Intercomponent Time Dynamics for Multivariate Functional Data

By

Cody James Carroll

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Hans-Georg Müller, Chair

Jane-Ling Wang

Ethan Anderes

Committee in Charge

2021

For JoAnn Marinelli and Indy.

Contents

List of Figures

List of Tables

Abstract

Intercomponent Time Dynamics
for Multivariate Functional Data

Multivariate functional data often present theoretical and practical complications which are not found in their univariate functional counterparts. One of these is a situation where the component functions of multivariate functional data are subject to mutual time warping. That is, the component processes exhibit a similar shape but are subject to systematic phase variation across their time domains. This dissertation addresses this previously unconsidered mode of warping with the introduction of multivariate time warping models which rely on either time-shifting or nonlinear time-distortion frameworks.

In the first chapter, we introduce a shift-warping model for multivariate time relations. This model is motivated by the Zürich longitudinal growth dataset, in which the growth trajectories for multiple body parts were observed from birth to adulthood for a sample of children. The proposed method differs from existing registration methods for functional data in a fundamental way. Namely, instead of focusing on the traditional approach to warping, where one aims to recover individual-specific registration, this technique focuses on shift registration across the components of a multivariate functional data vector on a population-wide level. After applying the method to the Zürich data, we find that there exists an archetypal ordering of pubertal growth spurts across modalities: on average, legs tend to experience peak growth velocity approximately a half a year before standing height and arms, which themselves tend to precede the growth spurt of the spine by half a year. Our proposed estimates for these shifts are identifiable, enjoy parametric rates of convergence, and often have intuitive physical interpretations, all in contrast to traditional curve-specific registration

approaches. Finite sample properties of these estimators were also investigated in simulation studies.

The second chapter of this dissertation introduces the Latent Transport Model for multivariate functional data. This model widens the class of possible cross-component warps from simple shifts to flexible and nonlinear transport functions. The proposed approach combines a random amplitude factor for each component with population based registration across the components of a multivariate functional data vector. It also includes a latent population function, which corresponds to a common underlying trajectory as well as a subject-specific warping component. This model allows for meaningful interpretation and is well suited to represent functional vector data. We also propose estimators for all components of the model, which not only lead to a novel representation for multivariate functional data, but can also be used in downstream analyses like Fréchet regression. Rates of convergence are established when curves are fully observed or observed with measurement error. The usefulness of the model, interpretations, and practical aspects are illustrated in simulations and with application to multivariate human growth curves as well as multivariate environmental pollution data.

Acknowledgments

I would like to thank my mentor, Prof. Hans-Georg Müller, for all of the guidance, knowledge, patience, and support that he has shared with me. Through his unmistakable dedication to his students and profound work ethic, he has been a true anchoring force for me while writing this dissertation in the midst of a pandemic. I have learned so much from him– about statistics and otherwise– and I feel deeply fortunate to have known him as my advisor.

I am also extremely grateful to Prof. Jane-Ling Wang for our discussions about teaching, research, and consulting. I am unquestionably a better thinker and communicator of statistics because of her.

My sincere thanks go to my Dissertation Committee, Prof. Ethan Anderes, Prof. Christiana Drake, and Prof. Emilio Ferrer, for their willingness to share their perspectives on my research.

I would also like to acknowledge the following professors and collaborators whose encouragement and mentorship have helped me grow as an academic: Prof. Wolfgang Polonik, Prof. Alois Kneip, Prof. Raymond Huey, Dr. Renee Gilbert-Levin, and Dr. Brent Covele.

Likewise, I am deeply indebted to my past teachers, Kathryn Rutledge, Josh Recio, and Prof. Alisa Havens Walch, who helped cultivate my interest and passion for math and statistics when I was starting out. Undoubtedly I would not have followed the same path were I not so fortunate to have had them as educators and role models.

Many thanks go to my colleagues and fellow researchers in Hans and Jane-Ling's group. The help and camaraderie shared by Dr. Paromita Dubey, Dr. Yaqing Chen, Alvaro Gajardo, Satarupa Bhattacharjee, Anthony Santistevan, and many others have been invaluable during our time together.

Lastly, I want to express my sincere appreciation for my biggest support group: my friends and family. I owe a great deal to my friends back home in Texas, my kickball teammates in Sacramento, and my friends from graduate school in Davis, all of whom have kept me sane during the ups and downs of my doctorate. I am also extremely grateful to my family for their unconditional love and belief in me. In particular, I would like to express my deepest thanks to the following people: my special aunts, Prof. Betty Travis, the first person to introduce me to the notion of a function, and Denise Janes, whose conversations and unconditional love helped me get through the struggles I experienced when I was younger; my copyeditors, Madison Janes and Jay Campbell, whose keen eyes saved this dissertation from countless typos; my Blueplex roommates, Michael Wallace, Lauren Heffernan, and Russell Fry, for looking out for me when I was home and my cats when I wasn't; my grandparents, James and Robbie Carroll, whose support has never wavered, and JoAnn Marinelli, whose hustle and spirit lives on in me; and finally my family, including my sweet cats, Indy and Red, my brother, Kyle, my sister, Kaylee, and my parents, Jim and Donna Carroll. All the love you gave has defined me and for that I am more grateful than words can express.

Chapter 1

Multivariate Functional Shift-Warping Models

1.1 Introduction

Multivariate functional data are often encountered in biological or chemical processes that are continuously measured for a group of subjects or observational units. Such processes arise in many longitudinal studies, especially in the biomedical sciences, the scopes of which range from human growth to time courses of protein levels during metabolic processes (Park & Ahn, 2017; Dubin & Müller, 2005). With the increasing ubiquity of multivariate functional data, the study of how to treat such data has recently become a very active field, particularly in the context of clustering (Brunel & Park, 2014; Jacques & Preda, 2014; Park & Ahn, 2017), functional regression (Chiou, 2012; Chiou et al., 2016), and in terms of general modeling of functional data (Claeskens et al., 2014). Common approaches for analyzing multivariate functional data have focused on dimension reduction via multivariate functional principal components (MFPCA) (Zhou et al., 2008; Chiou et al., 2014; Happ & Greven, 2018) or decomposition into component-specific processes and their interactions (Chiou et al., 2016).

In applications such as growth curves, if we view multivariate longitudinal data as generated by an underlying p-dimensional smooth stochastic process, the component curves of the functional vector may exhibit mutual time warping. If left unchecked, such vector component warping may distort principal components and inflate data variance, while if handled properly, it may yield intuitive physical interpretations and a more parsimonious representation of the data. As far as we know, the idea of explicitly modeling time relations between component processes has not yet been considered for multivariate functional data, which allows one to take advantage of repeated observations of a multivariate process for a cohort of subjects.

Typically, for each subject in longitudinal studies one has measurements on a grid of time points, where recordings are possibly contaminated with measurement error. Often these measurements are multivariate, notably in growth studies, which prompts consideration of functional methods which are geared towards repeatedly sampled multivariate functional data. The Zürich Longitudinal Growth Study motivated us to model such multivariate functional data by allowing the components to be mutually time-shifted against each other, as some components of growth may systematically precede others.

The idea of warping across components is most pragmatic when the component processes of multivariate functional data exhibit similarity in their shapes. In the case of growth studies, each body part's component process follows the same general pattern: a period of rapid development during infancy which then slows to a roughly constant rate of growth until puberty, at which time the growth velocity peaks (i.e., the pubertal growth spurt) before decreasing to zero as the subject reaches adulthood (Gasser et al., 1984c). The multivariate aspect of these growth curves allows us to compare the growth processes of different parts of the body. For example, it may be that legs undergo their growth spurt earlier in

life than arms do. It is an interesting biological question to search for a common process that ordinates the timings of growth spurts across body parts. Another situation where this phenomenon arises is in the above-mentioned recordings of protein levels during metabolic processes. Certain biological functions are associated with peaks and valleys of certain protein levels and their relative timings expose the order of the underlying enzymatic mechanisms at work.

Data from the Zürich Longitudinal Growth Study were previously used to investigate the timing of growth spurts across body parts using a phase-clustering model (Park & Ahn, 2017). Our study uses the same data but instead emphasizes the investigation of phase variations in the component growth velocity curves to establish time relations. In particular, we investigate mutual time warping in the derivatives across the components of the multivariate functional processes during a growth spurt window, as derivatives are more informative about human growth than the growth curves themselves. Specifically, we assume a model which uses relative time shifts between component processes to establish their pairwise time relations. Information about the relative shifts between pairs of components may then be combined to inform the full system of relative timings across body parts. We emphasize that our approach, while motivated by growth data, is by no means limited to this application and can also be implemented for multivariate functional data which has neither a well-defined time origin nor an endpoint, as in the case of blood protein time courses (e.g., Dubin & Müller (2005)).

1.2 A Shift-Warping Model for Multivariate Time Relations

To illustrate the idea of mutual component warping, consider the growth velocities for a handful of representative children in the Zürich Longitudinal Study

(Fig. 1.1), which will be revisited in its entirety in Section 1.5. We consider pubertal growth, i.e. growth curves are evaluated in the interval $\mathcal{I} = [9, 18]$ ranging from 9 to 18 years. Each child has four growth velocity curves, each corresponding to a different body part. The peaks represent the moment of maximal rate of growth and can be used as a crude measure of the timing of pubertal growth spurt for that modality. For ease of viewing we mark these locations in time with vertical lines in Figure 1.1.

A key observation is to recognize that regardless of when the child underwent puberty, the ordering of the spurts is consistent: legs undergo their growth spurts first, then arm length and standing height roughly together, followed by sitting height. This pattern in pubertal spurts was briefly discussed in the descriptive growth studies of Sheehy et al. (1999) and suggests that there is a population-wide mutual component warping occurring across the four modalities. Note also that the time differences between modalities are relatively consistent across children, despite individual differences in the age of the pubertal onset. This is worth highlighting for two reasons: (1) it motivates the estimation of a fixed population-wide set of shift parameters, and (2) it shows that cross-component registration makes sense even in the presence of subject-specific time warping, which is the usual mode of warping considered in univariate functional data. For a recent overview of traditional warping methods, we refer less familiar readers to Marron et al. (2015) and Wang et al. (2016).

To register these curves across components, we propose a shift-warping model, which provides a simple and interpretable method for cross-component alignment of growth data. From a methodological point of view, our approach builds on basic ideas in the literature on parametric and semi-parametric modeling of growth and related phenomena. In applied work on human growth, empirical studies often utilize parametric models (Milani, 2000).

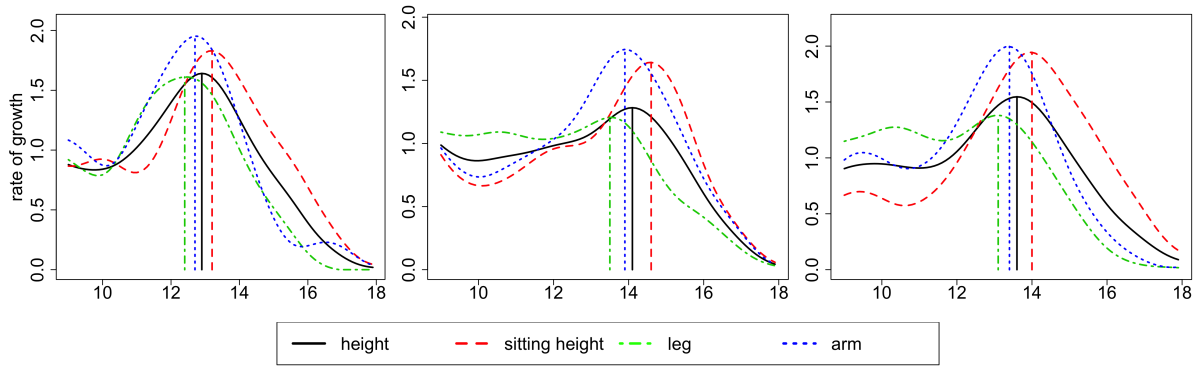

Figure 1.1: Three children's growth velocities for standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (blue, dotted). Peak velocity positions are marked with vertical lines and can be used as rough markers of pubertal onset for each modality.

One of the most popular classes of models has been proposed by Preece & Baines (1978); for a recent application, see, e.g. Banik et al. (2017). All of these models make use of shift parameters θ_{ij} to capture the main differences in individual timings. For p-dimensional multivariate functional data, $\{X_{i1}(t), \ldots, X_{ip}(t)\}^T$, $i = 1, \ldots, n$, which we consider here on a domain $\mathcal{I}$ that covers the pubertal period, an extension of the existing models to the multivariate case is as follows.

For some function G and some additional parameter vectors $\boldsymbol{\xi}_{ij}$ one posits that, with time shifts θ_{ij}, the growth curve for the j^{th} component of the i^{th} subject has the form

$$X_{ij}(t) = G(\boldsymbol{\xi}_{ij}, t - \theta_{ij}), \quad j = 1, \ldots, p, \ i = 1, \ldots, n, \ t \in \mathcal{I}, \tag{1.1}$$

where previously only cases with $p = 1$ have been considered. As fully parametric specifications were found to lack accuracy, various semi-parametric extensions have been proposed for the one-dimensional case. For example, for standing height, in the case $p = 1$, Kneip & Gasser (1995) assumed a shape-invariant model with $G(\boldsymbol{\xi}_{ij}, t - \theta_{ij}) = \xi_{ij;2} f\{\xi_{ij;1}(t - \theta_{ij})\} + \xi_{ij;3}$ for real-valued parameters $\xi_{ij;1}, \xi_{ij;2}, \xi_{ij;3}$, and an unknown real-valued function f which is estimated from the data. The k-mean alignment introduced by Sangalli et al. (2010) may be seen as a generalization of this framework, where it is assumed that the population can be decomposed into K disjoint clusters, and individual functions belonging to each cluster can be approximately described by a shape-invariant model with respect to a cluster-specific template function f_g, $g \in \{1, \ldots, K\}$.

In the following we assume that growth data follow a multivariate and flexible version of models of type (1.1), under the natural assumption that the shift parameters θ_{ij} can be decomposed in the form $\theta_{ij} = \theta_i + \theta_j$, where θ_i is specific for the individual, while θ_j is specific for the component. Then (1.1) may be rewrit-

ten in the form

$$X_{ij}(t) = G(\boldsymbol{\xi}_{ij}, t - \theta_i - \theta_j) \equiv G^*(\boldsymbol{\xi}^*_{ij}, t - \theta_j), \quad \text{where } \boldsymbol{\xi}^*_{ij} = (\boldsymbol{\xi}_{ij}, \theta_i). \tag{1.2}$$

Motivating our alignment procedure is that, for a given individual i, the component functions $X_{i1}(t), \ldots, X_{ip}(t)$ can be made more similar when removing the different shift parameters $\theta_1, \ldots, \theta_p$. The most favorable situation arises if shifts constitute the only important difference between components such that $\boldsymbol{\xi}_{ij} \equiv \boldsymbol{\xi}_i$ is independent of $j = 1, \ldots, p$. Then with $Z_i(s) := G^*(\boldsymbol{\xi}^*_i, s)$ we arrive at

$$X_{ij}(t) = Z_i(t - \theta_j), \quad j = 1, \ldots, p,\ i = 1, \ldots, n,\ t \in \mathcal{I}, \tag{1.3}$$

so that $E\int_{\mathcal{I}}\{X_{ij}(t + \theta_j) - X_{il}(t + \theta_l)\}^2 dt = E\int_{\mathcal{I}}\{Z_i(t) - Z_i(t)\}^2 dt = 0$ for all $j, l \in \{1, \ldots, p\}$. To apply this argument may require some pre-processing in order to eliminate scale differences between the different components (see Section 1.5).

In the context of growth curves, subject-specific alignment based on non-parametric monotonic warping functions $h_i : \mathcal{I} \to \mathcal{I}$ has been studied extensively (Gasser et al., 1990; Kneip & Gasser, 1995; Gervini & Gasser, 2004; Tang & Müller, 2008). Higher dimensional problems of subject-specific registration have been considered through the lens of elastic shape analysis (Srivastava et al., 2010; Srivastava & Klassen, 2016), or reduced to the problem of aligning a univariate curve generated from the component curves (Ramsay et al., 2014). It can be seen from (1.2) and (1.3) that in our context such functions h_i do not play any role and may simply be part of the parameter set $\boldsymbol{\xi}_i$. We therefore emphasize that in the non-traditional warping framework presented here, the pertinent issues are fundamentally different from those considered in the subject-specific warping framework discussed in the cited articles. In short, that it bypasses dealing with individual warping functions is a strength of our method and allows us to side-step the identifiability problems associated with subject-specific registration. It is especially noteworthy that we obtain a $\sqrt{n}$-rate of convergence for

the estimated time shifts to their targets under mild regularity conditions (see Section 1.7). Such fast convergence rates cannot be obtained in traditional warping approaches, since these focus on individual warps rather than component-specific warping and therefore require identification of n time alignments, where n is the sample size, whereas in our approach there are only p components that need to be considered, where p is the fixed dimension of the multivariate process. Of course, in some circumstances the model in (1.3) may just serve as a convenient approximation of a more nuanced warping relation between components. We discuss the potential for continuous analogues of cross-component shift-warping techniques in the Concluding Remarks.

A further distinction between cross-component warping as proposed here and the common subject-specific approach is that the latter traditionally views the presence of individual warping functions as a nuisance characteristic of the data to be accounted for in order to correctly analyze underlying functional features of interest; for example, curves will be registered first before conducting a functional principal component analysis (FPCA). In contrast, we argue that investigation of cross-component warping and the shift parameters $\theta_1, \ldots, \theta_p$ provide insight into intercomponent relationships and, when applicable, are an essential aspect of multivariate functional data that is of genuine interest rather than a nuisance.

1.3 Bivariate Cross-Component Registration

1.3.1 Pairwise-Shift Estimation

We introduce here the main idea of registering different component times across modalities, which we call Cross-Component Registration (XCR). As explained in the previous section, XCR differs in key aspects from traditional warping, which is also known as curve registration or alignment (Ramsay & Silverman, 2005;

Kneip & Gasser, 1992; Silverman, 1995), as it aims at a situation where the component curves of a multivariate functional process are time-shifted versions of one another. A major difference is that instead of estimating n individual warping functions, which align curves across subjects and the determination of which is the goal of traditional curve warping methods, our new approach targets a p-vector of shift parameters for the case of p-dimensional functional data. These component-wise shifts are then applied uniformly across all subjects to mutually align the component curves.

In the following, we write $(X_1, \ldots, X_p)^T$ to represent the generic underlying multivariate process and $\{X_{i1}(t), \ldots, X_{ip}(t)\}^T,\ i = 1, \ldots, n$, for a sample of realizations of the functional vector. One may assume *a priori* smoothness of curves or may preprocess the data with a smoothing method if the curves are subject to measurement error. In this subsection we consider the case of multivariate functional data with just $p = 2$ component curves to introduce the main ideas, and will then discuss the extension to $p > 2$. To fix the idea, consider a sample of bivariate functional processes, writing $\{X_{i1}(t), X_{i2}(t)\}_{i=1}^n$ for the observed i.i.d. realizations of the bivariate process (X_1, X_2), and assume that the domain of both component processes is a compact interval $\mathcal{T} = [0, T]$. As a criterion for alignment and to determine the optimal shift, we aim to minimize the $\mathcal{L}^2$-distance between functions on a subinterval $\mathcal{I} \subset \mathcal{T}$; see the discussion below. Using a simple shift-warp family under the $\mathcal{L}^2$-norm allows for a straightforward and clear interpretation of the relationship between two components and has been used previously in the context of shape-invariant modeling (Härdle & Marron, 1990; Kneip & Gasser, 1995; Silverman, 1995).

Specifically, we aim for the optimal value of the parameter τ, the pairwise *cross-component (XC) shift* as the minimizer of

$$\Lambda(\tau) = E \int_{\mathcal{I}} \{X_1(t) - X_2(t - \tau)\}^2 dt, \tag{1.4}$$

with associated sample version

$$L_n(\tau) = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{i2}(t-\tau)\}^2 dt \tag{1.5}$$

and sample-based shift parameter estimate

$$\widehat{\tau} = \underset{\tau}{\operatorname{argmin}}\, L_n(\tau), \tag{1.6}$$

targeting $\tau_0 = \operatorname{argmin}_\tau \Lambda(\tau)$.

Integrating over a subinterval $\mathcal{I}$ rather than the whole interval is a device that is necessary in order to ensure that both the shifted and unshifted curves are defined on the domain of integration. If we did not specify a suitable subinterval $\mathcal{I} \subset [0, T]$ that stays away from both 0 and T, shifting a curve forward or backward may result in a subinterval of integration in which one of the curves is defined while the other is not, making it impossible to compute their $\mathcal{L}^2$-distance. To be precise, we partition the data domain $\mathcal{T}$ into three disjoint intervals $\mathcal{T} = R_1 \cup \mathcal{I} \cup R_2$, where $\mathcal{I} = [r_1, r_2]$ is the subinterval of integration and $R_1 = [0, r_1)$ and $R_2 = (r_2, T]$ are the remaining intervals on the boundary. Note that this partitioning implies that the magnitude of pairwise shift estimates cannot exceed the length of the relevant remainder interval, depending on the direction of the shift. This subtlety suggests that the choice of subinterval of integration $\mathcal{I}$ is not trivial and should be done carefully and data-adaptively.

1.3.2 Subinterval Selection

We propose the following guidelines for subinterval selection: $\mathcal{I}$ should be chosen to (1) include the critical features of the sample curves, and (2) avoid censoring estimates of pairwise shifts. For example, in our application to the Zürich data, we choose $\mathcal{I}$ to range from the earliest age of pubertal onset to the age of adulthood. Doing so ensures the inclusion of the main pubertal growth spurt peaks which are the structural features to be aligned across components (Gasser

& Kneip, 1995). Unreasonable estimates may occur if the subinterval is too small, as an inappropriately narrow window may discard the features to be aligned for a subset of individuals.

The problem of subinterval selection was discussed previously in Kneip & Gasser (1995) and we follow their convention to seek an "overlapping interval" across all individuals, described as follows. Individual intervals $\mathcal{J}_i$ are chosen such that information about structural landmarks for the i^{th} individual are contained entirely in $\mathcal{J}_i$. Then the overlapping interval $\mathcal{J}$ is defined as $\mathcal{J} = \cup_i \mathcal{J}_i$ and guarantees that all individuals' structural features are included. One can then either simply use this overlapping interval as the subinterval of integration, i.e., let $\mathcal{I} = \mathcal{J}$, or choose $\mathcal{I}$ such that $\mathcal{J} \subset \mathcal{I}$ and $\mathcal{I}$ has some relevant physical meaning. An example for the latter case is demonstrated in the data application of Section 1.5.

In the more general setting with more than two components, we will encounter several pairwise time shifts between sets of component curves. To distinguish between these, we write τ_{jk} to denote the relative time shift which moves component k to component j. Note that the sample and population time shifts are symmetric in the sense that $\tau_{jk} = -\tau_{kj}$. The problem of estimating general cross-component shift parameters $\theta_1, \ldots, \theta_p$ can be solved after the estimation of all the pairwise shift parameters τ_{jk} for $1 \leq j < k \leq p$, as discussed in the following section.

1.4 General Cross-Component Registration

We now extend the methodology for bivariate cross-component registration (XCR) to the case of p-dimensional multivariate functional processes, aiming to align more than two component functions. Assume we observe p-variate functional data $\{X_{i1}(t), \ldots, X_{ip}(t)\}^T$ for $i = 1, \ldots, n$, now with $p > 2$. We search for a vector of

global XC shifts, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, such that when each modality $X_j(t),\ j = 1, \ldots, p$, is shifted by θ_j, all p curves are aligned. Here it is useful to introduce the idea of an underlying *latent process*, which may be seen as the Z_i component in model (1.3).

To fix the idea, consider only a single observation of simulated multivariate functional data where the components of the multivariate process are just time-shifted replicates. Figure 1.2 illustrates an example for $p = 4$. A simple approach would be to align the component curves by fixing one component curve and shifting the others via bivariate XCR to align them with the selected component. However, a major problem with this approach is that the resulting XC shifts depend on the choice of the fixed component.

These problems can be overcome by assuming that each curve is a shifted version of an unobserved and unshifted latent component, visualized as the solid curve in Figure 1.2. The observed components are then time-shifted with respect to this latent component and the shifts are subject to the constraint $\sum_{j=1}^{p} \theta_j = 0$, so that there is no net XC shift from the latent component curve. This assumption is necessary for the identifiability of the shift estimates.

A key observation is that there is a linear relationship between pairwise XC shifts, τ_{jk}, and the global XC shifts, θ_j and θ_k. Specifically, the pairwise shifts can be expressed as the difference of two global shifts as shown in Eq. 1.7. Thus, after estimating bivariate XC shifts τ_{jk} between component functions, we can infer the global XC vector $\boldsymbol{\theta}$, and importantly, the linear map between the two is invariant with respect to the choice of the latent process.

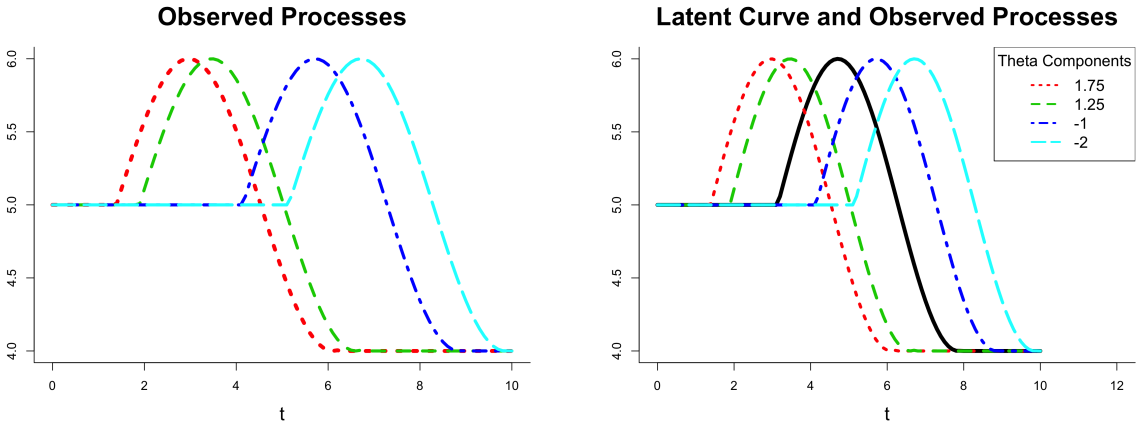

Figure 1.2: Observed components (dashed, left) and latent curve (solid, right) defined by identifiability constraint.

More explicitly, the linear map L is given by:

$$\tau_{jk} = \theta_j - \theta_k, \quad j,\ k = 1, \ldots, p, \quad j < k \tag{1.7}$$

with constraint $\sum_{j=1}^{p} \theta_j = 0$, so that

$$\boldsymbol{\tau}^* = L(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta}, \tag{1.8}$$

where $\boldsymbol{\tau}^* = (\boldsymbol{\tau}^T, 0)^T = (\tau_{12}, \tau_{13}, \ldots, \tau_{(p-1)p}, 0)^T$ is the pairwise shift parameter vector stacked with 0, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is the global shift vector of each component function with respect to the latent process, and $\mathbf{A}$ is the matrix of the linear map which corresponds to the contrasts in (1.8). Note that $\mathbf{A}$ is of dimension $(p(p-1)/2) \times p$, and is always of full column rank. Explicitly, we write

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & \ldots & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}.$$

To implement this approach, we must first estimate the stacked vector of bivariate XC shifts, $\widehat{\boldsymbol{\tau}}^* = (\widehat{\boldsymbol{\tau}}^T, 0)^T = (\widehat{\tau}_{12}, \widehat{\tau}_{13}, \ldots, \widehat{\tau}_{(p-1)p}, 0)^T$, leading to the model

$$\widehat{\boldsymbol{\tau}}^* = \mathbf{A}\boldsymbol{\theta} + \varepsilon, \tag{1.9}$$

where ε is a vector of random noise with mean 0 and finite variance. Once the pairwise shifts $\widehat{\tau}_{jk}$ are obtained, global shifts $\boldsymbol{\theta}$ can be estimated as

$$\widehat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widehat{\boldsymbol{\tau}}^* \tag{1.10}$$

by ordinary least squares. The p component curves will then be aligned (to the latent curve) once they are time-shifted with their respective estimated global XC shifts, $\widehat{\boldsymbol{\theta}}$, i.e., $X_{ij}(t + \hat{\theta}_j)$ for $j = 1, \ldots, p$.

