Extending Functional Principal Component Analysis to Model Weight Gain During Pregnancy

By

Lisa Shulman

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

BACKGROUND: Achieving an appropriate gestational weight gain (GWG) is important during pregnancy. Inadequate and excessive GWG have been linked to various negative pregnancy outcomes and future health issues for both the mother and fetus. One useful intervention is to provide individualized assessment and counselling at the beginning of pregnancy and throughout pregnancy as needed, to help women follow dietary and physical activity patterns that support optimal weight gain. Currently, pregnant women receive a personalized weight gain goal, usually a weight gain chart, that is based on their pre-pregnancy body mass index (BMI). However, these weight gain charts cannot be adjusted to account for individual weight measurements during pregnancy. Interventions focused on the progress of weight gain throughout pregnancy could thus benefit from a personalized weight growth trajectory.

A previous study used the functional principal component analysis (FPCA) approach and successfully estimated individual weight trajectories in pregnant women. However, the FPCA method borrows information from the whole cohort, and does not incorporate information on women's pre-pregnancy BMI. The objective of this thesis was to extend FPCA by incorporating additional BMI category-specific principal components (PCs) in trajectory modelling of gestational weight gain. The new proposed method, called JIVE–FPCA, applies the joint and individual variation explained (JIVE) algorithm to FPCA, and can be used to estimate individual trajectories from any sparse, longitudinal multi-block dataset.

METHODS: Weight data during pregnancy and early postpartum were collected from a large cohort ($n = 1648$) of pregnant and postpartum women. Longitudinal weight measurements were irregularly spaced and obtained from multiple data sources. Sparse, longitudinal multi-block data were simulated according to the JIVE–FPCA model, and closely mimicked the real, gestational weight data. The FPCA and JIVE–FPCA methods were then applied to both the simulated and real datasets. The performances of the two methods were compared on the basis of the mean squared error and average confidence bandwidth.

RESULTS: Studies with both simulated and real data showed that the JIVE–FPCA method provides a significantly better fit to individual trajectories than the original FPCA approach. This was evident both visually and numerically. The mean squared error and average confidence bandwidth were appreciably lower for the JIVE–FPCA method than for FPCA. In the application to gestational weight data, the JIVE–FPCA approach successfully captured differences in GWG patterns among different pre-pregnancy BMI categories. In particular, it was found that the weight trajectories for women with pre-pregnancy obesity were more gradually rising than those for women with lower pre-pregnancy BMI. The mean JIVE–FPCA weight trajectories for overweight and obese women were similar, and exhibited a different pattern than the trend displayed by the mean weight curves for underweight and normal weight women.

CONCLUSIONS: This thesis presents the JIVE–FPCA approach, which is an extension of the existing FPCA method. The advantage offered by JIVE–FPCA is that it can simultaneously capture patterns that are shared across multiple blocks of data and patterns that are specific to a particular block. Trajectory modelling of gestational weight gain indicated that the JIVE–FPCA method leads to a significant improvement in explaining the weight variation.

The new method also highlights the differences in GWG patterns between pre-pregnancy BMI classes.

To Meer

Thank you for your support through all the (countless) sleepless nights, and the ups and downs of research and classes. It's been a wild ride, and without you this thesis would not have been possible.

Acknowledgements

I would like to thank my supervisors Dr. Linglong Kong, Dr. Bei Jiang, and Dr. Yan Yuan at the University of Alberta for giving me the opportunity to perform this research. I appreciate their advice, support, and patience while conducting my research and writing this thesis.

I also gratefully acknowledge the funding received towards my MSc from my supervisors, the Department of Mathematical and Statistical Sciences at the University of Alberta, and from NSERC, through the Alexander Graham Bell Canada Graduate Scholarship.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Research Problem

Normal physiological adaptations favour weight gain and fat accretion during pregnancy to support fetal growth, followed by postpartum weight loss and fat mobilization to meet increased maternal energy demands during lactation [13, 38]. It is well-known that the risk of poor maternal and fetal health outcomes increases when women gain either too little or too much weight during pregnancy [5, 8, 24, 25, 34] . Gaining too little weight is associated with low birth weight and preterm birth, while gaining too much weight is associated with complications such as gestational diabetes, hypertension, caesarean sections, and macrosomia in infants [8, 27]. Recent studies indicate that, in developed countries, the majority of pregnant women gain more weight than recommended [2, 14, 23, 35, 43]. Excessive weight gain during pregnancy followed by inadequate postpartum weight loss can contribute to maternal long-term obesity and associated sequelae including cardiovascular disease, hypertension, diabetes, and degenerative joint disease [27, 34, 36, 42].

Research has shown that customized counselling and education, together with individual weight monitoring, improves adherence to gestational weight gain (GWG) recommendations [28–30, 32, 40]. At present, women are compared against a general weight gain trajectory chosen on the basis of pre-pregnancy body mass index (BMI). These weight gain charts depend only on pre-pregnancy BMI and cannot be adjusted to account for individual weight measurements during pregnancy. Adherence rates could be further improved,

however, if the weight trajectory could be individualized. The monitoring and modelling of individual weight growth trajectories during pregnancy may be useful for women and healthcare providers. In particular, gestational weight gain trajectories can be used to predict weight changes during pregnancy and thus may help identify women who are at high risk of suboptimal GWG.

A previous study used the functional principal component analysis (FPCA) approach and successfully estimated individual weight trajectories in pregnant women [7]. The traditional FPCA method borrows information from the entire cohort, and does not incorporate information on pre-pregnancy BMI. However, women of different pre-pregnancy BMI categories tend to exhibit different trends in gestational weight gain. For instance, women with lower pre-pregnancy BMI typically have higher gestational weight gain and weight growth trajectories that tend to be steeper than the mean weight growth trajectory across all BMI categories [23, 31]. Gestational weight trajectories of women with higher pre-pregnancy BMI exhibit the opposite trend: they tend to be more gradually rising than the mean weight growth trajectory.

Trajectories that have a different pattern than the sample mean are likely to have larger FPC scores on PCs which are not included in the estimation by FPCA. As a result, these estimated trajectories are closer to the mean function than they should be. This was observed in [7], where it was found that the estimated total GWG from FPCA trajectories underestimated the weight gain for underweight women, while overestimated the weight gain for women who were overweight or obese before pregnancy. Thus, there is a need to account for weight variation that is individual to each pre-pregnancy BMI category.

1.2 Background

1.2.1 Dimension Reduction

The analysis of high-dimensional data — datasets in which the dimensionality of each observation is comparable to or even larger than the number of observations — has drawn increasing attention in the last several decades [12, 16]. Text mining, image analysis, e-commerce, and computational biology, for ex-

ample, are fields of research that involve datasets where hundreds or thousands of variables are measured for each object of interest. These datasets present interesting challenges because they can be difficult to visualize and are often structurally complex. High-dimensional datasets are typically high-rank, meaning they contain a large number of patterns. The complexity of such datasets may be reduced by applying dimensionality reduction techniques, which decompose large high-rank matrices into smaller, low-rank components. Dimension reduction methods extract the features that explain most of the variability in the data, thus making the data much simpler to analyze. There are a number of matrix decomposition methods that can be used for dimension reduction. Some examples include principal component analysis (PCA) [1, 39, 46], partial least squares (PLS) [4, 20, 26], and alternating least squares (ALS) [9, 10, 18].

1.2.2 Multi-Block Data

Real-world data are often obtained as a collection of matrices rather than as a single matrix. Such *multi-block* data are naturally connected and usually share some common features while simultaneously exhibiting their own individual features, reflecting the underlying data generation mechanisms. Multi-block data frequently arise in biomedical studies, where multiple different types of high-dimensional data (e.g., gene expression, microRNA, genotype, protein abundance/activity) are commonly collected on the same set of organisms or tissue samples. Datasets of this structure, in which multiple large and fundamentally disparate sets of variables are available for a *common set of objects*, are said to be *vertically integrated*. In vertically integrated multi-block data, the columns of each matrix represent cases or objects, while the rows of different matrices represent variables measured on different platforms. Thus, each data matrix has the same number of columns, corresponding to a common set of objects, but a potentially different number of rows. For example, the rows of one matrix may contain gene expression measurements, while the rows of another may contain miRNA measurements for the same set of biological samples. In *horizontally integrated* multi-block data, disparate sets of samples

(e.g., sick and healthy patients) are available on the *same data type*. The data matrices thus have the same number of rows, each corresponding to a variable in a given measurement technology that does not vary from matrix to matrix. The number of columns in each data matrix may differ, however, since the matrices represent different sets of samples.

Until recently, there was little statistical methodology for integrative analyses of high-dimensional multi-block data. Two methods of treating such datasets were common in the past: in one approach, each data block was analyzed separately. In the second method, multi-block data were analyzed as a single combined dataset. However, both of these approaches have limitations. Individual analysis of each data block will not capture the critical associations and potential causal relationships between data blocks. Moreover, the important features that are unique to each data block may be missed when analyzing multi-block data as a single combined dataset. These limitations have motivated the need for new statistical methods that explore associations between multiple data blocks while recognizing that each data block can provide unique and important information. In recent years, a number of methods have been developed for analyzing multi-block data, which simultaneously model features that are shared across multiple blocks and features that are specific to a particular block. These methods extend well-established techniques such as partial least squares [22], canonical correlation analysis [47], non-parametric Bayesian modelling [33], non-negative factorization [44], factor analysis [11], and simultaneous components analysis [37, 41].

1.2.3 The JIVE Method

The joint and individual variation explained (JIVE) method was recently introduced for the integrated unsupervised analysis of multi-block high-dimensional data [21]. Developed as a multi-block extension of principal component analysis, JIVE quantifies the amount of joint (shared) variation between data blocks as well as the amount of individual (unique) variation specific to each data block. The JIVE method also reduces the dimensionality of the data and allows for visual examination of joint and individual (block-specific) structure.

JIVE was originally designed for the analysis of biomedical data from multiple technologies. The method has been successfully applied to the analysis of gene expression and miRNA data on glioblastoma multiforme tumor samples, providing better characterization of tumor types and better understanding of gene–miRNA associations [21]. However, the JIVE model and algorithm are very general and may be useful in other contexts. For example, the method has been used to analyze data that were processed using different computational pipelines [19], and in the analysis of rail commute patterns at different times of day [15]. In finance, JIVE can potentially improve upon current models that explain variation across and within disparate markets [3].

Although JIVE was primarily motivated by the analysis of vertically integrated multi-block data, where different types of data are measured on the same set of samples, the method may also be useful for the analysis of horizontally integrated data, in which a single data type is measured on different sets of samples. This approach is referred to as *horizontal* JIVE, since it involves integrating across samples (columns) rather than variables (rows). Thus, while a standard JIVE analysis identifies structure in the samples that is present across data types or individual to a data type, a horizontal JIVE analysis identifies structure in the variables that is present across sample sets or individual to a sample set. Some potential applications for such a horizontal analysis include: (1) the analysis of treatment and control samples in an experimental study, and (2) the analysis of samples for different species (e.g., cats and dogs), for which measurements are available on a common set of genes.

JIVE decomposes a multi-block dataset into three components: a low-rank matrix capturing joint variation across data blocks, low-rank matrices for structured variation individual to each block, and a residual matrix capturing variability in the data unattributed to joint or individual structures. Joint structure reveals patterns that explain substantial variability across multiple data blocks, whereas individual structure reveals patterns that explain substantial variation in one block but not others. With specified ranks, the JIVE algorithm iteratively estimates joint and individual structure to minimize the

overall residual sum of squares. Simultaneous estimation of individual structure allows for more accurate estimation of the underlying joint structure, and vice versa. Furthermore, the joint and individual structures are assumed to be orthogonal in order for the matrix decomposition to be identifiable.

1.2.4 Extending JIVE: The JIVE–FPCA Method

Motivated by the desire to individualize women's gestational weight gain trajectories, this thesis proposes a new method, called *JIVE–FPCA*, that applies the JIVE algorithm to FPCA for the analysis of irregularly spaced, sparse longitudinal multi-block data. In the JIVE–FPCA method, a subject's trajectory is decomposed into joint and individual components. These components account for the variation between and within grouped samples of functional data, respectively, and are iteratively estimated. The joint component of a given trajectory is calculated using the pooled data from all blocks, whereas the individual components are based on data from particular blocks. The individual components of a trajectory consist of principal components (PCs) that are unique to a specific feature set (vertically integrated data) or sample set (horizontally integrated data).

In this thesis, the JIVE–FPCA method is applied to simulated data and real, longitudinal gestational weight data. A horizontal JIVE–FPCA analysis is performed in both cases. This variation of JIVE–FPCA analyzes grouped observations on the same functional data type; however, JIVE–FPCA is also designed for analysis across disparate data types. The usefulness of the new methodology is illustrated by comparing JIVE–FPCA estimated trajectories with those estimated using the original FPCA approach.

1.3 Thesis Structure

Chapter 2 introduces the JIVE–FPCA model and notations used throughout this thesis. This section also presents the JIVE–FPCA algorithm for estimating the joint and individual components of a set of trajectories. Chapter 2 concludes with derivations of the asymptotic pointwise and simultaneous con-

fidence intervals for individual JIVE–FPCA trajectories. Chapter 3 discusses the design and results of the simulation experiment. The JIVE–FPCA method is applied to real, horizontally integrated, longitudinal gestational weight data in Chapter 4. The resulting weight growth trajectories are then compared with those obtained using the existing FPCA approach. A discussion and potential future work follow in Chapter 5.

Chapter 2

The JIVE–FPCA Method

2.1 Model

The FPCA method proposed by Yao et al. [45] was developed for sparse, irregularly measured longitudinal data, and is used as a basis for the JIVE–FPCA approach. In this thesis, the notation developed in [45] is extended to include principal components that are unique to a specific feature set (vertically integrated data) or sample set (horizontally integrated data).

