.1.1), requiring only canonical choices. These losses are generalizations of the total variation loss of curves in $\mathbb{R}^D$ to curves in $SO(3)$. We then show that these procedures satisfy the *inverse alignment property* (see Definition 3.0.3), which is also discussed in Srivastava et al. [2011a,b] and Vantini [2012].

Removing the marker placement effect. Section 3.2 explains our approach of *spatial alignment* of two curves $\gamma, \eta \in \mathcal{X}$ in order to remove the effect of different

marker placements. As discussed in Section 1.1, we have that if $\gamma, \eta \in \mathcal{X}$ differ only by marker placements, then there are $P, Q \in SO(3)$ such that

$$\eta(t) = P\gamma(t)Q^T$$

for all $t \in [0, 1]$. Estimating P, Q could be done based on least squares estimation using the Frobenius norm (see Prentice [1989], Chang and Rivest [2001] and Rivest and Chang [2006]). But the estimators for P and Q are numerically difficult to compute.

Our approach builds on Theorem 3.2.4. It states that choosing continuous lifts $\tilde{\gamma}$ and $\widetilde{P\gamma Q^T}$ of curves γ and $P\gamma Q^T$ having values in the group of unit quaternions S^3, a double cover of $SO(3)$, implies that there is a unique $R \in SO(4)$, the group of 4×4 rotation matrices, such that

$$\widetilde{P\gamma(t)Q^T} = R\tilde{\gamma}(t)$$

for all $t \in [0, 1]$. Using this observation and inspired by solutions to spherical (linear) regression (see Chang [1986]), we define a numerically easy to compute (see Theorem 3.2.11) and independent of the particular choice of the lifts (see Theorem 3.2.8) estimator (3.29) for P and Q. Moreover, in Section 3.2.2 we also prove that this estimator is strongly consistent, assuming we have samples from rGP models.

Chapter 4

This chapter introduces different two sample test for equality of PEMs. We introduce the *overlapping of simultaneous confidence sets test* (OCST) for rGP models, which is based on the simultaneous confidence sets constructed in Chapter 2. At this point this method is more of exploratory value, since the significance level cannot be tuned a priori as discussed for real valued Gaussian random variables in Schenker and Gentleman [2001] and Payton et al. [2003]. However, it can be used to visualize at which time points significant differences between the PEMs might appear.

The second test we introduce is a simultaneous Hotelling T^2 test for rGP models. Like the simultaneous confidence sets for rGP models, this test is based on the concentrated error assumption and again uses the Gaussian kinematic formula to estimate the critical value. An Euclidean version of this test based on Cao et al. [1999] and Taylor and Worsley [2008] was applied to muscle force curves in Pataky et al. [2013].

A third test proposed is a permutation test (ILLPerm). It is based on permuting an ILL of PESMs of two samples. Its advantage is that it is nonparametric. However, in contrast to the OCST and the Hotelling T^2 test, it cannot localize significant differences.

All of the above methods require spatial alignment and temporal registration of two sessions before testing, if the considered sessions are recorded with different marker placements and walking speeds. This is done using the procedures introduced in Chapter 3. Since these procedures build on estimators, which themselves

are random variables, this may influence the type I and II errors of these tests. Therefore, we propose a modified version of ILLPerm (viz., MILLPerm), which incorporates the estimation of the spatial alignment and temporal registration between the samples in the permutation scheme in order to take the variance of these estimators into account.

Chapter 5

Simultaneous confidence sets. The first part of this chapter (Section 5.1.1) compares using simulations the small sample size performance ($N \leq 50$) of three different methods for constructing simultaneous confidence bands of the mean curve in functional data. The compared methods are a bootstrap described in Degras [2011], a version of the multiplier bootstrap following Chernozhukov et al. [2013] and the Gaussian kinematic formula for t processes (see Taylor and Worsley [2007]). In our simulations only the Gaussian kinematic formula approach achieves the correct covering rates while for example the bootstrap method of Degras [2011] produces too conservative confidence bands for small sample sizes, which is compatible with the simulation results in Degras [2011]. The intention of these simulations is to justify the use of the Gaussian kinematic formula for computing simultaneous confidence sets for the PEM in rGP models for small sample sizes. However, this study is interesting itself, since most of the developed procedures for simultaneous confidence bands are asymptotic and therefore often only sample sizes $N \geq 30$ are simulated (among others, Krivobokova et al. [2012] and Cao et al. [2012]). In applications, however, it is not unusual to face small sample sizes.

Section 5.1.2 provides a simulation study of our simultaneous confidence sets for the PEM in rGP models. We show that under the small variance assumption our method produces confidence sets with the correct covering rate for different rGP models.

Type I and II errors. The second part of this chapter (Section 5.2) assesses the type I and II errors of the tests described in Chapter 4 using simulations. While the OCST turns out to be a conservative test, all the other tests achieve the given nominal significance level, if no marker placement and self-selected walking speed effects are present. If we include different marker placements and correct it using the methods described in Chapter 3, the type I error of the OCST and the simultaneous Hotelling T^2 test decreases dramatically, whereas ILLPerm becomes a slightly conservative test. Only MILLPerm still holds the given nominal significance level.

Including both experimental effects in our simulations yields that also the type I error of the ILLPerm and MILLPerm decreases slightly. However, due to computational complexity of the time warping procedure, we used a version of MILLPerm which takes only the variance of the marker placement into account. We believe that the complete MILLPerm would hold the given nominal significance level.

Chapter 6

This chapter explains the biomechanical experiment underlying this thesis in more detail and the previously developed statistical tools are applied to the data at hand.

Identification of volunteers. We show in Section 6.3.1.1 that temporal registration and spatial alignment is necessary in order to identify and distinguish volunteers. Indeed, using MILLPerm, we can identify all eight volunteers and distinguish between them for two different self-selected speeds called walk and fast walk.

Detecting changes due to kneeling. MILLPerm detects for many volunteers significant differences (p-value ≤ 0.05) between the control session and the sessions after a kneeling task. These differences can often be located near the gait events heel contact and toe off using the OCST. However, note that in our simulations, the type I error of OCST was unstable, if different marker placements and time warping effects were included. Hence, these localizations have to be taken with care. Kneeling effects are less visible in the fast walk data. This can be explained by the fact that the fast walk data was recorded directly after the walk data and hence the influence of kneeling started to fade away.

Different walking speeds. Interestingly, the individual gait pattern depends on the chosen walking speed. MILLPerm detects also significant differences between fast and walk data of the same volunteer for all volunteers. This is also visible in the the Euler angle curves. Here, OCST suggest that fast walk speed usually includes more flexion around mid stance, which is anatomical reasonable.

Bad marker placements. We also examined the effect of on purpose badly placed markers. We are rarely able to identify volunteers with these placements. However, each sample had only 5-9 valid gait cycles and it is not evident, how the chosen placements, which were not according to the protocol of the measurement device, influence the output of the proprietary software. Due to the chosen positioning of the markers, another possible explanation for the high rejection rate is the so called *soft tissue effect* (i.e., additional stochastic errors due to the movement of muscles or fat tissue beneath the markers). It is known that this reduces the reproducibility of gait patterns (among others, see Leardini et al. [2005] and Taylor et al. [2005b]).

Chapter 1

Statistical Modeling of Biomechanical Gait Data

In the introduction we briefly explained the process of recording biomechanical data of the knee joint by photogrammetry. Thereby we identified two effects – marker placement and different walking speeds – corrupting the data. In order to do statistical inference on this type of data it is necessary to include these effects into modeling.

Hence the next step is to provide spaces modeling these effects and statistical models compatible with them. Since it turns out (see Section 1.1) that marker placement and different walking speeds can be modeled by groups acting on a data space describing similarity transformations of the data space, we will call these effects *gait similarities*.

Hereinafter, we always assume to have an underlying probability space $(\Omega, \mathfrak{B}, \mathbb{P})$ and all topological spaces are endowed with its Borel σ-algebra.

1.1 Model Spaces and Gait Similarities

Recall from the introduction that from spatial observations of markers glued to the leg of a volunteer at each measurement time point $0 = t_1 \leq t_2 \leq ... \leq t_K$, $K \in \mathbb{N}$, positively oriented orthogonal coordinate frames (ONFs) of the tibia and the femur fixed at an estimated knee joint center are constructed by the software of the measurement device. The number K of measurement time points of the observed time series of a gait cycle (≈ 1 second for usual walk) is typically large (≈ 100) and depends on the data collection rate of the measurement device. We assume that this data stems from an unobserved continuous time process (i.e., ideally data would be observable at any point $t \in [0, t_K]$). Moreover, since the data collection rate of a measurement device is constant, but the walking speeds of volunteers may differ, we assume $t_K = 1$ to guarantee comparability of different gait cycles. The latter can always be achieved by linear scaling of the time domain.

Let us denote the orthogonal unit vectors of the constructed ONF at time t of the tibia with $e_1(t), e_2(t), e_3(t)$ and with $f_1(t), f_2(t), f_3(t)$ the corresponding unit

vectors of the femur. They define matrices

$$E(t) = (e_1(t), e_2(t), e_3(t)) \in SO(3) \quad \text{and} \quad F(t) = (f_1(t), f_2(t), f_3(t)) \in SO(3)\,. \tag{1.1}$$

Due to the observation that human gait is a continuous motion, which in principle could be modeled by Newton's law of motion[1] using a second order differential equation, it is reasonable to assume that the movements of markers in space without measuring noise are continuously differentiable, giving a continuously differentiable curve

$$\gamma:\ [0,1] \to SO(3)\,, \qquad t \mapsto \gamma(t) = F(t)E(t)^T \tag{1.2}$$

of relative rotations between the coordinate frames of the bones with respect to the laboratory coordinate system. This assumption is in accordance with the appearance of the data (e.g., Figure 1.1). Therefore we define our data space as follows.

Definition 1.1.1. *We denote with $\mathcal{X}$ the space of parametrized curves of continuously differentiable $SO(3)$-valued paths i.e., $\mathcal{X} = \mathcal{C}^1\big([0,1], SO(3)\big)$.*

Remark 1.1.2. *Later on, the assumption of differentiability is necessary to apply the Gaussian kinematic formula (see Section 2.3) and for removing the influence of different walking speeds. For the latter, absolute continuity of the curves suffices, but in order to have a unified presentation we keep the stronger assumption throughout this thesis.*

The effects of marker placements (MPs). Although in experimental practice trained technicians strive to place markers on approximately the same standard skin locations for each session and volunteer, marker placement causes varying ONFs (see Figure 2). Thus, we say that two trials $\gamma_A, \gamma_B \in \mathcal{X}$ from different sessions are equivalent up to MP, if there exist time independent rotations $Q, P \in SO(3)$ such that

$$QE(t) = \tilde{E}(t) \qquad \text{and} \qquad PF(t) = \tilde{F}(t) \text{ for all } t \in [0,1]\,, \tag{1.3}$$

where $E(t), F(t) \in SO(3)$ are the coordinate frames of the tibia and femur belonging to the original marker placement to obtain curve γ_A computed as in (1.2), similarly $\tilde{E}(t), \tilde{F}(t) \in SO(3)$ are computed after MP to obtain curve γ_B. The rotation matrices P, Q reflect the fact that computed ONFs may differ due to changes in marker placement. They are assumed time independent since the relative positions of the markers with respect to each other except for measurement noise do not change. Moreover, here we assume that soft tissue effects, which are errors due to the non rigid structure of muscles and skin on which the markers are attached, can be neglected or considered as measurement noise.

Using (1.2) in conjunction with (1.3) yields

$$\gamma_B(t) = \tilde{F}(t)\tilde{E}(t)^T = PF(t)E(t)^TQ^T = P\gamma_A(t)Q^T \tag{1.4}$$

relating the curves γ_A and γ_B in case of MP.

[1] More precisely, Euler's laws of motion, since the movement of the bones is modeled as movement of rigid bodies in space.

Remark 1.1.3. *In case that the MP of two trials is identical, but the volunteer moves into a different direction with respect to the recording camera system in each of the trials, an analogous argument shows that the trials are also related by* (1.4)*, now with the only difference that* $P = Q$*. Moreover, note that in contrast to MP this could introduce trial by trial rotations within a session. However, under laboratory conditions underlying the data of this thesis, this effect should be negligible, because the volunteers have been asked to move on a prescribed marked straight pathway.*

Modeling of marker placements. Recall that a group action of a group G on a space $\mathbb{X}$ is given by a map $G \times \mathbb{X} \to \mathbb{X} : \ (g, x) \mapsto g.x$ satisfying $e.x = x$, where e is the neutral element of G, and $g_1.(g_2.x) = (g_1 g_2).x$ for all $x \in \mathbb{X}$ and all $g_1, g_2 \in G$.

Therefore Equation (1.4) suggests that MP can be written as a group action of the Lie group $SO(3) \times SO(3)$ on the space $\mathcal{X}$ given by

$$SO(3) \times SO(3) \times \mathcal{X} \to \mathcal{X}\,, \quad \big((P, Q), \gamma\big) \mapsto (P, Q).\gamma = P\gamma Q^T \tag{1.5}$$

Indeed, this demonstrates that the data objects we are interested in do not belong to the data space $\mathcal{X}$, if MP has to be accounted for, but rather are naturally elements of the shape space

$$\mathfrak{X}_0 = \mathcal{X}/(SO(3) \times SO(3))\,,$$

where the equivalence classes are formed with respect to the group action (1.5).

We used the term *data objects* here. It was introduced and discussed in the articles Wang et al. [2007] and Marron and Alonso [2014]. The aim of those articles is quite different, but the author of this thesis believes that the distinction between data spaces and data object spaces offers an interesting philosophical point of view and is crucial in modeling specific practical problems.

Briefly, the distinction is the following: the data space is formed by the observed quantities and therefore randomness due to the measurement device and random variations of the population should be modeled in this space, whereas statistical descriptors or inference should be defined on the data object space, which is usually a quotient space, since descriptors or statistics of data sets are often considered equal under some similarity transformations. Note that the use of the notion *similarity transformation* here is unusual in the sense that we simply mean transformations describing, which elements of the data space should be considered to be similar or identical to each other despite their original disparity.

This point of view will be developed further and formalized in Section 1.2 using the equivariance principle.

Isometries of Riemannian manifolds. We now want to show that the relation (1.4) between curves with different MP can also be described as the action of the identity component of the isometry group of $SO(3)$ on the data space $\mathcal{X}$. In particular, this will demonstrate that the space $\mathfrak{X}_0$ is a natural generalization of a shape space of curves in $\mathbb{R}^D$ (see Srivastava et al. [2011a]) to curves in $SO(3)$.

To make this more precise we recall some basic notations and results from differential geometry. Recommendable introductions to Lie groups and differential geometry are Helgason [1962], do Carmo Valero [1992], Hilgert and Neeb [2011] and Lee [2013].

Definition 1.1.4. *Let $\mathcal{M}$ be a $\mathcal{C}^1$-Riemannian manifold with $\langle \cdot, \cdot \rangle : \Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \to \mathcal{C}^1(\mathcal{M}, \mathbb{R})$ its $\mathcal{C}^1$-Riemannian metric a differentiable map $f : \mathcal{M} \to \mathcal{M}$ is called isometry, if*

$$\langle d_p f(X), d_p f(Y) \rangle_{f(p)} = \langle X_{f(p)}, Y_{f(p)} \rangle_{f(p)}$$

for all differentiable vector fields $X, Y \in \Gamma(T\mathcal{M})$ on $\mathcal{M}$ and all $p \in \mathcal{M}$.

The space of all isometries $\mathcal{I}(\mathcal{M})$ is called the isometry space of the Riemannian manifold $\mathcal{M}$. Its connected component including the identity map $id_{\mathcal{M}} : p \mapsto p$ for $p \in \mathcal{M}$ is denoted with $\mathcal{I}_0(\mathcal{M})$.

Recall that any connected Riemannian manifold $\mathcal{M}$ can be given the structure of a metric space $(\mathcal{M}, d)$ (e.g., Lee [2013, Theorem 13.29, p.339]). Then one obtains that the set of isometries $\mathcal{I}(\mathcal{M})$ is identical with the set of distance preserving mappings, since $f : \mathcal{M} \to \mathcal{M}$ is an isometry if and only if $d(f(p), f(q)) = d(p, q)$ for all $p, q \in \mathcal{M}$ (see Helgason [1962, p. 60, §11]).

Example 1.1.5. *The Euclidean space $\mathbb{R}^D$, $D \in \mathbb{N}$, endowed with the Euclidean scalar product on its tangent space forms a Riemannian manifold. Its isometry group*

$$\mathcal{I}\big(\mathbb{R}^D\big) = \left\{ T : \mathbb{R}^D \to \mathbb{R}^D \mid \|T(x) - T(y)\| = \|x - y\| \text{ for all } x, y \in \mathbb{R}^D \right\},$$

where $\|\cdot\|$ denotes the Euclidean norm, is equal to the Euclidean motion group

$$E(D) = O(D) \ltimes \mathbb{R}^D$$

i.e., the semi-direct product of the orthogonal group $O(D)$ and $\mathbb{R}^D$. As a group this semi-direct product is given by the set $O(D) \times \mathbb{R}^D$ with group multiplication

$$(R_1, a_1)(R_2, a_2) = (R_1 R_2, R_2 a_1 + a_2)$$

for $R_1, R_2 \in O(D)$ and $a_1, a_2 \in \mathbb{R}^D$. The identity component of $\mathcal{I}\big(\mathbb{R}^D\big)$ is given by the special Euclidean motion group

$$\mathcal{I}_0\big(\mathbb{R}^D\big) = SE(D) = SO(D) \ltimes \mathbb{R}^D .$$

Of course, there is a natural action of $SE(D)$ on $\mathcal{C}^1\big([0,1], \mathbb{R}^D\big)$ given by the pointwise application of $g = (R, a) \in SE(D)$ to $\gamma \in \mathcal{C}^1\big([0,1], \mathbb{R}^D\big)$ i.e.,

$$(g.\gamma)(t) = R\gamma(t) + a$$

for all $t \in [0, 1]$.

Moreover, note that the set of isometries of a Riemannian manifolds has a useful structure:

Theorem 1.1.6 (Myers and Steenrod [1939]). *Let $\mathcal{M}$ be a Riemannian manifold. Then $\mathcal{I}(\mathcal{M})$ is a Lie group.*

By the *closed subgroup theorem* (see Lee [2013, p. 523, Theorem 20.12]) and the above theorem the closed subgroup $\mathcal{I}_0(\mathcal{M})$ is also a Lie group with the Lie group structure induced by $\mathcal{I}(\mathcal{M})$.

In this work the Lie group $SO(3)$ will always be endowed with its unique (up to a scalar) bi-invariant Riemannian metric (see Gilkey et al. [2015, Lemma 6.24, p. 76]). The identity component $\mathcal{I}_0\big(SO(3)\big)$ with respect to this Riemannian metric has a surprisingly simple structure as can be deduced using a more general result by Helgason [1962, p. 207 Theorem 4.1]:

Theorem 1.1.7. $\mathcal{I}_0\big(SO(3)\big) \cong SO(3) \times SO(3)$ *as Lie groups.*

Proof. Using Helgason [1962, Theorem 4.1 (i) on p. 207], we have to assert that $\mathcal{G} := SO(3) \times SO(3)$ and $\mathcal{K} := \operatorname{diag}(SO(3) \times SO(3))$ form a Riemannian symmetric pair where $\mathcal{G}$ is semisimple and acts effectively on $\mathcal{G}/\mathcal{K} = \{[g,h] : g,h \in SO(3)\}$, $[g,h] = \{(gk,hk) : k \in SO(3)\}$. Here the action is given by $(g',h') : [g,h] \mapsto [g'g,h'h]$. The fact that $(\mathcal{G},\mathcal{K})$ is a Riemannian symmetric pair is asserted in Helgason [1962, p. 207], $SO(3)$ is simple, hence semisimple and the effective action follows from the fact $\mathcal{G}$ has no trivial normal divisors $N \subset \mathcal{K}$ (see Helgason [1962, p. 110]). For if $\{(e,e)\} \neq N \subset \mathcal{K}$ would be a normal divisor of $\mathcal{G}$ then there would be a subgroup $\{e\} \neq H$ of $SO(3)$ with the property that for every $h \in H, g,k \in SO(3)$, in particular for $g \neq k$, there would be $h' \in H$ such that

$$(g,k)(h,h)(g^{-1},h^{-1}) = (h',h')\,,$$

i.e. $k^{-1}g$ would be in the center of $SO(3)$, which, however, is trivial because $SO(3)$ is non commutative and simple. Hence $g = k$, a contradiction. □

Note that by the definition of a bi-invariant metric the maps $R \mapsto PRQ^T$ are isometries of $SO(3)$ and therefore the action (1.5) is an action of isometries. However, the last Proposition additionally implies that this action coincides with the action of the group $\mathcal{I}_0\big(SO(3)\big)$ on $\mathcal{X}$ pointwise given by $(\psi.\gamma)(t) = \psi\big(\gamma(t)\big)$ for all $\gamma \in \mathcal{X}$ and all $\psi \in \mathcal{I}_0\big(SO(3)\big)$ i.e., for each $\psi \in \mathcal{I}_0\big(SO(3)\big)$ there is a unique $(P_\psi, Q_\psi) \in SO(3) \times SO(3)$ such that

$$\psi.\gamma = (P_\psi, Q_\psi).\gamma = P_\psi \gamma\, Q_\psi^T$$

for all $\gamma \in \mathcal{X}$ and vice versa.

These considerations establish that our data object space $\mathfrak{X}_0$ can equivalently be defined as the space

$$\mathfrak{X}_0 = \mathcal{X}/\mathcal{I}_0\big(SO(3)\big)\,,$$

revealing that it is simply a generalization of the shape space

$$\mathcal{C}^1\big([0,1],\mathbb{R}^D\big)/SE(D) = \mathcal{C}^1\big([0,1],\mathbb{R}^D\big)/\mathcal{I}_0\big(\mathbb{R}^D\big)$$

of curves without size and reflections in $\mathbb{R}^D$ to the Lie group $SO(3)$ (compare Example 1.1.5 and Srivastava et al. [2011a]).

Different walking speeds. If one observes processes in continuous time, they are often happening at different speeds introducing variability in the time domain of the process (also called phase variability). Children's growth, for example, is such a process, since it is a consequence of a complex sequence of hormonal events not happening at the same rate at the same age for every child. Nevertheless, there are patterns as the pubertal growth phase which are identifiable in all childrens (see Ramsay et al. [1995]). In fact, problems of this type are arising frequently in the analysis of functional data. Therefore it is often necessary to implement procedures such as *time warping* (also called curve registration or temporal registration) to reduce the variability in the time domain before any further analysis of the data is carried out (see e.g., Wang et al. [1997], Ramsay [2006, Chapter 7], Kneip and Ramsay [2008], Kutzner et al. [2010] and Su et al. [2014]).

In our application, variability in the time domain of the space $\mathcal{X}$ can be modeled as the action of the Lie group of monotonically increasing diffeomorphisms of $[0,1]$ on the data space $\mathcal{X}$ i.e.,

$$\mathrm{Diff}^+[0,1] = \Big\{\phi \in \mathcal{C}^\infty\big([0,1],[0,1]\big) \;\Big|\; \phi'(\mathrm{t}) > 0 \;\text{ for all } \mathrm{t} \in (0,1)\Big\} \tag{1.6}$$

with group multiplication given by composition. Here the action is given by

$$\mathrm{Diff}^+[0,1] \times \mathcal{X} \to \mathcal{X}\,, \quad (\phi,\gamma) \mapsto \phi.\gamma = \gamma \circ \phi\,.$$

Figure 1.1 demonstrates this effect using an example of the data set motivating this work. The only preprocessing done in this figure is a linear scaling of time as explained earlier such that the time domain of each trial is $[0,1]$. Apart from the visible spatial discrepancy due to MP, this figure suggests that the data could be corrupted by time warping effects in two different ways:

Let us denote the trials of a session with $\gamma_1, ...\gamma_N$, $N \in \mathbb{N}$. Moreover, assume that $\gamma_1^*, ..., \gamma_N^*$ are i.i.d. samples of a random curve $\gamma \in \mathcal{X}$. Then the influences of time variability (see e.g., Figure 1.1) can be modeled by

1. individual time warping (ITW) i.e., $\gamma_n = \phi_n.\gamma_n^*$ with unknown individual $\phi_n \in \mathrm{Diff}^+[0,1]$ for each $n \in \{1, ..., N\}$ and/or

2. sessionwise time warping (STW) i.e., $\gamma_n = \phi.\gamma_n^*$ for all $n \in \{1, ..., N\}$ with a common unknown $\phi \in \mathrm{Diff}^+[0,1]$.

The latter is interesting, if different sessions need to be compared, since it models that the average speed could vary between different sessions (even if they are samples from the same random curve). The right hand side of Figure 1.1 illustrates this point by showing mean curves of two sessions.

In this work we will only focus on STW for modeling and inference, since it turns out in Section 6.3.1.1 that STW is necessary for comparison of different sessions, whereas ITW actually should not be modeled, which will be detailed in Section 6.3.1.1 using the data underlying this thesis.

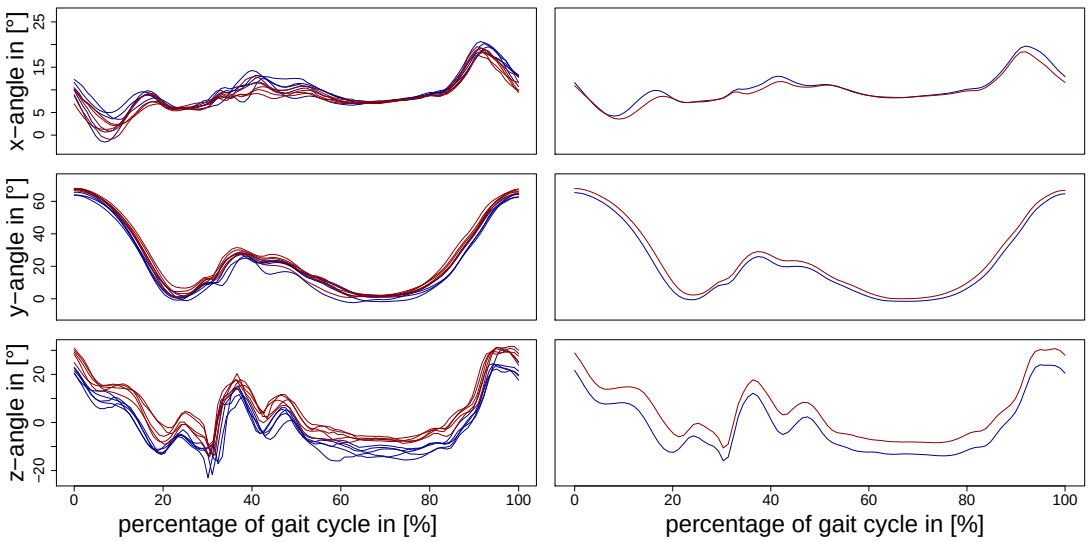

Figure 1.1: *left, ten gait cycles of two sessions (red and blue) of the same volunteer with different marker placement. Right, mean curves of these sessions. Around 20% of the gait cycle STW is visible, since extrema are not aligned.*

The group modeling gait similarities. We may now define the Lie group $\mathcal{S}$ modeling gait similarities between sessions as

$$\mathcal{S} = \mathcal{I}_0\big(SO(3)\big) \times \mathrm{Diff}^+[0,1]\,.$$

Since the actions of $\mathcal{I}_0\big(SO(3)\big)$ and $\mathrm{Diff}^+[0,1]$ on $\mathcal{X}$ commute, the action of $\mathcal{S}$ on $\mathcal{X}$ is given by

$$\mathcal{S} \times \mathcal{X} \to \mathcal{X}\,, \quad \big((\psi,\phi),\gamma\big) \mapsto (\psi,\phi).\gamma = \psi.\,(\phi.\gamma) = \phi.\,(\psi.\gamma) = \psi \circ \gamma \circ \phi$$

From the foregoing exposition it is evident that, if we want to compare different sessions, we cannot simply compute a descriptor of a session and compare it with the same descriptor computed from a different session, since the data may be corrupted by an unknown element of the group $\mathcal{S}$. Thus, we clearly have to make good for this unknown group element acting on the data, if we compare different sessions.

This leads to the idea that any useful descriptor of this data itself should be compatible with the action of the group $\mathcal{S}$. The next section will be devoted to the formal meaning of compatibility (i.e., equivariance) and thus to a class of meaningful statistical models and descriptors under the presence of groups acting on the data space.

1.2 Statistical Models and Equivariance

The previous section established that a reasonable model space for our data is the space $\mathcal{X}$ of differentiable curves taking values in the Lie group of rotations $SO(3)$, endowed with an action of the Lie group $\mathcal{I}_0\big(SO(3)\big) \times \mathrm{Diff}^+[0,1]$.

In order to properly include this group action into our statistical model, we will use the *equivariance principle* and *group invariant models*, which in more detail are discussed for example in Lehmann and Casella [1998], and combine it later on (see Chapter 3) with ideas from shape analysis, especially Procrustes analysis (e.g., Dryden and Mardia [1998, Chapter 5]).

We discuss this setup in an abstract fashion such that these considerations can also be applied to other practical problems involving similar group actions on a data space. Most of the following results and definitions are stated in Lehmann and Casella [1998] or can be deduced from his writings. For convenience, we restructure it differently and provide some of the proofs Lehmann omits.

Group invariant statistical models. Assume we have a space $\mathbb{X}$ and a group G acting on $\mathbb{X}$, i.e. there is a map $G \times \mathbb{X} \to \mathbb{X} : (g, x) \mapsto g.x$ satisfying $e.x = x$, where e is the neutral element of G, and $g_1.(g_2.x) = (g_1 g_2).x$ for all $x \in \mathbb{X}$ and all $g_1, g_2 \in G$.

We want to endow the space $\mathbb{X}$ with a family of probability distributions compatible with the action of G. Therefore let us recall the definition of a G-invariant statistical family (see Lehmann and Casella [1998, Def. 2.1]).

Definition 1.2.1. *Let G be a group acting on $\mathbb{X}$. Let Θ be a set. We call a statistical model $(\mathbb{X}, \{\mathbb{P}_\theta\}_{\theta\in\Theta})$ G-*invariant*, if for every $\theta \in \Theta$ and every $g \in G$ there exists $\theta' = \theta'(\theta, g) \in \Theta$ such that $g.X \sim \mathbb{P}_{\theta'}$, whenever $X \sim \mathbb{P}_\theta$. Here $g.X$ denotes the random variable taking the value $g.x$ whenever X has the value x.*

Additionally, we call the family identifiable *if $\theta' \neq \theta$ implies $\mathbb{P}_{\theta'} \neq \mathbb{P}_\theta$.*

In any G-invariant, identifiable statistical model the group G induces a transformation group $\bar{G}$ given by the transformations $\theta \mapsto \bar{g}\theta = \theta'(\theta, g)$ of Θ. Note that given $g \in G$ and $\theta \in \Theta$ the $\bar{g}\theta$ is unique due to the identifiability, implying that the map $\theta \mapsto \bar{g}\theta$ is well-defined.

Lemma 1.2.2. *Let $(\mathbb{X}, \{\mathbb{P}_\theta\}_{\theta\in\Theta})$ G-*invariant *and identifiable. Then the set of transformations*

$$\bar{G} = \{\theta \mapsto \bar{g}\theta \mid g \in G\}$$

form a group acting on Θ. Moreover, each $\bar{g} \in \bar{G}$ is bijective.

Proof. Define a multiplication on $\bar{G}$ by

$$(\bar{g}_1\bar{g}_2)\theta = \overline{g_1 g_2}\theta$$

for all $\theta \in \Theta$. Now, it follows at once that $\bar{e}$ is the neutral element, $\bar{g}^{-1} = \overline{g^{-1}}$ and the multiplication is associative.

To obtain the surjectivity of the transformations. Let $\theta' \in \Theta$ be arbitrary and $X \sim \mathbb{P}_{\theta'}$. Then $g.(g^{-1}.X) \sim \mathbb{P}_{\theta'}$ and there is a $\theta \in \Theta$ such that $g^{-1}.X \sim \mathbb{P}_\theta$. Thus by definition $\bar{g}\theta = \theta'$.

For the injectivity assume $\bar{g}\theta_1 = \bar{g}\theta_2$ for a $g \in G$. Then there exist random variables $X_1 \sim \mathbb{P}_{\theta_1}$ and $X_2 \sim \mathbb{P}_{\theta_2}$ and

$$X_1 \sim g^{-1}.(g.X_1) \sim g^{-1}.(g.X_2) \sim X_2$$

due to $\bar{g}\theta_1 = \bar{g}\theta_2$. This shows that $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ and hence identifiability implies $\theta_1 = \theta_2$. □

Remark 1.2.3. *Note that in general $G \ncong \bar{G}$ as groups even in identifiable statistical models, since the map $\pi\colon g \mapsto \bar{g}$ is only a surjective group homomorphism. However, of course $\bar{G} \cong G/\ker(\pi)$.*

Example 1.2.4. *1.) Let $\mathbb{X} = \mathbb{R}^D$ and $G = \{A \in \mathbb{R}^{D\times D} |\ A = \operatorname{diag}(\lambda_1, ..., \lambda_D)\,,\ \lambda_d > 0$ for all $d = 1, ..., D\}$ the group of invertible diagonal $D \times D$ matrices. The action of G on $\mathbb{R}^D$ is given by matrix multiplication, i.e. $A.x = Ax$ for all $A \in G$ and all $x \in \mathbb{R}^D$.*

Let $X \sim \mathcal{N}(0, \Sigma)$ for $\Sigma \in G$. The random variable $A.X$, for $A \in G$, is distributed according to $\mathcal{N}(0, A\Sigma A)$. Since G forms a group we have that $A\Sigma A \in G$. Thus, $\left(\mathbb{R}^D, \{\mathcal{N}(0, \Sigma)\}_{\Sigma\in G}\right)$ is an identifiable G-invariant statistical model. But $\bar{G} \ncong G$, since for any $R \in G \cap O(D)$, say $R = -I_{D\times D}$, we have that $R.X \sim X \sim \mathbb{P}_\Sigma$ for all $\Sigma \in G$.

2.) Let $\mathbb{X} = \mathcal{C}^1([0, 1], \mathbb{R}^2)$ and $G = SE(2)$. Then $g = (R, a) \in SE(2)$ acts on $\mathbb{X}$ by

$$(g.\gamma)(t) = R\gamma(t) + a$$

for all $t \in [0, 1]$ and for any $\gamma \in \mathcal{C}^1([0, 1], \mathbb{R}^2)$. We endow $\mathbb{X}$ with a family of probability measures in the following way:

Let Z denote a mean zero Gaussian process with index set $[0, 1]$ and values in $\mathbb{R}^2$, which has a continuously differentiable version. The set of all these processes is denoted with $\mathcal{G}$. We define the set $\Theta = \mathcal{C}^1([0, 1], \mathbb{R}^2) \times \mathcal{G}$. For any $\gamma_0 \in \mathbb{X}$ the random variable

$$\gamma = \gamma_0 + Z$$

induces a distribution $\mathbb{P}_\theta = \mathbb{P}_{\gamma_0,Z}$ on $\mathbb{X}$. Note that $(\mathbb{X}, \{\mathbb{P}_\theta\}_{\theta\in\Theta})$ is an identifiable $SE(2)$-invariant statistical model and that the induced action of $SE(2)$ on Θ is given by

$$\tilde{g}\theta = (R\gamma_0 + a, RZ)\ .$$

For G-invariant statistical models on $\mathbb{X}$ it is reasonable to restrict the class of statistics to those compatible with the group action in the following sense: a statistic of a random variable $g.X$ should be computable only by the knowledge of the corresponding statistic of X and the group element $g \in G$. Heuristically, this expresses the requirement that a statistical procedure should be independent of experimental effects. Mathematically, this is formalized using the notion of equivariance.

Definition 1.2.5. *Let $\mathbb{Y}$ be a measure space. A statistic $T : \mathbb{X} \to \mathbb{Y}$ (measurable map) is called* G-equivariant*, if there exists a map $F : G \times \mathbb{Y} \to \mathbb{Y}$ such that $F(g, T(X)) = T(g.X)$ for all $X \in \mathbb{X}$ and all $g \in G$.*

Proposition 1.2.6. *If T is an G-equivariant statistic, then the map F induces a group action of G on $T(\mathbb{X})$ given by $g.T(X) = F(g, T(X))$ for all $g \in G$ and $X \in \mathbb{X}$.*

Proof. First, note that the action is well-defined on $T(\mathbb{X})$, since

$$g.T(X) = F(g, T(X)) = T(g.X) \in T(\mathbb{X})\,, \tag{1.7}$$

for all $g \in G$ and all $X \in \mathbb{X}$ and, indeed, this defines a group action, since application of equation (1.7) yields $e.T(X) = T(e.X) = T(X)$ and

$$(g_1 g_2).T(X) = T((g_1 g_2).\, X) = T(g_1.(g_2.X)) = g_1.T(g_2.X) = g_1.\,(g_2.T(X))\ ,$$

for all $g_1, g_2 \in G$ and all $X \in \mathbb{X}$. □

Remark 1.2.7. *Note that vice versa, if the space $\mathbb{Y}$ admits a group action of the group G there is a canonical choice of F given by $F(g, y) = g.y$ for all $y \in \mathbb{Y}$ and all $g \in G$. In this case a statistic $T : \mathbb{X} \to \mathbb{Y}$ is equivariant, if $T(g.X) = g.T(X)$ for all $g \in G$ and all random variables $X \in \mathbb{X}$.*

Example 1.2.8. *In Example 1.2.4 2.) let $\gamma \sim \mathbb{P}_{\gamma_0,Z}$ be a random variable. Then the pointwise mean curve $\mathbb{E}[\gamma]$ and the pointwise variance curve $\mathrm{Var}[\gamma]$ given by*

$$\mathbb{E}[\gamma](t) = \mathbb{E}[\gamma(t)] \qquad \textit{and} \qquad \mathrm{Var}[\gamma](t) = \mathbb{E}\Big[(\gamma(t) - \mathbb{E}[\gamma(t)])(\gamma(t) - \mathbb{E}[\gamma(t)])^T\Big]$$

for all $t \in [0, 1]$ are equivariant statistics, since they satisfy

$$\mathbb{E}[(R, a).\gamma] = R\,\mathbb{E}[\gamma] + a\,, \qquad \mathrm{Var}[(R, a).\gamma] = R\,\mathrm{Var}[\gamma]R^T\,,$$

for all $(R, a) \in SE(2)$, due to the linearity of the expectation.

Testing in the data object space. So far we introduced G-invariant statistical models on a data space $\mathbb{X}$. We now want to elaborate at which point shape analysis will come into play to form together with a G-invariant statistical model an adequate statistical framework for comparison of samples possibly corrupted by elements of a similarity transformation group G. This is exactly the challenge we face, if we want to compare sessions from our gait data, since any session is collected with a slightly different marker placement and a different mean walking speed. By our discussion in Section 1.1 this results in an application of an unknown element (ψ, ϕ) of the gait similarity group $\mathcal{S}$ on all sample curves of a session, which prohibits that we can compare sessions directly.

Definition 1.2.9. *Let $\mathbb{X}$ be a data space and G a group acting on $\mathbb{X}$. We define the space of samples as*

$$\mathbb{S}(\mathbb{X}) = \big\{(x_1, \ldots, x_N) \in \mathbb{X}^N \mid N \in \mathbb{N}\big\} = \bigcup_{N \in \mathbb{N}} \mathbb{X}^N\ ,$$

on which the group G acts canonical by $g.\chi = (g.x_1, \ldots, g.x_N) \in \mathbb{S}$ for all $\chi \in \mathbb{S}$ and all $g \in G$.

In order to endow the space $\mathbb{S}(\mathbb{X})$ with a G-invariant family of distributions note that any G-invariant family of distributions $\mathbb{P}_\theta$ on $\mathbb{X}$ induces a G-invariant family of distributions on $\mathbb{S}(\mathbb{X})$ by

$$(X_1, \ldots, X_N) \sim \mathbb{P}_\theta \otimes \cdots \otimes \mathbb{P}_\theta$$

i.e., on each $\mathbb{X}^N$ we have the N-fold product measure of $\mathbb{P}_\theta$. This invariant family of distributions on $\mathbb{S}(\mathbb{X})$ models the case where a sample $\chi \in \mathbb{S}(\mathbb{X})$ consists of i.i.d. realizations from the distribution $\mathbb{P}_\theta$, which is the situation occurring most often in applications. In accordance with the notion from biomechanical experiments we define the following.

Definition 1.2.10. *We call a sample* $\chi = (x_1, \dots, x_N) \in \mathbb{S}(\mathbb{X})$, $N \in \mathbb{N}$, *with* $x_1, ..., x_N$ *i.i.d. a* session. *A single realization* $x_n \in \mathbb{X}$, $n \in \{1, ..., N\}$, *belonging to a session* χ *will be called a* trial.

Assume a session $\chi_1 \in \mathbb{S}(\mathbb{X})$ consisting of realizations from a distribution $\mathbb{P}_\theta$ and a session $\chi_2 \in \mathbb{S}(\mathbb{X})$ consisting of realizations from a distribution $\mathbb{P}_{\theta'}$ are given. One possibility is then to test whether there exists a $g \in G$ such that $\theta' = \bar{g}\theta$. Note that this is related to the idea of deformable models (see Del Barrio et al. [2015]) and references therein.

However, we are not necessarily interested in comparing the distributions, but comparing only descriptors of them. Therefore assume $T : \mathbb{X} \to \mathbb{Y}$ is a G-equivariant statistic. Our goal is to test the hypothesis

$$H_0 : \ \exists g \in G : \ T(X) = T(g.Y) \quad versus \quad H_1 : \ \forall g \in G : \ T(X) \neq T(g.Y)\,,$$

where $X \sim \mathbb{P}_\theta$ and $Y \sim \mathbb{P}_{\theta'}$ using the sessions χ_1 and χ_2. By the G-equivariance of our statistic we may rewrite this into

$$H_0 : \ [T(X)]_G = [T(Y)]_G \quad versus \quad H_1 : \ [T(X)]_G \neq [T(Y)]_G\,,$$

where the equivalence classes $[T(X)]_G = \{Y \in \mathbb{Y} \mid \exists g \in G : \ T(g.X) = Y\}$ are formed with respect to the group action of G on $\mathbb{Y}$ induced by T (see Proposition 1.2.6). Therefore we are left with a testing problem in the shape space $T(\mathbb{X})/G$.

In order to illustrate this point we close this paragraph with an example:

Example 1.2.11. *Assume we are in the setting of Example 1.2.4 2.) and let* $\chi_1 = (\gamma_1, .., \gamma_N)$ *a session of realizations from* $\gamma \sim \mathbb{P}_{\gamma_0, Z}$ *and* $\chi_2 = (\eta_1, .., \eta_M)$ *a session of realizations from* $\eta \sim \mathbb{P}_{\eta_0, Z'}$. *Recall from Example 1.2.8 that the pointwise mean is a* $SE(2)$*-equivariant statistic.*

Since the action of $SE(2)$ *could be present for example due to coordinate changes between the sessions* χ_1 *and* χ_2, *we could be interested in testing the hypothesis*

$$\begin{aligned} H_0 : \ & \exists (R, a) \in SE(2) : \ \mathbb{E}[\eta] = \mathbb{E}[R\gamma + a] \\ versus \quad H_1 : \ & \forall (R, a) \in SE(2) : \ \mathbb{E}[\eta] \neq \mathbb{E}[R\gamma + a]\,, \end{aligned}$$

which is equivalent to testing whether

$$H_0 : \ \big[\mathbb{E}[\eta]\big]_{SE(2)} = \big[\mathbb{E}[\gamma]\big]_{SE(2)} \quad versus \quad H_1 : \ \big[\mathbb{E}[\eta]\big]_{SE(2)} \neq \big[\mathbb{E}[\gamma]\big]_{SE(2)}\,,$$

where the equivalence classes belong to the space $\mathcal{C}^1\big(I, \mathbb{R}^2\big)/SE(2)$ *given by the pointwise action of* $SE(2)$ *on* $\mathcal{C}^1\big([0,1], \mathbb{R}^2\big)$. *Tests for these hypothesis could for example be based on the pointwise sample mean curves of the sessions* χ_1 *and* χ_2, *which are* $SE(2)$*-equivariant statistics on* $\mathbb{S}(\mathbb{X})$.

1.3 Gaussian Models for Gait Data

After we introduced in the previous section desirable properties of a statistical model in order to deal properly with shape effects, we will now discuss a model possessing these properties, which we will use in our statistical analysis of biomechanical gait data.

Definition 1.3.1. *A $\mathbb{R}^3$-valued stochastic process $A = \{A_t\}_{t\in I}$ (i.e., a family of $\mathbb{R}^3$-valued random variables on Ω indexed by $I = [0,1]$) is called* Gaussian process*, if for any $K \in \mathbb{N}$ and real numbers $0 \leq t_1 < ... < t_K \leq 1$ the random vector $\left(A_{t_1}, ..., A_{t_K}\right)$ has a (possibly degenerate) joint Gaussian distribution.*

We say that the Gaussian process A has a $\mathcal{C}^1$-version, if there is a Gaussian process $\tilde{A}$ satisfying $t \mapsto \tilde{A}_t(\omega) \in \mathcal{C}^1\big(I, \mathbb{R}^3\big)$ for almost all $\omega \in \Omega$ and $\tilde{A}$ is a version of A i.e.,

$$\mathbb{P}\Big(A_t = \tilde{A}_t\Big) = 1$$

for all $t \in I$.

Remark 1.3.2. *(i): Any $\mathbb{R}^3$-valued Gaussian process A is up to versions completely determined by its mean function $\mathbb{E}[A_t]$, $t \in I$, and its covariance function*

$$\Sigma_{s,t} = \mathbb{E}\Big[(A_s - \mathbb{E}[A_s])^T (A_t - \mathbb{E}[A_t])\Big] \ ,$$

as shown for example in Karatzas and Shreve [1988, p.103].

(ii): A general sufficient condition for the existence of a C^1-version can be found in Adler and Taylor [2009, Theorem 1.4.2., p. 23].

The way we define our statistical model on $\mathcal{X}$ is to define a family of stochastic process on $SO(3)$, which, in fact, are constructed from stochastic process on $\mathbb{R}^3$. Therefore recall that the Lie algebra $\mathfrak{so}(3)$ is isomorphic to $\mathbb{R}^3$ by the isomorphism $\iota : \mathbb{R}^3 \to \mathfrak{so}(3)$ explicitly given in (A.2). Moreover, recall that the Lie exponential $\mathrm{Exp} : \mathfrak{so}(3) \to SO(3)$ (see Definition A.3) is surjective. The processes are then constructed as follows.

Definition 1.3.3. *We say that a random curve $\gamma \in \mathcal{X}$ follows a* right Gaussian perturbation (rGP) *of a* center curve *$\gamma_0 \in \mathcal{X}$, if there is a zero-mean Gaussian processes $\{A_t\}_{t\in I}$ with values in $\mathbb{R}^3$ having almost surely continuously differentiable sample paths such that*

$$\gamma(t) = \gamma_0(t) \,\mathrm{Exp}(\iota \circ A_t) \ , \tag{1.8}$$

for all $t \in I$. The Gaussian process $\{A_t\}_{t\in I}$ will be called the generating Gaussian process*.*

Remark 1.3.4. *Note that γ indeed defines an element in $\mathcal{X}$, since the Lie group multiplication and the Lie exponential of $SO(3)$ are analytic.*

These processes have the favorable property that the pointwise action of an element of $\mathcal{S}$ on the sample paths of such a process, will produce sample paths also belonging to a rGP process. Hence the property of being a Gaussian perturbation model is preserved under the group action of $\mathcal{S}$ and therefore these processes will define a $\mathcal{S}$-invariant family as will be shown now.

Theorem 1.3.5 (Gaussian Perturbation Equivariance)**.** *Let $(\psi, \phi) \in \mathcal{S}$ and let the random curve $\gamma \in \mathcal{X}$ follow a rGP of $\gamma_0 \in \mathcal{X}$. Then $(\psi, \phi).\gamma$ follows a rGP of the center curve $(\psi, \phi).\gamma_0$.*

Proof. Let $(\psi, \phi) \in \mathcal{S}$ be arbitrary. Applying Proposition 1.1.7 we have that for any $\psi \in \mathcal{I}_0\big(SO(3)\big)$ there are $P_\psi, Q_\psi \in SO(3)$ with $\psi(R) = P_\psi R Q_\psi^T$ for all $R \in SO(3)$. Hence

$$\begin{aligned}(\psi,\phi).\gamma(t) &= P_\psi \gamma_0\big(\phi(t)\big)\operatorname{Exp}\Big(\iota \circ A_{\phi(t)}\Big) Q_\psi^T \\ &= P_\psi \gamma_0\big(\phi(t)\big) Q_\psi^T \sum_{j=1}^{\infty} \frac{\big(Q_\psi\, \iota\big(A_{\phi(t)}\big)\, Q_\psi^T\big)^j}{j!} \\ &= \big((\psi,\phi).\gamma_0\big)(t)\operatorname{Exp}\Big(\iota\big(Q_\psi A_{\phi(t)}\big)\Big)\end{aligned}$$

is a Gaussian perturbation model with center curve $(\psi, \phi).\gamma_0$ and zero-mean, continuously differentiable Gaussian process $\big\{Q_\psi A_{\phi(t)}\big\}_{t\in I}$. The third equality is due to Proposition A.1.2. □

Corollary 1.3.6. *The measures $\mathbb{P}_\theta$ with $\theta = (\gamma_0, \Sigma)$ on $\mathcal{X}$ induced by random variables γ following rGP models together with the set $\Theta = \mathcal{X} \times \Xi$ form an identifiable $\mathcal{S}$-invariant statistical model on $\mathcal{X}$, where Ξ denotes the set of all covariance functions Σ belonging to an $\mathbb{R}^d$-valued Gaussian process indexed by I and having a $\mathcal{C}^1$-version.*

Moreover, the induced action of $\mathcal{S}$ on the set Θ is given by

$$\mathcal{S} \times \Theta \to \Theta\,, \quad \big((\psi,\phi),(\gamma_0,\Sigma)\big) \mapsto \big((\psi,\phi).\gamma_0, Q_\psi^T \Sigma Q_\psi\big)\ .$$

Alternative Gaussian perturbation models. The non-commutativity of the Lie group $SO(3)$ suggests that one could also consider perturbation models involving a Gaussian process from the left or even from both sides i.e.,

$$\eta(t) = \operatorname{Exp}\big(\iota \circ B_t\big)\, \eta_0(t) \tag{1.9}$$

$$\delta(t) = \operatorname{Exp}\big(\iota \circ C_t\big)\, \delta_0(t) \operatorname{Exp}\big(\iota \circ D_t\big) \tag{1.10}$$

for all $t \in I$ and Gaussian processes $\{B_t\}_{t\in I}$, $\{C_t\}_{t\in I}$ and $\{D_t\}_{t\in I}$ having almost surely continuously differentiable sample paths.

It turns out that (1.8) and (1.9) are equivalent, whereas (1.10) is approximately equivalent to (1.8) and (1.9) up to first order, if the curves are assumed to be concentrated near zero. As our inferential statistics later on will rely on concentration asymptotics as well, these results show that our restriction to right Gaussian perturbations is reasonable and similar methods can be applied to the models (1.9) and (1.10).

Theorem 1.3.7. *Any right Gaussian model* (1.8) *can be rewritten into a left Gaussian model* (1.9) *with the same center curve $\gamma_0 \in \mathcal{X}$ i.e., for any continuously differentiable Gaussian process $\{A_t\}_{t\in I}$ there exists a continuously differentiable Gaussian process $\{B_t\}_{t\in I}$ such that*

$$\gamma_0(t)\operatorname{Exp}\big(\iota \circ A_t\big) = \operatorname{Exp}\big(\iota \circ B_t\big)\gamma_0(t)$$

and vice versa.

Proof. Using Proposition (A.4) (i.e., the naturality of the Lie exponential) and again Proposition A.1.2 we obtain at once

$$\begin{aligned}\gamma(t) &= \gamma_0(t)\mathrm{Exp}\big(\iota \circ A_t\big)\\ &= \gamma_0(t)\mathrm{Exp}\big(\iota \circ A_t\big)\gamma_0(t)^T\gamma_0(t)\\ &= \mathrm{Exp}\Big(\gamma_0(t)\iota(A_t)\gamma_0(t)^T\Big)\gamma_0(t)\\ &= \mathrm{Exp}\Big(\iota\big(\gamma_0(t)A_t\big)\Big)\gamma_0(t)\,.\end{aligned}$$

The claim follows now from the fact that $\{\gamma_0(t)A_t\}_{t\in I}$ is a Gaussian process with almost surely differentiable sample paths, whenever $\{A_t\}_{t\in I}$ is one. The other way around follows by a similar calculation. □

Theorem 1.3.8. *Let $\sigma \to 0$ be a concentration parameter. Consider a both-sided Gaussian perturbation of a center curve δ_0 given by*

$$\delta(t) = \mathrm{Exp}(\iota \circ C_t)\,\delta_0(t)\,\mathrm{Exp}(\iota \circ D_t)$$

with $\max_{t\in I}\|C_t\| = \mathcal{O}_p(\sigma)$ and $\max_{t\in I}\|D_t\| = \mathcal{O}_p(\sigma)$ for all $t \in I$.

Then $\delta(t)$ can be rewritten into a right Gaussian perturbation i.e.,

$$\delta(t) = \delta_0(t)\,\mathrm{Exp}\Big(\iota \circ A_t + \iota \circ \tilde{A}_t\Big) \tag{1.11}$$

with a zero-mean Gaussian process $A_t = \delta_0(t)^T C_t + D_t$ and a suitable zero-mean process $\tilde{A}_t$ satisfying $\max_{t\in I}\big\|\tilde{A}_t\big\| = \mathcal{O}_p\big(\sigma^2\big)$.

Proof. Note that the notation $\max_{t\in I}\|C_t\| = \mathcal{O}_p(\sigma)$ and $\sigma \to 0$ means that for all sequences $\mathbb{N} \ni l \mapsto \sigma_l > 0$ with $\sigma_l \to 0$ as $l \to \infty$ and a sequence of processes C_t^l we have that for for all $\varepsilon > 0$ there exists a $M > 0$ such that

$$\mathbb{P}\Big(\sigma_l^{-2}\max_{t\in I}\big\|C_t^l\big\| > M\Big) < \varepsilon$$

for all $l \in \mathbb{N}$.

Therefore in the following we assume that $\sigma_l \to 0$ and consider a sequence of Gaussian perturbation models with Gaussian processes $\big\{C_t^l\big\}_{t\in I}$ and $\big\{D_t^l\big\}_{t\in I}$ such that $\max_{t\in I}\big\|C_t^l\big\| = \mathcal{O}_p(\sigma_l) = \max_{t\in I}\big\|D_t^l\big\|$.

First observe that the proof of Theorem 1.3.7 shows that

$$\delta_0(t)^T\mathrm{Exp}\Big(\iota \circ C_t^l\Big)\,\delta_0(t)\,\mathrm{Exp}\Big(\iota \circ D_t^l\Big) = \mathrm{Exp}\left(\iota \circ \big(\delta_0(t)^T C_t^l\big)\right)\mathrm{Exp}\Big(\iota \circ D_t^l\Big)\,.$$

Thus, since $\max_{t\in I}\big\|\delta_0(t)^T C_t^l\big\| = \max_{t\in I}\big\|C_t^l\big\| = \mathcal{O}_p(\sigma_l)$, we may w.l.o.g. set $\delta_0(t) = I_{3\times 3}$ for all $t \in I$.