1.5 Application to the Zürich Longitudinal Growth Study

From 1954 to 1978, a longitudinal study on human growth and development was conducted at the University Children's Hospital in Zürich. Modalities of growth that were longitudinally measured on a dense regular time grid include standing height, sitting height, arm length, and leg length, so that the resulting data can be naturally viewed as multivariate functional data (Gasser et al., 1984a, 1989). The raw trajectories of the $p = 4$ component processes for the children measured are displayed in Figure 1.3, which also indicates the measurement grid. Component curves are initially observed on the domain $\mathcal{T} = [0, 20]$, which can be artificially extended to the right by assuming measurements stay constant in adulthood, since almost all subjects reach full maturation before age 20. We are especially interested in the timing of pubertal growth spurts, which occur for all individuals between ages 9 and 18 typically. We are using this time window as the subinterval of integration, $\mathcal{I}$, in accordance with the guidelines of Section 1.3. A common way to study growth velocities is to examine the derivatives of the growth curves instead of the curves themselves (Gasser et al., 1984c). The growth velocities have a peak during puberty, with the apex representing the instant when an individual's growth rate is at its maximum. Previous analysis of human growth curves indicates that there is a difference in the ways that boys and girls undergo puberty (Gasser et al., 1984c; Eiben et al., 2005). For example, it is widely known that girls begin puberty at younger ages than boys do on average. Accordingly, for the subsequent analysis we separate boys and girls and for

brevity display only the results for boys. We estimate the growth velocities, i.e., the derivatives of the growth trajectories, via local weighted linear smoothing using the package `fdapace` (Carroll et al., 2020a).

Because different body parts have different physical sizes, their velocities are also on different scales. We eliminate the majority of this amplitude variation by dividing each function by the total area under the curve, resulting in "relative velocities" for each modality. Relative velocities have been previously used in the growth curve literature (see, e.g. Sheehy et al. (1999)) and allow for the comparison of modalities which are on dissimilar scales. Figure 1.4 shows the rescaled derivative estimates for the four growth processes that we consider. After this pre-processing, we now have multivariate functional data with component functions such as those shown for the individuals in Figure 1.1. When we apply the proposed shift model to the growth velocities of the four growth modalities of the Zürich data, we obtain the following estimated global XC shifts (Table 1.1):

Component	Modality	Estimate
θ_1	Height	-0.0875
θ_2	Sitting Height	-0.5850
θ_3	Leg Length	0.5825
θ_4	Arm Length	0.0900

Table 1.1: Estimated global XC shifts for Zürich boys. These estimates imply the following ordering of growth spurts: (1) leg, (2) height, (3) arm, (4) sitting height.

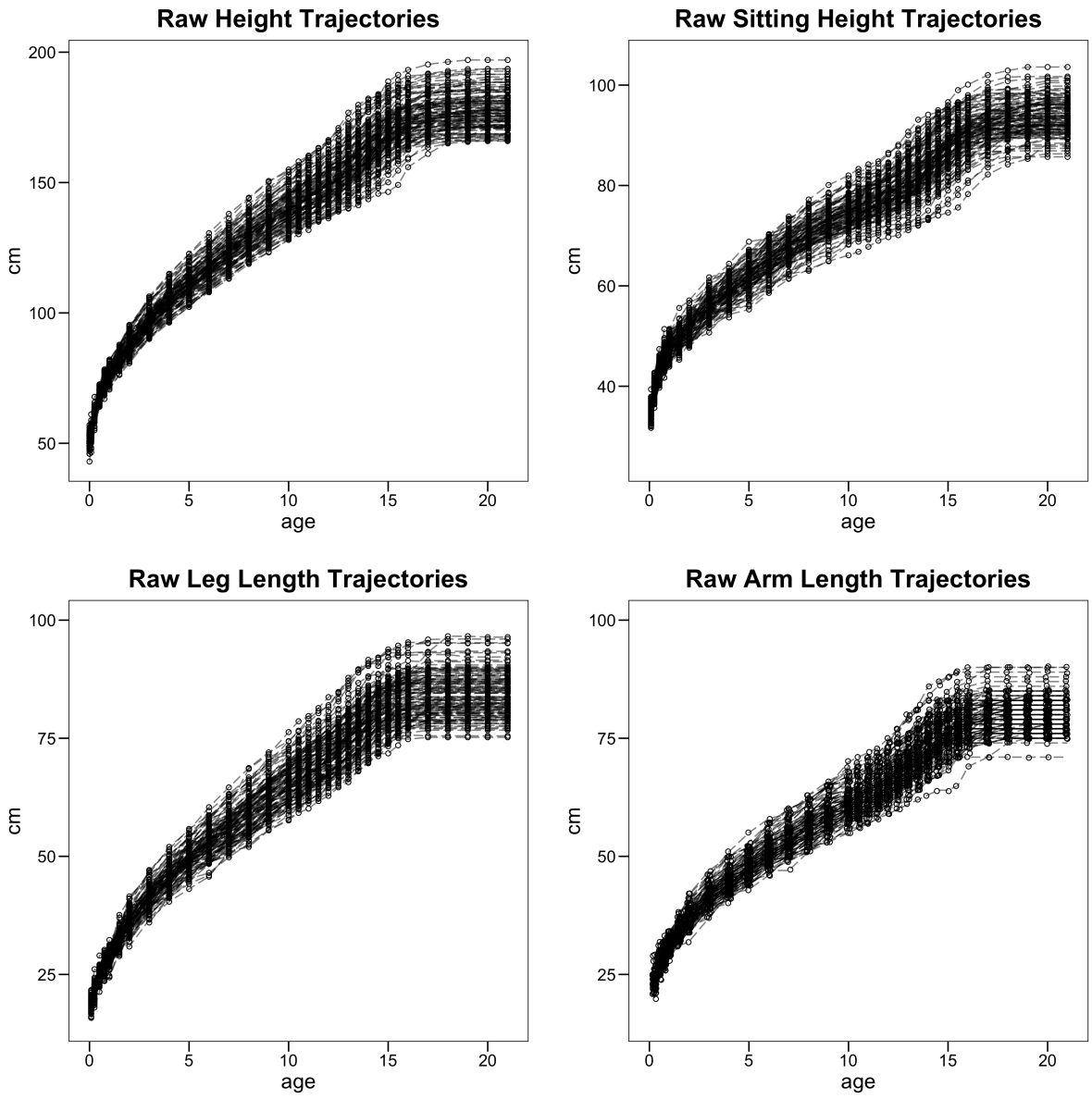

Figure 1.3: Raw growth trajectories for all Zürich boys.

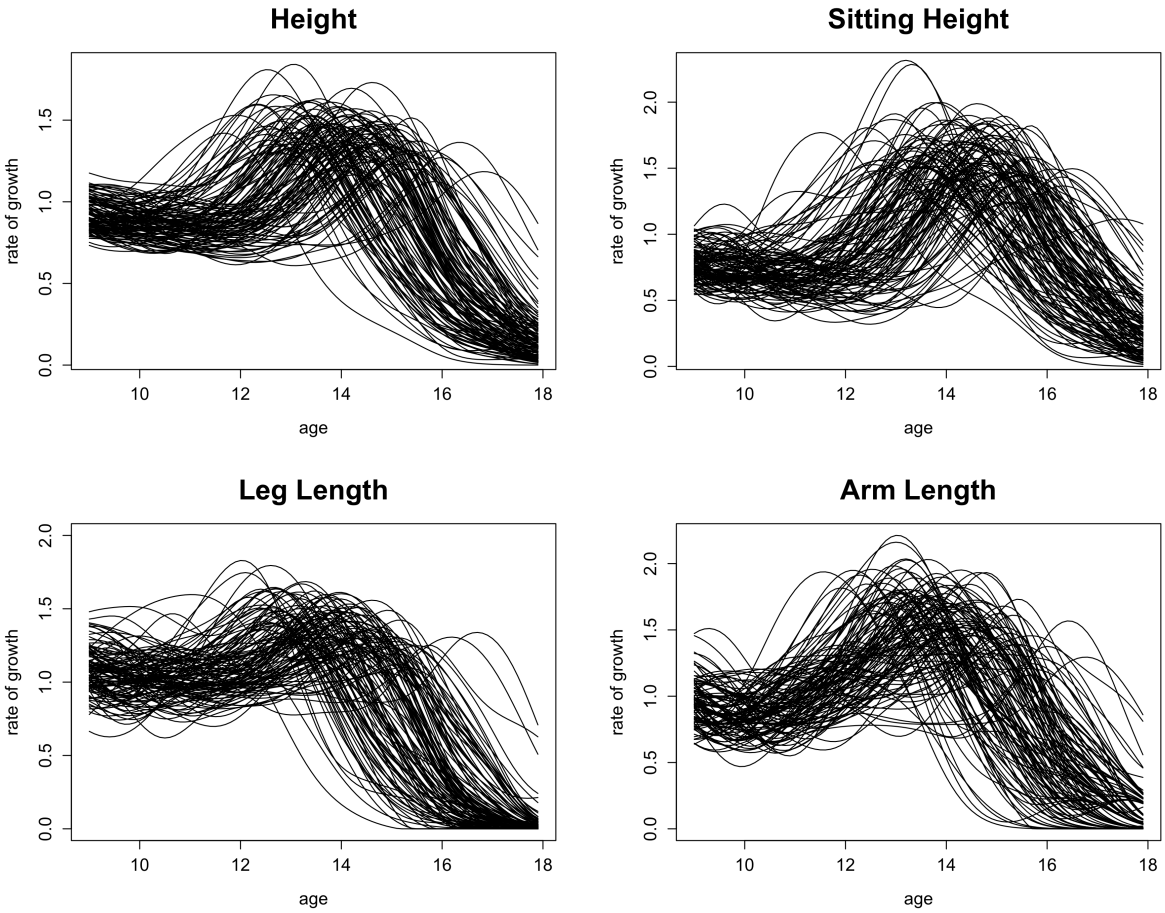

Figure 1.4: Scaled growth velocity curves for Zürich boys.

One can interpret these shift parameters in a pairwise manner. For example, legs tend to undergo their growth spurts roughly half a year before arms do ($\widehat{\tau}_{34} = \widehat{\theta}_3 - \widehat{\theta}_4 \approx 0.5$) and sitting height trails roughly half a year behind standing height ($\widehat{\tau}_{21} = \widehat{\theta}_2 - \widehat{\theta}_1 \approx -0.5$). Our shift estimates and their implied order of growth spurts is consistent with what is known about human growth patterns, as discussed in the descriptive longitudinal studies of Cameron et al. (1982) and Sheehy et al. (1999).

We next investigate some individuals before and after component alignment for a demonstration of how XC alignment affects the curves. Figure 1.5 (top) shows three individuals who are representative of the "average" ordering of growth spurts across modalities, whereas Figure 1.5 (bottom) displays those who generally went through pubertal spurts for whom the different body parts were already in sync before alignment. Individuals like those shown in Figure 1.5 (bottom) for whom alignment moved component curves further away from each other were very rare, and it was common for most individuals to have reduced $\mathcal{L}^2$-distance between the component curves after alignment.

To illustrate this further, we use the total cross-component $\mathcal{L}^2$-distance (XD) for an individual as a function of $\boldsymbol{\theta}$,

$$\text{XD}_i(\boldsymbol{\theta}) = \sum_{j<k} \int_{\mathcal{I}} \{X_{ij}(t+\theta_j) - X_{ik}(t+\theta_k)\}^2 dt, \tag{1.11}$$

noting that under a perfect model fit we would have $X_{ij}(t+\theta_j) = Z_i(t)$ for all $j = 1, \ldots, p$, and $\text{XD}_i(\boldsymbol{\theta}) = \sum_{j<k} \int_{\mathcal{I}} \{Z_i(t) - Z_i(t)\}^2 dt = 0$. Figure 1.6 displays the distribution of the difference in total cross-component $\mathcal{L}^2$-distance before and after shifting, i.e., $\text{XD}_i(\mathbf{0}) - \text{XD}_i(\widehat{\boldsymbol{\theta}})$. Here it is noteworthy that implementing component alignment reduced total $\mathcal{L}^2$-distance in the sample by about 40%.

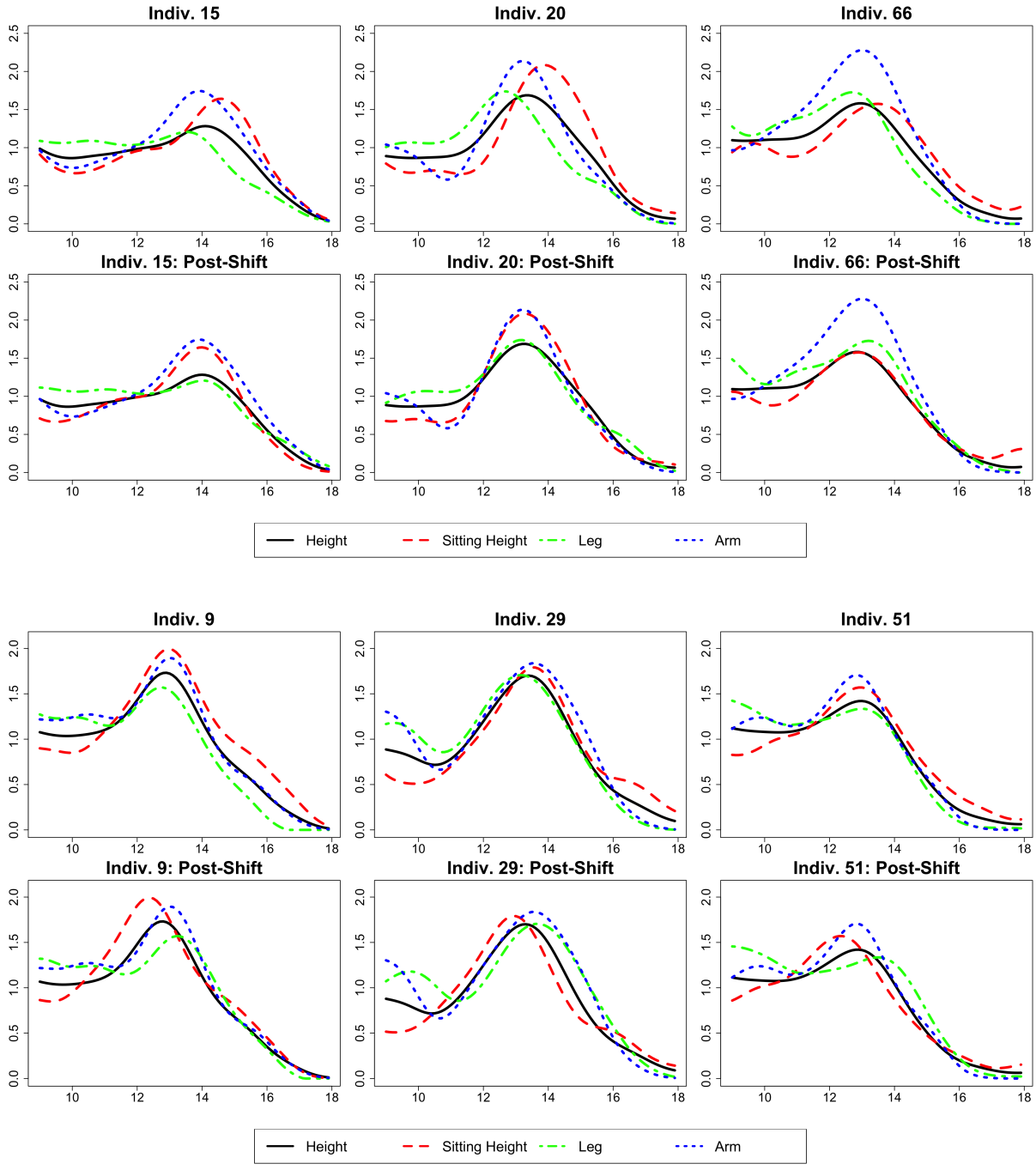

Figure 1.5: Well-aligned (top) and poorly aligned individuals (bottom) after component alignment. Growth modalities are standing height (black, solid), sitting height (red, dashed), leg length (green, dot-dashed), and arm length (blue, dotted).

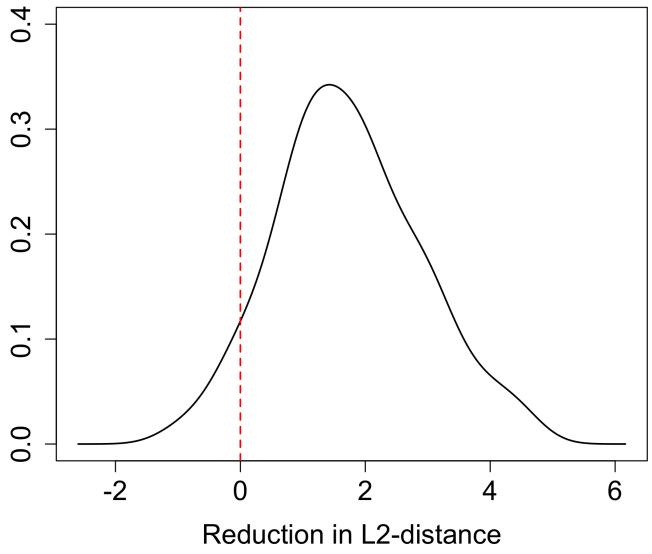

Figure 1.6: Kernel density estimate of the decrease in total $\mathcal{L}^2$-distance after performing XCR. The dashed line indicates no change.

1.6 Simulation Study

We demonstrate here the superior fit of curves aligned by cross-component registration prior to analysis through FPCA. We use the same base curve $Z(t) = 20 - .5t + 30e^{-\frac{(t-25)^2}{72}}$ on $t \in \mathcal{T} = [0, 50]$ as the underlying process dictating the common shape of the component curves and set $\boldsymbol{\theta} = (-5, -2.5, 2.5, 5)$ and $\mathcal{I} = [10, 40]$. We contaminate the curves with functional noise, measurement error, and noisy shift parameters by generating contaminated component curves

$$X_{ij}(t_k) = Z(t_k - \theta_j + \eta_{ij}) + \zeta_{ij} \sin(\frac{\pi t_k}{5}) + e_{ijk}, \tag{1.12}$$

where $\eta_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\eta^2)$, $\zeta_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\zeta^2)$, $e_{ijk} \overset{iid}{\sim} \mathcal{N}(0, 1)$, and k indexes the points on the data grid spanning $\mathcal{T}$ by increments of 0.5. Here the noise on the time domain is introduced through η_{ij}, while noise on the functional domain is controlled through ζ_{ij} and e_{ijk}, which correspond to a random amplitude sine wave and minor additive measurement error, respectively. The base multivariate process

is depicted in Figure 1.7.

One can consider each of the component curves as a single noisy warped realization of the underlying latent curve Z. We may try to estimate the latent curve by viewing all the component processes for all subjects as a noisy sample of Z and then analyzing them through an established method such as FPCA. We expect that failing to account for the component warping will inflate variances and result in a suboptimal fit, since the cross-component warping masks the features of Z, and this is indeed what the following simulations show. A sample of $N = 100$ curves were fit via FPCA using the first two eigenfunctions, both with and without incorporating XCR. When incorporating XCR, curves were first generated and used to estimate XC shifts, whereupon components were shifted according to these estimates, followed by an FPCA step applied to the thus aligned curves. The first two eigenfunctions were used to fit the sample of aligned curves, and after this fitting step the curves were shifted back to their original domains through the estimated shifts. To quantify the advantage of incorporating XCR, we obtained the integrated mean squared error for both approaches. The benefit of including XCR for various noise scenarios was measured through the percent decrease in integrated mean squared error for the sample.

This process was performed $B = 1000$ times under low, medium, and high functional noise settings ($\sigma_\zeta^2 = 25,\ 64,\ 100$), while letting the noise on the time domain start low and increase until it was on the same scale as the shifts themselves. Table 1.2 shows the average percent decreases across replications for various settings. The improvements in fit are relatively consistent across functional noise levels. It is noteworthy to observe that once the noise on the domain becomes comparable to that of the shifts themselves (i.e. $\sigma_\eta^2 > 0.5$), the advantage of XCR starts to decrease. It conforms with expectations that when the within-subject time ordering is highly noise-contaminated, the benefits of performing XCR are

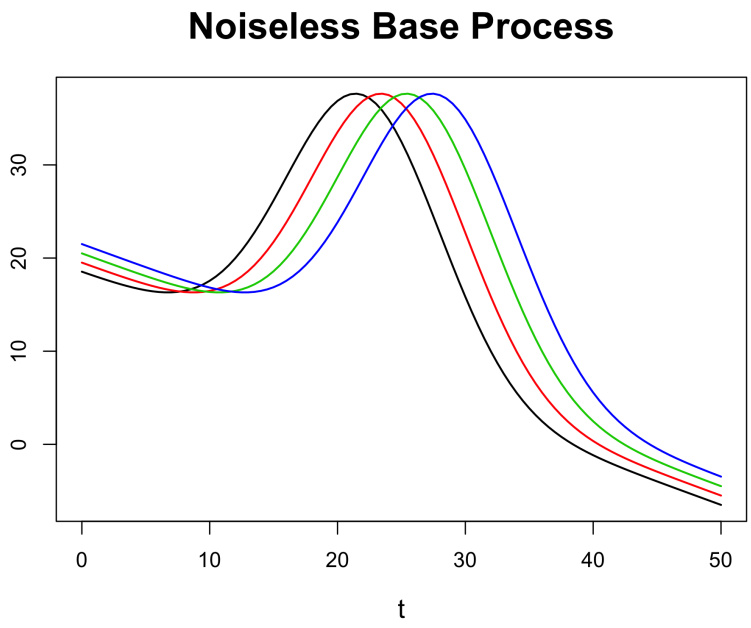

Figure 1.7: Base 4-dimensional process.

lost. At such high shift noise levels there would be little incentive to perform XCR, as exploratory data inspection would not likely indicate the presence of any systematic cross-component warping.

A visual comparison of performance for the two approaches can be seen for an example set of curves in Figure 1.8. Unmodified FPCA is ill-suited to account for sources of horizontal variation, like shift warping, as its eigenfunctions and their scores are geared towards representing vertical variation. In the presence of this horizontal variation, the estimated FPC scores then tend to over- or underestimate the actual amplitude variation, especially near the peaks, as seen in the left side of Figure 1.8. By accounting for component warping with XCR however, the burden of modeling time domain variation is lifted from FPCA, which can then focus on amplitude variability without the confounding phase noise.

We also evaluated the performance of the proposed approach at several levels of noise, using the same base shape and shifts, displayed in Table 1.3. Estimates remain unbiased through all noise settings we consider, indicating that the XCR

method can handle shape and amplitude variations in the main signal, the presence of nuisance peaks, and sizeable measurement error. For a visualization of simulated processes, see Figure 1.9.

Noise Level	$\sigma_\eta^2 = 0.1$	$\sigma_\eta^2 = 0.25$	$\sigma_\eta^2 = 0.5$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 2$
$\sigma_\zeta^2 = 25$	48.38	50.92	51.39	44.11	16.17
$\sigma_\zeta^2 = 64$	48.17	50.81	51.46	43.75	16.64
$\sigma_\zeta^2 = 100$	48.06	50.72	51.37	43.87	16.42

Table 1.2: Average percent decrease in IMSE after implementing XCR at various levels of contamination on the time (σ_η^2) and functional (σ_ζ^2) domains.

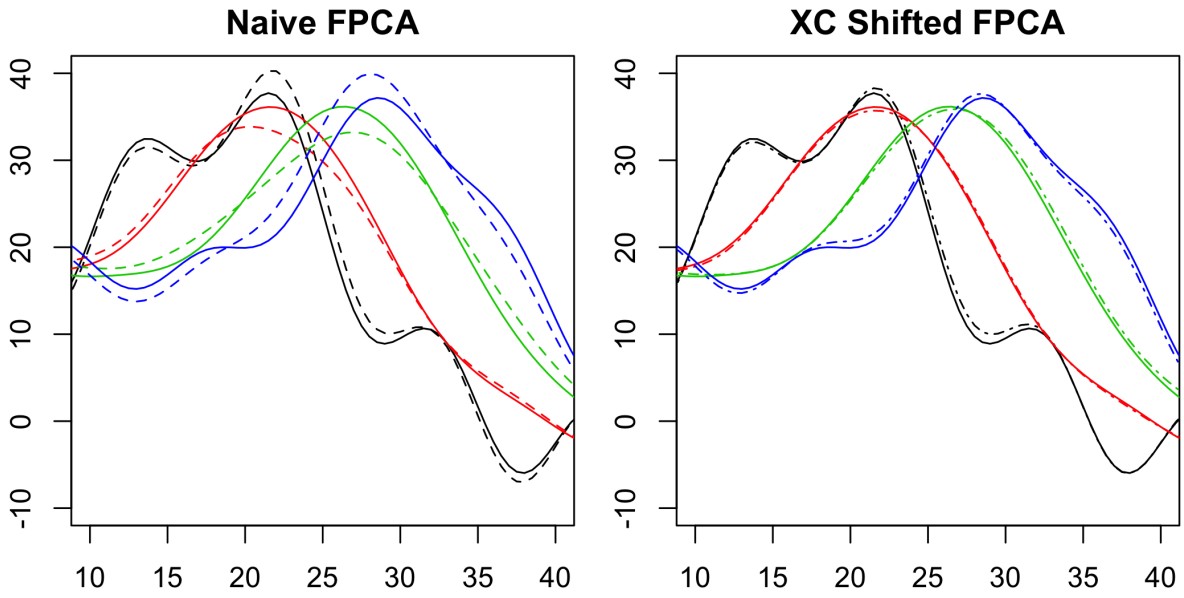

Figure 1.8: Differences in fit when using just naive FPCA (dashed, left) and FPCA and XCR together (dot-dashed, right), when $\sigma_\eta^2 = 0.25,\ \sigma_\zeta^2 = 25$. Solid lines represent the original data.

Noise Level	$\sigma_e^2 = 0$	$\sigma_e^2 = 9$	$\sigma_e^2 = 25$	$\sigma_e^2 = 64$	$\sigma_e^2 = 100$	component
$\sigma_\zeta^2 = 0$	0.0000	0.0010	0.0025	0.0036	0.0061	θ_1
	0.0000	0.0010	0.0038	0.0045	0.0066	θ_2
	0.0000	0.0012	0.0030	0.0040	0.0064	θ_3
	0.0000	0.0016	0.0026	0.0035	0.0060	θ_4
$\sigma_\zeta^2 = 9$	0.0004	0.0015	0.0034	0.0068	0.0051	θ_1
	0.0005	0.0016	0.0043	0.0072	0.0082	θ_2
	0.0009	0.0019	0.0038	0.0112	0.0062	θ_3
	0.0008	0.0019	0.0039	0.0109	0.0088	θ_4
$\sigma_\zeta^2 = 25$	0.0033	0.0035	0.0079	0.0037	0.0068	θ_1
	0.0042	0.0057	0.0070	0.0044	0.0069	θ_2
	0.0038	0.0059	0.0063	0.0055	0.0072	θ_3
	0.0036	0.0065	0.0087	0.0052	0.0076	θ_4
$\sigma_\zeta^2 = 64$	0.0018	0.0032	0.0035	0.0049	0.0091	θ_1
	0.0029	0.0037	0.0042	0.0055	0.0093	θ_2
	0.0031	0.0031	0.0053	0.0056	0.0085	θ_3
	0.0028	0.0032	0.0037	0.0054	0.0089	θ_4
$\sigma_\zeta^2 = 100$	0.0056	0.0069	0.0049	0.0122	0.0098	θ_1
	0.0080	0.0071	0.0094	0.0130	0.0125	θ_2
	0.0092	0.0081	0.0083	0.0121	0.0108	θ_3
	0.0052	0.0063	0.0100	0.0106	0.0123	θ_4

Table 1.3: MSEs of XC shift estimates under various noise settings. Estimates remain unbiased throughout.