Sparse functional data are modelled as noisy sampled points from a collection of trajectories that are assumed to be independent realizations of a smooth random function $X(t)$. The domain of $X(\cdot)$ is some bounded and closed time interval $\mathcal{T}$. Suppose that $X_i^{(\mathrm{J})}(t)$ is the joint component of the i^{th} random trajectory $X_i(t)$, $i = 1, \ldots, n$. Similarly, let $X_i^{(\mathrm{I},k)}(t)$ be the k^{th} individual component of the random curve $X_i(t)$, where k is used to denote the k^{th} feature set or sample set, with $k = 1, \ldots, m$ for some $m \in \mathbb{N}$. Let Y_{ij} be the j^{th} observation of the random function $X_i(\cdot)$, made at a random time $T_{ij} \in \mathcal{T}$, $j = 1, \ldots, N_i$, where N_i is the number of measurements made on the i^{th} subject and is considered random, reflecting sparse and irregular designs. The random variables N_i are assumed to be iid and independent of all other random variables. Furthermore, let $\tilde{\epsilon}_{ij}$ be uncorrelated measurement errors with mean zero and constant variance $\tilde{\sigma}^2$. Thus, the JIVE–FPCA model is given by

$$Y_{ij} = X_i(T_{ij}) + \tilde{\epsilon}_{ij} = X_i^{(\mathrm{J})}(T_{ij}) + \sum_{k=1}^{m} X_i^{(\mathrm{I},k)}(T_{ij}) + \tilde{\epsilon}_{ij}$$

$$= \underbrace{\mu(T_{ij}) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(T_{ij})}_{\text{Joint}} + \underbrace{\sum_{k=1}^{m} \left(\mu^{(k)}(T_{ij}) + \sum_{s=1}^{\infty} \eta_{is}^{(k)} \phi_s^{(k)}(T_{ij}) \right)}_{\text{Individual}} + \tilde{\epsilon}_{ij}, \quad (2.1)$$

where $\mu(t), \xi_{il}$, and $\phi_l(t)$ represent the mean function, functional principal component (FPC) scores, and eigenfunctions (PC functions), respectively, for the joint component of the i^{th} trajectory, which is calculated using the pooled data from all m blocks. These same quantities are, accordingly, represented by $\mu^{(k)}(t), \eta_{is}^{(k)}$, and $\phi_s^{(k)}(t)$ for the individual component that is based on data from the k^{th} feature set or sample set. Within each component of a given trajectory, it is assumed that the eigenfunctions are orthogonal, and that the FPC scores are uncorrelated random variables with mean zero. It is further assumed that the $\tilde{\epsilon}_{ij}$ are independent of the random coefficients ξ_{il} and $\eta_{is}^{(k)}$. Trajectories derived from vertically integrated datasets have a joint component and m individual components, where the k^{th} individual component is estimated from the k^{th} feature set. On the other hand, subjects from horizontally integrated datasets have trajectories that consist of a joint component and a single individual component, where the individual component is estimated using only the sample set data associated with that particular subject.

2.2 Estimation of the Model Components

The JIVE algorithm is applied to FPCA in order to estimate the joint and individual components of a set of trajectories $\{X_i(t)\}_{i=1}^n$. The general steps involved in the proposed JIVE–FPCA method are presented below, and may be applied to vertically or horizontally integrated datasets.

1. To estimate the joint component of each trajectory, the FPCA procedure in [45] is applied to all the data points from all m feature sets or sample sets. The fraction of variance explained (FVE) method is used to select the number of PC functions. With the FVE method, the smallest number of eigenfunctions that explain a pre-defined fraction (threshold) of the total variation in all the trajectories is chosen. Usually, the first few PCs explain the largest fraction of the total variance, and thus represent the

dominant modes of variation. Let L be the number of eigenfunctions chosen by the FVE approach. Then, the joint component of the i^{th} subject's trajectory is estimated by

$$\widehat{X}_i^{(\mathrm{J})}(t) = \hat{\mu}(t) + \sum_{l=1}^{L} \hat{\xi}_{il}\hat{\phi}_l(t), \tag{2.2}$$

where the $\hat{\xi}_{il}$ are estimated subject-specific FPC scores.

2. The residuals $\epsilon_{ij} = Y_{ij} - \widehat{X}_i^{(\mathrm{J})}(t)$ are computed next, using the estimate for $X_i^{(\mathrm{J})}(t)$ found in Step 1. This set of residuals may be represented as $\{\epsilon_{ij}\} = \varepsilon_{ij}^{(1)} \cup \varepsilon_{ij}^{(2)} \cup \cdots \cup \varepsilon_{ij}^{(m)}$, where $\varepsilon_{ij}^{(k)}$ are the residuals associated with the k^{th} feature set or sample set, $k = 1, \ldots, m$.

3. In Step 3, FPCA is applied to the residuals ϵ_{ij} obtained in Step 2 so as to estimate the individual component(s) of each trajectory. This process is computationally simpler to carry out for horizontally integrated data, since each trajectory is comprised of a single individual component.

   (a) First, consider the case of horizontally integrated data. Suppose that the i^{th} subject belongs to the k^{th} sample set. Thus, FPCA is applied to the residuals associated with the k^{th} sample set (denoted by $\varepsilon_{ij}^{(k)}$). Let $S^{(k)}$ be the number of eigenfunctions selected by the FVE criterion. Then, the estimate for the individual component of the i^{th} subject's trajectory is

   $$\widehat{X}_i^{(\mathrm{I},k)}(t) = \hat{\mu}^{(k)}(t) + \sum_{s=1}^{S^{(k)}} \hat{\eta}_{is}^{(k)}\hat{\phi}_s^{(k)}(t). \tag{2.3}$$

   (b) Estimation of the individual components proceeds in a similar fashion for vertically integrated (multi-datatype) datasets. The process begins with the application of FPCA to the residuals associated with the first feature set (denoted by $\varepsilon_{ij}^{(1)}$). Suppose that $S^{(1)}$ eigenfunctions are selected by the FVE method. Therefore, the first individual component of the i^{th} subject's trajectory is approximated

by

$$\widehat{X}_i^{(\mathrm{I},1)}(t) = \hat{\mu}^{(1)}(t) + \sum_{s=1}^{S^{(1)}} \hat{\eta}_{is}^{(1)} \hat{\phi}_s^{(1)}(t).$$

The residuals

$$Y_{ij} - \left(\widehat{X}_i^{(\mathrm{J})}(t) + \widehat{X}_i^{(\mathrm{I},1)}(t) \right)$$

are then calculated using the estimates for $X_i^{(\mathrm{J})}(t)$ and $X_i^{(\mathrm{I},1)}(t)$ found above. This set of residuals can be expressed as a union of m sets, where the k^{th} set represents the residuals associated with the k^{th} feature set, $k = 1, \ldots, m$. In general, to estimate the k^{th} individual component, for $k \geq 2$, the following residuals are computed

$$Y_{ij} - \left(\widehat{X}_i^{(\mathrm{J})}(t) + \widehat{X}_i^{(\mathrm{I},1)}(t) + \cdots + \widehat{X}_i^{(\mathrm{I},k-1)}(t) \right)$$

using previously obtained estimates for $X_i^{(\mathrm{J})}(t), X_i^{(\mathrm{I},1)}(t), \ldots, X_i^{(\mathrm{I},k-1)}(t)$. FPCA is then applied to the residuals associated with the k^{th} feature set. The estimate for the k^{th} individual component of the i^{th} subject's trajectory is given by Equation (2.3). This process is repeated until all m individual components have been estimated.

4. Combining the estimates for the joint and individual components obtained in Steps 1–3, we obtain the following estimate for the i^{th} subject's trajectory:

$$\begin{aligned} \widehat{X}_i(t) &= \widehat{X}_i^{(\mathrm{J})}(t) + \sum_{k=1}^{m} \widehat{X}_i^{(\mathrm{I},k)}(t) \\ &= \underbrace{\hat{\mu}(t) + \sum_{l=1}^{L} \hat{\xi}_{il} \hat{\phi}_l(t)}_{\text{Joint}} + \underbrace{\sum_{k=1}^{m} \left(\hat{\mu}^{(k)}(t) + \sum_{s=1}^{S^{(k)}} \hat{\eta}_{is}^{(k)} \hat{\phi}_s^{(k)}(t) \right)}_{\text{Individual}}. \end{aligned} \tag{2.4}$$

Note that for horizontally integrated data, Equation (2.4) simplifies to

$$\widehat{X}_i(t) = \hat{\mu}(t) + \sum_{l=1}^{L} \hat{\xi}_{il} \hat{\phi}_l(t) + \hat{\mu}^{(k)}(t) + \sum_{s=1}^{S^{(k)}} \hat{\eta}_{is}^{(k)} \hat{\phi}_s^{(k)}(t), \tag{2.5}$$

where k represents the sample set to which the i^{th} subject belongs.

Once the individual trajectories have been estimated, further steps may be taken to improve the fit of each trajectory and thus minimize the residual sum of squares (RSS). This may be achieved by iteratively estimating the joint and individual components. The JIVE–FPCA algorithm is an extension of the iterative method outlined in [21] for estimating joint and individual structures. Pseudocode for the JIVE–FPCA iterative estimation procedure is given below. The algorithm is written under the assumption that vertically integrated data are used, but can easily be modified for a horizontal JIVE–FPCA analysis.

JIVE–FPCA algorithm

- Obtain initial estimates for $X_i(t)$ by performing Steps 1–4 above.
- Compute the residuals $\tilde{\epsilon}_{ij} = Y_{ij} - \widehat{X}_i(t)$ and the RSS $= \sum_{i=1}^{n} \sum_{j=1}^{N_i} \tilde{\epsilon}_{ij}^2$.
- Loop:
  - Calculate $r_{ij}^{(\mathrm{J})} = Y_{ij} - \sum_{k=1}^{m} \widehat{X}_i^{(\mathrm{I},k)}(t)$ using estimates for $X_i^{(\mathrm{I},k)}(t)$ found in the previous iteration.
  - Apply FPCA to the residuals $r_{ij}^{(\mathrm{J})}$ to obtain new estimates for the joint component of each trajectory.
  - Perform Steps 2–4 to obtain new estimates for the individual components and for $X_i(t)$.
  - Update the RSS
- Repeat until some convergence criterion is satisfied.

The JIVE–FPCA algorithm is much simpler to implement when working with horizontally integrated data, since for every trajectory there is only one individual component that needs to be reestimated in each iteration. In terms of the above algorithm, the summation $\sum_{k=1}^{m} \widehat{X}_i^{(\mathrm{I},k)}(t)$ reduces to a single term $\widehat{X}_i^{(\mathrm{I},k)}(t)$.

The time needed to compute JIVE–FPCA curves depends primarily on the dimensions of the data and the number of PCs used to estimate the joint and individual components of each trajectory. A JIVE–FPCA analysis of the simulation described in Chapter 3 takes 15 minutes for 5 iterations on a 2.7 GHz CPU with 8GB RAM. The analysis of larger multi-block datasets, such as the

gestational weight data described in Chapter 4, takes longer (with 5 iterations requiring 23 minutes, for comparison). The difference in computing time between these particular analyses is attributed to the difference in the number of blocks (sample sets) associated with each dataset (2 for the simulated data versus 4 for gestational weight).

The application of JIVE–FPCA to simulated and real data (see Chapters 3 and 4, respectively) has shown that the iterative method is not monotone in the sense that the RSS does not necessarily decrease at each step. For example, in some experiments the RSS suddenly increased sharply after many iterations of steady decrease, and then decreased again for a long period of time. As a result, it is difficult to determine when the minimum RSS has been found. The JIVE–FPCA algorithm has shown to exhibit stable behavior (i.e., steadily decreasing RSS) for at least several iterations. The following stopping criteria are proposed in order to obtain optimal results:

- Run JIVE–FPCA until $\text{RSS}_{\text{JIVE–FPCA}} < \text{RSS}_{\text{FPCA}}$ and the percent change in RSS between two successive iterations is less than 1%.
- If $\text{RSS}_{\text{JIVE–FPCA}} < \text{RSS}_{\text{FPCA}}$ but the percent change in RSS between two successive iterations is greater than 1% after the 5^{th} iteration, stop after round 5 to prevent overfitting.

These stopping criteria are used in this thesis for both the simulated and gestational weight data, and are recommended for any general multi-block dataset.

2.3 Confidence Bands of Individual Trajectories

Pointwise and simultaneous confidence bands for individual JIVE–FPCA trajectories may be derived. The details on the construction of these bands are provided below.

This section extends the notation used by Yao et al. [45] in deriving the asymptotic pointwise and simultaneous confidence bands for FPCA predicted

individual trajectories. The notational conventions established throughout this thesis continue to be used here: for example, the subscripts l and s are used to associate quantities with the joint and individual components of a given trajectory, respectively. Some new notation is also introduced to distinguish between finite and infinite-dimensional processes.

In a finite-dimensional process, it is assumed that each component of a given trajectory can be approximated sufficiently well by a certain number of eigenfunctions. For such a process, the assumption is that the joint component has been truncated at the first L principal component functions of its expansion, and that the k^{th} individual component has been truncated at the first $S^{(k)}$ eigenfunctions, where $k = 1, \ldots, m$. In this section, the superscript (∞) is used to represent infinite-dimensional processes, so that the joint component of the i^{th} subject's trajectory is given by $X_i^{(\mathrm{J},\infty)}(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t)$, and the k^{th} individual component by $X_i^{(\mathrm{I},k,\infty)}(t) = \mu^{(k)}(t) + \sum_{s=1}^{\infty} \eta_{is}^{(k)}\phi_s^{(k)}(t)$. In general, $\{X_i(t)\}_{i=1}^n$ and $\{X_i^{(\infty)}(t)\}_{i=1}^n$ are used to represent the set of truncated and untruncated trajectories, respectively.

2.3.1 Asymptotic Pointwise Confidence Bands

The construction of asymptotic pointwise confidence intervals for individual JIVE–FPCA trajectories is first described. Let ξ_{il} be uncorrelated random variables with mean zero and variance $E(\xi_{il}^2) = \lambda_l$, where ξ_{il} are FPC scores, and λ_l are eigenvalues used in the orthogonal expansion of the covariance function of $X(t)$. Suppose that $\sum_l \lambda_l < \infty$, $\lambda_1 \geq \lambda_2 \geq \ldots$. Similarly, assume that $\eta_{is}^{(k)}$ are uncorrelated random variables with mean zero and variance $E\left[\left(\eta_{is}^{(k)}\right)^2\right] = \lambda_s^{(k)}$, where $\eta_{is}^{(k)}$ and $\lambda_s^{(k)}$ are, respectively, FPC scores and eigenvalues corresponding to the individual component that is based on data from the k^{th} feature set or sample set, $k = 1, \ldots, m$. Once again, suppose that $\sum_s \lambda_s^{(k)} < \infty$, $\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \ldots$ for all k.