Furthermore, we have $C^l, D^l \xrightarrow{\mathbb{P}} 0$ for $n \to \infty$ with respect to the maximums norm. This can be seen as follows:

By $\max_{t\in I}\big\|C_t^l\big\| = \mathcal{O}_p(\sigma_l)$ we obtain that for any $\epsilon > 0$ there exists a $M > 0$ such that

$$\mathbb{P}\left(\sigma_l^{-1}\max_{t\in I}\big\|C_t^l\big\| > M\right) < \epsilon$$

for all l. Thus, since $\sigma_l \to 0$, for any fixed $\delta > 0$ there is a $L > 0$ such that for all $l > L$, we have that $\sigma_l^{-1}\delta > M$, which implies

$$\mathbb{P}\Big(\max_{t\in I} \big\|C_t^l\big\| > \delta\Big) = \mathbb{P}\Big(\sigma_l^{-1} \max_{t\in I} \big\|C_t^l\big\| > \sigma_l^{-1}\delta\Big) \leq \mathbb{P}\Big(\sigma_l^{-1} \max_{t\in I} \big\|C_t^l\big\| > M\Big) < \epsilon$$

for all $l > L$.

Now, let us define the random process

$$h_t^l = \mathrm{Exp}\big(\iota \circ C_t^l\big)\, \mathrm{Exp}\big(\iota \circ D_t^l\big)\ ,$$

and let us introduce the process

$$\Big\{\tilde{h}_t^l\Big\}_{t\in I} = \Big\{\mathrm{Exp}|_{\mathfrak{V}}^{-1}\big(h_t^l\big)\Big\}_{t\in I}\,, \tag{1.12}$$

where $\mathfrak{V} \subset \mathcal{B}_\pi\big(0, \|\cdot\|_F\big) \subset \mathfrak{so}(3)$ is a set making the Lie exponential bijective. A specific choice of such a set is given in equation (2.21), and the discussion before. If we define the process

$$\tilde{A}_t^l = \iota^{-1} \circ \tilde{h}_t^l - C_t^l - D_t^l\,.$$

Then we obtain by definition

$$\mathrm{Exp}\Big(\iota \circ A_t^l + \iota \circ \tilde{A}_t^l\Big) = \mathrm{Exp}\big(\iota \circ C_t^l\big)\ \mathrm{Exp}\big(\iota \circ D_t^l\big)$$

for all $t \in I$ and $A_t^l = C_t^l + D_t^l$.

Thus, in order to prove (1.11) it suffices to show that for all $\epsilon > 0$ there exist a M such that

$$\mathbb{P}\Big(\sigma_l^{-2} \max_{t\in I} \big\|\tilde{A}_t\big\| > M\Big) < \epsilon\,. \tag{1.13}$$

Now, let $\epsilon > 0$ be given. We introduce the function value

$$f(X, Y) = \mathrm{Log}\big(\mathrm{Exp}(X)\,\mathrm{Exp}(Y)\big)$$

for deterministic vector fields $X, Y \in \mathfrak{so}(3)$, which gives a well-defined analytic function in a neighborhood U of $(0, 0) \in \mathfrak{so}(3) \times \mathfrak{so}(3)$. In particular, Taylor's formula is applicable yielding that

$$f(X, Y) = X + Y + \mathcal{O}\big(\|X\|^2 + \|Y\|^2\big) \tag{1.14}$$

for $(X, Y) \in U$. Moreover, we may choose a compact subset $K \subset U$ containing itself an open subset V on which equation (1.14) is valid.

We now may split the l.h.s. of (1.13) into the two summands

$$\mathbb{P}\Big(\mathfrak{K};\ \sigma_l^{-2} \max_{t\in I} \big\|\iota^{-1} \circ \mathrm{Log}\big(h_t^l\big) - C_t^l - D_t^l\big\| > M\Big)$$

and

$$\mathbb{P}\Big(\mathfrak{K}^C;\ \sigma_l^{-2} \max_{t\in I} \big\|\iota^{-1} \circ \tilde{h}_t^l - C_t^l - D_t^l\big\| > M\Big)\ ,$$

where $\mathfrak{K} = \{\omega \in \Omega \,|\, \forall t \in I : \ \big(\iota \circ C_t^l, \iota \circ D_t^l\big) \in K\}$. Note that we used in the first summand that $\mathrm{Log}\big(h_t^l\big) = \tilde{h}_t^l$ for all $t \in I$, if the complete sample path $t \mapsto (\iota \circ C_t^l, \iota \circ D_t^l)$ is contained in K.

Our goal is now to show that there exist M_1 and M_2 for each of the two summands such that the probabilities are smaller than $\epsilon/2$ for all $l \in \mathbb{N}$, since setting $M = \max(M_1, M_2)$ then proves (1.13).

First consider the probability containing $\mathfrak{K}$. Indeed, we obtain by Taylor's formula (1.14) that

$$\max_{t\in I} \big\|\tilde{A}_t^l\big\| \leq \Lambda \max_{t\in I}\Big(\big\|C_t^l\big\|^2 + \big\|D_t^l\big\|^2\Big) \leq \Lambda \max_{t\in I} \big\|C_t^l\big\|^2 + \Lambda \max_{t\in I} \big\|D_t^l\big\|^2 = \mathcal{O}_p\big(\sigma_l^2\big)\,,$$

where $\Lambda > 0$ can be chosen independent of $t \in I$, since the Hessian of f is bounded on K.

Thus, we have that $\max_{t\in I} \big\|\tilde{A}_t^l\big\| = \mathcal{O}_p\big(\sigma_l^2\big)$ on $\mathfrak{K}$ and therefore we find a $M_1 > 0$ such that for all $l \in \mathbb{N}$

$$\mathbb{P}\bigg(\mathfrak{K};\ \sigma_l^{-2} \max_{t\in I} \big\|\iota^{-1} \circ \mathrm{Log}\big(h_l\big) - C^l - D^l\big\| \geq M_1\bigg) < \frac{\epsilon}{2}\,,$$

proving our claim for this case.

Now, consider the second summand, i.e. there exists $t \in I$ such that $\big(\iota \circ C_t^l, \iota \circ D_t^l\big) \notin K$. By $C^l, D^l \xrightarrow{\mathbb{P}} 0$ with respect to the maximums norm it follows that $\mathbb{P}\big(\mathfrak{K}^C\big) \to 0$, if $l \to \infty$, and hence we find $L \in \mathbb{N}$ such that for all $l > L$

$$\mathbb{P}\Big(\mathfrak{K}^C\Big) < \frac{\epsilon}{2}$$

holds true. This yields that for any $M > 0$ and all $l > L$

$$\mathbb{P}\bigg(\mathfrak{K}^C;\ \sigma_l^{-2} \max_{t\in I} \big\|\iota^{-1} \circ \tilde{h}_t^l - C_t^l - D_t^l\big\| > M\bigg) \leq \mathbb{P}\big(\big(\iota \circ C^l, \iota \circ D^l\big) \in K^C\big) < \frac{\epsilon}{2}\,.$$

However, since any finite collection of random variables is bounded in probability, we also find for $\varepsilon/2$ a $M_2 > 0$ depending on L such that for all $l \leq L$

$$\mathbb{P}\bigg(\mathfrak{K}^C;\ \sigma_l^{-2} \max_{t\in I} \big\|\iota^{-1} \circ \tilde{h}_t^l - C_t^l - D_t^l\big\| > M_2\bigg) < \frac{\epsilon}{2}\,,$$

which shows the claim in this case and furthermore finishes the proof. □

Remark 1.3.9. *Note that any right Gaussian perturbation* (1.8) *or any left Gaussian perturbation* (1.9) *is also a both sided Gaussian perturbation* (1.10)*, since the deterministic process $C_t = 0$ or $D_t = 0$ for all $t \in I$ is a Gaussian process by definition.*

Chapter 2

Simultaneous Confidence Sets for Center Curves

In this chapter we will deal with the issue of estimating the center curve in Gaussian perturbation models and establish approximate simultaneous confidence regions for it. This is based on a generalization of the article Rancourt et al. [2000] who himself builds on Downs [1972] and showed in the non-functional case that for concentrated data certain extrinsic residuals are approximately Gaussian. In our approach we use more natural intrinsic residuals and prove similar concentration properties for right Gaussian perturbation models. Applying this result we then obtain approximative simultaneous confidence bands using the Gaussian kinematic formula (see Taylor [2006]).

2.1 Estimation of Center Curve and Residuals

Usually, given any probability distribution a first interesting descriptor is its *mean* or *expectation*. Unfortunately, the commonly used definition of the mean of an $\mathbb{R}^D$-valued random variable X, $D \in \mathbb{N}$, on a measure space $(\Omega, \mathbb{P})$

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P}$$

does rely on the linearity of $\mathbb{R}^D$ and can therefore not be generalized directly to random variables taking values in a non-linear manifold. However, on $\mathbb{R}^D$ the expectation, if it is finite, has also the property of being the unique minimizer of the functional

$$F_X(\mu) = \mathbb{E}\|X - \mu\|^2 = \mathbb{E}\left[d(X, \mu)^2\right]$$

and, if this is used as the definition of the mean, it can be generalized to arbitrary manifolds, since F_X does only rely on the existence of a distance d between two points.

Indeed, for any Riemannian manifold $\mathcal{M}$ isometrically embedded into $\mathbb{R}^D$ for some $D \in \mathbb{N}$ there are at least two canonical metrics: the intrinsic metric and the extrinsic metric (see Bhattacharya and Patrangenaru [2003]). Recall, however, that the resulting generalized means, called *Fréchet mean* and *extrinsic mean* are,

if existent, possibly not unique in contrary to the mean of an $\mathbb{R}^D$-valued random variable. A lot of research on properties (e.g., uniqueness results and central limit theorems) and applications of these means was done in the last decades among others by Ziezold [1977], Hendriks and Landsman [1998], Bhattacharya and Patrangenaru [2003], Bhattacharya and Patrangenaru [2005], Afsari [2011], Kendall et al. [2011], Huckemann et al. [2011] and Hotz and Huckemann [2015].

Interestingly, it turns out (see Theorem 2.1.7) that for right Gaussian perturbation models, which we introduced in Section 1.3, an extrinsic approach is more viable than the intrinsic approach, since it is easier to compute and will give a consistent estimator of the center curve of a right Gaussian perturbation model.

Definition 2.1.1. *For a random path γ in $\mathcal{X}$ and any session $\chi = (\gamma_1, ..., \gamma_N) \in \mathbb{S}(\mathcal{X})$ we have the non-empty (due to compactness of $SO(3)$) sets of population and sample minimizers at $t \in I = [0,1]$ i.e.,*

$$\mu(\gamma, t) = \operatorname*{argmin}_{\mu \in SO(3)} \mathbb{E}\big[\|\mu - \gamma(t)\|_F^2\big]\,, \qquad \hat{\mu}_N\big(\chi, t\big) = \operatorname*{argmin}_{\mu \in SO(3)} \frac{1}{N} \sum_{n=1}^{N} \|\mu - \gamma_n(t)\|_F^2\,.$$

Here $\|\cdot\|_F$ denotes the Frobenius norm on $\mathbb{R}^{3\times 3}$ (see Definition A.1). The (possibly) set valued map $t \mapsto \mu(\gamma, t)$ is called the pointwise extrinsic population mean (PEM) *of γ and the set valued map $t \mapsto \hat{\mu}_N(\chi, t)$ is called* pointwise extrinsic sample mean (PESM) *of χ.*

Indeed, as one might suspect from the invariance of the Frobenius norm under $\mathcal{I}_0\big(SO(3)\big)$, we have that the PEM and PESM are equivariant statistics (see Definition 1.2.5).

Theorem 2.1.2 (Equivariance of PEM and PESM)**.** *The PEM and PESM are equivariant statistics on $\mathcal{X}$, on $\mathbb{S}(\mathcal{X})$ respectively, with respect to the group actions of $\mathcal{S}$. More precisely,*

$$\mu\big((\psi,\phi).\gamma, t\big) = (\psi,\phi).\mu(\gamma, t) \quad \text{and} \quad \hat{\mu}\big((\psi,\phi).\chi, t\big) = (\psi,\phi).\hat{\mu}\big(\chi, t\big)\,,$$

for any $(\psi,\phi) \in \mathcal{S}$, random $\gamma \in \mathcal{X}$ and $\chi \in \mathbb{S}(\mathcal{X})$.

Proof. The result for the PESM is the special case of the PEM for a sum of point measures. Hence, the invariance of the Frobenius norm under $\mathcal{I}_0\big(SO(3)\big)$ yields

$$\begin{aligned}
\mu\big((\psi,\phi).\gamma, t\big) &= \operatorname*{argmin}_{\mu \in SO(3)} \mathbb{E}\Big[\big\|\mu - \psi.\gamma\big(\phi(t)\big)\big\|_F^2\Big] \\
&= \operatorname*{argmin}_{\mu \in SO(3)} \mathbb{E}\Big[\big\|\mu - Q_\psi \gamma\big(\phi(t)\big) P_\psi^T\big\|_F^2\Big] \\
&= \operatorname*{argmin}_{\mu \in SO(3)} \mathbb{E}\Big[\big\|Q_\psi^T \mu P_\psi - \gamma\big(\phi(t)\big)\big\|_F^2\Big] \\
&= \operatorname*{argmin}_{\mu \in SO(3):\, \tilde{\mu} = Q_\psi^T \mu P_\psi} \mathbb{E}\Big[\big\|\tilde{\mu} - \gamma\big(\phi(t)\big)\big\|_F^2\Big] \\
&= Q_\psi \operatorname*{argmin}_{\tilde{\mu} \in SO(3)} \mathbb{E}\Big[\big\|\tilde{\mu} - \gamma\big(\phi(t)\big)\big\|_F^2\Big]\, P_\psi^T \\
&= Q_\psi \mu\big(\phi(t)\big) P_\psi^T = (\psi,\phi).\mu(\gamma, t)\,.
\end{aligned}$$

□

In data applications or theoretical considerations we need more explicit representations of the PEM and the PESM. Luckily, for both there exists an elegant way to compute them based on the singular value decomposition (SVD) of a certain matrix. This is accomplished by showing that the minimization problems of the PEM and the PESM can be transformed into a well studied minimization problem known from rigid body motions in crystallography (see Mackenzie [1957], Stephens [1979]) spherical regression (see Chang [1986]) or Procrustes analysis of shape (see Dryden and Mardia [1998]). In order to restate this result tailored to the PEM and PESM, we introduce the matrix

$$\bar{\gamma}_N(t) = N^{-1} \sum_{n=1}^{N} \gamma_n(t) \in \mathbb{R}^{3\times 3}\,, \tag{2.1}$$

which we call the *pointwise Euclidean sample mean* of a sample $\gamma_1, ..., \gamma_N$, $N \in \mathbb{N}$.

Theorem 2.1.3. *Let γ be a random element of $\mathcal{X}$ and $(\gamma_1, ..., \gamma_N) \in \mathbb{S}(\mathcal{X})$. Fix $t \in I$. Then any element of the PEM (PESM, respectively) at time t is given by USV^T, where $U, V \in O(3)$ are obtained from a singular value decomposition UDV^T of $\mathbb{E}[\gamma(t)]$ ($\bar{\gamma}_N(t)$, respectively) with $D = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and*

$$S = \begin{cases} I_{4\times 4} & \text{if } \det(U)\det(V) = 1 \\ \operatorname{diag}(1, 1, -1) & \text{if } \det(U)\det(V) = -1 \end{cases}. \tag{2.2}$$

The PEM (PESM, respectively) at $t \in I$ is unique, if and only if $\operatorname{rank}\big(\mathbb{E}[\gamma(t)]\big) > 1$ ($\operatorname{rank}\big(\bar{\gamma}_N(t)\big) > 1$, respectively).

Proof. By expanding the square one obtains

$$\mu(\gamma, t) = \operatorname*{argmin}_{\mu\in SO(3)} \mathbb{E}\left[\|\mu - \gamma(t)\|_F^2\right] = \operatorname*{argmax}_{\mu\in SO(3)} \operatorname{tr}\Big(\mathbb{E}[\gamma(\mathrm{t})]\,\mu^{\mathrm{T}}\Big) = \operatorname*{argmin}_{\mu\in \mathrm{SO(3)}} \|\mu - \mathbb{E}[\gamma(\mathrm{t})]\|_{\mathrm{F}}^2\,.$$

Now, the results follow from Umeyama [1991, Lemma, p. 377], if we set $n = m$ and $B = I_{3\times 3}$ in their result. The proof for the PESM is the special case of a sum of point measures. □

Another characterization of the PEM and PESM can be deduced from Bhattacharya and Bhattacharya [2012]. In order to obtain it, we introduce the orthogonal projections $\mathfrak{pr} : \mathbb{R}^{3\times 3} \to SO(3)$ with respect to the Frobenius norm i.e.,

$$\mathfrak{pr}(M) = \operatorname*{argmin}_{Q\in SO(3)} \|Q - M\|_F^2\,, \tag{2.3}$$

for $M \in \mathbb{R}^{3\times 3}$, which possibly is multivalued. A pointwise application of Bhattacharya and Bhattacharya [2012, Proposition 4.2.a), p. 37] yields the following.

Theorem 2.1.4. *Let γ be a random element of $\mathcal{X}$ and $\chi = (\gamma_1, ..., \gamma_N) \in \mathbb{S}(\mathcal{X})$. Then for each $t \in I$ we have that the PEM of γ and the PESM of χ are given by*

$$\mu(\gamma, t) = \mathfrak{pr}\Big(\mathbb{E}\big[\gamma(t)\big]\Big) \quad \text{and} \quad \hat{\mu}\big(\chi, t\big) = \mathfrak{pr}\left(\frac{1}{N}\sum_{n=1}^{N} \gamma_n(t)\right) \tag{2.4}$$

This result combined with Theorem 2.1.3 can be used to establish that the PESM of $\gamma_1, ..., \gamma_N$ is itself an element belonging to $\mathcal{X}$, if the PESM is unique for all $t \in I$. Note that this is often the case in applications, since the data tends to be concentrated.

Corollary 2.1.5. *Let $\chi = (\gamma_1, ..., \gamma_N) \in \mathbb{S}(\mathcal{X})$. Assume that for all $t \in I$ we have that $\hat{\mu}_N\big(\chi, t\big)$ is unique. Then the PESM $t \mapsto \hat{\mu}_N\big(\chi, t\big)$ belongs to $\mathcal{X}$.*

Proof. By Theorem 2.1.4 we have that the PESM evaluated at $t \in I$ is given by

$$\mathfrak{pr}\big(\bar{\gamma}_N(t)\big) = \mathfrak{pr}\left(\frac{1}{N}\sum_{n=1}^{N}\gamma_n(t)\right),$$

for all $t \in I$. Since the embedding $SO(3) \hookrightarrow \mathbb{R}^{3\times 3}$ is smooth, we have that $\bar{\gamma}_N \in \mathcal{C}^1\big(I, \mathbb{R}^{3\times 3}\big)$. It remains to show that $\mathfrak{pr}$ is differentiable in a neighborhood of the image $\Gamma = \{B \in \mathbb{R}^{3\times 3} \mid \exists t \in I \colon B = \bar{\gamma}_N(t)\}$.

By Theorem 2.1.3 the PESM is unique for all $t \in I$ if and only if $\Gamma \cap \mathcal{F} = \emptyset$, where $\mathcal{F} = \{B \in \mathbb{R}^{3\times 3} \mid \operatorname{rank}(B) \leq 1\}$. Since Γ is compact and $\mathcal{F}$ is closed (see Lemma B.0.3) there exists an open neighborhood $U \supset \Gamma$ with $U \cap \mathcal{F} = \emptyset$. By Dudek and Holly [1994, Theorem 4.1, p. 6] we obtain that $\mathfrak{pr}$ is analytic on U and hence $t \mapsto \mathfrak{pr}\big(\bar{\gamma}_N(t)\big) \in \mathcal{C}^1\big(I, \mathbb{R}^{3\times 3}\big)$ as claimed. □

Remark 2.1.6. *The same proposition for the PEM is false in general. This is due to the fact that $t \mapsto \mathbb{E}[\gamma(t)]$ is not necessarily a $\mathcal{C}^1$-curve in $\mathbb{R}^{3\times 3}$. In order to achieve this one additionally needs to assume that for each $t \in I$ we have that $|\gamma'(t)| \leq Z(t)$ with $Z(t)$ an integrable random variable. Then the Lebesgue's dominated convergence theorem (see Loève [1955, Theorem C, p. 125]) ensures that $t \mapsto \mathbb{E}[\gamma(t)] \in \mathcal{C}^1\big(I, \mathbb{R}^{3\times 3}\big)$.*

We now want to study the PEM of Gausssian perturbation models. Interestingly, we can show that the PEM is identical to the center curve. Hence the PEM is a good descriptor of our model. Moreover, this will imply that the PESM is a consistent estimator of the center curve. Hence we can use it for inference on the center curve (see Section 2.2).

Theorem 2.1.7. *If a random curve $\gamma \in \mathcal{X}$ follows a right Gaussian perturbation of a center curve $\gamma_0 \in \mathcal{X}$ (see Definition 1.3.3), then its PEM is unique at any $t \in I$ and is identical to the center curve γ_0.*

Proof. Let $\mathbb{E}[\gamma(t)] = UDV^T$ be a singular value decomposition for $t \in I$, then with Theorem 2.1.3 the assertion is the claim that $\gamma_0(t) = USV^T$. Since $\mathbb{E}[\gamma(t)] = \gamma_0(t)\mathbb{E}\big[\operatorname{Exp}\big(\iota \circ A_t\big)\big]$, this claim holds if $\mathbb{E}\big[\operatorname{Exp}\big(\iota \circ A_t\big)\big]$ is symmetric and positive definite.

Indeed, let $\mathbb{E}\big[\operatorname{Exp}\big(\iota \circ A_t\big)\big] = \tilde{V}\Lambda\tilde{V}^T$ with $\tilde{V} \in SO(3)$ and diagonal non-degenerate Λ. Then $\Lambda = D$ (we assume that the eigenvalues are sorted from the highest to the lowest), since $\mathbb{E}[\gamma(t)]^T\mathbb{E}[\gamma(t)] = \mathbb{E}\big[\operatorname{Exp}\big(\iota \circ A_t\big)\big]^T\mathbb{E}\big[\operatorname{Exp}\big(\iota \circ A_t\big)\big]$ giving

$$UDV^T = \gamma_0(t)\tilde{V}D\tilde{V}^T.$$

Moreover, since the eigenspaces corresponding to same eigenvalues spanned by the columns of $\tilde{V}$ and V agree we can choose $\tilde{V} = V$, yielding $\gamma_0 = UV^T$ with $U \in SO(3)$, since $U = \gamma_0(t)\tilde{V} \in SO(3)$, which would prove the claim.

By Lemma B.0.2 it remains to prove that $A_t = X \sim \mathcal{N}(0, \Sigma)$ in $\mathbb{R}^3$ fulfills $\mathbb{E}[\cos \|X\|] \geq 0$. Indeed, making use of Fourier transformations of Gaussian densities, say

$$\begin{aligned}
\mathbb{E}[\cos \|X\|] &= \frac{1}{(2\pi)^{k/2}\sqrt{\nu_1 \cdot \ldots \cdot \nu_k}} \int_{\mathbb{R}^k} e^{-x^T \widetilde{\nu}^{-1} x/2} \cos(\|x\|)\, dx \\
&= \frac{1}{(2\pi)^{k/2}\sqrt{\nu_1 \cdot \ldots \cdot \nu_k}} \int_{\mathbb{R}^k} e^{-y^T \widetilde{\nu}^{-1} y/2} \cos(\|y\|)\, dy \\
&= \frac{1}{(2\pi)^{k/2}} \int_{S^{k-1}} \left(\int_0^\infty e^{-r^2/2} \cos\left(r\sqrt{\phi^T \widetilde{\nu} \phi} \right) dr \right) d\sigma(\phi) \\
&= \frac{1}{(2\pi)^{k/2}} \int_{S^{k-1}} \left(\frac{1}{2} \int_{-\infty}^\infty e^{-r^2/2} e^{ir\sqrt{\phi^T \widetilde{\nu} \phi}}\, dr \right) d\sigma(\phi) \\
&= \frac{1}{(2\pi)^{k/2-1}} \int_{S^{k-1}} \frac{1}{2}\, e^{-\phi^T \widetilde{\nu} \phi}\, d\sigma(\phi) \;>\; 0
\end{aligned}$$

with the spherical volume element $d\sigma(\phi)$ on the $k-1$ dimensional unit sphere S^{k-1}. Here we have used a svd $\Sigma = W \operatorname{diag}(\nu_1, \nu_2, \nu_3) W^T$ with $W = (w_1, w_2, w_3) \in SO(3)$, the smallest index $k \in \{1, 2, 3\}$ such that $\nu_k > 0$, $\widetilde{\nu} = \operatorname{diag}(\nu_1, ..., \nu_k)$ and $y = (w_1, ..., w_k)^T x \in \mathbb{R}^k$. □

Remark 2.1.8. *Let $\gamma(t) = \gamma_0(t)\operatorname{Exp}\big(\iota(A_t)\big)$ be any stochastic process on $SO(3)$ induced by a process $\{A_t\}_{t\in I}$ in $\mathbb{R}^3$. Then note that the only properties of the distribution of $\{A_t\}_{t\in I}$ used in the proofs are*

$$\mathbb{E}[A_t \operatorname{sinc}\|A_t\|] = 0 \quad \textit{and} \quad \mathbb{E}[\cos \|A_t\|] > 0\,. \tag{2.5}$$

Thus, any such process satisfying these conditions admits a unique PEM given by $\gamma_0(t)$.

Corollary 2.1.9. *Suppose that for each $N \in \mathbb{N}$ sessions $\chi_N \in \mathcal{X}^N \subset \mathbb{S}(\mathcal{X})$ (see Definition 1.2.10) of a rGP with center curve γ_0 are given. Fix $t \in I$ and choose a measurable selection $\hat{\gamma}_N(t) \in \hat{\mu}\big(\chi_N, t\big)$ of the PESM for each $N \in \mathbb{N}$. Then we have that $\hat{\gamma}_N(t) \to \gamma_0(t)$ almost surely.*

Proof. That the extrinsic sample mean set is a strongly consistent estimator of the extrinsic population mean set follows from a more general result by Ziezold [1977]. In case of uniqueness, guaranteed by the model we have that the center curve and PEM agree by virtue of Theorem 2.1.7. Hence, every measurable selection of the sample mean converges almost surely to the unique population mean yielding the assertion (see also Bhattacharya and Patrangenaru [2003]). □

Finally, we will prove that for rGP models we also have almost surely uniform convergence of the PESM to the center of perturbation, if an additional condition is fulfilled. Moreover, this will imply that the PESM of a session $\chi_N \in \mathbb{S}(\mathcal{X})$ consisting of i.i.d. realizations of an rGP model is with high probability a continuously differentiable curve, if N is large enough, and therefore techniques we will develop later for curves in $\mathcal{X}$ can be applied also to the PESM.

Theorem 2.1.10. *Let $\chi_N \in \mathcal{X}^N \subset \mathbb{S}(\mathcal{X})$, $N \in \mathbb{N}$, denote a session of an rGP γ with center curve γ_0. Moreover, assume that the Gaussian process $\{A_t\}_{t\in I}$ of the rGP satisfies that*

$$\mathbb{E}\left[\max_{t\in I} \|\partial_t A_t\|\right] < \infty\,. \tag{2.6}$$

Then we obtain that

(i) The map $t \mapsto \hat{\mu}\big(\chi_N, t\big) \in \mathcal{X}$ asymptotically almost surely i.e., there exists $\Omega' \subset \Omega$ with $\mathbb{P}(\Omega') = 1$ and for every $\omega \in \Omega'$ there exists N_ω such that $\hat{\mu}\big(\chi_N(\omega), t\big)$ is unique and $t \mapsto \hat{\mu}\big(\chi_N(\omega), t\big)$ is continuously differentiable curve in $SO(3)$ for all $N > N_\omega$.

(ii) $\max_{t\in I} \big\|\hat{\mu}\big(\chi_N, t\big) - \gamma_0\big\|_F \to 0$ for $N \to \infty$ almost surely.

(iii) $\max_{t\in I} d_{SO(3)}\big(\hat{\mu}\big(\chi_N, t\big), \gamma_0\big) \to 0$ for $N \to \infty$ almost surely.

Note, that in (ii) and (iii) as long as $\hat{\mu}\big(\chi_N, t\big)$ is not unique the distances denote distances between sets.

Proof. (i): As in the proof of Corollary 2.1.5 we have that the PESM is given by $\mathfrak{pr}\big(\bar{\gamma}_N(t)\big)$. Hence the PESM at $t \in I$ is unique if and only if the orthogonal projection of $\bar{\gamma}(t)$ does not belong to $\mathcal{F} = \{B \in \mathbb{R}^{3\times 3} \mid \operatorname{rank}(B) \leq 1\}$ (see Theorem 2.1.3). Hence, in order to prove (i) it suffices to show that

$$\max_{t\in I} \|\bar{\gamma}_N(t) - \mathbb{E}\left[\gamma(t)\right]\|_F \to 0 \quad a.s. \tag{2.7}$$

This is deduced as follows: by Theorem 2.1.7 the PEM of γ is unique and thus $\mathbb{E}\left[\gamma(t)\right] \notin \mathcal{F}$. Moreover, due to the fact that $\mathcal{F}$ is closed (see Lemma B.0.3) and the image of $t \mapsto \mathbb{E}\left[\gamma(t)\right]$ for $t \in I$ is compact there exists $\varepsilon_0 > 0$ such that

$$\big\|\mathbb{E}\left[\gamma(t)\right] - \mathcal{F}\big\|_F \geq \varepsilon_0\,,$$

for all $t \in I$. Now, if (2.7) is true, there exists a set $\tilde{\Omega}$ with $\mathbb{P}\big(\tilde{\Omega}\big) = 1$ such that there exists for all $\omega \in \tilde{\Omega}$ and $\varepsilon_0/2$ a natural number $N_\omega(\varepsilon_0) \in \mathbb{N}$ such that for all $N \geq N_\omega(\varepsilon_0)$ we have that

$$\big\|\bar{\gamma}_N(\omega, t) - \mathbb{E}\left[\gamma(t)\right]\big\|_F < \frac{\varepsilon_0}{2}$$

for all $t \in I$. This implies that $\bar{\gamma}_N(t) \notin \mathcal{F}$ for all $t \in I$, since

$$\big\|\bar{\gamma}_N(t) - \mathcal{F}\big\|_F \geq \big\|\mathbb{E}\left[\gamma(t)\right] - \mathcal{F}\big\|_F - \big\|\bar{\gamma}_N(t) - \mathbb{E}\left[\gamma(t)\right]\big\|_F > \frac{\varepsilon_0}{2}$$

for all $t \in I$ and all $N > N_\omega(\varepsilon_0)$. Therefore, $\hat{\mu}\big(\chi_N(\omega), t\big)$ is unique for all $N > N_\omega(\varepsilon_0)$ and the curve $t \mapsto \hat{\mu}\big(\chi_N(\omega), t\big)$ is $\mathcal{C}^1$, since $\bar{\gamma}_N(\omega, t)$ is $\mathcal{C}^1$ and $\mathfrak{pr} : \mathbb{R}^{3\times 3}\backslash\mathcal{F} \to SO(3)$ is analytic (see Dudek and Holly [1994, Theorem 4.1, p. 6]).

Thus, it remains to prove the uniform convergence (2.7). By Theorem B.0.9 it suffices to show that the sequence of processes $\bar{\gamma}_N(t)$ is stochastically equicontinuous, since we already have pointwise convergence by Corollary 2.1.9.

In order to establish this, recall that $\gamma_n(t) = \gamma_0(t)\mathrm{Exp}\big(\iota \circ A_t^n(\omega)\big)$ with a Gaussian process A^n with the same distribution as A (see Definition 1.3.3). Define Ω' with $\mathbb{P}(\Omega') = 1$ by $\Omega' = \bigcap_{n\in\{1,\dots,N\}} \Omega_n$, where Ω_n is the set for which $A^n(\omega) \in \mathcal{C}^1(I, \mathbb{R}^3)$. The triangle inequality then yields

$$\|\bar\gamma_N(\omega,t) - \bar\gamma_N(\omega,t')\|_F \le \frac{1}{N}\sum_{n=1}^{N} \big\|\gamma_0(s)\mathrm{Exp}\big(\iota\circ A_s^n(\omega)\big) - \gamma_0(t)\mathrm{Exp}\big(\iota\circ A_t^n(\omega)\big)\big\|_F \tag{2.8}$$

for all $\omega \in \Omega'$. Using that $\|R\|_F \le \sqrt{3}$ for all $R \in SO(3)$ and that γ_0 and $A^n(\omega)$ are Lipschitz continuous with Lipschitz constants $L_\gamma = \max_{t\in I}\|\gamma_0'(t)\|_F$ and $L_{A^n}(\omega) = \max_{t\in I}\|\partial_t A_t^n(\omega)\|$, we obtain for all $\omega \in \Omega'$

$$\begin{aligned}
&\big\|\gamma_0(s)\mathrm{Exp}\big(\iota\circ A_s^n(\omega)\big) - \gamma_0(t)\mathrm{Exp}\big(\iota\circ A_t^n(\omega)\big)\big\|_F \\
&\qquad\le \big\|\gamma_0(s)-\gamma_0(t)\big\|_F \big\|\mathrm{Exp}\big(\iota\circ A_s^n(\omega)\big)\big\|_F \\
&\qquad\quad + \big\|\gamma_0(t)\big\|_F \big\|\mathrm{Exp}\big(\iota\circ A_s^n(\omega)\big) - \mathrm{Exp}\big(\iota\circ A_t^n(\omega)\big)\big\|_F \\
&\qquad\le \sqrt{3}L_\gamma|s-t| + \sqrt{3}\big\|\mathrm{Exp}\big(\iota\circ A_s^n(\omega)\big) - \mathrm{Exp}\big(\iota\circ A_t^n(\omega)\big)\big\|_F \\
&\qquad\le M\big(1 + L_{A^n}(\omega)\big)|s-t|\,,
\end{aligned} \tag{2.9}$$

where $M > 0$ sufficiently large. Here the last inequality is due to Lipschitz continuity of $\mathrm{Exp} : \big(\mathfrak{so}(3), \|\cdot\|_F\big) \to \big(SO(3), \|\cdot\|_F\big)$ (see Lemma B.0.4) and Lipschitz continuity of $A^n(\omega)$. Putting (2.8) and (2.9) together yields

$$\|\bar\gamma_N(\omega,t) - \bar\gamma_N(\omega,t')\|_F \le M\left(1 + \frac{1}{N}\sum_{n=1}^{N} L_{A^n}(\omega)\right)|s-t| \tag{2.10}$$

for all $\omega \in \Omega'$. We still have to remove the dependency on ω of the right hand side. Therefore note that by Assumption (2.6) and the strong law of large numbers there exists $\Omega'' \subset \Omega$ with $\mathbb{P}(\Omega'') = 1$ such that

$$\frac{1}{N}\sum_{n=1}^{N} L_{A^n}(\omega) \to \mathbb{E}\left[\max_{t\in I}\|\partial_t A_t\|_F\right] = L < \infty \quad n\to\infty$$

for all $\omega \in \Omega''$. Finally, define the new set $\tilde\Omega = \Omega' \cap \Omega''$ still satisfying $\mathbb{P}\big(\tilde\Omega\big) = 1$. Now, the stochastical equicontinuity follows, since for any $\varepsilon > 0$ and any $\omega \in \tilde\Omega$ let $\delta = \frac{\varepsilon}{M(1+L+\epsilon')}$ with an arbitrary $\epsilon' > 0$, we obtain by the SLLN a $N_{\omega,\epsilon'}$ such that

$$\left|\frac{1}{N}\sum_{n=1}^{N} L_{A^n}(\omega) - \mathbb{E}\left[\max_{t\in I}\|\partial_t A_t^n\|_F\right]\right| < \epsilon'$$

for all $N > N_{\omega,\epsilon'}$. Indeed, combining this with equation (2.10) yields

$$\sup_{|s-t|\le\delta} \|\bar\gamma_N(\omega,t) - \bar\gamma_N(\omega,t')\|_F \le M\,(1 + L + \epsilon')\,\delta < \varepsilon\,.$$

for all $N > N_{\omega,\epsilon'}$, which proves the stochastical equicontinuity on $\tilde\Omega$, and therefore finishes the proof of this part.

(ii) Since $\mathbb{E}[\gamma(t)]$ is unique for all $t \in I$ as deduced in the proof of Theorem 2.1.7, we define analogously to Corollary 2.1.5 the compact set $\Gamma = \{B \in \mathbb{R}^{3\times 3} \mid \exists t \in I : B = \mathbb{E}[\gamma(t)]\}$. Then $\Gamma \cap \mathcal{F} = \emptyset$, where $\mathcal{F} = \{B \in \mathbb{R}^{3\times 3} \mid \operatorname{rank}(B) \leq 1\}$ and there exists $\varepsilon_0 > 0$ such that

$$\big\|\mathbb{E}\big[\gamma(t)\big] - \mathcal{F}\big\|_F \geq \varepsilon_0$$

for all $t \in I$. Define the compact set

$$K = \bigcup_{t\in I} \Big\{ B \in \mathbb{R}^{3\times 3} \;\big|\; \|\mathbb{E}[\gamma(t)] - B\| < \tfrac{\varepsilon_0}{2} \Big\},$$

which satisfies $K \cap \mathcal{F} = \emptyset$. Again by Dudek and Holly [1994, Theorem 4.1, p. 6] we obtain that the restriction of $\mathfrak{pr} : \big(\mathbb{R}^{3\times 3}, \|\cdot\|_F\big) \to (SO(3), \|\cdot\|_F)$ to K is analytic and hence Lipschitz continuous.

Since we proved in (i) that $\bar{\gamma}_N$ converges almost surely uniformly to $\mathbb{E}[\gamma]$, there exists N_K such that $\bar{\gamma}_N \in K$ almost surely for all $N > N_K$. Thus, by the Lipschitz continuity of $\mathfrak{pr}$ on K we obtain

$$\big\|\hat{\mu}\big(\chi_N, t\big) - \gamma_0(t)\big\|_F = \big\|\mathfrak{pr}\big(\bar{\gamma}_N(t)\big) - \mathfrak{pr}\big(\mathbb{E}[\gamma(t)]\big)\big\|_F \leq C_K \big\|\bar{\gamma}_N - \mathbb{E}[\gamma(t)]\big\|_F \quad (2.11)$$

for $C_K > 0$ sufficiently large and all $N > N_K$. Now, the almost sure uniform convergence $\bar{\gamma}_N \to \mathbb{E}[\gamma]$ for $N \to \infty$ implies the claimed almost sure uniform convergence of the left hand side of (2.11) to zero.

(iii) This follows from (ii) together with Proposition A.1.3. □

Corollary 2.1.11. *With the notations and assumptions of Theorem 2.1.10 we obtain*

$$\lim_{N\to\infty} \mathbb{P}\big(t \mapsto \hat{\mu}\big(\chi_N, t\big) \in \mathcal{X}\big) = 1\,.$$

Proof. By the arguments and notations of the proof of Theorem 2.1.10 we have that $t \mapsto \hat{\mu}\big(\chi_N, t\big) \in \mathcal{X}$ if for any $\varepsilon > 0$

$$\max_{t\in I} \|\bar{\gamma}_N(t) - \mathcal{F}\|_F > \varepsilon\,.$$

Choose ε_0 as in the proof of Theorem 2.1.10. Then for any $\varepsilon_0 > \varepsilon > 0$ the Markov inequality yields

$$\begin{aligned} \mathbb{P}\left(\max_{t\in I} \|\bar{\gamma}_N(t) - \mathcal{F}\|_F > \varepsilon\right) &\geq \mathbb{P}\left(\max_{t\in I} \|\bar{\gamma}_N(t) - \mathbb{E}[\gamma(t)]\|_F < \varepsilon_0 - \varepsilon\right) \\ &\geq 1 - \frac{\mathbb{E}\big[\max_{t\in I} \|\bar{\gamma}_N(t) - \mathbb{E}[\gamma(t)]\|_F\big]}{\varepsilon_0 - \varepsilon}\,. \end{aligned} \quad (2.12)$$

In the prove of Theorem 2.1.10 we showed that $\max_{t\in I} \|\bar{\gamma}_N(t) - \mathbb{E}[\gamma(t)]\|_F\big]$ converges to zero almost surely. Moreover, the triangle inequality yields $\|\bar{\gamma}_N(t) - \mathbb{E}[\gamma(t)]\|_F \leq 2\sqrt{3}$ for all $t \in I$. Thus, by Lebesgue's dominated convergence theorem (see Loève [1955, Theorem C, p. 125]) we obtain

$$\mathbb{E}\Big[\max_{t\in I} \|\bar{\gamma}_N(t) - \mathbb{E}[\gamma(t)]\|_F\Big] \to 0\,, \quad \text{for } n \to \infty\,,$$

which together with inequality (2.12) yields the claim. □

In order to show that there exists a mild sufficient condition on the derivative of a Gaussian processes such that it satisfies assumption (2.7) of the previous theorem, we need to introduce some new notations. Here we closely follow Chapter 1.3 of the book Adler and Taylor [2009], since we want to apply their Theorem 1.3.3. (p. 14), which is a well known result on bounding the expectation of the maximum of a Gaussian process.

Definition 2.1.12. *Let $\{A_t\}_{t\in I}$ be an $\mathbb{R}$-valued Gaussian process with $\mathbb{E}\big[A_t\big] = 0$ for all $t \in I$. The* canonical (pseudo-) metric *of I with respect to $\{A_t\}_{t\in I}$ is given by*

$$d_A(s,t) = \sqrt{\mathbb{E}\big[(A_s - A_t)^2\big]}$$

for all $s, t \in I$.

Definition 2.1.13. *Let $\{A_t\}_{t\in I}$ be an $\mathbb{R}$-valued Gaussian process with $\mathbb{E}\big[A_t\big] = 0$ for all $t \in I$ and assume that I is compact with respect to the topology induced by d_A. We define*

$$\mathcal{B}_\varepsilon\big(d_A, t\big) = \Big\{ s \in I \;\Big|\; d_A(s,t) \leq \varepsilon \Big\}$$

the centered closed d_A-ball with radius ε. Additionally, we denote with $N(I, d_A, \varepsilon)$ the smallest number of closed d_A-balls with radius ε necessary to cover I. The functions $\varepsilon \mapsto N(I, d_A, \varepsilon)$ and $\varepsilon \mapsto H\big(d_A, \varepsilon\big) = \log\big(N(I, d_A, \varepsilon)\big)$ are called the (metric) entropy function *and* log-entropy function*, respectively.*

Using the log-entropy function we can state the following bound on the expectation of the maximum of a centered Gaussian process.

Theorem 2.1.14. *Let $\{A_t\}_{t\in I}$ be an $\mathbb{R}$-valued Gaussian process with $\mathbb{E}\big[A_t\big] = 0$ for all $t \in I$ and assume that I is compact with respect to the topology induced by d_A. Then there exists an universal constant $L > 0$ such that*

$$\mathbb{E}\left[\max_{t\in I} |A_t|\right] \leq L \int_0^{\mathrm{diam(I)}/2} \sqrt{H\big(d_A, \varepsilon\big)}\, d\varepsilon\,. \tag{2.13}$$

Proof. The proof is given in Adler and Taylor [2009, p.16]. □

Remark 2.1.15. *Note that we replaced the supremum in Adler and Taylor [2009, Theorem 1.3.3.,p. 14] with a maximum. This is possible due the assumption that $\{A_t\}_{t\in I}$ has continuous sample paths and I is compact. They use the more general infimum, since they want to prove sufficient conditions that a Gaussian process has continuous sample paths based on their result.*

Finally, we can state and prove a simple sufficient condition on a Gaussian process such that Assumption (2.6) holds.

Proposition 2.1.16. *Let $\{A_t\}_{t\in I}$ be a Gaussian process with almost surely continuously differentiable sample paths. Let $\Sigma' : I \times I \to \mathbb{R}^3$ be the covariance function of the Gaussian process $\{\partial_t A_t\}_{t\in I}$. If Σ' is α-Hölder continuous for any $\alpha \in (0, 1]$ i.e.,*

$$\|\Sigma'_{s_1,t_1} - \Sigma'_{s_2,t_2}\|_F \leq L\big(|s_1 - s_2|^2 + |t_1 - t_2|^2\big)^{\frac{\alpha}{2}}$$

for all $s_1, s_2, t_1, t_2 \in I$, then A satisfies condition (2.6)*.*

Proof. Let us denote the Gaussian process $\{\partial_t A_t\}_{t\in I}$ with $\{B_t\}_{t\in I}$. Applying the triangle inequality we obtain

$$\mathbb{E}\left[\max_{t\in I}\|B_t\|\right] \leq \sum_{d=1}^{3}\mathbb{E}\left[\max_{t\in I}|\pi_d\circ B_t|\right], \tag{2.14}$$

where $\pi_d : \mathbb{R}^3 \to \mathbb{R}$ is defined by $\pi_d\big((x_1,x_2,x_3)\big) = x_d$ for $d \in \{1,2,3\}$. By the definition of Gaussian processes we have that $\pi_d \circ B_t$ is a one dimensional Gaussian process. Moreover, it holds true that

$$\mathbb{E}\left[\max_{t\in I}|\pi_d\circ B_t|\right] \leq 2\mathbb{E}\left[\max_{t\in I}\pi_d\circ B_t\right] + \mathbb{E}\big[|\pi_d\circ B_0|\big], \tag{2.15}$$

for $d \in \{1,2,3\}$. To prove this, note that for $d \in \{1,2,3\}$ it holds true that

$$\max_{t\in I}|\pi_d\circ B_t| \leq \max_{t\in I}\pi_d\circ B_t + |\pi_d\circ B_0|,$$

if $\max_{t\in I}|\pi_d\circ B_t| = \max_{t\in I}\pi_d\circ B_t$, or else

$$\max_{t\in I}|\pi_d\circ B_t| \leq \max_{t\in I} -\pi_d\circ B_t + |\pi_d\circ B_0|,$$

if $\max_{t\in I}|\pi_d\circ B_t| = \max_{t\in I} -\pi_d\circ B_t$. Thus, we obtain

$$\max_{t\in I}|\pi_d\circ B_t| \leq \max_{t\in I}\pi_d\circ B_t + \max_{t\in I} -\pi_d\circ B_t + |\pi_d\circ B_0|, \tag{2.16}$$

since, if $\max_{t\in I}|\pi_d\circ B_t| = \max_{t\in I}\pi_d\circ B_t$, we have

$$\max_{t\in I}\pi_d\circ B_t \leq \max_{t\in I}\pi_d\circ B_t + \max_{t\in I} -\pi_d\circ B_t + |\pi_d\circ B_0|,$$

which is true by $\max_{t\in I} -\pi_d\circ B_t \geq -\pi_d\circ B_0$, otherwise, if $\max_{t\in I}|\pi_d\circ B_t| = \max_{t\in I} -\pi_d\circ B_t$, we obtain

$$\max_{t\in I} -\pi_d\circ B_t \leq \max_{t\in I}\pi_d\circ B_t + \max_{t\in I} -\pi_d\circ B_t + |\pi_d\circ B_0|,$$

which is true by $\max_{t\in I}\pi_d\circ B_t \geq \pi_d\circ B_0$. Taking the expectation on both sides of inequality (2.16) together with the property that $\pi_d\circ B_t \sim -\pi_d\circ B_t$ yields (2.15).

Therefore, in order to prove that the right hand side of (2.14) is finite, it remains to show by Theorem 2.1.14 that

$$\int_0^{\mathrm{diam(I)}/2}\sqrt{H\big(d_{\pi_d B},\varepsilon\big)}\,d\varepsilon < \infty, \tag{2.17}$$

where $H(d_{\pi_d B},\varepsilon)$ is the log entropy function with respect to the induced metric $d_{\pi_d B}$ on I and diam(I) is the diameter of I with respect to this metric. A short calculation yields

$$d_{\pi_d B}(s,t) = \mathrm{Var}\big[\pi_d\circ B_s\big] - 2\mathrm{Cov}\big[\pi_d\circ B_s, \pi_d\circ B_t\big] + \mathrm{Var}\big[\pi_d\circ B_t\big]$$

and the Lipschitz continuity of Σ' implies

$$\begin{aligned}\big|\mathrm{Var}\big[\pi_d \circ B_s\big] - \mathrm{Cov}\big[\pi_d \circ B_s, \pi_d \circ B_t\big]\big| &< L|s-t|^\alpha \\ \big|\mathrm{Var}\big[\pi_d \circ B_t\big] - \mathrm{Cov}\big[\pi_d \circ B_s, \pi_d \circ B_t\big]\big| &< L|s-t|^\alpha\end{aligned}$$

yielding

$$d_{\pi_d B}(s,t) \leq 2L|s-t|^\alpha .$$

The latter estimate implies $\mathrm{diam(I)} \leq 2\mathrm{L}$ and $H\big(d_{\pi_d B}, \varepsilon\big) \leq H\big(|\cdot|, \sqrt[\alpha]{\varepsilon/L}\big)$, since for any $t \in I$ we have that

$$\mathcal{B}_{\sqrt[\alpha]{\frac{\varepsilon}{L}}}\big(|\cdot|, t\big) \subset \mathcal{B}_\varepsilon\big(d_{\pi_d B}, t\big) .$$

Thus, we obtain

$$\begin{aligned}\int_0^{\mathrm{diam(I)}/2} \sqrt{H\big(d_{\pi_d B}, \varepsilon\big)}\, d\varepsilon &\leq \int_0^L \sqrt{H\Big(|\cdot|, \sqrt[\alpha]{\varepsilon/L}\Big)}\, d\varepsilon \\ &\leq \int_0^L \sqrt{\log\left(\frac{1}{\sqrt[\alpha]{\varepsilon/L}+1}\right)}\, d\varepsilon < \infty ,\end{aligned}$$

where we used $N\big(I, |\cdot|, \sqrt[\alpha]{\varepsilon/L}\big) \leq \big(\sqrt[\alpha]{\varepsilon/L}+1\big)^{-1}$ in the last inequality. □

2.2 Simultaneous Confidence Sets for Center Curves

In this section we introduce a construction using a session $\chi = (\gamma_1, ..., \gamma_N)$ of a certain rGP model γ yielding simultaneous confidence sets for its center curve γ_0.

In few words our approach is as follows: given sample paths of certain Gaussian processes $\{A_t\}_{t\in I}$ with values in $\mathbb{R}^3$ the maximum of the pointwise Hotelling T^2 statistic H_t^A is well-defined and can be used to obtain simultaneous confidence sets (SCS) of the mean of the process. Here the *Gaussian kinematic formula* (Taylor and Worsley [2008]) can be utilized to estimate the quantiles of $\max_t H_t^A$ in order to compute the confidence sets. Other methods based on the multiplier bootstrap proposed in Chernozhukov et al. [2013] or on another bootstrap approach proposed in Degras [2011] to estimate these quantiles are introduced in Section 5.1.1, where we compare the small sample performance of these methods. Taking the maximum of a pointwise defined statistic to construct SCS dates back as early as Bickel and Rosenblatt [1973]. A more recent method and a short literature overview can be found in Krivobokova et al. [2012]. Often, however, these methods rely on asymptotic distributions for large sample sizes, whereas the Gaussian kinematic formula does not. We will shortly review the construction of SCS in the Gaussian process case and then construct a process $\big\{\tilde{H}_t^\gamma\big\}_{t\in I}$ for a right Gaussian perturbation model γ, which can be used to construct simultaneous confidence sets in $SO(3)$ for the center curve $\gamma_0 \in \mathcal{X}$ of γ.

In a last step we show that the computation of the quantiles of $\max_t \tilde{H}_t^\gamma$ reduces to estimating the quantiles of $\max_t H_t^A$ for the generating Gaussian process A of the considered rGP, if our data is sufficiently concentrated. Hence also in this case the Gaussian kinematic formula can be used to compute the confidence set.

Confidence sets for Gaussian processes. Let $\{A_t\}_{t\in I}$ be an $\mathbb{R}^3$-valued Gaussian process. Assume that A fulfills the following additional assumptions:

(*GKF 1*) $A = F_t(\xi_1, \xi_2, \xi_3)^T$. Here $\xi^1, \xi^2, \xi^3 \sim \xi$ independent $\mathbb{R}$-valued with almost surely $\mathcal{C}^2$-sample paths. $F_t \in \mathcal{C}^1\big(I, \mathbb{R}^{3\times 3}\big)$ is invertible for all $t \in I$.

(*GKF 2*) $\{\xi_t\}_{t\in I}$ is Gaussian with $\mathbb{E}[\xi_t] = 0$ and $\mathrm{Var}[\xi_t] = I_{3\times 3}$ for all $t \in I$.

(*GKF 3*) The joint distribution of $\big(\xi_t, \xi_t'\big)$ is nondegenerate for all $t \in I$. Here $\{\xi_t'\}_{t\in I} = \big\{\big(\frac{d\xi}{dt}\big)_t\big\}_{t\in I}$.

(*GKF 4*) Let $L, \alpha \in \mathbb{R}_{>0}$ and Σ' denote the covariance function of $\{\xi_t'\}_{t\in I}$. We assume

$$\Sigma'_{t,t} - 2\Sigma'_{s,t} + \Sigma'_{s,s} \le L\big|\log|s-t|\big|^{-(1+\alpha)}$$

for all $s, t \in I$.

These assumptions are necessary in order that the Gaussian kinematic formula (2.28) is valid (see Taylor and Worsley [2008, Section 3.1.] and Adler and Taylor [2009, Theorem 15.2.1., p.391]).

Given an i.i.d. sample $A^1, ..., A^N \sim A$, $N > 3$, where A fulfills Assumptions (*GKF 1*)–(*GKF 4*), we define the one-dimensional stochastic process

$$H_t^A = N\big(\bar{A}_t - \mathbb{E}[A_t]\big)^T \Big(\hat{\Sigma}_t^A\Big)^{-1} \big(\bar{A}_t - \mathbb{E}[A_t]\big) \tag{2.18}$$

called its *Hotelling T^2 process*, with

$$\hat{\Sigma}_t^A = \frac{1}{N-1}\sum_{n=1}^{N} \big(A_t^n - \bar{A}_t\big)\big(A_t^n - \bar{A}_t\big)^T. \tag{2.19}$$

This process is only well-defined, if $\hat{\Sigma}_t^A$ is non-singular for all $t \in I$. For some further restrictions on ξ in (*GKF 1*)–(*GKF 4*) a proof of non-singularity is given in Cao et al. [1999, Section 3.3]. For the general case, we could not find a proof in the literature. However, in our simulations we never experienced any singularity problems and from theoretical considerations, which are not yet a complete proof, we suspect that the non-singularity holds true in general given (*GKF 1*)–(*GKF 4*) for $N > 3$.

From now on, we assume that our processes have the property that $\{H_t^A\}_{t\in I}$ is well-defined as a process. Note that H_t^A is Hotelling $T^2_{3,N-1}$ distributed for each $t \in I$. Hence, since this distribution is pivotal, we obtain at once pointwise confidence sets for the mean $\mathbb{E}[A_t]$ at each $t \in I$.

Now, let $h \in \mathbb{R}$ such that $\mathbb{P}\big\{\sup_{t\in I} H_t^A \le h\big\} \ge \beta \in (0,1)$. Then the collection of sets

$$\mathcal{C}_\beta(t) = \left\{ a \in \mathbb{R}^3 \;\middle|\; N\big(\bar{A}_t - a\big)^T \Big(\hat{\Sigma}_t^A\Big)^{-1} \big(\bar{A}_t - a\big) \le h \right\}$$

form simultaneous $\beta\cdot 100\%$-confidence sets for $\mathbb{E}[A]$, since

$$\mathbb{P}\big(\,\mathbb{E}[A_t] \in \mathcal{C}_\beta(t) \text{ for all } t \in I\,\big) \ge \beta\,.$$

This simultaneous inference on the entire path, however, requires to include the covariance structure $(s,t) \mapsto \Sigma_{s,t} = \mathrm{Cov}[A_s, A_t]$ of A, since the distribution of $\sup_{t\in I} H_t^A$ is in general not pivotal. For convenience, we define the quantiles of this process. Its estimation as achieved in the literature is deferred to Section 2.3.

Definition 2.2.1 (Hotelling T^2 Process Quantiles)**.** *Let $A^1, ..., A^N \sim A$, $N > 3$, with a Gaussian process A on $\mathbb{R}^3$ satisfying Assumptions (GKF 1)–(GKF 4). For all levels $0 \leq \beta \leq 1$ we call*

$$h_{A,N,\beta} = \inf\left\{ h \in \mathbb{R}_{>0} \,\middle|\, \mathbb{P}\left(\sup_{t\in I} H_t^A \leq h \right) \geq \beta \right\}$$

the Hotelling process β-quantile.