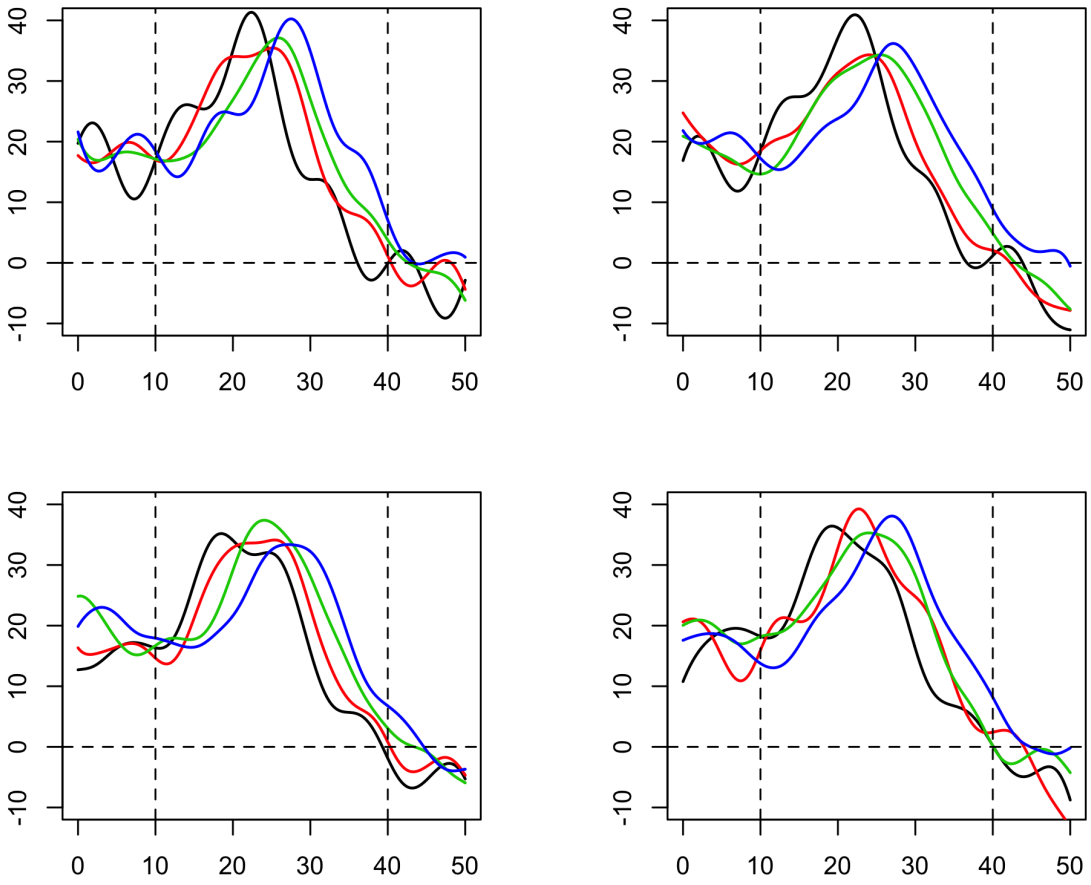

Figure 1.9: Four randomly selected simulated processes with under noise setting $\sigma_\zeta^2 = 25$, $\sigma_e^2 = 25$. The subinterval of integration $\mathcal{I} = [10, 40]$ is marked by dashed vertical lines.

1.7 Theoretical Results

For bivariate Cross-Component Registration, a key finding is that the centered process

$$Z_n(\tau) = \sqrt{n}\{L_n(\tau) - \Lambda(\tau)\}$$

converges weakly to a Gaussian limit process $Z(\tau)$, where L_n, Λ are as in (1.4), (1.5). The details of this result are shown in Lemma 1 of the Technical Proofs. To show weak convergence of the pairwise estimate $\widehat{\tau}$ as defined in (1.6), we require the following assumptions on Λ.

(P1) For any $\varepsilon > 0$, $\inf_{\tau: d(\tau,\tau_0)>\varepsilon} \Lambda(\tau) < \Lambda(\tau_0)$.

(P2) There exists $\eta > 0$, $C > 0$ and $\beta > 1$, such that, when $d(\tau, \tau_0) < \eta$, we have

$$\Lambda(\tau) - \Lambda(\tau_0) \geq C d(\tau, \tau_0)^{\beta}.$$

Assumption (P1) ensures that there exists a well-defined minimum, and assumption (P2) describes the local curvature of Λ at the true minimum τ_0. We also require the following assumptions for the observed random processes.

(A1) $X_j(t)$ is continuously twice differentiable for $j = 1, \ldots, p$,

(A2) $E\{\int_{\mathcal{I}} X_j^4(t)dt\} < \infty$, for $j = 1, \ldots, p$,

(A3) $E\{\int_{\mathcal{I}} X_j'^4(t)dt\} < \infty$, for $j = 1, \ldots, p$.

These assumptions are standard in the literature. They were, for example, previously stipulated in Hall & Horowitz (2007) and enable us to obtain asymptotic covariance matrices for our estimates and to derive some crucial bounds.

Theorem 1. *In the bivariate case, under assumptions (P1)-(P2), and (A1)-(A3), we have*

$$\widehat{\tau} - \tau_0 = \mathcal{O}_P(n^{-1/2(\beta-1)}).$$

In particular, when $\beta = 2$, the sequence $\sqrt{n}(\widehat{\tau} - \tau_0)$ is asymptotically normal with mean zero and variance $V = 4 \int_{\mathcal{I}} E[\{X_1(t) - X_2(t - \tau_0)\}X_2^{'}(t - \tau_0)]^2 dt / \{\Lambda^{''}(\tau_0)\}^2$, where $\Lambda(\tau) = E \int_{\mathcal{I}} \{X_1(t) - X_2(t - \tau)\}^2 dt$.

The proof is in Section 1.9 and utilizes results for M-estimators (Jain & Marcus, 1975; Van der Vaart & Wellner, 1996; van der Vaart, 1998). We note that when the local geometry around the minimum has a quadratic curvature, i.e. when $\beta = 2$, one obtains the parametric rate $n^{1/2}$.

Our main result for general Cross-Component Registration concerns the rate of convergence of the estimated global shift vector and its asymptotic distribution, as follows.

Theorem 2. *In the general case, under assumptions (P1)-(P2) and (A1)-(A3)*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_P(n^{-1/2(\beta-1)}).$$

In particular, when $\beta = 2$, the sequence $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma}_p = \frac{1}{p^2} \boldsymbol{A}^T \begin{bmatrix} \boldsymbol{V}_{\boldsymbol{\tau}_0}^{-1} E\left(\nabla \boldsymbol{m}_{\boldsymbol{\tau}_0} \nabla \boldsymbol{m}_{\boldsymbol{\tau}_0}^T\right) \boldsymbol{V}_{\boldsymbol{\tau}_0}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{A},$$

where $\boldsymbol{m}_{\tau_0} = \{L_n(\tau_{12}), L_n(\tau_{13}), \ldots, L_n(\tau_{(p-1)p})\}^T$ and $\boldsymbol{V}_{\tau_0}$ is the Hessian of $\boldsymbol{\Lambda}(\boldsymbol{\tau}) = E(\boldsymbol{m}_{\boldsymbol{\tau}})$ evaluated at $\boldsymbol{\tau}_0$.

1.8 Discussion

Cross-component registration seeks to address mutual component time warping that is often an issue for multivariate functional data arising from longitudinal studies in the biosciences. This issue does not manifest itself for univariate functional data. By focusing on time warping across components, and not on the traditional time warping between individual subjects, we are able to estimate

population-wide time shift parameters with fast parametric rates of convergence and obtain a limit distribution under suitable assumptions.

This new cross-component time warping approach leads to insights about the relative timings of the component processes, which is of interest for the analysis of growth data and also other multivariate longitudinal data. After cross-component shift warps have been identified and incorporated into the model, common methods such as functional principal component analysis for multivariate processes can be expected to lead to more meaningful outputs and the resulting principal component scores can be used for subsequent downstream analysis. The identification and estimation of the underlying latent process may also lead to a more parsimonious representation and is of interest in itself.

There are limitations of the framework we have established here. While the shift-warping model we develop in this chapter is appropriate for certain applications such as the Zürich Longitudinal Growth Study, the cross-component warping phenomena need not be restricted to shifts in general and may emerge in the form of non-linear distortions among components. Using the shift-warping methodology in such a situation may or may not yield satisfactory results, depending on the nature of the actual time warping. If the warp has a simple structure, a shift parameter may be a sufficient and parsimonious way to discover and approximate the component time relations, especially for practitioners who seek clear and concise interpretations. However, the situation for more pronounced or complicated warps is less auspicious. When the data at hand exhibit complex component warping beyond shifts, a more flexible warping paradigm should be adopted. The nonlinearity of such cross-component distortions may suggest that such problems warrant an alternative metric to the $\mathcal{L}^2$-norm.

In spite of this, we argue that the limitations of a shift-warping model are not necessarily tied to the general idea of cross-component registration which we

have presented here. While in this chapter we have used a shift-warping model to introduce the notion of cross-component registration, one can imagine more flexible extensions. The study of nonlinear warping models in cross-component registration is left for future research. Other potential directions of interest concern alternative representations of the cross-component warping problem.

1.9 Technical Proofs

Lemma 1. *Under assumptions (P1), (P2), and (A1)-(A3), it holds that*

$$Z_n(\tau) \rightsquigarrow Z(\tau),$$

where $\rightsquigarrow$ *denotes weak convergence and* $Z(\tau)$ *is a Gaussian process with mean zero and covariance* $G(\tau_1, \tau_2) = \int_{\mathcal{I}} \int_{\mathcal{I}} E\left[\{X_1(t) - X_2(t-\tau_1)\}^2 \{X_1(t) - X_2(t-\tau_2)\}^2\right] dtds - \Lambda(\tau_1)\Lambda(\tau_2)$.

Proof of Lemma 1.

In the following we consider compact intervals $\mathcal{T}$ and $\mathcal{I}$ such that $\mathcal{I} \subset \mathcal{T}$ where the functions are observed on $\mathcal{T}$ and $\mathcal{I}$ is the subinterval on which we take $\mathcal{L}^2$-distance of shifted curves. We also use C to represent a generic constant. We first establish a Central Limit Theorem for $Z_n(\tau) = \sqrt{n}\{L_n(\tau) - \Lambda(\tau)\}$ by applying Jain & Marcus (1975), so the proof reduces to verifying the following conditions:

(i.) $E\left[f\{Z_n(\tau)\}\right] = 0$ for all continuous linear functionals f.

(ii.) $\sup_\tau E\{Z_n(\tau)\} < \infty$.

(iii.) There exists a non-negative random variable M with finite variance such that
$|Z_n(\tau_1, \omega) - Z_n(\tau_2, \omega)| \leq M|\tau_1 - \tau_2|$.

(iv.) On the compact interval $\mathcal{T}$, $\int_{\mathcal{T}} \sqrt{\log N(\mathcal{T}, \varepsilon)} d\varepsilon < \infty$, where $N(\mathcal{T}, \varepsilon)$ is the covering number of $\mathcal{T}$.

Verifying these conditions,

(i.) follows as $E\left[f\{Z_n(\tau)\}\right] = f\left[E\{Z_n(\tau)\}\right] = f(0) = 0$.

(ii.) Since $Z_n(\tau)$ is centered,

$$E\{Z_n(\tau)\} = \text{Var}\left[\int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau)\}^2 dt\right]$$
$$\leq E\left(\left[\int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau)\}^2 dt\right]^2\right).$$

We examine the argument inside the square in more detail:

$$\begin{aligned}\int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau)\}^2 dt &\leq \left[\{\int_{\mathcal{I}} X_{i1}^2(t)dt\}^{1/2} + \{\int_{\mathcal{I}} X_{i2}^2(t-\tau)dt\}^{1/2}\right]^2 \\ &= (U_i + V_i)^2, \text{ so} \\ E\{Z_n^2(\tau)\} = E\{(U_i+V_i)^4\} &\leq C\{E(U_i^4) + E(V_i^4)\}.\end{aligned}$$

This is finite since

$$E(U_i^4) \leq \int_{\mathcal{I}} E\{X_{i1}^4(t)\}dt \leq E\left[\{\int_{\mathcal{I}} X_{i1}^2(t)dt\}^2\right] < \infty,$$

The same argument applies for $E(V_i^4)$. Then,

$$\sup_\tau E\{Z_n(\tau)\} \leq \sup_\tau C\left(\int_{\mathcal{I}} E\sup_t\{X_{i1}^4(t)\}dt + \int_{\mathcal{I}} E\sup_t\{X_{i2}^4(t-\tau)\}dt\right) < \infty,$$

where the right hand terms are finite under (A2).

(iii.) Consider differences of the Z_n with different arguments τ. Defining

$$D_i(\tau) := \int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau)\}^2 - E\left[\{X_{i1}(t) - X_{i2}(t-\tau)\}^2\right] dt$$

and applying the mean value theorem,

$$\begin{aligned}&|Z_n(\tau_1,\omega) - Z_n(\tau_2,\omega)| = \\ &\quad\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n \int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau_1)\}^2 dt - E\int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau_1)\}^2 dt\right. \\ &\quad\left. - \left[\frac{1}{n}\sum_{i=1}^n \int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau_2)\}^2 dt - E\int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau_2)\}^2 dt\right]\right|,\end{aligned}$$

$$|Z_n(\tau_1,\omega) - Z_n(\tau_2,\omega)| = \frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\{D_i(\tau_1) - D_i(\tau_2)\}\right|$$
$$\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|D_i'(\xi_i)\right||\tau_1-\tau_2| = |\tau_1-\tau_2|\,T_n$$

with $T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|D_i'(\xi_i)\right|$.

Now, due to the continuity of D and its derivatives in τ and t,

$$\begin{aligned}\frac{\partial}{\partial\xi_i}D_i(\xi_i) &= \int_{\mathcal{I}}\frac{\partial}{\partial\xi_i}\left(\{X_{i1}(t) - X_{i2}(t-\xi_i)\}^2 - E\left[\{X_{i1}(t) - X_{i2}(t-\xi_i)\}^2\right]\right)dt \\ &= 2\int_{\mathcal{I}}\Bigg(E\left[\{X_{i1}(t) - X_{i2}(t-\xi_i)\}X_{i2}'(t-\xi_i)\right] \\ &\qquad - \{X_{i1}(t) - X_{i2}(t-\xi_i)\}X_{i2}'(t-\xi_i)\Bigg)dt \\ &= 2\int_{\mathcal{I}}E[B_i(\xi_i,t)] - B_i(\xi_i,t)dt\end{aligned}$$

where $B_i(\xi_i,t) = \{X_{i1}(t) - X_{i2}(t-\xi_i)\}X_{i2}'(t-\xi_i)$.

Then,

$$\begin{aligned}\frac{1}{4}E\left[\{D_i'(\xi_i)\}^2\right] &= E\left(\left[\int_{\mathcal{I}}E\{B_i(\xi_i,t)\} - B_i(\xi_i,t)dt\right]^2\right) \\ &= \int_{\mathcal{I}}\int_{\mathcal{I}}\mathrm{Cov}\{B_i(\xi_i,u), B_i(\xi_i,v)\}dudv \leq \int_{\mathcal{I}}\mathrm{Var}\{B_i(\xi_i,t)\}dt\end{aligned}$$

Next, we bound this variance,

$$\begin{aligned}\mathrm{Var}\{B_i(\xi_i,t)\} \leq E\{B_i^2(\xi_i,t)\} &= E\left[\{X_{i1}(t) - X_{i2}(t-\xi_i)\}^2\{X_{i2}'(t-\xi_i)\}^2\right] \\ &\leq \left(C\left[E\{X_{i1}^4(t)\} + E\{X_{i2}^4(t-\xi_i)\}\right]E\{X_{i2}'(t-\xi_i)\}^4\right)^{1/2}\end{aligned}$$

so that

$$\begin{aligned}E(T_n^2) &\leq 4\sup_t\int_{\mathcal{I}}\mathrm{Var}\{B_i(\xi_i,t)\}dt \\ &\leq 4\left(C\left[E\{\int_{\mathcal{I}}X_{i1}^4(t)dt\} + E\{\int_{\mathcal{I}}X_{i2}^4(t-\xi_i)dt\}\right]E\left[\int_{\mathcal{I}}\{X_{i2}'(t-\xi_i)\}^4dt\right]\right)^{1/2} < \infty\end{aligned}$$

since each of the terms in the last line are finite by the integral conditions in (A2) and (A3).

(iv.) Is trivially satisfied.

With all 4 criterions checked, we apply the CLT of Jain & Marcus (1975) to $Z_n(\tau)$ and have the result.

Proof of Theorem 1. Our next result relies on Lemma 1 of Jain & Marcus (1975), which implies that

$$Z_n(\tau) = \sqrt{n}\{L_n(\tau) - \Lambda(\tau)\} \xrightarrow{D} Z(\tau), \text{ whence}$$
$$\sup_\tau |\sqrt{n}\{L_n(\tau) - \Lambda(\tau)\}| = \sup_\tau |Z_n(\tau)| \xrightarrow{p} \sup_\tau |Z(\tau)| = \mathcal{O}_p(1), \text{ and therefore}$$
$$\sup_\tau |L_n(\tau) - \Lambda(\tau)| = o_P(1) \tag{1.1}$$

From (1.1) and Theorem 3.2.3 in Van der Vaart & Wellner (1996), we then have

$$\widehat{\tau} - \tau_0 = o_P(1).$$

For the next part of the proof we consider the process $V_n(\tau) := L_n(\tau) - \Lambda(\tau)$. Then,

$$\begin{aligned} |V_n(\tau, \omega) - V_n(\tau_0, \omega)| &= |L(\tau) - \Lambda(\tau) - \{L(\tau_0) - \Lambda(\tau_0)\}| \\ &= |L(\tau) - L(\tau_0) - \{\Lambda(\tau) - \Lambda(\tau_0)\}| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} |W_i(\tau) - EW_i(\tau)| \end{aligned}$$

where $W_i(\tau) := \int_{\mathcal{I}} \{X_{i1}(t) - X_{i2}(t-\tau)\}^2 - (X_{i1}(t) - X_{i2}(t-\tau_0))^2 dt$. To control this uniformly over small $d(\tau, \tau_0)$, we define the function g as $g(\tau) = V_n(\tau) = L_n(\tau) - \Lambda(\tau)$ and the function class $\mathcal{M}_\delta := g(\tau) - g(\tau_0) : d(\tau, \tau_0) < \delta$. An envelope function for $\mathcal{M}_\delta$ is $G(\delta) = 2|\mathcal{T}|\delta$ and $E(G^2(\delta)) = \mathcal{O}(\delta^2)$. Define $J(\delta) = \int_0^1 \sqrt{1 + \log N(\varepsilon, \mathcal{T}, \delta)} d\varepsilon$ so $J(\delta) < \infty$. Then Theorems 2.7.11 and 2.14.2 of Van der Vaart and Wellner (1996) imply that for small enough δ,

$$E\left(\sup_{\tau: d(\tau,\tau_0)<\delta} |\sum_{i=1}^{n} W_i(\tau) - EW_i(\tau)| \right) \leq \frac{J(\delta) E[\{G^2(\delta)\}^{1/2}]}{\sqrt{n}} = \mathcal{O}(\delta n^{-1/2})$$

which implies

$$E\Big(\sup_{\tau:d(\tau,\tau_0)<\delta}|V_n(\tau)-V_n(\tau_0)|\Big)=\mathcal{O}(\delta n^{-1/2})$$

Finally, let $r_n = n^{\frac{\beta}{4(\beta-1)}}$ and

$$S_{j,n}=\{\tau : 2^{j-1} < r_n d(\tau,\tau_0)^{\beta/2} < 2^j\}.$$

Choose $\eta > 0$ to satisfy (P2) and also small enough that (P1) holds for all $\delta < \eta$ and set $\tilde{\eta} := \eta^{\beta/2}$. For any integer L,

$$P\{r_n d(\tau,\tau_0)^{\beta/2} > 2^L\} \le P\{d(\tau,\tau_0)\ge\eta\} + \sum_{\substack{j>L\\ 2^j<r_n\tilde{\eta}}} P\Big(\sup_{\tau\in S_{j,n}}|V_n(\tau)-V_n(\tau_0)|\ge C\frac{2^{2(j-1)}}{r_n^2}\Big).$$

For each j in the sum on the right hand side we have $d(\tau,\tau_0)\le\left(\frac{2^j}{r_n}\right)^{2/\beta}\le\eta$, so this sum is bounded by

$$\sum_{\substack{j>L\\ 2^j<r_n\tilde{\eta}}}\frac{2^{2j(1-\beta)/\beta}}{r_n^{2(1-\beta)/\beta}\sqrt{n}}\le\sum_{j>L}\left(\frac{1}{4^{(\beta-1)/\beta}}\right)^j.$$

Since $\beta > 1$, the last series converges and therefore the original probability can be made arbitrarily small by choosing a large enough L. This proves the desired result that $d(\widehat{\tau},\tau_0) = \mathcal{O}_P(r_n^{-2/\beta}) = \mathcal{O}_P(n^{-1/2(\beta-1)})$. Finally, asymptotic normality follows from Theorem 5.23 of Van der Vaart & Wellner (1996), where the Lipschitz condition and bounded second moment have been shown already in parts (ii.) and (iii.) of this proof.

Proof of Theorem 2: For convergence of $\widehat{\boldsymbol{\theta}}$, we recall that $\widehat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\widehat{\boldsymbol{\tau}}^*$,

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 & -1 \\ 0 & 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 \end{pmatrix}.$$

where the first $p-1$ rows contain all pairwise contrasts of the first column with the rest, the next $p-2$ rows contain all pairwise contrasts of the second column with the rest, and so on, until the second to last row contains the final pairwise contrast. The last line contains all 1's to represent the constraint that $\sum_{j=1}^{p}\theta_j = 0$. We observe that $\boldsymbol{A}^T\boldsymbol{A} = p\boldsymbol{I}_p$, which can be verified with a simple matrix multiplication. It then immediately follows that $\widehat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\widehat{\boldsymbol{\tau}}^* = \frac{1}{p}\boldsymbol{A}\widehat{\boldsymbol{\tau}}^* = \boldsymbol{B}\widehat{\boldsymbol{\tau}}^*$ and since the linear mapping induced by the matrix $\boldsymbol{B} = \frac{1}{p}\boldsymbol{A}^T$ is continuous, we can apply the continuous mapping theorem and observe that

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_P(1).$$

The proof for the rate of convergence follows the same arguments as those for $\widehat{\boldsymbol{\tau}}$. The asymptotic normality of $\widehat{\boldsymbol{\theta}}$ when $\beta = 2$ requires a closer examination. When estimating $\widehat{\boldsymbol{\tau}}^*$, we are estimating several pairwise XC shifts at the same time. This is done by minimizing several different $\mathcal{L}^2$ distances, which we stack in a vector,

$\boldsymbol{m}_{\tau_0}(X_1, \ldots, X_p)$, defined as:

$$\boldsymbol{m}_{\tau_0} = \boldsymbol{m}_{\tau_0}(X_1, \ldots, X_p) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{i2}(t-\tau_{12})\}^2 dt \\ \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{i3}(t-\tau_{13})\}^2 dt \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{ip}(t-\tau_{1p})\}^2 dt \\ \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i2}(t) - X_{i3}(t-\tau_{23})\}^2 dt \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i2}(t) - X_{ip}(t-\tau_{2p})\}^2 dt \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i(p-1)}(t) - X_{ip}(t-\tau_{(p-1)p})\}^2 dt \end{bmatrix}$$

By theorem 5.23 of Van der Vaart & Wellner (1996), then, we have

$$\sqrt{n}(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{V})$$

where $\boldsymbol{V} = \boldsymbol{V}_{\tau_0}^{-1} E\left(\nabla \boldsymbol{m}_{\tau_0} \nabla \boldsymbol{m}_{\tau_0}^T\right) \boldsymbol{V}_{\tau_0}^{-1}$ and $\nabla \boldsymbol{m}_{\tau_0}$ represents the gradient of $\boldsymbol{m}_\tau$ evaluated at $\boldsymbol{\tau}_0$, i.e.,

$$\nabla \boldsymbol{m}_{\tau_0} = \begin{bmatrix} \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{i2}(t-\tau_{12})\}X'_{i2}(t-\tau_{12})dt \\ \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{i3}(t-\tau_{13})\}X'_{i3}(t-\tau_{13})dt \\ \vdots \\ \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i1}(t) - X_{ip}(t-\tau_{1p})\}X'_{ip}(t-\tau_{1p})dt \\ \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i2}(t) - X_{i3}(t-\tau_{23})\}X'_{i3}(t-\tau_{23})dt \\ \vdots \\ \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i2}(t) - X_{ip}(t-\tau_{2p})\}X'_{ip}(t-\tau_{2p})dt \\ \vdots \\ \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{I}}\{X_{i(p-1)}(t) - X_{ip}(t-\tau_{34})\}X'_{ip}(t-\tau_{(p-1)p})dt \end{bmatrix}$$

and $\boldsymbol{V}_{\tau_0}$ is the Hessian of $\boldsymbol{\Lambda}(\tau) = E(\boldsymbol{m}_\tau)$ at τ_0. Finally, we apply the linear transformation $\widehat{\boldsymbol{\theta}} = \frac{1}{p}\boldsymbol{A}\widehat{\boldsymbol{\tau}}^*$ to obtain the result:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_p)$$

where $\boldsymbol{\Sigma}_p = \frac{1}{p^2}\boldsymbol{A}^T \begin{bmatrix} \boldsymbol{V}_{\tau_0}^{-1} E\left(\nabla \boldsymbol{m}_{\tau_0} \nabla \boldsymbol{m}_{\tau_0}^T\right) \boldsymbol{V}_{\tau_0}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{A}$.

Chapter 2

Latent Transport Models for Multivariate Functional Data

2.1 Introduction

The increased prevalence of complicated and high-dimensional data in recent years has brought along with it the need for improved statistical techniques to guide practical data analyses. One such type of data with complex structure is functional data, in which observations are functions $X_i(t),\ i = 1, \ldots, n$, where the index t (often representing time) ranges over some continuum $\mathcal{T} \subset \mathbb{R}$. Functional data analysis (FDA) has found applicability in several fields of research (e.g. biology, ecology, economics) and as such has spawned considerable methodological work as a subfield of statistics (Ramsay & Silverman, 2005; Jane-Wang et al., 2016; Ferraty & Vieu, 2006). In particular, the analysis of univariate functional data has driven the majority of developments in this area and has been studied extensively in the contexts of classical statistical methods such as principal components analysis (Kleffe, 1973), regression (Cardot et al., 1999; Yao et al., 2005), and clustering (Jacques & Preda, 2014). In contrast, functional data which arises in a multivariate setting (that is, data in which observations consist of a finite dimen-

sional vector whose elements are functions) has been the focus of much fewer studies, despite its pervasiveness and increased relevance in emerging fields.

Dimension reduction is a common approach to handle functional data with several components, with many studies focusing on extending univariate functional principal components analysis to the multivariate case (Chiou et al., 2014; Happ & Greven, 2018). Alternative methods have considered decomposition into marginal component processes and their interactions (Chiou et al., 2016). The majority of methodological work has focused on traditional amplitude variation-based models for dimension reduction, though phase variation-based methods for multivariate functional data have garnered attention in recent years. Brunel & Park (2014) have proposed a method for estimating multivariate structural means and Park & Ahn (2017) introduced a model for clustering multivariate functional data in the presence of phase variation. Carroll et al. (2020b) combined the notions of dimension reduction and phase variability through a multivariate version of the shape-invariant model (Kneip & Gasser, 1995), in which component processes share a common latent structure which is time-shifted across components. However, the assumption of a rigid shift-warping framework imposes a major parametric constraint on the warping structure and often the class of models described with simple shifts is not rich enough to apply to real-world datasets. Our main contribution is a less-restrictive alternative, in which time characterization of individual-specific temporal effects and component-specific effects is achieved through a fully non-parametric deformation-based model.

A major motivation for this framework is the fact that in many contexts, the component functions of a multivariate data vector may share a common structure which is subject to variation across modalities; the fundamental shape of growth curves is similar but not identical in timing patterns across body parts, for instance. This common latent structure can be seen as a crucial component of

the data-generating mechanism for multivariate functional data. In conjunction with this latent curve, a shared shape-based model also requires characterization of individual- and component-level variation. The idea of component-specific time effects has been previously explored through a time-shifting model, albeit only in a restrictive parametric framework which does not carry widespread applicability in practice (Carroll et al., 2020b). Allowing for more flexible and nuanced component effects would increase the viability of this approach and allow for improved data fidelity when describing component-specific effects which inform the time-dynamics of a larger system at work. To this end, we introduce a representation of multivariate functional data which uses tools from time warping (Marron et al., 2015) and optimal transport theory (Villani, 2003).

The organization of this chapter is as follows. Section 2 discusses existing approaches to univariate curve registration and introduces the Latent Transport Model for component-warped multivariate functional data. We derive estimators of model components in Section 3 and illustrate the utility and performance of our methodology through data analysis and simulation studies in Sections 4 and 5, respectively. Lastly, asymptotic results are established in Section 6.

2.2 Curve Registration and The Latent Transport Model

2.2.1 The Univariate Curve Registration Problem

The classical univariate curve registration problem is characterized by the observation of a sample of curves $X_i(t),\ i = 1, \ldots, n$, observed on an interval T which are realizations of a fixed template $\xi(t)$ subject to variation in their time domains. This domain variation is characterized by the monotonic time-warping functions $h_i(t)$ which act as random homeomorphisms of T.