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T$ be an $N_i \times 1$ column vector containing measurements for the i^{th} subject, $i = 1, \ldots, n$. Suppose that for horizontally integrated data, the i^{th} subject belongs to the k^{th} sample set, $k \in \{1, \ldots, m\}$. Let

$\boldsymbol{r}_i^{(k)} = \left(r_{i1}^{(k)}, \ldots, r_{iN_i}^{(k)}\right)^T$ be an $N_i \times 1$ vector of residuals obtained by calculating $Y_{ij} - \widehat{X}_i^{(\mathrm{J})}(t)$, where $\widehat{X}_i^{(\mathrm{J})}(t)$ is an estimate for the joint component of the i^{th} subject's trajectory. For vertically integrated data, let $\boldsymbol{r}_i^{(k)}$ be a column vector of length $N_i^{(k)}$ whose elements are residuals corresponding to the k^{th} feature set. Note that for vertically integrated data, $N_i^{(k)}$ is less than the number of measurements N_i made on the i^{th} subject, since the elements of $\boldsymbol{r}_i^{(k)}$ are a subset of the residuals obtained by subtracting joint component predicted values from subject i observations.

For arbitrary integers $L, S^{(k)} \geq 1$, let $\boldsymbol{\xi}_{L,i} = (\xi_{i1}, \ldots, \xi_{iL})^T$, $\tilde{\boldsymbol{\xi}}_{L,i} = (\tilde{\xi}_{i1}, \ldots, \tilde{\xi}_{iL})^T$, $\boldsymbol{\eta}_{S^{(k)},i} = \left(\eta_{i1}^{(k)}, \ldots, \eta_{iS^{(k)}}^{(k)}\right)^T$, and $\tilde{\boldsymbol{\eta}}_{S^{(k)},i} = \left(\tilde{\eta}_{i1}^{(k)}, \ldots, \tilde{\eta}_{iS^{(k)}}^{(k)}\right)^T$. The FPC scores for the joint component of the i^{th} subject's trajectory are given in [45] to be

$$\tilde{\xi}_{il} = E\left[\xi_{il} | \boldsymbol{Y}_i\right] = \lambda_l \boldsymbol{\phi}_{il}^T \boldsymbol{\Sigma}_{\boldsymbol{Y}_i}^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i), \tag{2.6}$$

where $\boldsymbol{\phi}_{il} = (\phi_l(T_{i1}), \ldots, \phi_l(T_{iN_i}))^T$, $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i} = Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_i)$, and $\boldsymbol{\mu}_i = (\mu(T_{i1}), \ldots, \mu(T_{iN_i}))^T$. For vertically integrated data, the FPC scores for the k^{th} individual component of the i^{th} subject's trajectory have the following form:

$$\tilde{\eta}_{is}^{(k)} = E\left[\eta_{is}^{(k)} | \boldsymbol{r}_i^{(k)}\right] = \lambda_s^{(k)} (\boldsymbol{\phi}_{is}^{(k)})^T \boldsymbol{\Sigma}_{\boldsymbol{r}_i^{(k)}}^{-1} (\boldsymbol{r}_i^{(k)} - \boldsymbol{\mu}_i^{(k)}), \tag{2.7}$$

where $\boldsymbol{\phi}_{is}^{(k)} = \left(\phi_s^{(k)}(T_{i1}), \ldots, \phi_s^{(k)}(T_{iN_i^{(k)}})\right)^T$, $\boldsymbol{\Sigma}_{\boldsymbol{r}_i^{(k)}} = Cov(\boldsymbol{r}_i^{(k)}, \boldsymbol{r}_i^{(k)})$, and $\boldsymbol{\mu}_i^{(k)} = \left(\mu^{(k)}(T_{i1}), \ldots, \mu^{(k)}(T_{iN_i^{(k)}})\right)^T$. Note that Equation (2.7) also holds when working with horizontally integrated data. The only differences are that for a horizontal JIVE–FPCA analysis: (a) there is only one individual component, so the superscript (k) represents the sample set to which subject i belongs; and (b) $N_i^{(k)}$ is replaced with N_i in the definitions of the $\boldsymbol{\phi}_{is}^{(k)}$ and $\boldsymbol{\mu}_i^{(k)}$ vectors.

Following the notation of Yao et al. [45], the covariance matrix of $\tilde{\boldsymbol{\xi}}_{L,i}$ is $Var(\tilde{\boldsymbol{\xi}}_{L,i}) = \boldsymbol{H} \boldsymbol{\Sigma}_{\boldsymbol{Y}_i}^{-1} \boldsymbol{H}^T$, where $\boldsymbol{H}$ is the $L \times N_i$ matrix given by $Cov(\boldsymbol{\xi}_{L,i}, \boldsymbol{Y}_i) = (\lambda_1 \boldsymbol{\phi}_{i1}, \ldots, \lambda_L \boldsymbol{\phi}_{iL})^T$. Similarly, for vertically integrated data, it follows that $Var(\tilde{\boldsymbol{\eta}}_{S^{(k)},i}) = \boldsymbol{H}^{(k)} \boldsymbol{\Sigma}_{\boldsymbol{r}_i^{(k)}}^{-1} (\boldsymbol{H}^{(k)})^T$ for the $S^{(k)} \times N_i^{(k)}$ matrix $\boldsymbol{H}^{(k)} = Cov(\boldsymbol{\eta}_{S^{(k)},i}, \boldsymbol{r}_i^{(k)}) = \left(\lambda_1^{(k)} \boldsymbol{\phi}_{i1}^{(k)}, \ldots, \lambda_{S^{(k)}}^{(k)} \boldsymbol{\phi}_{iS^{(k)}}^{(k)}\right)^T$. Note that $\boldsymbol{H}^{(k)}$ is an $S^{(k)} \times N_i$ matrix for the case of horizontally integrated data.

To take into account the variation of $\boldsymbol{\xi}_{L,i}$ and $\boldsymbol{\eta}_{S^{(k)},i}$, the quantities $Var(\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i})$ and $Var(\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i})$ are used to assess the estimation errors of $\tilde{\boldsymbol{\xi}}_{L,i}$ and

$\tilde{\boldsymbol{\eta}}_{S^{(k)},i}$, respectively. Since $\tilde{\boldsymbol{\xi}}_{L,i} = E\left[\boldsymbol{\xi}_{L,i}|\boldsymbol{Y}_i\right]$ is the projection of $\boldsymbol{\xi}_{L,i}$ onto the space spanned by linear functions of $\boldsymbol{Y}_i$, we have $E[\tilde{\boldsymbol{\xi}}_{L,i}\boldsymbol{\xi}_{L,i}^T] = E[\tilde{\boldsymbol{\xi}}_{L,i}\tilde{\boldsymbol{\xi}}_{L,i}^T]$, that is, $Var(\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i}) = Var(\boldsymbol{\xi}_{L,i}) - Var(\tilde{\boldsymbol{\xi}}_{L,i}) = \boldsymbol{\Omega}_L$, where $\boldsymbol{\Omega}_L = \boldsymbol{\Lambda} - \boldsymbol{H}\boldsymbol{\Sigma}_{\boldsymbol{Y}_i}^{-1}\boldsymbol{H}^T$ and $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_L)$. By the same principles, $Var(\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i}) = Var(\boldsymbol{\eta}_{S^{(k)},i}) - Var(\tilde{\boldsymbol{\eta}}_{S^{(k)},i}) = \boldsymbol{\Omega}_{S^{(k)}}$, where $\boldsymbol{\Omega}_{S^{(k)}} = \boldsymbol{\Lambda}^{(k)} - \boldsymbol{H}^{(k)}\boldsymbol{\Sigma}_{\boldsymbol{r}_i^{(k)}}^{-1}(\boldsymbol{H}^{(k)})^T$ and $\boldsymbol{\Lambda}^{(k)} = diag(\lambda_1^{(k)}, \ldots, \lambda_{S^{(k)}}^{(k)})$. Then, under Gaussian assumptions, it follows that $(\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i}) \sim Normal(\boldsymbol{0}, \boldsymbol{\Omega}_L)$ and $(\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i}) \sim Normal(\boldsymbol{0}, \boldsymbol{\Omega}_{S^{(k)}})$.

Let $\hat{\boldsymbol{\Omega}}_L = \hat{\boldsymbol{\Lambda}} - \hat{\boldsymbol{H}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}_i}^{-1}\hat{\boldsymbol{H}}^T$, where $\hat{\boldsymbol{\Lambda}} = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_L)$ and $\hat{\boldsymbol{H}} = (\hat{\lambda}_1\hat{\boldsymbol{\phi}}_{i1}, \ldots, \hat{\lambda}_L\hat{\boldsymbol{\phi}}_{iL})^T$. Furthermore, let $\boldsymbol{\phi}_{L,t} = (\phi_1(t), \ldots, \phi_L(t))^T$, $\hat{\boldsymbol{\phi}}_{L,t} = (\hat{\phi}_1(t), \ldots, \hat{\phi}_L(t))^T$, and $\widehat{X}_i^{(\mathrm{J})}(t) = \hat{\mu}(t) + \hat{\boldsymbol{\phi}}_{L,t}^T\hat{\boldsymbol{\xi}}_{L,i}$ for $t \in \mathcal{T}$. Thus, Theorem 4 in [45] establishes that the distribution of $\widehat{X}_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t)$ is asymptotically $Normal(0, \hat{\boldsymbol{\phi}}_{L,t}^T\hat{\boldsymbol{\Omega}}_L\hat{\boldsymbol{\phi}}_{L,t})$. Let $\hat{\boldsymbol{\Omega}}_{S^{(k)}} = \hat{\boldsymbol{\Lambda}}^{(k)} - \hat{\boldsymbol{H}}^{(k)}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}_i^{(k)}}^{-1}(\hat{\boldsymbol{H}}^{(k)})^T$, where $\hat{\boldsymbol{\Lambda}}^{(k)} = diag(\hat{\lambda}_1^{(k)}, \ldots, \hat{\lambda}_{S^{(k)}}^{(k)})$ and $\hat{\boldsymbol{H}}^{(k)} = \left(\hat{\lambda}_1^{(k)}\hat{\boldsymbol{\phi}}_{i1}^{(k)}, \ldots, \hat{\lambda}_{S^{(k)}}^{(k)}\hat{\boldsymbol{\phi}}_{iS^{(k)}}^{(k)}\right)^T$. Moreover, let $\boldsymbol{\phi}_{S^{(k)},t}^{(k)} = \left(\phi_1^{(k)}(t), \ldots, \phi_{S^{(k)}}^{(k)}(t)\right)^T$, $\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)} = \left(\hat{\phi}_1^{(k)}(t), \ldots, \hat{\phi}_{S^{(k)}}^{(k)}(t)\right)^T$, and $\widehat{X}_i^{(\mathrm{I},k)}(t) = \hat{\mu}^{(k)}(t) + (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T\hat{\boldsymbol{\eta}}_{S^{(k)},i}$. Once again, by Theorem 4 in [45] it follows that $\widehat{X}_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t)$ is asymptotically $Normal\left(0, (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T\hat{\boldsymbol{\Omega}}_{S^{(k)}}\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)}\right)$ for all $k = 1, \ldots, m$.

To construct asymptotic pointwise confidence intervals for individual JIVE–FPCA trajectories, it is necessary to assume that the joint and individual components are independent of each other. For the case of horizontally integrated data, this implies that for some given trajectory $X_i^{(\infty)}(t)$, the components $X_i^{(\mathrm{J},\infty)}(t)$ and $X_i^{(\mathrm{I},k,\infty)}(t)$ are independent for $k \in \{1, \ldots, m\}$. For vertically integrated data, the joint component is assumed to be independent of all m individual components. Intuitively, this means that patterns responsible for joint structure between data blocks are unrelated to patterns responsible for individual structure. For vertically integrated data, it is further assumed that all m individual components are independent of each other.

Having developed the necessary notation, asymptotic pointwise confidence intervals for $\{X_i^{(\infty)}(t)\}_{i=1}^n$ are now constructed. First, consider the case of vertically integrated data. By assumption, for some given trajectory $X_i^{(\infty)}(t)$, the joint component $X_i^{(\mathrm{J},\infty)}(t)$ is independent of all m individual components $X_i^{(\mathrm{I},k,\infty)}(t)$. Thus, it follows that $\widehat{X}_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t)$ and $\widehat{X}_i^{(\mathrm{I},k)}(t) -$

$X_i^{(\mathrm{I},k,\infty)}(t)$ are independent for all $k = 1, \ldots, m$. Furthermore, since all m individual components are assumed to be independent of each other, the $\left\{\widehat{X}_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t)\right\}_{k=1}^{m}$ are independent as well. Therefore,

$$\mathcal{X} = \left\{\widehat{X}_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t), \widehat{X}_i^{(\mathrm{I},1)}(t) - X_i^{(\mathrm{I},1,\infty)}(t), \ldots, \widehat{X}_i^{(\mathrm{I},m)}(t) - X_i^{(\mathrm{I},m,\infty)}(t)\right\}$$

is an independent set. From Theorem 4 in [45], the elements of $\mathcal{X}$ are asymptotically normally distributed, each with mean zero, and respective variances $\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t}, (\hat{\boldsymbol{\phi}}_{S^{(1)},t}^{(1)})^T \hat{\boldsymbol{\Omega}}_{S^{(1)}} \hat{\boldsymbol{\phi}}_{S^{(1)},t}^{(1)}, \ldots, (\hat{\boldsymbol{\phi}}_{S^{(m)},t}^{(m)})^T \hat{\boldsymbol{\Omega}}_{S^{(m)}} \hat{\boldsymbol{\phi}}_{S^{(m)},t}^{(m)}$. Using this result and Slutsky's theorem, it follows that the sum of these elements $\widehat{X}_i^{(\mathrm{J})}(t) + \sum_{k=1}^m \widehat{X}_i^{(\mathrm{I},k)}(t) - \left(X_i^{(\mathrm{J},\infty)}(t) + \sum_{k=1}^m X_i^{(\mathrm{I},k,\infty)}(t)\right) = \widehat{X}_i(t) - X_i^{(\infty)}(t)$ is asymptotically normal with mean zero and variance $\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t} + \sum_{k=1}^m (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T \hat{\boldsymbol{\Omega}}_{S^{(k)}} \hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)}$. Finally, since the joint component $X_i^{(\mathrm{J},\infty)}(t)$ can be approximated sufficiently well by the first L eigenfunctions, and the k^{th} individual component $X_i^{(\mathrm{I},k,\infty)}(t)$ is well-approximated by the first $S^{(k)}$ eigenfunctions, $k = 1, \ldots, m$, the $(1-\alpha)$ asymptotic pointwise confidence intervals for $X_i^{(\infty)}(t)$ are as follows:

$$\widehat{X}_i(t) \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t} + \sum_{k=1}^{m} (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T \hat{\boldsymbol{\Omega}}_{S^{(k)}} \hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)}}, \tag{2.8}$$

where Φ is the standard normal cumulative distribution function. Note that these confidence intervals are constructed by ignoring the bias that results from the truncation at L in $\widehat{X}_i^{(\mathrm{J})}(t)$ and at $S^{(k)}$ in $\widehat{X}_i^{(\mathrm{I},k)}(t)$, $k = 1, \ldots, m$. For horizontally integrated data, each trajectory has only one individual component and so Equation (2.8) is reduced to

$$\widehat{X}_i(t) \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t} + (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T \hat{\boldsymbol{\Omega}}_{S^{(k)}} \hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)}}, \tag{2.9}$$

where the superscript (k) represents the sample set to which the i^{th} subject belongs, $k \in \{1, \ldots, m\}$.