Intrinsic residuals. Since we do not observe the realizations of the Gaussian error processes $\{A_t\}_{t\in I}$ of a right Gaussian perturbation model γ directly, natural and computable residuals of a session $(\gamma_1, ..., \gamma_N)$ of γ could be defined using Corollary 2.1.9 as

$$\iota^{-1} \circ \mathrm{Log}\Big(\hat{\gamma}_N^T(t)\gamma_n(t) \Big), \tag{2.20}$$

where $\hat{\gamma}_N(t) \in \hat{\mu}\big((\gamma_1, ..., \gamma_N), t\big)$ is a measurable choice of an element from the PESM for each $t \in I$, if not unique. These residuals are, however, only almost surely well-defined for fixed $t \in I$ as the next proposition shows, but they do not need to be well-defined for all t simultaneously, since the map $t \mapsto \hat{\gamma}_N^T(t)\gamma_n(t)$ may still hit the cut locus $\mathcal{C}_{I_{3\times 3}}$ of the identity matrix for some $t \in I$ and the Lie logarithm is only well-defined on $SO(3)\backslash\mathcal{C}_{I_{3\times 3}}$ (see equation A.7 in Appendix A.1).

Proposition 2.2.2. *At each $t \in I$ the residuals defined in Equation* (2.20) *are almost surely well-defined.*

Proof. This follows directly from the fact that the cut locus $\hat{\gamma}_N^T(t)\mathcal{C}_{I_{3\times 3}}$ of $\hat{\gamma}_N(t)$ has measure zero with respect to the Riemannian volume (e.g.,Itoh and Tanaka [1998]) for every $t \in I$ together with the technical Lemma B.0.1. □

Usually, for applications this is enough, since the data is assumed to be concentrated and therefore, the cut locus does not cause any problems for real samples. Theoretically, however this is unsatisfying. Luckily, there exists a simple trick to circumvent this ambiguity, since we can define a logarithm even if $\hat{\gamma}_N^T(t)\gamma_n(t) \in \mathcal{C}_{I_{3\times 3}}$. Therefore, note that $\mathrm{Exp} : \overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0 \to \mathcal{C}_{I_{3\times 3}}$, where $\mathfrak{U}_0 = \{A \in \mathfrak{so}(3) \,|\, \|A\|_F < \pi\}$ is the open ball centered at 0, is a double cover (see Duistermaat and Kolk [1999, p. 25]). Thus, since $\mathrm{Exp} : \mathfrak{U}_0 \to SO(3)\backslash\mathcal{C}_{I_{3\times 3}}$ is bijective, we only have to find a set $\mathfrak{C} \subset \overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0$ such that $\mathrm{Exp} : \mathfrak{C} \to \mathcal{C}_{I_{3\times 3}}$ is bijective.

By the periodicity of the Lie exponential (see Equation (A.6)), we have that $\mathrm{Exp}(A) = \mathrm{Exp}(-A)$ for all $A \in \overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0 = \{A \in \mathfrak{so}(3) \,|\, \|A\|_F = \pi\}$. Hence, the upper hemisphere of the 2-sphere $\overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0$ together with an appropriate choice on the remaining S^1 (again the upper hemisphere together with one point) i.e.,

$$\mathfrak{C} = \Big\{ A \in \overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0 \,\Big|\, A_{21} > 0 \Big\} \cup \Big\{ A \in \overline{\mathfrak{U}}_0\backslash\mathfrak{U}_0 \,\Big|\, A_{21} = 0 \wedge A_{32} > 0) \Big\} \cup \left\{ \begin{pmatrix} 0 & 0 & \pi \\ 0 & 0 & 0 \\ -\pi & 0 & 0 \end{pmatrix} \right\},$$

makes $\mathrm{Exp} : \mathfrak{C} \to \mathcal{C}_{I_{3\times 3}}$ bijective. This implies that $\mathrm{Exp} : \mathfrak{V} = \mathfrak{U}_0 \cup \mathfrak{C} \to SO(3)$ is a bijection and we denote its inverse with $\mathfrak{L} : SO(3) \to \mathfrak{V}$. Finally, we can modify the residuals (2.20) by defining our intrinsic residuals as

$$X_t^n = \mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_n(t) \Big), \tag{2.21}$$

which are well-defined for all $t \in I$.

Confidence sets in SO(3). To construct simultaneous confidence sets we can proceed analogously to the case of Gaussian processes by defining the one-dimensional process

$$\tilde{H}_t^X = N X_t^T \Big(\hat{S}_t^X\Big)^{-1} X_t \tag{2.22}$$

with

$$X_t = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_0(t)\Big)$$
$$\hat{S}_t^X = \frac{1}{N-1}\sum_{n=1}^{N} X_t^n (X_t^n)^T .$$

Here $\hat{\gamma}_N(t) \in \hat{\mu}\big(\chi_N, t\big)$ is a measurable choice for all $t \in I$ and again we have to assume that $\big\{\hat{S}_t^X\big\}_{t\in I}$ is non-singular for all t. As before in the Gaussian case, we define the quantiles

$$\tilde{h}_{\gamma,N,\beta} = \inf\left\{ h \in \mathbb{R}_{\geq 0} \;\middle|\; \mathbb{P}\left(\sup_{t\in I} \tilde{H}_t^X \leq h \right) \geq \beta \right\} .$$

From these definitions we obtain at once $\beta \cdot 100\%$-confidence sets for γ_0.

Theorem 2.2.3. *Let $\chi_N = (\gamma_1, ..., \gamma_N)$ be a session of a rGP γ with center curve γ_0. Let $\hat{\gamma}_N(t) \in \hat{\mu}\big(\chi_N, t\big)$ a measurable choice for all $t \in I$. We define the sets*

$$\mathcal{C}_\beta\big(\chi_N; t\big) = \left\{ a \in \mathbb{R}^3 \;\middle|\; N a^T \Big(\hat{S}_t^X\Big)^{-1} a \leq \tilde{h}_{\gamma,N,\beta} \right\} .$$

Then it follows that

$$\mathbb{P}\Big(\gamma_0(t) \in \hat{\gamma}_N(t)\mathrm{Exp}\Big(\iota \circ \mathcal{C}_\beta\big(\chi_N; t\big)\Big) \text{ for all } t \in I \Big) \geq \beta$$

and hence we obtained simultaneous $\beta \cdot 100\%$-confidence sets for γ_0.

An important property of these confidence sets is that they are also equivariant with respect to the group $\mathcal{S}$ of gait similarities in the following sense:

Theorem 2.2.4. *Let $\chi_N = (\gamma_1, ..., \gamma_N)$ be a session of a rGP γ with center curve γ_0 and let $(\psi, \phi) \in \mathcal{S}$ be arbitrary. Define the sample $\eta_n = (\psi, \phi).\gamma_n$ for all $n \in \{1, ..., N\}$, which are samples of the rGP $(\psi, \phi).\gamma$ with center curve $\eta_0 = (\psi, \phi).\gamma_0$. Moreover, let $\psi(R) = P_\psi R Q_\psi^T$ with $R, P_\psi, Q_\psi \in SO(3)$ and*

$\hat{\gamma}_N(t) \in \hat{\mu}\big(\chi_N, t\big)$ *and* $\hat{\eta}_N = (\psi, \phi).\hat{\gamma}_N(t) \in \hat{\mu}\big((\psi, \phi).\chi_N, t\big)$, *seeTheorem 2.1.2, for all* $t \in I$.

Then the simultaneous confidence sets for $(\psi, \phi).\gamma_0$ *computed from* $\eta_1, ..., \eta_N$ *satisfy*

$$\hat{\eta}_N(t)\mathrm{Exp}\Big(\iota \circ \mathcal{C}_\beta\big((\eta_1, ..., \eta_N); t\big)\Big) = \big((\psi, \phi).\hat{\gamma}_N\big)(t)\mathrm{Exp}\Big(\iota \circ Q_\psi \mathcal{C}_\beta\big(\chi_N; \phi(t)\big)\Big)\,,$$

i.e., they can be computed by the knowledge of the confidence sets for γ_0 *and* $(\psi, \phi) \in \mathcal{S}$ *only.*

Proof. We start the proof restating the definitions of the intrinisic residuals for each of the samples:

$$X_t^n = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_n(t)\Big) \quad \text{and} \quad Y_t^n = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\eta}_N^T(t)\eta_n(t)\Big)\,,$$

Hence a simple computation using the equivariance of the PESM, Lemma B.0.5 and $\iota(Qv) = Q\iota(v)Q^T$ for all $v \in \mathbb{R}^3, Q \in SO(3)$ (see Lemma A.1.2(iv)) yields

$$\begin{aligned} Y_t^n &= \iota^{-1} \circ \mathfrak{L}\Big(\big((\phi, \psi).\hat{\gamma}_N(t)\big)^T(\phi, \psi).\gamma_n(t)\Big) \\ &= \iota^{-1} \circ \mathfrak{L}\Big(Q_\psi \hat{\gamma}_N^T\big(\phi(t)\big)\gamma_n\big(\phi(t)\big)Q_\psi^T\Big) \\ &= \pm Q_\psi\, \iota^{-1} \circ \mathfrak{L}\Big(\hat{\gamma}_N^T\big(\phi(t)\big)\gamma_n\big(\phi(t)\big)\Big) \\ &= \pm Q_\psi\, X_{\phi(t)}^n\,. \end{aligned}$$

Moreover, note that for $X_t = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_0(t)\Big)$ and $Y_t = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\eta}_N^T(t)\eta_0(t)\Big)$ we obtain by a similar calculation $Y_t = \pm Q_\psi X_{\phi(t)}$ yielding

$$\hat{S}_t^Y = Q\hat{S}_{\phi(t)}^X Q^T \quad \text{and} \quad H_t^Y = H_{\phi(t)}^X\,,$$

which immediately implies $\mathcal{C}_\beta\big((\eta_1, ..., \eta_N); t\big) = Q_\psi \mathcal{C}_\beta\big(\chi_N; \phi(t)\big)$ and $\tilde{h}_{\gamma,N,\beta} = \tilde{h}_{\eta,N,\beta}$. □

Concentrated data approximations. The main difficulty in computing simultaneous confidence sets for the center curve of a rGP is the estimation of the quantile $\tilde{h}_{\gamma,N,\beta}$. Additionally, it is further complicated in our case by the fact that we have to deal with small sample sizes of $N \approx 10$, such that bootstrap approaches do not seem viable, see Section 5.1.1.

Therefore, building on concentrated error approximations similar to the results in Rancourt et al. [2000] for extrinsic residuals, the main result of this section will show that in the case of concentrated errors, the residuals X_t^n are approximately the residuals of the Gaussian process $\{A_t\}_{t\in I}$ generating the rGP. This suggests that $\tilde{h}_{\gamma,N,\beta} \approx h_{A,N,\beta}$ and the latter is, indeed, computable using a Gaussian kinematic formula.

Theorem 2.2.5. *Suppose that* $\chi_N = (\gamma_1, ..., \gamma_N)$ *for* $N \in \mathbb{N}$ *be a session of a rGP model with center curve* γ_0 *such that the Gaussian error process satisfies*

$\max_{t\in I}\|A_t\| = \mathcal{O}_p(\sigma)$ *with* $0 < \sigma \to 0$ *for all* $t \in I$. *Let* $\hat{\gamma}_N(t)$ *be a measurable selection of the PESM* $\hat{\mu}_N\big(\chi_N, t\big)$ *at time* $t \in I$. *Setting*

$$\begin{aligned} X_t &= \iota^{-1}\circ\mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_0(t)\Big) \\ X_t^n &= \iota^{-1}\circ\mathfrak{L}\Big(\hat{\gamma}_N^T(t)\gamma_n(t)\Big) \\ \bar{A}_t &= \frac{1}{N}\sum_{n=1}^{N} A_t^n \end{aligned}$$

we have for every $t \in I$ *that*

$$X_t = \bar{A}_t + \mathcal{O}_p\left(\sigma^2\right) \tag{2.23}$$
$$X_t^n = A_t^n - \bar{A}_t + \mathcal{O}_p\left(\sigma^2\right), \tag{2.24}$$

where the $\mathcal{O}_p(\sigma^2)$ *holds uniform on* I.

Proof. As usual assume $\sigma_l \to 0$ and $\max_{t\in I}\|A_t^l\| = \mathcal{O}_p(\sigma_l)$. Let $\hat{\gamma}_{N,l}(t) \in \hat{\mu}_N\big(\chi, t\big)$. Then for each $t \in I$ there exist a $B_t^l \in \mathbb{R}^3$ such that $\hat{\gamma}_{N,l}(t) = \gamma_0\,\mathrm{Exp}\Big(\iota\circ B_t^l\Big)$. Then, making use of (B.1) and the property, $\mathrm{tr}\big(\iota(\mathrm{C})\iota(\mathrm{D})\big) = -2\mathrm{C^T D}$ for all $C, D \in \mathfrak{so}(3)$, we have that B_t^l must maximize for each $t \in I$

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N}\mathrm{tr}\left(\hat{\gamma}_{N,l}^T(t)\gamma_0(t)\mathrm{Exp}\Big(\iota\circ A_t^{n,l}\Big)\right) \\ &\quad= \mathrm{tr}\Bigg(\left(\mathrm{I}_{3\times3} - \iota\big(\mathrm{B_t^l}\big)\,\mathrm{sinc}\big(\big\|\mathrm{B_t^l}\big\|\big) + \frac{1-\cos\big(\big\|\mathrm{B_t^l}\big\|\big)}{\big\|\mathrm{B_t^l}\big\|^2}\iota\big(\mathrm{B_t^l}\big)^2\right) \\ &\qquad\cdot\Big(I_{3\times3} + \iota\big(\bar{A}_t^l\big) + \mathcal{O}_p\left(\sigma_l^2\right)\Big)\Bigg) \\ &\quad= 3 + 2\big(B_t^l\big)^T\bar{A}_t^l\,\mathrm{sinc}\big(\big\|B_t^l\big\|\big) - 2\frac{1-\cos\big(\big\|B_t^l\big\|\big)}{\big\|B_t^l\big\|^2}\big\|B_t^l\big\|^2 + \mathcal{O}_p\left(\sigma_l^2\right) \\ &\quad= 1 + 2\Big(\big(B_t^l\big)^T\bar{A}_t^l\,\mathrm{sinc}\big(\big\|B_t^l\big\|\big) + \cos\big(\big\|B_t^l\big\|\big)\Big) + \mathcal{O}_p\left(\sigma_l^2\right). \end{aligned}$$

Note that the $\mathcal{O}_p\left(\sigma_l^2\right)$ is indeed uniform in $t \in I$.

Writing $B_t^l = rE$ with a unit vector E and length $0 \le r \le \pi$, the first two summands above are maximized in B_t^l if

$$x\sin(r) + \cos(r)$$

is maximal under the side condition $-\|\bar{A}_t^l\| \le x = E^T\bar{A}_t^l \le \|\bar{A}_t^l\|$. Hence, for $0 \le r < \pi$ choose the maximizing $x = \|\bar{A}_t^l\|$ (as large as possible) and hence $r = \arctan\big(\|\bar{A}_l\|\big) \in (0, \pi/2)$. In consequence we have that

$$B_t^l = \bar{A}_t^l\,\frac{\arctan\|\bar{A}_t^l\|}{\|\bar{A}_t^l\|} + \mathcal{O}_p\left(\sigma_l^2\right) = \bar{A}_t^l + \mathcal{O}_p\left(\sigma_l^2\right) = \mathcal{O}_p\left(\sigma_l\right) \tag{2.25}$$

with the $\mathcal{O}_p(\sigma_l)$ and $\mathcal{O}_p(\sigma_l^2)$ uniform in $t \in I$. Now, define the set

$$\Omega_l = \left\{\omega \in \Omega \,\middle|\, B_t^l(\omega, t) \in \mathcal{B}_{\pi/2}(0, \|\cdot\|) \text{ for all } t \in I\right\}.$$

By (2.25) we have that $\mathbb{P}(\Omega_l) \to 1$ for $l \to \infty$. Moreover, we obtain by definition

$$\iota^{-1} \circ X_t^l = \mathfrak{L}\Big(\gamma_0^T(t)\hat{\gamma}_{N,l}(t)\Big) = \mathrm{Log}\Big(\gamma_0^T(t)\hat{\gamma}_{N,l}(t)\Big) = \iota \circ B_t^l = \iota \circ \bar{A}_t^l + \mathcal{O}_p\left(\sigma_l^2\right)$$

on Ω_l. Thus, using the same arguments for Ω_l instead of $\mathfrak{K}$ in the proof of Theorem 1.3.7 shows (2.23).

To establish equation (2.24) define the set

$$\tilde{\Omega}_l = \left\{\omega \in \Omega \,\middle|\, A_t^{n,l}(\omega, t) \in \mathcal{B}_{\pi/2}(0, \|\cdot\|) \text{ for all } t \in I\right\}.$$

Then $\mathbb{P}(\Omega_l \cap \tilde{\Omega}_l) \to 1$ for $l \to \infty$ and

$$\iota^{-1} \circ X_t^{n,l} = \mathfrak{L}\Big(\hat{\gamma}_{N,l}^T(t)\gamma_{n,l}(t)\Big) = \mathrm{Log}\Big(\hat{\gamma}_{N,l}^T(t)\gamma_{n,l}(t)\Big)$$

for all $t \in I$ on $\Omega_l \cap \tilde{\Omega}_l$. In consequence, in conjunction with (1.14) we have

$$\begin{aligned}\iota^{-1} \circ \mathfrak{L}\Big(\hat{\gamma}_{N,l}^T\gamma_{n,l}\Big) &= \iota^{-1} \circ \mathrm{Log}\left(\mathrm{Exp}\Big(-\iota \circ \bar{A}^l + \mathcal{O}_p\big(\sigma_l^2\big)\Big)\,\mathrm{Exp}\Big(\iota \circ A^{n,l}\Big)\right) \\ &= A^{n,l} - \bar{A}^l + \mathcal{O}_p(\sigma_l^2)\end{aligned}$$

on $\Omega_l \cap \tilde{\Omega}_l \cap \mathfrak{K}$, where $\mathfrak{K}$ is defined as in the proof of 1.3.7. Again using the same arguments as presented there, where only $\mathfrak{K}$ is replaced by $\Omega_l \cap \tilde{\Omega}_l \cap \mathfrak{K}$, proves (2.24).

□

Corollary 2.2.6. *With the assumptions and notations of Theorem 2.2.5, the one sample Hotelling T^2 statistic satisfies*

$$H_t^A = \tilde{H}_t^X + \mathcal{O}_p(\sigma),$$

if additionally $\mathrm{Var}\big[A_t\big] = \sigma^2\Sigma_t$ *with* Σ_t *non-singular for all* $t \in I$.

Proof. Recall the definitions

$$H_t^{A^l} = N\big(\bar{A}_t^l\big)^T\Big(\hat{\Sigma}_t^{A_l}\Big)^{-1}\bar{A}_t^l \quad \text{and} \quad \tilde{H}_t^{X^l} = N\big(X_t^l\big)^T\Big(\hat{S}_t^{X^l}\Big)^{-1}X_t^l,$$

see equation (2.18) and (2.22). By virtue of Theorem 2.2.5 we obtain

$$\hat{S}_t^{X^l} = \hat{\Sigma}_t^{A^l} + Z_t^l$$

with $\max_t \big\|Z_t^l\big\|_F = \mathcal{O}_p(\sigma_l^3)$. Using Henderson and Searle [1981, p. 58, eq. (24)] yields

$$\begin{aligned}\frac{1}{N}\tilde{H}_t^{X^l} &= \big(X_t^l\big)^T\Big(\hat{\Sigma}_t^{A^l} + Z_t^l\Big)^{-1}X_t^l \\ &= \big(X_t^l\big)^T\Big(\hat{\Sigma}_t^{A^l}\Big)^{-1}X_t^l - \big(X_t^l\big)^T\Big(\hat{\Sigma}_t^{A^l}\Big)^{-1}Z_t^l\Big(I_{3\times 3} + \Big(\hat{\Sigma}_t^{A^l}\Big)^{-1}Z_t^l\Big)^{-1}\Big(\hat{\Sigma}_t^{A^l}\Big)^{-1}X_t^l.\end{aligned}$$

Note that by the assumption $\mathrm{Var}\big[A_t^l\big] = \sigma_l^2 \Sigma_t$ we have that $\max_{t\in I} \left\| \big(\hat{\Sigma}_t^{A_l}\big)^{-1} \right\| = \mathcal{O}_p(\sigma_l^{-2})$, since $\sigma_l^{-2}\hat{\Sigma}_t^{A_l}$ is independent of l. Thus, we obtain

$$\big(X_t^l\big)^T \left(\hat{\Sigma}_t^{A_l}\right)^{-1} X_t^l = \frac{1}{N} H_t^{A^l} + \mathcal{O}_p(\sigma_l)$$

by equation (2.23). Moreover, we obtain that $\max_{t\in I} \left\| \left(\hat{\Sigma}_t^{A_l}\right)^{-1} Z_t^l \right\| = \mathcal{O}_p(\sigma_l)$ implying $\left(\hat{\Sigma}_t^{A_l}\right)^{-1} Z_t^l \xrightarrow{P} 0$ uniformly for $l \to \infty$. Therefore, the same technique as in the proof of Theorem 1.3.8 with $U = \left\{ \left\| \left(\hat{\Sigma}_t^{A_l}\right)^{-1} Z_t^l \right\| < 1 \right\}$ (instead of $\mathfrak{K}$) and the Taylor expansion on U given by the Von Neumann series

$$\left(I_{3\times 3} + \left(\hat{\Sigma}_t^{A^l}\right)^{-1} Z_t^l\right)^{-1} = \sum_{j=0}^{\infty} (-1)^j \left(\left(\hat{\Sigma}_t^{A^l}\right)^{-1} Z_t^l\right)^j$$

shows at once

$$\big(X_t^l\big)^T \left(\hat{\Sigma}_t^{A^l}\right)^{-1} Z_t^l \left(I_{3\times 3} + \left(\hat{\Sigma}_t^{A^l}\right)^{-1} Z_t^l\right)^{-1} \left(\hat{\Sigma}_t^{A^l}\right)^{-1} X_t^l = \mathcal{O}_p(\sigma_l)\,,$$

which finishes the proof. □

Remark 2.2.7. *The assumption that* $\mathrm{Var}\big[A_t^l\big] = \sigma_l^2 \Sigma_t$ *is natural in this case, since otherwise the sample covariance matrix* $\hat{\Sigma}_t^{A_l}$ *will become singular for* $l \to \infty$*, which does contradict our model assumption that the Gaussian process of a rGP model has an everywhere invertible variance matrix* $\Sigma_t = \mathrm{Var}\big[A_t\big]$*.*

2.3 Gaussian Kinematic Formula

As we have seen in the previous sections it remains to estimate the quantile $\tilde{h}_{\gamma,N,\beta}$ in order to construct simultaneous $\beta \cdot 100\%$-confidence sets for the center curve γ_0 of a rGP model.

Therefore, one has to estimate the probability that the real valued stochastic process $\big\{\tilde{H}_t^X\big\}_{t\in I}$ (see equation (2.22)) exceeds a certain value over the index set I. Usually, these probabilities cannot be computed explicitly.

However, if $\{A_t\}_{t\in I}$ is a Gaussian process the so called *expected Euler characteristic heuristic* (see Taylor et al. [2005a]) states that

$$\mathbb{P}\Big(\max_{t\in I} A_t > h\Big) \approx \mathbb{E}\Big[\mathfrak{x}\big(\{t \in I \mid A_t \geq h\}\big)\Big]\,, \tag{2.26}$$

where $\mathfrak{x}(\mathcal{U})$ denotes the topological invariant called *Euler characteristic* (EC) of a set $\mathcal{U} \subset I$. In our case, where $I = [0,1]$ the Euler characteristic $\mathfrak{x}(\mathcal{U})$ is equal to the number of connected components of $\mathcal{U}$. The above approximation of the exceedance probability of a Gaussian process by the expected Euler characteristic can be proven to be strikingly accurate for low probabilities. More precisely, the

approximation error is roughly of the order $e^{-\theta h^2/2}$ for some $\theta > 0$ (see Taylor et al. [2005a] and references therein).

Interestingly, the expected EC of the excursion set (i.e., the r.h.s. of equation (2.26)) can be computed analytically for many processes $\{A_t\}_{t\in I}$, which are derived from Gaussian processes (see among others Taylor [2006] and Taylor and Worsley [2008]). These formulas are known as *Gaussian kinematic formulas* and are used to estimate exceedance probabilities in order to do inference for example on linear models including a Gaussian process as noise (see Taylor and Worsley [2007]). Moreover, it is conjectured that even in these cases the expected EC is a good approximation of the exceedance probability of the process for low probabilities. A detailed overview of this topic can be found in Adler [2000] and the book Adler and Taylor [2009].

In order to estimate our quantiles $\tilde{h}_{\gamma,N,\beta}$ for the process $\left\{\tilde{H}^\gamma\right\}_{t\in I}$ derived from the rGP model γ. We will use the expected Euler characteristic heuristic and assume that

$$\mathbb{P}\Big(\max_{t\in I}\tilde{H}_t^X > h\Big) \approx \mathbb{E}\Big[\mathfrak{x}\Big(\big\{t\in I \mid \tilde{H}_t^X \geq h\big\}\Big)\Big] \approx \mathbb{E}\Big[\mathfrak{x}\Big(\big\{t\in I \mid H_t^A \geq h\big\}\Big)\Big]\,. \tag{2.27}$$

Here the second approximation is due to Corollary 2.2.6, which is expected to be valid in the case of concentrated errors. Although we cannot rigorously justify this approximation, we will show later in Chapter 5 using simulations that it works well in our application.

In the next sections we will restate the formulas for computation and estimation of the expected EC of Hotelling T^2-processes $\left\{H_t^A\right\}_{t\in I}$ as they are derived in Taylor and Worsley [2007] and Taylor and Worsley [2008].

Gaussian kinematic formula (GKF) for Hotelling T^2-processes. Let us consider a Hotelling T^2-process $\left\{H_t^A\right\}_{t\in I}$ (see Equation (2.18)) derived from a Gaussian process $\left\{A_t\right\}_{t\in I}$ fulfilling Assumptions (*GKF 1*)-(*GKF 4*). Upon knowledge of the covariance function Σ of the process A the expected EC of the excursion set $\left\{t\in I \mid H_t^A \geq h\right\}$ can be computed explicitly as shown in Taylor and Worsley [2007, 2008].

Their result known as the GKF for Hotelling T^2-processes is the following

$$\mathbb{E}\Big[\mathfrak{x}\big\{t\in I \mid H_t^A \geq h\big\}\Big] = \mathcal{L}_0(I)\rho_0^H(h) + \mathcal{L}_1(I)\rho_1^H(h) \tag{2.28}$$

with the so called Lipschitz-Killing curvatures

$$\mathcal{L}_0(I) = 1, \quad \mathcal{L}_1(I) = \int_0^1 \sqrt{\mathrm{Var}\Big[\tfrac{dA}{dt}(t)\Big]}\,dt$$

which are warped versions of intrinsic volumes converting the anisotropic case $\Sigma(t) \neq I_{3\times 3}$ to isotropicity (see Taylor and Worsley [2007, Sec. 3 and Appendix A.1]).

The *Euler characteristic densities* ρ_j^H for $j \in \{1,2\}$ appearing in the GKF (2.28) can be computed from the EC densities of a T-process with $N-1$ degrees

of freedom via Roy's union intersection principle (cf. Taylor and Worsley [2008, Sec. 3.1.]) using the formula

$$\rho_j^H(h) = \sum_{d=1}^{3} \mu_d(S^2)\rho_{j+d}^T\big(\sqrt{h}\big)\,, \qquad j = 0, 1\,.$$

Here $\mu_d(S^2)$ denotes the d-dimensional intrinsic volume of the two-sphere S^2 given by

$$\mu_0\big(S^2\big) = 2, \quad \mu_1\big(S^2\big) = 0 = \mu_3\big(S^2\big), \quad \mu_2\big(S^2\big) = 4\pi\,,$$

in Taylor and Worsley [2008, p. 23]. In contrast in the Stochastic Geometry literature, μ_0 gives the number of connected components and μ_2 would give the surface area of S^2 (e.g., Mecke and Stoyan [2000, p. 100])). Moreover, the EC densities of a T-process with $(N-1)$ degrees of freedom have the explicit representations

$$\begin{aligned}
\rho_0^T(t) &= \int_t^\infty \frac{\Gamma\left(\frac{N}{2}\right)}{\sqrt{N-1\pi}\Gamma\left(\frac{N-1}{2}\right)}\left(1+\tfrac{u^2}{N-1}\right)^{-N/2} du \\
\rho_1^T(t) &= (2\pi)^{-1}\left(1+\tfrac{t^2}{N-1}\right)^{1-N/2} \\
\rho_2^T(t) &= (2\pi)^{-3/2}\frac{\Gamma\left(\frac{N}{2}\right)}{\sqrt{\frac{N-1}{2}}\Gamma\left(\frac{N-1}{2}\right)} t\left(1+\tfrac{t^2}{N-1}\right)^{1-N/2} \\
\rho_3^T(t) &= (2\pi)^{-2}\left(\tfrac{N-2}{N-1}t^2-1\right)\left(1+\tfrac{t^2}{N-1}\right)^{1-N/2}\,,
\end{aligned}$$

given in Taylor and Worsley [2007, p. 915].

Estimation of the quantile $\tilde{h}_{\gamma,N,\beta}$. Using the GKF for Hotelling T^2-processes together with the EC heuristic (2.27) yields

$$\mathbb{P}\Big(\max_{t\in I}\tilde{H}_t^X > h\Big) \approx 2\rho_0^T\big(\sqrt{h}\big) - 4\pi\rho_2^T\big(\sqrt{h}\big) - \mathcal{L}_1(I)\left(2\rho_1^T\big(\sqrt{h}\big) + 4\pi\rho_3^T\big(\sqrt{h}\big)\right)\,,$$

which can be used, if $\mathcal{L}_1(I)$ is known, to estimate the value $\tilde{h}_{\gamma,N,\beta}$ for low probabilities $1-\beta$ by solving

$$2\rho_0^T(\sqrt{h}) - 4\pi\rho_2^T(\sqrt{h}) - \mathcal{L}_1(I)\left(2\rho_1^T(\sqrt{h}) + 4\pi\rho_3^T(\sqrt{h})\right) = \beta\,. \tag{2.29}$$

Thus, it remains to estimate the Lipschitz-Killing curvature $\mathcal{L}_1(I)$. This has been achieved for Gaussian processes in $\mathbb{R}^D$, $D \in N$, in Taylor and Worsley [2007, Sect. 4] and Taylor and Worsley [2008], where they also proved that their estimator is consistent.

Recall that by Theorem 2.2.5 the intrinsic residuals of a sample from an rGP model γ are in the case of concentrated errors close to the residuals of the generating Gaussian process $\{A_t\}_{t\in I}$. Since the estimator of Taylor and Worsley [2008, Equation (18)] is based only on the Gaussian residuals, we adapt their

estimator by simply replacing their residuals by our intrinsic residuals to obtain an estimator of the Lipschitz-Killing curvature $\mathcal{L}_1(I)$.

For convenience we restate their estimator tailored to our scenario. Assume we observe a session $(\gamma_1, \dots, \gamma_N)$ of an rGP model and the observation times are $0 = t_1 < t_2 < ... < t_K = 1$. Recall the definition of the intrinsic residuals $X^n_{t_k}$ from Theorem 2.2.5 and let us defne

$$R_{t_k} = \left(X^1_{t_k}, \dots, X^N_{t_k}\right)^T \in \mathbb{R}^{N\times 3}.$$

Further, denote by $R^d_{t_k}$ the d-th column of R_{t_k} and let us define the normalized residuals by

$$\hat{R}^d_{t_k} = \frac{R^d_{t_k}}{\|R^d_{t_k}\|}$$

for $d \in \{1, 2, 3\}$ and $k \in \{1, \dots, K\}$. Then the estimator of the Lipschitz-Killing curvature $\mathcal{L}_1(I)$ is given by

$$\hat{\mathcal{L}}_1(I) = \frac{1}{3}\sum_{k=1}^{K-1}\sum_{d=1}^{3}\left\|\hat{R}^d_{t_{k+1}} - \hat{R}^d_{t_k}\right\|. \tag{2.30}$$

Chapter 3

Estimation of Gait Similarities

In this chapter we will provide estimators for the gait similarities perturbing gait data as discussed in Section 1.1. This is a necessary step in order to compare different sessions. Since we build on estimation procedures introduced for somewhat similar problems treated in the literature, we will discuss them shortly as well. Our estimation procedures are closely related to Procrustes analysis (e.g., Carne [1990] and Dryden and Mardia [1998, Chapter 5 and Section 12.4]) or the concept of optimal positioning on manifolds with a Lie group action of Huckemann et al. [2010]. In the context of time warping of functional data related ideas are discussed in Srivastava et al. [2011b] and Vantini [2012].

Since some considerations and results can be stated in our general framework of Section 1.2, we will do this before discussing our particular estimators.

General methodology. As discussed in Section 1.2 G-equivariant descriptors $T : \mathbb{X} \to \mathbb{Y}$ of a G-invariant statistical model on $\mathbb{X}$ (e.g., for the $\mathcal{S}$-invariant rGP models the PEM is an equivariant descriptor by Theorem 2.1.2) induce a shape space $\mathbb{Y}/G$ (see Proposition 1.2.6). Our goal is to do inference on whether the descriptors of the underlying probability distribution of two sessions belong to the same equivalence class in $\mathbb{Y}/G$.

The main issue is, that even if the distributions $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta'}$ underlying two sessions χ_1 and χ_2 satisfy, that there exists a $g \in G$ such that if $X \sim \mathbb{P}_\theta$ then $g.X \sim \mathbb{P}_{\theta'}$, the group element g is usually unknown. Therefore g has to be estimated in order to compare equivariant descriptors of these distributions.

This can be done by bringing the descriptors into optimal position to each other with respect to a loss function L on $\mathbb{Y}$, i.e. estimating $g \in G$ by

$$\hat{g} \in \operatorname*{argmin}_{g \in G} L\big(g.T(\chi_1), T(\chi_2)\big)\,. \tag{3.1}$$

Recall that the action of g on the range of T is given by Lemma 1.2.6. In the terminology of shape analysis this is a Procrustes estimator on the shape space $\mathbb{Y}/G$.

Definition 3.0.1. *A function $L : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_{\geq 0}$ is called a* loss, *if $L(y_1, y_2) = 0$ for all $y_1, y_2 \in \mathbb{Y}$ if and only if $y_1 = y_2$. It is called* symmetric *if $L(y_1, y_2) = L(y_2, y_1)$ for all $y_1, y_2 \in \mathbb{Y}$ and it is called* G-invariant *if $L(g.y_1, g.y_2) = L(y_1, y_2)$ for all $y_1, y_2 \in \mathbb{Y}$ and all $g \in G$.*

Theorem 3.0.2. *Let L be a continuous loss and G compact. Then by setting*

$$\bar{L}\big([y_1],[y_2]\big) = \min_{g_1,g_2\in G} L(g_1.y_1, g_2.y_2) \tag{3.2}$$

for any $[y_1],[y_2] \in \mathbb{Y}/G$, L induces a well-defined loss $\bar{L}$ on the space $\mathbb{Y}/G$, i.e. for all $[y_1],[y_2] \in \mathbb{Y}/G$ we have that $\bar{L}\big([y_1],[y_2]\big) = 0$ if and only if $[y_1] = [y_2]$.

Additionally, if L is G-invariant, we even have

$$\bar{L}\big([y_1],[y_2]\big) = \min_{g\in G} L(g.y_1, y_2) \tag{3.3}$$

Proof. We first have to prove that the definition of $\bar{L}$ does not depend on the chosen representative. Therefore, let $\bar{y}_1 = g.y_1$ and $\bar{y}_2 = h.y_2$ for $g, h \in G$ be different representatives of the classes $[y_1],[y_2]$. We then obtain

$$\begin{aligned}
\bar{L}\big([\bar{y}_1],[\bar{y}_2]\big) &= \min_{g_1,g_2\in G} L(g_1.(g.y_1), g_2.(h.y_2)) \\
&= \min_{g_1,g_2\in G} L((g_1 g).y_1), (g_2 h).y_2) \\
&= \min_{g_1 g, g_2 h\in G} L((g_1 g).y_1), (g_2 h).y_2) \\
&= \bar{L}\big([y_1],[y_2]\big)
\end{aligned}$$

Now, it remains to prove that it defines a loss. First, let $\bar{L}\big([y_1],[y_2]\big) = 0$. Thus, since L is a loss there are $g_1, g_2 \in G$ such that $g_1.y_1 = g_2.y_2$. This, however, is equivalent to $[y_1] = [y_2]$.

The other way around follows from the observation that $[y_1] = [y_2]$ is only true, if it exists $h \in G$ such that $y_1 = h.y_2$ and hence

$$0 \le \bar{L}\big([y_1],[y_2]\big) = \min_{g_1,g_2\in G} L\big((g_1 h).y_2, g_2.y_2\big) \le L\big(y_2, y_2\big) = 0$$

□

The symmetry and invariance are necessary to assure that the Procrustes estimator does not depend on the choice of the reference shape, which will be formalized in the next definition. This property is also discussed in Vantini [2012] for time warping procedures.

Definition 3.0.3. *Let $L : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_{\ge 0}$ be a G-invariant loss. We call the Procrustes estimator* (3.1) inverse consistent*, if*

$$g^* \in \operatorname*{arginf}_{g\in G} L(g.y_1, y_2) \quad \Longrightarrow \quad (g^*)^{-1} \in \operatorname*{arginf}_{g\in G} L(g.y_2, y_1)\ .$$

Theorem 3.0.4. *Let L be continuous, symmetric and G-invariant and G compact. Then the Procrustes estimator* (3.1) *is inverse consistent.*

Proof. The statement follows directly from

$$\begin{aligned}
\min_{g\in G} L(g.y_2, y_1) &= \min_{g\in G} L\big(g^{-1}.y_1, g^{-1}.(g.y_2)\big) \\
&= \min_{g\in G} L\big(g^{-1}.y_1, y_2\big) \\
&= \min_{h=g^{-1}\in G} L(h.y_1, y_2)\ ,
\end{aligned}$$

where we used in the first equality the symmetry and invariance of L and in the second equality the properties of a group action. □

3.1 Time Warping: Removing of Different Velocities

Short overview of literature for Euclidean spaces. Removing variability in the time domain of functional data with values in $\mathbb{R}$ is a well studied problem. A good introduction to the problem using among other techniques landmark based registration is Ramsay and Li [1998]. Work on special semiparametric models including linear time shifts using M-estimation techniques can be found in Rønn [2001]. Nonparametric methods estimating warping function are Wang et al. [1997] using dynamic time warping or Kneip and Ramsay [2008] using registration to functional principal components. Unfortunately, these methods often fail to satisfy desirable properties as mentioned in Vantini [2012]. More criticism can be found in Srivastava et al. [2011a]. The latter introduced a method building on square root velocity functions and the Fisher-Rao metric, which is a sophisticated refinement of the dynamical time warping approach for speech recognition in Sakoe and Chiba [1978]. Its generalization to manifolds will be discussed in more detail in the next paragraph, since our time warping procedure uses some of their ideas.

Further readings can be found in the section about time warping of the review article Wang et al. [2015].

Curve registration using transported square-root vector fields. The methods proposed in the articles cited in the previous paragraph rely on a linear structure and hence cannot be used for temporal registration of curves in $SO(3)$. Until now only Su et al. [2014] provides a method applicable to curves with values in a manifold. We will shortly review their proposed method and discuss some unfavorable properties of this method. Note that there are a lot of recent publications (e.g., Celledoni and Eslitzbichler [2015], Bauer et al. [2015], Bauer et al. [2016], Amor et al. [2016]) applying and analyzing this method or slight modifications thereof. All these methods have in common that they build on the so called transported square-root vector field (TSRVF) representation of a curve, which in the opinion of the author of this thesis has some methodological shortcomings for curve registration in case of curved manifolds.

Let us start with introducing the TSRVF of a curve $\gamma : [0,1] \to \mathcal{M}$, where $(\mathcal{M}, g)$ is a Riemannian manifold. Choose a reference point $c \in \mathcal{M}$. The corresponding TSRVF is given by

$$q_\gamma(t) = \frac{P_{\gamma(t),c}\left(\dot{\gamma}(t)\right)}{\sqrt{\|\dot{\gamma}(t)\|_g}} \in T_c\mathcal{M}\,, \tag{3.4}$$

where $\dot{\gamma}(t)$ denotes the tangent vector at time t, $P_{\gamma(t),c} : T_{\gamma(t)}\mathcal{M} \to T_c\mathcal{M}$ denotes the parallel transport along the shortest geodesic between $\gamma(t)$ and c and $\|\cdot\|_g$ is the norm on $T_{\gamma(t)}\mathcal{M}$ induced by the Riemannian metric. Note that the TSRVF representation of a curve is well-defined as long as the reference point c does not belong to the cut locus of $\gamma(t)$ for all $t \in [0,1]$ and that it is possible to reconstruct the path γ from the tupel $(\gamma(0), q_\gamma(t))$.

The important property of the TSRVF is

$$q_{\gamma\circ\phi}(t) = q_\gamma(\phi(t))\sqrt{\dot{\phi}(t)}$$

for all $\phi \in \mathrm{Diff}^+[0,1]$. This identity follows from an application of the chain rule and implies that the metric on TSRVFs given by

$$d_T\big(q_\gamma(t), q_\eta(t)\big) = \sqrt{\int_0^1 |q_\gamma(t) - q_\eta(t)|_g^2 dt} \tag{3.5}$$

is invariant under $\mathrm{Diff}^+[0,1]$ and symmetric, which is the major point they emphasis about this so called *Fisher-Rao metric* for temporal registration of curves, since it implies

$$\inf_{\phi\in\mathrm{Diff}^+[0,1]} d_T\big(q_{\gamma\circ\phi}(t), q_\eta(t)\big) = \inf_{\phi\in\mathrm{Diff}^+[0,1]} d_T\big(q_\gamma(t), q_{\eta\circ\phi}(t)\big)\,.$$

The method they propose to find reparametrization between two curves γ and η is finding an approximation of

$$\operatorname*{arginf}_{\phi\in\mathrm{Diff}^+[0,1]} d_T\big(q_{\gamma\circ\phi}(t), q_\eta(t)\big)\,, \tag{3.6}$$

which is usually based on dynamic programming. Especially, note that, if $\gamma \neq \eta$, it is in general unknown, whether a minimizing $\phi \in \mathrm{Diff}^+[0,1]$ of (3.6) exists. However, if the curves γ, η have values in $\mathcal{M} = \mathbb{R}^D$ and at least one of the curves is piecewise linear, then there exists a minimizing $\phi \in \mathrm{Diff}^+[0,1]$ as was recently proven in Lahiri et al. [2015, Section 5, p.18]. In the case that a minimizer exists the same argument as in the proof of Theorem 3.0.4 yields the inverse alignment property (see Definition 3.0.3) for (3.6).

For $\mathcal{M} = \mathbb{R}^D$ this approach is identical with the one presented in Srivastava et al. [2011a]. However, in contrast to the $\mathbb{R}^D$ case this method does in general involve two types of non-trivial choices, which may lead to different estimated reparametrizations ϕ depending on those choices. The first type is the choice of a reference point $c \in \mathcal{M}$ and the second is the chosen identification of $T_{\gamma(t)}\mathcal{M}$ with $T_c\mathcal{M}$ using a parallel transport along a curve connecting $\gamma(t)$ and c. In $\mathbb{R}^D$ the TSRVF of a curve does neither depend on the curve, along which the parallel transport is carried out, nor on the reference point. In general, however, it is not evident why the parallel transport is chosen to be along geodesics. It seems also reasonable to first parallel transport the tangent vectors along the trajectory $\gamma(t)$ to the starting point $\gamma(0)$, say, and then use a parallel transport along geodesics to a reference point c.

Therefore we will propose in the next section an alternative time warping procedure for curves in Lie groups with bi-invariant metric fulfilling also the invariance under $\mathrm{Diff}^+[0,1]$ and the symmetry condition. The method still builds on the work of Srivastava et al. [2011a] and Su et al. [2014]. Its main advantage, however, is that the amount of arbitrary non-canonical choices is reduced to exactly two instead of having to choose from infinitely many as in the TSRVF framework, where the choice of a point $c \in \mathcal{M}$ and identifications of $T_{\gamma(t)}\mathcal{M}$ with $T_c\mathcal{M}$ are necessary. Moreover, we also propose a method without any choices (see Definition 3.1.1).

3.1.1 A New Time Warping Method for Lie-Group valued Curves

Note that all definitions and results presented in this section are valid for any Lie group admitting a bi-invariant metric without any changes in the proofs, but for the sake of a consistent presentation we will stay in the setting of curves in $SO(3)$.

An intrinsic loss function. In order to get a symmetric, $\mathcal{S}$-invariant and $\mathcal{S}$-compatible loss on $\mathcal{X}$ we will define a generalization of the total variation loss of functions to $SO(3)$.

Definition 3.1.1. *The* intrinsic length of difference curve losses *(ILLs) on $\mathcal{X}$ are given by*

$$\delta_{I,1}(\gamma,\eta) = \text{length}(\gamma\eta^{-1}) \quad \textit{and} \quad \delta_{I,2}(\gamma,\eta) = \text{length}(\gamma^{-1}\eta)\,, \tag{3.7}$$

for $\gamma,\eta \in \mathcal{X}$. Here the length is taken with respect to the bi-invariant metric on $SO(3)$.

In order to obtain a loss without the choice, where to place the inverse, we define the loss:

$$\delta_I(\gamma,\eta) = \tfrac{1}{2}\big(\delta_{I,1}(\gamma,\eta) + \delta_{I,2}(\gamma,\eta)\big) \tag{3.8}$$

Remark 3.1.2. *All these losses are indeed generalizations of the total variation loss between curves of the Lie group $\big(\mathbb{R}^D,+\big)$ with the standard Riemannian metric to curves in $SO(3)$, since for any curves $\gamma,\eta \in \mathcal{C}^1\big(I,\mathbb{R}^D\big)$ it holds that $\gamma\eta^{-1} = \gamma - \eta$ and $\gamma^{-1}\eta = \eta - \gamma$ and thus,*

$$\delta_{I,1}(\gamma,\eta) = \int_I \|\gamma'(t) - \eta'(t)\|\,dt = \delta_{I,2}(\gamma,\eta) = \delta_I(\gamma,\eta)\,. \tag{3.9}$$

In fact, the ILLs are symmetric and $\mathcal{S}$-invariant due to the biinvariance of the metric as will be derived now.

Theorem 3.1.3. *Let us denote with δ either $\delta_{I,1}$, $\delta_{I,2}$ or δ_I. Then the following hold*

(i) δ is symmetric.

(ii) δ is invariant under the action of $\mathcal{S}$, i.e.

$$\delta\big((\psi,\phi).\gamma,(\psi,\phi).\eta\big) = \delta(\gamma,\eta)$$

for all $(\psi,\phi) \in \mathcal{S}$ and all $\gamma,\eta \in \mathcal{X}$.

Proof. We only consider $\delta = \delta_{I,1}$, since the proof for the other two losses is analogous.

(i): Note that with the intrinsic distance $d_{SO(3)}$ on $SO(3)$ induced by the bi-invariant Riemannian metric the length $\delta(\gamma,\eta)$ of the differentiable curve $t \mapsto \zeta(t) = \gamma(t)\eta(t)^{-1}$ can be equivalently written as

$$\sup\left\{\sum_{k=0}^{K-1} d_{SO(3)}\big(\zeta(t_k),\zeta(t_{k+1})\big) \;\middle|\; K\in\mathbb{N},\ 0=t_0<t_1<\ldots<t_{k-1}<t_K=1\right\}.$$

The symmetry of δ is implied by $d_{SO(3)}\big(\zeta(t_k),\zeta(t_{k+1})\big) = d_{SO(3)}\big(\zeta(t_k)^{-1},\zeta(t_{k+1})^{-1}\big)$, which is a verified by the biinvariance of the metric.
(ii) follows directly from biinvariance and from the fact that the length of a curve does not depend on its parametrization (e.g., Lee [2013, Theorem 13.25, p.338]). □

Moreover, we can characterize the elements of $\mathcal{X}$ on which the ILLs attain the value zero.

Theorem 3.1.4. *Let $P \in SO(3)$ be arbitrary. Then*

$$\delta_{I,1}(P\gamma,\eta) = \delta_{I,1}(\gamma,\eta) \quad \text{and} \quad \delta_{I,2}(\gamma P,\eta) = \delta_{I,2}(\gamma,\eta)\,, \tag{3.10}$$

for all $\gamma,\eta \in \mathcal{X}$. Moreover, we have that $\delta_{I,1}(\gamma,\eta) = 0$ if and only if $\eta = \gamma P$ for some $P \in SO(3)$. Similarly, $\delta_{I,2}(\gamma,\eta) = 0$ if and only if $\eta = P\gamma$ for some $P \in SO(3)$.

Proof. We again without loss of generality consider the case $\delta = \delta_{I,1}$. Then $\delta(P\gamma,\eta) = \delta(\gamma,\eta)$ is again a direct consequence of the biinvariance of the metric.
For the second claim assume $\eta = P\gamma$. Then by Theorem 3.1.3(i) and the first statement of this theorem we obtain

$$\delta(\gamma,\eta) = \delta(\gamma,P\gamma) = \delta(\gamma,\gamma) = \operatorname{length}(\gamma\gamma^{-1}) = 0\,.$$

For the other way around assume that for all $P \in SO(3)$ we have that $\eta \neq P\gamma$. Then define the continuously differentiable curve $\zeta = \gamma\eta^{-1}$ and $\zeta(0) = \zeta_0$. The curve $\zeta\zeta_0^{-1}$ is continuously differentiable as well and by assumption $\zeta_0^{-1}\zeta$ cannot be constantly $I_{3\times3}$ for all $t \in I$. Therefore we find an $t' \in I$ such that $\zeta_0^{-1}\zeta(t') = R \neq I_{3\times3}$ for some $R \in SO(3)$. Thus, we have that

$$0 < d_{SO(3)}\big(I_{3\times3},R\big) \leq \operatorname{length}\big(\zeta_0^{-1}\zeta|_{[0,t']}\big) = \operatorname{length}\big(\zeta|_{[0,t']}\big) \leq \operatorname{length}\big(\zeta\big)\,,$$

where the equality stems from the biinvariance of the metric. □

Remark 3.1.5. *In view of the biomechanical application, the above theorem shows that $\delta_{I,1}$ and $\delta_{I,2}$ are invariant under any changes due to marker placement of one of the two orthogonal coordinate frames constructed from the measurement device!*

Registration of two curves. Following the general ideas presented in Kurtek et al. [2011], Srivastava et al. [2011b] and Srivastava et al. [2011a], which we already discussed earlier as we explained temporal registration using the TSRVF framework, temporal registration of two curves $\gamma_1,\gamma_2 \in \mathcal{X}$ is done by finding $\phi^* \in \mathrm{Diff}^+[0,1]$ approximating

$$\operatorname*{arginf}_{\phi\in\mathrm{Diff}^+[0,1]} \delta(\gamma_1,\gamma_2\circ\phi)\,. \tag{3.11}$$

Here again δ denotes either $\delta_{I,1}$, $\delta_{I,2}$ or δ_I. The specific loss is the only choice needed in our approach. In general one should pick δ_I, since it does treat left and

right translation of curves equally. In our gait data application we, however, use $\delta_{I,1}$ for convenience, since additional to the $\mathcal{S}$-invariance it is invariant against rotations of one of the coordinate systems, which are computed from the markers and which may vary due to replacement of the markers between sessions.

Note that using the same proof as Lemma 3.1.3 implies also the inverse alignment property for the ILLs, if the infimum is attained.

Theorem 3.1.6. *Let $\gamma, \eta \in \mathcal{X}$. Then we have that*

$$\phi^* \in \operatorname*{arginf}_{\phi \in \mathrm{Diff}^+[0,1]} \delta(\gamma_1, \gamma_2 \circ \phi) \Rightarrow (\phi^*)^{-1} \in \operatorname*{arginf}_{\phi \in \mathrm{Diff}^+[0,1]} \delta(\gamma_2, \gamma_1 \circ \phi)$$

Registration of multiple curves. The case of registration of multiple curves is done along the lines proposed in Srivastava et al. [2011b], where we again replace their loss derived from the Fisher-Rao metric by δ (i.e., one of the ILLs presented in Definition 3.1.1).

Given a set of curves $\gamma_1, ..., \gamma_N$ on $[0, 1]$, an initial candidate mean curve $\mu_0 = \gamma_1$, say, and a threshold $\varepsilon > 0$, the full algorithm of registration of these curves is given by

1. For $n \in \{1, 2, ..., N\}$, find ϕ_n by solving:

$$\operatorname*{argmin}_{\phi \in \mathrm{Diff}^+[0,1]} \delta\big(\mu_{k-1}, \gamma_n \circ \phi\big) .$$

   This is done using the dynamic program, which will be described in Section 3.1.2.

2. Update $t \mapsto \mu_k(t) \in \hat{\mu}_N\big(\gamma_1 \circ \phi, ..., \gamma_N \circ \phi, t\big)$ using the PESM (see Definition 2.1.1).

3. If $\min_{\phi \in \mathrm{Diff}^+[0,1]} \delta\big(\mu_k, \mu_{k-1} \circ \phi\big) < \varepsilon$, return μ_k and ϕ_n for $n \in \{1, ..., N\}$ else go to step 1.

Srivastava et al. [2011b] also included a step of finding a mean $\hat{\phi}_k$, which they call Karcher mean, of the warping functions $\phi_1, ...\phi_N$ in each iteration and would then solve

$$\operatorname*{argmin}_{\phi \in \mathrm{Diff}^+[0,1]} \delta\big(\mu_{k-1} \circ \hat{\phi}, \gamma_n \circ \phi\big).$$

We omit this step, since in our application any reparametrization of the mean μ_k suffices. Moreover, note that Karcher himself rejects the notion Karcher mean. Indeed, he dedicated an entire article (viz., Karcher [2014]) to this issue and insists on calling these means Riemannian center of mass.

3.1.2 Implementation Using a Dynamical Program

It is not straightforward to implement a solver of the minimization problem (3.11) and the article's Kurtek et al. [2011] and Su et al. [2014] only mention that their minimization problem, which differs from ours by the cost function, can be solved

using a dynamical program without explicitly providing this. Therefore we will now describe our implementation in detail, which is a variation of Dijkstra's algorithm (e.g., Cormen et al. [2001, Section 24.3]). For the sake of reducing computational time we implemented it in *C++* and use the *R* interface *Rcpp* to integrate it into our R-package *KneeStats*, which provides all algorithms and statistical test procedures reported in this thesis.

Our input data are two curves $\gamma, \eta \in \mathcal{X}$. Actually, due to discrete measurements we only have the values $\gamma(t'_1), ..., \gamma(t'_{K'}) \in SO(3)$ and $\eta(t''_1), ..., \eta(t''_{K''}) \in SO(3)$ for $0 = t'_1 < t'_2 < ... < t'_{K'} = 1$ and $0 = t''_1 < t''_2 < ... < t''_{K''} = 1$. Let γ^g and η^g denote piecewise geodesic interpolations of the discrete measurements, which are estimators of the curves γ and η. Here, of course we assume that the geodesic between two consecutive measurements is uniquely determined, which is always the case in our applications. The geodesic interpolation of the data could be replaced by a smoothing procedure using extrinsic kernel regression (see Lin et al. [2015]) or localized versions of the intrinsic polynomial regression proposed in Hinkle et al. [2014]. Since in our data example the observed curves are rather smooth, we did not implement such a procedure. However, note that in the literature for real valued functions (e.g., Wang et al. [1997] and Bigot et al. [2011]) it is recommended to smooth the data before applying the time warping procedure to obtain consistent estimators in some semiparametric models, if the observation error is not negligible.