A classical model for this scenario can be written,

$$X_i(t) = (\xi \circ h_i)(t), \quad \text{for all} \quad t \in \mathcal{T}, \quad i = 1, \ldots, n. \tag{2.1}$$

The goal of curve registration is to estimate the distortions, h_i, which are typically considered nuisance effects, in order to account for them before proceeding with further analysis, e.g. estimation of ξ, functional principal component analysis, etc. This problem was first considered in the context of aligning spoken-word data (Sakoe & Chiba, 1978) which used a version of the dynamic time warping algorithm which Wang & Gasser (1997) later studied in a statistical context. A major branch of time-warping techniques is based on the idea of aligning processes to some reference curve which carries the main features in common across subjects. This reference curve is referred to as a template function and is employed by landmark-based registration methods (Kneip & Gasser, 1992, 1995), pairwise curve alignment (Tang & Müller, 2008) or the Procrustes approach (Ramsay & Li, 1998), among many others. For a comprehensive review and additional references, we refer to Marron et al. (2015).

While the curve registration literature is varied and rich in methodology, no single method has prevailed as a silver bullet in all warping contexts. Indeed, the debate over desirable properties of existing and future registration techniques continues in search of a gold-standard. With this in mind, we emphasize that our aim here is not to advocate for one alignment method over another, but rather extend the ideas present in univariate registration to a multivariate model which considers the presence of a composite warping function with fixed and random effects, which characterize the distortion of a template function across vector components as well as subjects. In practice, any suitable registration method may be employed in the estimation of the proposed Latent Transport Model (see Estimation).

2.2.2 A Unified Model for Multivariate Time Dynamics

Let $\{X_j\}_{j=1}^p$ denote a generic set of random functions with each component process X_j in $L_2(\mathcal{T})$ for an interval $\mathcal{T} = [T_1, T_2],\ T_1, T_2 \in \mathbb{R}$. Suppose further that each component is positive-valued, i.e. $X_j(t) > 0$ for all $t \in \mathcal{T}$, $j = 1, \ldots, p$; the assumption of positivity is made to make estimation of model components more straightforward. Without loss of generality we consider hereafter the unit domain case $\mathcal{T} = [0, 1]$ which can always be achieved by a simple shift and rescale of the original domain. Our terminology takes inspiration from the case where the domain is a time interval, though the general framework may be applied even when $\mathcal{T}$ is some other continuum. In terms of notation, we reserve the use of Greek letters for fixed, unknown population quantities, while Latin letters represent random, individual-specific ones.

The Latent Transport Model (LTM) is motivated by the situation when the functional forms of the component processes $X_j,\ j \in \{1, \ldots, p\}$ (or any subset thereof) exhibit a structural similarity among one another. This connection across variables suggests that information in one component may be harnessed to improve the modeling of another. Furthermore, information from all relevant components may be combined to estimate and analyze the mutual structure shared among them. Denoting a random sample from a p-dimensional stochastic process by $\{\boldsymbol{X}_i\}_{i=1}^n$, where $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{ip}(t))^T$, we model this shared structure through a latent curve λ, which characterizes the component curves through the relation

$$X_{ij}(G_{ij}^{-1}(t)) = A_{ij}\lambda(t), \quad i = 1, \ldots, n,\ j = 1, \ldots, p, \tag{2.2}$$

where λ is a fixed function, and the random amplitude factors A and random time distortion functions G reflect differences in realized curves across components and individuals. Without loss of generality we assume $\sup_{t \in \mathcal{T}} |\lambda(t)| = ||\lambda||_\infty = 1$ since it is always possible to rescale the latent curve without changing the

model by defining new amplitude factors $\tilde{A}_{ij} := A_{ij}||\lambda||_\infty$ and scaled curve $\tilde{\lambda}(t) = \lambda(t)/||\lambda||_\infty$.

The distortion functions G are elements of the convex space of all smooth, strictly increasing functions with common endpoints, i.e., $\mathcal{W} := \{g : \mathcal{T} \to \mathcal{T} \mid g \in C^2(\mathcal{T}),\ g(T_1) = T_1,\ g(T_2) = T_2,\ g \text{ is a strictly increasing homeomorphism}\}$. The elements of this space represent random homeomorphisms of the time domain and capture the presence of non-linear phase variation. We further assume that these functions G may be decomposed into the mixed-effects form:

$$G_{ij}(t) = (\Psi_j \circ H_i)(t), \quad i = 1, \ldots, n,\ j = 1, \ldots, p, \tag{2.3}$$

where the deterministic functions Ψ describe the component-based effects of time distortion and the random functions H describe the subject-level phase variation. In other words, this decomposition suggests that the mapping Ψ_j conveys the internal time scale of the j^{th} component, while H_i carries the internal time scale of the i^{th} subject. These mappings may be seen as transports from standard clock time, $id(t) \equiv t$, to the system time of a given component or individual. As such we refer to the collection of functions $\Psi = \{\Psi_j : j = 1, \ldots, p\}$ as *component-level transport functions* and the collection of functions $\mathcal{H} = \{H_i : i = 1, \ldots, n\}$ as *subject-level transport functions*.

The random warping functions H_i obey some law on the convex space $\mathcal{W}$ where EH_i^{-1} is assumed to exist. We further assume that they are distributed according to a law such that these functions induce no net distortion on average, i.e., $EH_i^{-1}(t) = t$ for $t \in \mathcal{T}$. This assumption is sometimes referred to as "standardizing" the registration procedure (Kneip & Ramsay, 2008). It is a mild assumption, since were it the case that $EH_i^{-1}(t) = h_0^{-1}(t),$ with $h_0^{-1} \neq id$, then a standardized registration procedure is given by reparameterizing the warping functions as $\tilde{H}_i = h_0^{-1} \circ H_i$ so that $E\tilde{H}_i^{-1}(t) = E(H_i^{-1} \circ h_0)(t) = t$. Component transport functions are also assumed to be standardized, but because they are

deterministic and not random, the assumption becomes $\frac{1}{p}\sum_{j=1}^{p}\Psi_j^{-1}(t) = t$ for $t \in \mathcal{T}$. Together these conditions imply $E(\frac{1}{p}\sum_{j=1}^{p}G_{ij}^{-1}(t)) = t$ so that there is no net distortion from the latent curve λ across components or individuals.

Combining (2.1) and (2.2) yields the Latent Transport Model for multivariate functional data:

$$X_{ij}(t) = A_{ij}\left(\lambda \circ \Psi_j \circ H_i\right)(t), \quad i = 1, \ldots, n,\ j = 1, \ldots, p. \tag{2.4}$$

In practice, it may be useful to pose the model in an equivalent form, defining the component-warped versions of the latent curve as $\gamma_j = \lambda \circ \Psi_j$ so that

$$X_{ij}(t) = A_{ij}\left(\gamma_j \circ H_i\right)(t), \quad i = 1, \ldots, n,\ j = 1, \ldots, p. \tag{2.5}$$

In this form, the curves $\gamma_j(t)$ convey the "typical" time progression of the latent curve according to the j^{th} component's system time, so we refer to this composition as the j^{th} *component tempo function*. The component tempo functions can be viewed as the synchronized processes for each component after accounting for random subject-level time distortions. Figure 2.1 provides a schematic of the data generating mechanism assumed by the LTM for a simulated dataset.

2.2.3 Cross-Component Transport Maps

Marginal Cross-Component Transports

To understand and quantify the relative timings between any pair of components, $j, k \in \{1, \ldots, p\}$, it is useful to define their *cross-component transport* T_{jk}, which is the transport that, when applied to j^{th} component, maps its tempo to that of the k^{th} component. Explicitly, we write,

$$T_{jk} = \Psi_j^{-1} \circ \Psi_k, \tag{2.6}$$

so that $\gamma_j(T_{jk}) = \lambda \circ \Psi_j \circ \Psi_j^{-1} \circ \Psi_k = \lambda \circ \Psi_k = \gamma_k$. Because the component transports Ψ_k can be represented as distribution functions and are closed under composi-

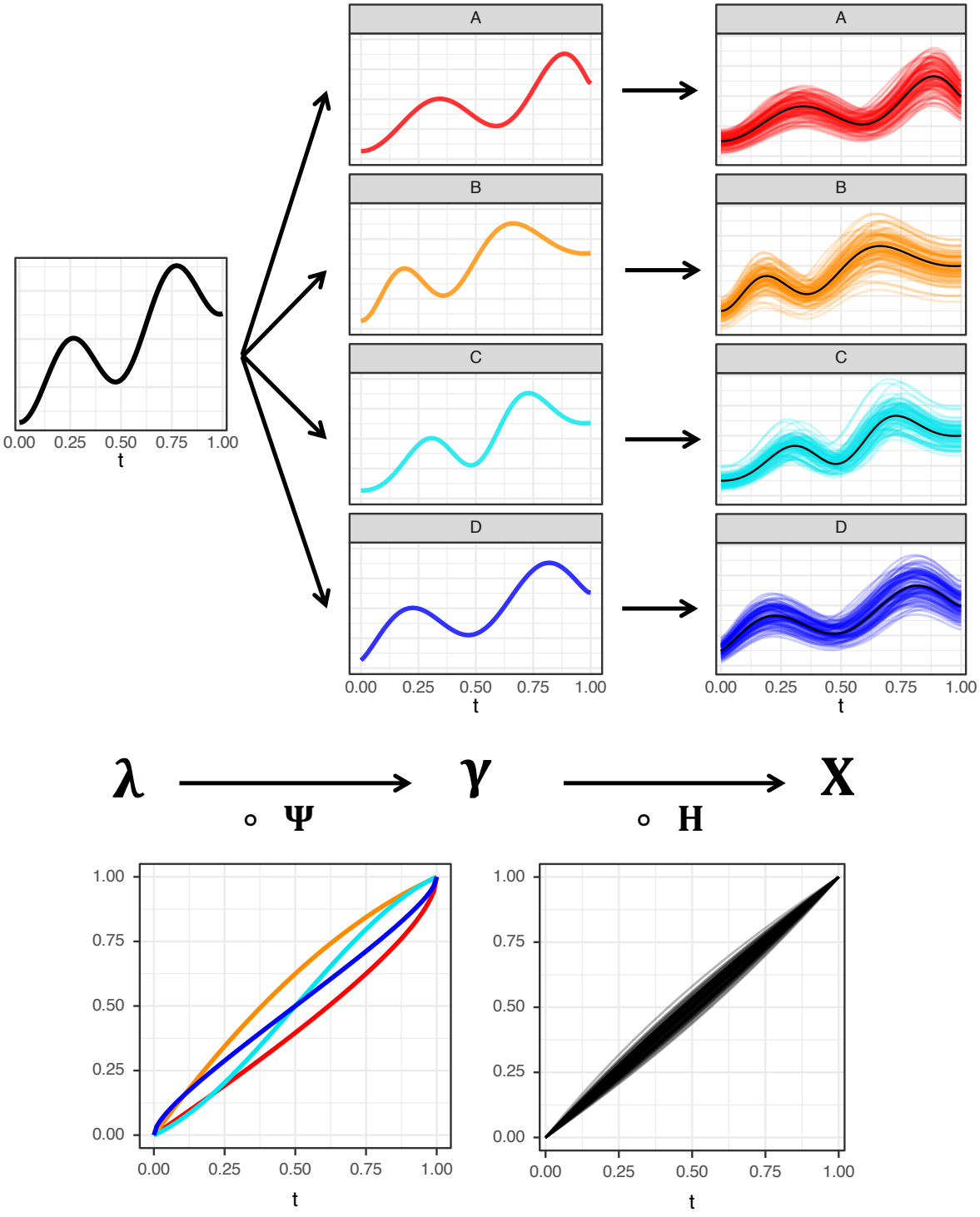

Figure 2.1: Schematic of the Latent Transport Model, where λ denotes the latent base curve (top-left), Ψ denotes component transports (bottom-left), γ denotes component tempos (top-center), H denotes random subject-wise time distortion functions (bottom-right), and X denotes the observed multivariate curve data (top-right) resulting from the complete data generating mechanism.

tion, the cross-component transport (XCT) may also be represented as a distribution function and is interpreted similarly to an ordinary component tempo. While the component tempo Ψ_k expresses the k^{th} component's timing patterns in terms of clock time, the cross-component transport T_{jk} expresses the same patterns relative to the tempo of the j^{th} component.

For example, consider a pair of component processes, Component A and Component B, for which Component A tends to lag behind the latent curve, while the Component B precedes it. An example of this can be seen in the red and orange curves, respectively, in Figure 2.1. The corresponding red transport, Ψ_A, falls below the diagonal and conveys the lagged tempo, while the orange transport, Ψ_B lies above the diagonal and expresses an accelerated system time. The transport function T_{AB} then sits above the diagonal and represents the time-acceleration needed to bring the red tempo in line with the orange component.

Subject-Level Cross-Component Transports

While the marginal XCTs describe the general time relations between components on a population level, we may also be interested to see how an individual's component processes relate to one another. This perspective may be especially useful when trying to understand intercomponent dynamics which are mediated by covariate effects. Conceptually it is straightforward to extend the notion of cross-component transports to individuals by searching for the warping function $T_{jk}^{(i)}$ which brings the i^{th} individual's j^{th} component in line with the k^{th}. A natural definition under the LTM is then

$$T_{jk}^{(i)} = G_{ij}^{-1} \circ G_{ik}, \tag{2.7}$$

since this choice gives $X_{ij} \circ T_{jk}^{(i)} \propto A_{ij}(\lambda \circ G_{ij} \circ G_{ij}^{-1} \circ G_{ik}) \propto (\lambda \circ G_{ik}) \propto X_{ik}$. In practice, this proportionality will become equality once random amplitude factors are dealt with during estimation.

The proposed cross-component transport acts as a flexible, non-linear tool for representing intercomponent time dynamics and does not rely on an absolute clock time. Statistics based on the XCT model can be used in downstream analyses like hypothesis testing and regression. Several data illustrations are given in the applications of Section 2.4

2.3 Model Estimation and Curve Reconstruction

2.3.1 Pairwise Warping

The model estimation procedure proposed here relies on solving several univariate warping problems of type (1). Any of the traditional methods described in Section 2 may be used to solve these problems in practice. For the sake of developing convergence rates for our model, we illustrate the estimation using the pairwise alignment method of Tang & Müller (2008) which provides an explicit representation of warping functions and is briefly outlined using the notation of the LTM as follows. Throughout our applications, the pairwise warping algorithm was implemented using the package `fdapace` (Carroll et al., 2020a). Similar rates of convergence are anticipated if one adopts an alternative warping algorithm.

Fix a component j and to simplify notation write $U_i = X_{ij}/||X_{ij}||_\infty$ to denote a curve normalized by its maximum. For any two curves $U_i, U_{i'},\ i, i' \in \{1, \ldots, n\}$, the pairwise warping function $V_{i'i} : \mathcal{T} \to \mathcal{T}$ is the homeomorphism of the time domain which aligns $U_{i'}$ to U_i and defined by

$$V_{i'i}(t) = H_{i'}^{-1}(H_i(t)), \quad \text{for all } t \in \mathcal{T}. \tag{2.8}$$

We assume that a linear spline representation may parametrize the approximate warping functions. Let $t_\ell = \ell/(L+1)$, for $\ell = 1, \ldots, L$, denote L equidistant knots over $\mathcal{T}$ with $t_0 = 0$ and $t_{L+1} = 1$. For any function v, we define $\theta = [v(t_1), \ldots, v(t_\ell)]^T$

so that the piecewise linear approximation of v can be written

$$v(t) = \theta^T \alpha(t), \quad \text{for all } t \in \mathcal{T}, \tag{2.9}$$

where $\alpha(t) = [\alpha_1(t), \ldots, \alpha_{L+1}(t)]^T$ and $\alpha_\ell(t)$ are linear basis functions defined as $\alpha_\ell(t) = \alpha_\ell^{(1)}(t) - \alpha_{\ell+1}^{(2)}(t)$ with $\alpha_\ell^{(1)}(t) = \dfrac{t - t_{\ell-1}}{t_\ell - t_{\ell-1}} \mathbf{1}_{[t_{\ell-1}, t_\ell)}$ and $\alpha_\ell^{(2)}(t) = \dfrac{t - t_\ell}{t_\ell - t_{\ell-1}} \mathbf{1}_{[t_{\ell-1}, t_\ell)}$, for $\ell = 1, .., L+1$. The parameter space Θ of the coefficient vector is

$$\Theta = \{\theta \in \mathbb{R}^{L+1} : 0 < \theta_1 < \cdots < \theta_{L+1} = 1\}$$

so that the monotonicity constraint is satisfied. Finally, we express the pairwise warping function between subjects i and i' as

$$V_{i'i}(\cdot) = \theta^T \alpha(\cdot), \qquad \text{with } \theta \in \Theta.$$

Then, the spline coefficient vector is estimated by solving the minimization problem,

$$\tilde{\theta}_{V_{i'i}} = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \mathscr{C}_\eta(\theta, U_i, U_{i'}), \tag{2.10}$$

$$\text{where } \mathscr{C}_\eta(\theta, U_i, U_{i'}) = \int_{\mathcal{T}} d^2\left(U_{i'}(\theta^T \alpha(t)), U_i(t)\right) dt + \eta \int_{\mathcal{T}} (\theta^T \alpha(t) - t)^2 dt, \tag{2.11}$$

where d is a distance well-suited for the type of alignment desired and η is a regularization parameter which penalizes the magnitude of warping. In the original pairwise warping framework of Tang & Müller (2008), the default distance was taken as $d(f(t), g(t)) = (f - g)(t)$. With this choice of alignment criteria, the penalty parameter plays an important role in avoiding excessive warping, in particular the so-called "pinching effect," in which curves are squeezed together in an exaggerated fashion to minimize $\mathcal{L}^2-$distance, but are not actually well-aligned (Marron et al., 2015). The choice of an appropriate alignment criteria is specific to the data application in question; other potential choices include the semi-norm $d_r(f(t), g(t)) = \left| \dfrac{f(t)}{||f||_r} - \dfrac{g(t)}{||g||_r} \right|$, where $||\cdot||_r$ is the $\mathcal{L}^r-$norm for

proportional alignment (Marron et al., 2015). The semi-norm approach aims at reducing the effect of amplitude variation across subjects; note that had we not normalized our curves in the initial formulation of the problem, we could have equivalently worked with the original curves and chosen the modified semi-norm $d_\infty(f(t), g(t)) = \left| \frac{f(t)}{||f||_\infty} - \frac{g(t)}{||g||_\infty} \right|$.

Once $\tilde{\theta}_{V_{i'i}}$ is obtained, an estimate of the pairwise warping problem immediately follows as

$$\tilde{V}_{i'i}(t) = \tilde{\theta}^T_{V_{i'i}} \alpha(t) \quad \text{for all } t \in \mathcal{T}. \tag{2.12}$$

Since we assume $E(H_i^{-1}(t)) = t$ for all $t \in \mathcal{T}$, we have the identity $E[V_{i'i}(t)|H_i] = E[H_{i'}^{-1}(H_i(t))|H_i] = H_i(t)$, the empirical version of which leads to the estimate

$$\tilde{H}_i(t) = \frac{1}{n} \sum_{i'=1}^{n} \tilde{V}_{i'i}(t) \quad \text{for all } t \in \mathcal{T}. \tag{2.13}$$

2.3.2 Component-wise Alignment

With the notation of pairwise warping established, we now turn our attention to estimation of model components. We claim that under the LTM, each component $j = 1, \ldots, p$, yields their own univariate warping problem for H_i. To see this, consider for a fixed component j the sample of univariate curves, $S_j := \{X_{ij}\}_{i=1}^n$. By restricting our attention to S_j, we limit our scope to the estimation of a single component tempo γ_j at a time. For each collection of curves S_j, there exists a univariate warping problem based on the normalized curves $X^*_{ij} = X_{ij}/||X_{ij}||_\infty$. Estimation of γ_j and H_i for the j^{th} component is achieved through the model,

$$X^*_{ij}(t) = (\lambda \circ \Psi_j \circ H_i)(t) \tag{2.14}$$

which coincides with a warping framework of type (2.1) with $\xi = \lambda \circ \Psi_j$, and $h_i = H_i$. Replacing X by X^* in (2.14) is necessary in order to eliminate the random amplitude factors A_{ij}. To see this, note that since the random functions G_{ij} are

homeomorphisms, we have $||X_{ij}||_\infty = A_{ij}||\lambda \circ G_{ij}||_\infty = A_{ij}$. Thus the normalized curves $X^*_{ij}(t) = (\lambda \circ \Psi_j \circ H_i)(t)$ do not depend on the amplitude factors A_{ij}.

Applying an estimation method like pairwise warping for each of the subcollections $S_1, \ldots, S_p$, results in p estimates of the subject-level warping function, $\tilde{H}^{(1)}_i(t), \ldots, \tilde{H}^{(p)}_i(t)$. In each of these registration procedures the penalty parameter can either be taken as the default choice, $\eta_{1j} = 10^{-4} \times \{n^{-1} \sum_{i=1}^n \int_{\mathcal{T}} (X_{ij}(t) - \bar{X}_j(t))^2 dt\}$, as per Tang & Müller (2008) or chosen by cross-validation. Taking the mean of the resulting p warping functions gives an estimate for the subject-specific warp,

$$\hat{H}_i = p^{-1} \sum_{j=1}^{p} \tilde{H}^{(j)}_i, \quad i = 1, \ldots, n. \tag{2.15}$$

Under this definition, the associated overall penalty parameter is $\eta_1 := \sup_{1 \le j \le p} \eta_{1j}$. With an individual warping estimator in hand, a plug-in estimate of γ_j is obtained by averaging the component-aligned curves,

$$\hat{\gamma}_j = n^{-1} \sum_{i=1}^{n} (X_{ij} \circ \hat{H}^{-1}_i)/||X_{ij}||_\infty, \quad \text{for } j = 1, \ldots, p. \tag{2.16}$$

2.3.3 Global Alignment and Latent Curve Estimation

A central idea in the estimation of the LTM is the fact that *any* univariate curve X_{ij} contains information about the latent curve, regardless of its component. This motivates the adoption of a perspective in which we temporarily ignore the multivariate structure of the data and expand our scope to the full collection of curves, $S = \cup_{j=1}^p S_j$. For each individual i, select one of its component curves at random as a representative. Call this representative curve Z_i and denote its normalized counterpart by Z^*_i. Selecting one of the components at random ensures that we have $P(Z_i = X_{ij}) = 1/p$ for all $i = 1, \ldots, n,\ j = 1, \ldots, p$. The collection of curves $\{Z_i,\ i = 1, \ldots, n\}$ can be thought of as realizations of λ subject to some random distortion D_i, where $D_i = G_{ij}$ if the j^{th} component curve is selected. Define I_{ij} as the event that the curve Z_i comes from the collection of j^{th} component

curves, S_j. Conditional on the event I_{ij} (which happens with probability $1/p$ for all $i = 1, \ldots, n$), it follows that $D_i = G_{ij} = \Psi_k \circ H_i$. Then, on average there is no net warping from the latent curve, as

$$E[D_i^{-1}] = E\{E[D_i^{-1}|I_{ij}]\} = \sum_{j=1}^{p} E[H_i^{-1} \circ \Psi_k^{-1}]P(I_{ij}) = p^{-1}\sum_{j=1}^{p} \Psi_j^{-1} = id. \tag{2.17}$$

This observation motivates the warping problem

$$Z_i^* = \lambda \circ D_i, \quad \text{for } i = 1, \ldots, n. \tag{2.18}$$

The critical implication of this relation is that if we expand our scope to the full collection S and apply a traditional method like pairwise warping to obtain $\hat{D}_i$ for all $i = 1, \ldots, n$, the latent curve can be estimated by averaging the globally-aligned curves,

$$\hat{\lambda} = n^{-1}\sum_{i=1}^{n}(Z_i \circ \hat{D}_i^{-1})/||Z_i||_\infty \tag{2.19}$$

The estimators of the component transports are motivated by recalling that

$$\gamma_j = \lambda \circ \Psi_j \quad \text{for} \quad j = 1, \ldots., p.$$

Using a spline representation similar to (2.9), we write

$$\Psi_j(t) = \theta^T \alpha(t) \tag{2.20}$$

and estimate the component warps by solving the penalized minimization problem,

$$\begin{aligned} \tilde{\theta}_{\Psi_j} &= \underset{\theta \in \Theta}{\operatorname{argmin}}\, \mathscr{C}_{\eta_2}(\theta; \hat{\gamma}_j, \hat{\lambda}), \\ \text{where } \mathscr{C}_{\eta_2}(\theta; \hat{\gamma}_j, \hat{\lambda}) &= \int_{\mathcal{T}} d^2\left(\hat{\gamma}_j, \hat{\lambda}(\theta^T\alpha(t))\right) dt + \eta_2 \int_{\mathcal{T}} (\theta^T\alpha(t) - t)^2 dt. \end{aligned} \tag{2.21}$$

Again the penalty parameter can be set to $\eta_2 = 10^{-4} \times \{p^{-1}\sum_{j=1}^{p}\int_{\mathcal{T}}(\hat{\gamma}_j(t) - \hat{\lambda}(t))^2 dt\}$ in line with Tang & Müller (2008) or tuned by cross-validation. Finally, we estimate the component warps as

$$\hat{\Psi}_j(t) = \tilde{\theta}_{\Psi_j}^T \alpha(t). \tag{2.22}$$

Note that under the assumption of fully observed curves without measurement error, the amplitude factors $A_{ij} = ||X_{ij}||_\infty$ are known. Often in practice, this is not realistic, and the factors must be estimated by, e.g., $\hat{A}_{ij} = ||\tilde{X}_{ij}||_\infty$ where $\tilde{X}$ denotes a smoothing estimate, as described in the following section.

2.3.4 Measurement Error

In practice, the functions X_{ij} are often contaminated with measurement error and available only on a discrete grid. In this situation, an initial step is to perform smoothing on the discrete and noisy observations before applying the model estimation method of Sections 3.2 and 3.3. A model for describing contamination by measurement error is posed as follows.

Let $\{\boldsymbol{X}_i,\ i = 1, \ldots, n\}$ be a random sample of a p-dimensional stochastic process $\boldsymbol{X}$ in $L_2(\mathcal{T}) \times \cdots \times L_2(\mathcal{T})$, where $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{ip}(t))^T$. We assume that these processes are observed for m equispaced points at the discrete times $t_s,\ s = 1, \ldots, m$. We write $\boldsymbol{Y}_{is} = (Y_{i1s}, \ldots, Y_{ips})^T$ to represent the i^{th} subject's observation at time t_s subject to measurement errors, $\boldsymbol{\epsilon}_{is} = (\epsilon_{i1s}, \ldots, \epsilon_{ips})^T$, which are independent of $\boldsymbol{X}_i$:

$$\boldsymbol{Y}_{is} = \boldsymbol{X}_i(t_{is}) + \boldsymbol{\epsilon}_{is}. \tag{2.23}$$

The measurement errors, ϵ_{ijs}, are assumed to be independent and identically distributed with mean zero and component-specific variances $\sigma_j^2 < \infty$ for $j = 1, \ldots, p$. We consider the case of dense component curves so that the recording times t_s can be viewed as a grid on the supports of each component. Estimation of the smooth component curves, $\tilde{X}_{ij}$, may be performed by any smoothing technique, e.g. kernel methods, smoothing splines, or local polynomial fitting via locally weighted least squares.

Example 1. Let $K : \mathbb{R} \to \mathbb{R}$ represent a non-negative kernel function. We apply local linear smoothing with bandwidth b across each individual so that

$\tilde{X}_{ij}(t) = \hat{\beta}_{ij}^{(0)}$ where

$$(\hat{\beta}_{ij}^{(0)}, \hat{\beta}_{ij}^{(1)}) = \underset{\beta_{ij}^{(0)}, \beta_{ij}^{(1)}}{\operatorname{argmin}} \sum_{s=1}^{m} K\left(\frac{t_s - t}{b}\right) (Y_{ijs} - \beta_{ij}^{(0)} - \beta_{ij}^{(1)}(t - t_s))^2, \tag{2.24}$$

for all $i = 1, \ldots, n,\ j = 1, \ldots, p$. Bandwidth selection procedures may be data-adaptive (see e.g., Rice & Silverman (1991)) but a subjective choice often suffices in practice.