2.3.2 Asymptotic Simultaneous Confidence Bands

The construction of asymptotic simultaneous confidence bands is described in this section. This derivation continues to make use of the notation developed in Section 2.3.1. It is further supposed that the assumptions of Theorem 5 in

[45] hold for the joint and individual components of any given trajectory. The bands are first derived for the case of vertically integrated data.

Let

$$X_i(t) = X_i^{(\mathrm{J})}(t) + \sum_{k=1}^{m} X_i^{(\mathrm{I},k)}(t) = \mu(t) + \sum_{l=1}^{L} \xi_{il}\phi_l(t) + \sum_{k=1}^{m} \left(\mu^{(k)}(t) + \sum_{s=1}^{S^{(k)}} \eta_{is}^{(k)} \phi_s^{(k)}(t) \right)$$

and

$$\tilde{X}_i(t) = \tilde{X}_i^{(\mathrm{J})}(t) + \sum_{k=1}^{m} \tilde{X}_i^{(\mathrm{I},k)}(t) = \mu(t) + \sum_{l=1}^{L} \tilde{\xi}_{il}\phi_l(t) + \sum_{k=1}^{m} \left(\mu^{(k)}(t) + \sum_{s=1}^{S^{(k)}} \tilde{\eta}_{is}^{(k)} \phi_s^{(k)}(t) \right),$$

where $\tilde{\xi}_{il}$ and $\tilde{\eta}_{is}^{(k)}$ are defined in Equations (2.6) and (2.7), respectively. It follows that

$$\begin{aligned} \tilde{X}_i(t) - X_i(t) &= \left(\tilde{X}_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J})}(t) \right) + \sum_{k=1}^{m} \left(\tilde{X}_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k)}(t) \right) \\ &= \boldsymbol{\phi}_{L,t}^T (\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i}) + \sum_{k=1}^{m} (\boldsymbol{\phi}_{S^{(k)},t}^{(k)})^T (\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i}). \end{aligned}$$

Due to orthogonality, $\mathcal{F} = \{\boldsymbol{\phi}_{L,t} : t \in \mathcal{T}\}$ is an L-dimensional compact set and $\mathcal{G}^{(k)} = \{\boldsymbol{\phi}_{S^{(k)},t}^{(k)} : t \in \mathcal{T}\}$ is an $S^{(k)}$-dimensional compact set, for $k = 1, \ldots, m$. Since $\boldsymbol{\Omega}_L$ is positive definite, there exists an $L \times L$ nonsingular matrix $\boldsymbol{V}$ such that $\boldsymbol{V}\boldsymbol{\Omega}_L\boldsymbol{V}^T = \boldsymbol{I}_L$. Let $\boldsymbol{\theta} = \boldsymbol{V}\boldsymbol{\xi}_{L,i}$ and $\tilde{\boldsymbol{\theta}} = \boldsymbol{V}\tilde{\boldsymbol{\xi}}_{L,i}$. Since it is assumed that $(\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i}) \sim Normal(\mathbf{0}, \boldsymbol{\Omega}_L)$, it follows that $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{V}(\tilde{\boldsymbol{\xi}}_{L,i} - \boldsymbol{\xi}_{L,i}) \sim Normal(\mathbf{0}, \boldsymbol{I}_L)$, and so $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \chi_L^2$. Similarly, since $\boldsymbol{\Omega}_{S^{(k)}}$ is positive definite for $k = 1, \ldots, m$, there exists an $S^{(k)} \times S^{(k)}$ nonsingular matrix $\boldsymbol{W}^{(k)}$ such that $\boldsymbol{W}^{(k)}\boldsymbol{\Omega}_{S^{(k)}}(\boldsymbol{W}^{(k)})^T = \boldsymbol{I}_{S^{(k)}}$. Let $\boldsymbol{\vartheta}^{(k)} = \boldsymbol{W}^{(k)}\boldsymbol{\eta}_{S^{(k)},i}$ and $\tilde{\boldsymbol{\vartheta}}^{(k)} = \boldsymbol{W}^{(k)}\tilde{\boldsymbol{\eta}}_{S^{(k)},i}$. Since $(\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i}) \sim Normal(\mathbf{0}, \boldsymbol{\Omega}_{S^{(k)}})$, then $(\tilde{\boldsymbol{\vartheta}}^{(k)} - \boldsymbol{\vartheta}^{(k)}) = \boldsymbol{W}^{(k)}(\tilde{\boldsymbol{\eta}}_{S^{(k)},i} - \boldsymbol{\eta}_{S^{(k)},i}) \sim Normal(\mathbf{0}, \boldsymbol{I}_{S^{(k)}})$, so that $(\tilde{\boldsymbol{\vartheta}}^{(k)} - \boldsymbol{\vartheta}^{(k)})^T(\tilde{\boldsymbol{\vartheta}}^{(k)} - \boldsymbol{\vartheta}^{(k)}) \sim \chi_{S^{(k)}}^2$ for $k = 1, \ldots, m$. By assumption, all the components of any given trajectory are independent. Thus, the FPC scores associated with the joint component are independent of those associated with each of the individual components. Furthermore, all m sets of individual component FPC scores are independent as well. That is, $\left\{ \{\xi_{il}\}_{l=1}^{L}, \{\eta_{is}^{(1)}\}_{s=1}^{S^{(1)}}, \ldots, \{\eta_{is}^{(m)}\}_{s=1}^{S^{(m)}} \right\}$ is a collection of independent sets, for some $i \in \{1, \ldots, n\}$. Therefore, $\Psi = (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) +$

$\sum_{k=1}^{m}(\tilde{\boldsymbol{\vartheta}}^{(k)}-\boldsymbol{\vartheta}^{(k)})^T(\tilde{\boldsymbol{\vartheta}}^{(k)}-\boldsymbol{\vartheta}^{(k)}) \sim \chi^2_{L+\sum_{k=1}^{m} S^{(k)}}$ and $P\left(\Psi \leq \chi^2_{L+\sum_{k=1}^{m} S^{(k)},1-\alpha}\right) = 1-\alpha$, where $\chi^2_{L+\sum_{k=1}^{m} S^{(k)},1-\alpha}$ is the $100(1-\alpha)^{th}$ percentile of the chi-squared distribution with $L+\sum_{k=1}^{m} S^{(k)}$ degrees of freedom.

Next, recall the following well-known result from linear algebra: For a fixed $p\times 1$ vector $\boldsymbol{y}$ and a constant $c>0$, $\boldsymbol{y}^T\boldsymbol{y} \leq c^2$ if and only if $|\boldsymbol{a}^T\boldsymbol{y}| \leq c\sqrt{\boldsymbol{a}^T\boldsymbol{a}}$, for all $\boldsymbol{a} \in \mathbb{R}^p$. Applying this result with $\boldsymbol{y} = \left((\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}), (\tilde{\boldsymbol{\vartheta}}^{(1)}-\boldsymbol{\vartheta}^{(1)}), \ldots, (\tilde{\boldsymbol{\vartheta}}^{(m)}-\boldsymbol{\vartheta}^{(m)})\right)^T \in \mathbb{R}^{L+\sum_{k=1}^{m} S^{(k)}}$ and $c^2 = \chi^2_{L+\sum_{k=1}^{m} S^{(k)},1-\alpha}$, it follows that

$$P\left(|\boldsymbol{a}^T\boldsymbol{y}| \leq \sqrt{\chi^2_{L+\sum_{k=1}^{m} S^{(k)},1-\alpha}\boldsymbol{a}^T\boldsymbol{a}} : \text{for all } \boldsymbol{a} \in \mathbb{R}^{L+\sum_{k=1}^{m} S^{(k)}}\right) = 1-\alpha.$$

Now, let

$$\mathcal{E} = \{\boldsymbol{a} \in \mathbb{R}^{L+\sum_{k=1}^{m} S^{(k)}} : \boldsymbol{\phi}_{L,t} = \boldsymbol{V}^T\boldsymbol{a}_1, \boldsymbol{\phi}^{(1)}_{S^{(1)},t} = (\boldsymbol{W}^{(1)})^T\boldsymbol{a}_2, \ldots, \boldsymbol{\phi}^{(m)}_{S^{(m)},t} = (\boldsymbol{W}^{(m)})^T\boldsymbol{a}_{m+1}, t \in \mathcal{T}\},$$

where $\boldsymbol{a} = (\boldsymbol{a}_1|\boldsymbol{a}_2|\ldots|\boldsymbol{a}_{m+1})^T$ and $\boldsymbol{a}_1 \in \mathbb{R}^L, \boldsymbol{a}_2 \in \mathbb{R}^{S^{(1)}}, \ldots, \boldsymbol{a}_{m+1} \in \mathbb{R}^{S^{(m)}}$. Note that $\mathcal{E}$ is a compact subset of $\mathbb{R}^{L+\sum_{k=1}^{m} S^{(k)}}$. Therefore,

$$P\left(|\boldsymbol{a}^T\boldsymbol{y}| \leq \sqrt{\chi^2_{L+\sum_{k=1}^{m} S^{(k)},1-\alpha}\boldsymbol{a}^T\boldsymbol{a}} : \text{for all } \boldsymbol{a} \in \mathcal{E}\right) \geq 1-\alpha.$$

Since

$$\boldsymbol{a}^T\boldsymbol{y} = \left(\boldsymbol{\phi}^T_{L,t}\boldsymbol{V}^{-1}, (\boldsymbol{\phi}^{(1)}_{S^{(1)},t})^T(\boldsymbol{W}^{(1)})^{-1}, \ldots, (\boldsymbol{\phi}^{(m)}_{S^{(m)},t})^T(\boldsymbol{W}^{(m)})^{-1}\right)\begin{pmatrix}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})\\(\tilde{\boldsymbol{\vartheta}}^{(1)}-\boldsymbol{\vartheta}^{(1)})\\\vdots\\(\tilde{\boldsymbol{\vartheta}}^{(m)}-\boldsymbol{\vartheta}^{(m)})\end{pmatrix}$$

$$= \left(\boldsymbol{\phi}^T_{L,t}\boldsymbol{V}^{-1}, (\boldsymbol{\phi}^{(1)}_{S^{(1)},t})^T(\boldsymbol{W}^{(1)})^{-1}, \ldots, (\boldsymbol{\phi}^{(m)}_{S^{(m)},t})^T(\boldsymbol{W}^{(m)})^{-1}\right)\begin{pmatrix}\boldsymbol{V}(\tilde{\boldsymbol{\xi}}_{L,i}-\boldsymbol{\xi}_{L,i})\\\boldsymbol{W}^{(1)}(\tilde{\boldsymbol{\eta}}_{S^{(1)},i}-\boldsymbol{\eta}_{S^{(1)},i})\\\vdots\\\boldsymbol{W}^{(m)}(\tilde{\boldsymbol{\eta}}_{S^{(m)},i}-\boldsymbol{\eta}_{S^{(m)},i})\end{pmatrix}$$

$$= \left(\boldsymbol{\phi}^T_{L,t}(\tilde{\boldsymbol{\xi}}_{L,i}-\boldsymbol{\xi}_{L,i}) + (\boldsymbol{\phi}^{(1)}_{S^{(1)},t})^T(\tilde{\boldsymbol{\eta}}_{S^{(1)},i}-\boldsymbol{\eta}_{S^{(1)},i}) + \cdots + (\boldsymbol{\phi}^{(m)}_{S^{(m)},t})^T(\tilde{\boldsymbol{\eta}}_{S^{(m)},i}-\boldsymbol{\eta}_{S^{(m)},i})\right),$$

for some $\boldsymbol{a} \in \mathcal{E}$, it follows that

$$P\Bigg(\left|\boldsymbol{\phi}^T_{L,t}(\tilde{\boldsymbol{\xi}}_{L,i}-\boldsymbol{\xi}_{L,i}) + \sum_{k=1}^{m}(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T(\tilde{\boldsymbol{\eta}}_{S^{(k)},i}-\boldsymbol{\eta}_{S^{(k)},i})\right|$$

$$\leq \sqrt{\chi^2_{L+\sum_{k=1}^m S^{(k)},1-\alpha}\left[\boldsymbol{\phi}_{L,t}^T\boldsymbol{V}^{-1}(\boldsymbol{V}^T)^{-1}\boldsymbol{\phi}_{L,t}+\sum_{k=1}^m(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T(\boldsymbol{W}^{(k)})^{-1}((\boldsymbol{W}^{(k)})^T)^{-1}\boldsymbol{\phi}^{(k)}_{S^{(k)},t}\right]} : \text{for all } t\in\mathcal{T}\Bigg)$$

$\geq 1-\alpha$, which implies that

$$P\Bigg(\left|\tilde{X}_i(t)-X_i(t)\right|$$

$$\leq \sqrt{\chi^2_{L+\sum_{k=1}^m S^{(k)},1-\alpha}\left[\boldsymbol{\phi}_{L,t}^T\boldsymbol{\Omega}_L\boldsymbol{\phi}_{L,t}+\sum_{k=1}^m(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T\boldsymbol{\Omega}_{S^{(k)}}\boldsymbol{\phi}^{(k)}_{S^{(k)},t}\right]} : \text{for all } t\in\mathcal{T}\Bigg)$$

$\geq 1-\alpha$, since $\tilde{X}_i(t)-X_i(t)=\boldsymbol{\phi}_{L,t}^T(\tilde{\boldsymbol{\xi}}_{L,i}-\boldsymbol{\xi}_{L,i})+\sum_{k=1}^m(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T(\tilde{\boldsymbol{\eta}}_{S^{(k)},i}-\boldsymbol{\eta}_{S^{(k)},i})$, $\boldsymbol{V}\boldsymbol{\Omega}_L\boldsymbol{V}^T=\boldsymbol{I}_L$, and $\boldsymbol{W}^{(k)}\boldsymbol{\Omega}_{S^{(k)}}(\boldsymbol{W}^{(k)})^T=\boldsymbol{I}_{S^{(k)}}$, for $k=1,\ldots,m$.