In order to use a dynamical program to solve for an optimal diffeomorphism ϕ^* in continuous time we approximate ϕ^* using an optimal piecewise linear function

$$\hat{\phi}^*(t) = \sum_{k=1}^{K} \mathbb{1}_{[t_{k-1},t_k]}(t) \left(\frac{\tau_{j_k} - \tau_{j_{k-1}}}{t_k - t_{k-1}} (t - t_{k-1}) + \tau_{j_{k-1}} \right) . \tag{3.12}$$

Here $0 = t_1 < ... < t_{K-1} < t_K = 1$, $K \in \mathbb{N}$, is an equidistant partition of $I = [0, 1]$, $0 = \tau_1 < \tau_2 < ... < \tau_{J-1} < \tau_J = 1$ with $J = K+(K-1)w$, $w \in \mathbb{N}$, is an equidistant partition with $\tau_{k+(k-1)w} = t_k$ for all $k \in \{1, ..., K\}$. We will refer later on to these partitions as the t-partition and τ-partition and $\{j_1, ..., j_K\} \subset \{1, ..., J\}$, $K \in \mathbb{N}$. In order to have a good recovering of the diffeomorphism ϕ^* the t-partition has to be fine enough to capture the features of the curves γ and η. Moreover, we have to assume that $J \gg K$ such that ϕ^* can be approximated well by a function of the form (3.12).

Using the discretization of loss (3.7) (for convenience we use $\delta_{I,2}$) given by

$$\hat{\delta}_{I,2}\big(\gamma, \eta \circ \phi\big) = \sum_{k=1}^{K-1} d_{SO(3)}\Big(\gamma^g(t_k)^T \eta^g\big(\phi(t_k)\big), \gamma^g(t_{k+1})^T \eta^g\big(\phi(t_{k+1})\big)\Big)$$

for the t-partition and minimizing over piecewise linear functions $\phi = \hat{\phi}^*$ given by (3.12) reduces problem (3.11) to the problem of finding a minimizing sequence $1 = j_1 < j_2 < ... < j_{K-1} < j_K = J$ of

$$V_{K,J} = \min_{1=j_1<j_2<...<j_{K-1}<j_K=J} \sum_{k=1}^{K-1} d_{SO(3)}\Big(\gamma^g(t_k)^T \eta^g(\tau_{j_k}), \gamma^g(t_{k+1})^T \eta^g(\tau_{j_{k+1}})\Big) . \tag{3.13}$$

To shorten our notation let us define

$$f(t_k, \tau_{j_k}, t_{k+1}, \tau_{j_{k+1}}) = d_{SO(3)}\Big(\gamma^g(t_k)^T \eta^g(\tau_{j_k}), \gamma^g(t_{k+1})^T \eta^g(\tau_{j_{k+1}})\Big)\,.$$

Finding the minimizing sequence $\{j_k\}_{k\in\{1,\dots,K\}}$ of (3.13) can be achieved by a dynamical program, since rewriting (3.13) into

$$\begin{aligned} V_{K,J} &= \min_{1=j_1<j_2<\dots<j_{K-1}<j_K=J} \sum_{k=1}^{K-1} f(t_k, \tau_{j_k}, t_{k+1}, \tau_{j_{k+1}}) \\ &= \min_{1=j_1<j_2<\dots<j_{K-1}<j_K=J} \sum_{k=1}^{K-2} f(t_k, \tau_{j_k}, t_{k+1}, \tau_{j_{k+1}}) + f(t_{K-1}, \tau_{j_{K-1}}, t_K, \tau_J) \\ &= \min_{L<J} \min_{1=j_1<j_2<\dots<j_{K-1}=L} \sum_{k=1}^{K-2} f(t_k, \tau_{j_k}, t_{k+1}, \tau_{j_{k+1}}) + f(t_{K-1}, \tau_L, t_K, \tau_J) \\ &= \min_{L<J} V_{K-1,L} + f(t_{K-1}, \tau_L, t_K, \tau_J) \end{aligned} \tag{3.14}$$

yields a recursive formula for the minimal value $V_{K,J}$. Note that such a recursive formula is called the Bellmann equation of a minimization problem, which can be solved by dynamical programming.

Using the recursive formula (3.14) we construct a matrix $V = (V_{k,j})_{k,j=1}^{K,J} \in \mathbb{R}^{K\times J}$. As starting values for the recursion we use $V_{1,j} = 0$ for all $j \in \{1, \dots, J\}$. Moreover, note that we require that $j_1 = 1$ and necessarily $j > k$ in order to obtain an monotonically increasing function. Thus, we obtain the additional requirements:

- $V_{2,j} = f(t_1, \tau_1, t_2, \tau_j)$, for $j = 1, \dots, J$
- $V_{k,j} = \infty$ if $j > k$.

By construction the entry $V_{K,J}$ of V contains the minimal value of the discretization of loss (3.7) over all possible piecewise linear functions of the form (3.12) for a given t- and τ-partition. In order to obtain also the optimal choice $1 = j_1 < j_2 < \dots < j_{K-1} < j_K = J$ corresponding to the minimal value $V_{K,J}$, we have to store the index L obtaining the minimum of (3.14) in each construction step in a second matrix $W \in \mathbb{N}^{K\times J}$. Then backwards iteration starting with $j_{K-1} = W_{K,J}$ and computing the other values of $\{1, j_2, \dots, j_{K-2}, W_{K,J}, J\}$ using

$$j_k = W_{(k+1),j_{k+1}}$$

for $1 < k < K-1$, reconstructs the optimal sequence $\{j_k\}_{k\in\{1,\dots,K\}}$ leading to the minimal $V_{K,J}$.

Example 3.1.7. *In order to clarify the presentation assume that* $K = 5$ *and* $J = 9$. *Using the recursion* (3.14) *we obtain*

$$V = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \infty & V_{2,2} & V_{2,3} & V_{2,4} & V_{2,5} & V_{2,6} & V_{2,7} & V_{2,8} & V_{2,9} & V_{2,10} \\ \infty & \infty & V_{3,3} & V_{3,4} & V_{3,5} & V_{3,6} & V_{3,7} & V_{3,8} & V_{3,9} & V_{3,10} \\ \infty & \infty & \infty & V_{4,4} & V_{4,5} & V_{4,6} & V_{4,7} & V_{4,8} & V_{4,9} & V_{4,10} \\ \infty & \infty & \infty & \infty & V_{5,5} & V_{5,6} & V_{5,7} & V_{5,8} & V_{5,9} & V_{5,10} \end{pmatrix}.$$

Here the orange entries are initial values. The blue entries are computed using the orange entries. The purple entries are computed using the blue entries and the red entry is computed using the purple entries. All gray entries are not necessary, since they either do not produce a monotone increasing sequence $\{j_k\}_{k\in\{1,\dots,K\}}$ *of length* $K = 5$*, or they do not fulfill the boundary condition* $j_5 = 10$*. A possible matrix* W *could be*

$$W = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 4 & 4 & 6 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 2 & 6 & 4 & 6 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix}.$$

giving the optimal sequence $1 < 2 < 4 < 7 < 10$ (red entries).

3.2 Spatial Alignment: Removing of MP Effect

In this section we introduce our method to remove the marker placement effect between two curves $\gamma, \eta \in \mathcal{X}$ stemming from our gait analysis application. Thus, our task is to estimate a spatially optimally aligning $\psi \in \mathcal{I}_0\big(SO(3)\big)$ such that $\psi(\gamma)$ is close to η with respect to some loss L on $\mathcal{C}\big(I, SO(3)\big)$ (compare Section 1.1).

We will start this section with reviewing some approaches in the literature dealing with similar problems. Afterwards in Section 3.2.1 we will introduce our solution to spatial alignment problem (or marker placement problem) based on a transformation into a spherical (linear) regression problem (e.g., Mackenzie [1957]). In Section 3.2.2 we will prove that this procedure yields consistent estimators correcting the MP effect between two curves stemming from rGP models.

Approaches in the literature. The marker placement problem (also known as kinematic "cross-talk" effect) and its consequences for reliability and reproducibility of gait analysis experiments is well known in the biomechanics community (e.g., Woltring [1994], Besier et al. [2003], McGinley et al. [2009] and Osis et al. [2015]). Moreover, there are several procedures in the biomechanical literature trying to reduce this effect in the pre-recording phase (e.g., by improving the marker placement process using machines or different marker position protocols), among others Ehrig et al. [2007], Schache et al. [2006] and Noehren et al. [2010]. However, it seems that there are only relatively few articles trying to remove the MP effect in the post-recording phase (i.e., Woltring [1994], Rivest [2005], Ball and Greiner [2012] and Baudet et al. [2014]). Woltring proposed to estimate two rotations $(P, Q) \in \mathcal{I}_0\big(SO(3)\big)$ such that at the gait event of maximum knee flexion (see Figure 3) the x- and z-Euler angles are zeroed. Since these elements are in general not unique he chooses the smallest rotation in a specific sense (see Woltring [1994, p.1406]). Here the non uniqueness problem can in general be solved, if one takes more than a single observation from a gait cycle into account. This was done in Rivest [2005]. They estimate the two rotations minimizing the

quadratic variations in the x-and z-Euler angles over a complete gait cycle. However, they assume that the misalignments have small Euler angles and that the y-angle is already aligned. Ball and Greiner [2012] give a different purely algorithmic approach, which again assumes that the y-Euler angle is already aligned. The method of Baudet et al. [2014] performs a principal component analysis of the Euler angles and uses the transformation matrix to the eigenbasis to get a different representation of the Euler angles and thereby neglecting the geometry of the problem.

Note that our approach to the MP problem will differ conceptually from above approaches. All presented methods have in common that they estimate for each gait cycle $\gamma \in \mathcal{X}$ rotations $(P_\gamma, Q_\gamma) \in \mathcal{I}_0\big(SO(3)\big)$ such that $Q_\gamma \gamma P_\gamma$ fulfills special requirements like for example the variation in x- and z-Euler angles over a complete gait cycle are minimal, and then they compare the curves $Q_\gamma \gamma P_\gamma$ and $Q_\eta \eta P_\eta$ for $\gamma, \eta \in \mathcal{X}$.[1] This corresponds to the choice of a specific representative of the equivalence class $[\gamma] = \{\zeta \in \mathcal{X} \mid \exists \psi \in \mathcal{I}_0\big(SO(3)\big) : \ \zeta = \psi(\gamma)\}$ and is similar to the approach of using Bookstein coordinates in shape analysis (e.g., Bookstein [1997]). Our approach, however, aims to find $\psi = (P, Q) \in \mathcal{I}_0\big(SO(3)\big)$ such that $Q\gamma P$ is close to η in order to directly compare the gait cycles $\gamma, \eta \in \mathcal{X}$, which corresponds to a Procrustes approach to shape analysis.

In the statistical literature a scenario similar to the MP problem considered as a Procrustes problem, appears in Prentice [1989]. His objective is to estimate $\psi \in \mathcal{I}_0\big(SO(3)\big)$ given fixed and known $R_1, ..., R_N \in SO(D)$, $D \in \mathbb{N}$, and independent random variables $S_1, ..., S_N \in SO(D)$ satisfying $\mathbb{E}[S_n] = \psi(R_n)$ and some further assumptions on their distributions (e.g., isotropicity). He proves that the M-estimator

$$\hat{\psi} = (\hat{P}, \hat{Q}) \in \underset{(P,Q) \in \mathcal{I}_0\left(SO(3)\right)}{\operatorname{argmin}} \sum_{n=1}^{N} \|P R_n Q - S_n\|_F^2 \tag{3.15}$$

is a consistent estimator of ψ and derives its asymptotic distribution. Unfortunately, some of his proofs and results turned out to be wrong. Corrections are given in Chang and Rivest [2001]. Moreover, in Rivest and Chang [2006] these results are generalized for the special case of $D = 3$ to cover also some classes of non-isotropic distributions of the S_n's.

In fact, if we assume that the curves $\gamma, \eta \in \mathcal{X}$ are observed on a grid $0 = t_1 < t_2 < ... < t_K$, $K \in \mathbb{N}$, we could use (3.15) to compute a spatially aligning $\hat{\psi}$ between γ and η by

$$\hat{\psi} = (\hat{P}, \hat{Q}) \in \underset{(P,Q) \in \mathcal{I}_0\left(SO(3)\right)}{\operatorname{argmin}} \sum_{k=1}^{K} \|P\gamma(t_k)Q - \eta(t_k)\|_F^2 \,. \tag{3.16}$$

Solving this minimization problem, however, is challenging. Some gradient descent methods are derived in Prentice [1989] and are improved in Rivest and Chang [2006].

[1] Baudet et al. [2014] is a small exception, since he does not estimate $\psi \in \mathcal{I}_0\big(SO(3)\big)$. However, also his approach follows the Boosteinian paradigm.

3.2.1 Transformation to Spherical Regression

To the best of our knowledge the only method present in the literature, which could be used to solve the MP problem treated as a Procrustean problem is the method (3.16) inspired by the M-estimator (3.15) of Prentice [1989]. However, since its minimization is numerically challenging, we propose a different method, which is numerically easier to handle. This will be achieved by showing that the problem of finding $\psi \in \mathcal{I}_0\big(SO(3)\big)$ such that $\eta = \psi(\gamma)$ for $\gamma, \eta \in \mathcal{X}$ can be transformed using Theorem 3.2.4 into a spherical regression problem, which can be solved using a singular value decomposition.

To this end we want to lift a curve $\gamma \in \mathcal{X}$ to a continuous curve in S^3. Note that S^3 can be identified with the multiplicative group of unit quaternions via

$$S^3 \ni (x_1, x_2, x_3, x_4)^T \quad \leftrightarrow \quad x_1 + ix_2 + jx_3 + kx_4 \tag{3.17}$$

with $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$, $ki = -ik = j$, $jk = -kj = i$ (e.g., Chirikjian and Kyatkin [2000]). Moreover, the map

$$\begin{aligned} &\pi : \ S^3 \to SO(3) \\ &\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \mapsto \begin{pmatrix} 1-2x_3^2-2x_4^2 & 2(x_2x_3+x_1x_4) & 2(x_2x_4-x_1x_3) \\ 2(x_2x_3-x_1x_4) & 1-2x_2^2-2x_4^2 & 2(x_3x_4+x_1x_2) \\ 2(x_2x_4+x_1x_3) & 2(x_3x_4-x_1x_2) & 1-2x_2^2-2x_3^2 \end{pmatrix} \end{aligned} \tag{3.18}$$

is a double (even universal) cover of $SO(3)$ and a smooth surjective group homomorphism with the property $\pi(x) = \pi(-x)$ for all $x \in S^3$ (see Stuelpnagel [1964]). Thus, by the lifting property of covering maps (e.g., Lee [2013, Proposition A.77, p.616]) any curve $\gamma \in \mathcal{X}$ has exactly two continuous lifts $\tilde{\gamma}$ in S^3, each uniquely determined by the choice of the starting element from $\pi^{-1}\big(\gamma(0)\big)$.

Given a right inverse $r : SO(3) \to S^3$ of π (i.e., $\pi \circ r = id_{SO(3)}$) these lifts can be constructed explicitly. Let $\mathcal{D}_{r,\gamma} \subset [0,1]$ denote the discontinuity points of the curve $r(\gamma(t))$. Then there is a function $\epsilon_\gamma : [0,1] \to \{-1, 1\}$ with $\epsilon_\gamma(0) = 1$ and changing sign at each $t \in \mathcal{D}_{r,\gamma}$ such that $t \mapsto \epsilon_\gamma(t) r\big(\gamma(t)\big)$ is continuous in S^3. Obviously, the other continuous lift is given by $t \mapsto -\epsilon_\gamma(t) r\big(\gamma(t)\big)$.

In order to prove the Main Theorem 3.2.4 of this section, we restate some well known results connecting unit quaternions and the rotation group $SO(4)$ (see e.g., Mebius [2005]).

Lemma 3.2.1. *Let $p = (p_1, p_2, p_3, p_4)^T \in S^3$ and $q = (q_1, q_2, q_3, q_4)^T \in S^3$ be unit quaternions. Then there are unique*

$$R_p^l = \begin{pmatrix} p_1 & -p_2 & -p_3 & -p_4 \\ p_2 & p_1 & -p_4 & p_3 \\ p_3 & p_4 & p_1 & -p_2 \\ p_4 & -p_3 & p_2 & p_1 \end{pmatrix} \in SO(4)\,, \quad R_q^r = \begin{pmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{pmatrix} \in SO(4)$$

such that $p \cdot v = R_p^l v$ and $v \cdot x = R_q^r v$ for all $v \in \mathbb{R}^4$. Here "$\cdot$" denotes the quaternion multiplication.

Proof. Follows from Mebius [2005, Section 4.1]. □

Interestingly, it turns out that there is a converse to Lemma 3.2.1, i.e. any $R \in SO(4)$ can be represented as the multiplication of unit quaternions from the left and from the right and this representation is unique up to simultaneous change of sign of the quaternions. This result was discovered almost two centuries ago by van Elfrinkhof [1897] and rediscovered in the beginning of the last century by Bouman [1932]. We will restate it now.

Theorem 3.2.2. *For every $R \in SO(4)$ there exist unique up to simultaneous change of sign unit quaternions $p_R, q_R \in S^3$ such that $Rv = p_R \cdot v \cdot q_R$.*

Moreover, there is a smooth surjective group homomorphism

$$\begin{aligned} \pi_{SO(4)} : S^3 \times S^3 &\to SO(4) \\ (p,q) &\mapsto R^l_p R^r_q = R^r_q R^l_p \end{aligned} \tag{3.19}$$

with $\pi_{SO(4)}(p,q) = \pi_{SO(4)}(-p,-q)$.

Proof. A recent proof of the first part by pure matrix computations can be found in Mebius [2005]. The second claim follows from Lemma 3.2.1. □

Remark 3.2.3. *By the above theorem we have $SO(4) \cong \big(S^3 \times S^3\big)/\ker\big(\pi_{SO(4)}\big) = \big(S^3 \times S^3\big)/\{I_{4\times4}, -I_{4\times4}\}$ as groups.*

We are now ready to state and prove the main result of this section, which shows that the action of $\mathcal{I}_0\big(SO(3)\big)$ on $\mathcal{X}$ lifts to the canonical left action of $SO(4)$ on continuously lifted curves $\tilde{\gamma} \in \mathcal{C}\big(I, S^3\big)$.

Theorem 3.2.4. *Let $(P,Q) \in \mathcal{I}_0\big(SO(3)\big)$ and $\gamma \in \mathcal{X}$ be arbitrary and let $\widetilde{P\gamma Q^T}$, $\tilde{\gamma}$ be fixed continuous lifts of the two paths. Then there exists a unique $R \in SO(4)$ with the property*

$$\widetilde{P\gamma Q^T}(t) = R\tilde{\gamma}(t) \tag{3.20}$$

for all $t \in [0,1]$.

Proof. Let $\gamma \in \mathcal{X}$ and $(P,Q) \in SO(3)\times SO(3)$ be arbitrary and choose continuous lifts $\tilde{\gamma}$ and $\widetilde{P\gamma Q^T}$ of γ and $P\gamma Q^T$.

Choose any right inverse r of the homomorphism π given in (3.18). Then there are $p, q \in S^3$ such that $p = r(P)$ and $q^{-1} = r(Q^T)$ and there is a function $\epsilon_\gamma : [0,1] \to \{-1,1\}$ such that $\epsilon_\gamma r(\gamma)$ is the continuous lift $\tilde{\gamma}$. Further, since π is a group homomorphism and $\pi(x) = \pi(-x)$ for all $x \in S^3$, we obtain

$$r\big(P\gamma Q^T\big)(t) = \varepsilon(t) p \cdot r\big(\gamma(t)\big) \cdot q^{-1} \tag{3.21}$$

with $\varepsilon(t) \in \{-1,1\}$ for all $t \in [0,1]$. Thus, multiplication of (3.21) on both sides with $\varepsilon(t)\epsilon_\gamma(t)$ implies that the path $\epsilon_\gamma \varepsilon r\big(P\gamma Q^T\big)$ is continuous in S^3. Since $\pi\Big(\epsilon_\gamma \varepsilon r\big(P\gamma Q^T\big)\Big) = P\gamma Q^T$ using the property $\pi(x) = \pi(-x)$ for all $x \in S^3$, there exists by the unique lifting property of covering maps a time independent

$\kappa \in \{-1,1\}$ such that $\widetilde{P\gamma Q^T} = \kappa\epsilon_\gamma \varepsilon r\big(P\gamma Q^T\big)$. Finally, by Theorem 3.2.2 we obtain

$$\widetilde{P\gamma Q^T} = \kappa p\tilde{\gamma}q^{-1} = R\tilde{\gamma} \tag{3.22}$$

with a unique $R = R^l_{\kappa p} R^r_{q^{-1}}$. □

Inspired by the above theorem, our goal is to estimate for two curves $\gamma, \eta \in \mathcal{X}$ an $(P,Q) \in \mathcal{I}_0\big(SO(3)\big)$ optimally aligning these curves with respect to some continuous loss $L : \mathcal{C}\big(I, S^3\big) \times \mathcal{C}\big(I, S^3\big) \to \mathbb{R}_{\geq 0}$ by choosing continuous lifts $\tilde{\gamma}$ and $\tilde{\eta}$ to S^3 and computing

$$\hat{R} \in \operatorname*{argmin}_{R \in SO(4)} L\big(R\tilde{\gamma}, \tilde{\eta}\big)\,. \tag{3.23}$$

This, however, in order to be reasonable requires that we can map the estimator $\hat{R} \in SO(4)$ to an estimator $\big(\hat{P}, \hat{Q}\big) \in \mathcal{I}_0\big(SO(3)\big)$ of (P,Q) and that the resulting estimator $\big(\hat{P}, \hat{Q}\big)$ does not depend on the choice of the lifts. To achieve this we introduce the following map.

Definition 3.2.5. *Let us define*

$$\Pi : SO(4) \xrightarrow{r_{SO(4)}} S^3 \times S^3 \xrightarrow{\pi\times\pi} SO(3) \times SO(3)\,, \tag{3.24}$$

where $r_{SO(4)}$ is any right inverse of the map $\pi_{SO(4)}$ given in (3.19)*. Note that by $\pi(x) = \pi(-x)$ for all $x \in S^3$ any choice of a right inverse $r_{SO(4)}$ results in the same map Π.*

The following properties of Π will imply that our yet to be defined estimator $\big(\hat{P}, \hat{Q}\big)$ does not depend on the choice of the lifts (see Theorem 3.2.8) and that this estimator fulfills the inverse alignment property (see Definition 3.0.3 and Theorem 3.2.9).

Lemma 3.2.6. *For any $R \in SO(4)$ we have that $\Pi(R) = \Pi(-R)$ and $\Pi(R^{-1}) = \Pi(R)^{-1}$.*

Proof. Let $R \in SO(4)$ be arbitrary. Fix a right inverse $r_{SO(4)}$ of $\pi_{SO(4)}$ and let $r_{SO(4)}(R) = (p,q) \in S^3 \times S^3$. Since $\pi_{SO(4)}$ is a group homomorphism with $\pi_{SO(4)}(p,q) = \pi_{SO(4)}(-p,-q)$ for all $(p,q) \in S^3$, we obtain for every $R, S \in SO(4)$ that there is $\kappa \in \{1,-1\}$ such that

$$r_{SO(4)}(RS) = \kappa r_{SO(4)}(R) r_{SO(4)}(S)\,. \tag{3.25}$$

The claim $\Pi(R) = \Pi(-R)$ can now be deduced using equation (3.25), since

$$r_{SO(4)}(-R) = \big((\pm\kappa,0,0,0)^T, (\mp\kappa,0,0,0)^T\big) r_{SO(4)}(R) = (\pm\kappa p, \mp\kappa q)\,,$$

where we used the matrices R^l_p and R^r_q of Lemma 3.2.1 to obtain $r_{SO(4)}\big(-I_{4\times 4}\big) = \big((\pm 1,0,0,0),(\mp 1,0,0,0)\big)$. The particular sign depends on the choice of the right inverse. However, since $\big(\pi(\kappa p), \pi(-\kappa q)\big) = \big(\pi(-\kappa p), \pi(\kappa q)\big) = \big(\pi(p), \pi(q)\big)$ the claim follows.

We now want to prove $\Pi(R^{-1}) = \Pi(R)^{-1}$. Therefore, note that equation (3.25) implies $r_{SO(4)}\big(R^{-1}\big) = \kappa\big(r_{SO(4)}(R)\big)^{-1}$ for $\kappa \in \{1, -1\}$. Thus,

$$\Pi\big(R^{-1}\big) = (\pi\times\pi)\big(r(R)^{-1}\big) = \Big(\pi\big(\kappa p^{-1}\big), \pi\big(\kappa q^{-1}\big)\Big) = \Big(\pi(p)^{-1}, \pi(q)^{-1}\Big) = \Pi(R)^{-1}\,.$$

Here we used in the third equality $\pi(x^{-1}) = \pi(x)^{-1}$, which is a direct consequence of π being a group homomorphism, and $\pi(x) = \pi(\kappa x)$ for all $x \in S^3$. □

Theorem 3.2.7. *Let $(P, Q) \in \mathcal{I}_0\big(SO(3)\big)$ and $\gamma \in \mathcal{X}$ be arbitrary and let $\widetilde{P\gamma Q^T}$, $\tilde{\gamma}$ be fixed continuous lifts of the two paths and let $R \in SO(4)$ be the unique element satisfying equation* (3.20) *in Theorem 3.2.4. Then $\Pi(R) = (P, Q^T)$.*

Proof. Using the notation of Theorem 3.2.4 we can by equation (3.22) choose a right inverse $r_{SO(4)}$ of $\pi_{SO(4)}$ such that $r_{SO(4)}(R) = (\kappa p, q^T)$. Further by definition of p and q we obtain that $\pi(p) = P$ and $\pi(q^{-1}) = Q^T$, which yields the required statement. □

Now, Theorem 3.2.4 and 3.2.7 suggest that instead of minimizing a loss over $\mathcal{I}_0\big(SO(3)\big) = SO(3) \times SO(3)$ we equivalently can choose continuous lifts of the curves γ and η to S^3, solve the minimization problem given in (3.23) with some loss $L : \mathcal{C}\big(I, S^3\big) \times \mathcal{C}\big(I, S^3\big) \to \mathbb{R}_{\geq 0}$ and mapping the solution with Π into $\mathcal{I}_0\big(SO(3)\big)$, i.e. our Procrustes estimator for optimally spatially aligning $\gamma, \eta \in \mathcal{X}$ with respect to $\mathcal{I}_0\big(SO(3)\big)$ is given by

$$\big(\hat{P}, \hat{Q}\big) \in \Pi\left(\operatorname*{argmin}_{R\in SO(4)} L\big(R\tilde{\gamma}, \tilde{\eta}\big)\right). \tag{3.26}$$

Since the continuity of L implies the continuity of $R \mapsto L(R\gamma, \eta)$ and $SO(4)$ is compact, the set on the right hand side of (3.26) is never empty. Moreover, this estimators, indeed, fulfill the requirement that they do not depend on the choice of the continuous lifts.

Theorem 3.2.8. *The Procrustes estimators given in* (3.26) *do not depend on the particular choices of the continuous lifts.*

Proof. Without loss of generality we consider the case, where we have two continuous lifts $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ of $\gamma \in \mathcal{X}$. Note that $\tilde{\gamma}_1 = -\tilde{\gamma}_2$. For any fixed continuous lift $\tilde{\eta}$ of $\eta \in \mathcal{X}$ we obtain

$$\operatorname*{argmin}_{R\in SO(4)} L\big(R\tilde{\gamma}_1(t), \tilde{\eta}(t)\big) = \operatorname*{argmin}_{R\in SO(4)} L\big(-R\tilde{\gamma}_2(t), \tilde{\eta}(t)\big) = -\operatorname*{argmin}_{R\in SO(4)} L\big(R\tilde{\gamma}_2(t), \tilde{\eta}(t)\big)\,,$$

since $-I_{4\times 4} \in SO(4)$. Hence any minimizer R of the left hand side with respect to the lift $\tilde{\gamma}_1$ yields a minimizer $-R$ of the Procrustes loss with the lift $\tilde{\gamma}_2$. Thus, by Lemma 3.2.6 we obtain that $\Pi(R) = \Pi(-R)$, which proves the independence of the estimator (3.26) from the particular chosen lift. □

In the beginning of this chapter we stated that any $\mathcal{I}_0\big(SO(3)\big)$-invariant loss satisfies the inverse alignment property, Definition 3.0.3 and Theorem 3.0.4. We will now state an analogous result for the estimator (3.26).

Theorem 3.2.9. *Let $L : \mathcal{C}\big(I, S^3\big) \times \mathcal{C}\big(I, S^3\big) \to \mathbb{R}_{\geq 0}$ be $SO(4)$-invariant and symmetric. Then the estimator* (3.26) *fulfills the inverse alignment property, i.e.*

$$\big(\hat{P}, \hat{Q}\big) \in \Pi\left(\operatorname*{argmin}_{R \in SO(4)} L\big(R\tilde{\gamma}, \tilde{\eta}\big)\right) \quad \Rightarrow \quad \big(\hat{P}^{-1}, \hat{Q}^{-1}\big) \in \Pi\left(\operatorname*{argmin}_{R \in SO(4)} L\big(R\tilde{\eta}, \tilde{\gamma}\big)\right).$$

Proof. Theorem 3.0.4 implies

$$R^* \in \operatorname*{argmin}_{R \in SO(4)} L\big(R\tilde{\gamma}, \tilde{\eta}\big) \quad \Rightarrow \quad \big(R^*\big)^{-1} \in \operatorname*{argmin}_{R \in SO(4)} L\big(R\tilde{\eta}, \tilde{\gamma}\big)$$

and by Lemma 3.2.6 we have $\Pi\Big(\big(R^*\big)^{-1}\Big) = \Pi(R^*)^{-1}$ giving the claim. □

We now want to introduce the $SO(4)$-invariant, symmetric loss L, which we use in our application. Note that the above discussion reveals that estimating an optimally spatially aligning $\psi \in \mathcal{I}_0\big(SO(3)\big)$ for two curves $\gamma, \eta \in \mathcal{X}$ is closely related to spherical (linear) regression. To make this point more clear and since our loss is inspired by a solution of the spherical (linear) regression problem, we will shortly explain spherical regression.

The problem in spherical (linear) regression is the following (e.g., Mackenzie [1957]): let $x = (x_1, ..., x_K) \in S^3$ and $y = (y_1, ..., y_K) \in S^3$ such that $y_k = R^* x_k$ for $k \in \{1, ..., K\}$ and $R^* \in SO(4)$, can we compute R^* knowing only x and y? Up to the fact, that R^* is not necessarily unique the answer is yes. In Mackenzie [1957] and Stephens [1979] it is shown that

$$R^* \in \operatorname*{argmin}_{R \in SO(3)} \sum_{k=1}^{K} \|Rx_k - y_k\|^2, \tag{3.27}$$

and that R^* is unique, if the matrix $K^{-1} \sum_{k=1}^{K} y_k x_k^T$ has rank greater than two. Later Chang [1986] considered the same problem, but assumed that the y_k are i.i.d. random variables having a distribution centered at $R^* x_k$ and proved that (3.27) under some suitable conditions on the distributions of the y_k's yields a consistent estimator for R^*. He also generalized this observations to the errors in variables case (i.e., additionally assuming that also $x_1, ..., x_K$ are i.i.d. random variables) in Chang [1989].

After this discussion it is obvious that our approach of continuously lifting paths $\gamma, \eta \in \mathcal{X}$ to paths in S^3 and using the estimator (3.26) two bring them into optimal position to each other, transforms estimating $(P, Q) \in \mathcal{I}_0\big(SO(3)\big)$ into a spherical (linear) regression problem for curves in S^3. Therefore in order to get a generalization of (3.27) to curves, we propose to use the loss

$$\begin{aligned} L_2 : \mathcal{C}\big(I, S^3\big) \times \mathcal{C}\big(I, S^3\big) &\to \mathbb{R}_{\geq 0} \\ (\tilde{\gamma}, \tilde{\eta}) &\mapsto \int_0^1 \|\tilde{\gamma}(t) - \tilde{\eta}(t)\|^2 \, dt, \end{aligned} \tag{3.28}$$

which yields the Procrustes estimator

$$\big(\hat{P}, \hat{Q}\big) \in \Pi\left(\operatorname*{argmin}_{R \in SO(4)} \int_0^1 \|R\tilde{\gamma}(t) - \tilde{\eta}(t)\|^2 \, dt\right) \tag{3.29}$$

estimating optimally spatially aligning $(P,Q) \in \mathcal{I}_0\big(SO(3)\big)$ for two curves $\gamma, \eta \in \mathcal{X}$. As before the set on the right hand side of (3.29) is never empty, since $R \mapsto L_2(R\gamma, \eta)$ is continuous and $SO(4)$ is compact. A uniqueness result will be given in Theorem 3.2.11. Moreover, note that L_2 is $SO(4)$-invariant and symmetric. Therefore the assumptions of Theorem 3.2.8 and Theorem 3.0.4 are fulfilled implying that (3.29) does not depend on the choice of the continuous lifts and is inverse consistent with respect to the group action of $\mathcal{I}_0\big(SO(3)\big)$ on $\mathcal{X}$.

A remarkable property of the estimators (3.27) for the spherical (linear) regression problem is that they are easy to compute using a singular value decomposition as was proven in Stephens [1979]. A better accessible version of the proof can be found in Umeyama [1991]. A similar result for the Procrustes estimator (3.29) can be obtained using their result as we will deduce now.

Lemma 3.2.10. *Let $\tilde{\gamma}, \tilde{\eta} \in \mathcal{C}\big(I, S^3\big)$ and let $X = \int_0^1 \tilde{\eta}(t)\tilde{\gamma}(t)^T \, dt$. Then the minimizers of*

$$\operatorname*{argmin}_{R \in SO(4)} \int_0^1 \|R\tilde{\gamma}(t) - \tilde{\eta}(t)\|^2 \, dt$$

are given by

$$\hat{R} = USV^T,$$

where $U, V \in O(4)$ together with $D = \operatorname{diag}(\lambda_1, ..., \lambda_4)$, $\lambda_1 \geq ... \geq \lambda_4 \geq 0$, form a SVD of X, i.e. $X = UDV^T$, and

$$S = \begin{cases} I_{4\times 4} & \text{if } \det(U)\det(V) = 1 \\ \operatorname{diag}(1,1,1,-1) & \text{if } \det(U)\det(V) = -1 \end{cases}. \tag{3.30}$$

Moreover, if additionally $\operatorname{rank}(X) > 2$, then $\hat{R}$ is unique.

Proof. By expanding the square we obtain

$$\begin{aligned} \operatorname*{argmin}_{R \in SO(4)} \int_0^1 \|R\tilde{\gamma}(t) - \tilde{\eta}(t)\|^2 \, dt &= \operatorname*{argmax}_{R \in SO(4)} \int_0^1 \tilde{\eta}(t)^T R \tilde{\gamma}(t) \, dt \\ &= \operatorname*{argmax}_{R \in SO(4)} \operatorname{tr}\left(\mathrm{R} \int_0^1 \tilde{\gamma}(\mathrm{t})\tilde{\eta}(\mathrm{t})^{\mathrm{T}} \, \mathrm{dt} \right) \\ &= \operatorname*{argmin}_{R \in SO(4)} \left\| I_{4\times 4} - R \int_0^1 \tilde{\gamma}(t)\tilde{\eta}(t)^T \, dt \right\|_F^2 . \end{aligned}$$

Thus, setting $m = n = 4$, $A = I_{4\times 4}$ and $B = \int_0^1 \tilde{\gamma}(t)\tilde{\eta}(t)^T \, dt$ the claim follows from Umeyama [1991, Lemma 1]. □

Theorem 3.2.11. *Let $\gamma, \eta \in \mathcal{C}\big(I, SO(3)\big)$ and $\tilde{\gamma}, \tilde{\eta} \in \mathcal{C}\big(I, S^3\big)$ be any choice of continuous lifts to S^3 and $X = \int_0^1 \tilde{\eta}(t)\tilde{\gamma}(t)^T \, dt$. Then the Procrustes estimator (3.29) consists of solutions of the form $\Pi\big(USV^T\big)$, where $U, V \in O(4)$ together with $D = \operatorname{diag}(\lambda_1, ..., \lambda_4)$, $\lambda_1 \geq ... \geq \lambda_4 \geq 0$, form a SVD of X, i.e. $X = UDV^T$, and*

$$S = \begin{cases} I_{4\times 4} & \text{if } \det(U)\det(V) = 1 \\ \operatorname{diag}(1,1,1,-1) & \text{if } \det(U)\det(V) = -1 \end{cases}. \tag{3.31}$$

If additionally $\operatorname{rank}(X) > 2$, then the Procrustes estimator (3.29) is unique.

Proof. This follows from Lemma 3.2.10 and the fact that $\operatorname{rank}(X) > 2$ does not depend on the chosen continuous lifts, since X and $-X$ have the same rank. □

In our application we do not observe the complete paths $\gamma, \eta \in \mathcal{X}$, but their values $\gamma(t_1), ..., \gamma(t_K), \eta(t_1), ..., \eta(t_K)$ on an equidistant partition $0 = t_1 < t_2 < ... < t_K = 1$. Note that, if the equidistant partition is fine enough, geodesical interpolation between $\gamma(t_{k-1})$ and $\gamma(t_k)$ ($\eta(t_{k-1})$ and $\eta(t_k)$ respectively) yields a continuous curve $\gamma^g \in \mathcal{C}\big(I, SO(3)\big)$, which satisfies $\tilde{\gamma}^g(t_k) = \tilde{\gamma}(t_k)$ for all $k \in \{1, ..., K\}$, if the continuous lifts $\tilde{\gamma}^g$ and $\tilde{\gamma}$ to S^3 are chosen such that $\tilde{\gamma}^g(0) = \tilde{\gamma}(0)$. Therefore discretization of the loss L_2 yields the estimators

$$\big(\hat{P}, \hat{Q}\big) = \Pi\left(\operatorname*{argmin}_{R \in SO(4)} K^{-1} \sum_{k=1}^{K} \|R\tilde{\gamma}(t_k) - \tilde{\eta}(t_k)\|^2 \right). \tag{3.32}$$

Here, of course, the estimator $\big(\hat{P}, \hat{Q}\big)$ depends on $K \in \mathbb{N}$. However, if we assume that K tends to infinity, i.e. the equidistant partition of $I = [0, 1]$ becomes finer, we have that

$$\lim_{K \to \infty} K^{-1} \sum_{k=1}^{K} \|R\tilde{\gamma}(t_k) - \tilde{\eta}(t_k)\|^2 = \int_0^1 \|R\tilde{\gamma}(t) - \tilde{\eta}(t)\|^2 \, dt \,, \tag{3.33}$$

and hence in that limit the estimators $\big(\hat{P}, \hat{Q}\big)$ do not depend on the measurement points.

Finally, note that without any major changes in the proofs analogue versions of Theorems 3.2.8, 3.2.9 and 3.2.11 are true for the estimators (3.32) with the only change that we use

$$X = K^{-1} \sum_{k=1}^{K} \tilde{\eta}(t_k) \tilde{\gamma}(t_k)^T \,.$$

3.2.2 A Strongly Consistent Estimator in rGP Models

We will now show that having two sessions $\gamma_1, ..., \gamma_N$ and $\eta_1, ..., \eta_M$, $N, M \in \mathbb{N}$, the estimator (3.32) applied to the PESMs $\hat{\gamma}_N$ and $\hat{\eta}_M$ leads to a strongly consistent estimator for the rotation (P, Q) aligning those sessions.

The next Lemma is a useful tool showing strong consistency, if one has M-estimators, and was stated in the prove of Chang [1986, Lemma 2] without a proof. For convenience we will give a proof here.

Lemma 3.2.12. *Let $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ be continuous and $\mathbb{X}$ compact and assume $f(x, y_0)$ has a unique maximum for a specific $y_0 \in \mathbb{Y}$. Suppose y_n converges to y_0 and each x_n is a choice of a maximum of $f(x, y_n)$. Then $x_n \to x_0$.*

Proof. Let (x_{n_k}) be any convergent subsequence of x_n. By the definition of x_n we have that $f(x_{n_k}, y_{n_k}) \geq f(x, y_{n_k})$ for all k and all $x \in X$. Hence if k tends to infinity by the continuity of f and the uniqueness of x_0 we have that x_{n_k} converges to x_0.

Now assume x_n does not converge to x_0. Then there is a subsequence (x_{n_j}) such that $|x_{n_j} - x_0| > \varepsilon$. But this is a contradiction, since by compactness this sequence again has a convergent subsequence, which by the above argumentation, converges to x_0. Thus x_n must converge to x_0. □

Theorem 3.2.13. *Assume $\gamma_1, ..., \gamma_N$, $N \in \mathbb{N}$, is a sample of rGP model (see Definition 1.3.3) with center curve γ_0 and $\eta_1, ..., \eta_M$, $M \in \mathbb{N}$, a sample of a rGP with center curve η_0. Assume that the Gaussian processes of the rGP models fulfill Assumption* (2.6) *of Theorem 2.1.10. Let us denote with $\tilde{\gamma}_0, \tilde{\eta}_0 \in \mathcal{C}\big(I, S^3\big)$ fixed continuous lifts of the center curves and with $0 = t_1 < t_2 < ... < t_K = 1$, $K \in \mathbb{N}$, an equidistant partition of I.*

If additionally the following Assumption is fulfilled

$$\operatorname{rank}\left(\sum_{k=1}^{K} \tilde{\eta}_0(t_k)\tilde{\gamma}_0(t_k)^T\right) > 2\,,$$

then for any choice of the estimators proposed in equation (3.32) *aligning the PEMs we obtain*

$$\Pi\left(\operatorname*{argmin}_{R\in SO(4)} \sum_{k=1}^{K} \left\| R\widetilde{\hat{\gamma}_N}(t_k) - \widetilde{\hat{\eta}_M}(t_k)\right\|^2\right) \ni \left(\hat{P}_{N,M}, \hat{Q}_{N,M}\right) \xrightarrow{N,M\to\infty} \left(P^*, Q^*\right) \;\; a.s.\,,$$

where $\widetilde{\hat{\gamma}_N}, \widetilde{\hat{\eta}_M}$ are continuous lifts to S^3 of the PEMs and

$$\left(P^*, Q^*\right) = \Pi\left(\operatorname*{argmin}_{R\in SO(4)} \sum_{n=1}^{N} \|R\tilde{\gamma}_0(t_n) - \tilde{\eta}_0(t_n)\|^2\right).$$

Note that $\left(\hat{P}_{N,M}, \hat{Q}_{N,M}\right)$ and $\left(P^*, Q^*\right)$ do not depend on the particular choices of the continuous lifts.

Proof. By the assumption we have that Theorem 2.1.10 yields that the PESMs $\hat{\gamma}_N$ and $\hat{\eta}_M$ are strongly consistent estimators of γ and η for $N, M \to \infty$. Therefore there is a $\Omega' \subset \Omega$ with $\mathbb{P}\big(\Omega'\big) = 1$, on which the consistency holds for both curves simultaneously.

Let $\omega \in \Omega'$ be arbitrary. Then there exists by part (i) of Theorem 2.1.10 N_ω and M_ω such that $\hat{\gamma}_N(\omega, \cdot), \hat{\eta}_M(\omega, \cdot) \in \mathcal{X}$ for all $N > N_\omega$ and all $M > M_\omega$. Therefore the continuous lifts $\widetilde{\hat{\gamma}_N}(\omega, \cdot)$ and $\widetilde{\hat{\eta}_M}(\omega, \cdot)$ are well-defined then.

Therefore in order to prove the almost sure convergence

$$\Pi\left(\operatorname*{argmin}_{P\in SO(4)} \sum_{k=1}^{K} \left\| P\widetilde{\hat{\gamma}_N}(\omega, t_k) - \widetilde{\hat{\eta}_M}(\omega, t_k)\right\|^2\right) \xrightarrow{N,M\to\infty} \left(P^*, Q^*\right)$$

it remains to show that we can apply Lemma 3.2.12 for the chosen $\omega \in \Omega'$.

To this end with the notations of Lemma 3.2.12 let $\mathbb{X} = SO(4)$, which in fact is compact, and $\mathbb{Y} = \mathbb{R}^{4K} \times \mathbb{R}^{4K}$. Define the continuous map

$$f\big(R, (x_1, ..., x_K), (y_1, ..., y_K)\big) = -\sum_{k=1}^{K} \|Rx_k - y_k\|^2 \,.$$

Since $\operatorname{rank}\big(\sum \tilde{\gamma}_0(t_k)\tilde{\eta}_0(t_k)^T\big) > 2$ there is a unique solution

$$R^* = \underset{R\in SO(4)}{\operatorname{argmax}} f\Big(R, \big(\tilde{\gamma}_0(t_1), ..., \tilde{\gamma}_0(t_K)\big), \big(\tilde{\eta}_0(t_1), ..., \tilde{\eta}_0(t_K)\big)\Big)$$

by Lemma 3.2.10. Moreover, by Theorem 2.1.10 we have $\hat{\gamma}_N(\omega, t) \to \gamma_0(t)$ and $\hat{\eta}_M(\omega, t) \to \eta_0(t)$ for $N, M \to \infty$ and all $t \in I$. The latter implies that we can choose continuous lifts such that $\widetilde{\hat{\gamma}_N}(\omega, t_k) \to \tilde{\gamma}_0(t_k)$ and $\widetilde{\hat{\eta}_M}(\omega, t_k) \to \tilde{\eta}_0(t_k)$ for $N, M \to \infty$ and all $k \in \{1, ..., K\}$. Let

$$R_{N,M}(\omega) \in \underset{R\in SO(4)}{\operatorname{argmax}} f\Big(R, \big(\widetilde{\hat{\gamma}_N}(\omega, t_1), ..., \widetilde{\hat{\gamma}_N}(t_K)\big), \big(\widetilde{\hat{\eta}_M}(\omega, t_1), ..., \widetilde{\hat{\eta}_M}(t_K)\big)\Big), \tag{3.34}$$

be any choice. Then by Lemma 3.2.12 we obtain $R_{N,M}(\omega) \to R^*$ and hence $\Pi\big(R_{N,M}(\omega)\big) \to \Pi\big(R^*\big)$ for $N, M \to \infty$. □

Remark 3.2.14. *1. An analogous version of Theorem 3.2.13 holds true for the estimator* (3.29)*, if* $Y = \mathcal{C}^1\big(I, S^3\big) \times \mathcal{C}^1\big(I, S^3\big)$ *endowed with the maximum norm is used in the proof and* $f\big(R, (\gamma, \eta)\big) = -\int_0^1 \|R\gamma(t) - \eta(t)\|^2 \, dt$.

2. The only ingredient of the rGP model we need in the proof of Theorem 3.2.13 is the uniform convergence of the PESM to the unique PEM. Thus, any process fulfilling a Theorem similar to 2.1.10 *will yield a consistent estimator for the aligning element* $\psi \in \mathcal{I}_0\big(SO(3)\big)$.

Corollary 3.2.15. *Additionally to the assumptions and notations of Theorem 3.2.13 assume that* $\eta_0 = P\gamma_0 Q^T$ *for* $P, Q \in SO(3)$*, i.e.* γ_0 *and* η_0 *only differ by marker placement. Then we have that*

$$\Pi\left(\underset{R\in SO(4)}{\operatorname{argmin}} \sum_{k=1}^{K} \left\|R\widetilde{\hat{\gamma}_N}(t_k) - \widetilde{\hat{\eta}_M}(t_k)\right\|^2\right) \ni \left(\hat{P}_{N,M}, \hat{Q}_{N,M}\right) \xrightarrow{N,M\to\infty} \left(P, Q^T\right) \quad a.s.$$

Proof. Applying Theorem 3.2.13 we have to compute

$$\begin{aligned}
\big(P^*, Q^*\big) &= \Pi\left(\underset{R\in SO(4)}{\operatorname{argmin}} \sum_{n=1}^{N} \|R\tilde{\gamma}_0(t_n) - \tilde{\eta}_0(t_n)\|^2\right) \\
&= \Pi\left(\underset{R\in SO(4)}{\operatorname{argmin}} \sum_{n=1}^{N} \|R\tilde{\gamma}_0(t_n) - \widetilde{P\gamma_0 Q^T}(t_n)\|^2\right) \\
&= \Pi\left(\underset{R\in SO(4)}{\operatorname{argmin}} \sum_{n=1}^{N} \|R\tilde{\gamma}_0(t_n) - R_{P,Q^T}\tilde{\gamma}_0(t_n)\|^2\right).
\end{aligned}$$

In the last equality we used Theorem 3.2.4. This implies that $R = R_{P,Q^T}$ is the unique minimizer and by Theorem 3.2.7 we obtain $\Pi\big(R_{P,Q^T}\big) = \big(P, Q^T\big)$. □

Chapter 4

Tests of Equality of Center Curves

This chapter introduces our two-sample tests for equality of the center curve of rGP models under perturbation by gait similarities. Hence, given rGP models γ, η our hypotheses of equal or unequal shape of the center curves γ_0 and η_0 are

$$H_0: \ \exists g \in \mathcal{S}: \ \eta_0 = g.\gamma_0 \quad vs. \quad H_1: \ \forall g \in \mathcal{S}: \ \eta_0 \neq g.\gamma_0 \tag{4.1}$$

However, our first three testing procedures are tests for the hypotheses

$$H_0: \ \eta_0 = \gamma_0 \quad vs. \quad H_1: \ \eta_0 \neq \gamma_0$$

which consider the effect of (ψ, ϕ) as removable by preprocessing, whereas the last test, indeed, tries to test the hypotheses (4.1) directly.

We assume in this chapter that the PESM of all considered samples is unique and thus by Corollary 2.1.5 an element of $\mathcal{X}$. This is usually not a restriction, since in applications the data often is concentrated, which ensures unique PESMs.

The preprocessing step. We start with describing the preprocessing step of optimally aligning two sessions $\chi_1 = (\gamma_1, ..., \gamma_N)$ and $\chi_2 = (\eta_1, ..., \eta_M)$, $N, M \in \mathbb{N}$. It consists of the following steps.

Preprocessing 4.0.16.

1. *Compute the PESMs $\hat{\gamma}_N$ and $\hat{\eta}_M$ of χ_1 and χ_2 (see Theorem 2.1.3).*
2. *Estimate $\hat{\psi} = \left(\hat{P}_\psi, \hat{Q}_\psi\right)$ using the estimator* (3.29) *for the curves $\hat{\gamma}_N$ and $\hat{\eta}_M$.*
3. *Estimate $\hat{\phi}$ using estimator* (3.11) *for the curves $\hat{\psi} \circ \hat{\gamma}_N$ and $\hat{\eta}_M$.*
4. *Compute the new session $(\hat{\psi} \circ \gamma_1 \circ \hat{\phi}, ..., \hat{\psi} \circ \gamma_N \circ \hat{\phi})$*
5. *Replace $\hat{\gamma}_N$ in Step 1 by $\hat{\psi} \circ \hat{\gamma}_N \circ \hat{\phi}$ and iterate step 2-3 until convergence of $\hat{\psi}$ and $\hat{\phi}$.*

In our data application we usually had to perform Step 1-4 only once.

The overlapping SO(3) confidence sets test (OCST). In Section 2.2 we constructed simultaneous confidence sets for the PEM in rGP models. We use them to test equality of the PEMs of two sessions χ_1 and χ_2. Here the idea is that we reject the null hypothesis of equal PEMs, if the simultaneous confidence sets for the two sessions do not intersect for all times $t \in I$ and otherwise we accept the hypothesis.

Unfortunately, the significance level of such a testing procedures is not obvious a priori and may depend on the particular distribution, for example, for 1D Gaussian random variables Schenker and Gentleman [2001] and Payton et al. [2003] study a test based on the overlap of confidence intervals and they show using simulations and heuristic arguments that if 85%-confidence intervals of the means are used, the size of the test will be approximately 0.05, if the quotient of the variances of the Gaussians is close to one.

The testing procedure for equality of PEMs using simultaneous $\beta \cdot 100\%$-confidence sets is as follows.

Test 4.0.17 (OCST). *Let $\beta \in (0,1)$ and two sessions $\chi_1 = (\gamma_1, ..., \gamma_N)$, $N \in \mathbb{N}$, consisting of trials from a rGP with center curve γ_0 and $\chi_2 = (\eta_1, ..., \eta_M)$, $M \in \mathbb{N}$, consisting of trials from a rGP with center curve η_0 be given.*

1. *Compute the PESMs $\hat{\gamma}_N$ and $\hat{\eta}_M$ of χ_1 and χ_2 (see Theorem 2.1.3).*
2. *Compute $\mathcal{C}_\beta(\chi_1; t)$ and $\mathcal{C}_\beta(\chi_2; t)$ as explained in Section 2.2 and Section 2.3.*
3. *If for all $t \in I$*

$$\hat{\gamma}_N(t)\mathrm{Exp}\Big(\mathcal{C}_\beta(\chi_1; t)\Big) \cap \hat{\eta}_N(t)\mathrm{Exp}\Big(\mathcal{C}_\beta(\chi_2; t)\Big) \neq \emptyset$$

accept the hypothesis of equal shape of γ_0 and η_0, else reject it.

Remark 4.0.18. *The preprocessing 4.0.16 has to be carried out before applying this test, if one wants to remove marker placement or time warping effects. Note, however, that by Theorem 2.2.4 it is equivalent to compute $Q_{\hat{\psi}}^T \mathcal{C}_\beta(\chi_1; t)$ instead of $\mathcal{C}_\beta\Big((\hat{\psi}, \hat{\phi}).\chi_1; \hat{\phi}(t)\Big)$. Therefore one could either first align the sessions (i.e., use $(\hat{\psi}, \hat{\phi}).\chi_1$ and χ_2 to compute the two confidence bands) or compute first the confidence bands for χ_1 and χ_2 and then align the confidence sets using Theorem 2.2.4 with the estimators $\hat{\psi}$ and $\hat{\phi}$ from 4.0.16. Both procedures will produce the same outcome.*

Simultaneous approximated Hotelling T^2 test. Suppose that the session $\chi_1 = (\gamma_1, ..., \gamma_N)$, $N \in \mathbb{N}$, consists of trials from a rGP with center curve γ_0 and the session $\chi_2 = (\eta_1, ..., \eta_M)$, $M \in \mathbb{N}$, consists of trials from a rGP with center curve η_0. Building on the observation from Theorem 2.2.5 that in the case of concentrated errors the residuals of such sessions are approximatively the intrinsic residuals of the generating Gaussian processes (see equation (2.21)), we will now propose a version of a Hotelling T^2 test. Note that for $\mathbb{R}^3$ an analogous test is proposed in Pataky et al. [2013].

Therefore, if one wants to remove marker placement or self-selected walking speed effects, replace χ_1 by its aligned version stemming from Preprocessing 4.0.16. Let us denote with $\hat{\mu}$ the pooled PESM (i.e., $\hat{\mu}(t) = \hat{\mu}_{N+M}\big((\chi_1, \chi_2), t\big)$ for all $t \in I = [0,1]$). We define residuals analogous to the residuals defined in (2.21) by

$$X^{n,pool} = \iota^{-1} \circ \mathfrak{L}\big(\hat{\mu}^T \gamma_n\big) \qquad \text{and} \qquad Y^{m,pool} = \iota^{-1} \circ \mathfrak{L}\big(\hat{\mu}^T \eta_m\big)\,,$$

for $n \in \{1, ..., N\}$ and $m \in \{1, ..., M\}$. If the corresponding weighted sum of the sample covariance matrices is invertible for all $t \in I$, this yields the two sample Hotelling T^2-statistic

$$\hat{W}_t = \frac{NM}{N+M} \left(\bar{X}_t - \bar{Y}_t\right)^T \left(\frac{1}{N+M-2}\Big((N-1)\hat{\Sigma}_t^X + (M-1)\hat{\Sigma}_t^Y\Big)\right)^{-1} \left(\bar{X}_t - \bar{Y}_t\right)$$

for $t \in I$. Here we have

$$\bar{X} = N^{-1}\sum_{n=1}^{N} X^{n,pool}\,, \qquad \bar{Y} = N^{-1}\sum_{n=1}^{N} Y^{n,pool}\,,$$

$$\hat{\Sigma}^X = \frac{1}{N-1}\sum_{n=1}^{N} X^{n,pool}(X^{n,pool})^T\,, \qquad \hat{\Sigma}^Y = \frac{1}{N-1}\sum_{n=1}^{N} Y^{n,pool}(Y^{n,pool})^T\,.$$

Under the null hypothesis that $\gamma_0 = \eta_0$ and the underlying Gaussian processes generating the samples are identical, we have that $\hat{W}_t$ is approximatively Hotelling T^2-distributed with $N + M - 2$ degrees of freedom for each $t \in I$. In order to circumvent multiple testing and to take the covariance structure of the processes into account, we estimate h_α such that

$$\mathbb{P}\Big(\max_{t \in I} \hat{W}_t > h_\alpha\Big) = \alpha \tag{4.2}$$

under the null hypothesis. This is achieved by using the GKF for Hotelling T^2 processes as described in Section 2.3. Since under the null hypothesis the trials of χ_1 and χ_2 have the same distribution, we will use χ_1 to estimate the Lipschitz killing curvature.