After smoothing is performed, estimation may be carried out by substituting $\tilde{X}_{ij}$ for X_{ij} and implementing the procedure described in Sections 3.2 and 3.3. Once all model components are estimated, plug-in estimates of the composite distortion functions and marginal and subject-level component transport functions follow immediately:

$$\hat{G}_{ij} = \hat{\Psi}_j \circ \hat{H}_i, \tag{2.25}$$

$$\hat{T}_{jk} = \hat{\Psi}_j^{-1} \circ \hat{\Psi}_k, \quad \text{and} \tag{2.26}$$

$$\hat{T}_{jk}^{(i)} = \hat{G}_{ij}^{-1} \circ \hat{G}_{ik}, \quad \text{for} \quad i = 1, \ldots, n, \quad j, k = 1, \ldots, p. \tag{2.27}$$

Additionally, fitted curves based on the LTM can be obtained as

$$\begin{aligned} \hat{X}_{ij}(t) &= \hat{A}_{ij}(\hat{\lambda} \circ \hat{G}_{ij})(t) \\ &= \hat{A}_{ij}(\hat{\lambda} \circ \hat{\Psi}_j \circ \hat{H}_i)(t), \end{aligned} \tag{2.28}$$

for $i = 1, \ldots, n,\ j = 1, \ldots, p$. These fits can be viewed through the lens of dimension reduction as their calculation require only $n + p + 1$ estimated functions as opposed to np curves in the original data. In Section 6, we derive asymptotic results for the proposed estimates in the cases of curves observed with and without measurement error.

2.4 Data Applications

2.4.1 Zürich Growth Study

Our first application is a reexamination of the Zürich longitudinal growth study from 1954-1978. The sitting heights, arm lengths, and leg lengths of a cohort of children were measured on a dense time grid such that the resulting data can be viewed as multivariate functional data. We focus on the timing of pubertal growth spurts, which usually occur between ages 9 and 18. It is standard in growth curve literature to examine the derivatives of the growth curves, i.e. the growth velocities, instead of the curves themselves (Gasser et al., 1984b). The growth velocities have a peak during puberty, with the crest location representing the age when an individual is growing fastest. The timings and curvatures of these peaks are critical in informing growth patterns. We estimate these growth velocities, displayed in Fig. 2.2, via local linear smoothing. It is known among physicians and laypeople alike that there is a difference in the ways that boys and girls undergo puberty. This distinction is clear from just a simple inspection of the growth velocities in Figure 1, but researchers may want to further quantify these differences. For example, medical practitioners may wish to evaluate differences in the size of body parts between boys and girls at specific times or quantify the differential between onsets of puberty for boys and girls. The proposed representation provides a path forward to answer simple questions like these as well as ones of a more complicated nature.

It is clear that girls start puberty earlier than boys do: girls' system times accelerate rapidly between the ages of 9 and 12, whereas boys typically must wait until ages 12 or 13 to experience substantial deviation from clock time (represented by the black dashed line on the diagonal). Component tempos for boys and girls are a simple way to summarize these differences (Fig. 2.2, dashed and dotted lines, respectively), as they serve as the structural means of the timing

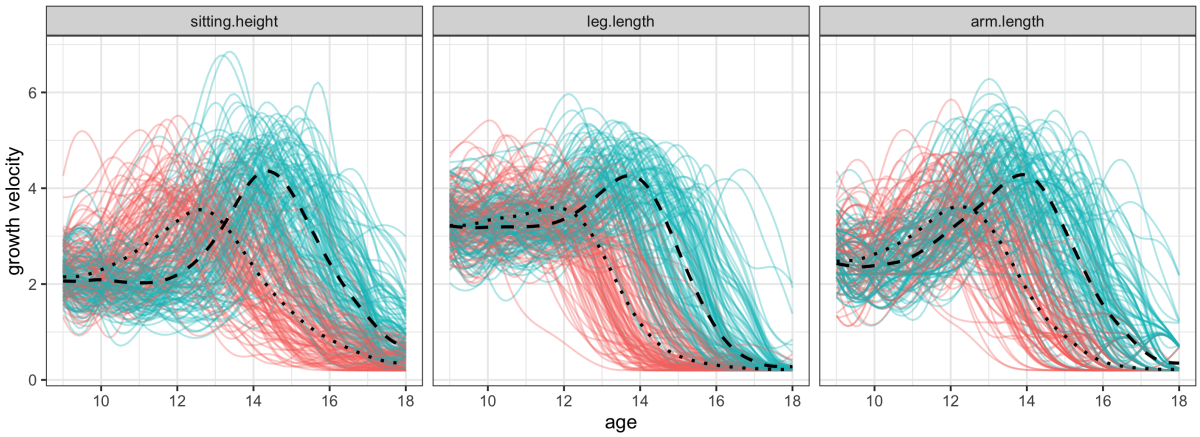

Figure 2.2: Growth velocities (in cm/year) during puberty for boys (blue) and girls (red). Scaled component tempo functions are marked for boys and girls with dashed and dotted lines, respectively.

functions.

Turning our attention to the joint time dynamics of the $p = 3$ modalities, we restrict our analysis to the boys for the sake of brevity. A natural place to start when comparing growth patterns is the component tempos, which are displayed for each modality in the left panel of Fig. 2.3. The dynamics of joint development can be seen through by examining the order of peaks across modalities. In this case leg length is first, followed by arm length, while sitting height lags behind. The tempos obtain similar slopes during puberty, though leg length has the most gradual spurt and sitting height has the sharpest increase, This may reflect the fact that its lagged onset results in a smaller window between the onset of its growth spurt and the maturation date of 18 years. We note that it is of course possible for an individual to experience some minor growth past the age of 18, but in the Zürich study such cases were rare and so this complication was ignored. The component transports displayed in Fig. 2.3 (right) further illustrate the nature of each body parts tempo relative to the baseline latent system time. Remarkably, the tempo of arm length is nearly identical to the latent curve. This suggests that the arm can be used a representative modality which mirrors a

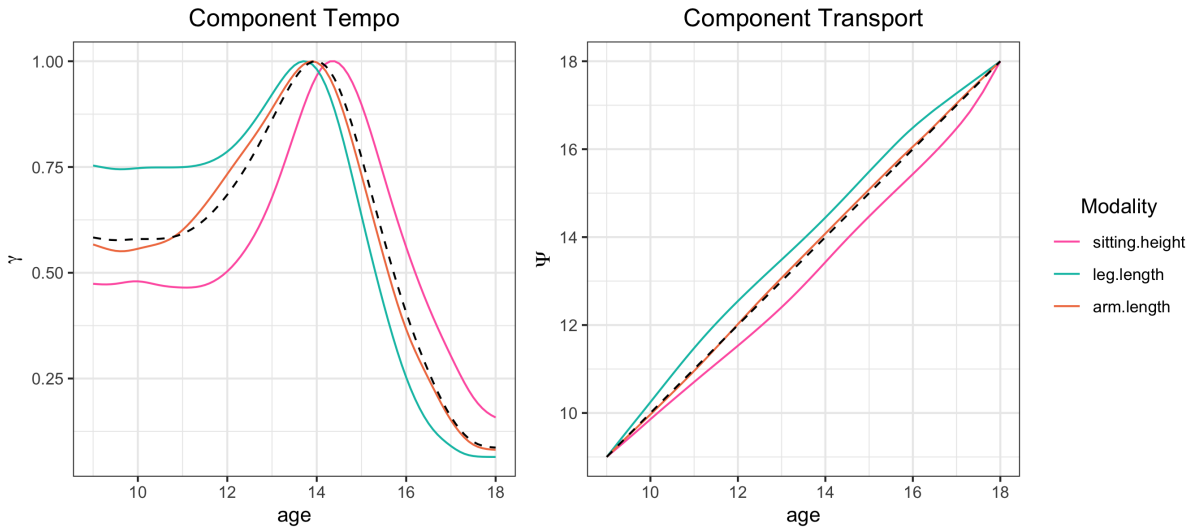

Figure 2.3: Component tempos γ (left) and transports Ψ (right) for growth modalities. The dashed line represents the tempo and transport for the latent tempo, λ.

child's overall development.

The estimated cross-component transports, $\hat{T}_{jk},\ j,k \in \{1,\ldots,p\}$ as per (2.26), are depicted in Figure 2.4 and illustrate the intercomponent time dynamics for the three growth modalities. An XCT map can be interpreted as the transport required in order to accelerate or decelerate the tempo of component j to that of component k. Taking the transport between sitting height and leg length, $\hat{T}_{23}$ as an example, we see that the XCT map falls above the identity, indicating that sitting height's tempo must be accelerated in order to synchronize with that of leg length. This matches the interpretation that individuals tend to experience growth spurts in leg lengths before sitting height, reflected in Fig. 2.3 and previous descriptive analyses of pubertal growth spurts (Sheehy et al., 1999).

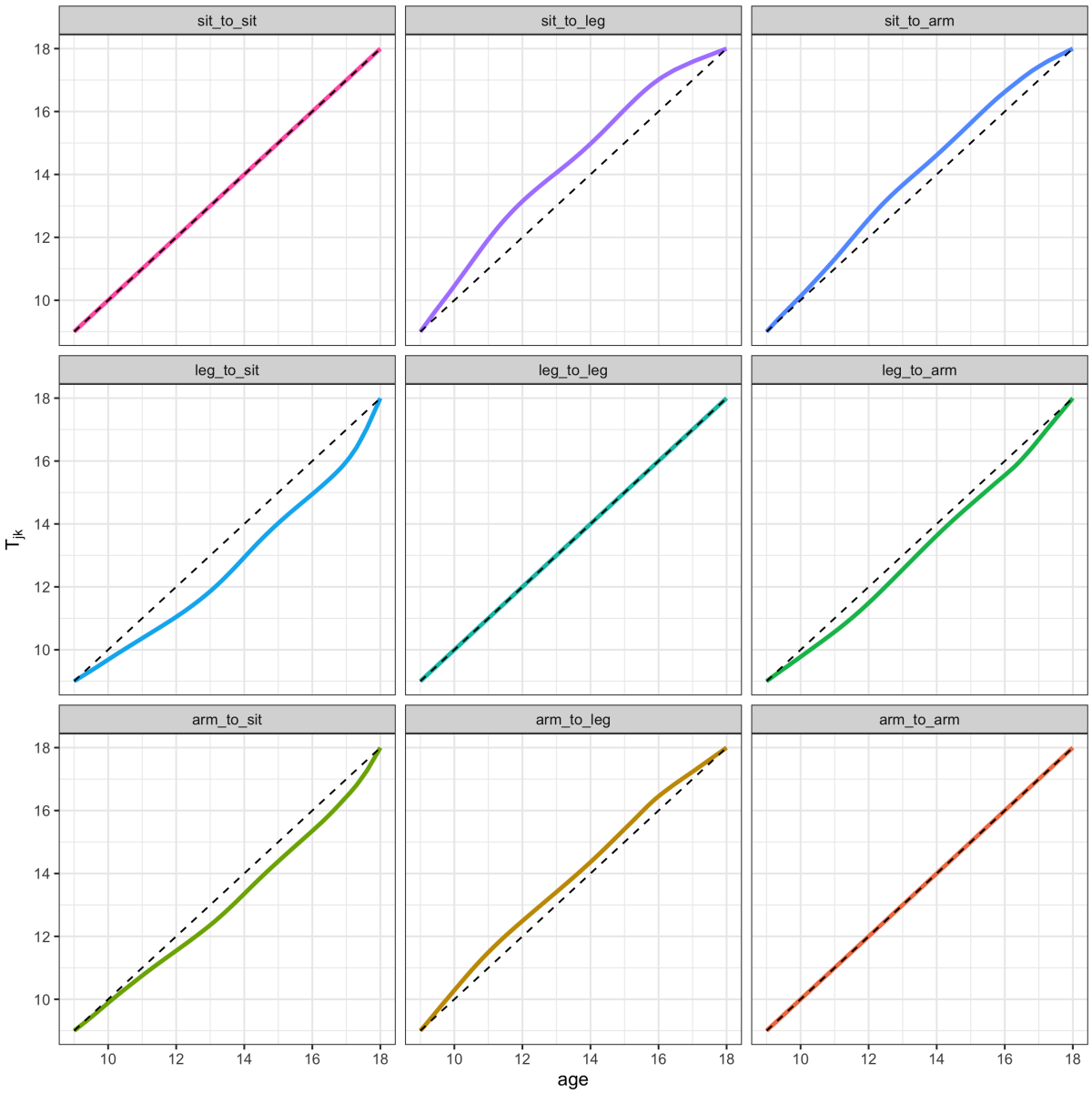

Figure 2.4: The cross-component transport matrix, which characterizes pairwise time relations between components.

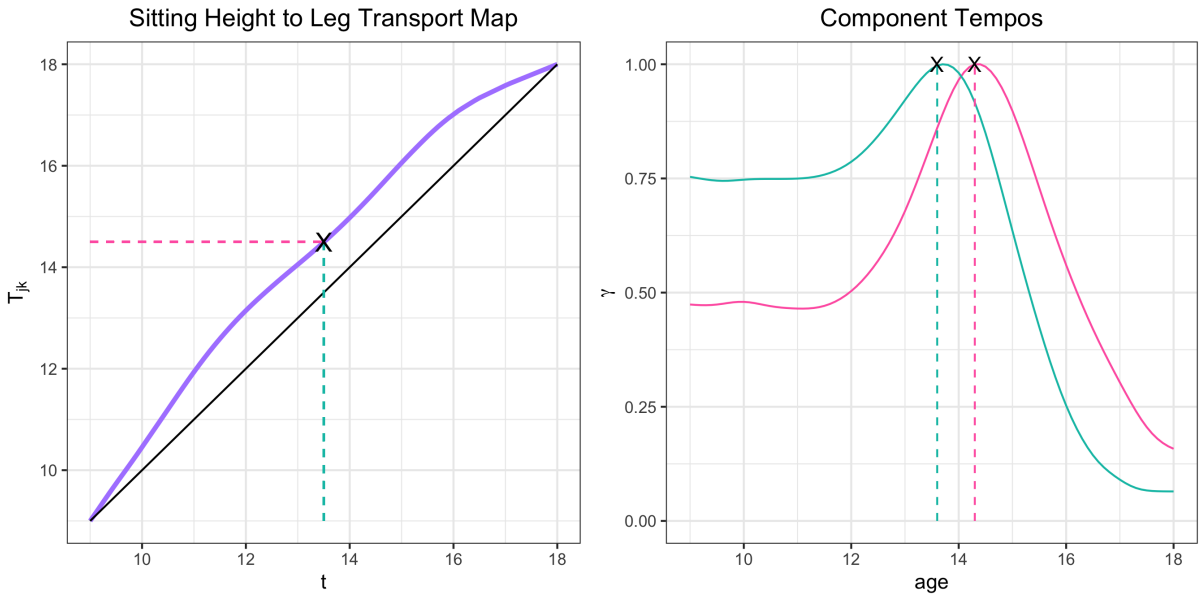

Figure 2.5: The cross-component transport map T_{12} which expresses the sitting height's timing patterns relative to the leg length's as a baseline. The peak of pubertal growth rate for the leg occurs at approximately age 13.5, while the maximum growth velocity for sitting height growth occurs at approximately $T_{12}(13.5) \approx 14.5$ years old.

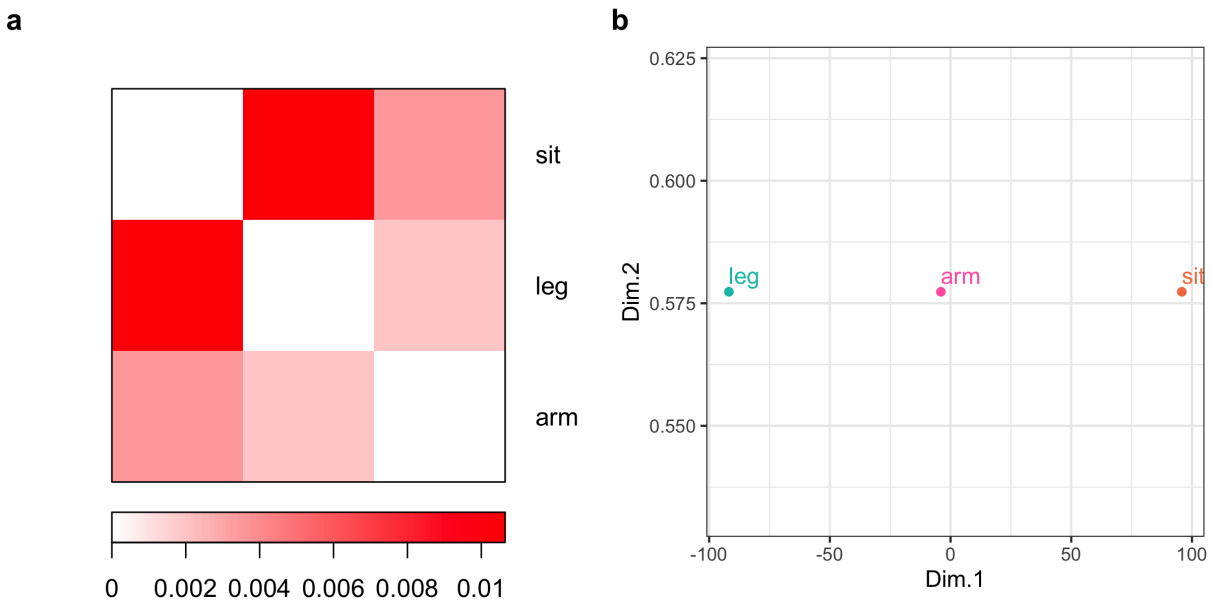

Figure 2.6: The (a) dissimilarity matrix heatmap and (b) MDS-generated representation of cross-component tempos for growth modalities. The tempos of sitting height and leg length are least similar, while arm length lies in between. MDS suggests the dominant variation among the tempos is one-dimensional and mirrors the ordering of growth spurts across modalities.

The magnitude of the XCT map's deviation from the identity shows how dissimilar the two components are. For example, sitting height and leg length are the most distinct modalities of growth among those considered here, and their XCT map exhibits the most pronounced departure from the identity. An intuitive interpretation of the map is that T_{jk} expresses the k^{th} component's timing patterns relative to the j^{th} component's as a baseline. For example, when the leg tempo is at time $t = 13.5$, the comparable time point for the sitting height tempo is approximately at $T_{jk}(13.5) \approx 14.5$ as illustrated in Fig. 2.5.

To quantify the differences in tempos across components as a scalar, one may take the Wasserstein distance (Villani, 2003) between $\hat{T}_{jk}(t)$ and $id(t) \equiv t$ to create a dissimilarity matrix for the component tempos. Fig. 2.6 illustrates the dissimilarity matrix for the Zürich data and the results of performing multidimensional scaling (MDS), which results in a configuration of the three components in an abstract coordinate system. Remarkably, this representation shows that the component tempos can be represented well by a single dimension of variation. Additionally, their relative positions on this axis match the timing of the growth spurts for the three modalities.

2.4.2 Air Pollutants in Sacramento, CA

The study of air pollutants has been a topic of interest for atmospheric scientists and environmentalists alike for several decades. In particular, increased ground-level ozone (O_3) concentrations have been shown to have harmful effects on human health. Unlike many air pollutants, surface ozone is not directly emitted by sources of air pollution (e.g. road traffic); it is formed as a result of interactions between nitrogen oxides and volatile organic compounds in the presence of sunlight (Abdul-Wahab, 2001). Because of this interaction, compounds such as nitrogen dioxide are known and important precursors of increased ozone concentrations (Tu et al., 2007).

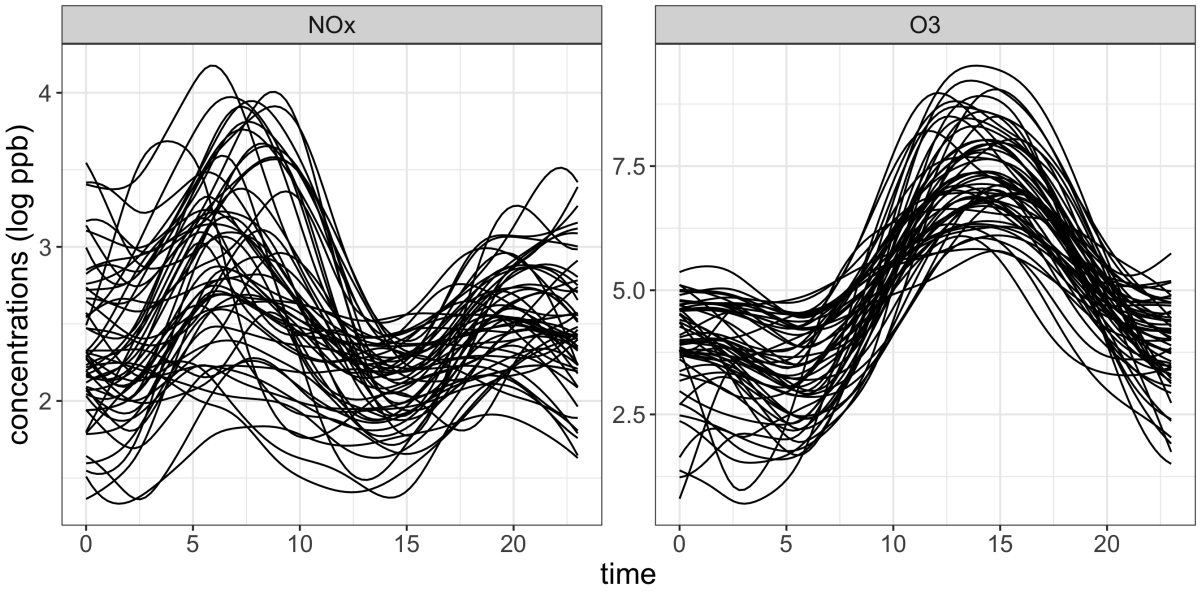

Figure 2.7: 24-hour trajectories of NO_x (left) and ozone (right), concentrations in parts per billion (ppb) on a log scale.

The California Environmental Protection Agency has monitored hourly air pollutant concentrations at several station locations since the 1980s. Here we consider the sample of weekday trajectories of ozone (O_3), and nitrogen oxides (NO_x) concentrations during the summer of 2005 in Sacramento (Fig. 2.7). Smooth trajectories were obtained from raw data using locally linear weighted least squares. Gervini (2015) has previously investigated a similar dataset in the context of warped functional regression, where the primary aim was to model phase variation explicitly in order to relate the timing of peak concentrations of NO_x to those of O_3.

The chemistry of the compounds as well as a visual inspection of the curves suggests that we have two classes of pollutants. NO_x concentrations tend to peak around 8 a.m., reflecting standard morning commute hours and the impact of traffic emissions on air quality. On the other hand, ozone levels peak around 2 to 3 p.m., indicating that the synthesis mechanism induces a lag of up to approximately 6 hours.

It is an interesting question to ask if meteorological factors might affect the rate of ozone synthesis. Individual component transports combined with Fréchet regression for distributions provide a natural framework to study this query (Petersen & Müller, 2019). Subject-specific transports from NO_x concentrations to ozone concentrations, $T^{(i)}_{NO_x \to O_3}$, were calculated as per (2.27) for each day.

Global Fréchet regression was then applied through the model

$$\begin{aligned} \hat{m}_{\oplus}(x) &= \underset{T \in \mathcal{W}}{\operatorname{argmin}}\, M_n(T, x) \\ \text{with } M_n(T, x) &= n^{-1} \sum_{i=1}^{n} q_{in} d_W^2(T_i, T), \end{aligned} \tag{2.29}$$

where $m_{\oplus}$ denotes the conditional Fréchet mean of the transport given the covariate x, the wind speed recorded a given day. Here, d_W denotes the $2-$Wasserstein distance (Villani, 2003) and the weights are defined as $q_{in} = 1 + (x_i - \bar{x})(x - \bar{x})/\hat{s}_x^2$,

as per (Petersen & Müller, 2019), where $\bar{x}$ and $\hat{s}_x^2$ represent the sample mean and variance of the observed wind speeds, respectively. The model was fit using the `R` package `frechet`, observing that the transports can be represented as distribution functions (Chen et al., 2020).

Figure 2.8 displays the observed transports and the fits obtained from Fréchet regression using windspeed as a predictor. The rainbow gradient corresponds to windspeeds ranging from 3 to 10 knots and their associated fitted transports are overlaid upon the original data. The regression fits suggest that days with lower windspeeds correspond with transports which are further from the diagonal, indicating an exaggerated lag between peak concentrations of NO_x and ozone. On the other hand, days with high wind speeds have fitted transports very near the diagonal, which suggests that windier settings accelerate the synthesis process. Intuitively this is a reasonable result in terms of the physical interpretation, as more wind will result in a higher rate of collisions of the particles, and thus quicker production of ozone after peak NO_x emission. The Fréchet $R^2_\oplus$ value was 0.44, which suggests that wind speed explains a considerable amount of variation in the observed transports.

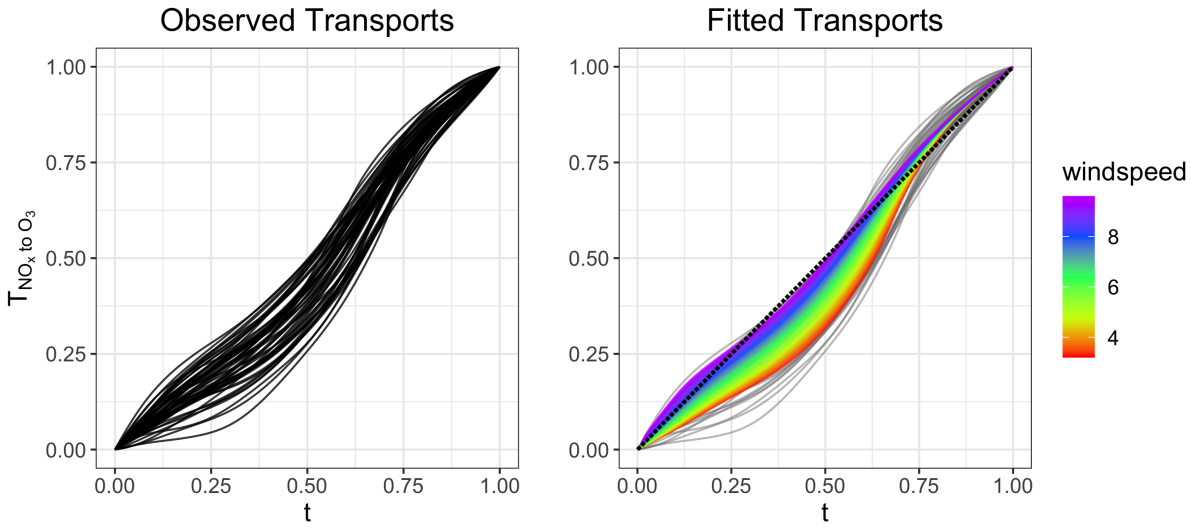

Figure 2.8: Fréchet regression of NO_x-to-O_3 cross-component transports onto daily max windspeeds in knots. Windier days correspond to more linear transport functions, which suggests O_3 synthesis more closely follow NO_x emission. Less windy days are associated with more pronounced lags between the pollutants.

2.5 Simulation Study

In this section we illustrate the latent transport model through a simulated $p = 4-$ dimensional dataset which exhibits intercomponent warping of a base latent curve. Component curves are simulated on a grid $\mathcal{T} = [0, 1]$ traversed by increments of 0.05 as

$$X_{ij}(t_s) = A_{ij}(\lambda \circ \Psi_j \circ H_i \circ R_{ij})(t_s) + \epsilon_{ijs}, \tag{2.30}$$

where the latent curve is defined as $\lambda(t) = \lambda_0(t)/||\lambda_0||_\infty$, for $\lambda_0(t) = 20 + 15t^2 - 5\cos(4\pi t) + 3\sin(\pi t^2)$, the random amplitudes are distributed as $A_{ij} \overset{iid}{\sim} \mathcal{N}(100, \sigma_A^2)$, the component distortion functions $\Psi_j,\ j = 1, \ldots, 4$ are mixtures of Beta distributions, and H_i are random distortions of the unit interval. Specifically, we set $\Psi_j(t) = \vartheta B_t(a_j, b_j) + (1 - \vartheta)t$ for $j = 1, 2$, where B_t denotes the regularized incomplete Beta function and $a = (2, 1)^T$, $b = (2, 1/2)^T$, and $\vartheta = 0.5$. Then we define $\Psi_{j+2}^{-1}(t) = 2t - \Psi_j^{-1}(t),\ j = 1, 2$, so the net identity warp constraint $p^{-1}\sum_{j=1}^{p} \Psi_j^{-1}(t) = t$ is satisfied. The sources of random phase variation are introduced by the subject-level warping functions defined by $H_i^{-1}(t) = \dfrac{\exp(tw_i) - 1}{\exp(w_i) - 1}$ where $w_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_W^2)$ and a random nuisance distortion function defined by $R_{ij}^{-1}(t) = \dfrac{\exp(td_{ij}) - 1}{\exp(d_{ij}) - 1}$ where $d_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_D^2)$. Finally, measurement error is also added to contaminate the raw observations through the random variables $\epsilon_{ijs} \overset{iid}{\sim} \mathcal{N}(0, \sigma_E^2)$. The levels of noise varied such that $\sigma_W = 0, .5, 1$, $\sigma_D = 0, .5, 1$, and $\sigma_E = 0, 1, 5, 10$. Figure 2.1 displays an example of simulated data and component warping functions for the case where $\sigma_W = 0.5$, $\sigma_D = 0$, and $\sigma_E = 1$.