This proves the following result:

$$P\left(\sup_{t\in\mathcal{T}}\frac{|\tilde{X}_i(t)-X_i(t)|}{\sqrt{\boldsymbol{\phi}_{L,t}^T\boldsymbol{\Omega}_L\boldsymbol{\phi}_{L,t}+\sum_{k=1}^m(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T\boldsymbol{\Omega}_{S^{(k)}}\boldsymbol{\phi}^{(k)}_{S^{(k)},t}}}\leq\sqrt{\chi^2_{L+\sum_{k=1}^m S^{(k)},1-\alpha}}\right)\geq 1-\alpha. \tag{2.10}$$

Let $\omega(t)=\boldsymbol{\phi}_{L,t}^T\boldsymbol{\Omega}_L\boldsymbol{\phi}_{L,t}+\sum_{k=1}^m(\boldsymbol{\phi}^{(k)}_{S^{(k)},t})^T\boldsymbol{\Omega}_{S^{(k)}}\boldsymbol{\phi}^{(k)}_{S^{(k)},t}$ for $t\in\mathcal{T}$. By the triangle inequality,

$$|\widehat{X}_i(t)-X_i(t)|=|\widehat{X}_i(t)-\tilde{X}_i(t)+\tilde{X}_i(t)-X_i(t)|\leq|\widehat{X}_i(t)-\tilde{X}_i(t)|+|\tilde{X}_i(t)-X_i(t)|.$$

Therefore, it follows that

$$\sup_{t\in\mathcal{T}}\frac{|\widehat{X}_i(t)-X_i(t)|}{\sqrt{\omega(t)}}\leq\left(\sup_{t\in\mathcal{T}}\frac{|\widehat{X}_i(t)-\tilde{X}_i(t)|}{\sqrt{\omega(t)}}+\sup_{t\in\mathcal{T}}\frac{|\tilde{X}_i(t)-X_i(t)|}{\sqrt{\omega(t)}}\right)\sup_{t\in\mathcal{T}}\sqrt{\frac{\omega(t)}{\hat{\omega}(t)}}.$$

Let $A=\sup_{t\in\mathcal{T}}|\widehat{X}_i(t)-\tilde{X}_i(t)|/\sqrt{\omega(t)}$, $B=\sup_{t\in\mathcal{T}}|\tilde{X}_i(t)-X_i(t)|/\sqrt{\omega(t)}$, and $C=\sup_{t\in\mathcal{T}}\sqrt{\omega(t)/\hat{\omega}(t)}$. Note that $\omega(t)$ may be expressed as

$$\omega(t)=\underbrace{\boldsymbol{\phi}_{L,t}^T\boldsymbol{\Omega}_L\boldsymbol{\phi}_{L,t}}_{\omega_L(t,t)}+\underbrace{(\boldsymbol{\phi}^{(1)}_{S^{(1)},t})^T\boldsymbol{\Omega}_{S^{(1)}}\boldsymbol{\phi}^{(1)}_{S^{(1)},t}}_{\omega_{S^{(1)}}(t,t)}+\cdots+\underbrace{(\boldsymbol{\phi}^{(m)}_{S^{(m)},t})^T\boldsymbol{\Omega}_{S^{(m)}}\boldsymbol{\phi}^{(m)}_{S^{(m)},t}}_{\omega_{S^{(m)}}(t,t)}$$

$$= \omega_L(t,t) + \sum_{k=1}^{m} \omega_{S^{(k)}}(t,t).$$

Now, since $\omega_L(t,t)$ and $\omega_{S^{(1)}}(t,t), \ldots, \omega_{S^{(m)}}(t,t)$ are continuous, positive definite functions on the bounded interval $\mathcal{T}$, it follows that $\omega(t) = \omega_L(t,t) + \sum_{k=1}^{m} \omega_{S^{(k)}}(t,t)$ is also continuous and positive definite on $\mathcal{T}$. Therefore, $\omega(t)$ is bounded from above and below, say $0 < a \leq \omega(t) \leq b < \infty$. Observe that $\widehat{X}_i(t) - \tilde{X}_i(t) = \left(\widehat{X}_i^{(\mathrm{J})}(t) - \tilde{X}_i^{(\mathrm{J})}(t)\right) + \sum_{k=1}^{m} \left(\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right)$. According to Equations (12), (17), and (20) in [45], it follows that

$$\lim_{n\to\infty} \sup_{t\in\mathcal{T}} |\widehat{X}_i^{(\mathrm{J})}(t) - \tilde{X}_i^{(\mathrm{J})}(t)| \xrightarrow{p} 0,$$

and that $\lim_{n\to\infty} \sup_{t\in\mathcal{T}} |\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)| \xrightarrow{p} 0$, for all $k = 1, \ldots, m$.

Thus, by Slutsky's theorem,

$$\lim_{n\to\infty} \sum_{k=1}^{m} \sup_{t\in\mathcal{T}} |\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)| \xrightarrow{p} 0.$$

Furthermore, since

$$0 \leq \sup_{t\in\mathcal{T}} \left|\sum_{k=1}^{m} \left(\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right)\right|$$

$$\leq \sup_{t\in\mathcal{T}} \sum_{k=1}^{m} \left|\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right| \leq \sum_{k=1}^{m} \sup_{t\in\mathcal{T}} \left|\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right|,$$

it follows that

$$\lim_{n\to\infty} \sup_{t\in\mathcal{T}} \left|\sum_{k=1}^{m} \left(\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right)\right| \xrightarrow{p} 0.$$

Therefore, since

$$0 \leq \sup_{t\in\mathcal{T}} |\widehat{X}_i(t) - \tilde{X}_i(t)| = \sup_{t\in\mathcal{T}} \left|\left(\widehat{X}_i^{(\mathrm{J})}(t) - \tilde{X}_i^{(\mathrm{J})}(t)\right) + \sum_{k=1}^{m} \left(\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right)\right|$$

$$\leq \sup_{t\in\mathcal{T}} \left|\left(\widehat{X}_i^{(\mathrm{J})}(t) - \tilde{X}_i^{(\mathrm{J})}(t)\right)\right| + \sup_{t\in\mathcal{T}} \left|\sum_{k=1}^{m} \left(\widehat{X}_i^{(\mathrm{I},k)}(t) - \tilde{X}_i^{(\mathrm{I},k)}(t)\right)\right|,$$

then

$$\lim_{n\to\infty} \sup_{t\in\mathcal{T}} |\widehat{X}_i(t) - \tilde{X}_i(t)| \xrightarrow{p} 0,$$

which implies that $\lim_{n\to\infty} A \xrightarrow{p} 0$.

Now, the proof of Equation (22) in Yao et al. [45] establishes that $\lim_{n\to\infty}\hat{\omega}_L(t,t) \xrightarrow{p} \omega_L(t,t)$ and that $\lim_{n\to\infty}\hat{\omega}_{S^{(k)}}(t,t) \xrightarrow{p} \omega_{S^{(k)}}(t,t)$ for all $k = 1,\ldots,m$. Therefore, applying Slutsky's theorem, it follows that $\lim_{n\to\infty}\sum_{k=1}^{m}\hat{\omega}_{S^{(k)}}(t,t) \xrightarrow{p} \sum_{k=1}^{m}\omega_{S^{(k)}}(t,t)$. Moreover, since $\hat{\omega}(t) = \hat{\omega}_L(t,t) + \sum_{k=1}^{m}\hat{\omega}_{S^{(k)}}(t,t)$, another application of Slutsky's theorem gives the following result:

$$\lim_{n\to\infty}\hat{\omega}(t) \xrightarrow{p} \omega_L(t,t) + \sum_{k=1}^{m}\omega_{S^{(k)}}(t,t) = \omega(t).$$

It then follows that $\lim_{n\to\infty} C \xrightarrow{p} 1$. Next, note that

$$\begin{aligned}
&\left\{(A+B)C \geq \left(\epsilon + \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right)(1+\epsilon)\right\} \\
&\subseteq \left\{(A+B) \geq \left(\epsilon + \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right)\right\} \cup \left\{C \geq (1+\epsilon)\right\} \\
&\subseteq \left\{A \geq \epsilon\right\} \cup \left\{B \geq \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right\} \cup \left\{C \geq (1+\epsilon)\right\}.
\end{aligned}$$

Now, since $\lim_{n\to\infty} A \xrightarrow{p} 0$ and $\lim_{n\to\infty} C \xrightarrow{p} 1$, it follows that $P(A \geq \epsilon) \leq \delta/3$ and $P(C - 1 \geq \epsilon) \leq \delta/3$, for sufficiently large n. It was shown that $P\left(B \geq \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right) \leq \alpha$ in Equation (2.10). This implies

$$\lim_{n\to\infty} P\left((A+B)C \geq \left(\epsilon + \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right)(1+\epsilon)\right) \leq \alpha, \tag{2.11}$$

and then

$$\lim_{n\to\infty} P\left(\sup_{t\in\mathcal{T}} \frac{|\widehat{X}_i(t) - X_i(t)|}{\sqrt{\hat{\boldsymbol{\phi}}^T_{L,t}\hat{\boldsymbol{\Omega}}_L\hat{\boldsymbol{\phi}}_{L,t} + \sum_{k=1}^{m}(\hat{\boldsymbol{\phi}}^{(k)}_{S^{(k)},t})^T\hat{\boldsymbol{\Omega}}_{S^{(k)}}\hat{\boldsymbol{\phi}}^{(k)}_{S^{(k)},t}}} \leq \sqrt{\chi^2_{L+\sum_{k=1}^{m}S^{(k)},1-\alpha}}\right) \geq 1-\alpha, \tag{2.12}$$

by letting $\epsilon \to 0$.

Equation (2.12) provides the asymptotic simultaneous bands for $\widehat{X}_i(t) - X_i(t)$, for given and fixed L and $S^{(k)}$, $k = 1,\ldots,m$. From the Karhunen-Loève theorem, it follows that $|X_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t)| \xrightarrow{p} 0$ as $L \to \infty$, and that $|X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t)| \xrightarrow{p} 0$ as $S^{(k)} \to \infty$, for all $k = 1,\ldots,m$. Therefore, by Slutsky's theorem,

$$\sum_{k=1}^{m}|X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t)| \xrightarrow{p} 0 \text{ as } S^{(1)},\ldots,S^{(m)} \to \infty.$$

Now, note that

$$0 \leq \left| \sum_{k=1}^{m} \left(X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t) \right) \right| \leq \sum_{k=1}^{m} \left| X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t) \right|,$$

and so it follows that

$$\left| \sum_{k=1}^{m} \left(X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t) \right) \right| \xrightarrow{p} 0 \text{ as } S^{(1)}, \ldots, S^{(m)} \to \infty.$$

Thus,

$$\begin{aligned} 0 &\leq \left| \left(X_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t) \right) + \sum_{k=1}^{m} \left(X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t) \right) \right| \\ &= \left| \left(X_i^{(\mathrm{J})}(t) + \sum_{k=1}^{m} X_i^{(\mathrm{I},k)}(t) \right) - \left(X_i^{(\mathrm{J},\infty)}(t) + \sum_{k=1}^{m} X_i^{(\mathrm{I},k,\infty)}(t) \right) \right| \\ = |X_i(t) - X_i^{(\infty)}(t)| &\leq \left| X_i^{(\mathrm{J})}(t) - X_i^{(\mathrm{J},\infty)}(t) \right| + \left| \sum_{k=1}^{m} \left(X_i^{(\mathrm{I},k)}(t) - X_i^{(\mathrm{I},k,\infty)}(t) \right) \right|, \end{aligned}$$

from which it follows that

$$|X_i(t) - X_i^{(\infty)}(t)| \xrightarrow{p} 0 \text{ as } L, S^{(1)}, \ldots, S^{(m)} \to \infty,$$

after applying Slutsky's theorem. The Karhunen-Loève theorem implies that $\sup_{t \in \mathcal{T}} E[X_i(t) - X_i^{(\infty)}(t)]^2 \to 0$ as $L, S^{(1)}, \ldots, S^{(m)} \to \infty$. Therefore, ignoring a remaining approximation error that may be interpreted as bias, the $(1-\alpha)$ asymptotic simultaneous bands for $X_i^{(\infty)}(t)$ are given by

$$\widehat{X}_i(t) \pm \sqrt{\chi^2_{L+\sum_{k=1}^{m} S^{(k)}, 1-\alpha} \left[\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t} + \sum_{k=1}^{m} (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T \hat{\boldsymbol{\Omega}}_{S^{(k)}} \hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)} \right]}. \tag{2.13}$$

For a given trajectory, the asymptotic simultaneous confidence interval is always wider than the corresponding asymptotic pointwise confidence interval, since $\sqrt{\chi^2_{L+\sum_{k=1}^{m} S^{(k)}, 1-\alpha}} > \Phi^{-1}(1-\frac{\alpha}{2})$ for all $L + \sum_{k=1}^{m} S^{(k)} \geq 1$. The derivation of the asymptotic simultaneous confidence bands is very similar but less complicated for horizontally integrated data, since each trajectory only consists of a single individual component. In this case, Equation (2.13) simplifies to

$$\widehat{X}_i(t) \pm \sqrt{\chi^2_{L+S^{(k)}, 1-\alpha} \left[\hat{\boldsymbol{\phi}}_{L,t}^T \hat{\boldsymbol{\Omega}}_L \hat{\boldsymbol{\phi}}_{L,t} + (\hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)})^T \hat{\boldsymbol{\Omega}}_{S^{(k)}} \hat{\boldsymbol{\phi}}_{S^{(k)},t}^{(k)} \right]}, \tag{2.14}$$

where the superscript (k) represents the sample set associated with the i^{th} subject, $k \in \{1, \ldots, m\}$.

Chapter 3

Simulation Study

The simulation study performed in this thesis is described below. The study illustrates the advantage of using JIVE–FPCA over the traditional FPCA method in analyzing multi-block data. Horizontally integrated data were specifically generated for the simulation study so as to mimic the gestational weight dataset described in Chapter 4; however, similar studies may be carried out with vertically integrated multi-block data.