The complete simultaneous approximated Hotelling T^2 test is then the following.

Test 4.0.19 (Simultaneous Approximated Hotelling T^2 Test.)**.** *Given two sessions $\chi_1 = (\gamma_1, ..., \gamma_N)$, $N \in \mathbb{N}$, consisting of trials of a distribution with unique PEM γ_0 and $\chi_2 = (\eta_1, ..., \eta_M)$, $M \in \mathbb{N}$, consisting of trials of a distribution with unique PEM η_0. Let $\alpha \in (0,1)$ be a given significance level.*

1. *If one wants to remove marker placement or self-selected walking speed effects, replace χ_1 by its temporally registered and aligned version obtained from Preprocessing 4.0.16.*

2. *Using χ_1 estimate with* (2.30) *the Lipschitz killing curvature and use this and the GKF for Hotelling T^2 processes with $N + M - 2$ degrees of freedom (see equation* (2.28)*) to obtain the quantile h_α given by* (4.2)*.*

3. *Compute the process* $\{\hat{W}_t\}_{t\in I}$ *from the sessions* χ_1 *and* χ_2.

4. *If* $\hat{W}_t > h_\alpha$ *for any* $t \in I$ *reject the null hypothesis of equal shapes of* γ_0 *and* η_0*, else accept.*

Intrinsic length loss permutation tests (ILLPerm & MILLPerm). The above two tests require that our sessions are consisting of trials drawn from rGP models. The following two permutation tests in contrast are distribution-free tests in the sense that they do not assume any distributional properties about the populations. In what follows δ denotes either $\delta_{I,1}$, $\delta_{I,2}$ or δ_I (see Definition 3.1.1).

Test 4.0.20 (ILLPerm)**.** *Given two sessions* $\chi_1 = (\gamma_1, ..., \gamma_N)$, $N \in \mathbb{N}$, *consisting of trials of a distribution with unique PEM* γ_0 *and* $\chi_2 = (\eta_1, ..., \eta_M)$, $M \in \mathbb{N}$, *consisting of trials of a distribution with unique PEM* η_0*. Let* $\alpha \in (0, 1)$ *be a given significance level.*

1. *If one wants to remove marker placement or self-selected walking speed effects, replace* χ_1 *by its temporally registered and aligned version obtained from Preprocessing 4.0.16.*

2. *Let* $\chi = (\chi_1, \chi_2)$ *be the pooled session. Denote with* $\chi_1^{(l)}$, $l \in \left\{1, ..., \binom{N+M}{N}\right\}$, *all possible choices with* N *elements from* χ *and with* $\chi_2^{(l)}$ *its complement in* χ *with* M *elements. Moreover, we assume* $\chi_1^{(1)} = \chi_1$ *and* $\chi_2^{(1)} = \chi_2$*. For each* l *compute the PESMs* $t \mapsto \hat{\gamma}_{N,l}(t) = \hat{\mu}_N\left(\chi_1^{(l)}, t\right)$ *and* $t \mapsto \hat{\eta}_{M,l}(t) = \hat{\mu}_M\left(\chi_2^{(l)}, t\right)$ *and compute*

$$d_l = \delta\left(\hat{\gamma}_{N,l}, \hat{\eta}_{M,l}\right)$$

3. *Let* $r = \#\{l \mid d_l > d_1\}$*. We reject the null hypothesis of equal shape of* γ_0 *and* η_0*, if p-value* $= r/\binom{N+M}{N} < \alpha$.

Remark 4.0.21. *In practice if* N *and* M *are large it is not feasible to compute all* $\binom{N+M}{N}$ *permutations of the data, since it is computationally costly. Therefore one approximates the p-value of the permutation test by drawing randomly* K *permutations of the data and computing the rank from this* K *permutations only (see [Edgington and Onghena, 2007, Chapter 3.6 , p. 40]).*

The previous approaches all consider the temporal registration and spatial alignment of the compared sessions as a preprocessing step. However, since it includes estimates of an aligning gait similarity (ψ, ϕ), the variance of the estimators may as well influence the error of the first and second kind. Therefore, we propose a modified version of the above permutation test taking this variation into account by including the preprocessing into each permutation.

Test 4.0.22 (MILLPerm)**.** *Given two sessions* $\chi_1 = (\gamma_1, ..., \gamma_N)$, $N \in \mathbb{N}$, *consisting of trials of a distribution with unique PEM* γ_0 *and* $\chi_2 = (\eta_1, ..., \eta_M)$, $M \in \mathbb{N}$, *consisting of trials of of a distribution with unique PEM* η_0*. Let* $\alpha \in (0, 1)$ *be a given significance level.*

1. *Let $\chi = (\chi_1, \chi_2)$ be the pooled session. Denote with $\chi_1^{(l)}$, $l \in \{1, ..., \binom{N+M}{N}\}$, all possible choices with N elements from χ and with $\chi_2^{(l)}$ its complement with M elements in χ. Moreover, we assume $\chi_1^{(1)} = \chi_1$ and $\chi_2^{(1)} = \chi_2$. For $i, j \in \{1, 2\}$ let $\chi_{ij}^{(l)}$ denote the session consisting of all trials of $\chi_i^{(l)}$ also belonging to χ_j. For each l, i compute the PESMs $t \mapsto \hat{\gamma}_{i,l}(t) = \hat{\mu}_N\Big(\chi_{i1}^{(l)}, t\Big)$ and $t \mapsto \hat{\eta}_{i,l}(t) = \hat{\mu}_M\Big(\chi_{i2}^{(l)}, t\Big)$.*

2. *For each i, l apply Preprocessing 4.0.16 to $\hat{\gamma}_{i,l}$ and $\hat{\eta}_{i,l}$ to obtain an aligned version $\hat{\gamma}_{i,l}^{aligned}$. Compute the PESM $\hat{\omega}_{i,l}$ of $\hat{\eta}_{i,l}$ and $\hat{\gamma}_{i,l}^{aligned}$.*

3. *Apply Preprocessing 4.0.16 to $\hat{\omega}_{1,l}$ and $\hat{\omega}_{2,l}$ to get an aligned version $\hat{\omega}_{1,l}^{aligned}$ of $\hat{\omega}_{1,l}$ and compute*
$$d_k = \delta\big(\hat{\omega}_{1,l}^{aligned}, \hat{\omega}_{2,l}\big)$$

4. *Let $r = \#\{l \mid d_l > d_1\}$. We reject the null hypothesis of equal shape of γ_0 and η_0 at a significance level $\alpha \in (0, 1)$, if p-value $= r/\binom{N+M}{N} < \alpha$.*

Chapter 5

Simulations

5.1 Assessing Covering Rates of Confidence Sets

5.1.1 Small Sample Behavior of Multiplier Bootstrap, Asymptotic Confidence Bands and GKF for Simultaneous Confidence Bands

The aim of this section is to analyze the small sample behavior of different methods for estimating simultaneous confidence bands for the pointwise mean function in functional data. These simulations are included in this thesis, since they justify the use of the GKF as our tool for computing the α-quantile of the maximum of the Hotelling T^2 statistic of rGP models. Let $\mu : [0,1] \to \mathbb{R}$ be a deterministic function. We consider in this section the following data model

$$X(t) = \mu(t) + Z_t\,, \qquad t \in I \tag{5.1}$$

where $\{Z_t\}_{t\in I}$ is a real valued Gaussian process on $I = [0,1]$ with almost surely $\mathcal{C}^2$-sample paths, $\mathbb{E}[Z_t] = 0$, $\mathrm{Var}[Z_t] = \sigma_t^2 > 0$ and Z_t/σ_t fulfills the Assumptions (*GKF 3*) and (*GKF 4*). Moreover, we assume that we observe the process X only at times $0 = t_1 < t_2 < ... < t_K = 1$.

Let $X^1, \ldots, X^N$ be a sample from model (5.1), then we want to find two functions $l_N, u_N : [0,1] \to \mathbb{R}$ depending on the sample such that

$$\mathbb{P}\big(l_N(t) \le \mu(t) \le u_N(t) \text{ for all } t \in I\big) \ge 1 - \alpha\,.$$

using the observed values $X^n(t_k)$ for $k \in \{1, ..., K\}$ and $n \in \{1, ..., N\}$. This can be achieved using the stochastic process

$$T_t = \sqrt{N}\frac{\bar{X}(t) - \mu(t)}{\hat{\sigma}_t}\,, \tag{5.2}$$

where we define

$$\bar{X}(t) = N^{-1} \sum_{n=1}^{N} X^n(t)$$
$$\hat{\sigma}_t^2 = \frac{1}{N-1} \sum_{n=1}^{N} R_n(t)^2$$
$$R_n(t) = X^n(t) - \bar{X}(t)\,.$$

Note that this process is as shown in Adler and Taylor [2009, Section 15.10.3, p.430] well-defined for all $N \geq 2$.

Now, given $h_\alpha \in R_{>0}$ such that

$$\mathbb{P}\left(\max_{t\in[0,1]} |T_t| > h_\alpha \right) \leq \alpha.$$

we obtain that the collection of intervals

$$\left[\bar{X}(t) - h_\alpha \tfrac{\hat{\sigma}_t}{\sqrt{N}}, \bar{X}(t) + h_\alpha \tfrac{\hat{\sigma}_t}{\sqrt{N}} \right], \quad \text{for } t \in [0,1]$$

form a simultaneous $(1-\alpha)\cdot 100\%$ confidence band for μ i.e.,

$$\mathbb{P}\left(\mu(t) \in \left[\bar{X}(t) - h_\alpha \tfrac{\hat{\sigma}_t}{\sqrt{N}}, \bar{X}(t) + h_\alpha \tfrac{\hat{\sigma}_t}{\sqrt{N}} \right] \text{ for all } t \in I \right) \geq 1-\alpha\,.$$

We will now describe three different methods for estimating the threshold h_α and explore their small sample performance using simulations.

Naive bootstrap approach. The first method, which we will use in our comparison, is proposed in Degras [2011] called the naive bootstrap. The main result of Degras [2011] is a functional asymptotic normality result for the local linear estimator for dense functional data. Although this result allows for constructing (asymptotically correct) $(1-\alpha)\%$ confidence bands of the mean curve, Degras proposes to use the naive bootstrap for small sample sizes. The naive bootstrap works as follows.

1. Resample with replacement from a sample $X^1, ..., X^N$ of model (5.1) to produce a bootstrap sample $X^{1,*}, ..., X^{N,*}$.
2. Compute the pointwise empirical mean $\bar{X}^*$ and variance $(\hat{\sigma}^*)^2$ functions of the bootstrap sample $X^{1,*}, ..., X^{N,*}$.
3. Compute $Z^* = \sqrt{N} \max_{t\in I} \|(\bar{X}^* - \bar{X})/\hat{\sigma}^*\|$.
4. Repeat steps 1 to 3 many times to approximate the conditional law $\mathcal{L}^* = \mathcal{L}\big(Z^* \mid X^1, ..., X^N\big)$ and take the $(1-\alpha)\cdot 100\%$ quantile of $\mathcal{L}^*$ to estimate h_α.

Note that in Degras [2011] instead of $\bar{X}$ the local linear estimator is used, which smooths the data. This simplification can be done, since we do not include an additional observation error in model (5.1) as done in Degras [2011] and therefore smoothing is not necessary.

Multiplier bootstrap. The second method builds on a version of the multiplier (or Wild) bootstrap (e.g., Mammen [1993]) designed for the maximum of sums of N independent random variables in high dimensions as discussed in detail by Chernozhukov et al. [2013]. More precisely, let $Y_1, ..., Y_N$ be independent random vectors in $\mathbb{R}^K$, $N, K \in \mathbb{N}$ with $\mathbb{E}\big[Y_n\big] = 0$ and finite covariance $\mathbb{E}\big[Y_n Y_n^T\big]$ for all $n \in \{1, ..., N\}$. We define $Y_n^T = \big(Y_{n1}, ..., Y_{nK}\big)$ and assume there are $c, C \in \mathbb{R}_{>0}$ such that $c < \mathbb{E}\big[Y_{nk}^2\big] < C$ for all $n \in \{1, ..., N\}$ and all $k \in \{1, ..., K\}$. Under these assumptions it is shown in Chernozhukov et al. [2013, Theorem 3.1] that the quantiles of the distribution of

$$\max_{k\in\{1,...,K\}} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_{nk}$$

can be asymptotically consistently estimated by the quantiles of the multiplier bootstrap i.e., by the distribution of

$$\max_{k\in\{1,...,K\}} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} g_k Y_{nk}$$

with i.i.d. multipliers $g_1, ..., g_N \sim \mathcal{N}(0, 1)$ given the data $Y_1, ..., Y_N$.

In order to apply Chernozhukov et al. [2013, Theorem 3.1], note that we can rewrite

$$\max_{k\in\{1,...,K\}} T(t_k) = \max_{k\in\{1,...,K\}} \frac{1}{\sqrt{N}\hat{\sigma}_{t_k}} \sum_{n=1}^{N} \sqrt{\tfrac{N}{N-1}} R_n(t_k)$$

by

$$\sqrt{N}\big(\bar{X} - \mu\big) \sim \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sqrt{\tfrac{N}{N-1}} R_n \,,$$

which follows from $\mathbb{E}[R_n(t)R_n(s)] = \frac{N-1}{N}\mathbb{E}[X^n(t)X^n(s)]$ for all $t, s \in I$. Since $R_1, ..., R_N, \hat{\sigma}_t$ are independent as stochastic processes, the random vectors $Y_n = \sqrt{\frac{N}{N-1}}\big(R_n(t_1)/\hat{\sigma}_{t_1}, ..., R_n(t_K)/\hat{\sigma}_{t_K}\big) \in \mathbb{R}^K$, $n \in \{1, ..., N\}$, satisfies the assumptions of Chernozhukov et al. [2013, Theorem 3.1] and therefore the multiplier bootstrap is applicable to estimate quantiles of the distribution of the maximum of the random vector $T = \big(T_{t_1}, ..., T_{t_K}\big)$.

Since Chernozhukov and co-authors show that the multiplier bootstrap works also for $K \gg N$, we apply this method without further theoretical justification to the functional case and the process $\{T_t\}_{t\in I}$. The same reasoning as above, then yields that we can use the multiplier bootstrap to estimate h_α given by

$$\mathbb{P}\left(\max_{t\in[0,1]} |T_t| > h_\alpha\right) = \mathbb{P}\left(\max_{t\in[0,1]} \left|\frac{1}{\sqrt{N}}\hat{\sigma}_t^{-1} \sum_{n=1}^{N} \sqrt{\frac{N}{N-1}} R_n\right| > h_\alpha\right).$$

Hence we introduce i.i.d. multipliers $g_1, ..., g_N \sim \mathcal{N}(0, 1)$ and estimate h_α using bootstrap replicates of

$$W = \max_{t\in[0,1]} \frac{1}{\sqrt{N}\hat{\sigma}_t^g} \left|\sum_{n=1}^{N} \sqrt{\frac{N}{N-1}} g_n R_n(t)\right|$$

with

$$\big(\hat{\sigma}_t^g\big)^2 = \frac{1}{N-1}\sum_{n=1}^{N}\big(g_n R_n(t)\big)^2 .$$

The estimator of h_α is then given by

$$\hat{h}_\alpha = \inf_{h\in\mathbb{R}}\left\{\mathbb{P}_g\big(W \le h\big) \ge 1-\alpha\right\},$$

where $\mathbb{P}_g$ is the probability measure induced by the multipliers g holding $R_n(t)$ fixed i.e., $\mathbb{P}_g\big(W \le h\big) = \mathbb{P}\big(W \le h | R_1, ..., R_N\big)$.

Gaussian kinematic formula of T-statistic. Analogously to our approach in Section 2.2 we can estimate the threshold h_α using the GKF for the process $\{T_t\}_{t\in[0,1]}$ given in equation (5.2), which has pointwise a t-distribution with $(N-1)$-degrees of freedom. By the expected Euler characteristic heuristic and the Gaussian kinematic formula (see Adler and Taylor [2009, Theorem 15.10.3.]), we obtain

$$\begin{aligned}\mathbb{P}\left(\max_{t\in[0,1]}|T(t)| > h\right) &= 2\mathbb{P}\left(\max_{t\in[0,1]} T(t) > h\right) \approx 2\mathbb{E}\Big[\chi\big(\{t\in[0,1] \mid T_t \ge h\}\big)\Big] \\ &= 2\left(\mathcal{L}_1\big([0,1]\big)\frac{\big(1+\frac{h^2}{N-1}\big)^{1-\frac{N}{2}}}{2\pi} + \big(1-F_{N-1}(u)\big)\right).\end{aligned} \tag{5.3}$$

Here F_{N-1} denotes the cumulative distribution function of a Student's t-distribution with $(N-1)$-degrees of freedom and the first equality is due to the fact that the processes $\{T_t\}_{t\in I}$ and $\{-T_t\}_{t\in I}$ have the same distribution; hence

$$\begin{aligned}\mathbb{P}\left(\max_{t\in[0,1]}|T(t)| > h\right) &= \mathbb{P}\left(\max_{t\in[0,1]} T(t) > h\right) + \mathbb{P}\left(\max_{t\in[0,1]} -T(t) > h\right) \\ &= 2\mathbb{P}\left(\max_{t\in[0,1]} T(t) > h\right).\end{aligned}$$

By equation 5.3 we only have to estimate the Lipschitz killing curvature from the observations to construct simultaneous confidence bands. The Lipschitz killing curvature is given by

$$\mathcal{L}_1\big([0,1]\big) = \int_0^1 \sqrt{\mathrm{Var}\Big[\tfrac{d}{dt}\big(\tfrac{Z_t}{\sigma_t}\big)\Big]}\,dt\,,$$

which can be found in Taylor and Worsley [2007, Section 3.3]. Given an i.i.d. sample $X^1, ..., X^N$ of model (5.1) evaluated on a partition $0 = t_1 < t_2 < ... < t_K = 1$ we use a discretized version of $\mathcal{L}_1\big([0,1]\big)$ by replacing the integral by its Riemann sum and the derivative by finite differences. Moreover, we use that

$$\mathrm{Var}\Big[\tfrac{d}{dt}\Big(\sigma_t^{-1}Z_t\Big)\Big] = \mathrm{Var}\Big[\tfrac{d}{dt}\Big(\sigma_t^{-1}\big(X_t-\mu(t)\big)\Big)\Big].$$

This yields the estimator

$$\hat{\mathcal{L}}_1\big([0,1]\big) = \sum_{k=1}^{K-1} \widehat{d\Xi_k}(t_{k+1} - t_k)\,,$$

where we defined

$$\widehat{d\Xi_k}^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(dX_k^n - N^{-1}\sum_{n=1}^{N} dX_k^n\right)^2$$

with

$$dX_k^n = \left(\frac{X^n(t_{k+1})}{\hat{\sigma}_{t_{k+1}}} - \frac{X^n(t_k)}{\hat{\sigma}_{t_k}}\right)(t_{k+1} - t_k)^{-1}$$

for $k \in \{1, ..., K-1\}$.

Error processes for 1D confidence bands simulation. In the simulations of the covering rate of 1D confidence bands constructed using the methods proposed above, we assume for simplicity that $\mu(t) = 0$ for all $t \in [0,1]$. The performance of the presented methods is tested using for the error processes Z in model 5.1 the processes

$$\begin{aligned}
\varepsilon_t^{1,l} &= f_l(t)\Big(a_1 \sin\left(\tfrac{\pi}{2}t\right) + a_2 \cos\left(\tfrac{\pi}{2}t\right)\Big) \\
\varepsilon_t^{2,l} &= f_l(t)\left(\frac{\sum_{i=1}^{10} a_i e^{-\frac{\left(x-\frac{i-1}{9}\right)^2}{0.2}}}{\sqrt{\sum_{i=1}^{10} e^{-2\frac{\left(x-\frac{i-1}{9}\right)^2}{0.2}}}}\right) \\
\varepsilon_t^{3,l} &= f_l(t)\left(a_0 e^{-5t} + \sqrt{10}\int_0^t e^{5(s-t)} dW_t\right)
\end{aligned} \tag{5.4}$$

with i.i.d. $a_i \sim \mathcal{N}(0,1)$ for $i \in \{0, ..., 10\}$, $\{W_t\}_{t\in I}$ a Wiener process, and for $l \in \{1,2,3\}$ we have

$$f_1(t) = 1\,, \qquad f_2(t) = 4\,, \qquad f_3(t) = \sin(4\pi t) + 1.5\,.$$

Note that the processes satisfy $\mathrm{Var}\left[\varepsilon_t^{\nu,l}\right] = f_l(t)^2$ for all $t \in [0,1]$ and $\nu \in \{1,2,3\}$. Moreover, the sample paths of the processes $\varepsilon^{1,l}$ and $\varepsilon^{2,l}$ have $\mathcal{C}^\infty$-sample paths, whereas the sample paths of $\varepsilon^{3,l}$, which is a Ornstein-Uhlenbeck process (e.g., Iacus [2008, p.43]) are only continuous implying that the GKF is not applicable for this process. However, since the estimator of the Lipschitz killing curvature is computable also for the Ornstein-Uhlenbeck process, we studied also confidence sets using the GKF approach for the Ornstein-Uhlenbeck process. We expect, that this does not work well, since the estimation of the Lipschitz killing curvature relies on the estimation of the variance of the derivative of the process, which does not exists in this case.

Design of 1D confidence bands simulation. We use the proposed methods to construct confidence bands for the mean function $\mu \equiv 0$. To obtain the covering rates for small sample sizes, we do the following: simulate $N \in \{5, 10, 15, 20, 30, 50\}$ realizations of the process $\varepsilon_t^{\nu,l}$ for $\nu, l \in \{1, 2, 3\}$ on the equidistant time grid $\mathcal{T}$ with $\Delta t = 0.01$ and compute the simultaneous confidence band with the selected method at these points. Then check whether $\mu(t)$ is contained in the constructed confidence band for all $t \in \mathcal{T}$. We repeat this $M = 5000$ times and the relative frequency between the trials such that μ is always within the constructed confidence band and the number of simulations approximates the true covering rate.

For the bootstrap methods we used 2000 bootstrap replicates.

Results of 1D confidence bands simulation. The results of this simulation are collected in the Tables C.1, C.2 and C.3 in Appendix C. The conclusions are the following: in the case of small sample sizes ($\approx$ 10-20) the only reliably working method is the Gaussian kinematic formula approach, which is surprisingly accurate and only systematically overestimates the covering rate for the Ornstein Uhlenbeck error process, which, anyway, does not satisfy the assumptions of the GKF. While the naive bootstrap yields too conservative confidence bands for small samples sizes, we discovered that the multiplier bootstrap underestimates the covering rate. For larger sample sizes (≥ 50) both bootstrap methods start to perform well.

Note that another advantage of the GKF is that it is computational very fast. Due to these observations we will only use the GKF approach to construct simultaneous confidence bands of the PEM in rGP models.

5.1.2 Covering Rates of Simultaneous Confidence Sets for rGP models

Since the estimation of the quantile $\tilde{h}_{\gamma,N,\alpha}$ of the process (2.22), which we used in Section 2.2 to construct simultaneous confidence sets, relies on an approximation for concentrated data given in Theorem 2.2.5, we study the actual covering rate of this method also with simulations.

Error processes for rGP confidence sets simulations. We asses the covering rates by generating data from a rGP model (1.8). This is done by simulating a Gaussian process $A = \{A_t\}_{t\in I}$ in $\mathbb{R}^3$ and mapping it to $SO(3)$ using the Lie exponential (see A.3). For simplicity, we assume that the mean curve γ_0 of the rGP model is given by $\gamma_0(t) = I_{3\times 3}$ for all $t \in [0, 1]$. This can be done without loss of generality, since the PESM $\hat{\gamma}_N$ of a sample $\gamma_1, \ldots, \gamma_N$ from a rGP model with realizations $A^1, ..., A^N$ of the process A is equivariant and hence we obtain

$$\hat{\gamma}_N(t) = \gamma_0(t) \operatorname*{argmin}_{\mu\in SO(3)} \sum_{n=1}^{N} \left\| \operatorname{Exp}\big(\iota(A_t^n)\big) - \mu(t) \right\|_F^2 = \gamma_0(t)\hat{A}_N(t)\,,$$

where $\hat{A}_N$ is the PESM of the error process $\operatorname{Exp}\big(\iota(A^1)\big), ..., \operatorname{Exp}\big(\iota(A^N)\big)$.

In our simulations studying the covering rates of the simultaneous confidence sets given in Theorem 2.2.3, we use the error processes $\{\varepsilon_t^{i,l}\}_{t\in I}$, $i \in \{1,2,3\}$, $l \in \{1,3\}$, defined in (5.4) to construct the Gaussian process A by the following formula

$$A_t^{i,l,j,\sigma} = W_j\big(\sigma\varepsilon_{1,t}^{i,l}, \sigma\varepsilon_{2,t}^{i,l}, \sigma\varepsilon_{3,t}^{i,l}\big)^T, \tag{5.5}$$

for $i \in \{1,2,3\}$, $j \in \{1,2\}$, $l \in \{1,3\}$ and $\sigma \in \mathbb{R}_{>0}$. Here we denote with $\varepsilon_{s,t}^{i,l}$ for $s = 1,2,3$ independent realizations of $\{\varepsilon_t^{i,l}\}_{t\in I}$. The matrices

$$W_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

are introduced to include the case of dependencies between the coordinates. Moreover, it introduces different variances in the coordinates, since for $j = 2$ the second component will have half the variance than the other two components.

Design of rGP confidence sets simulations. We simulate $N \in \{10, 15, 30\}$ realizations of the process $\{A_t^{i,l,j,\sigma}\}$ on the equidistant time grid $\mathcal{T}$ with $\Delta t = 0.01$ for $i \in \{1,2,3\}$, $j \in \{1,2\}$, $l \in \{1,3\}$ and $\sigma \in \{0.005, 0.05, 0.1, 0.6\}$. Note that the variance of our real gait data is $\sigma \leq 0.05$, such that we cover this case in our simulations. Then simultaneous 95%-confidence sets are constructed as described in Section 2.2, where we estimate the quantile $\tilde{h}_{\gamma,N,\alpha}$ as proposed in Section 2.3 and check whether $\gamma \equiv I_{3\times 3}$ is contained in the constructed confidence set for all $t \in \mathcal{T}$. We repeat this $M = 5000$ times and the relative frequency between the trials such that γ is always contained in the constructed confidence set and the number of simulations approximates the true covering rate.

Results of rGP confidence sets simulations. We report the results of the simulations in Tables C.4 and C.5. The simulations have a simple message: For variance $\sigma \leq 0.1$ the simulated covering rate is close to 95%. Only in the case of the Ornstein Uhlenbeck error process we have slightly too high covering rates. For high variance $\sigma = 0.6$ we underestimate the covering rate. This is expected, since the proposed estimator is designed for concentrated data and the map $v \mapsto \mathrm{Log}\big(\mathrm{Exp}(v)\big)$ is only the identity on $\|v\| < \pi$ and we have the inequality

$$\Big\|\mathrm{Log}\Big(\mathrm{Exp}\big(\iota(v)\big)\Big)\Big\|_F \leq \|v\|. \tag{5.6}$$

The latter implies that our estimated covariance matrix has smaller eigenvalues then the covariance matrix of the sample and hence our confidence sets will become smaller. This effect is even more visible if the sample size is large.

Recall that the real gait data, which we will analyze in Section 6 is far away from these problematic region, since it is concentrated with $\sigma \leq 0.05$.

5.2 Assessing Type I and Type II Error of Two-Sample Tests

In Chapter 4 we introduced different two-sample tests for the equality of PEMs. In this section we will simulate their performance for rGP models. In a first step we will simulate their type I and type II error rate under the hypothesis that no gait similarities are perturbing the data. In a second step we will introduce perturbations by an element of $\mathcal{I}_0\big(SO(3)\big)$ and remove it as explained in Section 4 by using the estimators introduced in Section 3. A third step will include perturbation by an element from $\mathcal{S}$. However, since the latter simulations are extremely time consuming due to the estimation of the temporal alignment we present a relatively small simulation study.

Error processes for two-sample test simulations. The rGP processes used in this simulation are constructed from the processes (5.4) and the center curves $\gamma_0^\lambda(t)$, $\lambda \in \mathbb{R}$, given by

$$\begin{aligned}
\alpha_x^\lambda\big(t\big) &= 80t^2 - 80t + 20 + \lambda \frac{e^{-\frac{1}{2}\left(\frac{t-0.5}{0.08}\right)^2}}{0.08\sqrt{2\pi}} - 35 \\
\alpha_y\big(t\big) &= 70t \sin\left(4\pi t^{0.7}\right) + 5 \\
\alpha_z\big(t\big) &= 10 \cos\left(13\pi\right),
\end{aligned}$$

where $\alpha_x^\lambda, \alpha_y, \alpha_z$ are the Euler angles representation in degrees (see Section A.2) of $\gamma_0^\lambda(t)$. For $\lambda \in \{0, 0.5, 1, 2, 2.5\}$ these curves are shown in Figure D.1.

In all simulations of the type I and II error of our two-sample tests from Section 4 we will use the following five processes, which are selected from the processes (5.5) used for the simulations of the covering rates of our simultaneous confidence sets,

$$\begin{aligned}
\gamma_A^0 &= \gamma_0^0(t)\mathrm{Exp}\left(\iota\Big(A_t^{1,1,1,0.05}\Big)\right) \\
\gamma_B^\lambda &= \gamma_0^\lambda(t)\mathrm{Exp}\left(\iota\Big(A_t^{2,3,2,0.05}\Big)\right) \quad \text{for } \lambda \in \{0.5, 1, 2, 2.5\}.
\end{aligned} \tag{5.7}$$

Examples of simulations from these processes are shown in Figure D.2, D.3 and D.4.

5.2.1 Performances Without Pertubation by Gait Similarities

Design of the two-sample test in $\mathcal{X}$ simulations. Let γ and η each be distributed according to one of the processes given in (5.7). We simulate $\gamma_1, ..., \gamma_N$ and $\eta_1, ..., \eta_N$, $N \in \{10, 15, 30\}$, realizations on the equidistant time grid $\mathcal{T}$ with $\Delta t = 0.01$. We then apply the OCST, Hotelling T^2-test, ILLPerm and MILLPerm ($\beta = 0.9$ for the OCST and $\alpha = 0.05$ for the other tests) to the two samples $\gamma_1, ..., \gamma_N$ and $\eta_1, ..., \eta_N$.

We perform $M = 2000$ simulations of this type and report the acceptance rate. If the distributions of γ and η are equal, then one minus the acceptance rate gives an approximation of the type I error, else, if the distributions are different, the acceptance rate is an approximation of the type II error. For the permutation approaches we use $M_{perm} = 5000$ permutations.

Results of the two-sample test in $\mathcal{X}$ simulations. The results of this simulation are given in Tables 5.1 and 5.2. ILLPerm, MILLPerm and the simultaneous Hotelling T^2 test seem to achieve the correct significance level $\alpha = 0.05$. For the considered error processes the best power is offered by the Hotelling T^2-test. This is in accordance with the observation that the permutation tests are distribution free, while the Hotelling T^2 test is specifically designed for Gaussian processes.

As expected, the overlap of the simultaneous 90%-confidence sets (OCST) produces a very conservative test, which has also a higher type II error than the other considered tests.

Note that for small sample sizes relatively small perturbations of the tested distributions do lead to a high type II error, which explains the high values on the second diagonal of Tables 5.1 and 5.2. This is mainly due to the small sample sizes. Thus, as expected the type II error always decreases if the sample size increases.

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	99.9 99.9 99.8	98.0 91.6 52.7	77.9 23.0 0.0	1.2 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		100 99.9 100	86.8 22.3 0.0	0.0 0.0 0.0	0.0 0.0 0.0
γ_B^1			100 100 100	1.8 0.0 0.0	4.8 0.0 0.0
γ_B^2				100 100 100	86.9 21.4 0.0
$\gamma_B^{2.5}$					100 100 100

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	95.3 95.8 95.1	71.5 54.6 17.1	16.3 0.9 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		94.3 94.4 94.8	36.0 6.5 0.0	0.0 0.0 0.0	0.0 0.0 0.0
γ_B^1			95.3 95.3 95.4	0.0 0.0 0.0	0.0 0.0 0.0
γ_B^2				95.9 95.3 94.7	35.5 6.8 0.0
$\gamma_B^{2.5}$					94.6 95.0 95.3

Table 5.1: Acceptance rate in percentage of H_0 of OCST (left) and simultaneous Hotelling T^2 (right).

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	95.8 94.5 94.5	80.7 72.8 45.1	29.8 8.6 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		95.4 94.6 94.8	88.1 81.6 63.0	8.4 0.2 0.0	0.1 0.0 0.0
γ_B^1			94.2 95.3 93.65	49.7 26.4 0.3	7.1 0.3 0.0
γ_B^2				95.3 95.3 95.4	87.9 81.2 61.8
$\gamma_B^{2.5}$					95.3 95.2 95.4

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	94.8 95.1 95.1	82.9 73.6 40.4	23.2 3.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		94.8 94.8 95.3	90.3 85.5 69.4	0.0 0.0 0.0	0.0 0.0 0.0
γ_B^1			95.5 95.7 95.2	59.6 30.5 0.4	9.0 0.0 0.0
γ_B^2				96.4 94.5 95.2	90.0 85.6 68.3
$\gamma_B^{2.5}$					95.8 95.3 94.2

Table 5.2: Acceptance rate in percentage of H_0 of ILLPerm (left) and MILLPerm (right) without application of P and Q to the η sample

5.2.2 Performances Including Perturbation by $\mathcal{I}_0\big(SO(3)\big)$

Design of the two-sample test in $\mathcal{X}/\mathcal{I}_0\big(SO(3)\big)$ simulations. Let γ and η each be distributed according to one of the processes given in (5.7). We simulate $\gamma_1, ..., \gamma_N$ and $\eta_1, ..., \eta_N$, $N \in \{10, 15, 30\}$, realizations on the equidistant time grid $\mathcal{T}$ with $\Delta t = 0.01$. Contrary to the simulation discussed in the last paragraph, we compute the realization $P\eta_1 Q^T, ..., P\eta_N Q^T$ with a rotation $P \in SO(3)$ with Euler angles $\alpha_x = -0.5°$, $\alpha_y = 13°$, $\alpha_z = -9°$ and a rotations $Q \in SO(3)$ with Euler angles $\alpha_x = 12°$, $\alpha_y = 0°$, $\alpha_z = 5°$. Denote this rotated sample by $\tilde{\eta}_1, ..., \tilde{\eta}_N$.

The application of P and Q to the simulated samples is performed in order to explore the robustness of the proposed testing procedures, if one has to correct for the perturbation by an element from $\mathcal{I}_0\big(SO(3)\big)$. In terms of our biomechanical application it mimics the robustness of the test procedure against different marker placements. To this end we apply the OCST, Hotelling T^2 test and ILLPerm with the preprocessing step of estimating $\hat{P}$ and $\hat{Q}$ as described in Preprocessing 4.0.16 ($\beta = 0.9$ for OCST and $\alpha = 0.05$ for the other tests). Note that we replace $\hat{\phi}$ by $id_{[0,1]}$ in Step 3 of Preprocessing 4.0.16, since temporal registration is not necessary for the data. MILLPerm ($\alpha = 0.05$) is applied to the two samples $\gamma_1, ..., \gamma_N$ and $\tilde{\eta}_1, ..., \tilde{\eta}_N$. Note MILLPerm does not require a preprocessing step, since it is constructed to test equality even under a perturbations by $\mathcal{I}_0\big(SO(3)\big)$. For the same reason mentioned before we replace within MILLPerm the estimator $\hat{\phi}$ by $id_{[0,1]}$.

We used $M = 2000$ simulations and report the acceptance rate. If the distribution of γ and η is equal, then one minus the acceptance rate approximates the type I error rate, else, if the distributions are different, the acceptance rate is an approximation of the type II error. For the permutation approaches we use $M_{perm} = 5000$ permutations.

Results of the two-sample test in $\mathcal{X}/\mathcal{I}_0\big(SO(3)\big)$ simulations. The results of this simulation are given in Tables 5.3 and 5.5. Interestingly, under the perturbation by (P, Q) only MILLPerm seems to achieve the correct significance level $\alpha = 0.05$ and its power is similar to the power without perturbation, compare Table 5.2. The type I error of the Hotelling T^2 test and the OCST seem to depend on the chosen error process and is sometimes higher and sometimes lower than the 5%. Even worse it increases, if the sample size increases. We suspect that this behavior is due to the observation that the estimator for $\hat{Q}$ and $\hat{P}$ are never exact and hence the null hypothesis that $\hat{P}\gamma_1\hat{Q}^T, ..., \hat{P}\gamma_N\hat{Q}^T$ and $\tilde{\eta}_1, ..., \tilde{\eta}_N$ have the same PEM is never true for the two samples and finite $N \in \mathbb{N}$. For the Hotelling T^2 test this is in accordance with the observation that it has a good power and hence is able to detect such a difference even better with increasing N. For the OSCT this is unsuspected, since it is a very conservative procedure without a perturbation from $\mathcal{I}_0\big(SO(3)\big)$ and we have no simple explanation. ILLPerm does perform relatively well in our simulations. However, for the tested error processes it has a lower type I error than expected, which is preferable, but also a higher type II error, which is not preferable.

Therefore we conclude that MILLPerm is the best choice within the compared

methods, if the data is corrupted by a perturbation of an element of $\mathcal{I}_0\big(SO(3)\big)$.

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	100 100 100	100 100 99.9	99.8 99.0 69.4	85.2 18.9 0.0	51.8 5.6 0.0
$\gamma_B^{0.5}$		95.4 88.3 79.9	81.3 44.3 5.3	1.3 0.0 0.0	0.0 0.0 0.0
γ_B^1			95.5 89.7 79.7	27.7 7.4 0.0	1.3 0.0 0.0
γ_B^2				95.7 88.3 80.6	82.7 49.6 6.7
$\gamma_B^{2.5}$					96.2 87.8 80.6

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	99.8 99.7 99.9	97.8 97.6 94.8	88.9 77.1 8.9	13.7 0.3 0.0	1.9 0.0 0.0
$\gamma_B^{0.5}$		74.2 66.5 64.0	48.6 29.2 4.1	0.0 0.0 0.0	0.0 0.0 0.0
γ_B^1			72.6 68.1 63.1	6.3 0.3 0.0	0.0 0.0 0.0
γ_B^2				72.1 68.4 62.4	49.9 30.3 4.5
$\gamma_B^{2.5}$					74.8 70.1 61.1

Table 5.3: Acceptance rate in percentage of H_0 of OCST (left) and simultaneous Hotelling T^2 (right). The samples are perturbed by an element of $\mathcal{I}_0\big(SO(3)\big)$.

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	99.9 99.9 99.9	93.6 90.2 73.1	50.1 16.5 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		96.1 95.9 96.5	93.1 88.4 75.9	11.5 0.8 0.0	0.2 0.0 0.0
γ_B^1			95.8 96.8 96.1	63.4 39.0 0.8	12.4 0.3 0.0
γ_B^2				96.4 96.6 96.3	91.7 89.1 75.1
$\gamma_B^{2.5}$					96.7 96.4 97.1

10 N=15 30	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2	$\gamma_B^{2.5}$
γ_A^0	94.9 94.9 95.4	81.8 74.0 40.0	24.1 2.8 0.0	0.0 0.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		95.3 95.2 94.3	89.3 86.1 70.5	9.4 0.2 0.0	0.0 0.0 0.0
γ_B^1			95.1 95.3 95.8	59.7 30.2 0.4	8.0 0.2 0.0
γ_B^2				95.2 94.9 95.0	89.1 85.4 67.8
$\gamma_B^{2.5}$					95.7 94.8 95.0

Table 5.4: Acceptance rate in percentage of H_0 of ILLPerm (left) and MILLPerm (right). The samples are perturbed by an element of $\mathcal{I}_0\big(SO(3)\big)$.

5.2.3 Performances Including Pertubation by $\mathcal{S} = \mathcal{I}_0\big(SO(3)\big) \times \mathrm{Diff}^+[0,1]$

Design of the two-sample test in $\mathcal{X}/\mathcal{S}$ simulations. In this simulation we want to asses the influence of the temporal registration on the performance of ILLPerm and MILLPerm. The used error processes are identical to the error processes, which we used in the simulations with perturbation by $\mathcal{I}_0\big(SO(3)\big)$. However, we additionally introduce a time warping of the sample $\tilde{\eta}_1, ..., \tilde{\eta}_N$ i.e., we simulate the sample $\tilde{\eta}_1 \circ \phi, ..., \tilde{\eta}_N \circ \phi$, where

$$\phi(t) = t + \mathbb{1}_{[0.11,0.41]}(t) \frac{e^{-\frac{1}{2}\left(\frac{t-0.4}{0.1}\right)^2}}{0.1\sqrt{2\pi}} .$$

Before application of the ILLPerm the perturbation by ϕ and (P, Q) must be estimated and removed. This is done as described in 4.0.16. In principle, we could apply MILLPerm as described in Test 4.0.22 to deal with both the pertubation by $\mathcal{I}_0\big(SO(3)\big)$ and $\mathrm{Diff}^+[0,1]$ simultaneously. Estimating the the inverse of ϕ is,

however, computationally costly and therefore we apply MILLPerm only to the samples $\gamma_1, ..., \gamma_N$ and $\tilde{\eta}_1 \circ \phi \circ \hat{\phi}^{-1}, ..., \tilde{\eta}_N \circ \phi \circ \hat{\phi}^{-1}$, where $\hat{\phi}^{-1}$ is estimated using Preprocessing 4.0.16.

We perform $M = 600$ simulations with $M_{perm} = 2000$ and report as before the acceptance rates.

Results of the two-sample test in $\mathcal{X}/\mathcal{S}$ simulations. We observe the same behavior, but less drastically, as in the case of the Hotelling T^2 test and the OCST and and a pertubation from $\mathcal{I}_0\big(SO(3)\big)$. The type I error increases with increasing sample size. However, recall that due to computational complexity we can not simulate the version of MILLPerm, which takes also the variance of the temporal alignment into account. We believe that the complete MILLPerm would perform well also in the setup of this simulation. This motivates further research.

10 **N=15** **30**	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2
γ_A^0	91.8 86.7 76.8	76.5 63.7 18.8	20.3 2.0 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		95.5 93.7 93.3	89.0 87.2 61.2	12.2 0.2 0.0
γ_B^1			93.5 94.3 92.8	60.8 24.2 0.4
γ_B^2				93.8 94.3 92.0

10 **N=15** **30**	γ_A^0	$\gamma_B^{0.5}$	γ_B^1	γ_B^2
γ_A^0	90.0 85.5 78.8	80.3 67.3 16.2	16.7 2.3 0.0	0.0 0.0 0.0
$\gamma_B^{0.5}$		95.8 95.5 92.2	88.2 85.2 56.8	9.8 0.3 0.0
γ_B^1			96.5 93.5 92.2	59.8 26.8 0.0
γ_B^2				93.7 94.5 90.5

Table 5.5: Acceptance rate in percentage of H_0 of ILLPerm (left) and MILLPerm correcting for $\mathcal{I}_0\big(SO(3)\big)$ (right). The samples are perturbed by an element of $\mathcal{S}$.

Chapter 6

Applications to Biomechanical Gait Data

"The theory of induction is the despair of philosophy – and yet all our activities are based upon it."

– Alfred N. Whitehead

In this chapter we apply previously developed statistical procedures to real gait data. The data within this thesis was collected in an experiment, designed and carried out by Michael Pierrynowski and Jodi Gallant, McMaster University, Canada. We organize this chapter as follows: first we will introduce the reader to the experimental protocol in Section 6.1 such that we can connect the results of our statistical analysis to the experimental setup. In Section 6.2 we explain the data processing, since we could not use all of the recorded data, and preprocessing steps like extraction of gait cycles are necessary. In Section 6.3 we discuss the results of our statistical analysis of the data and draw conclusions, among others, on identifiability of individuals (especially, after marker replacement), the influence of self-selected walking speeds and the effect of kneeling prior to recording walking trials.

6.1 Experimental Setup

Volunteers. Eight volunteers participated in this study. They were purposely selected to balance gender (male, female) and age (younger: 20-30 years; older: 50-60 years). Details are provided in Table 6.1. Volunteer selection was limited to healthy adults who reported that they had no prior history of knee ligament or meniscal injury, connective tissue disorders, or neurological disorders; additionally, no other injuries to the lower limbs must have occurred within the previous six months. The volunteers were asked not to perform vigorous activity resulting in excessive fatigue or strain of the lower limb muscles and joints for at least 24 hours prior to testing. The volunteers were provided with a verbal and written explanation of the study's protocol. After this briefing, the volunteers were given the option to continue with the study or to withdraw. Those wishing to continue signed a combined Hospital and University Research Ethics Board approved consent form of McMaster University, Canada.

Table 6.1: *volunteer characteristics*

Volunteer	Gender	Age (years)	Mass (kg)	Height (cm)	ASIS distance (cm)	Knee width (cm)
1	M	27	83	183	21.8	10.6
2	M	57	88	186	24.1	10.5
3	F	27	47	167	21.3	8.7
4	F	59	58	158	26.0	10.1
5	M	25	75	176	21.7	10.8
6	M	53	76	171	24.5	10.3
7	F	23	66	171	24.9	10.6
8	F	56	47	151	21.9	8.6

Lower extremity model and marker placement. Lower extremity anthropometric data and the mass and height of the volunteers were collected. Each volunteer's pelvis, thighs, lower legs and feet were instrumented with 16 reflective markers according to the modified Helen Hayes marker set Davis et al. [1991] described in the VCM protocol (Vicon, Oxford Metrics, London, UK). The VCM protocol combined the anthropometric data and a static view of the 14 markers plus four *temporary* markers placed on the medial sides of the knees and ankle axes while the volunteer quietly stood to construct a link-segment model of the lower extremity. The anthropometric measures and the marker placements were collected by the same technician for all volunteers during a four week interval. The marker locations were collected using an eight camera kinematic data acquisition system (Vicon, Oxford Metrics, London, UK) at 100 Hz.

General setup of a session. In each session, each volunteer had to complete 14 barefoot walking trials at a self-selected comfortable walking speed over an 11 m straight-line distance starting at opposite ends of the walkway each trial. Directly, after collecting these trials they had to perform additional 14 trials at a self-selected fast walking speed.

In the middle of the walkway three in-line force plates (AMTI, Watertown, MA) were installed flush to the floor. The force plates provided the ground reaction force applied to the floor by the volunteer's feet and therefore provided knowledge of when the volunteer's right and left feet were in contact with the floor (force threshold: 25 N). It was intended to use these data to extract single gait cycles from trials.

Once before the sessions were collected, a minimum of five practice walks were performed to ensure participants' comfort.

Collected sessions. The experiment consisted of six different sessions named A, B, C, D, E and F. After the completion of the 14 comfortable walking speed trials and 14 fast walking speed trials of Sessions A, B, C and D, the eight *lateral* markers were removed and the skin cleaned to remove knowledge of the location of where these markers were placed to examine the sensitivity of knee orientation

estimates on marker placement. Moreover, between the sessions the technician processed previously collected data while the volunteer performed unrelated activities such as reading or game playing for approximately 10 minutes. At the start of the next sessions (B to E) the eight lateral and four VCM medial markers were reapplied, the static standing trial collected, the VCM model constructed, the four medial markers removed, and the walking trials collected.

Sessions A and B were designed to understand the effect of bad marker placement. Some of the markers were purposely placed wrongly. Note that the same wrong positioning of the markers was used for all evenly numbered volunteers and a different wrong, but fixed, marker placement was used for all odd numbered volunteers.

Sessions C and D were replicate sessions used to examine the sensitivity of the data to marker placement and volunteer inter-session variations. Ideally, a statistical analysis should be able to identify and distinguish volunteers between sessions C and D.

Sessions E (kneeling) and F (prolonged kneeling) were similar to Sessions A to D with the following two exceptions. Firstly, the markers were not removed/replaced between these sessions. This was done to maximize the kneeling effect by minimizing the delay between the sessions. Secondly, a kneeling challenge was introduced just prior to recording the walking trials. The volunteer knelt on a stiff foam mat for 15 minutes. This kneeling posture required that the dorsal aspect of the volunteer's feet were in contact with the floor surface, their buttocks rested on their heels, and their torso held perpendicular to the ground. Immediately following the 15 minute kneeling interval the volunteer walked about 3 m to perform the 28 walking trials. The delay from the end of kneeling to the start of the first walking trials was between 20 and 50 seconds.

At the completion of Session F all of the markers were removed and the volunteer thanked for her/his participation. A summary of session activities which includes the marker placement/replacement and kneeling timing is provided in Table 6.2.

6.2 Data Processing.

The constructed lower extremity model (VCM model) and the motions of the markers were used by the Plug-in-Gait Software (Vicon, Oxford Metrics, London, UK) to calculate and output orientations of the left and right lower leg relative to the thigh, during each trial. This $SO(3)$-valued time series was parameterized using the Euler angle sequence y-x-z (see Appendix A.2), which is the typical used sequence by biomechanists, see Grood and Suntay [1983]. Here y denotes the flexion/extension about the thigh's medial-lateral axis, x is the adduction/abduction about an anterior-posterior axis and z gives axial rotation about the lower leg's longitudinal axis (see Figure 2).

Gait cycle extraction. The recorded $SO(3)$-valued time series of a walking trial usually contains 2-3 gait cycles. Since consecutive gait cycles are dependent,

we have to extract for each trial exactly one gait cycle in order to obtain independent data. Here, we define a gait cycle as the time between two consecutive

Activity	A	B	C	D	E	F
volunteer consent obtained	X	-	-	-	-	-
8 non-lateral markers applied	X	-	-	-	-	-
8 lateral markers applied	X	X	X	X	X	-
4 medial markers applied	X	X	X	X	X	-
standing trial collected	X	X	X	X	X	-
4 medial markers removed	X	X	X	X	X	-
15 minute kneeling intervention	-	-	-	-	X	X
walking trials collected	X	X	X	X	X	X
8 lateral markers removed	X	X	X	X	-	X
8 non-lateral markers removed	-	-	-	-	-	X
volunteer thanked	-	-	-	-	-	X

Table 6.2: *ordered data collection activities during each session (read column-wise)*

maximal flexion events (see Figure 3). These events correspond to mid-swing, when the knee was maximally flexed. If possible, we used the gait cycle closest to the middle of the walkway.

This atypical gait presentation is introduced to automatically select gait cycles. Moreover, this definition does not produce inconsistencies as other definitions we tried: Figure D.5 shows an example of automatically detected consecutive heel contacts using the force plate data, which is a usually used definition of a gait cycle. Around heel contacts this definition seems to produce too many inconsistent gait cycles. A second approach using analytical defined heel contacts via a local minimum of knee flexion (i.e., y-Euler angle) does produce similar inconsistencies. For our data set our new definition was reliable in extracting consistent gait cycles. This improvement might be due to the absence of external forces (ground contact in case of heel contact) in the swing phase.

Although in each session 14 comfortable walking speed trials and 14 fast walking speed trials were collected, it was not possible to extract always a complete properly recorded gait cycle. There are mainly two different sources of errors leading to removal of a trial from the data set. Firstly, since the recording time in each trial was manually initiated and terminated by the experimenter, it happened that some trials did not contain a full maximal flexion to maximal flexion gait cycle, because the original aim was to obtain at least one heel contact to heel contact gait cycle. Secondly, markers are sometimes hidden from camera sight for a short time period due to other extremities. These missing marker positions are usually predicted by an interpolation method. However, this method produced sometimes unrealistic discontinuities, in which case we chose another gait cycle within this trial. If this was not possible, we removed the complete trial from our data set.

6.3 Results of Statistical Analysis

In Chapter 4 we discussed different statistical tests and evaluated in Chapter 5 their performances. Since only MILLPerm (see Test 4.0.22) performs well under perturbation by gait similarities, we base our analysis mainly on this permutation test. Some questions, however, do not require the estimation of an aligning gait similarity between the compared sessions. In these cases we use ILLPerm (see Test 4.0.20), because our simulations showed that overall it has a slightly better power in this scenario than MILLPerm. However, MILLPerm produces no qualitative differences. The OCST will be used exclusively as a visualization tool to cross check the results of the permutation tests and provide insights at which time points the PEMs may differ, since its type I error rate decreased in our simulations, if marker placement effects are present.

For ILLPerm and MILLPerm we always used $M_{perm} = 20000$ permutations. In order to get an impression of the dependence of the p-values on the fact that we do not use all possible permutations, our modus operandi is to apply the permutation test ten times and report the mean p-value of these tests and its standard deviation.

Additionally, we try to improve the readibility of the tables by introducing a color code. We will report a *significant* p-value ($\alpha \leq 0.05$) in cyan and a *highly significant* p-value ($\alpha \leq 0.01$) in red.

6.3.1 Walk Data

6.3.1.1 Session C vs D: Identification of Volunteers

In this section we apply the permutation tests (ILLPerm, MILLPerm) as well as the OCST (see Test 4.0.17) to Sessions C and D and draw conclusion about modeling biomechanical gait data and the effect of marker placement and self-selected walking speeds. Note that there is no difference in the experimental setup of these sessions other than slightly different self-selected walking speeds and marker placements.

Sessionwise spatial alignment (SSA) necessary. The first important observation is that, although the markers where placed by a trained technican on pre-specified identifiable locations of the lower limb, we cannot identify all volunteers after marker replacement without any preprocessing. In fact, applying the overlapping simultaneous 90%-confidence regions test (OCST) or the permutation test (ILLPerm) to Sessions C and D without Preprocessing 4.0.16 results in extremely low identification rates. Actually, we cannot identify any volunteer using OCST (see Table 6.3(a)-(b)). Moreover, the identification rate for ILLPerm using a significance level of $\alpha = 0.01$ is also below 50% (see Table 6.3(c)-(d)). Better identification rates for ILLPerm stem from the observation that it is a feature of the ILLs that they automatically correct for the MP used to compute the coordinate system of one part of the leg (see Proposition 3.1.4). Hence, if the other coordinate system is only slightly changed by the corresponding MP, we observe a small value of the ILL resulting in a relatively high p-value.

Using the tests including SSA (i.e., Preprocessing 4.0.16 without estimating temporal registration) does improve the identification rate drastically. The OCST only does reject the equality of means for the right side for volunteer 3 and 5 after we corrected the confidence sets by estimated isometries which spatially align the compared sessions. However, visual inspection of the data reveals that especially for volunteer 3 the rejection could also stem from a STW effect (see Figures D.6 and D.7 in Appendix D). MILLPerm (we used Test 4.0.17, but always with $\hat{\phi}(t) = t$ for all $t \in I$) does also improve the identification rate. Here the improvement is not as drastic as for the OCST and it is not at all satisfactory for identification of volunteers.

Note that we do not have any false positives for the OCST, ILLPerm and MILLPerm in our tested scenarios and therefore we do not report these results in our tables.