The component warping functions were chosen to induce both distortions which accelerate or delay the latent trajectory uniformly and more complicated distortions which alternate between stretching and shrinking the time domain. The former type can be considered a warping structure with a simpler effect, as the warping functions remain on either side of the diagonal. In some cases, these

kinds of warps may be approximated adequately with simple shift parameters. On the other hand, a component transport which oscillates back and forth across the diagonal corresponds to a nuanced, non-rigid effect which needs added flexibility beyond that of shift-warping. The latent transport model proposed here provides this increased freedom, as we illustrate in this simulation. We also provide details on the finite sample performance of our methods on the fits of both the latent curves and the realizations of the full process.

Pre-smoothing of curves was performed using local linear regression with the Epanechnikov kernel to handle the presence of measurement error. The number of knots was set at $K = 4$, though we note that this hyperparameter did not drastically change the quality of estimates in practice unless it was set at a blatantly low value ($K \leq 2$), which resulted in lack of fit, or raised excessively high ($K \geq 7$), which distorted the estimates of warping functions and resulted in prohibitively high computation time.

Estimates were obtained from fitting the model on a sample of $n = 50$ $4-$dimensional processes at each noise level. This process was repeated for $B = 250$ Monte Carlo runs and the results are summarized in terms of four performance measures. We calculated the integrated squared error for the estimated latent curve, component warps, and fitted processes as

$$\begin{aligned} \textit{LISE} = \int_0^1 \{\hat{\lambda}(t) - \lambda(t)\}^2 dt, \ \textit{HMISE} = \frac{1}{n}\sum_{i=1}^{n} \int_0^1 \{\hat{H}_i(t) - H_i(t)\}^2 dt, \text{ and} \\ \textit{XMISE} = \frac{1}{np}\sum_{j=1}^{p}\sum_{i=1}^{n} \int_0^1 \{\hat{X}_{ij}(t) - X_{ij}(t)\}^2 dt, \quad \text{for} \quad j,k = 1,\ldots,p. \end{aligned} \tag{2.31}$$

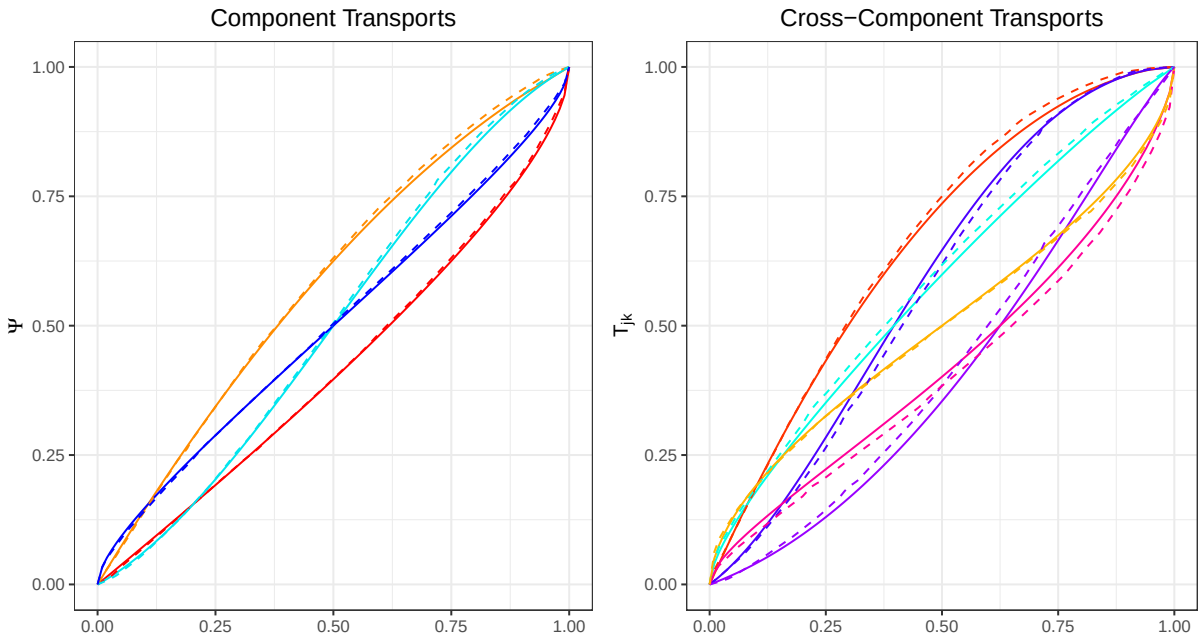

Figure 2.9: Estimated (dashed) vs. true (solid) component transport functions (left) and cross-component transport functions (right) under the latent transport model at the noise setting with $\sigma_W = 0.5$, $\sigma_D = 0.5$, and $\sigma_E = 1$.

Tables $2.1 - 2.3$ depict these quantities based on samples generated at each noise level. For ease of reading, the quantities *LISE* and *HMISE* are magnified by a factor of 10^2 as the actual values have several leading zeros. Table 2.1 shows that estimation of the latent curve λ generally degrades with increased warping and nuisance distortion levels in all cases. Interestingly the effect of these distortions is comparable between levels where standard deviations are 0 and 0.5, the latter corresponds to a typical degree of warping, but jumps when increased to 1, which reflects a severe amount of phase variation. Effects on the integrated squared error for the latent curve are similar for both sources of distortion, σ_W and σ_D. Increases in measurement error on the other hand do not seem to affect the quality of the estimate of λ much. A potential explanation for this is the fact that the estimator of λ is an average, so additive measurement errors tend to cancel each other out.

As for the estimates of the subject-specific warping functions H, we see a similar trend across both kinds of distortion (see Table 2.2). As σ_W increases, the warping function targets have a greater likelihood to be more extreme warps, which are harder to estimate by nature. In addition, the bias imposed by regularization is greater for these extreme cases. As expected, greater nuisance distortion also degrades performance for similar reasons. Measurement error is also associated with worse performance, which is in contrast to its effect on the estimation of λ. Since the warping functions H are subject-specific, and not averaged over the sample, they do not enjoy the robustness to measurement error that the latent curve displays.

Finally, performance measures for the sample of fitted curves are displayed in Table 2.3. Mean integrated squared errors for fitted curves X_{ij} can be thought of as a composite measure of the errors in estimates of the latent curve λ, component transports Ψ_j, individual warps H_i, and amplitude factors A_{ij}. Note that the

amplitude factors are randomly scattered about 100 for each component, so the scale of errors for fitted curves are much larger than that of the latent curve and warping functions.

The sources of noise have a slightly different effect on *XMISE* as they do on *LISE* and *HMISE*. Again, more variable warping and nuisance distortions result in worse performance, however in the case of fitted curves, there is a difference between warping and nuisance distortion. Increasing the level of warping distortion σ_W has less of an impact on *XMISE* than does increasing that of nuisance σ_D. For example, in the case of no measurement error, increasing σ_D from 0 to 0.5 results in a roughly 175-fold increase in *XMISE*, while a similar change in σ_W increases the error by a factor of 6.5. A similar pattern occurs over all levels of measurement error. This observation suggests that increased nuisance distortion degrades the estimation of the component transports Ψ_j, this trend was not seen in the estimation of λ or H, and amplitude factors are invariant to distortion effects of all kinds. Measurement errors have a similar effect as they did on *HMISE*, which again may be explained by the fact that fitted curves are subject-specific and not marginalized.

Table 2.1: Integrated squared error of the latent curve estimates under simulated noise settings, magnified by factor of 100.

LISE $\times 10^2$		*Measurement Error*			
Warping Distortion	*Nuisance Distortion*	$\sigma_E = 0$	$\sigma_E = 1$	$\sigma_E = 5$	$\sigma_E = 10$
$\sigma_W = 0$	$\sigma_D = 0.0$	0.01	0.01	0.01	0.01
	$\sigma_D = 0.5$	0.00	0.00	0.01	0.01
	$\sigma_D = 1.0$	0.09	0.08	0.06	0.08
$\sigma_W = 0.5$	$\sigma_D = 0.0$	0.01	0.01	0.01	0.01
	$\sigma_D = 0.5$	0.01	0.01	0.01	0.02
	$\sigma_D = 1.0$	0.10	0.09	0.07	0.09
$\sigma_W = 1$	$\sigma_D = 0.0$	0.11	0.10	0.08	0.10
	$\sigma_D = 0.5$	0.13	0.11	0.09	0.11
	$\sigma_D = 1.0$	0.32	0.31	0.28	0.31

Table 2.2: Mean integrated squared error (magnified by 10^2) of the subject-level warping estimates $\hat{H}_i(t)$ under simulated noise settings.

HMISE $\times 10^2$		*Measurement Error*			
Warping Distortion	*Nuisance Distortion*	$\sigma_E = 0$	$\sigma_E = 1$	$\sigma_E = 5$	$\sigma_E = 10$
$\sigma_W = 0$	$\sigma_D = 0.0$	0.00	0.00	0.01	0.02
	$\sigma_D = 0.5$	0.02	0.02	0.02	0.03
	$\sigma_D = 1.0$	0.15	0.15	0.12	0.10
$\sigma_W = 0.5$	$\sigma_D = 0.0$	0.00	0.00	0.01	0.02
	$\sigma_D = 0.5$	0.02	0.02	0.02	0.03
	$\sigma_D = 1.0$	0.15	0.15	0.12	0.11
$\sigma_W = 1$	$\sigma_D = 0.0$	0.12	0.12	0.13	0.21
	$\sigma_D = 0.5$	0.14	0.13	0.15	0.22
	$\sigma_D = 1.0$	0.29	0.28	0.31	0.38

Table 2.3: Mean integrated squared error of the fitted curve estimates $\hat{X}_{ij}(t)$ under simulated noise settings. Note the increase in scale of the *XMISE* when compared to *LISE* and *HMISE* is due to the fact that curves are scaled up by the amplitude factor which varies randomly around the value 100.

XMISE		*Measurement Error*			
Warping Distortion	*Nuisance Distortion*	$\sigma_E = 0$	$\sigma_E = 1$	$\sigma_E = 5$	$\sigma_E = 10$
$\sigma_W = 0$	$\sigma_D = 0.0$	0.04	0.50	13.09	57.19
	$\sigma_D = 0.5$	6.97	7.35	20.05	63.65
	$\sigma_D = 1.0$	80.98	81.50	92.78	140.97
$\sigma_W = 0.5$	$\sigma_D = 0.0$	0.26	0.74	13.32	57.37
	$\sigma_D = 0.5$	7.11	7.54	20.14	63.81
	$\sigma_D = 1.0$	80.75	81.84	93.75	143.37
$\sigma_W = 1$	$\sigma_D = 0.0$	17.35	19.22	34.28	91.97
	$\sigma_D = 0.5$	25.48	25.20	42.68	98.38
	$\sigma_D = 1.0$	104.60	102.80	120.65	181.99

2.6 Theoretical Results

Our results focus on convergence of the components of the XCT model described in (4) as the number of curves n and the number of observations per curve m tend to infinity. In order to achieve asymptotic results, we require the following assumptions on (L) the components of the Latent Transport Model and (S) the smoothing methodology in the presence of discretely observed curves.

(L1) The latent curve $\lambda(t) \in C^2(\mathcal{T})$ is a bounded function. For any non-degenerate interval $\mathcal{T}_0 \subset \mathcal{T},\ 0 < \int_{\mathcal{T}_0} \lambda'(t)^2 dt < \infty$.

(L2) For $j = 1, \ldots, p$, $\sup_{1 \le i \le n} A_{ij} = \mathcal{O}_P(1)$ and $\sup_{1 \le i \le n} A_{ij}^{-1} = \mathcal{O}_P(1)$.

Assumption (L1) bounds the latent curves and its derivatives and ensures there are no flat stretches. These conditions are made to avoid behavior which would undermine the asymptotic results and uniqueness of the component estimates. (L2) ensures that the ranges of the random processes are bounded away from zero and infinity with high probability; this condition is needed for the uniform convergence of the smoothing estimate.

(S0) The time points $t_1, \ldots, t_m$, where the sequence m is assumed to depend on the sample size n, $m = m(n)$, correspond to a dense regular design with smooth design density f with $\inf_{t \in \mathcal{T}} f(t) > 0$ that generates the time points according to $t_s = F^{-1}(\frac{s-1}{m-1}),\ s = 1, \ldots, m$, where F^{-1} denotes the quantile function associated with f. The second derivative f'' is bounded, $\sup_{t \in \mathcal{T}^\circ} |f''(t)| < \infty$.

(S1) The kernel function K is a probability density function with support $[-1, 1]$. It is also symmetric around zero and uniformly continuous on its support with $\int_{-1}^{1} K^2(u) du < \infty$.

(S2) For each $j = 1, \ldots, p$, the sequences $m = m(n)$ and $b = b(n)$ satisfy (1) $0 < b < \infty$, and (2) $m \to \infty$, $b \to 0$, and $mb^2(\log b)^{-1} \to \infty$ as $n \to \infty$.

These assumptions guarantee the consistent estimation of n curves simultaneously, as shown in the following Proposition. We observe that (S2) is for example satisfied if the bandwidth sequence is chosen such that $b = b(n) \sim m(n)^{-1/6}$. We next establish the rate of convergence for the smoothed curves and also the normalized versions to the true underlying processes.

Proposition 1. *Under assumptions (S0−S2), if* $E||X^{(\nu)}(t)||^2_\infty < \infty,\ \nu = 0, 1, 2$, *we have the uniform convergence*

$$\sup_{t\in\mathcal{T}}|\tilde{X}_{ij}(t) - X_{ij}(t)| = \mathcal{O}_P(m^{-1/3}). \tag{2.32}$$

The rate also extends to the standardized versions $X^*_{ij} = X_{ij}/||X_{ij}||_\infty$,

$$\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| = \mathcal{O}_P(m^{-1/3}). \tag{2.33}$$

The estimators of the latent curve and component transports involve averages of the smoothing estimates over the sample of curves as $n \to \infty$. The corresponding rates of convergence will thus rely on the uniform summability of the difference between the smoothed and true curves over n and we then have a uniform rate of $\tau_m = m^{-(1-\delta)/3}$ for an arbitrarily small $\delta > 0$ in lieu of the above rate $m^{-1/3}$; see Lemma 1 in the Technical Proofs. The proposed estimators also rely on the mechanics of the pairwise warping methods, whose convergence properties have been established in a general form in Tang & Müller (2008) and Chen & Müller (2020). Lemma 2, provided in the Technical Proofs section, states these rates in the specific framework of the Latent Transport Model. We are now in a position to state our main result, which establishes rates of convergence for the estimators of the components of the Latent Transport Model as follows.

Theorem 1. *Under assumptions (L1), (L2), and (S0−S2), with* $\tau_m = m^{-(1-\delta)/3}$ *for an arbitrarily small* $\delta > 0$ *and penalty parameters* η_1, η_2 *as in (2.11), (2.21), we have for all* $i = 1, \ldots, n,\ j = 1, \ldots, p$*,*

a. $\sup_{t \in \mathcal{T}} |\hat{H}_i(t) - H(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2})$,

b. $\sup_{t \in \mathcal{T}} |\hat{\gamma}_j(t) - \gamma_j(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2})$

c. $\sup_{t \in \mathcal{T}} |\hat{\lambda}(t) - \lambda(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2})$,

d. $\sup_{t \in \mathcal{T}} |\hat{\Psi}_j(t) - \Psi_j(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2})$,

e. $\sup_{t \in \mathcal{T}} |\hat{G}_{ij}(t) - G_{ij}(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2})$, *and*

f. $|\hat{A}_{ij} - A_{ij}| = \mathcal{O}_P(m^{-1/6})$.

The three terms in the rates correspond, in order, to (1) the parametric rate achieved through the standard central limit theorem, (2) the smoothing rate which is dependent on the number of observations per curve m, and (3) a rate due to the well-known bias introduced by the penalty parameters used in the regularization steps. Additionally, if we suppose that m is bounded below by a multiple of $n^{3(1-\delta)^{-1}}$, then the rates corresponding to the smoothing steps are bounded above by $n^{-1/2}$. If we take the penalty parameters to be $\eta_1 \sim \eta_2 = \mathcal{O}(n^{-1})$, a $n^{-1/2}$ rate of convergence can be achieved for each of the estimators in Theorem 1 $a.$-$e.$ Otherwise, if $m \sim n^{\Delta(1-\delta)^{-1}}$, for any $\Delta < 3$, the convergence is limited by the smoothing step and achieves the rate of $n^{-\Delta/6}$. This line of reasoning gives the following result for curves which are observed fully or on sufficiently dense designs.

Corollary 1. *Suppose the penalty parameters $\eta_1 \sim \eta_2 = \mathcal{O}(n^{-1})$. If the random trajectories are fully observed without error or the trajectories are recorded with at least a multiple of $m \sim n^{\Delta(1-\delta)^{-1}}$ observations per curve, with $\Delta > 3$, then under the assumptions of Theorem 1, we have for all $i = 1, \ldots, n,\ j = 1, \ldots, p$,*

a. $\sup_{t \in \mathcal{T}} |\hat{H}_i(t) - H(t)| = \mathcal{O}_P(n^{-1/2})$,
b. $\sup_{t \in \mathcal{T}} |\hat{\gamma}_j(t) - \gamma_j(t)| = \mathcal{O}_P(n^{-1/2})$
c. $\sup_{t \in \mathcal{T}} |\hat{\lambda}(t) - \lambda(t)| = \mathcal{O}_P(n^{-1/2})$,
d. $\sup_{t \in \mathcal{T}} |\hat{\Psi}_j(t) - \Psi_j(t)| = \mathcal{O}_P(n^{-1/2})$,
e. $\sup_{t \in \mathcal{T}} |\hat{G}_{ij}(t) - G_{ij}(t)| = \mathcal{O}_P(n^{-1/2})$, *and*
f. $|\hat{A}_{ij} - A_{ij}| = \mathcal{O}_P(n^{-1/2})$.

The asymptotic results for the cross-component transports then follow immediately from the rates established in Theorem 1.

Theorem 2. *Under assumptions of Theorem 1 for $i = 1, \ldots, n,\ 1 \leq j, k \leq p$,*

a. $\sup_{t \in \mathcal{T}} |\hat{T}_{jk}(t) - T_{jk}(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2})$, *and*
b. $\sup_{t \in \mathcal{T}} |\hat{T}_{jk}^{(i)}(t) - T_{jk}^{(i)}(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2})$.

A similar corollary for cross-component transports follows in the case of fully observed curves or dense enough designs.

Corollary 2. *Suppose the penalty parameters $\eta_1 \sim \eta_2 = \mathcal{O}(n^{-1})$. If the random trajectories are fully observed without error or are recorded with at least a multiple of $m \sim n^{\Delta(1-\delta)^{-1}}$ observations per curve, with $\Delta > 3$, then under the assumptions of Theorem 1, we have for $i = 1, \ldots, n,\ 1 \leq j, k \leq p$,*

a. $\sup_{t \in \mathcal{T}} |\hat{T}_{jk}(t) - T_{jk}(t)| = \mathcal{O}_P(n^{-1/2})$, *and*
b. $\sup_{t \in \mathcal{T}} |\hat{T}_{jk}^{(i)}(t) - T_{jk}^{(i)}(t)| = \mathcal{O}_P(n^{-1/2})$.

Corollaries 1 and 2 suggest that, on dense enough measurement schedules, parametric rates of convergence are achievable for the quantities described by the

LTM. The collection of cross-component transports also exhibit the following algebraic structure.

Remark 1. *For any cycle of components indexed by the sequence,*

$$\pi_1 \to \pi_2 \to \pi_3 \to \cdots \to \pi_L \to \pi_1,$$

with arbitrary length L *and* $\pi_1, \ldots, \pi_L \in \{1, \ldots, p\}$*, their respective cross-component transports satisfy*

$$T_{\pi_1\pi_2} \circ T_{\pi_2\pi_3} \circ \cdots \circ T_{\pi_L\pi_1} = id.$$

This result ensures that the system of cross-component transport maps prevents inconsistencies within itself. For example, if for three components A, B, and C, the pairwise transports T_{AB} and T_{BC} suggest that Component A tends to precede Component B which itself tends to precede Component C, this implies that the transport T_{AC} must indicate that Component A tends to precede Component C. Furthermore, mapping a component tempo through other components and then back to itself will result in the original component tempo, unchanged. Next we consider rates of convergence for the reconstructed curves as per Eq. (2.28), putting all model components together.

Theorem 3. *Under assumptions of Theorem 1 for* $i = 1, \ldots, n,\ j = 1, \ldots, p$*,*

$$\sup_{t \in \mathcal{T}} |\hat{X}_{ij}(t) - X_{ij}(t)| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2}).$$

Again a parametric rate is achievable on dense enough designs.

Corollary 3. *Suppose the penalty parameters* $\eta_1 \sim \eta_2 = \mathcal{O}(n^{-1})$*. If the random trajectories are fully observed without error or the trajectories are recorded with at least a multiple of* $m \sim n^{\Delta(1-\delta)^{-1}}$ *observations per curve, with* $\Delta > 3$*, then under the assumptions of Theorem 1, we have for* $i = 1, \ldots, n,\ j = 1, \ldots, p$*,*

$$\sup_{t \in \mathcal{T}} |\hat{X}_{ij}(t) - X_{ij}(t)| = \mathcal{O}_P(n^{-1/2}).$$

2.7 Concluding Remarks

In this chapter, we have introduced the Latent Transport Model (LTM) as a means of decomposing multivariate functional data and quantifying their intercomponent time dynamics. The model provides a simple representation for multivariate funcitonal data: information concerning their time dynamics is synthesized and compressed into two fixed effect terms (the latent curve and a collection of component-level warping functions) and two random effect terms (a random amplitude vector and a collection of subject-level warping functions). This representation requires the estimation of only one random warping function and amplitude vector per subject, in addition to $p+1$ deterministic functions overall.

In some cases these components may be reduced even further. For example, when subject-level warping is negligible or part of a pre-processing step, a special case of the model arises in which time-dynamics are fully characterized by the $p+1$ fixed effect curves and one random scalar per component. Alternatively, if subject-level time warping is present but further dimension reduction is desired, transformation of warps by the LQD transform (Petersen & Müller, 2016) or others (see, e.g. Happ et al. (2019)) will permit a Karhunen-Loève expansion in $\mathcal{L}^2-$space. Applying the LTM and truncating this expansion at an appropriate number of eigenfunctions, say K_0, creates a representation of multivariate functional data using only $p + K_0$ random scalars, as opposed to, for example, a standard MFPCA representation which requires $p \times K_0$ variables.

The LTM serves both as an extension of existing univariate functional warping methods, as well as a stepping stone for many new potential models for multivariate functional data analysis and registration. Future directions of note include implementing the LTM for alternative alignment algorithms besides pairwise warping, harnessing the flexibility of cross-component transport maps for imputation of components in partially observed multivariate functional data, or

relaxing structural assumptions to allow for more flexible functional relationships between different latent curves for distinct subsets of components. Spatiotemporal applications are also promising for the LTM, in which the vector function components are indexed by location. In such a situation, component warping functions may reveal time trends across geographic regions. In areas like spatiotemporal data analysis where the number of vector components is high, downstream application of dimension reduction techniques like multidimensional scaling may prove useful in comparing and understanding the mutual time dynamics across a large number of components.

2.8 Technical Proofs

2.8.1 Intermediate Lemmata

The following two Lemmata provide rates of convergence for the smoothing procedure and the pairwise warping estimates which are needed in the proofs of Theorems 1 and 2.

Lemma 1. *Under the conditions of Proposition 1, for any* $\delta \in (0,1)$ *if* $b \sim m^{-(1-\delta)/6}$ *for all* $i = 1, \ldots, n$, *and* $\limsup_{n\to\infty} nm^{-\delta/2} < \infty$, *then for all* $j = 1, \ldots, p$,

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{t\in\mathcal{T}} |\tilde{X}_{ij}(t) - X_{ij}(t)| = \mathcal{O}_P(\tau_m), \tag{2.34}$$

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{t\in\mathcal{T}} \left| \frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}(t)}{||X_{ij}||_\infty} \right| = \mathcal{O}_P(\tau_m), \tag{2.35}$$

where $\tau_m = m^{-(1-\delta)/3}$.

Lemma 2. *For all* $j = 1, \ldots, p$, *Under assumptions (L1), (L2), and (S0-S2), with bandwidth* $b(m) \sim m^{-1/6}$ *and* $\eta \to 0$, *the spline coefficients of the warping functions* $\tilde{\theta}^{(j)}_{V_{i'i}}$ *satisfy*

$$||\tilde{\theta}^{(j)}_{V_{i'i}} - \theta^{(j)}_{V_{i'i}}|| = \mathcal{O}_P(m^{-1/6}) + \mathcal{O}(\eta_1^{1/2}),$$

and the corresponding estimates of the pairwise warping function $\tilde{V}_{i'i}^{(j)}$ in (2.12) satisfy

$$\sup_{t\in\mathcal{T}}|\tilde{V}_{i'i}^{(j)}(t) - V_{i'i}^{(j)}(t)| = \mathcal{O}_P(m^{-1/6}) + \mathcal{O}(\eta_1^{1/2}).$$

2.8.2 Proofs of Theoretical Results

Throughout these proofs we write C, C', C'' and so on to denote generic constants.

Proof of Prop. 1:

Let $(\Omega, \mathscr{F}, P)$ be the probability space on which the observed data $Y_{ijs} = X_{ij}(t_s) + \epsilon_{ijs}$, $i = 1, \ldots, n,\ j = 1, \ldots, p,\ s = 1, \ldots, m$, are defined, where Ω is the sample space, $\mathscr{F}$ is the σ-algebra of events, and $P : \mathscr{F} \to [0, 1]$ is the probability measure. Because of the independence between (A_{ij}, H_i), the sources of amplitude and phase variation, and ϵ_{ijs}, the measurement errors, the probability space $(\Omega, \mathscr{F}, P)$ can be considered as a product space of two probability spaces, $(\Omega_1, \mathscr{F}_1, P_1)$ where the amplitude factors, warping functions, and hence the multivariate processes, $\mathbf{X}$, are defined, and $(\Omega_2, \mathscr{F}_2, P_2)$ where the measurement errors are defined. Fixing an element $\omega_1 \in \Omega_1$ corresponds to a realization of the amplitude factor A, the warping function H, and the multivariate process $\mathbf{X}$. We note that the errors $\{\epsilon_{ijs}\}_{s=1}^m$ are independent in $(\Omega_2, \mathscr{F}_2, P_2)$. Hence, we may substitute the index i in the subscripts of $A_{ij}, H_i, X_{ij}, Y_{ijs}$ and ϵ_{ijs} by ω_1; specifically, we use $A_{\omega_1,j}, H_{\omega_1}, X_{\omega_1,j}, Y_{\omega_1,js}$ and $\epsilon_{\omega_1,js}$, respectively, in the equations that follow.

For any fixed $\omega_1 \in \Omega_1$, $\{\epsilon_{\omega_1,js}\}_{s=1}^m$ are i.i.d. realizations of $\epsilon_{\omega_1,j}$, with $Y_{\omega_1,j}(t_s) = X_{\omega_1,j}(t_s) + \epsilon_{\omega_1,js}$. To show the rates of convergence it will be useful to frame local linear smoothing as a special case of local Fréchet regression so we can apply Theorem 2 of Chen & Müller (2020).

To satisfy one of the assumptions of this theorem, we must restrict the range of our curves to a bounded set. Assumption (L2) guarantees that for arbitrarily small $\rho > 0$, there exist constants $0 < c_\rho < C_\rho < \infty$ and an event $Q_\rho := \{c_\rho < \sup_{1\le j\le p} \sup_{1\le i\le n} A_{ij}(t) < C_\rho\}$ such that $P(Q_\rho) > 1 - \rho$. When the event Q_ρ holds, the

range of $X_{ij}(t)$ is the bounded set $R^* = [c_\rho, C_\rho]$ for all $j = 1, \ldots, p$. We may then ignore the event Q_ρ^C which happens with arbitrarily small probability.