3.1 Experimental Set-up and Design

The data for the simulation study were generated according to the model given in Equation (2.1). A horizontally integrated dataset was constructed in which a single data type was measured on two different sets (groups) of samples (subjects). Each subject was either a member of Class 1 or Class 2. In total, $n = 1000$ iid normal samples were constructed, 500 from each class. The joint component mean function $\mu(t)$ was set to zero, as were the Class 1- and Class 2-specific (individual component) mean functions, $\mu^{(1)}(t)$ and $\mu^{(2)}(t)$, respectively, $0 \leq t \leq 40$. For simplicity, only one joint PC function, $\phi(t) = 4\cos(2\pi t/40)/\sqrt{20}$, was simulated. Similarly, a single PC function was generated for each class: $\phi^{(1)}(t) = 4\cos(\pi t/40)/\sqrt{20}$ and $\phi^{(2)}(t) = -4\cos(3\pi t/40)/\sqrt{20}$ for Class 1 and Class 2, respectively. Taken together, the joint and individual PC functions form an orthogonal set. The joint PC function represents the joint structure between the two sample sets (groups of subjects), whereas the class-specific PC functions represent structure individ-

ual to a given class.

The joint FPC scores ξ_i were generated from a $Normal(0, \lambda)$ distribution, where $\lambda = 25$. The class-specific FPC scores $\eta_i^{(k)}$ were generated from a $Normal(0, \lambda^{(k)})$ distribution, where $\lambda^{(1)} = 10$ and $\lambda^{(2)} = 8$ for Class 1 and Class 2, respectively. The measurement errors $\tilde{\epsilon}_{ij}$ in Equation (2.1) were assumed to be $Normal(0, 0.2)$ and independent of the random coefficients ξ_i and $\eta_i^{(k)}$, where $k \in \{1, 2\}$. Let $\mathcal{S} = \{s_1, \ldots, s_{100}\}$ be an equally spaced grid on $[0, 40]$ with $s_1 = 0$ and $s_{100} = 40$. Each curve was sampled at a random number of points, chosen from a discrete uniform distribution on $\{8, \ldots, 12\}$, and the locations of the measurements were randomly chosen from $\mathcal{S}$ without replacement. The sparsity of the simulated data reflected that of the gestational weight dataset in Chapter 4. Likewise, the simulated time points ranged from 0 to 40 to match the duration of a typical pregnancy.

3.2 Methodology

3.2.1 FVE Threshold and Algorithm Stopping Point

The principal analysis by conditional expectation (PACE) [45] package in MATLAB was used to implement FPCA. The original FPCA and JIVE–FPCA methods were applied to the simulated data. The number of eigenfunctions was chosen by the FVE method in each case. One purpose of the simulation study was to compare how well the two methods could recover the true PC functions. Careful selection of the FVE threshold was thus required so that precisely three PC functions were obtained when each method was applied. An FVE threshold of 90% resulted in the selection of 3 PCs for the FPCA method, with the first, second, and third PC functions accounting for 65.3%, 17.2%, and 10.1% of the total variation, respectively. Four iterations of JIVE–FPCA were necessary to satisfy the stopping criteria outlined in Section 2.2. Different FVE thresholds were required for each application of FPCA in the JIVE–FPCA algorithm in order to recover exactly one PC function for the joint component and each individual component. These thresholds were found through experimentation.

3.2.2 Proportion of Variance Explained

Another objective of the simulation study was to examine how well the JIVE–FPCA method could account for variation in the data unexplained by FPCA. The total variability within the set of observations Y_{ij} is measured by the total sum of squares (SST), which is given by

$$\text{SST} = \sum_{i=1}^{n} \sum_{j=1}^{N_i} (Y(T_{ij}) - \bar{Y})^2, \tag{3.1}$$

where $Y(T_{ij})$ is the j^{th} measurement for the i^{th} subject made at time $T_{ij} \in \mathcal{S}$, and $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{N_i} Y(T_{ij})$ is the overall mean of all observations. For the JIVE–FPCA model, the total sum of squares can be decomposed into two parts, a sum of squares indicating how much of the total variation in the data is explained by the joint component ($\text{SSR}_{\text{Joint}}$), and a sum of squares indicating what amount is unexplained by the joint component ($\text{SSE}_{\text{Joint}}$):

$$\begin{aligned} \text{SST} &= \text{SSR}_{\text{Joint}} + \text{SSE}_{\text{Joint}} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{N_i} (\widehat{Y}^{(\text{J})}(T_{ij}) - \bar{Y})^2 + \sum_{i=1}^{n} \sum_{j=1}^{N_i} (Y(T_{ij}) - \widehat{Y}^{(\text{J})}(T_{ij}))^2, \end{aligned} \tag{3.2}$$

where $\widehat{Y}^{(\text{J})}(T_{ij})$ is the estimate for $Y(T_{ij})$ given by the joint component of the i^{th} subject's JIVE–FPCA trajectory. The joint residual sum of squares $\text{SSE}_{\text{Joint}}$ may be further broken down into two terms, one for each class:

$$\text{SSE}_{\text{Joint}} = \text{SSE}_{\text{Joint},1} + \text{SSE}_{\text{Joint},2}, \tag{3.3}$$

where $\text{SSE}_{\text{Joint},k}$ is the joint residual sum of squares corresponding to the set of Class k observations, $k \in \{1, 2\}$. More specifically,

$$\text{SSE}_{\text{Joint},k} = \sum_{i=1}^{n^{(k)}} \sum_{j=1}^{N_i^{(k)}} (Y^{(k)}(T_{ij}) - \widehat{Y}^{(\text{J},k)}(T_{ij}))^2, \tag{3.4}$$

where the superscript (k) denotes the subject class, $n^{(k)}$ is the number of subjects in Class k, and $N_i^{(k)}$ is the number of measurements made on the i^{th} Class k subject. The sum of squares $\text{SSE}_{\text{Joint},k}$ may be decomposed even further into two components:

$$\text{SSE}_{\text{Joint},k} = \text{SSR}_{\text{Individual},k} + \text{SSE}_{\text{Joint + Individual},k}, \tag{3.5}$$

where $\text{SSR}_{\text{Individual},k}$ is a sum of squares term indicating how much of the variation unexplained by the Class k joint component is accounted for by the Class k individual component. The sum of squared differences between the observed values and their corresponding JIVE–FPCA estimates, for the set of Class k subjects, is given by $\text{SSE}_{\text{Joint + Individual},k}$. In summary, the total variability in the observations is partitioned as follows:

$$\begin{aligned}\text{SST} &= \text{SSR}_{\text{Joint}} + \underbrace{\underbrace{\text{SSR}_{\text{Individual},1} + \text{SSE}_{\text{Joint + Individual},1}}_{\text{SSE}_{\text{Joint},1}} + \underbrace{\text{SSR}_{\text{Individual},2} + \text{SSE}_{\text{Joint + Individual},2}}_{\text{SSE}_{\text{Joint},2}}}_{\text{SSE}_{\text{Joint}}} \\ &= \underbrace{\left(\text{SSR}_{\text{Joint}} + \text{SSR}_{\text{Individual},1} + \text{SSR}_{\text{Individual},2}\right)}_{\textbf{SSR}} + \underbrace{\left(\text{SSE}_{\text{Joint + Individual},1} + \text{SSE}_{\text{Joint + Individual},2}\right)}_{\textbf{SSE}}.\end{aligned} \tag{3.6}$$

The quantity

$$\text{PVE}_{\text{Individual},k} = \left(\frac{\text{SSR}_{\text{Individual},k}}{\text{SSE}_{\text{Joint},k}}\right) \times 100\% \tag{3.7}$$

gives the percentage of the variation unexplained by the Class k joint component that is explained by the Class k individual component. This value was calculated for both classes in the simulation study to demonstrate the benefit of including class-specific principal components in trajectory modelling.

3.2.3 Performance Measures: Mean Squared Error and Average Bandwidth

After estimating the trajectory of each subject using the FPCA and JIVE–FPCA procedures, the performance of the two methods was evaluated by comparing the corresponding mean squared errors (MSEs) and average bandwidths (ABWs). The MSE was calculated using the following equation:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_i}(Y(T_{ij}) - \widehat{Y}(T_{ij}))^2, \tag{3.8}$$

where n is the total number of observations in the dataset, N_i is the number of measurements made on the i^{th} subject, $Y(T_{ij})$ is the j^{th} measurement for the i^{th} subject made at time $T_{ij} \in \mathcal{S}$, and $\widehat{Y}(T_{ij})$ is the corresponding model-based estimate.

The ABW measures the average width of the set of 95% pointwise confidence intervals obtained from the FPCA and JIVE–FPCA methods. The 95% simultaneous confidence bands were not computed in the simulation study. The ABW is calculated by averaging the differences between the upper and lower confidence bounds of each subject's trajectory, and then taking the average of these means by summing them and dividing by the number of observations. That is,

$$\text{ABW} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{N_i}\sum_{j=1}^{N_i}(UCB(T_{ij}) - LCB(T_{ij}))\right], \tag{3.9}$$

where $UCB(T_{ij})$ and $LCB(T_{ij})$ are, respectively, the upper and lower confidence bounds of the i^{th} subject's trajectory at time $T_{ij} \in \mathcal{S}$. The distribution of averaged differences was found to be symmetric and contained no outliers, thus justifying the use of the mean as the measure of central tendency in Equation (3.9).

3.3 Results and Discussion

In Figure 1, the true PC functions are overlaid with the FPCA and JIVE–FPCA estimated PC functions. The graphs in this figure indicate that the JIVE–FPCA method is better at recovering the true PC functions than the original FPCA approach. This is particularly evident from the upper right and lower panels, wherein the JIVE–FPCA estimates are smoother, and more closely resemble the true Class 1- and Class 2-specific PC functions than the corresponding FPCA estimates. The fact that the JIVE–FPCA estimated class-specific PC functions are much closer to the ground-truth curves is expected, since this method is specifically designed to be able to uncover the patterns that are unique to each class (sample set).

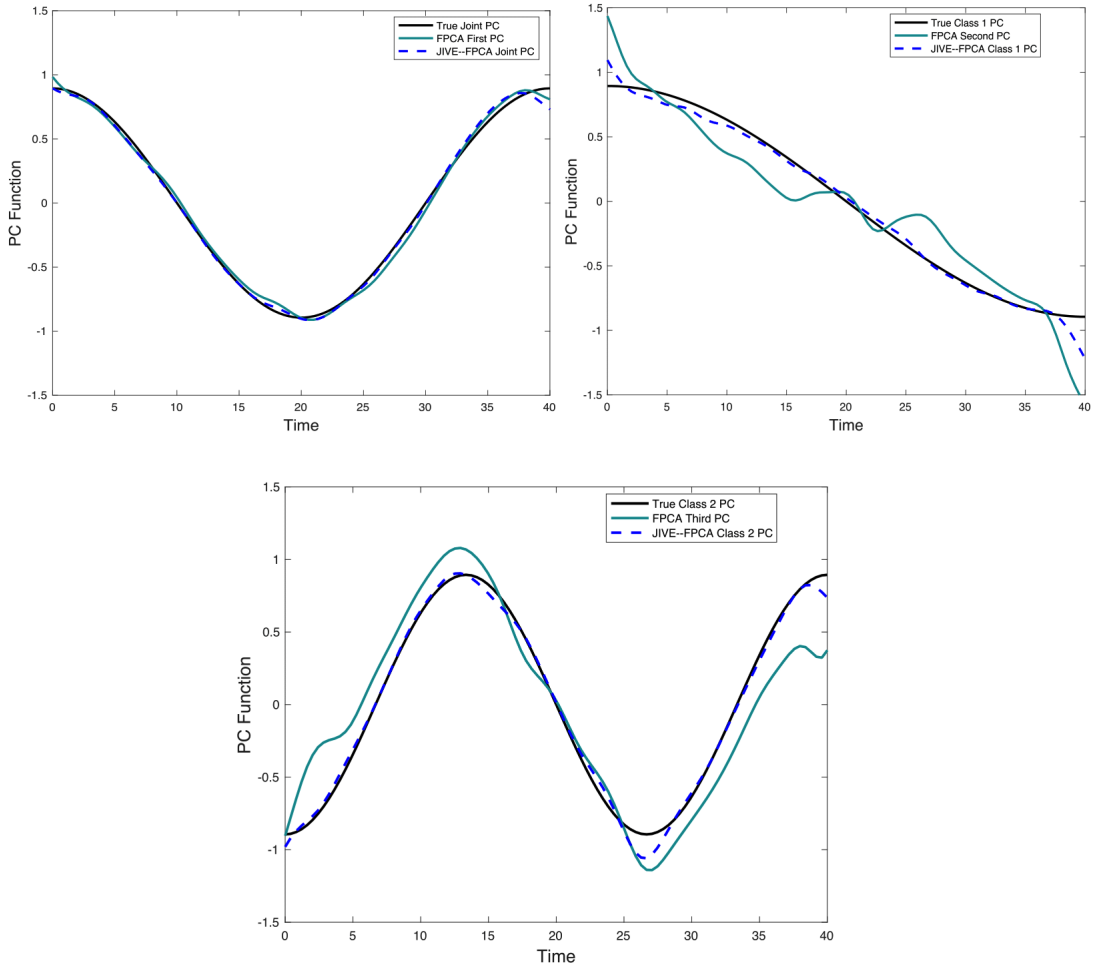

Figure 1: True PC Functions Overlaid with the FPCA and JIVE–FPCA Estimated PC Functions. The JIVE–FPCA estimates are smoother, and more closely resemble the ground-truth curves than the corresponding FPCA estimates. This is especially evident in the upper right and lower panels depicting the class-specific PC functions.

Figure 2 shows the FPCA and JIVE–FPCA estimated trajectories overlaid with the true trajectories of two subjects, one from each class. The graphs in this figure demonstrate that the JIVE–FPCA method provides a significantly better fit to individual trajectories than the original FPCA approach. Indeed, the JIVE–FPCA estimated trajectories are evidently smoother, and visually seem to more closely approximate the true trajectories. The 95% pointwise confidence intervals obtained from JIVE–FPCA also appear to be smoother and narrower than the corresponding FPCA estimates. Observe that

the JIVE–FPCA confidence interval is completely contained within the FPCA confidence interval at several time points in each graph.