(a) raw data (b) with SSA (c) raw data (d) with SSA

Vol	L	R	Vol	L	R
1	0	0	1	X	X
2	0	0	2	X	X
3	0	0	3	X	0
4	0	0	4	X	X
5	0	0	5	X	0
6	0	0	6	X	X
7	0	0	7	X	X
8	0	0	8	X	X

Vol	L	R
1	0.6 ± 0.1	2.3 ± 0.1
2	0.0 ± 0.0	0.3 ± 0.0
3	2.7 ± 0.1	0.0 ± 0.0
4	0.1 ± 0.0	0.1 ± 0.0
5	0.0 ± 0.0	0.1 ± 0.0
6	6.5 ± 0.2	29.2 ± 0.2
7	6.8 ± 0.2	1.0 ± 0.2
8	1.6 ± 0.1	1.8 ± 0.1

Vol	L	R
1	3.2 ± 0.1	5.8 ± 0.2
2	0.0 ± 0.0	0.8 ± 0.1
3	5.6 ± 0.1	0.0 ± 0.0
4	4.1 ± 0.2	0.2 ± 0.0
5	0.0 ± 0.0	0.8 ± 0.0
6	20.4 ± 0.2	38.6 ± 0.4
7	55.8 ± 0.4	10.3 ± 0.3
8	10.3 ± 0.2	57.4 ± 0.5

Table 6.3: *walk, Sessions C vs D, L=left knee, R=right knee. (a)-(b): OCST with* $\beta = 0.9$. *X = accepted, 0 = rejected. (c): p-values in* [%] *of ILLPerm applied to the data without spatial alignment. (d): p-values in* [%] *of MILLPerm.*

The influence of sessionwise time warping (STW). The results for the OCST are satisfactory for identification. However, we have seen in our simulations that this test does not perform as well as MILLPerm. The latter still detects after SSA differences between session C and D. Our suspicion is that these differences are due to STW effects. In order to remove them we estimate the STW effect between these sessions using Preprocessing 4.0.16. Afterwards we apply MILLPerm to the temporally registered sessions without taking the variance of the STW estimator into account. We decided to do it this way, since estimation of the STW is sensitive to the grid sizes and discretization and thus extremely time consuming, if performed in each permutation step. Moreover, our simulation suggests that for small sample sizes the type I error does not decrease dramatically, if we treat STW only as a preprocessing step.

C vs D: SSA/STW

Vol	L	R
1	23.6 ± 0.3	48.0 ± 0.4
2	39.2 ± 0.4	5.4 ± 0.2
3	66.3 ± 0.3	5.1 ± 0.1
4	53.8 ± 0.2	7.8 ± 0.2
5	6.2 ± 0.1	18.5 ± 0.3
6	53.7 ± 0.4	64.1 ± 0.5
7	97.2 ± 0.1	66.7 ± 0.3
8	60.7 ± 0.2	89.2 ± 0.3

Table 6.4: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Now, MILLPerm performs very well. We can identify all eight volunteers after removing SSA and STW with MILLPerm at a significance level $\alpha = 0.01$ and

$\alpha = 0.05$. Moreover, note that the p-values of volunteer 2 and 3 on the right side are close to 5%. More precisely 5% is within the 2σ region, which is expected since approximatively one out of 16 should have a p-value less than 5% under the null hypothesis. Note that all p-values for tests between different volunteers are zero.

We also mention that the OCST does slightly improve after STW. But we still have a rejection of the null hypothesis on the right side for volunteer 4, whom we would not have rejected without STW. However, the PESMs of these sessions visually look temporally better aligned (see Figures D.9 and D.10). Moreover, note that the null hypothesis for volunteer 3 is not rejected anymore, which by visual inspection could very well have been an artifacts of STW (see Figures D.7 and D.8). These are –additionally to the results of MILLPerm– reasonable arguments supporting the use of STW.

Individual time warping. Another question is whether there is an influence of slightly different speeds between the trials of a session. In order to incorporate this possibility we include as a preprocessing step the estimation of an ITW effect as presented in Section 3.1.1. Afterwards we again apply MILLPerm. Of course, in order to appropriately take this preprocessing step into account one would also have to include an estimation step into the permutation test, but this is computationally not feasible. The results given in Table 6.5 are somewhat surprising, since we would reject the null hypothesis of equal PEMs for the same volunteer in 62.5% of the cases using 0.01 as our critical value. There are two possible explanations for this observation. First, if we assume that ITW is an effect, which has to be taken into account for comparison of PEMs, we have to conclude that it is not possible to identify volunteers, since the human gait pattern varies significantly, even if there was no intervention between the sessions. In this case there is not much hope to detect differences in the gait pattern due to interventions, since these differences could also be due to the variation of the gait pattern in normal gait. However, our interpretation is that ITW should not be used to model human gait. The low identification rates are then due to artificial removal of natural variance of the data and thereby emphasizing differences existing only between the particular samples, but not between the populations. Figures D.12 and D.13 visualize this argument.

C vs D: SSA/STW/ITW

Vol	L	R
1	0.0 ± 0.0	0.0 ± 0.0
2	0.0 ± 0.0	0.0 ± 0.0
3	0.1 ± 0.0	0.0 ± 0.0
4	1.1 ± 0.1	0.0 ± 0.0
5	0.0 ± 0.0	0.0 ± 0.0
6	1.1 ± 0.1	0.5 ± 0.1
7	32.6 ± 0.2	2.5 ± 0.1
8	4.8 ± 0.2	6.5 ± 0.2

Table 6.5: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Individual spatial alignment (ISA). The last question about modeling human gait data inspired by the data at hand is, whether slightly different walking directions with respect to the camera system do influence the testing procedure (see Remark 1.1.3). Since testing after removing SSA and STW does well and since the volunteers had to walk between two prescribed lines we expect that this effect is negligible. Indeed, this is the case as shown in Table 6.6. Additionally, note that the p-values of MILLPerm using only SSA and STW are mostly within the 2σ regions of the p-values of MILLPerm including also ISA and vice versa. This certainly suggests that SSA does not need to be modeled in experiments with a similar design as the one we study here, where the positioning of the camera is stable with respect to the walking direction.

C vs D: SSA/ISA/STW

Vol	L	R
1	23.1 ± 0.2	47.8 ± 0.2
2	39.1 ± 0.4	5.2 ± 0.1
3	66.4 ± 0.3	5.1 ± 0.1
4	53.9 ± 0.3	7.9 ± 0.2
5	6.2 ± 0.2	18.4 ± 0.1
6	53.5 ± 0.4	64.5 ± 0.3
7	97.1 ± 0.1	66.7 ± 0.3
8	60.7 ± 0.4	89.1 ± 0.2

Table 6.6: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Conclusions. Our statistical analysis suggests that, if one wants to compare human gait data of two volunteers stemming from different sessions, it is necessary to include STW and SSA in order to remove unwanted differences due to slightly different marker placements and self-selected walking speeds. On this basis we cannot reject the null hypothesis of equal PEMs of Sessions C and D for the same volunteer using MILLPerm, whereas between volunteer comparison always rejects this null hypothesis. However, it would be interesting to verify these findings on a larger population than available in this particular study. Furthermore, it would be desirable to have a higher sample size per volunteer, since our simulations (see Table 5.5) suggests, that for small sample sizes of $N \approx 10$-15 the type II error for small deviations in the PEM is relatively high, but decreases quickly, if the sample size grows.

Last but not least, we recommend not to do individual time warping, if the self-selected walking speeds are similar, since a slightly different timing between trials of a sessions seems to be an expression of the natural variability of human gait. Lastly, we observed that under laboratory conditions the effect of the walking direction with respect to the camera system is negligible.

6.3.1.2 Session E and F: The Influence of Kneeling

The observation in the previous section that it is possible to reliably identify volunteers in different sessions is a necessary preliminary step in order to detect changes in human gait after the kneeling task. Hence, we will now apply MILLPerm with STW preprocessing and OCST with Preprocessing 4.0.16 to sessions E and F. Recall that there was a kneeling task of approximatively 15 minutes before Session E and Session F and that the markers were not replaced between these sessions.

The effect of kneeling. In order to detect changes in the gait pattern of our volunteers due to the kneeling task we use MILLPerm and the OCST to find differences between Session C/D and E for the short kneeling and between Session C/D and F for prolonged kneeling respectively. The results of MILLPerm are summarized in Table 6.8, whereas the results of the OCST are given in Table 6.7. Both tests agree that we have differences in the walking pattern for volunteer 2, 4 and 6 after kneeling. In fact, using MILLPerm these differences are often highly significant ($\alpha \leq 0.01$). Moreover, volunteer 7 has a mostly consistent significant kneeling effect ($\alpha \leq 0.05$) detected by MILLPerm. However, these differences are not detected by the OCST. Inspection of the data reveals, that especially in Session F the variance of this volunteer is extremely high (see Figure D.14 and D.15). This is probably due to the reported discomfort of volunteer 7, while recording the data after the prolonged kneeling task. Since for small sample sizes the simultaneous confidence regions usually include most of the sample curves, it is not surprising that a difference in the PEM due to kneeling using the OCST cannot be detected in the case of high variances.

various sessions: SSA/STW

Vol	EC		ED		FC		FD	
	L	*R*	*L*	*R*	*L*	*R*	*L*	*R*
1	X	0	X	X	0	0	0	X
2	0	0	0	0	0	0	0	0
3	X	X	X	0	0	X	X	0
4	X	0	X	0	0	0	0	0
5	X	X	X	X	X	X	X	X
6	0	0	0	X	0	0	0	0
7	X	X	0	X	X	X	X	X
8	X	X	X	0	X	X	X	0

Table 6.7: *walk, rejection/acceptance using OCST with* $\beta = 0.9$.

various sessions: SSA/STW

Vol	E vs C		E vs D		F vs C		F vs D	
	L	R	L	R	L	R	L	R
1	20.4 ± 0.3	22.5 ± 0.3	15.8 ± 0.3	86.4 ± 0.2	2.9 ± 0.1	3.3 ± 0.1	12.7 ± 0.2	24.9 ± 0.4
2	4.6 ± 0.1	1.2 ± 0.1	0.2 ± 0.0	0.7 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
3	87.2 ± 0.3	3.8 ± 0.2	30.7 ± 0.3	0.2 ± 0.0	19.1 ± 0.2	6.7 ± 0.1	31.1 ± 0.2	4.9 ± 0.1
4	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
5	21.4 ± 0.4	22.7 ± 0.3	73.5 ± 0.4	39.5 ± 0.6	55.9 ± 0.3	10.8 ± 0.1	35.5 ± 0.3	48.7 ± 0.3
6	0.0 ± 0.0	2.7 ± 0.1	0.0 ± 0.0	0.6 ± 0.1	1.0 ± 0.1	1.0 ± 0.1	0.8 ± 0.1	0.3 ± 0.0
7	0.0 ± 0.0	2.3 ± 0.1	0.0 ± 0.0	0.8 ± 0.1	2.7 ± 0.1	6.4 ± 0.1	4.2 ± 0.1	3.7 ± 0.1
8	46.7 ± 0.5	4.0 ± 0.1	70.5 ± 0.3	0.3 ± 0.0	10.2 ± 0.3	3.1 ± 0.1	14.9 ± 0.2	1.3 ± 0.1

Table 6.8: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Volunteer 3 and 8 seem to have a kneeling effect only on the right limb, whereas drawing any conclusions about a kneeling effect for volunteer 1 is difficult. For this volunteer both tests agree that there is a kneeling effect between Session F and C, but to be consistent this effect should also be visible between Session F and D, which is not consistently the case. Interestingly, the only volunteer seemingly

not at all affected by the kneeling task is volunteer 5, who was reported to be a plumber whose occupation required that he work on his knees frequently..

The OCST allows to localize the time points, where the two confidence regions do not overlap and thereby indicates time regions, where significant differences may have appeared. It is noteworthy that often this differences seem to be located around heel contact and toe off (i.e., in the transition phase between swing and stance phase). Examples showing localizations of differences are given in Figures D.16, D.17, D.18 and D.19. Moreover, Tables C.10 and C.10 summarize these localizations for all volunteers.

Last but not least, we mention that the p-values of MILLPerm between volunteers are always zero, which indicates that within-individual effects are much smaller than between-individual effects.

Short kneeling versus prolonged kneeling. After we compared Sessions E and F, which are recorded after a kneeling intervention, with Sessions C and D consisting of trajectories of normal gait, we still have to answer the question, whether there is an effect of prolonged kneeling. Note that between Session E and F the markers were not replaced. Hence we expect that we do not need to take SSA into account. This is also suggested by ILLPerm, where we only corrected for STW this time (see Table 6.9). However, inspection of the data and the OCST suggest that a MP effect is also visible between these sessions, especially for the left limb of volunteer 2 (see Figure D.20). A possible explanation is that the markers stayed attached to the legs during the kneeling task. Thus, small perturbations of their positions may have occurred due to stretching of the skin. Hence, we propose to use MILLPerm with a STW preprocessing even, if the markers between sessions where not replaced. Again note that it is not peculiar that ILLPerm may not detect these effects due to the previously discussed partial invariance of ILLs. The results of MILLPerm do not change the observation from ILLPerm, that we are not able to detect a significant difference between kneeling and prolonged kneeling (see Table C.6).

E vs F: STW

Vol	L	R
1	43.2 ± 0.5	66.8 ± 0.3
2	16.6 ± 0.2	12.9 ± 0.3
3	85.5 ± 0.2	61.7 ± 0.4
4	82.6 ± 0.2	20.3 ± 0.4
5	84.4 ± 0.2	83.5 ± 0.2
6	65.3 ± 0.3	64.7 ± 0.2
7	50.3 ± 0.3	69.4 ± 0.2
8	94.8 ± 0.2	31.6 ± 0.2

Table 6.9: *walk, p-values in* [%] *of ILLPerm with standard deviation.*

Conclusions. In a nutshell our analysis of the kneeling intervention often does detect significant differences between gait patterns of volunteers before and after kneeling. However, there is also a volunteer, who seems to be unaffected by this task implying that the influence of kneeling on the gait pattern may depend on the individual. Using the OCST we can localize the transition between swing and stance phase as a possible candidate, where differences between the PEMs occur.

An effect due to prolonged kneeling (i.e., testing Session E against F), is not detectable and, in fact, visual inspection of the data also seems not to support such a hypothesis.

Surprisingly, we also observed and discussed that due to the experimental setup

SSA is even necessary, if the markers are not replaced between sessions. Therefore we advice to use SSA before any comparison of human gait data with a similar experimental designs.

Lastly, we mention that, although it is from a statistical point of view interesting to have more trials within a session, it is in the case of studying interventions like kneeling expected that effects due to the intervention slowly fade away during sequential trials (see also the upcoming Section 6.3.2). This is a major challenge for any analysis trying to detect these effects.

6.3.2 Fast Data

As described in Section 6.1 directly after each walking speed session there were sessions collected with a self-selected fast walking speed. We will not repeat using this data the detailed modeling discussion given previously, since the conclusions using this data would be the same. Therefore, we shorten this section and focus only on the results using MILLPerm together with the STW preprocessing step.

various sessions: SSA/STW

Vol	C vs D		E vs F	
	L	R	L	R
1	69.4 ± 0.4	10.4 ± 0.4	69.6 ± 0.2	10.9 ± 0.2
2	33.9 ± 0.3	10.8 ± 0.2	95.5 ± 0.1	88.4 ± 0.2
3	99.4 ± 0.1	50.8 ± 0.3	58.8 ± 0.2	85.3 ± 0.2
4	79.2 ± 0.3	21.1 ± 0.2	80.0 ± 0.3	85.2 ± 0.2
5	67.8 ± 0.4	41.3 ± 0.4	98.9 ± 0.0	90.6 ± 0.2
6	32.0 ± 0.3	95.9 ± 0.1	33.7 ± 0.3	98.4 ± 0.0
7	97.3 ± 0.1	77.9 ± 0.4	13.1 ± 0.3	1.7 ± 0.1
8	43.9 ± 0.3	98.1 ± 0.1	57.8 ± 0.3	87.9 ± 0.3

Table 6.10: *fast, p-values in* [%] *of MILLPerm with standard deviation.*

First, we observe that we are again able to detect all eight volunteers using MILLPerm (see Table 6.10). Due to recording of the fast walk data directly after the walk session, we expect that the kneeling effect may have slowly faded away resulting in less significant differences in the fast walk data between the sessions recorded after the kneeling task (E and F) and the control sessions (C and D). A comparison of Table 6.11 and Table 6.8 reveals that we have 27 highly significant and 40 significant p-values between kneeling and control sessions in the walk data, whereas we only have 14 highly significant and 25 significant p-values for the fast walk data. This comparison, however, has to be taken with care, since the observed differences could also be for example, due to different speeds. It is at least imaginable, that a higher speed may alter the individual gait pattern and/or the effect of kneeling on the gait pattern may depend on the walking speed. The latter hypothesis is mildly supported by the observation that volunteer 1 and volunteer 8 have some significant p-values, which are not consistently observed in walk and fast walk data simultaneously. However, this could also be a random effect and in order to answer this question another experiment specifically targeting this question has to be designed.

various sessions: SSA/STW

Vol	E vs C		E vs D		F vs C		F vs D	
	L	R	L	R	L	R	L	R
1	23.3 ± 0.4	0.4 ± 0.0	3.7 ± 0.2	11.2 ± 0.2	14.6 ± 0.2	6.1 ± 0.1	7.8 ± 0.1	8.5 ± 0.2
2	26.1 ± 0.4	6.1 ± 0.2	2.1 ± 0.1	1.6 ± 0.1	20.3 ± 0.3	7.8 ± 0.2	0.9 ± 0.1	3.6 ± 0.1
3	96.2 ± 0.1	41.8 ± 0.5	99.7 ± 0.0	77.2 ± 0.3	19.9 ± 0.1	14.2 ± 0.3	35.6 ± 0.4	5.6 ± 0.1
4	0.0 ± 0.0	7.2 ± 0.2	0.0 ± 0.0	1.0 ± 0.1	0.0 ± 0.0	18.1 ± 0.2	0.0 ± 0.0	5.5 ± 0.1
5	47.3 ± 0.3	58.6 ± 0.5	80.7 ± 0.2	40.7 ± 0.3	35.9 ± 0.3	12.7 ± 0.2	74.7 ± 0.3	4.1 ± 0.1
6	0.1 ± 0.0	3.1 ± 0.1	0.3 ± 0.0	4.1 ± 0.2	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.3 ± 0.0
7	2.9 ± 0.1	17.4 ± 0.4	8.8 ± 0.2	2.8 ± 0.1	2.4 ± 0.1	48.3 ± 0.4	0.7 ± 0.1	25.6 ± 0.3
8	7.9 ± 0.2	24.7 ± 0.4	39.1 ± 0.2	12.4 ± 0.2	2.4 ± 0.0	33.0 ± 0.3	17.7 ± 0.2	25.0 ± 0.4

Table 6.11: *fast, p-values in* [%] *of MILLPerm with standard deviation.*

6.3.3 Walk vs Fast Data

In this section we explore the influence of the walking speed on the gait pattern. For this purpose we compare Sessions C and D of the walk and fast walk data. We study this question again using MILLPerm with STW preprocessing. The results of this comparison are reported in Table 6.12. With the exception of volunteer 4 the gait patterns of walk and fast walk of a volunteer are always significantly different. Therefore it is possible that also the kneeling task has a different effect, if the speed changes. Moreover, the OCST and visual inspection of the data suggests that the differences between the sessions can often be located in the stance phase (see Figures D.21 and D.22).

various sessions: SSA/STW

Vol	Walk C vs Fast D		Fast C vs Walk F	
	L	R	L	R
1	0.0 ± 0.0	0.5 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
2	1.4 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
4	12.5 ± 0.2	0.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
5	0.0 ± 0.0	0.1 ± 0.4	0.0 ± 0.0	0.0 ± 0.0
6	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
7	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
8	1.4 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Table 6.12: *walk versus fast, p-values in* [%] *of MILLPerm with standard deviation.*

The mean and standard deviation of the recorded frames of a volunteer for all sessions are reported in Table C.8 and C.9.

6.3.4 Session A and B: Improper Marker Placement

Application of MILLPerm to the sessions with on purpose badly placed markers (Sessions A and B) reveals that we only rarely accept the null hypothesis of equal PEMs (see Table C.7). This could have two explanations: firstly, it is not evident, how bad marker placements, which are not according to the protocol of the measurement device, influence the output of the proprietary software. Secondly, due to the chosen positioning of the markers, the high rejection rate could be due to soft tissue effects, since it is known that these effects reduce the reproducibility of gait patterns (among others, see Leardini et al. [2005] and Taylor et al. [2005b]).

Future Perspectives

Questions concerning rGP models. From a purely mathematical point of view there exists an interesting open question about rGP models. What is the pointwise Fréchet mean of an rGP and is it always unique and identical to the center curve of the PEM? Numerical computations of the Fréchet functional and some analytic calculations suggests that the uniqueness may depend on the variance matrix of the generating Gaussian process in the sense that if the eigenvalues are large uniqueness fails to hold. This could be related to the observation in Hotz and Huckemann [2015] that uniqueness of the Fréchet mean of the circle depends on the value of the density at the antipodal point of critical values of the Fréchet functional.

Another important question also for applications is the following: is it possible to improve Corollary 2.1.11? It would be desirable to prove that given an rGP γ with generating Gaussian process A there exists N' such that the PESM of a session $(\gamma_1, ..., \gamma_N)$ following γ is almost surely unique for all $N > N'$. In order to prove this one needs to show that

$$\mathbb{P}\left(\operatorname{rank}\left(\sum_{n=1}^{N} \operatorname{Exp}\Big(\iota\big(A_t^n\big)\Big)\right) > 1 \quad \text{for all } t \in [0,1]\right) = 1\,.$$

Especially, this result would imply that the PESM is almost surely a differentiable curve in $SO(3)$.

Hotelling process on $SO(3)$. There exists a open mathematical questions, which would be nice to answer. Are the Assumptions (*GKF 1*)–(*GKF 4*) (see Section 2.2) sufficient in order that the process Hotelling process on $SO(3)$ is always well-defined for all $t \in I = [0,1]$. To the best of our knowledge, the corresponding result for the Gaussian case and Hotelling T^2 processes seems to be open too. In applications, this is assumed to hold true in the literature (e.g., Taylor and Worsley [2008], Pataky et al. [2013]). However, it seems not to be trivial, since a special case including among other assumptions isotropicity of the Gaussian processes is proved in Cao et al. [1999].

ILLs and $\mathrm{Diff}^+[0,1]$-invariant estimators of the MP effect. Methodological there is a drawback to the numerical simple approach of estimating $(Q,P) \in \mathcal{I}_0\big(SO(3)\big)$ using the loss L_2 (viz., the estimator (3.29)). To see this assume that we want to match curves $\gamma, \eta \in \mathcal{X}$ with respect to the group $\mathcal{S}$ (i.e., finding

$(\psi, \phi) \in \mathcal{S}$ such that $\psi \circ \gamma \circ \phi$ is close to η) as is required for our gait data in order to remove the different walking speeds and marker placements between sessions.

Now, assume the perturbing element $\phi \in \mathrm{Diff}^+[0,1]$ is known. Then we obtain that in general

$$\Pi\left(\operatorname*{argmin}_{R\in SO(4)} L_2(R\tilde{\gamma}\circ\phi, \tilde{\eta})\right) \neq \Pi\left(\operatorname*{argmin}_{R\in SO(4)} L_2(R\tilde{\eta}\circ\phi^{-1}, \tilde{\gamma})\right)^{-1}, \tag{6.1}$$

since an application of the chain rule yields that in general

$$L_2(\tilde{\gamma}\circ\phi, \tilde{\eta}\circ\phi) \neq L_2(\tilde{\gamma}, \tilde{\eta})$$

for $\phi \in \mathrm{Diff}^+[0,1]$. Note that we call $\tilde{\eta}$ on the left hand side of (6.1) the reference curve, since $\tilde{\gamma}$ is rotated and time scaled such that it matches $\tilde{\eta}$. Then, however, equation (6.1) shows that the element $\psi \in \mathcal{I}_0\big(SO(3)\big)$ matching $\gamma, \eta \in \mathcal{X}$ using the loss L_2 depends in general on the choice, which of the two curve we use as the reference curve. This is of course an unfavorable property.

Circumventing this dependence requires a loss L, which is not only symmetric and $\mathcal{I}_0\big(SO(3)\big)$-invariant, but also $\mathrm{Diff}^+[0,1]$-invariant, since this implies $L(R\gamma \circ \phi, \eta) = L(R^{-1}\eta \circ \phi^{-1}, \gamma)$. Notably, it is possible to obtain a estimators for $\psi = (Q, P) \in \mathcal{I}_0\big(SO(3)\big)$ using the ILLs (see Definition 3.1.1), which are $\mathcal{S}$-invariant by Theorem 3.1.3 and hence do not suffer from the explained methodological shortcoming. To obtain these estimators note that by Theorem 3.1.4 we obtain

$$\begin{aligned}
\min_{P,Q\in SO(3)} \delta_{I,1}\big(P\gamma Q^T, \eta\big) &= \min_{P\in SO(3)} \delta_{I,1}\big(P\gamma, \eta\big) \\
\min_{P,Q\in SO(3)} \delta_{I,2}\big(P\gamma Q^T, \eta\big) &= \min_{P\in SO(3)} \delta_{I,2}\big(\gamma Q^T, \eta\big)
\end{aligned}$$

and hence estimators for the spatially aligning rotations Q and P are

$$\begin{aligned}
\hat{P}_{ILL} &\in \operatorname*{argmin}_{P\in SO(3)} \delta_{I,1}\big(P\gamma, \eta\big) \\
\hat{Q}_{ILL} &\in \operatorname*{argmin}_{Q\in SO(3)} \delta_{I,2}\big(\gamma Q^T, \eta\big)\,,
\end{aligned} \tag{6.2}$$

Interestingly, since P and Q can be estimated independently this reduces the problem of minimization over $SO(3) \times SO(3)$ to two minimization problems over $SO(3)$.

Since we could not prove consistency results for the estimators (6.2) and do not have a simple algortihm to solve the minimization problems, we used the loss L_2 in our application despite this rather important observation. Future work, however, should include a closer investigation of this estimators and further properties of the $\mathrm{Diff}^+[0,1]$-invariant ILLs, for example, whether always minimizing diffeomorphisms between two curves matched by an ILL exist (see e.g., Lahiri et al. [2015] and Bruveris [2015] in the framework of Srivastava et al. [2011b]) Moreover, a comparison of the ILL method with the TSRVF framework is desirable. Lastly, one could ask whether it is possible to extend the ILL framework to general manifolds?

Localizing differences between PEMs. Can we replace the OCST by simultaneuos confidence sets for the difference of PEMs? Does such a procedure perform better than the OCST, if time registration and spatial alignment is necessary?

Given two sessions $(\gamma_1, ..., \gamma_N)$ and $(\eta_1, ..., \eta_N)$ of rGPs γ and η with center curves $\gamma_0, \eta_0 \in \mathcal{X}$ and generating Gaussian processes A, B one possibility is to try to construct sets $\mathcal{V}_\beta(t) \subset SO(3)$ for $t \in I = [0,1]$ and $\beta \in (0,1)$ such that

$$\mathbb{P}\Big(\gamma_0(t)^T\eta_0(t) \in \mathcal{V}_\beta(t) \text{ for all } t \in I\Big) = \beta\,.$$

In principal this can be done following the same ideas as for the construction of SCS for the PEM in rGP models, since

$$\gamma\eta^t = \mathrm{Exp}\big(-\iota \circ A\big)\gamma_0^T\eta_0\mathrm{Exp}\big(\iota \circ B\big) \approx \gamma_0^T\eta_0\mathrm{Exp}\big(\iota \circ Z\big)$$

with an Gaussian process Z. Here we used Theorem 1.3.8 under the concentrated error assumption. The main problem is to obtain independent samples of Z in order to apply the Gaussian kinematic formula to estimate a critical value. A possibility, if $N = M$, could be to use the residual processes

$$Z^n = \iota^{-1} \circ \mathfrak{L}\Big(\hat{\eta}_N^T\hat{\gamma}_N\gamma_n\eta_n^T\Big)\,,$$

where $\hat{\gamma}_n, \hat{\eta}$ are a choice of PESMs. However, its validity and performance must be investigated further.

Statistics of temporal registration. In our simulations we did use only MILLPerm correcting for the variance in the spatial alignment procedure. It would be desirable to simulate the performance of MILLPerm including also the variation of temporal registration, since the version without temporal registration worked very well in the simulations including only marker placement effects. The problem here is mainly computational and faster implementations or improvements of the curve matching algorithm given in Section 3.1.2 could make simulations feasible.

Independent of the MILLPerm framework it is desirable to further investigate, how preprocessing influence statistical testing and find general remedies for the – at least in our simulations observed– phenomenon of increasing type I error rates, if preprocessing techniques are applied prior to testing.

Biomechanical questions. First of all, it would be desirable to apply the statistical methodologies presented in this thesis to data sets containing more volunteers. Moreover, there are plenty of questions about human gait, which could be explored, for example, does the kneeling effect depend on the chosen speed or is there a model for the effect of kneeling on the gait pattern? For speed related questions, hypothesis on the influence of speed on the gait pattern can be stated easily by describing the change of the flexion pattern of the knee in the stance phase. Unfortunately, for kneeling effects in the analyzed data there does not exist such a simple visible change of the gait pattern across volunteers. Hence, further research is necessary to explore this question.

Appendix A

Some Facts about Rotations

A.1 Properties of the Rotation Group

In what follows we collect some useful facts about the compact, connected Lie group of three-dimensional rotations $SO(3)$, which are all well-known in the literature. Its Lie algebra $\mathfrak{so}(3) = \{A \in \mathbb{R}^{3\times 3} : A^T = -A\}$ consists of 3×3 skew symmetric matrices. It is a three-dimensional linear subspace of all 3×3 matrices and thus carries the natural structure of $\mathbb{R}^3$. Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^3$ and let us endow the space of 3×3 matrices with the *Frobenius norm*

$$\|A\|_F = \frac{1}{2}\sqrt{\mathrm{tr}(\mathrm{A}\mathrm{A}^{\mathrm{T}})}\,. \tag{A.1}$$

We define the map $\iota : \mathbb{R}^3 \to \mathfrak{so}(3)$ defined by

$$\iota(a) = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}, \quad \text{for } a = (a_1, a_2, a_3)^T \in \mathbb{R}^3\,. \tag{A.2}$$

It follows at once by simple calculations that this map is an isometry.

Proposition A.1.1.

$$\iota : (\mathbb{R}^3, \|\cdot\|) \to (\mathfrak{so}(3), \|\cdot\|_F)$$

is an isometry

The isometry ι satisfies some useful properties.

Proposition A.1.2. *Let $a, b \in \mathbb{R}^3$ and $Q \in SO(3)$. Then*

(i) $\iota(a)\iota(b) = ba^T - a^T b I_{3\times 3}$

(ii) $\iota(a)^2 = aa^T - \|a\|^2 I_{3\times 3}$

(iii) $\iota(a)b = a \times b$

(iv) $Q\iota(a)Q^T = \iota(Qa)$

Proof. (i)-(iii): These are simple matrix computation. (iv): Note that the cross product satisfies $R(a \times b) = Ra \times Rb$ for all $a, b \in \mathbb{R}^3$ and all $R \in SO(3)$. Thus, we have by (iii) for all $b \in \mathbb{R}^3$

$$\iota(a)b = Q^T Q(a \times b) = Q^T\big(Qa \times Qb\big) = Q^T \iota(Qa)Qb\,,$$

which yields the claim. □

The *Lie exponential* of $SO(3)$ is given by the usual matrix exponential

$$\mathrm{Exp}:\ \mathfrak{so}(3) \to SO(3)\,, \quad A \mapsto \sum_{i=1}^{\infty} \frac{A^i}{i!} \tag{A.3}$$

which, in fact, is a surjection (see Price [1977, Corollary 4.3.5, p.96]). Using that $Q^TQ = I_{3\times3}$, where $I_{3\times3}$ denotes the identity matrix, we obtain the *naturality of the Lie exponential* i.e.,

$$Q\mathrm{Exp}(A)Q^T = \mathrm{Exp}(QAQ^T) \tag{A.4}$$

for all $A \in \mathfrak{so}(3)$ and $Q \in SO(3)$. Due to skew symmetry of $A \in \mathfrak{so}(3)$ the matrix exponential simplifies to

$$\mathrm{Exp}(A) = I_{3\times3} + \frac{\sin(\|A\|_F)}{\|A\|_F}\,A + \frac{1-\cos(\|A\|_F)}{\|A\|_F^2}\,A^2\,, \tag{A.5}$$

which is known as the *Rodriguez formula* (see Chirikjian and Kyatkin [2000]), and it yields that

$$\mathrm{Exp}(A) = \mathrm{Exp}\big(A + 2\pi k A/\|A\|_F\big) \text{ for all } A \in \mathfrak{so}(3),\ k \in \mathbb{Z}\,. \tag{A.6}$$

Hence, the inverse Log, called the *Lie logarithm*, is well-defined on

$$\mathcal{Q}_0 = \Big\{Q \in SO(3) \,\Big|\, \mathrm{tr}\big(\mathrm{Q}\big) > -1\Big\} \tag{A.7}$$

and maps $\mathcal{Q}_0$ onto

$$\mathfrak{U}_0 = \Big\{A \in \mathfrak{so}(3) \,\Big|\, \|A\|_F < \pi\Big\}\,.$$

For $Q \in \mathcal{Q}_0$ this inverse is explicitly given by

$$\mathrm{Log}(Q) = \frac{1}{2\mathrm{sinc}\big(\Theta(Q)\big)}\Big(Q - Q^T\Big),\ \text{where } \Theta(Q) = \mathrm{acos}\left(\frac{\mathrm{tr}(Q)-1}{2}\right) \tag{A.8}$$

as shown in Chirikjian and Kyatkin [2000, p. 121]. The set

$$\mathcal{C}_{I_{3\times3}} = SO(3)\backslash\,\mathcal{Q}_0 = \Big\{Q \in SO(3) \,\Big|\, \mathrm{tr}(\mathrm{Q}) = -1\Big\}$$

is called the *cut locus* of the identity matrix $I_{3\times3}$. Using the Lie logarithm we can define a distance on $SO(3)$ as follows

$$d_{SO(3)}(P,Q) = \begin{cases} \big\|\mathrm{Log}\big(PQ^T\big)\big\|_F, & \text{if } PQ^T \notin \mathcal{C}_{I_{3\times3}} \\ \pi & \text{else} \end{cases} \tag{A.9}$$

There is a useful relation between this intrinsic distance and the Frobenius distance.

Proposition A.1.3. *Let $P, Q \in SO(3)$ be arbitrary. Then*

$$\|P - Q\|_F = 2\sqrt{2}\sin\big(d_{SO(3)}(P,Q)/2\big) \tag{A.10}$$

Proof. See for example Stanfill et al. [2013, Section 2.2. & Supplementary]. □

A.2 Euler and Cardan angles

In the biomechanics community rotations are usually represented using Euler angles or Cardan angles. Euler angles represent a rotation by factorizing it into a series of three sequential rotations $R_x(\alpha), R_y(\beta), R_z(\upsilon)$ around the canonical coordinate axes given by

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$$
$$R_y(\beta) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix}$$
$$R_z(\upsilon) = \begin{pmatrix} \cos(\upsilon) & \sin(\upsilon) & 0 \\ -\sin(\upsilon) & \cos(\upsilon) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Any sequence of these three rotations such that no two consecutive rotations are around the same axis can be used as a possible Euler angle convention. There are twelve such sets in total – the first six use all three coordinate axes and are sometimes known as Cardan angles: x-y-z, x-z-y, z-x-y, z-y-x, y-z-x and y-x-z, while the other six have its first and last rotations about the same axis: x-y-x, y-z-y, z-x-z, x-z-x, y-x-y, z-y-z.

In our biomechanical application the y-x-z convention is used. Hence, a rotations R is factorized into

$$\begin{aligned} R &= R_z(\upsilon)R_x(\alpha)R_y(\beta) \\ &= \begin{pmatrix} c(\upsilon)c(\beta) - s(\upsilon)s(\alpha)s(\beta) & s(\upsilon)c(\alpha) & c(\upsilon)s(\beta) - s(\upsilon)s(\alpha)c(\beta) \\ s(\upsilon)c(\beta) - c(\upsilon)s(\alpha)s(\beta) & c(\upsilon)c(\alpha) & -s(\upsilon)s(\beta) + c(\upsilon)s(\alpha)c(\beta) \\ -c(\alpha)s(\beta) & s(\alpha) & c(\alpha)c(\beta) \end{pmatrix}. \end{aligned}$$

Here, we used $\sin(x) = s(x)$ and $\cos(x) = c(x)$ for all $x \in \mathbb{R}$. Note that this factorization is not always unique. This phenomenon is known as *gimbal lock* and appears for all Euler conventions. From a differential geometry point of view it is a realization of the fact, that there exists no global chart of $SO(3)$.

Appendix B

Other Technical Tools

In this appendix we will give some technical lemmas and theorems, which we used in the thesis. If they are well-known, it will be clarified before the result is stated and proved. Otherwise the results are lemmas, which we did not include directly into the thesis for the sake of a clear presentation.

Lemma B.0.1. *Let γ follow a rGP model. Then for each t the measure on $SO(3)$ induced by the random variable $\gamma(t)$ is absolutely continuous with respect to the volume measure induced by the biinvariant Riemannian metric of $SO(3)$.*

Proof. We have to show that any set of measure zero with respect to the volume measure vol_g of $SO(3)$ has measure zero with respect to $\mathbb{P}^{\gamma(t)}$. Let $\mathcal{D} \subset SO(3)$ be a set with $\mathrm{vol}_g(\mathcal{D}) = 0$. By definition of the volume measure there is a set

$$D \subset \iota^{-1} \circ \mathrm{Log}_{\gamma_0(t)}\big(SO(3) \backslash \mathcal{C}_{\gamma_0(t)}\big) \subset \mathbb{R}^3$$

with B Lebesgue measure zero such that $\gamma_0(t)\mathrm{Exp}(\iota \circ D) = \mathcal{D} \backslash \mathcal{C}_{\gamma_0(t)}$, where $\mathcal{C}_{\gamma_0(t)}$ denotes the cut locus of $\gamma_0(t)$. Now note that the pre-image $\mathrm{Exp}^{-1}(\mathrm{Exp}(\iota \circ D))$ is the union of countable many null sets with respect to the Lebesgue measure and hence a null set with respect to $\mathbb{P}^{\gamma(t)}$, since this measure has a density with respect to the Lebesgue measure. If $\mathcal{D}_0 = C_{\gamma(t)} \cap \mathcal{D} \neq \emptyset$, then again $\mathrm{Exp}^{-1}(\mathcal{D}_0)$ is a countable union of sets with Lebesgue measure zero. Thus $\mathbb{P}^{\gamma(t)}(\mathcal{D}) = 0$. □

Lemma B.0.2. *Suppose that the distribution of a random variable $X \in \mathbb{R}^3$ with existing first moment $\mathbb{E}[X]$ is even i.e. that $\mathbb{P}\{X \in M\} = \mathbb{P}\{-X \in M\}$ for all Borel sets $M \subset \mathbb{R}^3$ and $\mathbb{E}\left[\cos \|X\|\right] > 0$. Then $\mathbb{E}\big[\mathrm{Exp}(\iota \circ X)\big]$ is symmetric positive definite.*

Proof. With the Rodriguez formula (A.5) we have

$$\begin{aligned}\mathrm{Exp}(\iota \circ X) &= I_{3\times 3} + \iota \circ X \,\mathrm{sinc}\|X\| + (\iota \circ X)^2 \frac{1 - \cos \|X\|}{\|X\|^2} \\ &= \cos(\|X\|) I_{3\times 3} + \iota \circ X \,\mathrm{sinc}\|x\| + (1 - \cos(\|X\|)) \frac{XX^T}{\|X\|^2}\,,\end{aligned}$$

where the second equality is due to

$$\left(\iota\left(\frac{x}{\|x\|}\right)\right)^2 = \frac{xx^T}{\|x\|^2} - I_{3\times 3}\,. \tag{B.1}$$

By hypothesis, X is even, hence $\mathbb{E}[\iota \circ X \,\mathrm{sinc}\|X\|] = \iota \circ \mathbb{E}[X \,\mathrm{sinc}\|X\|] = 0$ which yields that

$$\mathbb{E}[\mathrm{Exp}(\iota \circ X)] = \mathbb{E}[\cos \|X\|] + \mathbb{E}\left[(1 - \cos \|X\|)\frac{XX^T}{\|X\|^2}\right]$$

is symmetric. Let $V \in \mathbb{R}^3$ be arbitrary with $\|V\| = 1$. Positive definiteness follows from the inequality

$$V^T \mathbb{E}\left[\mathrm{Exp}(\iota \circ X)\right] V = \mathbb{E}[\cos \|X\|] + \mathbb{E}\left[(1 - \cos \|X\|)\frac{(V^T X)^2}{\|X\|^2}\right] \geq \mathbb{E}[\cos \|X\|]\,,$$

where the last inequality is due to the Cauchy-Schwartz inequality and $1 - \cos \|X\| > 0$. □

The following two lemma (viz., Lemma B.0.3 and B.0.4) are well-known. We could not find good references and therefore we include their proofs.

Lemma B.0.3. *The set* $\mathcal{F} = \{A \in \mathbb{R}^{3\times 3} \mid \operatorname{rank}(A) \leq 1\}$ *is closed in* $\mathbb{R}^{3\times 3}$.

Proof. Let us denote by $m_{ij}(A)$ the 2×2 minor of a matrix $A \in \mathbb{R}^{3\times 3}$, where the i-th row and the j-th column is removed. Moreover, we define the map

$$f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}, \quad A \mapsto \begin{pmatrix} m_{11}(A) & m_{12}(A) & m_{13}(A) \\ m_{21}(A) & m_{22}(A) & m_{23}(A) \\ m_{31}(A) & m_{32}(A) & m_{33}(A) \end{pmatrix}$$

this map is continuous, since it is a composition of projections and the determinante map. Finally, note that $\mathrm{rk(A)} < \mathrm{k} + 1$ if and only if all $k+1$-minors are zero. Therefore we have $\mathcal{F} = f^{-1}(0)$ and thus $\mathcal{F}$ is closed, since f is continuous and $\{0\}$ is closed in $\mathbb{R}^{3\times 3}$.

□

Lemma B.0.4. *There exist an* $L > 0$ *such that* $\|\mathrm{Exp}(A) - \mathrm{Exp}(B)\|_F \leq L\|A - B\|_F$ *for all* $A, B \in \mathfrak{so}(3)$.

Proof. For each $A, B \in \mathfrak{so}(3)$ there is by the periodicity of Exp, cf. (A.6), an $\tilde{A}, \tilde{B} \in \overline{\mathfrak{U}}_0$ such that

$$\|\mathrm{Exp}(A) - \mathrm{Exp}(B)\|_F = \|\mathrm{Exp}(\tilde{A}) - \mathrm{Exp}(\tilde{B})\|_F \leq L\|\tilde{A} - \tilde{B}\|_F \leq L\|A - B\|_F ,$$

where the first inequality is due to $\mathrm{Exp} : \overline{\mathfrak{U}}_0 \to SO(3)$ is analytic and hence Lipschitz and the second is obvious from the definition of $\tilde{A}$ and $\tilde{B}$. □

The following is a small Lemma about our intrinsic residuals.

Lemma B.0.5. *Let* $P, Q \in SO(3)$ *be arbitrary, then*

$$\mathfrak{L}\big(QPQ^T\big) = \tau Q \mathfrak{L}\big(P\big)Q^T \tag{B.2}$$

for a $\tau \in \{1, -1\}$.

Proof. We can calculate using the naturality of the Lie exponential (see equation (A.4))

$$QPQ^T = Q\mathrm{Exp}\Big(\mathfrak{L}\big(P\big)\Big)Q^T = \mathrm{Exp}\Big(Q\mathfrak{L}\big(P\big)Q^T\Big)$$

Application of $\mathfrak{L} : SO(3) \to \mathfrak{V}$ to both sides implies that $\mathfrak{L}\big(QPQ^T\big) = \tau Q\mathfrak{L}\big(P\big)Q^T$, since

$$\mathfrak{L}\Big(\mathrm{Exp}\Big(Q\mathfrak{L}\big(P\big)Q^T\Big)\Big) = \begin{cases} Q\mathfrak{L}\big(P\big)Q^T & , \text{ if } Q\mathfrak{L}\big(P\big)Q^T \in \mathfrak{V} \\ -Q\mathfrak{L}\big(P\big)Q^T & , \text{ else}. \end{cases}$$

□

More general versions of the well-known Theorem B.0.9 on generic uniform convergence of stochastic process can be found in Newey [1991], Andrews [1992] and the book Davidson [1994]. For completeness we provide a proof of the statement tailored to our setting. In what follows $\|\cdot\|$ denotes a norm on $\mathbb{R}^D$, $D \in \mathbb{N}$.

Definition B.0.6 (Stochastic Equicontinuity)**.** *A sequence of stochastic processes* $\{X^n\}_{n\in\mathbb{N}}$ *is said to be stochastically equicontinuous, if there exists a set* $\Omega' \subset \Omega$ *with* $\mathbb{P}(\Omega') = 1$ *such that for every* $\omega \in \Omega'$ *and every* $\varepsilon > 0$ *there exist* $\delta = \delta(\omega, \varepsilon) > 0$ *and* $N_\varepsilon(\omega) \in \mathbb{N}$ *such that*

$$\sup_{|s-t|\leq\delta} \|X_t^n(\omega) - X_s^n(\omega)\| < \varepsilon$$

for all $n > N_\varepsilon(\omega)$.

Remark B.0.7. *A sufficient condition for stochastic equicontinuity of* $\{X^n\}_{n\in\mathbb{N}}$ *is that there exists a random variable* L*, which is almost surely bounded, such that for all* $\omega \in \Lambda$ *we have*

$$\sup_{s,t\in I} \|X_t^n(\omega) - X_s^n(\omega)\| < L(\omega)|s-t|$$

for all $n \in \mathbb{N}$.

Lemma B.0.8. *Let* $\{X^n\}_{n\in\mathbb{N}}$ *be stochastically equicontinuous and* $f \in C^1(I, \mathbb{R}^D)$*. Assume for every* $t \in I$ *that* $X_t^n \to f(t)$ *almost surely for* $n \to \infty$*. Then there exists a set* $\Omega' \subset \Omega$ *with* $\mathbb{P}(\Omega') = 1$ *such that* $X_t^n(\omega) \to f(t)$ *for* $n \to \infty$ *and every* $\omega \in \Omega'$ *and every* $t \in I$*.*

Proof. Let $t' \in \mathbb{Q} \cap I$. Then there exists a set $\Omega_{t'}$ with $\mathbb{P}(\Omega_{t'})$ such that $X_{t'}^n(\omega) \to f(t')$ for all $\omega \in \Omega_{t'}$. Define the set $\bar{\Omega} = \bigcap_{t\in\mathbb{Q}} \Omega_{t'}$. Since the intersection is countable we have that $\mathbb{P}(\bar{\Omega}) = 1$ and $X_{t'}^n(\omega) \to f(t')$ for all $\omega \in \Omega_{t'}$. Let $\tilde{\Omega} \subset \Omega$ be the set of equicontinuity and define $\Omega' = \bar{\Omega} \cap \tilde{\Omega}$, which again satisfies $\mathbb{P}(\Omega') = 1$.

We will now prove that the almost sure convergence of X_t^n to $f(t)$ holds true on all of Ω'. By equicontinuity we obtain that for all $\epsilon/3$ it exists a $0 < \delta$ such that

$$\sup_{|s-t|\leq\delta} \|X_t^n(\omega) - X_s^n(\omega)\| < \varepsilon/3$$

for all $\omega \in \Omega'$ and all $n > N_\omega$. Additionally, the definition of Ω' yields

$$\|X_{t'}^n(\omega) - f(t)\| < \varepsilon/3$$

for any $t' \in \mathbb{Q} \cap I$, all $\omega \in \Omega'$ and all $n > N'_\omega$.

Now, for any $t \in I$ it exists $t' \in \mathbb{Q} \cap I$ such that $|t - t'| < \min\{\delta, \varepsilon/3C\}$, where $C = \max_{t\in I} \|f'(t)\|$. Putting all this together results in

$$\|X_t^n(\omega) - f(t)\| \leq \|X_t^n(\omega) - X_{t'}^n(\omega)\| + \|X_{t'}^n(\omega) - f(t')\| + \|f(t') - f(t)\| < \varepsilon$$

for all $n > \max\{N_\omega, N'_\omega\}$ and $\omega \in \Omega'$. □

We are now able to prove a sufficient condition on almost sure uniform convergence (see also Davidson [1994, Theorem 21.8]).

Theorem B.0.9. *Let* $\{X^n\}_{n\in\mathbb{N}}$ *be stochastically equicontinuous and* $f \in C^1(I, \mathbb{R}^D)$*. Assume for every* $t \in I$ *that* $X_t^n \to f(t)$ *almost surely for* $n \to \infty$*. Then*

$$\sup_{t\in I} \|X_t^n - f(t)\| \to 0 \quad a.s. \tag{B.3}$$

for $n \to \infty$.

Proof. By Lemma B.0.8 we have that X^n is equicontinuos and pointwise convergent for all $t \in I$ on a set Ω' with $\mathbb{P}(\Omega') = 1$. We will now prove that for $\omega \in \Omega'$ the process is even uniformly convergent.

Let $\omega \in \Omega'$ be arbitrary. Choose an arbitrary $\varepsilon/3 > 0$. Then by equicontinuity we have that it exists $0 < \delta < \varepsilon/3C$, where $C = \max_{t\in I} \|f'(t)\|$, such that

$$\sup_{|s-t|\leq\delta} \|X_t^n(\omega) - X_s^n(\omega)\| < \varepsilon/3 \tag{B.4}$$

for all $n > N_\omega$.