Next it is necessary to explicitly define the objective functions involved local linear estimation. For any $t \in \mathcal{T}$, the unobserved true process $X_{\omega_1,j}$ satisfies,

$$\begin{aligned} X_{\omega_1,j}(t) &= E(Y_{\omega_1,j}(t)|X_{\omega_1,j}) = \underset{x \in R^*}{\operatorname{argmin}}\ M_{\omega_1}^{(j)}(x,t), \\ M_{\omega_1}^{(j)}(x,t) &= E_{\Omega_2}\left[(Y_{\omega_1,j}(t) - x)^2 | X_{\omega_1,j}\right]. \end{aligned} \tag{2.36}$$

Furthermore, the target of local linear estimation is,

$$\beta_{\omega_1,j}(t) = (\beta_{\omega_1,j}^{(0)}(t), \beta_{\omega_1,j}^{(1)}(t))^T \tag{2.37}$$

which minimizes

$$\check{M}_{\omega_1,b}^{(j)}(\beta_j, t) = E_{\Omega_2}\left[(Y_{\omega_1,j} - Z(t)\beta_j)^T W (Y_{\omega_1,j} - Z(t)\beta_j) | X_{\omega_1,j}\right], \tag{2.38}$$

where $Z(t) = \begin{bmatrix} 1 & t_0 - t \\ 1 & t_1 - t \\ \vdots & \\ 1 & t_m - t \end{bmatrix}$, $W = diag(K_b(t_s - t))$ for $s = 1, \ldots, m$, $K_b(\cdot) = b^{-1}K(\cdot/b)$, and $\check{X}_{\omega_1,j}(t) = \beta_{\omega_1,j}^{(0)}(t)$. Finally, the local linear estimates can be expressed through the empirical version,

$$\hat{\beta}_{\omega_1,j} = (\hat{\beta}_{\omega_1,j}^{(0)}, \hat{\beta}_{\omega_1,j}^{(1)})^T \tag{2.39}$$

which minimizes

$$\tilde{M}_{\omega_1,m}^{(j)}(\beta_{\omega_1,j}, t) = (Y_{\omega_1,j} - Z(t)\beta_{\omega_1,j})^T W (Y_{\omega_1,j} - Z(t)\beta_{\omega_1,j}). \tag{2.40}$$

with the final smoothing estimate taken as $\tilde{X}_{\omega_1,j}(t) = \hat{\beta}_{\omega_1,j}^{(0)}$.

With these quantities defined, the proof reduces to verifying the following conditions:

(i) The marginal density f is bounded away from zero on $\mathcal{T}$ and the second derivative is bounded, $\sup_{t\in\mathcal{T}^\circ}|f''(t)| < \infty$.

(ii) For all $\omega_1 \in \Omega_1$ and $t \in \mathcal{T}$, the minimizers $X_{\omega_1,j}(t)$, $\beta_{\omega_1,j}$ and $\hat{\beta}_{\omega_1,j}$ exist and are unique, the last P_{Ω_2}-almost surely. In addition for any $\epsilon > 0$,

$$\inf_{\omega_1\in\Omega_1,t\in\mathcal{T}} \inf_{|X_{\omega_1,j}(t)-x|>\epsilon} \left(M^{(j)}_{\omega_1}(x,t) - M^{(j)}_{\omega_1}(X_{\omega_1,j}(t),t)\right) > 0,$$

$$\liminf_{b\to 0} \inf_{\omega_1\in\Omega_1,t\in\mathcal{T}} \inf_{||\beta_{\omega_1,j}-\beta_j||>\epsilon} \left(\check{M}^{(j)}_{\omega_1,b}(\beta_j,t) - \check{M}^{(j)}_{\omega_1,b}(\beta_{\omega_1,j},t)\right) > 0.$$

(iii) There exist constants $r_{1j}, r_{2j} > 0,\ C_{1j}, C_{2j} > 0$, such that for $j = 1,\ldots,p$,

$$\inf_{\omega_1\in\Omega_1,t\in\mathcal{T}} \inf_{|X_{\omega_1,j}(t)-x|<r_{1j}} \left[M^{(j)}_{\omega_1}(x,t) - M^{(j)}_{\omega_1}(X_{\omega_1,j}(t),t) - C_{1j}|x - X_{\omega_1,j}(t)|^2\right] > 0,$$

$$\liminf_{b\to 0} \inf_{\omega_1\in\Omega_1,t\in\mathcal{T}} \inf_{||\beta_{\omega_1,j}-\beta||<r_{2j}} \left[\check{M}^{(j)}_{\omega_1,b}(\beta_j,t) - \check{M}^{(j)}_{\omega_1,b}(X_{\omega_1,j}(t),t) - C_{2j}||\beta_j - \beta_{\omega_1,j}||^2\right] > 0.$$

(iv) Let $B_r(X_{\omega_1,j}(t)) \subset R_C$ be a ball of radius r centered at $X_{\omega_1,j}(t)$ and denote be its covering number using balls of radius ϵ as $N(r\epsilon, B_r(X_{\omega_1,j}(t)), ||\cdot||_\infty)$. Then

$$\sup_{r>0}\sup_{\omega_1\in\Omega_1} \int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N(r\epsilon, B_r(X_{\omega_1,j}(t)), ||\cdot||_\infty)}d\epsilon < \infty.$$

First, (i) is satisfied by assumption (S0). Then for the remaining conditions, without loss of generality, fix j and apply the same argument for each component. Conditions (ii) and (iii) rely on the following observations and similar calculations. Since $M^{(j)}_{\omega_1}(x,t)$ and $\check{M}^{(j)}_{\omega_1,b}(x,t)$ are concave for all $t \in \mathcal{T}$, $\tilde{M}^{(j)}_{\omega_1,m}(x,t)$, is P_{Ω_2}-almost surely concave for all $t \in \mathcal{T}$, their respective minimizers exist and are unique. Then we verify that the minimum difference in objective functions outside of an $\epsilon-$neighborhood of the minimizer is positive.

By independence of $X_{\omega_1,j}$ and $\varepsilon_{\omega_1,j}$, we have

$$\begin{aligned} M^{(j)}_{\omega_1}(x,t) &= E_{\Omega_2}[(Y_{\omega_1,j,b}(t)-x)^2|X_{\omega_1,j}(t)] \\ &= E_{\Omega_2}[(X_{\omega_1,j}(t)-x+\varepsilon_{\omega_1,j})^2|X_{\omega_1,j,b}(t)] \\ &= E_{\Omega_2}[(X_{\omega_1,j}(t)-x)^2|X_{\omega_1,j}(t)] + E_{\Omega_2}[\varepsilon^2_{\omega_1,j}|X_{\omega_1,j}(t)] \\ &= (X_{\omega_1,j}(t)-x)^2+\sigma_j^2, \end{aligned}$$

and

$$\begin{aligned} M^{(j)}_{\omega_1}(X_{\omega_1,j}(t),t) &= E_{\Omega_2}[(Y_{\omega_1,j}(t)-X_{\omega_1,j}(t))^2|X_{\omega_1,j}(t)] \\ &= E_{\Omega_2}[\varepsilon^2_{\omega_1,j}|X_{\omega_1,j}(t)] = \sigma_j^2. \end{aligned}$$

Together these imply

$$M^{(j)}_{\omega_1}(x,t) - M^{(j)}_{\omega_1}(X_{\omega_1,j}(t),t) = (X_{\omega_1,j}(t)-x)^2 \tag{2.41}$$

and so condition (ii) is satisfied as

$$\inf_{\omega_1\in\Omega_1, t\in\mathcal{T}} \inf_{|X_{\omega_1,j}(t)-x|>\epsilon} (X_{\omega_1,j}(t)-x)^2 > \epsilon^2 > 0.$$

Similarly, for the localized target, we have,

$$\begin{aligned} \check{M}^{(j)}_{\omega_1,b}(\beta_j,t) &= E_{\Omega_2}\left[(Y_{\omega_1,j}-Z(t)\beta_j)^T W(Y_{\omega_1,j}-Z(t)\beta_j)|X_{\omega_1,j}\right] \\ &= E_{\Omega_2}\left[\sum_{s=1}^m K(\frac{t_s-t}{b})(Y_{\omega_1,js}-\beta_0-\beta_1(t-t_s))^2|X_{\omega_1,j}\right] \\ &= E_{\Omega_2}\left[\sum_{s=1}^m K(\frac{t_s-t}{b})(X_{\omega_1,js}-\beta_0-\beta_1(t-t_s))^2|X_{\omega_1,j}\right] \\ &\qquad + E_{\Omega_2}\left[\sum_{s=1}^m K(\frac{t_s-t}{b})\varepsilon^2_{\omega_1,js}|X_{\omega_1,j}\right] \\ &= (X_{\omega_1,j}-Z(t)\beta_j)^T W(X_{\omega_1,j}-Z(t)\beta_j) + \sigma_j^2\sum_{s=1}^m K(\frac{t_s-t}{b}), \end{aligned}$$

and

$$\begin{aligned}
\check{M}^{(j)}_{\omega_1,b}(\beta_{\omega_1,j},t) &= E_{\Omega_2}\left[(Y_{\omega_1,j}-Z(t)\beta_{\omega_1,j})^T W(Y_{\omega_1,j}-Z(t)\beta_{\omega_1,j})|X_{\omega_1,j}\right] \\
&= E_{\Omega_2}\left[\sum_{s=1}^{m} K(\frac{t_s-t}{b})(X_{\omega_1,js}-\beta^{(0)}_{\omega_1,j}-\beta^{(1)}_{\omega_1,j}(t-t_s))^2|X_{\omega_1,j}\right] \\
&\qquad + E_{\Omega_2}\left[\sum_{s=1}^{m} K(\frac{t_s-t}{b})\varepsilon^2_{\omega_1,js}|X_{\omega_1,j}\right] \\
&= (X_{\omega_1,j}-Z(t)\beta_{\omega_1,j})^T W(X_{\omega_1,j}-Z(t)\beta_{\omega_1,j}) + \sigma_j^2\sum_{s=1}^{m} K(\frac{t_s-t}{b}).
\end{aligned}$$

Together these give

$$\check{M}^{(j)}_{\omega_1,b}(\beta,t)-\check{M}^{(j)}_{\omega_1,b}(\beta_{\omega_1,j},t) = [Z(t)(\beta_{\omega_1,j}-\beta_j)]^T W[Z(t)(\beta_{\omega_1,j}-\beta_j)] \tag{2.42}$$

$$= ||W^{1/2}Z(t)(\beta_{\omega_1,j}-\beta)||^2, \tag{2.43}$$

and thus

$$\liminf_{b\to 0}\inf_{\omega_1\in\Omega_1,t\in\mathcal{T}}\inf_{||\beta_{\omega_1,j}-\beta||>\epsilon}\check{M}^{(j)}_{\omega_1,b}(\beta,t)-\check{M}^{(j)}_{\omega_1,b}(\beta_{\omega_1,j},t) =$$

$$\liminf_{b\to 0}\inf_{\omega_1\in\Omega_1,t\in\mathcal{T}}\inf_{||\beta_{\omega_1,j}-\beta||>\epsilon}||W^{1/2}Z(t)(\beta_{\omega_1,j}-\beta_j)||^2 > \liminf_{b\to 0}||W^{1/2}Z(t)||^2\epsilon^2 > 0.$$

so condition (ii) is met. Equations (2.41) and (2.43) also imply that the inequality in (iii) holds for $C_{1j}=1$, and $C_{2j}=||W^{1/2}Z(t)||^2$.

Finally, condition (iv) is met since by Lemma 2.7.8 of Van der Vaart & Wellner (1996), there exists some constant $C>0$,

$$\log N(r\epsilon, B_r(X_{\omega_1,j}(t)), ||\cdot||_\infty) \le C\epsilon^{-1}$$

where C depends only on the bounded set of $\mathbb{R}$ on which $X_{\omega_1,j}$ takes values and is thus uniform over $\omega_1\in\Omega_1$. Then, for any $r>0$, we may bound the integral by

$$\begin{aligned}
&\sup_{\omega_1\in\Omega_1}\int_0^1 \sup_{t\in\mathcal{T}}\sqrt{1+\log N(r\epsilon, B_r(X_{\omega_1,j}(t)), ||\cdot||_\infty)}d\epsilon \\
&\qquad = \sup_{\omega_1\in\Omega_1}\int_0^1 \sqrt{1+C\epsilon^{-1}}d\epsilon \le 1+2\sqrt{C}<\infty,
\end{aligned}$$

whence (iv) holds. This concludes the proof.

Proof of Prop. 2: The argument to extend the convergence result to the normalized curves is outlined as follows. We have

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| &\leq \sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{\tilde{X}_{ij}(t)}{||X_{ij}||_\infty}\right|+\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||X_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| \\ &:= I+II.\end{aligned}$$

The difference in sup-norms can be bounded using the reverse triangle inequality,

$$\begin{aligned}\left|||\tilde{X}_{ij}||_\infty-||X_{ij}||_\infty\right| &\leq ||\tilde{X}_{ij}-X_{ij}||_\infty \\ &= \mathcal{O}_P(m^{-1/3})\end{aligned}$$

Note also that this implies the rate for the amplitude factors,

$$|\hat{A}_{ij}-A_{ij}|=\left|||\tilde{X}_{ij}||_\infty-||X_{ij}||_\infty\right|=\mathcal{O}_P\left(m^{-1/3}\right).$$

Then we bound I as follows, using the fact that $A_{ij}^{-1}=1/||X_{ij}(t)||_\infty=\mathcal{O}_P(1)$,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{\tilde{X}_{ij}(t)}{||X_{ij}||_\infty}\right| &= \sup_{t\in\mathcal{T}}\left|\frac{(||X_{ij}||_\infty-||\tilde{X}_{ij}||_\infty)\tilde{X}_{ij}(t)}{||X_{ij}||_\infty||\tilde{X}_{ij}||_\infty}\right| \\ &\leq \frac{\left|||X_{ij}||_\infty-||\tilde{X}_{ij}||_\infty\right|}{||X_{ij}||_\infty} \\ &= \mathcal{O}_P\left(m^{-1/3}\right).\end{aligned}$$

For II, we have

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||X_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| &\leq ||\tilde{X}_{ij}(t)-X_{ij}(t)||_\infty \\ &= \mathcal{O}_P\left(m^{-1/3}\right).\end{aligned}$$

And finally,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| &\leq I+II \\ &= \mathcal{O}_P\left(m^{-1/3}\right).\end{aligned}$$

Proof of Lemma 1: Under a similar notation used in the proof of Proposition 1, where $\tilde{X}_{ij}$ is as per (2.40) and denotes the smoothing estimator and $\check{X}_{ij}$ is as per (2.38) and denotes the local target of $\tilde{X}_{ij}$, we bound the sup-norm between the smoothed function and the true curve by

$$n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\tilde{X}_{ij}(t)| \leq n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)|+n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{X}_{ij}(t)-\check{X}_{ij}(t)|,$$

and turn our attention to each of these terms separately, the first corresponding to the bias part and the latter to the stochastic part.

For the bias, observing that

$$\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)| \leq \sup_{1\leq i\leq n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)|,$$

and we may immediately apply Theorem 2 of Chen & Müller (2020) to see that

$$\sup_{1\leq i\leq n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)| = \mathcal{O}(b^2), \tag{2.44}$$

and thus

$$n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)| = \sup_{1\leq i\leq n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\check{X}_{ij}(t)| = \mathcal{O}(b^2).$$

For the stochastic part, letting $a_m = b\sqrt{m}$ we observe that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\check{X}_{ij}(t)-\tilde{X}_{ij}(t)| > C(a_m m^{-\delta/2})^{-1}\right)\\
&\quad\leq \limsup_{n\to\infty}\sum_{i=1}^{n} P\left(\sup_{t\in\mathcal{T}}|\check{X}_{ij}(t)-\tilde{X}_{ij}(t)| > C(a_m m^{-\delta/2})^{-1}\right)\\
&\quad\leq \limsup_{n\to\infty}\sum_{i=1}^{n}\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_m m^{-\delta/2}\sup_{t\in\mathcal{T}}|\check{X}_{ij}(t)-\tilde{X}_{ij}(t)| > C\right).
\end{aligned}$$

And so the stochastic term will converge at the rate $a_m^{-1}m^{\delta/2} = b^{-1}m^{(\delta-1)/2}$ if we show,

$$\limsup_{n\to\infty}\sum_{i=1}^{n}\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_m m^{-\delta/2}\sup_{t\in\mathcal{T}}|\check{X}_{ij}(t)-\tilde{X}_{ij}(t)| > C\right) \to 0, \quad \text{as} \quad C\to\infty. \tag{2.45}$$

Again following the argument in Chen & Müller (2020) with $\beta_1 = \beta_2 = 2$ results in the following bound for any $q \in \mathbb{N}^+$,

$$\limsup_{n\to\infty} \sum_{i=1}^{n} \sup_{\omega_1 \in \Omega_1} P_{\Omega_2}\left(a_m m^{-\delta/2} \sup_{t\in\mathcal{T}} |\check{X}_{ij}(t) - \tilde{X}_{ij}(t)| > 2^q \right) \tag{2.46}$$

$$\leq 4C \sum_{k>q} 2^{-k} \limsup_{n\to\infty} m^{-\delta/2} + 4C \sum_{k>q} 2^{-k} \limsup_{n\to\infty} n m^{-\delta/2} \tag{2.47}$$

thus the probability tends to 0 as $q \to \infty$ uniformly over $\omega_1 \in \Omega_1$ if $\limsup_{n\to\infty} n m^{-\delta/2} < \infty$, whence (2.45) follows. Then with $b \sim m^{-(1-\delta)/6}$, we have $b^2 \sim b^{-1} m^{(\delta-1)/2} \sim m^{-(1-\delta)/3}$ and we have shown (2.34). Lastly, (2.35) can be seen by following the same reverse triangle inequality argument used in Proposition 2 in the context of the rate found in Lemma 1.

Proof of Lemma 2: Without loss of generality, consider a fixed component j. Define the unpenalized cost function

$$\mathscr{C}_{\gamma_j}(\theta, H_{i'}, H_i) = \int_{\mathcal{T}} d^2 \left[(\gamma_j \circ H_{i'})[\theta^T \alpha(t)], (\gamma_j \circ H_i)(t) \right] dt, \quad j = 1, \ldots, p,$$

where in our case $d = d(f, g) = (f - g)(t)$. Then the arguments of Theorem 3 of Chen & Müller (2020) imply that, for some positive constants $C, C', C'' < \infty$,

$$\begin{aligned} ||\tilde{\theta}^{(j)}_{V_{i'i}} - \theta^{(j)}_{V_{i'i}}|| &\leq C \sup_{\theta\in\Theta} |\mathscr{C}_\eta(\theta, H'_i, H_i) - \mathscr{C}_{\gamma_j}(\theta, H'_i, H_i)|^{1/2} \\ &\leq C' \left[\sup_{t\in\mathcal{T}} |\tilde{X}_{ij}(t) - X_{ij}(t)|^{1/2} + \sup_{t\in\mathcal{T}} |\tilde{X}_{i'j}(t) - X_{i'j}(t)|^{1/2} \right] + C'' \eta_1^{1/2} \\ &= \mathcal{O}_P(m^{-1/6}) + \mathcal{O}(\eta_1^{1/2}). \end{aligned}$$

Then since $\sup_{t\in\mathcal{T}} |\alpha_\ell(t)| \leq 1$ for all $\ell = 1, \ldots, L+1$, it follows that

$$\sup_{t\in\mathcal{T}} |\tilde{V}^{(j)}_{i'i}(t) - V^{(j)}_{i'i}(t)| = \sup_{t\in\mathcal{T}} |(\tilde{\theta}^{(j)}_{V_{i'i}} - \theta^{(j)}_{V_{i'i}})^T \alpha(t)| \leq ||\tilde{\theta}^{(j)}_{V_{i'i}} - \theta^{(j)}_{V_{i'i}}|| \tag{2.48}$$

$$= \mathcal{O}_P(m^{-1/6}) + \mathcal{O}(\eta_1^{1/2}). \tag{2.49}$$

Proof of Theorem 1:

a. We seek a uniform rate of convergence of the individual distortion estimates $\hat{H}_i$. From the definition of $\hat{H}_i$, we observe,

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{H}_i(t) - H_i(t)| &= \sup_{t\in\mathcal{T}}|p^{-1}\sum_{j=1}^{p}(\tilde{H}_i^{(j)}(t) - H_i(t))| \\ &\leq p^{-1}\sum_{j=1}^{p}\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t) - H_i(t)| \\ &\leq \sup_{1\leq j\leq p}\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t) - H_i(t)|.\end{aligned}$$

Then for each $j = 1, \ldots, p$,

$$\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t) - H_i(t)| \leq \frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{V}_{i'i}^{(j)}(t) - V_{i'i}^{(j)}(t)| + \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}V_{i'i}^{(j)}(t) - H_i(t)\right|.$$

For the first term, an application of Lemmas 1 and 2 gives,

$$\begin{aligned}\frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{V}_{i'i}^{(j)}(t) - V_{i'i}^{(j)}(t)| &\leq \frac{1}{n}\sum_{i'=1}^{n}||\tilde{\theta}_{i'i}^{(j)} - \theta_{i'i}^{(j)}|| \\ &\leq C\left[\sup_{t\in\mathcal{T}}|\tilde{X}_{ij}(t) - X_{ij}(t)|^{1/2} + \frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{X}_{i'j}(t) - X_{i'j}(t)|^{1/2} + \eta_1^{1/2}\right] \\ &= \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\eta_1^{1/2}).\end{aligned}$$

For the second term, by Theorem 2.7.5 of Van der Vaart & Wellner (1996),

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}V_{i'i}^{(j)}(t) - H_i(t)\right| &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}H_{i'}^{-1}(t) - t\right| \\ &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}H_{i'}^{-1}(t) - EH_{i'}^{-1}(t)\right| = \mathcal{O}_P(n^{-1/2}),\end{aligned}$$

and therefore

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{H}_i(t) - H_i(t)| &\leq p\sup_{1\leq j\leq p}\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t) - H_i(t)| \\ &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\eta_1^{1/2}),\ i = 1, \ldots, n.\end{aligned}$$

b. For all $j = 1, \ldots, p$, we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{\gamma}_j(t) - \gamma_j(t)| &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{X}_{ij}\circ\hat{H}_i^{-1}(t)}{||\tilde{X}_{ij}||_\infty} - \gamma(t)\right| \\
&\leq \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\tilde{X}_{ij}\circ\hat{H}_i^{-1}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}\circ\hat{H}_i^{-1}(t)}{||X_{ij}||_\infty}\right]\right| \\
&\quad + \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ij}\circ\hat{H}_i^{-1}(t)}{||X_{ij}||_\infty} - \gamma(t)\right|.
\end{aligned}$$

The first term relies on the uniform rate of the smoother,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\tilde{X}_{ij}\circ\hat{H}_i^{-1}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}\circ\hat{H}_i^{-1}(t)}{||X_{ij}||_\infty}\right]\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}\circ\hat{H}_i^{-1}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}\circ\hat{H}_i^{-1}(t)}{||X_{ij}||_\infty}\right| \\
&= \frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty} - \frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| \\
&= \mathcal{O}_P(m^{-(1-\delta)/3}),
\end{aligned}$$

by Lemma 1.

The bounding argument for the second term is as follows and uses the Lipschitz continuity of λ and Ψ_j,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ij}\circ\hat{H}_i^{-1}(t)}{||X_{ij}||_\infty} - \gamma_j(t)\right| &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_{ij}\lambda\circ\Psi_j\circ H_i\circ\hat{H}_i^{-1}(t)}{A_{ij}} - (\lambda\circ\Psi_j)(t)\right]\right| \\
&= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[(\lambda\circ\Psi_j)(H_i(t)) - (\lambda\circ\Psi_j)(\hat{H}_i(t))\right]\right| \\
&\leq \frac{C}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\hat{H}_i(t) - H_i(t)| \\
&\leq C'\sup_{1\leq j\leq p}\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t) - H_i(t)|.
\end{aligned}$$

The rate of this last term follows by a similar argument used in part a. Specifi-

cally, we observe that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{H}_i^{(j)}(t)-H_i(t)| &\leq \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\tilde{V}_{i'i}^{(j)}(t)-V_{i'i}^{(j)}(t)|\right.\\
&\qquad\left.+\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}V_{i'i}^{(j)}(t)-H_i(t)\right|\right]\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}||\tilde{\theta}_{i'i}^{(j)}-\theta_{i'i}^{(j)}||+\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}H_{i'}^{-1}(t)-t\right|\\
&\leq C\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|X_{ij}(t)-\tilde{X}_{ij}(t)|^{1/2}\\
&\quad+C\frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|X_{i'j}(t)-\tilde{X}_{i'j}(t)|^{1/2}+C\eta_1^{1/2}\\
&\quad+\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n}H_{i'}^{-1}(t)-t\right|\\
&= \mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2}),
\end{aligned}$$

where the final rates come from Lemma 1 and Theorem 2.7.5 of Van der Vaart & Wellner (1996). Then altogether,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{\gamma}_j(t)-\gamma_j(t)| &= \mathcal{O}_P(m^{-(1-\delta)/3})+\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2})\\
&= \mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2}).
\end{aligned}$$

c. The argument for the latent curve mirrors the proof of part b., as we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}}|\hat{\lambda}(t)-\lambda(t)| = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\tilde{Z}_i\circ\hat{D}_i^{-1}(t)}{||\tilde{Z}_i||_\infty}-\lambda(t)\right|\\
&\leq \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\tilde{Z}_i\circ\hat{D}_i^{-1}(t)}{||\tilde{Z}_i||_\infty}-\frac{Z_i\circ\hat{D}_i^{-1}(t)}{||Z_i||_\infty}\right]\right|+\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i\circ\hat{D}_i^{-1}(t)}{||Z_i||_\infty}-\lambda(t)\right|.
\end{aligned}$$

Defining $\mathbf{1}_{ij}$ as the indicator that event I_{ij} occurs (that is, the j^{th} curve is selected as the representative curve for the i^{th} observation), the first term relies on the

uniform rate of the smoother,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\tilde{Z}_i\circ\hat{D}_i^{-1}(t)}{||\tilde{Z}_i||_\infty}-\frac{Z_i\circ\hat{D}_i^{-1}(t)}{||Z_i||_\infty}\right]\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{Z}_i\circ\hat{D}_i^{-1}(t)}{||\tilde{Z}_i||_\infty}-\frac{Z_i\circ\hat{D}_i^{-1}(t)}{||Z_i||_\infty}\right| \\
&= \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n}\mathbf{1}_{ij}\left(\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right|\right) \\
&\leq \frac{p}{n}\sup_{1\leq j\leq p}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}\left|\frac{\tilde{X}_{ij}(t)}{||\tilde{X}_{ij}||_\infty}-\frac{X_{ij}(t)}{||X_{ij}||_\infty}\right| \\
&= \mathcal{O}_P(m^{-(1-\delta)/3}),
\end{aligned}$$

where the rate is given by by Lemma 1 and the finiteness of p.

The second term is bounded as follows, using the Lipschitz continuity of λ and applying Lemma 2 and the argument used in Theorem 1 part b. for the uniform rate of $\hat{D}_i$,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i\circ\hat{D}_i^{-1}(t)}{||X_{ij}||_\infty}-\lambda(t)\right| &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_{ij}\lambda\circ D_i\circ\hat{D}_i^{-1}(t)}{A_{ij}}-\lambda(t)\right]\right| \\
&= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\lambda(D_i(t))-\lambda(\hat{D}_i(t))\right]\right| \\
&\leq \sup_{1\leq j\leq p}\frac{C}{n}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}}|\hat{D}_i(t)-D_i(t)| \\
&= \mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2}).
\end{aligned}$$

Then altogether,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{\lambda}(t)-\lambda(t)| &= \mathcal{O}_P(m^{-(1-\delta)/3})+\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2}) \\
&= \mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\eta_1^{1/2}).
\end{aligned}$$

d. Define the unpenalized objective function for the component warps

$$\mathscr{C}^*(\theta;\gamma_j,\lambda)=\int_{\mathcal{T}}(\gamma_j(t)-\lambda(\theta^T\alpha(t))^2dt. \tag{2.50}$$

Then a similar argument to that used in Theorem 3 of Chen & Müller (2020) bounds the discrepancy between estimated and true spline coefficients. Specifically, there exists a constant $C_0 > 0$ such that

$$||\tilde{\theta}_{\Psi_j} - \theta_{\Psi_j}||^2 \leq C_0 \ \left[\mathscr{C}^*(\tilde{\theta}_{\Psi_j}; \gamma_j, \lambda) - \mathscr{C}^*(\theta_{\Psi_j}; \gamma_j, \lambda) - \mathscr{C}_{\eta_2}(\tilde{\theta}_{\Psi_j}; \hat{\gamma}_j, \hat{\lambda}) + \mathscr{C}_{\eta_2}(\theta_{\Psi_j}; \hat{\gamma}_j, \hat{\lambda})\right]$$

Then we note that

$$\begin{aligned}
&\mathscr{C}^*(\tilde{\theta}_{\Psi_j}; \gamma_j, \lambda) - \mathscr{C}^*(\theta_{\Psi_j}; \gamma_j, \lambda) - \mathscr{C}_{\eta_2}(\tilde{\theta}_{\Psi_j}; \hat{\gamma}_j, \hat{\lambda}) + \mathscr{C}_{\eta_2}(\theta_{\Psi_j}; \hat{\gamma}_j, \hat{\lambda}) \\
&\qquad = \int_{\mathcal{T}} \left((\gamma_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)))^2 - (\hat{\gamma}_j(t) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)))^2\right) dt \\
&\qquad\qquad + \int_{\mathcal{T}} \left((\gamma_j(t) - \lambda(\theta_{\Psi_j}^T \alpha(t)))^2 - (\hat{\gamma}_j(t) - \hat{\lambda}(\theta_{\Psi_j}^T \alpha(t)))^2\right) dt \\
&\qquad\qquad + \eta_2 \left(\int_{\mathcal{T}} (\tilde{\theta}_{\Psi_j}^T \alpha(t)) - t)^2 - (\theta_{\Psi_j}^T \alpha(t)) - t)^2\right) dt \\
&\qquad := I + II + III,
\end{aligned}$$

where we let I, II, and III denote the three terms in the sum and bound each of them separately, using $C_1, C_2, ...$ to denote generic constants.