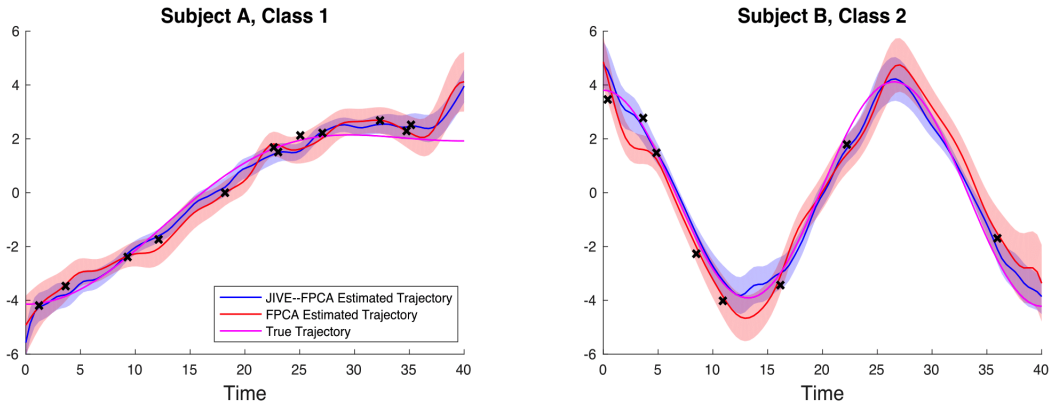

Figure 2: FPCA and JIVE–FPCA estimated trajectories overlaid with the true trajectories of two subjects, one from each class. The JIVE–FPCA estimated trajectories are evidently smoother, and visually seem to better approximate the true trajectories. The 95% pointwise confidence intervals obtained from JIVE–FPCA (shaded in blue) also appear to be smoother and narrower than the corresponding FPCA estimates (shaded in red).

These observations are supported well by numerical calculations, which show that after four iterations, the JIVE–FPCA method resulted in a 4.35% reduction in the MSE and a 37.1% decrease in the ABW compared to FPCA. To demonstrate the superior performance of the JIVE–FPCA method, Table 1 reports the MSE and ABW values obtained by both approaches.

Table 1: Performance comparison of the FPCA and JIVE–FPCA methods applied to simulated data. The mean squared error and average bandwidth are lower for the JIVE–FPCA method, thus illustrating the superior performance of this method compared to the original FPCA approach.

Method	Mean Squared Error (MSE)	Average Bandwidth (ABW)
FPCA	2.30	1.43
JIVE–FPCA (After 4 iterations)	2.20	0.90

Further calculations showed that the percentage of the total variation in the data explained by the joint component was

$$\text{PVE}_{\text{Joint}} = \left(\frac{\text{SSR}_{\text{Joint}}}{\text{SST}} \right) \times 100\% = 70.6\%. \tag{3.10}$$

Therefore, 29.4% of the total variation was unexplained by the joint component. That is,

$$\text{PVUE}_{\text{Joint}} = \left(\frac{\text{SSE}_{\text{Joint}}}{\text{SST}} \right) \times 100\% = 29.4\%. \tag{3.11}$$

The percentage of this unexplained variation associated with Class k is defined as

$$\text{PVUE}_{\text{Joint},k} = \left(\frac{\text{SSE}_{\text{Joint},k}}{\text{SSE}_{\text{Joint}}} \right) \times 100\%. \tag{3.12}$$

In the simulation study, it was found that 54.7% of the variation unexplained by the joint component was associated with Class 1, while 45.3% was associated with Class 2. In other words,

$$\text{PVUE}_{\text{Joint},1} = 54.7\% \quad \text{and} \quad \text{PVUE}_{\text{Joint},2} = 45.3\%.$$

More interestingly, the Class 1 individual component explained 95.1% of the variation unexplained by the Class 1 joint component. That is, $\text{PVE}_{\text{Individual},1} = 95.1\%$. The value of $\text{PVE}_{\text{Individual},2}$ was also very high at 94.0%, which indicates that the Class 2 individual component accounted for nearly all of the variation unexplained by the Class 2 joint component. These findings demonstrate the usefulness of the JIVE–FPCA method: the class-specific PCs that are incorporated in trajectory modelling explain a significant amount of unexplained variation in the data.

Chapter 4

Application to Gestational Weight Data

This chapter describes the application of JIVE–FPCA to real-world, horizontally integrated gestational weight data. Each data block contained sparse, longitudinal gestational weight measurements for subjects of a specific pre-pregnancy BMI category. Since four pre-pregnancy BMI categories were considered, there were four blocks of data (groups of subjects) in total.

4.1 Materials: Data Collection and Inclusion Criteria

Pregnant women less than 27 weeks gestation and greater than 16 years of age were enrolled in a prospective longitudinal cohort, the Alberta Pregnancy Outcomes and Nutrition (APrON) study, between 2009–2012 [17]. Detailed descriptions of participants and the study are published elsewhere [2, 17]. Ethics approval for the APrON study was obtained from the Health Research Ethics Boards at the University of Alberta (Pro 00002954) and the University of Calgary (E22101). Women provided written informed consent prior to enrollment.

Data were collected at 2–3 study visits during pregnancy, spaced to coincide with each trimester, and at one follow-up visit approximately 3 months postpartum. Upon recruitment, each woman's pre-pregnancy weight and due date were self-reported. Women recruited before 13 weeks gestation were as-

sessed in each trimester, whereas those recruited between 14 and 27 weeks gestation were assessed in trimesters 2 and 3. Each assessment included a weight and height measurement. The highest weight during pregnancy and delivery date were self-reported during the postpartum visit. Women were classified as underweight (< 18.5), normal (18.5–24.9), overweight (< 25.0–29) or obese (≥ 30) according to their pre-pregnancy BMI [6]. Gestational age (GA) was calculated based on due dates.

Subjects with a singleton live birth were included in the dataset. Clinical weight measurements taken at regular prenatal visits were also obtained from the included APrON study participants. The GAs corresponding to APrON and clinical weight records ranged from 0 to 43 weeks, with the minimum clinical weight GA being 4 weeks. For statistical modelling, the subjects were required to have a known pre-pregnancy weight, a height measurement, and at least one of each type of weight record (APrON and clinical) with corresponding GAs.

Each weight trajectory could have a maximum of 5 APrON body weight data points (pre-pregnancy weight, highest weight during pregnancy, and one for each trimester of pregnancy). Several APrON data points were often missing from the trajectories, however, due to skipped study visits. Clinical weight measurements often occurred at irregularly spaced time points as well, due to missing or incorrectly recorded data. On average, there were 7 clinical weight data points per subject. The pre-pregnancy weight was assumed to be the weight measured at time $t = 0$, and the highest weight during pregnancy was the weight associated with the GA at birth.

4.2 Methods

As in Chapter 3, the FPCA and JIVE–FPCA methods were applied to the gestational weight data ($n = 1648$), with the number of eigenfunctions chosen by the FVE method. An FVE threshold of 99% was used for FPCA, which resulted in the selection of 3 PCs. The first, second, and third PC functions for the weight data accounted for 94.8%, 2.8%, and 1.6% of the total varia-

tion, respectively. Four iterations of the JIVE–FPCA algorithm were required to meet the stopping criteria defined in Section 2.2. Several experiments were performed to determine how many PCs should be selected for the joint and individual components. Ultimately, the goal was to find the PCs that explained the largest proportion of variation in the data. This objective was met by extracting one PC function for the joint component; two PC functions for the normal and overweight individual components, and 3 PCs for the underweight and obese individual components. Higher principal components primarily described noise and were thus not extracted. Different FVE thresholds were required for each application of FPCA in the JIVE–FPCA algorithm in order to obtain the number of PCs specified above. These thresholds were found through trial and error.

Once the predicted trajectories and 95% pointwise confidence bands were computed for each subject, the FPCA and JIVE–FPCA methods were compared on the basis of their MSEs and ABWs. The estimated mean weight trajectory from FPCA was then compared with the mean weight functions obtained for each pre-pregnancy BMI category after four iterations of JIVE–FPCA. For a given pre-pregnancy BMI class, the mean weight trajectory was calculated by averaging the weight values predicted by JIVE–FPCA at each time point.

4.3 Results and Discussion

Figure 3 displays the estimated weight trajectories of four subjects, one from each pre-pregnancy BMI class. The FPCA and JIVE–FPCA estimated trajectories are superimposed in each panel. Overall, this figure shows that for each subject, the trajectory estimated by JIVE–FPCA gives much better estimates for the observed weights. This is particularly evident for subjects with normal and overweight pre-pregnancy BMI (Subjects B and C, respectively). Observe that for Subject B, the JIVE–FPCA estimated trajectory is significantly smoother at lower gestational age values than the trajectory predicted by FPCA. Similarly, the FPCA fitted trajectory for Subject C gives a poor

estimate for the first clinical weight observation, which yields a large residual.

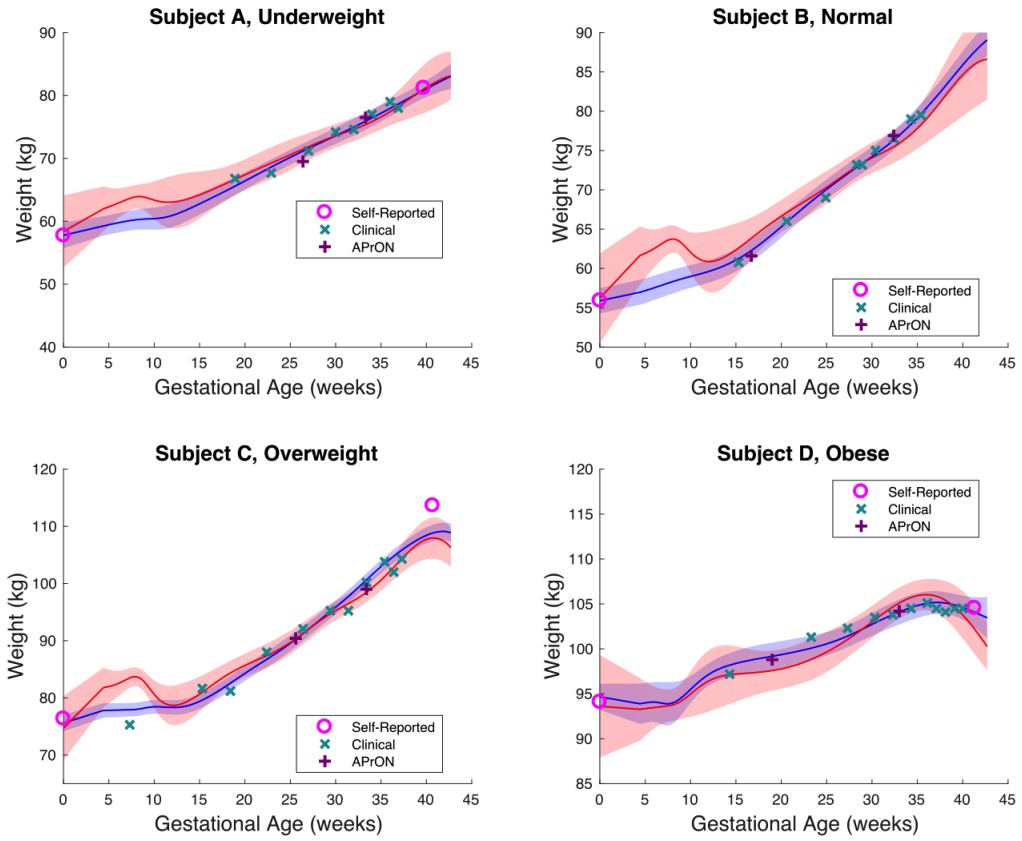

Figure 3: Estimated weight trajectories of four subjects, one from each pre-pregnancy BMI category. The estimated trajectories (solid lines) and corresponding 95% pointwise confidence intervals (shaded areas) from the FPCA (red) and JIVE–FPCA (blue) methods are superimposed in each panel. Different symbols are used to represent different types of weight observations, as indicated by the legend.

The 95% pointwise confidence intervals estimated from JIVE–FPCA are evidently smoother and narrower for each subject, and are largely contained within the corresponding FPCA confidence intervals. The superior model fit using the JIVE–FPCA method is confirmed numerically in Table 2, which shows that the MSE and ABW after four iterations of JIVE–FPCA are, respectively, 15.4% and 46.0% lower than the corresponding values obtained from the original FPCA approach. These results are consistent with those found in the simulation study.

Table 2: Performance comparison of the FPCA and JIVE–FPCA methods applied to gestational weight data.

Method	Mean Squared Error (MSE)	Average Bandwidth (ABW)
FPCA	16.34	5.59
JIVE–FPCA (After 4 iterations)	13.83	3.02

The mean weight trajectory from FPCA is shown in Figure 4, along with the mean weight functions for each pre-pregnancy BMI category, obtained after four iterations of JIVE–FPCA. Together, Figures 3 and 4 illustrate that the JIVE–FPCA method successfully captures differences in gestational weight gain patterns among different pre-pregnancy BMI groups. For instance, Figure 3 shows that Subject D, with obese pre-pregnancy BMI, appears to exhibit an overall slower rate of weight gain than Subjects A–C, and even lost weight at the end of the third trimester.

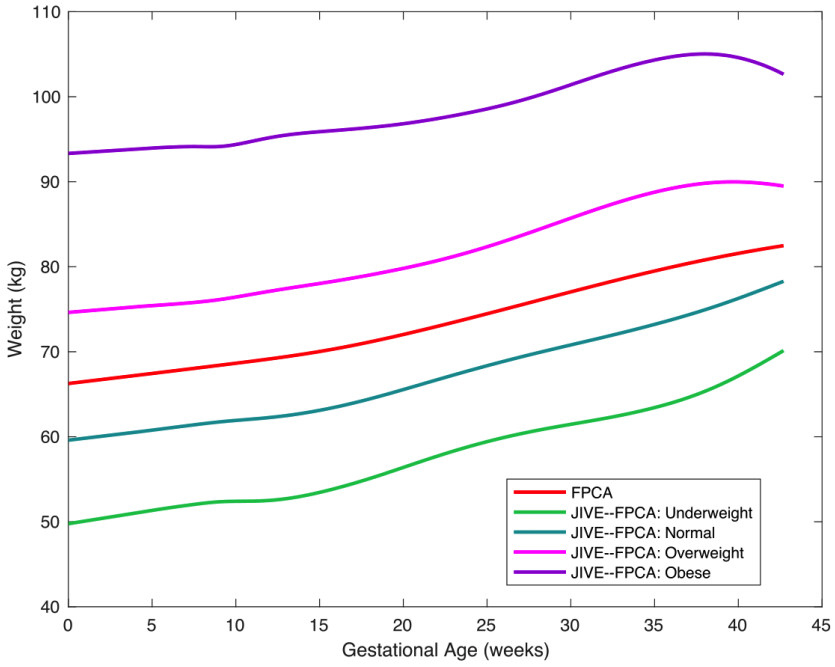

Figure 4: Estimated mean function from FPCA alongside the mean weight trajectories for each pre-pregnancy BMI group, obtained after four iterations of JIVE–FPCA. The FPCA mean function was estimated based on all the observations from all subjects, and was directly extracted. For a given pre-pregnancy BMI category, the mean weight trajectory was computed by averaging the weight values predicted by JIVE–FPCA at each time point.