Since I is compact we may divide I into a finite number of open sets $\{U_p\}_{p=1,...,P}$ such that $I \subset \bigcup_{p=1}^P U_p$ and $\sup_{s,t\in U_p} |t - s| \leq \delta$. Let $\{t_p\}_{p=1,...,P} \subset I$ such that $t_p \in U_p$. Pointwise convergence ensures the existence of an N_ω^p such that

$$\|X_{t_p}^n(\omega) - f(t_p)\| < \varepsilon/3 \tag{B.5}$$

for all $n > N_\omega^p$. Moreover, define $\tilde{N}_\omega = \max\{N_\omega^1, ..., N_\omega^P\}$. Then

$$\max_{p=1,...,P} \|X_{t_p}^n(\omega) - f(t_p)\| < \varepsilon/3 \tag{B.6}$$

for all $n > \tilde{N}_\omega$

Now, for any $t \in I$ we find an $p \in \{1, ..., P\}$ such that $t \in U_p$. By equation (B.4) we have that

$$\|X_t^n(\omega) - X_{t_p}^n(\omega)\| < \varepsilon/3 \tag{B.7}$$

for all $n > N_\omega$ and altogether we obtain

$$\|X_t^n(\omega) - f(t)\| \leq \|X_t^n(\omega) - X_{t_p}^n(\omega)\| + \|X_{t_p}^n(\omega) - f(t_p)\| + \|f(t_p) - f(t)\| < \varepsilon \tag{B.8}$$

for all $n > \max\{N_\omega, \tilde{N}_\omega\}$. Since $\max\{N_\omega, \tilde{N}_\omega\}$ and ϵ does not depend on t we have that $\sup_{t\in I} \|X_t^n(\omega) - f(t)\| < \varepsilon$ for all $n > \max\{N_\omega, \tilde{N}_\omega\}$ and thus we obtained uniform convergence for all $\omega \in \Omega'$. □

Appendix C

Tables

β	Cov.Rate	N	E.P.
85	95.30	5	$\varepsilon_{1,1}$
85	87.80	10	$\varepsilon_{1,1}$
85	86.86	15	$\varepsilon_{1,1}$
85	86.08	20	$\varepsilon_{1,1}$
85	85.78	30	$\varepsilon_{1,1}$
85	85.44	50	$\varepsilon_{1,1}$
85	95.82	5	$\varepsilon_{1,2}$
85	87.04	10	$\varepsilon_{1,2}$
85	85.88	15	$\varepsilon_{1,2}$
85	85.52	20	$\varepsilon_{1,2}$
85	85.26	30	$\varepsilon_{1,2}$
85	86.06	50	$\varepsilon_{1,2}$
85	95.10	5	$\varepsilon_{1,3}$
85	87.88	10	$\varepsilon_{1,3}$
85	85.90	15	$\varepsilon_{1,3}$
85	85.92	20	$\varepsilon_{1,3}$
85	85.32	30	$\varepsilon_{1,3}$
85	85.50	50	$\varepsilon_{1,3}$
85	99.06	5	$\varepsilon_{2,1}$
85	91.38	10	$\varepsilon_{2,1}$
85	88.34	15	$\varepsilon_{2,1}$
85	87.78	20	$\varepsilon_{2,1}$
85	84.82	30	$\varepsilon_{2,1}$
85	85.36	50	$\varepsilon_{2,1}$
85	99.38	5	$\varepsilon_{2,2}$
85	91.34	10	$\varepsilon_{2,2}$
85	88.46	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
85	85.98	20	$\varepsilon_{2,2}$
85	85.32	30	$\varepsilon_{2,2}$
85	85.26	50	$\varepsilon_{2,2}$
85	99.00	5	$\varepsilon_{2,3}$
85	91.56	10	$\varepsilon_{2,3}$
85	87.88	15	$\varepsilon_{2,3}$
85	86.80	20	$\varepsilon_{2,3}$
85	86.78	30	$\varepsilon_{2,3}$
85	84.72	50	$\varepsilon_{2,3}$
85	99.64	5	$\varepsilon_{3,1}$
85	94.82	10	$\varepsilon_{3,1}$
85	90.68	15	$\varepsilon_{3,1}$
85	88.36	20	$\varepsilon_{3,1}$
85	87.20	30	$\varepsilon_{3,1}$
85	85.06	50	$\varepsilon_{3,1}$
85	99.64	5	$\varepsilon_{3,2}$
85	94.24	10	$\varepsilon_{3,2}$
85	90.30	15	$\varepsilon_{3,2}$
85	88.74	20	$\varepsilon_{3,2}$
85	86.92	30	$\varepsilon_{3,2}$
85	84.24	50	$\varepsilon_{3,2}$
85	99.62	5	$\varepsilon_{3,3}$
85	94.64	10	$\varepsilon_{3,3}$
85	90.34	15	$\varepsilon_{3,3}$
85	88.70	20	$\varepsilon_{3,3}$
85	86.36	30	$\varepsilon_{3,3}$
85	85.86	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
90	98.58	5	$\varepsilon_{1,1}$
90	92.78	10	$\varepsilon_{1,1}$
90	90.50	15	$\varepsilon_{1,1}$
90	90.78	20	$\varepsilon_{1,1}$
90	90.54	30	$\varepsilon_{1,1}$
90	90.26	50	$\varepsilon_{1,1}$
90	98.12	5	$\varepsilon_{1,2}$
90	93.14	10	$\varepsilon_{1,2}$
90	92.20	15	$\varepsilon_{1,2}$
90	90.60	20	$\varepsilon_{1,2}$
90	90.18	30	$\varepsilon_{1,2}$
90	90.08	50	$\varepsilon_{1,2}$
90	98.34	5	$\varepsilon_{1,3}$
90	92.98	10	$\varepsilon_{1,3}$
90	91.14	15	$\varepsilon_{1,3}$
90	90.96	20	$\varepsilon_{1,3}$
90	90.42	30	$\varepsilon_{1,3}$
90	89.44	50	$\varepsilon_{1,3}$
90	99.78	5	$\varepsilon_{2,1}$
90	95.76	10	$\varepsilon_{2,1}$
90	93.16	15	$\varepsilon_{2,1}$
90	92.14	20	$\varepsilon_{2,1}$
90	90.62	30	$\varepsilon_{2,1}$
90	89.84	50	$\varepsilon_{2,1}$
90	99.74	5	$\varepsilon_{2,2}$
90	96.14	10	$\varepsilon_{2,2}$
90	93.38	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
90	92.02	20	$\varepsilon_{2,2}$
90	90.22	30	$\varepsilon_{2,2}$
90	90.30	50	$\varepsilon_{2,2}$
90	99.84	5	$\varepsilon_{2,3}$
90	95.50	10	$\varepsilon_{2,3}$
90	93.52	15	$\varepsilon_{2,3}$
90	92.62	20	$\varepsilon_{2,3}$
90	91.08	30	$\varepsilon_{2,3}$
90	90.74	50	$\varepsilon_{2,3}$
90	99.94	5	$\varepsilon_{3,1}$
90	98.06	10	$\varepsilon_{3,1}$
90	95.28	15	$\varepsilon_{3,1}$
90	93.36	20	$\varepsilon_{3,1}$
90	91.56	30	$\varepsilon_{3,1}$
90	91.12	50	$\varepsilon_{3,1}$
90	99.94	5	$\varepsilon_{3,2}$
90	97.58	10	$\varepsilon_{3,2}$
90	95.24	15	$\varepsilon_{3,2}$
90	93.16	20	$\varepsilon_{3,2}$
90	91.44	30	$\varepsilon_{3,2}$
90	91.32	50	$\varepsilon_{3,2}$
90	99.88	5	$\varepsilon_{3,3}$
90	97.80	10	$\varepsilon_{3,3}$
90	95.16	15	$\varepsilon_{3,3}$
90	93.40	20	$\varepsilon_{3,3}$
90	91.74	30	$\varepsilon_{3,3}$
90	90.30	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
95	99.80	5	$\varepsilon_{1,1}$
95	97.40	10	$\varepsilon_{1,1}$
95	96.00	15	$\varepsilon_{1,1}$
95	95.58	20	$\varepsilon_{1,1}$
95	95.62	30	$\varepsilon_{1,1}$
95	95.64	50	$\varepsilon_{1,1}$
95	99.82	5	$\varepsilon_{1,2}$
95	97.70	10	$\varepsilon_{1,2}$
95	96.32	15	$\varepsilon_{1,2}$
95	95.16	20	$\varepsilon_{1,2}$
95	95.38	30	$\varepsilon_{1,2}$
95	94.76	50	$\varepsilon_{1,2}$
95	99.70	5	$\varepsilon_{1,3}$
95	97.26	10	$\varepsilon_{1,3}$
95	96.46	15	$\varepsilon_{1,3}$
95	95.96	20	$\varepsilon_{1,3}$
95	95.44	30	$\varepsilon_{1,3}$
95	95.42	50	$\varepsilon_{1,3}$
95	100.0	5	$\varepsilon_{2,1}$
95	98.86	10	$\varepsilon_{2,1}$
95	97.42	15	$\varepsilon_{2,1}$
95	96.40	20	$\varepsilon_{2,1}$
95	96.50	30	$\varepsilon_{2,1}$
95	95.06	50	$\varepsilon_{2,1}$
95	100.0	5	$\varepsilon_{2,2}$
95	98.84	10	$\varepsilon_{2,2}$
95	97.38	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
95	96.76	20	$\varepsilon_{2,2}$
95	95.26	30	$\varepsilon_{2,2}$
95	95.08	50	$\varepsilon_{2,2}$
95	100.0	5	$\varepsilon_{2,3}$
95	98.98	10	$\varepsilon_{2,3}$
95	97.12	15	$\varepsilon_{2,3}$
95	96.80	20	$\varepsilon_{2,3}$
95	95.54	30	$\varepsilon_{2,3}$
95	95.60	50	$\varepsilon_{2,3}$
95	100.0	5	$\varepsilon_{3,1}$
95	99.56	10	$\varepsilon_{3,1}$
95	98.14	15	$\varepsilon_{3,1}$
95	97.34	20	$\varepsilon_{3,1}$
95	96.50	30	$\varepsilon_{3,1}$
95	95.14	50	$\varepsilon_{3,1}$
95	100.0	5	$\varepsilon_{3,2}$
95	99.58	10	$\varepsilon_{3,2}$
95	98.62	15	$\varepsilon_{3,2}$
95	97.52	20	$\varepsilon_{3,2}$
95	95.88	30	$\varepsilon_{3,2}$
95	95.76	50	$\varepsilon_{3,2}$
95	100.0	5	$\varepsilon_{3,3}$
95	99.40	10	$\varepsilon_{3,3}$
95	98.40	15	$\varepsilon_{3,3}$
95	97.50	20	$\varepsilon_{3,3}$
95	96.12	30	$\varepsilon_{3,3}$
95	95.30	50	$\varepsilon_{3,3}$

Table C.1: Simulated covering rate of simultaneous β%-confidence bands constructed using the naive bootstrap for 1D Gaussian processes obtained from $M = 5000$ simulations

β	Cov.Rate	N	E.P.
85	78.28	5	$\varepsilon_{1,1}$
85	82.24	10	$\varepsilon_{1,1}$
85	84.32	15	$\varepsilon_{1,1}$
85	83.74	20	$\varepsilon_{1,1}$
85	84.96	30	$\varepsilon_{1,1}$
85	84.30	50	$\varepsilon_{1,1}$
85	79.80	5	$\varepsilon_{1,2}$
85	83.82	10	$\varepsilon_{1,2}$
85	83.16	15	$\varepsilon_{1,2}$
85	83.68	20	$\varepsilon_{1,2}$
85	84.64	30	$\varepsilon_{1,2}$
85	85.38	50	$\varepsilon_{1,2}$
85	78.86	5	$\varepsilon_{1,3}$
85	83.06	10	$\varepsilon_{1,3}$
85	84.48	15	$\varepsilon_{1,3}$
85	84.08	20	$\varepsilon_{1,3}$
85	85.70	30	$\varepsilon_{1,3}$
85	85.86	50	$\varepsilon_{1,3}$
85	71.02	5	$\varepsilon_{2,1}$
85	76.74	10	$\varepsilon_{2,1}$
85	79.98	15	$\varepsilon_{2,1}$
85	81.78	20	$\varepsilon_{2,1}$
85	83.48	30	$\varepsilon_{2,1}$
85	83.56	50	$\varepsilon_{2,1}$
85	71.00	5	$\varepsilon_{2,2}$
85	78.56	10	$\varepsilon_{2,2}$
85	79.62	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
85	80.54	20	$\varepsilon_{2,2}$
85	82.14	30	$\varepsilon_{2,2}$
85	84.06	50	$\varepsilon_{2,2}$
85	70.90	5	$\varepsilon_{2,3}$
85	77.34	10	$\varepsilon_{2,3}$
85	80.30	15	$\varepsilon_{2,3}$
85	81.64	20	$\varepsilon_{2,3}$
85	82.74	30	$\varepsilon_{2,3}$
85	83.88	50	$\varepsilon_{2,3}$
85	56.92	5	$\varepsilon_{3,1}$
85	71.52	10	$\varepsilon_{3,1}$
85	76.32	15	$\varepsilon_{3,1}$
85	78.60	20	$\varepsilon_{3,1}$
85	80.76	30	$\varepsilon_{3,1}$
85	81.94	50	$\varepsilon_{3,1}$
85	56.84	5	$\varepsilon_{3,2}$
85	71.50	10	$\varepsilon_{3,2}$
85	76.26	15	$\varepsilon_{3,2}$
85	79.62	20	$\varepsilon_{3,2}$
85	79.76	30	$\varepsilon_{3,2}$
85	82.54	50	$\varepsilon_{3,2}$
85	58.02	5	$\varepsilon_{3,3}$
85	72.04	10	$\varepsilon_{3,3}$
85	75.82	15	$\varepsilon_{3,3}$
85	76.92	20	$\varepsilon_{3,3}$
85	81.02	30	$\varepsilon_{3,3}$
85	82.54	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
90	83.42	5	$\varepsilon_{1,1}$
90	86.82	10	$\varepsilon_{1,1}$
90	89.00	15	$\varepsilon_{1,1}$
90	88.72	20	$\varepsilon_{1,1}$
90	89.12	30	$\varepsilon_{1,1}$
90	88.86	50	$\varepsilon_{1,1}$
90	84.12	5	$\varepsilon_{1,2}$
90	87.08	10	$\varepsilon_{1,2}$
90	87.94	15	$\varepsilon_{1,2}$
90	88.74	20	$\varepsilon_{1,2}$
90	89.46	30	$\varepsilon_{1,2}$
90	89.06	50	$\varepsilon_{1,2}$
90	83.88	5	$\varepsilon_{1,3}$
90	88.34	10	$\varepsilon_{1,3}$
90	88.20	15	$\varepsilon_{1,3}$
90	88.26	20	$\varepsilon_{1,3}$
90	88.56	30	$\varepsilon_{1,3}$
90	90.06	50	$\varepsilon_{1,3}$
90	77.64	5	$\varepsilon_{2,1}$
90	83.26	10	$\varepsilon_{2,1}$
90	84.66	15	$\varepsilon_{2,1}$
90	86.06	20	$\varepsilon_{2,1}$
90	88.22	30	$\varepsilon_{2,1}$
90	88.48	50	$\varepsilon_{2,1}$
90	77.36	5	$\varepsilon_{2,2}$
90	82.80	10	$\varepsilon_{2,2}$
90	86.04	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
90	85.82	20	$\varepsilon_{2,2}$
90	88.40	30	$\varepsilon_{2,2}$
90	88.72	50	$\varepsilon_{2,2}$
90	77.72	5	$\varepsilon_{2,3}$
90	83.32	10	$\varepsilon_{2,3}$
90	85.28	15	$\varepsilon_{2,3}$
90	86.74	20	$\varepsilon_{2,3}$
90	87.80	30	$\varepsilon_{2,3}$
90	88.48	50	$\varepsilon_{2,3}$
90	67.50	5	$\varepsilon_{3,1}$
90	77.90	10	$\varepsilon_{3,1}$
90	81.94	15	$\varepsilon_{3,1}$
90	84.76	20	$\varepsilon_{3,1}$
90	85.14	30	$\varepsilon_{3,1}$
90	87.60	50	$\varepsilon_{3,1}$
90	66.68	5	$\varepsilon_{3,2}$
90	77.34	10	$\varepsilon_{3,2}$
90	82.00	15	$\varepsilon_{3,2}$
90	83.86	20	$\varepsilon_{3,2}$
90	86.12	30	$\varepsilon_{3,2}$
90	86.70	50	$\varepsilon_{3,2}$
90	68.12	5	$\varepsilon_{3,3}$
90	78.28	10	$\varepsilon_{3,3}$
90	82.02	15	$\varepsilon_{3,3}$
90	83.86	20	$\varepsilon_{3,3}$
90	86.80	30	$\varepsilon_{3,3}$
90	86.50	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
95	90.54	5	$\varepsilon_{1,1}$
95	91.84	10	$\varepsilon_{1,1}$
95	92.74	15	$\varepsilon_{1,1}$
95	93.34	20	$\varepsilon_{1,1}$
95	94.32	30	$\varepsilon_{1,1}$
95	95.16	50	$\varepsilon_{1,1}$
95	90.00	5	$\varepsilon_{1,2}$
95	92.44	10	$\varepsilon_{1,2}$
95	93.60	15	$\varepsilon_{1,2}$
95	93.82	20	$\varepsilon_{1,2}$
95	94.42	30	$\varepsilon_{1,2}$
95	94.22	50	$\varepsilon_{1,2}$
95	89.64	5	$\varepsilon_{1,3}$
95	92.86	10	$\varepsilon_{1,3}$
95	93.62	15	$\varepsilon_{1,3}$
95	93.44	20	$\varepsilon_{1,3}$
95	93.38	30	$\varepsilon_{1,3}$
95	94.78	50	$\varepsilon_{1,3}$
95	86.46	5	$\varepsilon_{2,1}$
95	89.72	10	$\varepsilon_{2,1}$
95	91.44	15	$\varepsilon_{2,1}$
95	92.50	20	$\varepsilon_{2,1}$
95	92.90	30	$\varepsilon_{2,1}$
95	94.08	50	$\varepsilon_{2,1}$
95	85.98	5	$\varepsilon_{2,2}$
95	88.82	10	$\varepsilon_{2,2}$
95	91.28	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
95	91.90	20	$\varepsilon_{2,2}$
95	93.26	30	$\varepsilon_{2,2}$
95	93.38	50	$\varepsilon_{2,2}$
95	85.52	5	$\varepsilon_{2,3}$
95	90.08	10	$\varepsilon_{2,3}$
95	90.96	15	$\varepsilon_{2,3}$
95	91.74	20	$\varepsilon_{2,3}$
95	93.40	30	$\varepsilon_{2,3}$
95	93.72	50	$\varepsilon_{2,3}$
95	79.62	5	$\varepsilon_{3,1}$
95	85.78	10	$\varepsilon_{3,1}$
95	88.40	15	$\varepsilon_{3,1}$
95	90.58	20	$\varepsilon_{3,1}$
95	92.06	30	$\varepsilon_{3,1}$
95	93.04	50	$\varepsilon_{3,1}$
95	80.00	5	$\varepsilon_{3,2}$
95	85.40	10	$\varepsilon_{3,2}$
95	88.62	15	$\varepsilon_{3,2}$
95	90.50	20	$\varepsilon_{3,2}$
95	91.90	30	$\varepsilon_{3,2}$
95	92.72	50	$\varepsilon_{3,2}$
95	79.94	5	$\varepsilon_{3,3}$
95	85.76	10	$\varepsilon_{3,3}$
95	88.36	15	$\varepsilon_{3,3}$
95	90.16	20	$\varepsilon_{3,3}$
95	92.84	30	$\varepsilon_{3,3}$
95	93.06	50	$\varepsilon_{3,3}$

Table C.2: Simulated covering rate of simultaneous β%-confidence bands constructed using the variance modified multiplier bootstrap for 1D Gaussian processes obtained from $M = 5000$ simulations

β	Cov.Rate	N	E.P.
85	84.54	5	$\varepsilon_{1,1}$
85	85.28	10	$\varepsilon_{1,1}$
85	84.82	15	$\varepsilon_{1,1}$
85	85.62	20	$\varepsilon_{1,1}$
85	85.00	30	$\varepsilon_{1,1}$
85	85.62	50	$\varepsilon_{1,1}$
85	84.80	5	$\varepsilon_{1,2}$
85	84.92	10	$\varepsilon_{1,2}$
85	85.06	15	$\varepsilon_{1,2}$
85	84.74	20	$\varepsilon_{1,2}$
85	85.78	30	$\varepsilon_{1,2}$
85	85.14	50	$\varepsilon_{1,2}$
85	85.82	5	$\varepsilon_{1,3}$
85	84.46	10	$\varepsilon_{1,3}$
85	85.46	15	$\varepsilon_{1,3}$
85	85.74	20	$\varepsilon_{1,3}$
85	84.72	30	$\varepsilon_{1,3}$
85	84.64	50	$\varepsilon_{1,3}$
85	86.50	5	$\varepsilon_{2,1}$
85	85.38	10	$\varepsilon_{2,1}$
85	85.84	15	$\varepsilon_{2,1}$
85	86.06	20	$\varepsilon_{2,1}$
85	86.40	30	$\varepsilon_{2,1}$
85	85.22	50	$\varepsilon_{2,1}$
85	86.78	5	$\varepsilon_{2,2}$
85	84.44	10	$\varepsilon_{2,2}$
85	85.54	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
85	84.98	20	$\varepsilon_{2,2}$
85	86.02	30	$\varepsilon_{2,2}$
85	85.98	50	$\varepsilon_{2,2}$
85	86.60	5	$\varepsilon_{2,3}$
85	85.14	10	$\varepsilon_{2,3}$
85	85.80	15	$\varepsilon_{2,3}$
85	86.32	20	$\varepsilon_{2,3}$
85	85.96	30	$\varepsilon_{2,3}$
85	86.02	50	$\varepsilon_{2,3}$
85	91.24	5	$\varepsilon_{3,1}$
85	90.44	10	$\varepsilon_{3,1}$
85	90.44	15	$\varepsilon_{3,1}$
85	90.50	20	$\varepsilon_{3,1}$
85	90.04	30	$\varepsilon_{3,1}$
85	90.52	50	$\varepsilon_{3,1}$
85	91.76	5	$\varepsilon_{3,2}$
85	90.98	10	$\varepsilon_{3,2}$
85	90.92	15	$\varepsilon_{3,2}$
85	90.28	20	$\varepsilon_{3,2}$
85	90.76	30	$\varepsilon_{3,2}$
85	90.42	50	$\varepsilon_{3,2}$
85	92.02	5	$\varepsilon_{3,3}$
85	90.38	10	$\varepsilon_{3,3}$
85	90.02	15	$\varepsilon_{3,3}$
85	90.46	20	$\varepsilon_{3,3}$
85	89.84	30	$\varepsilon_{3,3}$
85	90.44	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
90	90.28	5	$\varepsilon_{1,1}$
90	90.56	10	$\varepsilon_{1,1}$
90	89.18	15	$\varepsilon_{1,1}$
90	90.36	20	$\varepsilon_{1,1}$
90	90.86	30	$\varepsilon_{1,1}$
90	90.54	50	$\varepsilon_{1,1}$
90	89.76	5	$\varepsilon_{1,2}$
90	90.02	10	$\varepsilon_{1,2}$
90	90.24	15	$\varepsilon_{1,2}$
90	90.30	20	$\varepsilon_{1,2}$
90	90.24	30	$\varepsilon_{1,2}$
90	89.98	50	$\varepsilon_{1,2}$
90	89.68	5	$\varepsilon_{1,3}$
90	89.98	10	$\varepsilon_{1,3}$
90	89.92	15	$\varepsilon_{1,3}$
90	89.70	20	$\varepsilon_{1,3}$
90	90.16	30	$\varepsilon_{1,3}$
90	89.60	50	$\varepsilon_{1,3}$
90	91.50	5	$\varepsilon_{2,1}$
90	90.38	10	$\varepsilon_{2,1}$
90	90.76	15	$\varepsilon_{2,1}$
90	90.26	20	$\varepsilon_{2,1}$
90	90.34	30	$\varepsilon_{2,1}$
90	89.84	50	$\varepsilon_{2,1}$
90	90.34	5	$\varepsilon_{2,2}$
90	90.52	10	$\varepsilon_{2,2}$
90	90.70	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
90	90.50	20	$\varepsilon_{2,2}$
90	90.92	30	$\varepsilon_{2,2}$
90	90.58	50	$\varepsilon_{2,2}$
90	90.86	5	$\varepsilon_{2,3}$
90	90.90	10	$\varepsilon_{2,3}$
90	91.12	15	$\varepsilon_{2,3}$
90	90.68	20	$\varepsilon_{2,3}$
90	90.36	30	$\varepsilon_{2,3}$
90	90.24	50	$\varepsilon_{2,3}$
90	95.32	5	$\varepsilon_{3,1}$
90	93.86	10	$\varepsilon_{3,1}$
90	93.60	15	$\varepsilon_{3,1}$
90	93.30	20	$\varepsilon_{3,1}$
90	93.60	30	$\varepsilon_{3,1}$
90	93.32	50	$\varepsilon_{3,1}$
90	94.74	5	$\varepsilon_{3,2}$
90	93.50	10	$\varepsilon_{3,2}$
90	93.10	15	$\varepsilon_{3,2}$
90	93.24	20	$\varepsilon_{3,2}$
90	93.06	30	$\varepsilon_{3,2}$
90	93.36	50	$\varepsilon_{3,2}$
90	95.08	5	$\varepsilon_{3,3}$
90	93.38	10	$\varepsilon_{3,3}$
90	93.30	15	$\varepsilon_{3,3}$
90	93.44	20	$\varepsilon_{3,3}$
90	93.32	30	$\varepsilon_{3,3}$
90	93.38	50	$\varepsilon_{3,3}$

β	Cov.Rate	N	E.P.
95	95.56	5	$\varepsilon_{1,1}$
95	95.18	10	$\varepsilon_{1,1}$
95	94.84	15	$\varepsilon_{1,1}$
95	94.72	20	$\varepsilon_{1,1}$
95	95.06	30	$\varepsilon_{1,1}$
95	95.46	50	$\varepsilon_{1,1}$
95	95.32	5	$\varepsilon_{1,2}$
95	94.94	10	$\varepsilon_{1,2}$
95	95.38	15	$\varepsilon_{1,2}$
95	95.34	20	$\varepsilon_{1,2}$
95	95.00	30	$\varepsilon_{1,2}$
95	95.38	50	$\varepsilon_{1,2}$
95	94.82	5	$\varepsilon_{1,3}$
95	94.80	10	$\varepsilon_{1,3}$
95	94.78	15	$\varepsilon_{1,3}$
95	94.62	20	$\varepsilon_{1,3}$
95	95.14	30	$\varepsilon_{1,3}$
95	95.24	50	$\varepsilon_{1,3}$
95	95.64	5	$\varepsilon_{2,1}$
95	95.28	10	$\varepsilon_{2,1}$
95	95.54	15	$\varepsilon_{2,1}$
95	94.60	20	$\varepsilon_{2,1}$
95	95.32	30	$\varepsilon_{2,1}$
95	95.24	50	$\varepsilon_{2,1}$
95	95.96	5	$\varepsilon_{2,2}$
95	94.90	10	$\varepsilon_{2,2}$
95	95.28	15	$\varepsilon_{2,2}$

β	Cov.Rate	N	E.P.
95	95.18	20	$\varepsilon_{2,2}$
95	94.84	30	$\varepsilon_{2,2}$
95	94.64	50	$\varepsilon_{2,2}$
95	95.52	5	$\varepsilon_{2,3}$
95	95.24	10	$\varepsilon_{2,3}$
95	95.16	15	$\varepsilon_{2,3}$
95	95.42	20	$\varepsilon_{2,3}$
95	95.12	30	$\varepsilon_{2,3}$
95	94.90	50	$\varepsilon_{2,3}$
95	97.96	5	$\varepsilon_{3,1}$
95	96.58	10	$\varepsilon_{3,1}$
95	96.88	15	$\varepsilon_{3,1}$
95	96.68	20	$\varepsilon_{3,1}$
95	96.08	30	$\varepsilon_{3,1}$
95	96.54	50	$\varepsilon_{3,1}$
95	97.98	5	$\varepsilon_{3,2}$
95	96.48	10	$\varepsilon_{3,2}$
95	96.24	15	$\varepsilon_{3,2}$
95	96.68	20	$\varepsilon_{3,2}$
95	96.84	30	$\varepsilon_{3,2}$
95	96.96	50	$\varepsilon_{3,2}$
95	97.88	5	$\varepsilon_{3,3}$
95	96.62	10	$\varepsilon_{3,3}$
95	96.68	15	$\varepsilon_{3,3}$
95	96.76	20	$\varepsilon_{3,3}$
95	96.46	30	$\varepsilon_{3,3}$
95	96.90	50	$\varepsilon_{3,3}$

Table C.3: Simulated covering rate of simultaneous β%-confidence bands constructed using the GKF for 1D Gaussian processes obtained from $M = 5000$ simulations

β	Cov.Rate	N	σ	E.P.
85/90/95	84.7/90.3/94.6	10	0.005	$A^{1,1,1,\sigma}$
85/90/95	84.5/90.0/95.1	15	0.005	$A^{1,1,1,\sigma}$
85/90/95	84.9/90.1/94.9	30	0.005	$A^{1,1,1,\sigma}$
85/90/95	84.5/90.6/95.0	10	0.005	$A^{1,1,2,\sigma}$
85/90/95	85.8/90.3/95.0	15	0.005	$A^{1,1,2,\sigma}$
85/90/95	85.4/89.7/95.2	30	0.005	$A^{1,1,2,\sigma}$
85/90/95	85.7/90.9/95.0	10	0.005	$A^{1,3,1,\sigma}$
85/90/95	84.8/89.4/95.0	15	0.005	$A^{1,3,1,\sigma}$
85/90/95	84.5/90.2/94.7	30	0.005	$A^{1,3,1,\sigma}$
85/90/95	84.2/90.4/95.3	10	0.005	$A^{1,3,2,\sigma}$
85/90/95	84.7/89.9/95.0	15	0.005	$A^{1,3,2,\sigma}$
85/90/95	85.4/90.3/94.8	30	0.005	$A^{1,3,2,\sigma}$
85/90/95	86.1/91.0/95.0	10	0.05	$A^{1,1,1,\sigma}$
85/90/95	85.0/90.1/95.4	15	0.05	$A^{1,1,1,\sigma}$
85/90/95	85.1/91.0/94.9	30	0.05	$A^{1,1,1,\sigma}$
85/90/95	85.3/89.9/94.6	10	0.05	$A^{1,1,2,\sigma}$
85/90/95	85.4/89.8/95.4	15	0.05	$A^{1,1,2,\sigma}$
85/90/95	85.0/90.2/95.6	30	0.05	$A^{1,1,2,\sigma}$
85/90/95	84.8/90.0/95.3	10	0.05	$A^{1,3,1,\sigma}$
85/90/95	84.3/89.9/95.2	15	0.05	$A^{1,3,1,\sigma}$
85/90/95	84.7/90.1/95.2	30	0.05	$A^{1,3,1,\sigma}$
85/90/95	86.0/90.6/95.0	10	0.05	$A^{1,3,2,\sigma}$
85/90/95	84.9/90.0/94.7	15	0.05	$A^{1,3,2,\sigma}$
85/90/95	85.1/89.7/95.3	30	0.05	$A^{1,3,2,\sigma}$
85/90/95	84.7/90.8/94.9	10	0.1	$A^{1,1,1,\sigma}$
85/90/95	84.9/89.8/95.1	15	0.1	$A^{1,1,1,\sigma}$
85/90/95	85.0/90.5/95.1	30	0.1	$A^{1,1,1,\sigma}$
85/90/95	85.5/90.4/94.5	10	0.1	$A^{1,1,2,\sigma}$
85/90/95	85.4/89.9/94.7	15	0.1	$A^{1,1,2,\sigma}$
85/90/95	85.1/89.6/95.0	30	0.1	$A^{1,1,2,\sigma}$
85/90/95	85.4/90.1/96.0	10	0.1	$A^{1,3,1,\sigma}$
85/90/95	84.1/89.6/94.7	15	0.1	$A^{1,3,1,\sigma}$
85/90/95	85.4/90.3/94.9	30	0.1	$A^{1,3,1,\sigma}$
85/90/95	84.6/90.5/95.1	10	0.1	$A^{1,3,2,\sigma}$
85/90/95	85.2/90.2/95.1	15	0.1	$A^{1,3,2,\sigma}$
85/90/95	85.7/89.6/95.0	30	0.1	$A^{1,3,2,\sigma}$
85/90/95	82.4/87.7/93.9	10	0.6	$A^{1,1,1,\sigma}$
85/90/95	79.9/85.7/92.7	15	0.6	$A^{1,1,1,\sigma}$
85/90/95	79.4/85.5/92.4	30	0.6	$A^{1,1,1,\sigma}$
85/90/95	81.5/87.7/93.8	10	0.6	$A^{1,1,2,\sigma}$
85/90/95	81.9/86.8/93.1	15	0.6	$A^{1,1,2,\sigma}$
85/90/95	80.0/85.7/91.9	30	0.6	$A^{1,1,2,\sigma}$
85/90/95	83.0/88.7/94.7	10	0.6	$A^{1,3,1,\sigma}$
85/90/95	81.9/88.5/93.5	15	0.6	$A^{1,3,1,\sigma}$
85/90/95	80.2/86.7/93.1	30	0.6	$A^{1,3,1,\sigma}$
85/90/95	84.3/89.7/94.4	10	0.6	$A^{1,3,2,\sigma}$
85/90/95	81.5/86.8/93.5	15	0.6	$A^{1,3,2,\sigma}$
85/90/95	81.3/86.6/92.4	30	0.6	$A^{1,3,2,\sigma}$

β	Cov.Rate	N	σ	E.P.
85/90/95	87.1/90.1/95.0	10	0.005	$A^{2,1,1,\sigma}$
85/90/95	86.2/90.4/95.4	15	0.005	$A^{2,1,1,\sigma}$
85/90/95	86.3/90.5/94.7	30	0.005	$A^{2,1,1,\sigma}$
85/90/95	85.9/90.3/95.0	10	0.005	$A^{2,1,2,\sigma}$
85/90/95	86.5/90.6/95.1	15	0.005	$A^{2,1,2,\sigma}$
85/90/95	87.0/89.7/95.1	30	0.005	$A^{2,1,2,\sigma}$
85/90/95	86.1/89.5/94.9	10	0.005	$A^{2,3,1,\sigma}$
85/90/95	87.0/90.2/94.9	15	0.005	$A^{2,3,1,\sigma}$
85/90/95	85.4/90.5/95.6	30	0.005	$A^{2,3,1,\sigma}$
85/90/95	85.3/91.3/95.0	10	0.005	$A^{2,3,2,\sigma}$
85/90/95	85.9/90.7/95.2	15	0.005	$A^{2,3,2,\sigma}$
85/90/95	85.7/90.6/95.2	30	0.005	$A^{2,3,2,\sigma}$
85/90/95	85.3/90.1/95.6	10	0.05	$A^{2,1,1,\sigma}$
85/90/95	85.7/90.7/94.9	15	0.05	$A^{2,1,1,\sigma}$
85/90/95	86.4/90.6/94.7	30	0.05	$A^{2,1,1,\sigma}$
85/90/95	86.1/90.9/95.4	10	0.05	$A^{2,1,2,\sigma}$
85/90/95	85.9/90.5/94.9	15	0.05	$A^{2,1,2,\sigma}$
85/90/95	85.9/89.8/94.9	30	0.05	$A^{2,1,2,\sigma}$
85/90/95	86.2/90.9/95.5	10	0.05	$A^{2,3,1,\sigma}$
85/90/95	86.2/90.6/95.0	15	0.05	$A^{2,3,1,\sigma}$
85/90/95	86.6/90.8/94.9	30	0.05	$A^{2,3,1,\sigma}$
85/90/95	85.4/90.3/95.5	10	0.05	$A^{2,3,2,\sigma}$
85/90/95	85.4/90.5/95.3	15	0.05	$A^{2,3,2,\sigma}$
85/90/95	85.9/90.7/94.9	30	0.05	$A^{2,3,2,\sigma}$
85/90/95	85.2/91.4/95.4	10	0.1	$A^{2,1,1,\sigma}$
85/90/95	86.1/90.4/95.1	15	0.1	$A^{2,1,1,\sigma}$
85/90/95	85.8/91.1/95.5	30	0.1	$A^{2,1,1,\sigma}$
85/90/95	86.3/90.8/95.1	10	0.1	$A^{2,1,2,\sigma}$
85/90/95	86.1/89.9/95.3	15	0.1	$A^{2,1,2,\sigma}$
85/90/95	85.4/90.7/95.7	30	0.1	$A^{2,1,2,\sigma}$
85/90/95	85.4/90.2/94.6	10	0.1	$A^{2,3,1,\sigma}$
85/90/95	86.0/90.5/95.0	15	0.1	$A^{2,3,1,\sigma}$
85/90/95	85.3/90.1/95.3	30	0.1	$A^{2,3,1,\sigma}$
85/90/95	86.5/91.0/95.3	10	0.1	$A^{2,3,2,\sigma}$
85/90/95	86.2/89.8/95.3	15	0.1	$A^{2,3,2,\sigma}$
85/90/95	85.1/90.6/95.5	30	0.1	$A^{2,3,2,\sigma}$
85/90/95	81.6/87.3/93.6	10	0.6	$A^{2,1,1,\sigma}$
85/90/95	80.7/86.4/92.9	15	0.6	$A^{2,1,1,\sigma}$
85/90/95	78.7/84.8/92.3	30	0.6	$A^{2,1,1,\sigma}$
85/90/95	82.0/88.6/93.8	10	0.6	$A^{2,1,2,\sigma}$
85/90/95	81.0/87.1/93.2	15	0.6	$A^{2,1,2,\sigma}$
85/90/95	80.9/85.6/92.1	30	0.6	$A^{2,1,2,\sigma}$
85/90/95	84.2/88.8/94.2	10	0.6	$A^{2,3,1,\sigma}$
85/90/95	80.9/87.2/93.8	15	0.6	$A^{2,3,1,\sigma}$
85/90/95	80.0/86.3/92.8	30	0.6	$A^{2,3,1,\sigma}$
85/90/95	84.2/89.0/94.9	10	0.6	$A^{2,3,2,\sigma}$
85/90/95	81.6/87.2/94.0	15	0.6	$A^{2,3,2,\sigma}$
85/90/95	81.8/86.7/92.4	30	0.6	$A^{2,3,2,\sigma}$

Table C.4: Simulated covering rates of simultaneous β%-confidence sets for rGP models obtained from $M = 5000$ simulations. All the error process fulfill the assumptions necessary for application of the GKF.

β	Cov.Rate	N	σ	E.P.
85/90/95	90.6/93.5/96.8	10	0.005	$A^{3,1,1,\sigma}$
85/90/95	89.3/93.1/96.3	15	0.005	$A^{3,1,1,\sigma}$
85/90/95	90.3/93.3/96.1	30	0.005	$A^{3,1,1,\sigma}$
85/90/95	90.2/93.6/96.2	10	0.005	$A^{3,1,2,\sigma}$
85/90/95	90.7/92.5/96.6	15	0.005	$A^{3,1,2,\sigma}$
85/90/95	89.4/93.2/96.1	30	0.005	$A^{3,1,2,\sigma}$
85/90/95	90.5/93.9/96.5	10	0.005	$A^{3,3,1,\sigma}$
85/90/95	90.2/93.0/96.6	15	0.005	$A^{3,3,1,\sigma}$
85/90/95	90.2/92.7/96.7	30	0.005	$A^{3,3,1,\sigma}$
85/90/95	90.4/92.8/96.6	10	0.005	$A^{3,3,2,\sigma}$
85/90/95	89.7/93.0/96.0	15	0.005	$A^{3,3,2,\sigma}$
85/90/95	90.5/93.5/96.1	30	0.005	$A^{3,3,2,\sigma}$
85/90/95	90.4/93.9/96.6	10	0.05	$A^{3,1,1,\sigma}$
85/90/95	89.4/93.0/96.6	15	0.05	$A^{3,1,1,\sigma}$
85/90/95	90.1/93.5/96.5	30	0.05	$A^{3,1,1,\sigma}$
85/90/95	90.1/93.1/97.2	10	0.05	$A^{3,1,2,\sigma}$
85/90/95	90.3/93.0/96.7	15	0.05	$A^{3,1,2,\sigma}$
85/90/95	90.2/92.9/96.6	30	0.05	$A^{3,1,2,\sigma}$
85/90/95	91.0/93.6/97.1	10	0.05	$A^{3,3,1,\sigma}$
85/90/95	90.3/93.0/96.2	15	0.05	$A^{3,3,1,\sigma}$
85/90/95	90.0/92.6/96.5	30	0.05	$A^{3,3,1,\sigma}$
85/90/95	90.3/93.3/96.9	10	0.05	$A^{3,3,2,\sigma}$
85/90/95	90.1/93.5/97.3	15	0.05	$A^{3,3,2,\sigma}$
85/90/95	89.9/92.9/96.5	30	0.05	$A^{3,3,2,\sigma}$
85/90/95	90.3/93.4/96.7	10	0.1	$A^{3,1,1,\sigma}$
85/90/95	89.5/91.6/96.6	15	0.1	$A^{3,1,1,\sigma}$
85/90/95	89.9/92.7/96.3	30	0.1	$A^{3,1,1,\sigma}$
85/90/95	90.3/93.3/96.4	10	0.1	$A^{3,1,2,\sigma}$
85/90/95	89.9/93.1/95.9	15	0.1	$A^{3,1,2,\sigma}$
85/90/95	89.9/93.1/96.4	30	0.1	$A^{3,1,2,\sigma}$
85/90/95	90.1/93.6/97.0	10	0.1	$A^{3,3,1,\sigma}$
85/90/95	88.9/92.9/96.5	15	0.1	$A^{3,3,1,\sigma}$
85/90/95	88.9/93.4/96.5	30	0.1	$A^{3,3,1,\sigma}$
85/90/95	89.9/93.4/96.3	10	0.1	$A^{3,3,2,\sigma}$
85/90/95	89.8/93.1/96.2	15	0.1	$A^{3,3,2,\sigma}$
85/90/95	90.9/93.2/96.6	30	0.1	$A^{3,3,2,\sigma}$
85/90/95	87.1/91.2/95.5	10	0.6	$A^{3,1,1,\sigma}$
85/90/95	85.2/90.2/94.6	15	0.6	$A^{3,1,1,\sigma}$
85/90/95	82.8/87.6/92.9	30	0.6	$A^{3,1,1,\sigma}$
85/90/95	88.1/92.1/96.0	10	0.6	$A^{3,1,2,\sigma}$
85/90/95	86.3/90.5/94.7	15	0.6	$A^{3,1,2,\sigma}$
85/90/95	85.2/87.6/93.9	30	0.6	$A^{3,1,2,\sigma}$
85/90/95	88.1/91.6/96.0	10	0.6	$A^{3,3,1,\sigma}$
85/90/95	86.0/90.5/95.1	15	0.6	$A^{3,3,1,\sigma}$
85/90/95	85.0/89.5/94.0	30	0.6	$A^{3,3,1,\sigma}$
85/90/95	87.4/92.5/96.2	10	0.6	$A^{3,3,2,\sigma}$
85/90/95	86.2/89.7/95.2	15	0.6	$A^{3,3,2,\sigma}$
85/90/95	85.8/89.2/93.2	30	0.6	$A^{3,3,2,\sigma}$

Table C.5: Simulated covering rates of simultaneous β%-confidence sets for rGP models obtained from $M = 5000$ simulations. The error processes do not fulfill the assumptions necessary for application of the GKF.

E vs F: SSA/STW

Vol	L	R
1	49.0 ± 0.4	80.4 ± 0.3
2	24.8 ± 0.3	16.0 ± 0.3
3	87.9 ± 0.2	71.0 ± 0.4
4	88.1 ± 0.2	22.0 ± 0.3
5	84.7 ± 0.2	87.8 ± 0.3
6	72.5 ± 0.4	71.5 ± 0.3
7	60.2 ± 0.2	74.0 ± 0.3
8	94.2 ± 0.1	28.6 ± 0.2

Table C.6: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Improper Marker Placements: SSA/STW

Vol	A vs B		A vs C		A vs D	
	L	R	L	R	L	R
1	4.9 ± 0.1	0.1 ± 0.0	1.1 ± 0.1)	13.5 ± 0.3	0.1 ± 0.0	2.6 ± 0.1
2	0.1 ± 0.0	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.4 ± 0.0	0.0 ± 0.0
3	0.9 ± 0.1	0.0 ± 0.0	0.9 ± 0.1	0.0 ± 0.0	0.4 ± 0.1	0.0 ± 0.0
4	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
5	14.8 ± 0.3	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
6	51.1 ± 0.3	49.7 ± 0.4	0.4 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
7	47.6 ± 0.4	23.6 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
8	1.6 ± 0.1	14.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Vol	B vs C		B vs D	
	L	R	L	R
1	0.2 ± 0.1	2.7 ± 0.1	1.3 ± 0.1	5.1 ± 0.2
2	0.0 ± 0.0	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0
3	0.7 ± 0.0	9.5 ± 0.2	0.0 ± 0.0	0.2 ± 0.0
4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
6	1.2 ± 0.1	0.3 ± 0.0	0.5 ± 0.0	0.1 ± 0.0
7	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
8	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Table C.7: *walk, p-values in* [%] *of MILLPerm with standard deviation.*

Vol	A		B		C	
	L	R	L	R	L	R
1	99.2 ± 1.4	99.3 ± 1.7	98.8 ± 1.2	99.0 ± 1.3	98.6 ± 1.2	98.2 ± 0.8
2	116.8 ± 2.2	115.3 ± 1.9	119.1 ± 3.0	118.1 ± 2.5	118.8 ± 2.5	118.2 ± 2.2
3	108.1 ± 2.2	109.9 ± 2.8	104.8 ± 1.3	105.3 ± 2.1	103.0 ± 1.9	103.5 ± 1.9
4	90.9 ± 2.7	90.1 ± 2.4	88.5 ± 1.2	88.6 ± 1.4	90.4 ± 1.7	90.1 ± 1.4
5	103.1 ± 1.7	102.9 ± 2.3	100.0 ± 2.2	99.1 ± 2.3	101.1 ± 1.7	100.6 ± 1.8
6	104.0 ± 1.6	105.3 ± 1.7	102.9 ± 1.4	102.7 ± 0.5	101.3 ± 1.7	102.2 ± 1.5
7	103.8 ± 2.0	104.3 ± 3.1	104.5 ± 1.8	104.7 ± 1.7	103.4 ± 2.1	103.0 ± 1.9
8	101.4 ± 1.9	101.3 ± 2.3	99.7 ± 1.1	100.4 ± 1.3	97.4 ± 1.6	97.5 ± 1.4

Vol	D		E		F	
	L	R	L	R	L	R
1	96.3 ± 1.9	96.9 ± 2.0	101.3 ± 1.3	101.6 ± 2.2	100.8 ± 3.2	100.8 ± 2.4
2	115.3 ± 2.7	115.0 ± 2.2	118.6 ± 1.5	117.9 ± 1.7	119.3 ± 2.1	118.0 ± 2.0
3	97.6 ± 2.4	97.5 ± 3.7	101.8 ± 2.9	103.2 ± 3.1	99.4 ± 3.0	100.4 ± 2.6
4	88.8 ± 2.4	89.8 ± 2.0	88.5 ± 3.7	89.2 ± 3.1	91.6 ± 3.2	92.5 ± 1.9
5	98.4 ± 1.5	97.2 ± 2.3	100.3 ± 1.9	99.5 ± 1.5	100.9 ± 2.7	100.3 ± 3.3
6	100.6 ± 1.1	100.9 ± 1.3	101.1 ± 1.6	101.8 ± 2.2	101.5 ± 2.1	102.1 ± 2.1
7	103.2 ± 1.7	102.8 ± 1.8	100.3 ± 3.0	100.6 ± 2.1	105.2 ± 9.3	106.1 ± 7.7
8	98.3 ± 1.1	97.8 ± 1.9	98.6 ± 1.4	98.2 ± 2.1	96.3 ± 1.7	95.6 ± 2.1

Table C.8: *walk, mean number of frames per gait cycle with standard deviation.*

Vol	A		B		C	
	L	R	L	R	L	R
1	92.4 ± 1.4	91.5 ± 1.4	90.3 ± 1.5	90.5 ± 1.8	89.1 ± 1.8	88.5 ± 1.7
2	104.8 ± 2.0	103.3 ± 2.0	106.8 ± 2.8	106.2 ± 2.2	108.5 ± 1.0	108.4 ± 1.1
3	87.3 ± 2.1	85.8 ± 2.7	84.2 ± 1.7	84.6 ± 1.4	81.3 ± 2.5	81.5 ± 2.2
4	79.1 ± 1.5	78.4 ± 2.5	79.3 ± 1.3	78.9 ± 1.6	81.1 ± 2.7	80.8 ± 3.0
5	90.8 ± 1.3	90.8 ± 1.5	89.7 ± 1.1	90.0 ± 1.8	91.5 ± 1.4	90.6 ± 2.4
6	97.6 ± 1.4	97.4 ± 1.4	96.8 ± 1.4	97.5 ± 1.0	94.8 ± 1.8	95.8 ± 1.5
7	91.9 ± 1.9	92.2 ± 2.1	93.8 ± 2.1	94.4 ± 1.8	92.5 ± 1.3	92.8 ± 1.5
8	90.6 ± 2.8	90.7 ± 2.8	89.3 ± 1.8	89.8 ± 1.8	88.2 ± 2.4	88.3 ± 1.9

Vol	D		E		F	
	L	R	L	R	L	R
1	89.1 ± 1.7	88.9 ± 1.8	90.1 ± 1.2	90.0 ± 1.6	88.5 ± 1.9	86.9 ± 2.0
2	107.2 ± 2.2	107.1 ± 2.0	109.3 ± 2.4	108.2 ± 1.8	108.3 ± 1.4	107.9 ± 1.0
3	81.6 ± 1.8	82.5 ± 2.4	81.9 ± 3.6	82.1 ± 3.1	80.4 ± 3.1	80.0 ± 3.0
4	79.5 ± 3.1	79.5 ± 2.6	83.5 ± 3.0	83.5 ± 3.3	79.8 ± 3.0	79.0 ± 2.7
5	90.8 ± 1.5	90.9 ± 1.4	90.9 ± 1.4	90.6 ± 1.9	90.8 ± 1.6	90.6 ± 1.9
6	95.0 ± 1.7	95.6 ± 1.2	96.8 ± 1.9	96.7 ± 1.9	94.6 ± 1.1	95.0 ± 1.5
6	92.8 ± 1.6	93.6 ± 1.9	92.6 ± 2.1	93.2 ± 2.1	90.9 ± 1.3	91.9 ± 1.8
8	87.6 ± 1.4	87.0 ± 1.8	87.1 ± 2.1	86.8 ± 2.2	85.7 ± 2.0	85.3 ± 1.7

Table C.9: *fast walk, mean number of frames per gait cycle with standard deviation.*

Left Knee						
Vol	C vs D	E vs C	E vs D	F vs C	F vs C	E vs F
1	X	X	X	HC	HC,MS	X
2	X	HC	HC,MS,TO	HC,TO	HC,MS,TO,MF	X
3	X	X	X	HC	X	X
4	X	X	X	MF	MF	X
5	X	X	X	X	X	X
6	X	TO	MS	MS,TO	MS	X
7	X	X	HC	X	X	X
8	X	X	X	X	X	X

Right Knee						
Vol	C vs D	E vs C	E vs D	F vs C	F vs C	E vs F
1	X	MS,TO	X	HC,MS,TO	X	X
2	X	HC	MF	HC,TO	TO,MF	X
3	X	X	HC,MS	X	HC,MS	X
4	HC	HC,TO	TO	MF	MF	X
5	X	X	X	X	X	X
6	X	MS	X	TO,MF	MS	X
7	X	X	X	X	X	X
8	X	X	MF	X	MF	X

Table C.10: *walk, OCST with $\beta = 0.9$, locations where confidence bands do not overlap, everywhere overlapping (X).*

Left Knee						
Vol	**C vs D**	**E vs C**	**E vs D**	**F vs C**	**F vs C**	**E vs F**
1	X	X	X	X	X	X
2	MS,TO	MF,HC	MS,TO	X	MS,TO,MF	X
3	X	X	MS	MS	MS	X
4	X	X	MS	MF	X	X
5	MS	X	X	X	X	X
6	X	HC,TO	HC	HC,TO	HC	X
7	X	MS	X	X	X	X
8	X	X	X	X	X	X

Right Knee						
Vol	**C vs D**	**E vs C**	**E vs D**	**F vs C**	**F vs C**	**E vs F**
1	MS	HC,TO	X	TO	MF	X
2	X	X	X	X	MS	X
3	X	MS	X	MS	X	X
4	X	X	X	X	X	X
5	MS	X	X	X	X	X
6	X	TO	HC	HC	X	X
7	X	X	MS	HC	X	X
8	X	X	MS	X	X	X

Table C.11: *fast walk, OCST with* $\beta = 0.9$*, locations where confidence bands do not overlap, everywhere overlapping (X).*

Appendix D

Figures

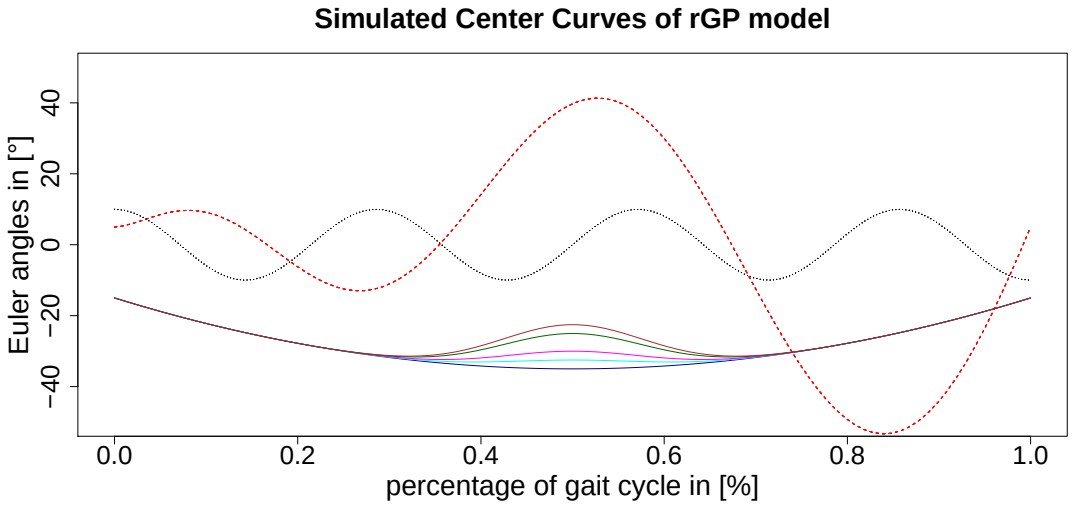

Figure D.1: *center curves* γ_0^λ, $\lambda \in \{0, 0.5, 1, 2, 2.5\}$, *of the rGP models* γ_A^λ *and* γ_B^λ.