For I, we have

$$\begin{aligned}
I &= \int_{\mathcal{T}} \left((\gamma_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)))^2 - (\hat{\gamma}_j(t) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)))^2\right) dt \\
&= \int_{\mathcal{T}} \Bigg[\left(\gamma_j(t) - \hat{\gamma}_j(t) + \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right) \\
&\qquad\qquad \times \left(\gamma_j(t) + \hat{\gamma}_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right) \Bigg] dt \\
&\leq \left(\int_{\mathcal{T}} \left|\gamma_j(t) - \hat{\gamma}_j(t) + \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2} \\
&\qquad\qquad \times \left(\int_{\mathcal{T}} \left|\gamma_j(t) + \hat{\gamma}_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2}.
\end{aligned}$$

The first term in the product is straightforward to bound:

$$\begin{aligned}
&\left(\int_{\mathcal{T}} \left|\gamma_j(t) - \hat{\gamma}_j(t) + \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2} \\
&\leq \left(2\int_{\mathcal{T}} |\gamma_j(t) - \hat{\gamma}_j(t)|^2 dt + 2\int_{\mathcal{T}} \left|\hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2} \\
&\leq \left(C_1 ||\hat{\gamma}_j - \gamma_j||_\infty^2 + C_2 ||\hat{\lambda} - \lambda||_\infty^2\right)^{1/2} \\
&\leq C_3 \left(||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty\right)
\end{aligned}$$

The bound for the second term involves the same terms in addition to distance between estimated and true spline coefficients. The argument uses the fact that $\gamma_j(t) = \lambda(\Psi_j(t)) = \lambda(\theta_{\Psi_j}^T \alpha(t))$:

$$\begin{aligned}
&\left(\int_{\mathcal{T}} \left|\gamma_j(t) + \hat{\gamma}_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2} \\
&= \left(\int_{\mathcal{T}} \left|\hat{\gamma}_j(t) - \gamma_j(t) + \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t)) + 2(\gamma_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)))\right|^2 dt\right)^{1/2} \\
&\leq \left(2\int_{\mathcal{T}} \left|\hat{\gamma}_j(t) - \gamma_j(t) + \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt \right. \\
&\qquad\qquad \left. + 8\int_{\mathcal{T}} |\lambda(\theta_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))|^2 dt\right)^{1/2}
\end{aligned}$$

From here it just remains to bound the second term, which relies on the Lipschitz continuity of λ and the fact that and $\sup_t |\alpha_\ell(t)| \leq 1$ for all $\ell = 1, ..., L+1$:

$$\begin{aligned}
\int_{\mathcal{T}} |\lambda(\theta_{\Psi_j}^T \alpha(t)) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t))|^2 &\leq C_4 ||\theta_{\Psi_j}^T \alpha(t) - \tilde{\theta}_{\Psi_j}^T \alpha(t)||_\infty^2 \\
&\leq C_5 ||\theta_{\Psi_j} - \tilde{\theta}_{\Psi_j}||^2.
\end{aligned}$$

Thus the second term in the product is bounded as:

$$\begin{aligned}
&\left(\int_{\mathcal{T}} \left|\gamma_j(t) + \hat{\gamma}_j(t) - \lambda(\tilde{\theta}_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\tilde{\theta}_{\Psi_j}^T \alpha(t))\right|^2 dt\right)^{1/2} \\
&\leq C_6 \left(||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty + ||\tilde{\theta}_{\Psi_j} - \theta_{\Psi_j}||\right)
\end{aligned}$$

So altogether I is bounded by

$$\begin{aligned} I &\leq C_3 \left[||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty \right] \\ &\qquad \times C_6 \left[||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty + ||\theta_{\Psi_j} - \tilde{\theta}_{\Psi_j}|| \right] \\ &\leq C_7 \left(||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty \right)^2 \\ &\qquad + C_8 \left(||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty \right) \times ||\tilde{\theta}_{\Psi_j} - \theta_{\Psi_j}||. \end{aligned}$$

Next, II can be simplified, again by noting that $\gamma_j(t) = \lambda(\Psi_j(t)) = \lambda(\theta_{\Psi_j}^T \alpha(t))$ so that

$$\int_{\mathcal{T}} \left((\gamma_j(t) - \lambda(\theta_{\Psi_j}^T \alpha(t)))^2 - (\hat{\gamma}_j(t) - \hat{\lambda}(\theta_{\Psi_j}^T \alpha(t)))^2 \right) dt = \int_{\mathcal{T}} \left(\hat{\gamma}_j(t) - \hat{\lambda}(\theta_{\Psi_j}^T \alpha(t)) \right)^2 dt.$$

Then,

$$\begin{aligned} II &= \int_{\mathcal{T}} \left(\hat{\gamma}_j(t) - \lambda(\theta_{\Psi_j}^T \alpha(t)) + \lambda(\theta_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\theta_{\Psi_j}^T \alpha(t)) \right)^2 dt \\ &\leq 2 \int_{\mathcal{T}} \left(\hat{\gamma}_j(t) - \gamma_j(t) \right)^2 dt + 2 \int_{\mathcal{T}} \left(\lambda(\theta_{\Psi_j}^T \alpha(t)) - \hat{\lambda}(\theta_{\Psi_j}^T \alpha(t)) \right)^2 dt \\ &\leq C_9 ||\hat{\gamma}_j - \gamma_j||_\infty^2 + C_{10} ||\hat{\lambda} - \lambda||_\infty^2 \\ &\leq C_{11} \left(||\hat{\gamma}_j - \gamma_j||_\infty + ||\hat{\lambda} - \lambda||_\infty \right)^2. \end{aligned}$$

Lastly III is bounded as follows:

$$\begin{aligned} &\eta_2 \left(\int_{\mathcal{T}} (\tilde{\theta}_{\Psi_j}^T \alpha(t)) - t)^2 - (\theta_{\Psi_j}^T \alpha(t)) - t)^2 \right) dt \\ &\qquad \leq \eta_2 \left(\int_{\mathcal{T}} (\tilde{\theta}_{\Psi_j}^T \alpha(t)) - t)^2 + (\theta_{\Psi_j}^T \alpha(t)) - t)^2 \right) dt \\ &\qquad \leq 2\eta_2 \int_{\mathcal{T}} t^2 dt \\ &\qquad \leq 2\eta_2 |\mathcal{T}|^3 / 3. \end{aligned}$$

Then combining these bounds, we have

$$\begin{aligned}\mathscr{C}^*(\tilde{\theta}_{\Psi_j};\gamma_j,\lambda)-\mathscr{C}^*(\theta_{\Psi_j};\gamma_j,\lambda)-\mathscr{C}_{\eta_2}(\tilde{\theta}_{\Psi_j};\hat{\gamma}_j,\hat{\lambda})+\mathscr{C}_{\eta_2}(\theta_{\Psi_j};\hat{\gamma}_j,\hat{\lambda})&\\ \leq I+II+III&\\ \leq (C_7+C_{11})\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)^2&\\ +C_8\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)\times||\tilde{\theta}_{\Psi_j}-\theta_{\Psi_j}||+2\eta_2|\mathcal{T}|^3/3&\end{aligned}$$

which implies

$$\begin{aligned}||\tilde{\theta}_{\Psi_j}-\theta_{\Psi_j}||^2-C_0^2C_8\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)\times||\tilde{\theta}_{\Psi_j}-\theta_{\Psi_j}||&\\ \leq C_0\left[(C_7+C_{11})\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)^2+2\eta_2|\mathcal{T}|^3/3\right].&\end{aligned}$$

By completing the square,

$$\begin{aligned}\left|||\tilde{\theta}_{\Psi_j}-\theta_{\Psi_j}||-\frac{C_0C_8}{2}\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)\right|^2&\\ \leq C_0\left[(C_7+C_{11})\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)^2+2\eta_2|\mathcal{T}|/3\right]&\\ +\frac{C_0^2C_7^2}{4}\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)^2&\\ \leq C_{12}\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)^2+2C_0\eta_2|\mathcal{T}|^3/3,&\end{aligned}$$

and taking the square root, we have

$$\begin{aligned}\left|||\tilde{\theta}_{\Psi_j}-\theta_{\Psi_j}||-C^*\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)\right|&\\ \leq C_{13}\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)+C_{14}\eta_2^{1/2}.&\end{aligned}$$

Then by Theorem 1b. and 1c., the right hand side is such that

$$C_{13}\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)+C_{14}\eta_2^{1/2}=\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(\tau_m^{1/2})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2}),$$

and so is the term,

$$C^*\left(||\hat{\gamma}_j-\gamma_j||_\infty+||\hat{\lambda}-\lambda||_\infty\right)=\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(\tau_m^{1/2})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2}),$$

so we conclude that

$$||\tilde{\theta}_{\Psi_j} - \theta_{\Psi_j}|| = \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2}).$$

Finally, since $\sup_t |\alpha_\ell(t)| \leq 1$ for all $\ell = 1, ..., L+1$, the result for the component warps immediately follows, as:

$$\begin{aligned}
\sup_{t \in \mathcal{T}} |\hat{\Psi}_j(t) - \Psi_j(t)| &= \sup_{t \in \mathcal{T}} |(\tilde{\theta}_{\Psi_j}(t) - \theta_{\Psi_j})^T \alpha(t)| \\
&\leq ||\tilde{\theta}_{\Psi_j} - \theta_{\Psi_j}|| \\
&= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2}).
\end{aligned}$$

e. For the overall warping functions G_{ij}, we again break the difference into two parts which we already know the rates for and use the Lipschitz continuity of Ψ_j:

$$\begin{aligned}
\sup_{t \in \mathcal{T}} |\hat{G}_{ij}(t) - G_{ij}(t)| &= \sup_{t \in \mathcal{T}} |(\hat{\Psi}_j \circ \hat{H}_i)(t) - (\Psi_j \circ \hat{H}_i)(t) + (\Psi_j \circ \hat{H}_i)(t) - (\Psi_j \circ H_i)(t)| \\
&\leq \sup_{t \in \mathcal{T}} |(\hat{\Psi}_j \circ \hat{H}_i)(t) - (\Psi_j \circ \hat{H}_i)(t)| + \sup_{t \in \mathcal{T}} |(\Psi_j \circ \hat{H}_i)(t) - (\Psi_j \circ H_i)(t)| \\
&\leq \sup_{t \in \mathcal{T}} |\hat{\Psi}_j(t) - \Psi_j(t)| + C \sup_{t \in \mathcal{T}} |\hat{H}_i(t) - H_i(t)| \\
&= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\max(\eta_1, \eta_2)^{1/2})
\end{aligned}$$

where the rates follow from Theorem 1a. and 1d.

f. The amplitude factors' convergence rate was shown in the proof of Proposition 1.

Proof of Theorem 2:

a. Consider bounding the maximum difference between the estimated XCT and its target by

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{T}_{jk}(t) - T_{jk}(t)| &= \sup_{t\in\mathcal{T}}\left|(\hat{\Psi}_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\Psi_k)(t)\right| \\ &\leq \sup_{t\in\mathcal{T}}\left|(\hat{\Psi}_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\hat{\Psi}_k)(t)\right| \\ &\quad + \sup_{t\in\mathcal{T}}\left|(\Psi_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\Psi_k)(t)\right|\end{aligned}$$

The rate of the first term can be obtained in a straightforward manner by observing,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|(\hat{\Psi}_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\hat{\Psi}_k)(t)\right| &= \sup_{t\in\mathcal{T}}\left|(\hat{\Psi}_j^{-1}\circ\hat{\Psi}_k)(\hat{\Psi}_k^{-1}(t)) - (\Psi_j^{-1}\circ\hat{\Psi}_k)(\hat{\Psi}_k^{-1}(t))\right| \\ &= \sup_{t\in\mathcal{T}}\left|\hat{\Psi}_j^{-1}(t) - \Psi_j^{-1}(t)\right| \\ &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\max(\eta_1,\eta_2)^{1/2})\end{aligned}$$

where the rates are given by the argument found in Theorem 1d.

For the second term, we use the Lipschitz continuity of Ψ_j^{-1} to arrive at

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|(\Psi_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\Psi_k)(t)\right| &\leq C\sup_{t\in\mathcal{T}}\left|\hat{\Psi}_k(t) - \Psi_k(t)\right| \\ &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\max(\eta_1,\eta_2)^{1/2})\end{aligned}$$

again by Theorem 1d. Therefore, the cross-component transports' rate of convergence is

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{T}_{jk}(t) - T_{jk}(t)| &\leq \sup_{t\in\mathcal{T}}\left|(\hat{\Psi}_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\hat{\Psi}_k)(t)\right| \\ &\quad + \sup_{t\in\mathcal{T}}\left|(\Psi_j^{-1}\circ\hat{\Psi}_k)(t) - (\Psi_j^{-1}\circ\Psi_k)(t)\right| \\ &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\max(\eta_1,\eta_2)^{1/2}),\end{aligned}$$

for all $j,k = 1,\ldots,p$.

b. The argument for the subject-specific cross-component transports follows the same argument but instead involves the composition of G_{ij} and G_{ik}. Note that since Ψ_j and H_i are Lipschitz continuous, then $G_{ij} = \Psi_j \circ H_i$ is as well. Then we have,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{T}_{jk}^{(i)}(t) - T_{jk}^{(i)}(t)| &= \sup_{t\in\mathcal{T}}\Big|(\hat{G}_{ij}^{-1}\circ\hat{G}_{ik})(t) - (G_{ij}^{-1}\circ\hat{G}_{ik})(t) \\
&\qquad + (G_{ij}^{-1}\circ\hat{G}_{ik})(t) - (G_{ij}^{-1}\circ G_{ik})(t)\Big| \\
&\leq \sup_{t\in\mathcal{T}}\left|(\hat{G}_{ij}^{-1}\circ\hat{G}_{ik})(t) - (G_{ij}^{-1}\circ\hat{G}_{ik})(t)\right| \\
&\quad + \sup_{t\in\mathcal{T}}\left|(G_{ij}^{-1}\circ\hat{G}_{ik})(t) - (G_{ij}^{-1}\circ G_{ik})(t)\right| \\
&\leq \sup_{t\in\mathcal{T}}\left|\hat{G}_{ij}^{-1}(t) - G_{ij}(t)\right| + C\sup_{t\in\mathcal{T}}\left|\hat{G}_{ik}(t) - G_{ik}(t)\right| \\
&= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(m^{-(1-\delta)/6}) + \mathcal{O}(\max(\eta_1,\eta_2)^{1/2}),
\end{aligned}$$

for all $i = 1,\ldots,n,\ j,k = 1,\ldots,p$, by Theorem 1e. This completes the proof.

Proof of Theorem 3:

The reconstructed curves depends on the rates of convergence for the A_{ij}, λ, and G_{ij} terms, which are established in Theorem 1. We bound the sup-norm between the estimated and true curves in terms of the quantities I and II as follows,

$$\begin{aligned}
\sup_{t\in\mathcal{T}}|\hat{X}_{ij}(t) - X_{ij}(t)| &= \sup_{t\in\mathcal{T}}|\hat{A}_{ij}(\hat{\lambda}\circ\hat{G}_{ij})(t) - A_{ij}(\lambda\circ G_{ij})(t)| \\
&\leq \sup_{t\in\mathcal{T}}|\hat{A}_{ij}(\hat{\lambda}\circ\hat{G}_{ij})(t) - \hat{A}_{ij}(\lambda\circ G_{ij})(t)| \\
&\quad + \sup_{t\in\mathcal{T}}|\hat{A}_{ij}(\lambda\circ G_{ij})(t) - A_{ij}(\lambda\circ G_{ij})(t)| \\
&= \hat{A}_{ij}\sup_{t\in\mathcal{T}}|(\hat{\lambda}\circ\hat{G}_{ij})(t) - (\lambda\circ G_{ij})(t)| + \sup_{t\in\mathcal{T}}|(\hat{A}_{ij} - A_{ij})(\lambda\circ G_{ij})(t)| \\
&:= I + II.
\end{aligned}$$

Bounding I can be done using the familiar Lipschitz strategy used in the proofs

of Theorems 1 and 2:

$$\begin{aligned}\hat{A}_{ij}\sup_{t\in\mathcal{T}}|(\hat{\lambda}\circ\hat{G}_{ij})(t)-(\lambda\circ G_{ij})(t)| &= \hat{A}_{ij}\sup_{t\in\mathcal{T}}\Big|(\hat{\lambda}\circ\hat{G}_{ij})(t)-(\lambda\circ\hat{G}_{ij})(t)\\ &\qquad +(\lambda\circ\hat{G}_{ij})(t)-(\lambda\circ G_{ij})(t)\Big|\\ &\leq \hat{A}_{ij}\left(\sup_{t\in\mathcal{T}}|(\hat{\lambda}-\lambda)(t)|+C\sup_{t\in\mathcal{T}}|\hat{G}_{ij}(t)-G_{ij}(t)|\right)\end{aligned}$$

The estimated amplitude factor can be dealt with by adding and subtracting the true A_{ij} and using the triangle inequality:

$$\begin{aligned}&\hat{A}_{ij}\left(\sup_{t\in\mathcal{T}}|(\hat{\lambda}-\lambda)(t)|+C\sup_{t\in\mathcal{T}}|\hat{G}_{ij}(t)-G_{ij}(t)|\right)\\ &\quad\leq A_{ij}\left(\sup_{t\in\mathcal{T}}|(\hat{\lambda}-\lambda)(t)|+C\sup_{t\in\mathcal{T}}|\hat{G}_{ij}(t)-G_{ij}(t)|\right)+\\ &\qquad|\hat{A}_{ij}-A_{ij}|\left(\sup_{t\in\mathcal{T}}|(\hat{\lambda}-\lambda)(t)|+C\sup_{t\in\mathcal{T}}|\hat{G}_{ij}(t)-G_{ij}(t)|\right)\\ &\quad=\mathcal{O}_P(1)\left[\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2})\right]+\\ &\qquad\mathcal{O}_P(m^{-1/6})\left[\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2})\right]\\ &\quad=\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2}).\end{aligned}$$

Then II follows by noting that G_{ij} is a homeomorphism and $||\lambda||_\infty = 1$ by assumption.

$$\begin{aligned}\sup_{t\in\mathcal{T}}|(\hat{A}_{ij}-A_{ij})(\lambda\circ G_{ij})(t)| &\leq |\hat{A}_{ij}-A_{ij}|\sup_{t\in\mathcal{T}}|(\lambda\circ G_{ij})(t)|\\ &=|\hat{A}_{ij}-A_{ij}|=\mathcal{O}_P(m^{-1/6}).\end{aligned}$$

Combining these rates gives

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{X}_{ij}(t)-X_{ij}(t)| &= \mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2})+\mathcal{O}_P(m^{-1/6})\\ &=\mathcal{O}_P(n^{-1/2})+\mathcal{O}_P(m^{-(1-\delta)/6})+\mathcal{O}(\max(\eta_1,\eta_2)^{1/2})\end{aligned}$$

which completes the proof.

Proof of Corollaries 1-3: The arguments for all Corollaries 1-3 are identical so here we show Corollary 1 and omit the others.

If $\eta_1 \sim \eta_2 = \mathcal{O}(n^{-1})$ and $m(n) = \mathcal{O}(n^{\Delta(1-\delta)^{-1}})$, with $\Delta > 3$, then

$$\begin{aligned}\sup_{t\in\mathcal{T}}|\hat{H}_i(t) - H(t)| &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(\tau_m^{1/2}) + \mathcal{O}(\eta_1^{1/2}) \\ &= \mathcal{O}_P(n^{-1/2}) + \mathcal{O}_P(n^{-\Delta/6}) + \mathcal{O}_P(n^{-1/2}).\end{aligned}$$

The result follows by observing that $\Delta/6 > 1/2$ if $\Delta > 3$ and thus,

$$\sup_{t\in\mathcal{T}}|\hat{H}_i(t) - H(t)| = \mathcal{O}_P(n^{-1/2}).$$

This same argument holds for the terms in Theorems 1b.-e., 2, and 3. For Theorem 1f., the argument is very similar, as $\Delta(1-\delta)^{-1}/6 > \Delta/6 > 1/2$, so we have

$$\begin{aligned}|\hat{A}_{ij} - A_{ij}| &= \mathcal{O}_P(m^{-1/6}) \\ &= \mathcal{O}_P(n^{-\Delta(1-\delta)^{-1}/6}) \\ &= \mathcal{O}_P(n^{-1/2}),\end{aligned}$$

and the result follows.

References

Abdul-Wahab, S. A. (2001). Ier photochemical smog evaluation and forecasting of short-term ozone pollution levels with artificial neural networks. *Process Safety and Environmental Protection*, *79*(2), 117–128.

Banik, S. D., Salehabadi, S. M., & Dickinson, F. (2017). Preece-Baines model 1 to estimate height and knee height growth in boys and girls from Merida, Mexico. *Food and Nutrition Bulletin*, *38*(2), 182–195.

Brunel, N. J.-B., & Park, J. (2014). Removing phase variability to extract a mean shape for jugg trajectories. *Electron. J. Statist.*, *8*(2), 1848–1855.

Cameron, N., Tanner, J., & Whitehouse, R. (1982). A longitudinal analysis of the growth of limb segments in adolescence. *Annals of Human Biology*, *9*(3), 211–220.

Cardot, H., Ferraty, F., & Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, *45*, 11–22.

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P., Han, K., Ji, H., Lin, S., & Dubey, P. (2020a). fdapace: functional data analysis and empirical dynamics. *R package version 0.5.6*.

Carroll, C., Müller, H.-G., & Kneip, A. (2020b). Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics*.

Chen, Y., Gajardo, A., Fan, J., Zhong, Q., Dubey, P., Han, K., Bhattacharjee, S., & Müller, H.-G. (2020). *frechet: Statistical Analysis for Random Objects and Non-Euclidean Data*. R package version 0.2.0.
URL `https://CRAN.R-project.org/package=frechet`

Chen, Y., & Müller, H.-G. (2020). Uniform convergence of local fréchet regression and time warping for metric-space-valued trajectories. *arXiv preprint arXiv:2006.13548*.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *Annals of Applied Statistics*, *6*(4), 1588–1614.

Chiou, J.-M., Chen, Y.-T., & Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, *24*, 1571–1596.

Chiou, J.-M., Yang, Y.-F., & Chen, Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, *146*, 301–312.

Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, *109*(505), 411–423.

Dubin, J. A., & Müller, H.-G. (2005). Dynamical correlation for multivariate longitudinal data. *Journal of the American Statistical Association*, *100*(471), 872–881.

Eiben, O., Barabás, A., & Németh, Á. (2005). Comparison of growth, maturation, and physical fitness of hungarian urban and rural boys and girls. *Journal of Human Ecology*, *17*(2), 93–100.

Ferraty, F., & Vieu, P. (2006). *Nonparametric Functional Data Analysis.*. New York: Springer, New York.

Gasser, T., & Kneip, A. (1995). Searching for structure in curve samples. *Journal of the American Statistical Association*, *90*, 1179–1188.

Gasser, T., Kneip, A., Binding, A., Largo, R., Prader, A., & Molinari, L. (1989).

Flexible methods for nonparametric fitting of individual and sample growth curves. *Auxology*, *88*, 23–30.

Gasser, T., Kneip, A., Ziegler, P., Largo, R., & Prader, A. (1990). A method for determining the dynamics and intensity of average growth. *Annals of Human Biology*, *17*(6), 459–474.

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., & Molinari, L. (1984a). Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology*, *11*(5), 397–411.

Gasser, T., Köhler, W., Müller, H.-G., Kneip, A., Largo, R., Molinari, L., & Prader, A. (1984b). Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology*, *11*(5), 397–411.

Gasser, T., Müller, H.-G., Köhler, W., Molinari, L., & Prader, A. (1984c). Nonparametric regression analysis of growth curves. *Annals of Statistics*, *12*, 210–229.

Gervini, D. (2015). Warped functional regression. *Biometrika*, (pp. 000–000).

Gervini, D., & Gasser, T. (2004). Self-mode warping functions. *Journal of the Royal Statistical Society: Series B*, *66*, 959–971.

Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, *35*, 70–91.

Happ, C., & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, *113*, 649–659.

Happ, C., Scheipl, F., Gabriel, A.-A., & Greven, S. (2019). A general framework for multivariate functional principal component analysis of amplitude and phase variation. *Stat*, *8*(1), e220.

Härdle, W., & Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, *18*(1), 63–89.

Jacques, J., & Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis*, *71*, 92–106.

Jain, N. C., & Marcus, M. B. (1975). Central limit theorems for C(S)-valued random variables. *Journal of Functional Analysis*, *19*(3), 216–231.

Jane-Wang, Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its Application*, *3*, 257–295.

Kleffe, J. (1973). Principal components of random variables with values in a separable Hilbert space. *Statistics: A Journal of Theoretical and Applied Statistics*, *4*, 391–406.

Kneip, A., & Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, *20*, 1266–1305.

Kneip, A., & Gasser, T. (1995). Model estimation in nonlinear regression under shape invariance. *Annals of Statistics*, *23*, 551–570.

Kneip, A., & Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association*, *103*, 1155–1165.

Marron, J. S., Ramsay, J. O., Sangalli, L. M., & Srivastava, A. (2015). Functional data analysis of amplitude and phase variation. *Statistical Science*, *30*(4), 468–484.

Milani, S. (2000). Kinetic models for normal and impaired growth. *Annals of Human Biology*, *27*(1), 1–18.

Park, J., & Ahn, J. (2017). Clustering multivariate functional data with phase variation. *Biometrics*, *73*(1), 324–333.

Petersen, A., & Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, *44*(1), 183–218.

Petersen, A., & Müller, H.-G. (2019). Fréchet regression for random objects with euclidean predictors. *Annals of Statistics*, *47*(2), 691–719.

Preece, M., & Baines, M. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology*, *5*(1), 1–24.

Ramsay, J. O., Gribble, P., & Kurtek, S. (2014). Description and processing of functional data arising from jugg trajectories. *Electronic Journal of Statistics*, *8*(2), 1811–1816.

Ramsay, J. O., & Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society: Series B*, *60*(2), 351–363.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. New York: Springer, second ed.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B*, *53*(1), 233–243.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*, 43–49.

Sangalli, L. M., Secchi, P., Vantini, S., & Vitelli, V. (2010). K-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, *54*(5), 1219–1233.

Sheehy, A., Gasser, T., Molinari, L., & Largo, R. (1999). An analysis of variance of the pubertal and midgrowth spurts for length and width. *Annals of Human Biology*, *26*(4), 309–331.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, *57*(4), 673–689.

Srivastava, A., Klassen, E., Joshi, S. H., & Jermyn, I. H. (2010). Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(7), 1415–1428.

Srivastava, A., & Klassen, E. P. (2016). *Functional and shape data analysis*. Springer.

Tang, R., & Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, *95*, 875–889.

Tu, J., Xia, Z.-G., Wang, H., & Li, W. (2007). Temporal variations in surface ozone and its precursors and meteorological effects at an urban site in china. *Atmospheric Research*, *85*(3-4), 310–337.

Van der Vaart, A., & Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Villani, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society.

Wang, J., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*, 257–295.

Wang, K., & Gasser, T. (1997). Alignment of curves by dynamic time warping. *Annals of Statistics*, *25*(3), 1251–1276.

Yao, F., Müller, H.-G., & Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590.

Zhou, L., Huang, J., & Carroll, R. (2008). Joint model of paired sparse functional data using principal components. *Biometrika*, *95*, 601–619.