Figure 4 confirms that the mean weight trajectory for women with pre-pregnancy obesity is more gradually rising than (a) the mean weight growth trajectory estimated for all subjects using the FPCA method, and (b) the mean JIVE–FPCA weight trajectories for women with lower pre-pregnancy BMI. In general, the mean weight trajectories for overweight and obese women appear to exhibit a similar trend that is different from the trend exhibited by the mean weight curves for women with underweight and normal pre-pregnancy BMI. The graph shows that on average, women with higher pre-pregnancy BMI tend to exhibit weight loss at the end of the third trimester and experience a smaller total GWG. Underweight and normal weight women have mean growth curves that are steeper than the mean weight trajectory estimated by FPCA, and lack the dip at approximately 38 weeks gestation present in the curves for overweight and obese women. The different trends displayed by the different pre-pregnancy BMI classes in Figure 4 agree with those reported in the literature [23, 31]. These results could be useful in developing targeted interventions aimed at optimizing gestational weight gain for long-term maternal and infant health.

As expected, the results also show that the JIVE–FPCA method leads to a significant improvement in explaining the weight variation. Table 3 below summarizes information on the percentage of variation explained by the joint and individual components, using quantities previously defined in Equations (3.7), (3.10)–(3.12). The subscripts $1, 2, 3$, and 4 associated with some of these quantities are used to represent the underweight, normal, overweight, and obese subject classes, respectively.

Table 3: Percentage of weight variation explained by the joint and individual components after four iterations of JIVE–FPCA.

$\text{PVE}_{\text{Joint}}$	97.2%	$\text{PVUE}_{\text{Joint}}$	2.8%
$\text{PVE}_{\text{Individual},1}$	83.6%	$\text{PVUE}_{\text{Joint},1}$	3.0%
$\text{PVE}_{\text{Individual},2}$	78.0%	$\text{PVUE}_{\text{Joint},2}$	51.6%
$\text{PVE}_{\text{Individual},3}$	81.3%	$\text{PVUE}_{\text{Joint},3}$	27.3%
$\text{PVE}_{\text{Individual},4}$	85.5%	$\text{PVUE}_{\text{Joint},4}$	18.1%

The advantage offered by JIVE–FPCA is evident in the left column of Table

3. Observe that the $\text{PVE}_{\text{Individual},k}$ values are large for all $k \in \{1, 2, 3, 4\}$ (i.e., for all pre-pregnancy BMI classes). This implies that each BMI class-specific individual component explains a large percentage of the weight variation unexplained by the corresponding class-specific joint component. These results agree with those obtained in the simulation study, where the $\text{PVE}_{\text{Individual},k}$ values were also found to be high (i.e., in the mid-90s) for both simulated classes.

4.4 Future Work: Vertical JIVE–FPCA

As a final point, it is worth noting that efforts were made to carry out a vertical JIVE–FPCA analysis on the gestational weight dataset. The fact that the longitudinal weight measurements were obtained from multiple data sources made such an analysis viable. In this analysis, each data block consisted of weight measurements of a particular type. The self-reported and APrON weight measurements were pooled together in one block, while the clinical weight measurements were included in another. Since the subjects in this study were required to have at least one of each type of weight record, data were included for every subject in each block. Each subject's JIVE–FPCA trajectory consisted of a joint component and two individual components, one for each data type.

While it was anticipated that incorporating additional weight category-specific principal components would result in an improved fit to individual gestational weight gain trajectories, preliminary work showed the opposite result. The application of JIVE–FPCA to vertically integrated gestational weight data actually resulted in a larger MSE and ABW compared to the FPCA method. Moreover, the model fit kept worsening with additional iterations of the JIVE–FPCA algorithm (i.e., the MSE became larger with every iteration). Further investigations showed that this was due to the bias inherent in self-reported weight measurements, which tend to be under-reported. In the future, it would be interesting to examine how the JIVE–FPCA algorithm performs on a more suitable, vertically integrated multi-block dataset.

Chapter 5

Conclusion

In summary, this thesis presents a new method, called JIVE–FPCA, which is an extension of the traditional FPCA method. The JIVE–FPCA approach was developed out of the desire to more accurately estimate individual gestational weight gain trajectories; however, the method can be used with any sparse, longitudinal multi-block dataset. This novel approach is useful because it simultaneously captures patterns that are shared among multiple data blocks and patterns that are unique to a particular block.

The JIVE–FPCA advantage was confirmed by the results obtained with simulated and real, gestational weight data. Both studies showed that the JIVE–FPCA method accounts for a significant proportion of the variation unexplained by FPCA. Visually, the estimated JIVE–FPCA trajectories were much smoother, and more closely fit the observed data points than the corresponding FPCA estimates. The JIVE–FPCA pointwise confidence bands were noticeably smoother and narrower as well. These findings were supported by numerical calculations, which showed that the MSE and ABW were significantly smaller for JIVE–FPCA compared to FPCA. The JIVE–FPCA results with gestational weight data particularly emphasize the different trends in GWG among different pre-pregnancy BMI categories, and can be used for clinical counselling. As the studies in this thesis were primarily focused on the application of JIVE–FPCA to horizontally integrated data, future work could include investigating the performance of this new method on vertically integrated multi-block datasets.

Bibliography

1. Agarwal, A., El-Ghazawi, T., El-Askary, H. & Le-Moigne, J. Efficient hierarchical-PCA dimension reduction for hyperspectral imagery. *2007 IEEE International Symposium on Signal Processing and Information Technology,* 353–356 (2007).
2. Begum, F., Colman, I., McCargar, L. J. & Bell, R. C. Gestational weight gain and early postpartum weight retention in a prospective cohort of Alberta women. *Journal of Obstetrics and Gynaecology Canada* **34,** 637–647 (2012).
3. Bekaert, G., Hodrick, R. J. & Zhang, X. International stock return comovements. *The Journal of Finance* **64,** 2591–2626 (2009).
4. Boulesteix, A.-L. PLS dimension reduction for classification with microarray data. *Statistical Applications in Genetics and Molecular Biology* **3,** 1–30 (2004).
5. Campbell, S. Fetal macrosomia: a problem in need of a policy. *Ultrasound in Obstetrics and Gynecology* **43,** 3–10 (2014).
6. *Canadian guidelines for body weight classification in adults.* (Ottawa: Health Canada, 2003).
7. Che, M., Kong, L., Bell, R. C. & Yuan, Y. Trajectory modeling of gestational weight: a functional principal component analysis approach. *PLoS ONE* **12,** 1–15 (2017).
8. Cogswell, M. E., Serdula, M. K., Hungerford, D. W. & Yip, R. Gestational weight gain among average-weight and overweight women – what is excessive? *American Journal of Obstetrics and Gynecology* **172,** 705–712 (1995).
9. De Lathauwer, L. & Nion, D. Decompositions of a higher-order tensor in block terms - part III: alternating least squares algorithms. *Tensor Decompositions and Applications* **30,** 1067–1083 (2008).
10. De Leeuw, J., Young, F. W. & Takane, Y. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika* **41,** 471–503 (1976).

11. De Tayrac, M., Lê, S., Aubry, M., Mosser, J. & Husson, F. Simultaneous analysis of distinct omics data sets with integration of biological knowledge: multiple factor analysis approach. *BMC Genomics* **10,** 32 (2009).

12. Donoho, D. L. *High-dimensional data analysis: the curses and blessings of dimensionality* in *AMS conference on Mathematical Challenges of the 21st Century* (2000).

13. Hediger, M. L., Scholl, T. O., Schall, J. I., Healey, M. F. & Fischer, R. L. Changes in maternal upper arm fat stores are predictors of variation in infant birth weight. *The Journal of Nutrition* **124,** 24–30 (1994).

14. Helms, E., Coulson, C. C. & Galvin, S. L. Trends in weight gain during pregnancy: a population study across 16 years in North Carolina. *American Journal of Obstetrics and Gynecology* **194,** e32–e34 (2006).

15. Jere, S. *et al. Extracting commuting patterns in railway networks through matrix decompositions* in *13th International Conference on Control, Automation, Robotics & Vision, ICARCV 2014, Singapore, December 10-12, 2014* (2014), 541–546.

16. Johnstone, I. M. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics* **29,** 295–327 (2001).

17. Kaplan, B. J. *et al.* The Alberta Pregnancy Outcomes and Nutrition (APrON) cohort study: rationale and methods. *Maternal and Child Nutrition* **10,** 44–60 (2014).

18. Kim, H. & Park, H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* **23,** 1495–1502 (2007).

19. Kuligowski, J. *et al.* Analysis of multi-source metabolomic data using joint and individual variation explained (JIVE). *The Analyst* **140,** 4521–4529 (2015).

20. Li, G.-Z., Zeng, X.-Q., Yang, J. Y. & Yang, M. Q. Partial least squares based dimension reduction with gene selection for tumor classification. *2007 IEEE 7th International Symposium on BioInformatics and BioEngineering,* 1439–1444 (2007).

21. Lock, E. F., Hoadley, K. A., Marron, J. S. & Nobel, A. B. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics* **7,** 523–542 (2013).

22. Löfstedt, T. & Trygg, J. OnPLS – a novel multiblock method for the modelling of predictive and orthogonal variation. *Journal of Chemometrics* **25,** 441–455 (2011).

23. Lowell, H. & Miller, D. C. Weight gain during pregnancy: adherence to Health Canada's guidelines. *Health Reports* **21,** 31–36 (2010).

24. McCormick, M. C. The contribution of low birth weight to infant mortality and childhood morbidity. *The New England Journal of Medicine* **312,** 82–90 (2 1985).

25. McIntire, D. D., Bloom, S. L., Casey, B. M. & Leveno, K. J. Birth weight in relation to morbidity and mortality among newborn infants. *The New England Journal of Medicine* **340,** 1234–1238 (1999).

26. Nguyen, D. V. & Rocke, D. M. On partial least squares dimension reduction for microarray-based classification: a simulation study. *Computational Statistics and Data Analysis* **46,** 407–425 (2004).

27. Nohr, E. A. *et al.* Combined associations of prepregnancy body mass index and gestational weight gain with the outcome of pregnancy. *The American Journal of Clinical Nutrition* **87,** 1750–1759 (2008).

28. Olson, C. M. Achieving a healthy weight gain during pregnancy. *Annual Review of Nutrition* **28,** 411–423 (2008).

29. Olson, C. M., Strawderman, M. S. & Reed, R. G. Efficacy of an intervention to prevent excessive gestational weight gain. *American Journal of Obstetrics and Gynecology* **191,** 530–536 (2004).

30. Polley, B. A., Wing, R. R. & Sims, C. J. Randomized controlled trial to prevent excessive weight gain in pregnant women. *International Journal of Obesity* **26,** 1494–1502 (2002).

31. Rasmussen, K. M. & Yaktine, A. L. *Weight gain during pregnancy: reexamining the guidelines* (Washington, DC: National Academies Press, 2009).

32. Rauh, K. *et al.* Safety and efficacy of a lifestyle intervention for pregnant women to prevent excessive maternal weight gain: a cluster-randomized controlled trial. *BMC Pregnancy and Childbirth* **13,** 1–11 (2013).

33. Ray, P., Zheng, L., Lucas, J. & Carin, L. Bayesian joint analysis of heterogeneous genomics data. *Bioinformatics* **30,** 1370–1376 (2014).

34. Rooney, B. L. & Schaugerger, C. W. Excess pregnancy weight gain and long-term obesity: one decade later. *Obstetrics and Gynecology (New York. 1953)* **100,** 245–252 (2002).

35. Schieve, L. A., Cogswell, M. E. & Scanlon, K. S. Trends in pregnancy weight gain within and outside ranges recommended by the Institute of Medicine in a WIC population. *Maternal and Child Health Journal* **2,** 111–116 (1998).

36. Schmitt, N. M., Nicholson, W. K. & Schmitt, J. The association of pregnancy and the development of obesity: results of a systematic review and meta-analysis on the natural history of postpartum weight retention. *International Journal of Obesity* **31,** 1642–1651 (2007).

37. Schouteden, M., Van Deun, K., Wilderjans, T. F. & Van Mechelen, I. Performing DISCO-SCA to search for distinctive and common information in linked data. *Behavior Research Methods* **46,** 576–587 (2014).

38. Sidebottom, A. C., Brown, J. E. & Jacobs, D. R. Pregnancy-related changes in body fat. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* **94,** 216–223 (2001).

39. Subasi, A. & Gursoy, I. M. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications* **37,** 8659–8666 (2010).

40. Tanentsapf, I., Heitmann, B. L. & Adegboye, A. R. Systematic review of clinical trials on dietary interventions to prevent excessive weight gain during pregnancy among normal weight, overweight and obese women. *BMC Pregnancy and Childbirth* **11,** 81–92 (2011).

41. Van Deun, K. *et al.* DISCO-SCA and properly applied GSVD as swinging methods to find common and distinctive processes. *PLoS ONE* **7,** 1–13 (2012).

42. Viswanathan, M. *et al.* Outcomes of maternal weight gain. *Evidence Report/Technology Assessment,* 1–223 (2008).

43. *What Mothers Say: The Canadian Maternity Experiences Survey.* (Ottawa: Public Health Agency of Canada, 2009).

44. Yang, Z. & Michailidis, G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* **32,** 1–8 (2016).

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

46. Zhang, D. & Zhou, Z.-H. Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing* **69,** 224–231 (2005).

47. Zhou, G., Cichocki, A., Zhang, Y. & Mandic, D. P. Group component analysis for multiblock data: common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems* **27,** 2426–2439 (2016).