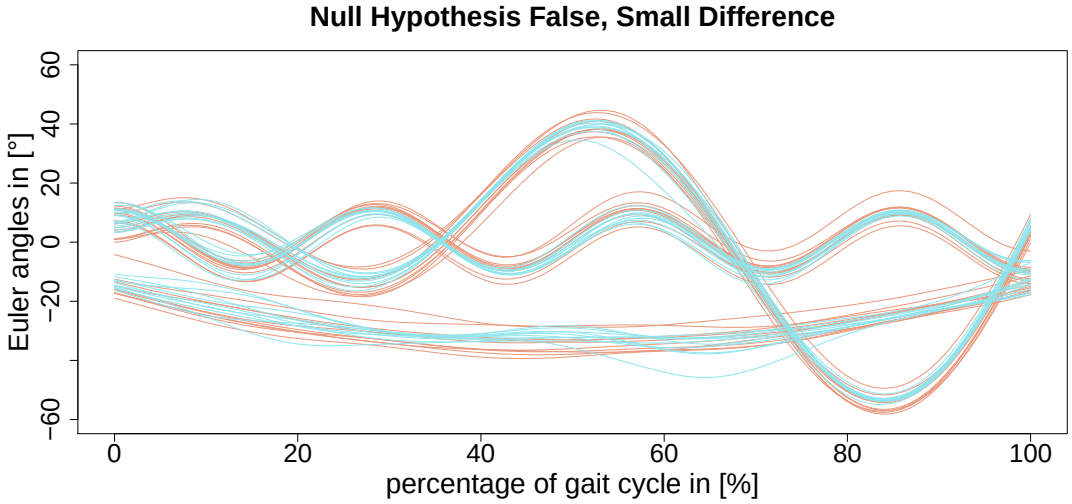

Figure D.2: *ten samples of rGP model* γ_A^0 *(salmon) and* γ_B^1 *(cyan).*

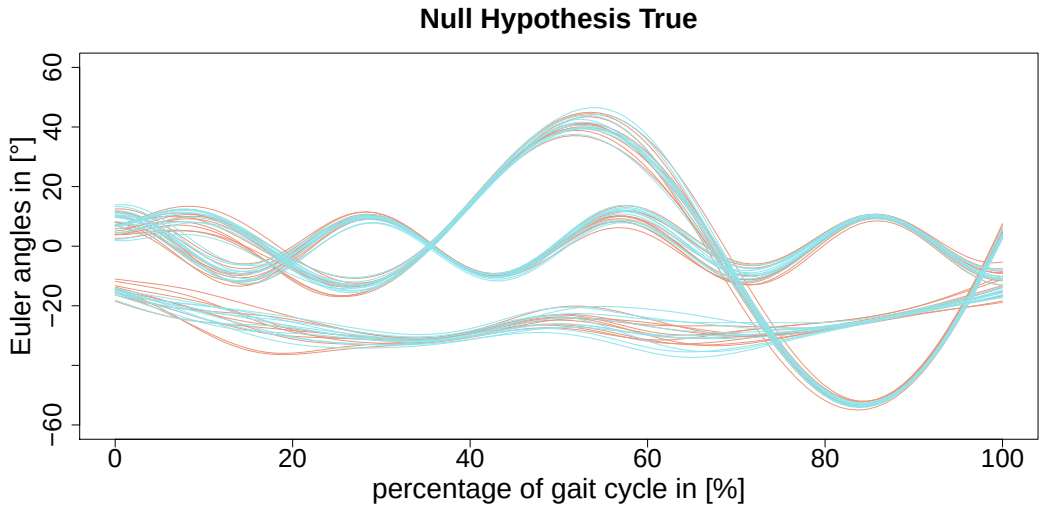

Figure D.3: *ten samples of rGP model γ_B^2 (salmon) and γ_B^2 (cyan).*

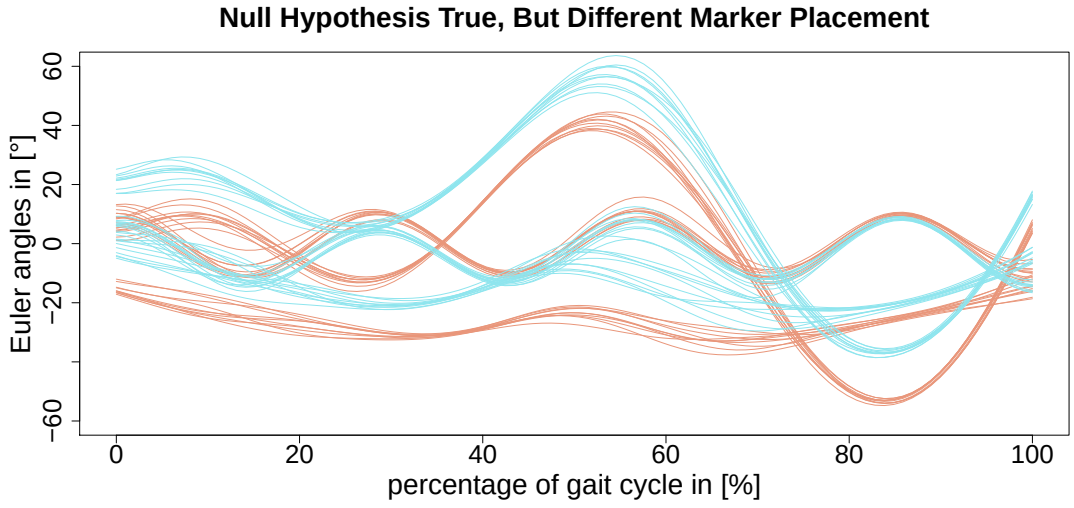

Figure D.4: *ten samples of rGP model $\gamma_B^{2.5}$ (salmon) and $P\gamma_B^{2.5}Q$ (cyan).*

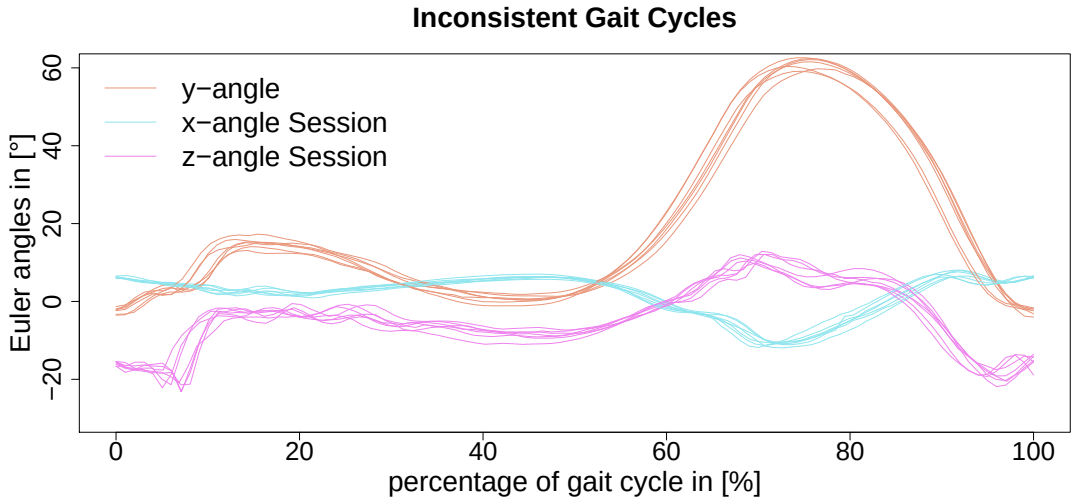

Figure D.5: *walk, example of inconsistencies in gait cycle extraction using force plate data. At the boundaries the shapes of the angles are not always the same.*

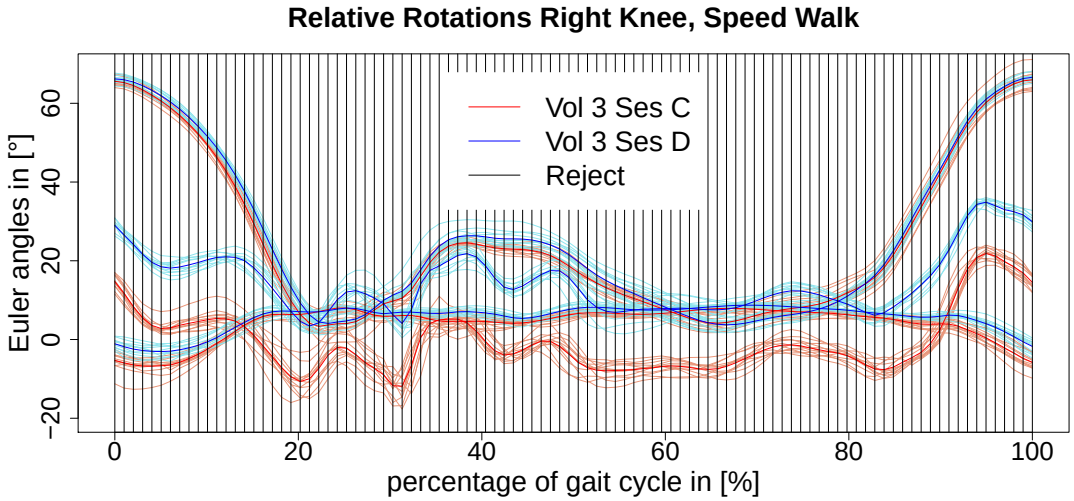

Figure D.6: *raw data of two sessions. Vertical black lines indicate that the* 90%*-confidence sets for the PEMs do not overlap at this time points.*

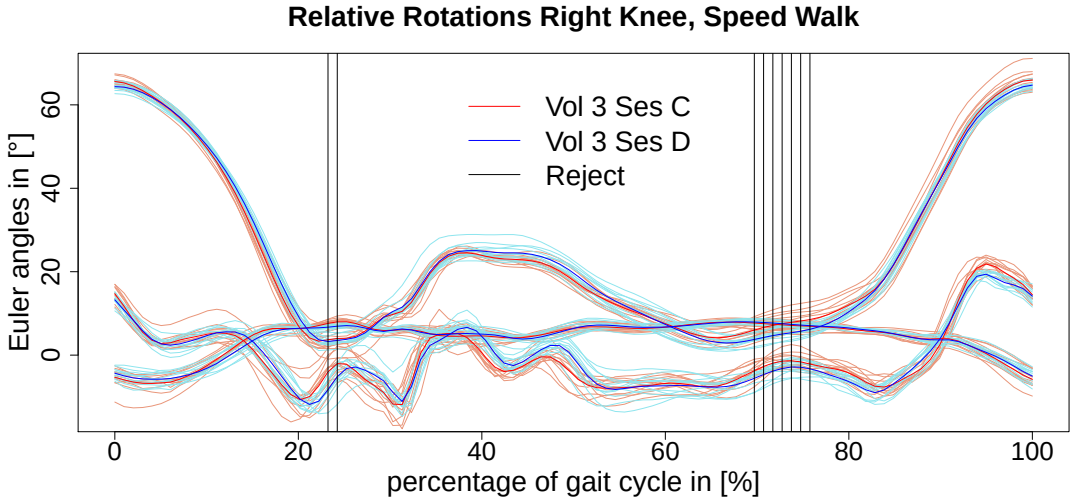

Figure D.7: *data of two sessions corrected by SSA. Vertical black lines indicate that the* 90%*-confidence sets for the PEMs do not overlap at these time points.*

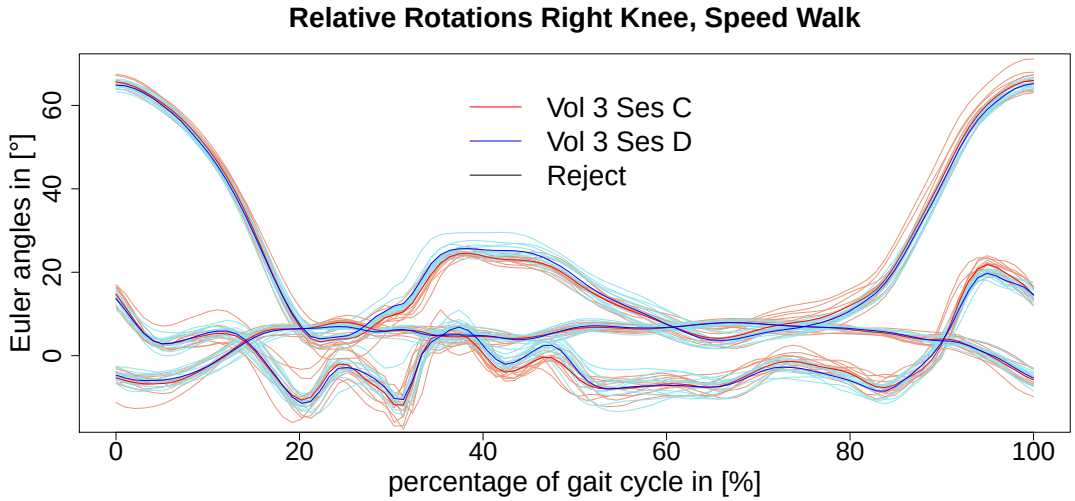

Figure D.8: *data of two sessions corrected by SSA and STW. No rejections anymore around* 20% *and* 80% *(compare D.7). Also MILLPerm does not reject the null hypothesis of equal center curves.*

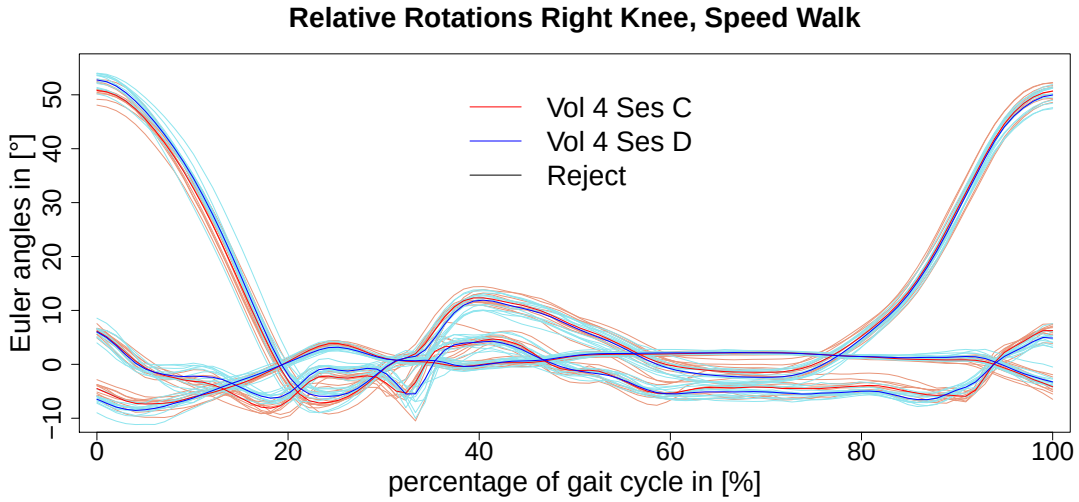

Figure D.9: *data of two sessions corrected by SSA. Also MILLPerm does not reject the null hypothesis of equal PEMs.*

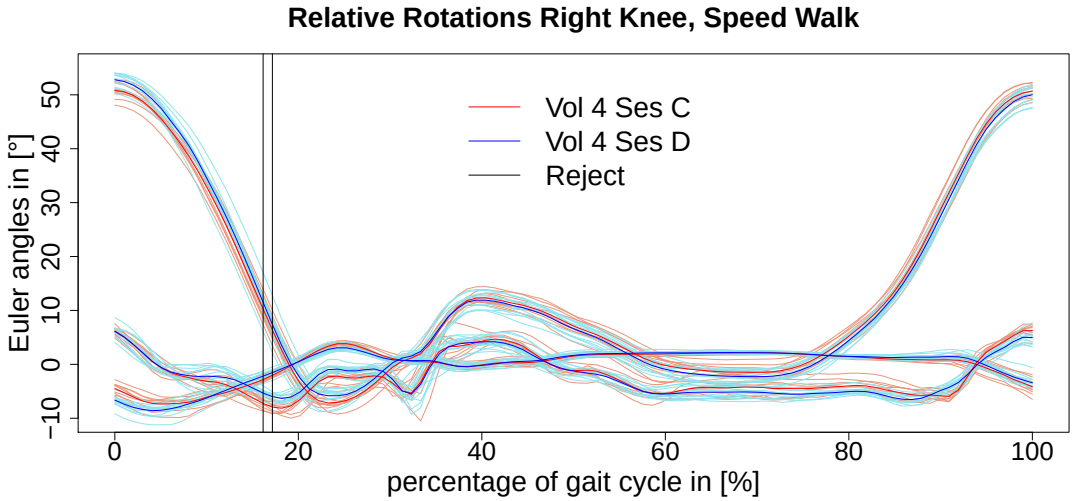

Figure D.10: *data of two sessions corrected by SSA and STW. Rejections near* 20%*, but also better temporally registered around* $20\% - 40\%$ *than in D.9.*

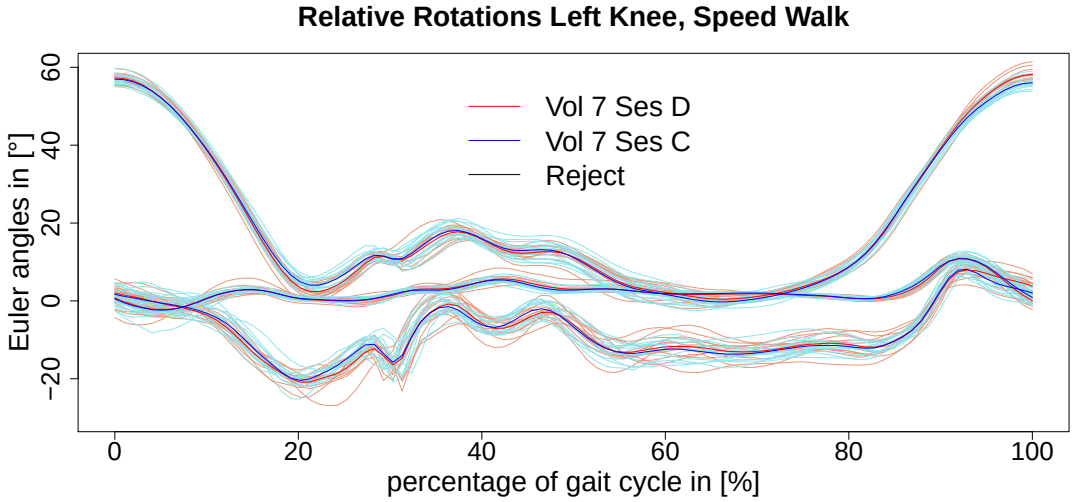

Figure D.11: *data corrected of two sessions by SSA and STW. No rejections and similar variance in both sessions.*

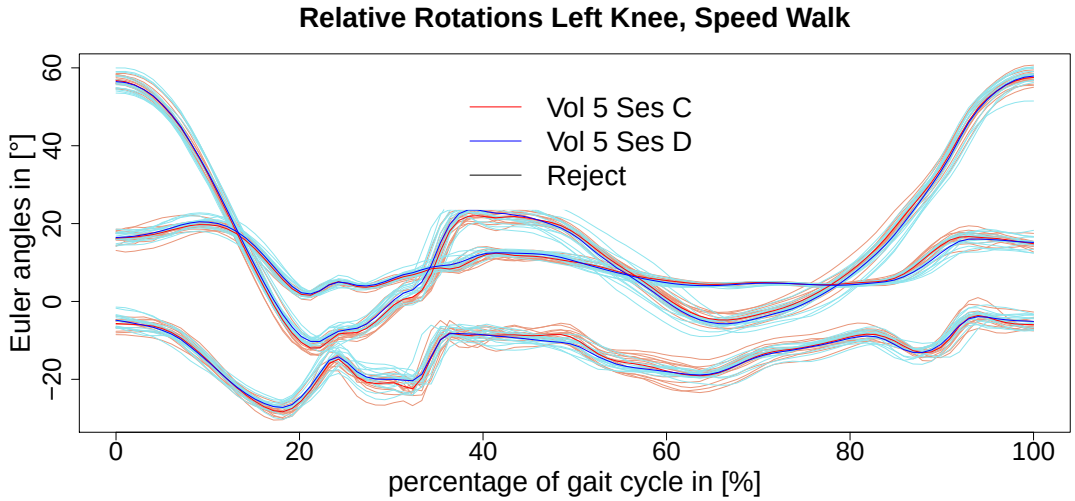

Figure D.12: *data of two sessions corrected by SSA and STW. MILLPerm does not reject the null hypothesis of equal PEMs.*

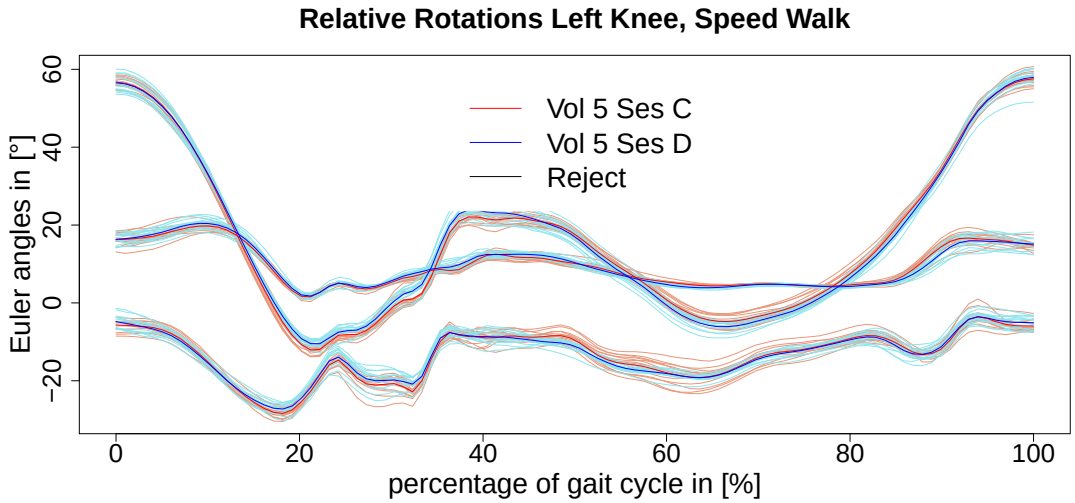

Figure D.13: *data of two sessions corrected by SSA, STW and ITW. Although within a session the curves are better temporally registered than in Figure D.12, we have that the MILLPerm rejects the null hypothesis of equal PEMs.*

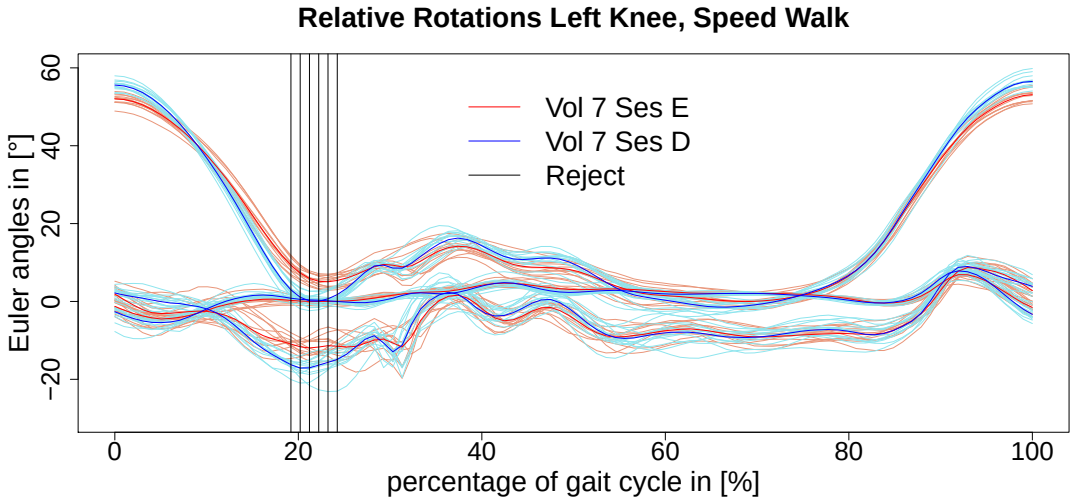

Figure D.14: *data of two sessions corrected by SSA and STW. Rejections near* 20%*, variance is still similar.*

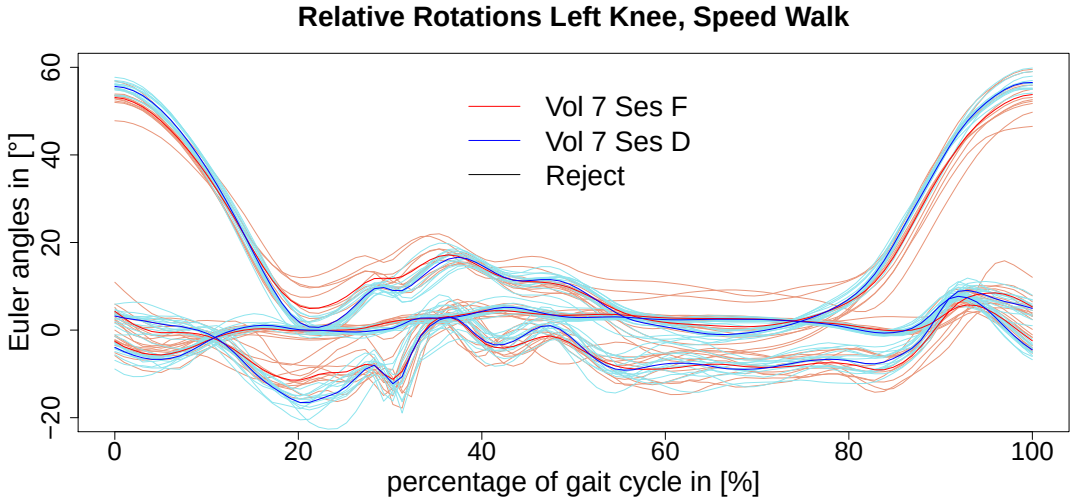

Figure D.15: *data of two sessions corrected by SSA and STW. Rejections near* 20%*, compare D.14, dissappeared due to high variance (e.g., y-angle).*

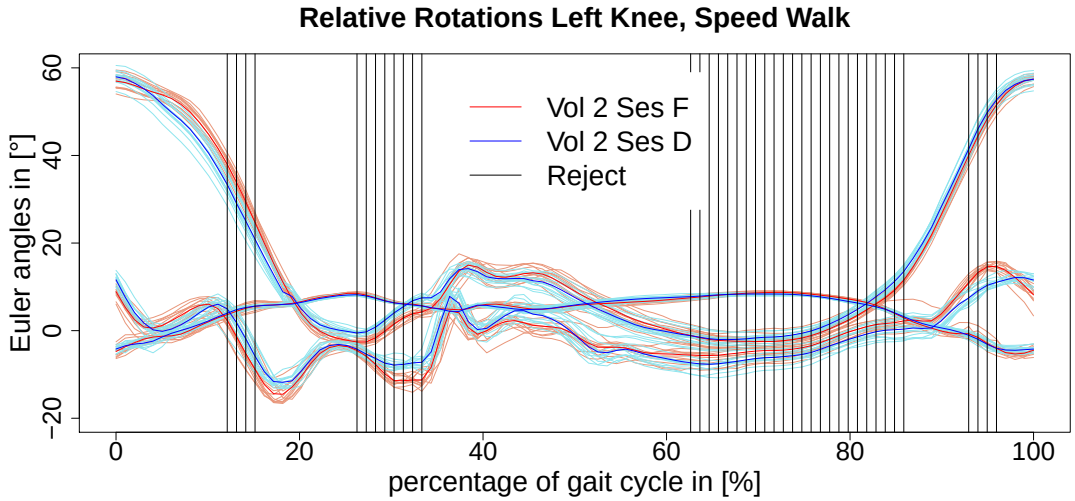

Figure D.16: *data of two sessions corrected by SSA and STW. OCST and MILLPerm agree that this volunteer has a kneeling effect.*

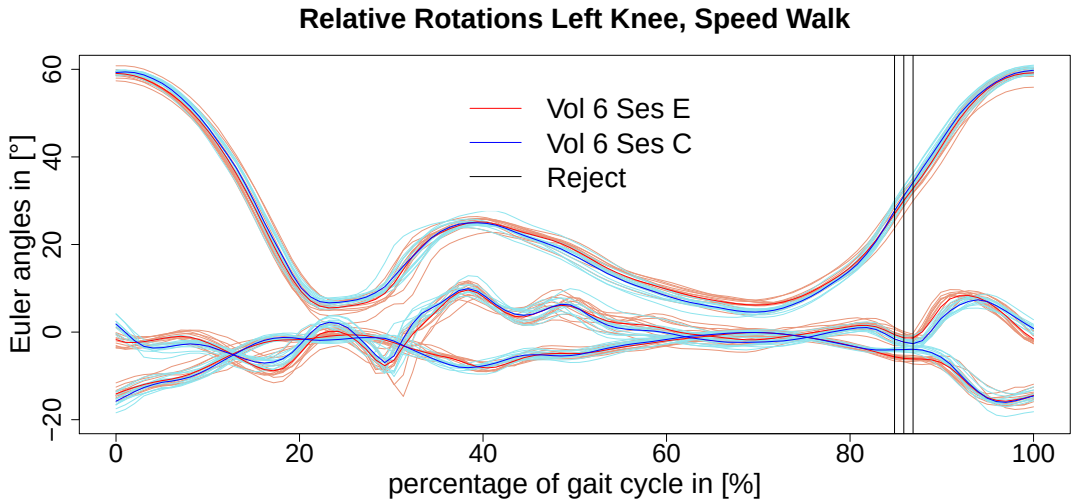

Figure D.17: *data corrected of two sessions by SSA and STW. OCST and MILLPerm agree that this volunteer has a kneeling effect.*

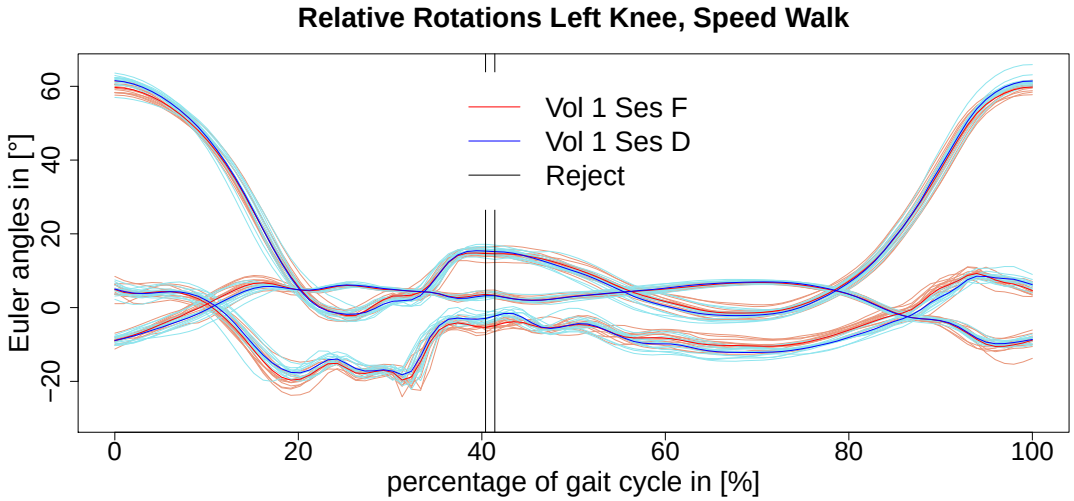

Figure D.18: *data of two sessions corrected by SSA and STW. OCST rejects the equality of means, whereas MILLPerm accepts it.*

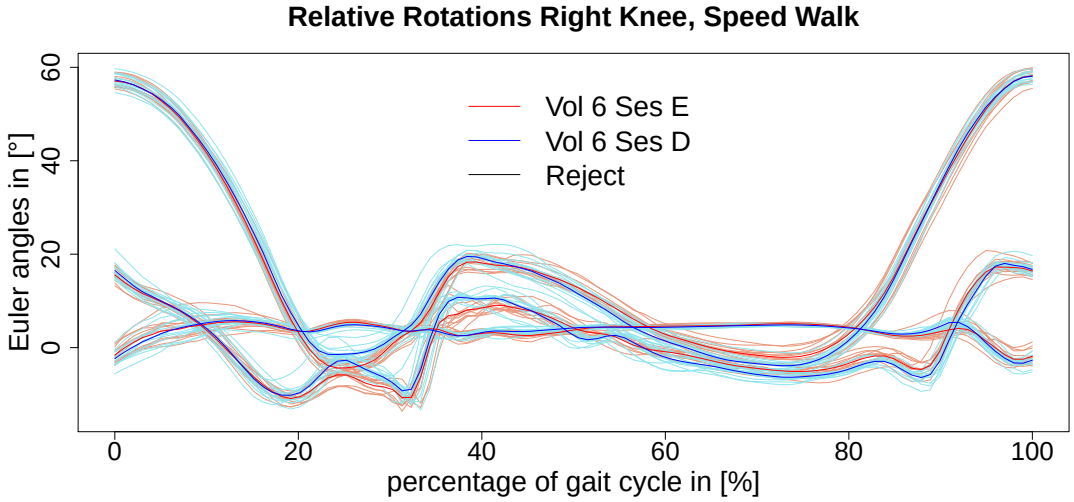

Figure D.19: *data of two sessions corrected by SSA and STW. OCST does accepts the equality of means, whereas MILLPerm rejects it.*

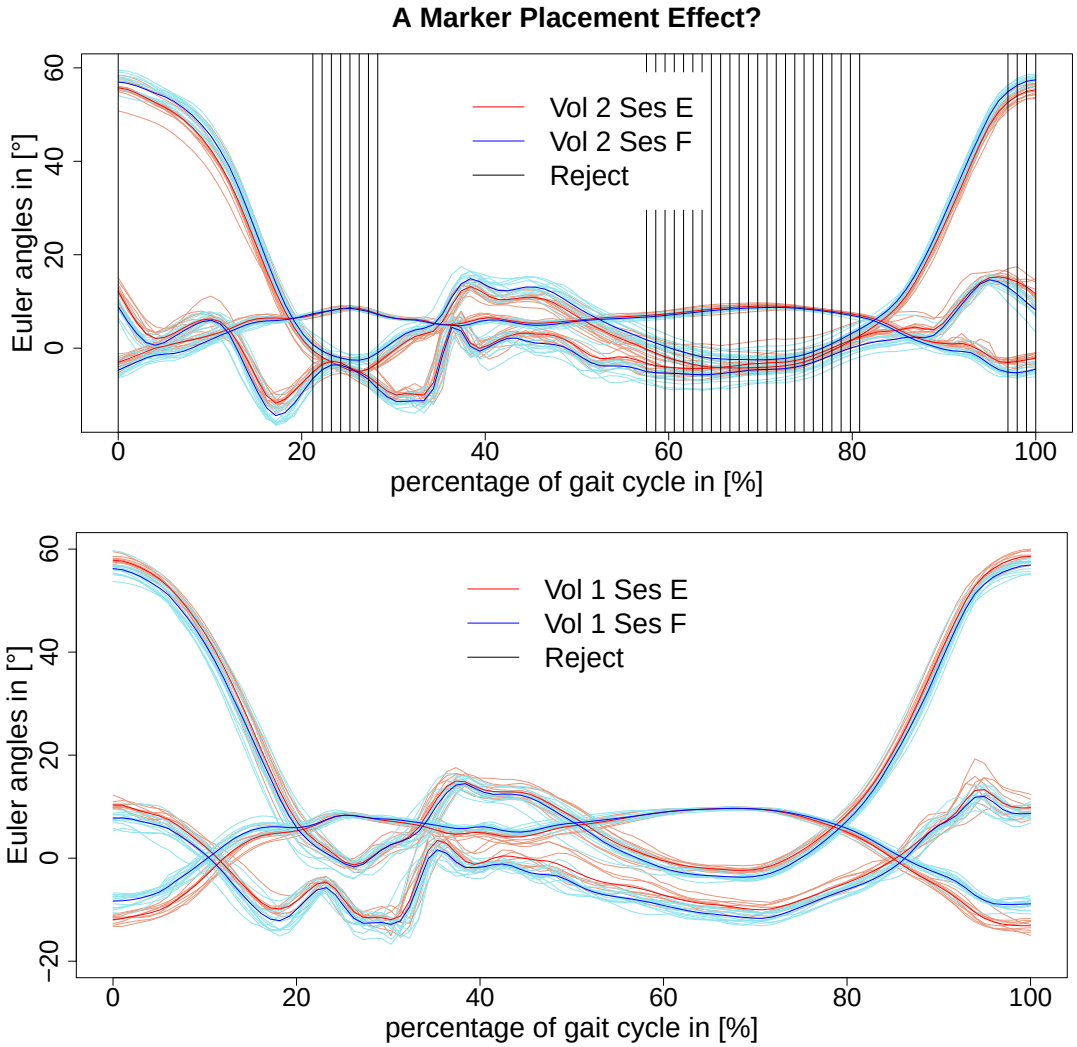

Figure D.20: *left knee walk (above) and right knee walk (below), data corrected by STW. It is likely that a MP effect is visible here, although no marker replacement was performed. ILLPerm accepts the null hypothesis of equal means, suggesting that the difference is due to the action of an element of* $\mathcal{I}_0\big(SO(3)\big)$*, since the ILL is invariant with respect to a subgroup of* $\mathcal{I}_0\big(SO(3)\big)$*.*

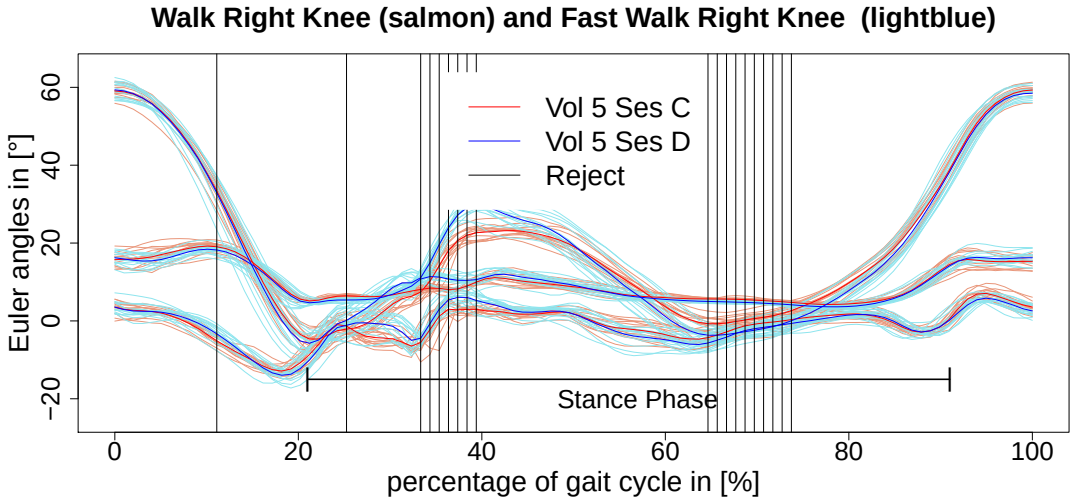

Figure D.21: *data of two sessions corrected by SSA and STW. Rejections mainly in stance phase.*

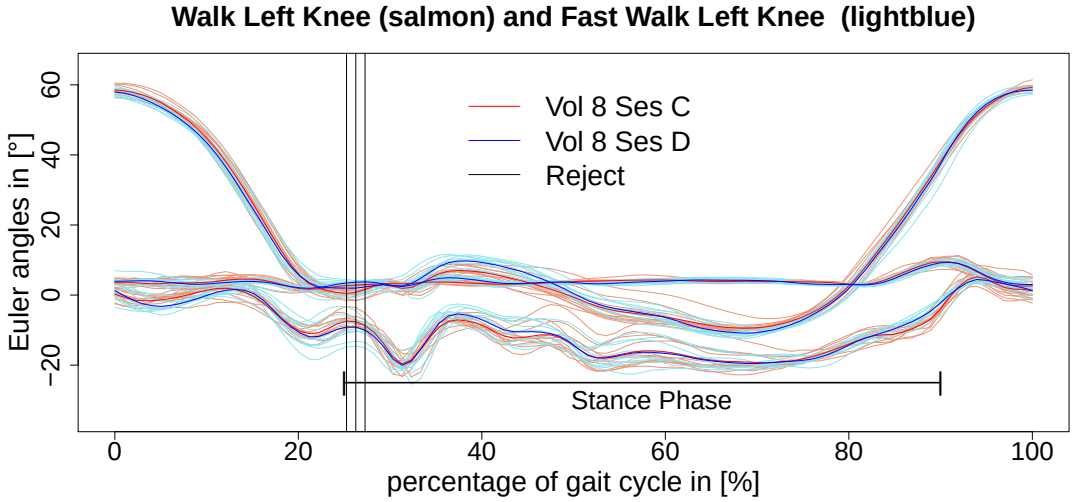

Figure D.22: *data of two sessions corrected by SSA and STW. Rejections mainly in stance phase.*

Nomenclature

$(s,t) \mapsto \Sigma_{s,t}$ ..... covariance function of a process, page 24
$\{H_t^A\}_{t\in I}$ ..... Hotelling T^2 process of a Gaussian process A, page 40
χ ..... a session, page 23
δ_I, $\delta_{I,1}$, $\delta_{I,2}$ ..... intrinsic length losses, page 55
$\mathrm{Diff}^+[0,1]$ ..... diffeomorphisms of $[0,1]$, page 18
Exp ..... Lie exponential of $SO(3)$, page 106
γ, η ..... (random) curves in $SO(3)$
γ^g, η^g ..... piecewise geodesic curves in $SO(3)$
γ_0, η_0 ..... center curves of right Gaussian models, page 24
$\hat{\Sigma}_t^A$ ..... sample covariance at time t of a sample of a Gaussian process A, page 40
$\hat{S}_t^X$ ..... sample covariance matrix of intrinsic residuals, page 42
$\hat{\gamma}_N(t)$ ..... element from the PESM at t of a sample $\gamma_1, ..., \gamma_N$
$\mathcal{I}_0\big(SO(3)\big)$ ..... identity component of isometry group of $SO(3)$, page 17
ι ..... an isomorphism between $\mathfrak{so}(3)$ and $\mathbb{R}^3$, page 106
Log ..... Lie logarithm of $SO(3)$, page 107
$\mathcal{C}, \mathcal{C}^\nu$ ..... continuous, ν-times continuously differentiable
$\mathcal{C}^1\big([0,1], SO(3)\big)$. differentiable $SO(3)$-valued curves, page 14
$\mathcal{C}_\beta$ ..... confidence set, page 40
$\mathcal{C}_{I_{3\times 3}}$ ..... cut locus of $I_{3\times 3}$, page 107
$\mathcal{L}_0, \mathcal{L}_1$ ..... Lipschitz killing curvature, page 47
$\mathfrak{L}$ ..... a right inverse of the Lie exponential, page 42
$\mathfrak{x}(X)$ ..... Euler characteristic of X, page 46
$\mu(\gamma,t)$, $\hat{\mu}_N\big(\chi,t\big)$ .. pointwise extrinsic population and sample mean at t, page 30
ϕ ..... an element of $\mathrm{Diff}^+[0,1]$
Π ..... a map $SO(4) \to SO(3) \times SO(3)$, page 64
π ..... a covering map $S^3 \to SO(3)$, page 62
$\pi_{SO(4)}$ ..... covering map $S^3 \times S^3 \to SO(4)$, page 63
ψ ..... an element of $\mathcal{I}_0\big(SO(3)\big)$
$\mathbb{S}(\mathbb{X})$ ..... space of samples in $\mathcal{X}$, page 22
$\mathcal{S}$ ..... group of gait similarities, page 19
$\mathfrak{so}(3)$ ..... Lie algebra of $SO(3)$, page 106
$\tilde{h}_{\gamma,N,\beta}$ ..... quantile of Hotelling T^2 process of γ, page 42
$\tilde{\gamma}, \tilde{\eta}$ ..... continuous lifts of curves in $SO(3)$ to S^3
$\|A\|_F$ ..... the Frobenius norm of a matrix A, page 106
$\mathcal{X}$ ..... data space (i.e., $\mathcal{C}^1\big([0,1], SO(3)\big)$), page 14
A, $\{A_t\}_{t\in I}$ ..... a Gaussian process, page 24

$d_{SO(3)}(\cdot,\cdot)$ ..... metric induced by bi-invariant Riemannian metric on $SO(3)$, page 107
$h_{A,N,\beta}$ ..... quantile of Hotelling T^2 process of A, page 41
I ..... the unit interval $[0,1]$
$I_{3\times3}$ ..... 3×3 unit matrix
$r_{SO(4)}$ ..... a right inverse of Π, page 64
S^D ..... D-dimensional sphere
$SO(3)$ ..... group of 3×3 rotation matrices, page 3
X_t^n ..... intrinsic residuals, page 42
ILLPerm ..... intrinsic length loss permutation test, page 74
ILLPerm ..... modified intrinsic length loss permutation test, page 74
ITW ..... individual time warping, page 18
OCST ..... overlapping confidence sets test, page 72
PEM, PESM ..... pointwise extrinsic population and sample mean, page 30
rGP ..... right Gaussian pertubation model, page 24
STW ..... sessionwise time warping, page 18

Bibliography

R. J. Adler. On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, 10(1):1–74, 2000.

R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

B. Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.

B. B. Amor, J. Su, and A. Srivastava. Action recognition using rate-invariant analysis of skeletal shape trajectories. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 38(1):1–13, 2016.

D. W. Andrews. Generic uniform convergence. *Econometric theory*, 8(02):241–257, 1992.

P. Baker, D. Coggon, I. Reading, D. Barrett, M. McLaren, and C. Cooper. Sports injury, occupational physical activity, joint laxity, and meniscal damage. *The Journal of Rheumatology*, 29(3):557–563, 2002.

F. G. Ball, I. L. Dryden, and M. Golalizadeh. Brownian motion and Ornstein–Uhlenbeck processes in planar shape space. *Methodology and Computing in Applied Probability*, 10(1):1–22, 2008.

K. A. Ball and T. M. Greiner. A procedure to refine joint kinematic assessments: functional alignment. *Computer methods in Biomechanics and biomedical engineering*, 15(5):487–500, 2012.

A. Baudet, C. Morisset, P. d'Athis, J.-F. Maillefert, J.-M. Casillas, P. Ornetti, and D. Laroche. *PloS one*, 9(7):e102098, 2014.

M. Bauer, M. Eslitzbichler, and M. Grasmair. Landmark-guided elastic shape analysis of human character motions. *arXiv preprint arXiv:1502.07666*, 2015.

M. Bauer, M. Bruveris, and P. W. Michor. Why use Sobolev metrics on the space of curves. In *Riemannian Computing in Computer Vision*, pages 233–255. Springer, 2016.

T. F. Besier, D. L. Sturnieks, J. A. Alderson, and D. G. Lloyd. Repeatability of gait data using a functional hip joint centre and a mean helical knee axis. *Journal of Biomechanics*, 36(8):1159–1168, 2003.

A. Bhattacharya and R. Bhattacharya. *Nonparametric inference on manifolds: with applications to shape spaces*. Cambridge University Press, 2012.

R. N. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics*, 31(1):1–29, 2003.

R. N. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics*, 33(3):1225–1259, 2005.

P. J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095, 1973.

J. Bigot, B. Charlier, et al. On the consistency of Fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, 5:1054–1089, 2011.

F. L. Bookstein. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, 1997.

J. Bouman. Over quaternionen en hunne toepassing in de meetkunde der vierdimensionale ruimte. *Nieuw Archief voor Wiskunde*, 2(3):240–266, 1932.

D. R. Brillinger and B. S. Stewart. Elephant-seal movements: modelling migration. *Canadian Journal of Statistics*, 26(3):431–443, 1998.

M. Bruveris. Optimal reparametrizations in the square root velocity framework. *arXiv preprint arXiv:1507.02728*, 2015.

G. Cao, L. Yang, and D. Todem. Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2): 359–377, 2012.

J. Cao, K. J. Worsley, et al. The detection of local shape changes via the geometry of hotelling's T^2 fields. *The Annals of Statistics*, 27(3):925–942, 1999.

T. Carne. The geometry of shape spaces. *Proceedings of the London Mathematical Society*, 61:407–432, 1990.

E. Celledoni and M. Eslitzbichler. Shape analysis on lie groups with applications in computer animation. *arXiv preprint arXiv:1506.00783*, 2015.

T. Chang. Spherical regression. *The Annals of Statistics*, 14(3):907–924, 1986.

T. Chang. Spherical regression with errors in variables. *The Annals of Statistics*, 17(1):293–306, 1989.

T. Chang and L.-P. Rivest. M-estimation for location and regression parameters in group models: A case study using Stiefel manifolds. *The Annals of Statistics*, 29(3):784–814, 2001.

V. Chernozhukov, D. Chetverikov, K. Kato, et al. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 2013.

L. Chiari, U. Della Croce, A. Leardini, and A. Cappozzo. Human movement analysis using stereophotogrammetry: part 2: Instrumental errors. *Gait & Posture*, 21(2):197–211, 2005.

G. S. Chirikjian and A. B. Kyatkin. *Engineering applications of noncommutative harmonic analysis: with emphasis on rotation and motion groups*. CRC press, 2000.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*, volume 6. MIT press Cambridge, 2001.

A. Cutti, I. Parel, M. Raggi, E. Petracci, A. Pellegrini, A. Accardo, R. Sacchetti, and G. Porcellini. Prediction bands and intervals for the scapulo-humeral coordination based on the bootstrap and two gaussian methods. *Journal of Biomechanics*, 47(5):1035–1044, 2014.

J. Davidson. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

R. B. Davis, S. Ounpuu, D. Tyburski, and J. R. Gage. A gait analysis data collection and reduction technique. *Human movement science*, 10(5):575–587, 1991.

D. A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765, 2011.

E. Del Barrio, L. Hélène, and L. Jean-Michel. A statistical analysis of a deformation model with wasserstein barycenters: estimation procedure and goodness of fit test. *arXiv preprint arXiv:1508.06465*, 2015.

U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo. Human movement analysis using stereophotogrammetry: part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait & Posture*, 21(2): 226–237, 2005.

M. P. do Carmo Valero. *Riemannian geometry*. Birkhäuser, 1992.

T. D. Downs. Orientation statistics. *Biometrika*, 59(3):665–676, 1972.

I. L. Dryden and K. V. Mardia. *Statistical shape analysis*, volume 4. J. Wiley Chichester, 1998.

E. Dudek and K. Holly. Nonlinear orthogonal projection. In *Annales Polonici Mathematici*, volume 59, pages 1–31. Institute of Mathematics Polish Academy of Sciences, 1994.

A. Duhamel, J. Bourriez, P. Devos, P. Krystkowiak, A. Destee, P. Derambure, and L. Defebvre. Statistical tools for clinical gait analysis. *Gait & Posture*, 20 (2):204–212, 2004.

J. J. Duistermaat and J. A. Kolk. *Lie groups*. Springer Science & Business Media, 1999.

E. Edgington and P. Onghena. *Randomization tests*. CRC Press, 2007.

R. M. Ehrig, W. R. Taylor, G. N. Duda, and M. O. Heller. A survey of formal methods for determining functional joint axes. *Journal of Biomechanics*, 40 (10):2150–2157, 2007.

P. T. Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105(2):171–185, 2013.

N. Gaudreault, N. Hagemeister, S. Poitras, and J. A. de Guise. Comparison of knee gait kinematics of workers exposed to knee straining posture to those of non-knee straining workers. *Gait & Posture*, 38(2):187–191, 2013.

P. Gilkey, J. Park, and R. Vázquez-Lorenzo. Aspects of differential geometry ii. *Synthesis Lectures on Mathematics and Statistics*, 7(2):1–157, 2015.

C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339, 1991.

U. Grenander. *General pattern theory: a mathematical study of regular structures*. Clarendon Press, 1993.

E. S. Grood and W. J. Suntay. A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *Journal of biomechanical engineering*, 105(2):136–144, 1983.

M. Haddou, L.-P. Rivest, and M. Pierrynowski. A nonlinear mixed effects directional model for the estimation of the rotation axes of the human ankle. *The Annals of Applied Statistics*, pages 1892–1912, 2010.

S. Helgason. *Differential Geometry and Symmetric Spaces*. Academic Press, New York, 1962.

H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60, 1981.

H. Hendriks and Z. Landsman. Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *Journal of Multivariate Analysis*, 67(2):227–243, 1998.

J. Hilgert and K.-H. Neeb. *Structure and geometry of Lie groups*. Springer Science & Business Media, 2011.

J. Hinkle, P. T. Fletcher, and S. Joshi. Intrinsic polynomials for regression on Riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 50(1-2): 32–52, 2014.

T. Hotz and S. Huckemann. Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics*, 67(1):177–193, 2015.

S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, 20:1–100, 2010.

S. F. Huckemann et al. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics*, 39(2):1098–1124, 2011.

S. M. Iacus. *Simulation and inference for stochastic differential equations: with R examples*. Springer Science & Business Media, 2008.

J.-i. Itoh and M. Tanaka. The dimension of a cut locus on a smooth Riemannian manifold. *Tohoku Mathematical Journal, Second Series*, 50(4):571–575, 1998.

S. Johansen and I. M. Johnstone. Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *The Annals of Statistics*, 18(2):652–684, 1990.

T. Kajaks and P. Costigan. The effect of sustained static kneeling on kinetic and kinematic knee joint gait parameters. *Applied ergonomics*, 46:224–230, 2015.

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer Science & Business Media, 1988.

H. Karcher. Riemannian center of mass and so called Karcher mean. *arXiv preprint arXiv:1407.2087*, 2014.

W. S. Kendall. Barycenters and hurricane trajectories. *Geometry Driven Statistics*, pages 146–159, 2015.

W. S. Kendall, H. Le, et al. Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, 25(3):323–352, 2011.

J. T. Kent and K. V. Mardia. Consistency of procrustes estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):281–290, 1997.

A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, 20(3):1266–1305, 1992.

A. Kneip and J. O. Ramsay. Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165, 2008.

J. Knoop, M. Steultjens, M. Van der Leeden, M. Van der Esch, C. Thorstensson, L. Roorda, W. Lems, and J. Dekker. Proprioception in knee osteoarthritis: a narrative review. *Osteoarthritis and Cartilage*, 19(4):381–388, 2011.

T. Krivobokova, T. Kneib, and G. Claeskens. Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, 105(490):852–863, 2012.

S. A. Kurtek, A. Srivastava, and W. Wu. Signal estimation under random time-warpings and nonlinear signal alignment. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 675–683. Curran Associates, Inc., 2011.

I. Kutzner, B. Heinlein, F. Graichen, A. Bender, A. Rohlmann, A. Halder, A. Beier, and G. Bergmann. Loading of the knee joint during activities of daily living measured in vivo in five subjects. *Journal of Biomechanics*, 43(11): 2164–2173, 2010.

S. Lahiri, D. Robinson, and E. Klassen. Precise matching of pl curves in $\mathbb{R}^n$ in the square root velocity framework. *arXiv preprint arXiv:1501.00577*, 2015.

H. Le. On the consistency of procrustean mean shapes. *Advances in Applied Probability*, 30(1):53–63, 1998.

A. Leardini, L. Chiari, U. Della Croce, and A. Cappozzo. Human movement analysis using stereophotogrammetry: part 3. soft tissue artifact assessment and compensation. *Gait & Posture*, 21(2):212–225, 2005.

J. M. Lee. *Introduction to Smooth manifolds*. Springer Verlag, New York, 2013.

E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Science & Business Media, 1998.

S. Lele. Euclidean distance matrix analysis (edma): estimation of mean form and mean form difference. *Mathematical Geology*, 25(5):573–602, 1993.

M. W. Lenhoff, T. J. Santner, J. C. Otis, M. G. Peterson, B. J. Williams, and S. I. Backus. Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. *Gait & Posture*, 9(1):10–17, 1999.

L. Lin, B. S. Thomas, H. Zhu, and D. B. Dunson. Extrinsic local regression on manifold-valued data. *arXiv preprint arXiv:1508.02201*, 2015.

M. Loève. *Probability Theory: Foundations, Random Sequences*. Van Nostrand, 1955.

J. Mackenzie. The estimation of an orientation relationship. *Acta Crystallographica*, 10(1):61–62, 1957.

E. Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285, 1993.

F. Marin, H. Mannel, L. Claes, and L. Dürselen. Correction of axis misalignment in the analysis of knee rotations. *Human movement science*, 22(3):285–296, 2003.

J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.

J. L. McGinley, R. Baker, R. Wolfe, and M. E. Morris. The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & Posture*, 29(3):360–369, 2009.

J. E. Mebius. A matrix-based proof of the quaternion representation theorem for four-dimensional rotations. *arXiv preprint math/0501249*, 2005.

K. R. Mecke and D. Stoyan. *Statistical physics and spatial statistics: the art of analyzing and modeling spatial structures and pattern formation*, volume 554. Springer Science & Business Media, 2000.

D. Mumford and A. Desolneux. *Pattern theory: the stochastic analysis of real-world signals*. CRC Press, 2010.

S. B. Myers and N. Steenrod. The group of isometries of a Riemannian manifold. *Annals of Mathematics*, 40(2):400–416, 1939.

W. K. Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 59(4):1161–1167, 1991.

B. Noehren, K. Manal, and I. Davis. Improving between-day kinematic reliability using a marker placement device. *Journal of Orthopaedic Research*, 28(11): 1405–1410, 2010.

R. A. Olshen, E. N. Biden, M. P. Wyatt, and D. H. Sutherland. Gait analysis and the bootstrap. *The Annals of Statistics*, 17(4):1419–1440, 1989.

S. T. Osis, B. A. Hettinga, S. L. Macdonald, and R. Ferber. A novel method to evaluate error in anatomical marker placement using a modified generalized procrustes analysis. *Computer methods in Biomechanics and biomedical engineering*, 18(10):1108–1116, 2015.

T. C. Pataky. Generalized n-dimensional biomechanical field analysis using statistical parametric mapping. *Journal of Biomechanics*, 43(10):1976–1982, 2010.

T. C. Pataky, M. A. Robinson, and J. Vanrenterghem. Vector field statistical analysis of kinematic and force trajectories. *Journal of Biomechanics*, 46(14): 2394–2401, 2013.

T. C. Pataky, J. Vanrenterghem, and M. A. Robinson. Zero-vs. one-dimensional, parametric vs. non-parametric, and confidence interval vs. hypothesis testing procedures in one-dimensional biomechanical trajectory analysis. *Journal of Biomechanics*, 48(7):1277–1285, 2015.

M. E. Payton, M. H. Greenstone, and N. Schenker. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *Journal of Insect Science*, 3(1):34, 2003.

M. R. Pierrynowski, P. A. Costigan, M. R. Maly, and P. T. Kim. Patients with osteoarthritic knees have shorter orientation and tangent indicatrices during gait. *Clinical Biomechanics*, 25(3):237–241, 2010.

M. J. Prentice. Spherical regression on matched pairs of orientation statistics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, pages 241–248, 1989.

J. F. Price. *Lie groups and compact groups*, volume 25. Cambridge University Press, 1977.

J. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):351–363, 1998.

J. Ramsay, R. Bock, and T. Gasser. Comparison of height acceleration curves in the fels, zurich, and berkeley growth data. *Annals of Human Biology*, 22(5): 413–426, 1995.

J. O. Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

D. Rancourt, L.-P. Rivest, and J. Asselin. Using orientation statistics to investigate variations in human kinematics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(1):81–94, 2000.

L.-P. Rivest. A directional model for the statistical analysis of movement in three dimensions. *Biometrika*, 88(3):779–791, 2001.

L.-P. Rivest. A correction for axis misalignment in the joint angle curves representing knee movement in gait analysis. *Journal of Biomechanics*, 38(8): 1604–1611, 2005.

L.-P. Rivest and T. Chang. Regression and correlation for 3×3 rotation matrices. *Canadian Journal of Statistics*, 34(2):187–202, 2006.

L.-P. Rivest, S. Baillargeon, and M. Pierrynowski. A directional model for the estimation of the rotation axes of the ankle joint. *Journal of the American Statistical Association*, 2012.

B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63 (2):243–259, 2001.

H. Sadeghi, P. A. Mathieu, S. Sadeghi, and H. Labelle. Continuous curve registration as an intertrial gait variability reduction technique. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(1):24–30, 2003.

H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

A. G. Schache, R. Baker, and L. W. Lamoreux. Defining the knee joint flexion–extension axis for purposes of quantitative gait analysis: an evaluation of methods. *Gait & Posture*, 24(1):100–109, 2006.

N. Schenker and J. F. Gentleman. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3):182–186, 2001.

L. Sloot, M. Van der Krogt, and J. Harlaar. Self-paced versus fixed speed treadmill walking. *Gait & Posture*, 39(1):478–484, 2014.

M. Solomonow. Ligaments: a source of work-related musculoskeletal disorders. *Journal of Electromyography and Kinesiology*, 14(1):49–60, 2004.

A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(7):1415–1428, 2011a.

A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. Marron. Registration of functional data using fisher-rao metric. *arXiv preprint arXiv:1103.3817*, 2011b.

B. Stanfill, U. Genschel, and H. Hofmann. Point estimation of the central orientation of random rotations. *Technometrics*, 55(4):524–535, 2013.

M. A. Stephens. Vector correlation. *Biometrika*, 66(1):pp. 41–48, 1979.

J. Stuelpnagel. On the parametrization of the three-dimensional rotation group. *SIAM review*, 6(4):422–430, 1964.

J. Su, S. Kurtek, E. Klassen, A. Srivastava, et al. Statistical analysis of trajectories on Riemannian manifolds: bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics*, 8(1):530–552, 2014.

J. E. Taylor. A Gaussian kinematic formula. *The Annals of Probability*, 34(1): 122–158, 2006.

J. E. Taylor and K. J. Worsley. Detecting sparse signals in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, 102(479):913–928, 2007.

J. E. Taylor and K. J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *The Annals of Statistics*, 36(1):1–27, 2008.

J. E. Taylor, A. Takemura, and R. J. Adler. Validity of the expected Euler characteristic heuristic. *The Annals of Probability*, 33(4):1362–1396, 2005a.

W. R. Taylor, R. M. Ehrig, G. N. Duda, H. Schell, P. Seebeck, and M. O. Heller. On the influence of soft tissue coverage in the determination of bone kinematics using skin markers. *Journal of Orthopaedic Research*, 23(4):726–734, 2005b.

S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(4):376–380, 1991.

L. van Elfrinkhof. Eene eigenschap van de orthogonale substitutie van de vierde orde. In *Handelingen van het zesde Nederlandsch Natuuren Geneeskundig Congres*, pages 237–240, 1897.

S. Vantini. On the definition of phase and amplitude variability in functional data analysis. *Test*, 21(4):676–696, 2012.

H. Wang, J. Marron, et al. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.

J.-L. Wang, J.-M. Chiou, and H.-G. Mueller. Review of functional data analysis. *arXiv preprint arXiv:1507.05135*, 2015.

K. Wang, T. Gasser, et al. Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3):1251–1276, 1997.

H. J. Woltring. 3D attitude representation of human joints: a standardization proposal. *Journal of Biomechanics*, 27(12):1399–1414, 1994.

H. Ziezold. Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes*, A: 591–602, 1977.