Principal Differential Analysis With Covariates: A Simulation Study On The Effect Of The Smoothing Parameters

Indika Varuna Mallawaarachchi

University of Texas at El Paso, ivmallawaarachchi@miners.utep.edu

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd

Part of the Statistics and Probability Commons

Recommended Citation


https://digitalcommons.utep.edu/open_etd/2333
PRINCIPAL DIFFERENTIAL ANALYSIS WITH COVARIATES:
A SIMULATION STUDY ON THE EFFECT OF THE SMOOTHING PARAMETERS

INDIKA VARUNA MALLAWAARACHCHI

Department of Mathematical Sciences

APPROVED:

Joan Staniswalis, Ph.D., Chair

Ming-Ying Leung, Ph.D.

James Wood, Ph.D.

Benjamin Flores, Ph.D.
Acting Dean of the Graduate School
to my

FAMILY

with love
PRINCIPAL DIFFERENTIAL ANALYSIS WITH COVARIATES:
A SIMULATION STUDY ON THE EFFECT OF THE SMOOTHING PARAMETERS
by
INDIKA VARUNA MALLAWAARACHCHI

THESIS
Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

August 2011
Acknowledgements

First of all I would like to express my gratitude to my advisor, Dr. Joan Staniswalis from the Mathematical Sciences Department at The University of Texas at El Paso, for her advice, encouragement and financial support. She always advised and helped me when I was struggling. I was so lucky to have her as my advisor.

I wish to thank the other members of my committee, Dr. M. Leung from the Mathematical Sciences Department and Dr. James Wood from the Psychology Department, both at The University of Texas at El Paso. Their comments, guidance and flexibility were so helpful for the completion of this work.

I also wish to thank Dr. Ori Rosen, Dr. Naijun Sha, Dr. Amy Wagler, Dr. Panagis Moschopoulos and Dr. Hamdie Dogan from the Mathematical Sciences Department at The University of Texas at El Paso for stimulating my knowledge. I also want to thank all of the other professors and staff of the Mathematical Sciences Department at The University of Texas at El Paso for all their support.

And finally, I must thank my dear family and friends for all the support and encouragement throughout the program.

NOTE: This thesis was submitted to my Supervising Committee on the July 22, 2011.
Abstract

Principal Differential Analysis deals with functional data. The word functional data refers to a collection of curves that are independent and measured on a dense grid of time points in an interval. These time points can be equally or unequally spaced. A differential equation is believed capable of capturing the features of these $n$ curves.

Ramsay (1996) [13] first introduced Principal Differential Analysis (PDA) as an alternative to the Principal Component Analysis (PCA). PDA finds a linear differential equation that captures features of a collection of curves, in order to have a low dimensional approximation for functional data. PDA is based on the theorem: the span of the functions $f_1, \ldots, f_m$ with $m$ derivatives has an annihilating linear differential operator (LDO) of the form $L = w_0 I + w_1 D + \ldots, w_{m-1} D^{m-1} + w_m D^m$.

In PDA the coefficients $w_0, \ldots, w_m$ of the LDO are estimated using data for a specified $m$. The sum of squared norms of the residuals with penalty is used as the fitting criterion. Here the residual is that part of the data curve that is not annihilated by the LDO. The penalty of Eilers and Marx (1996) [3] is used to impose the smoothness. A low-dimensional approximation of the curve data can be obtained by a linear combination of the null space basis functions.

Jin (2006) [6] developed the theory for PDA with covariates when functional data collected from experimental units are described by covariates. This thesis is a Monte-Carlo simulation study to identify the effect of the smoothing parameters for the bias and variance of the estimators in PDA with covariates. Implementations of Jin (2006) [6] were translated from Splus to R and were made to run more efficiently. A hearing data set taken from Wood (2007) [16] was analyzed using PDA with covariates.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2 Nonparametric Regression with P-Splines</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Introduction to Nonparametric Regression</td>
<td>3</td>
</tr>
<tr>
<td>2.2 Smoothing Splines</td>
<td>4</td>
</tr>
<tr>
<td>2.3 B-Splines</td>
<td>7</td>
</tr>
<tr>
<td>2.4 P-Splines</td>
<td>8</td>
</tr>
<tr>
<td>3 Principal Component Analysis</td>
<td>10</td>
</tr>
<tr>
<td>3.1 Principal Component Analysis for Multivariate Data</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Best linear model for continuous processes</td>
<td>19</td>
</tr>
<tr>
<td>4 Principal Differential Analysis</td>
<td>29</td>
</tr>
<tr>
<td>4.1 Linear Differential Operator</td>
<td>29</td>
</tr>
<tr>
<td>4.2 Low Dimensional Approximation for Functional Data</td>
<td>33</td>
</tr>
<tr>
<td>5 Principal Differential Analysis with Covariates</td>
<td>40</td>
</tr>
<tr>
<td>5.1 Linear Differential Operator with a Covariate</td>
<td>40</td>
</tr>
<tr>
<td>6 Simulations</td>
<td>44</td>
</tr>
<tr>
<td>7 Analysis of Data</td>
<td>69</td>
</tr>
<tr>
<td>8 Conclusion</td>
<td>88</td>
</tr>
<tr>
<td>References</td>
<td>89</td>
</tr>
</tbody>
</table>
Appendix

A More Simulation Results .................................................... 91
A.1 Without Noise ................................................................. 92
A.2 With Noise ................................................................. 104
Curriculum Vitae ............................................................... 116
List of Tables

6.1a $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_1$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0$) ........................................ 46

6.1b $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_2$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0$) ........................................ 47

6.2a $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_1$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0.15$) ........................................ 48

6.2b $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_1$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0.15$) ........................................ 49
List of Figures

1.1 Twenty records of lip position ................................................. 2
2.1 Linear, Quadratic and Cubic B-Splines ................................. 8
3.1 Daily ozone values for 47 weeks. ............................................ 16
3.2 Scree Plot for the ozone data. ............................................... 17
3.3 First two PCs for ozone data: Dashed line is the first PC, Solid line is the
second PC. ............................................................................. 17
3.4 Data approximation using the first two PCs for ozone data. .......... 18
3.5 First PC Score for ozone data. ............................................... 18
3.6 Twenty records of lip position after smoothing with smoothing parameter
value $1 \times e^{-12}$. ................................................................... 22
3.7 First two PC functions for lip data. ......................................... 23
3.8 Data approximation using the first two PC functions for lip data: Solid
curves are the observations and dashed curves are their approximations. ... 23
3.9 Simulated data after smoothing with smoothing parameter value $1 \times e^{-3}$. 24
3.10 First two PC functions for simulated data. ............................. 25
3.11 Data approximation using the first two PC functions for simulated data
(Observations and approximations are overlapped). ..................... 25
3.12 Evoked potential of the individuals below 7 years of age with normal hearing. 27
3.13 First two PC functions for hearing data. ............................... 27
3.14 Data approximation using the first two PC functions for hearing data: Solid
curves are the observations and dashed curves are their approximations. ... 28
4.1 The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 1 \times e^{-6}$ (lip data). .... 34
4.2 The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 1 \times e^{-6}$ (lip data). .... 35
4.3 Solid curves are observations and dashed curves are their approximations with $\lambda = 1 \times e^{-6}$ (lip data).

4.4 The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 1 \times e^{-6}$ (simulated data).

4.5 The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 1 \times e^{-6}$ (simulated data).

4.6 Observations and their approximations with $\lambda = 1 \times e^{-6}$ for simulated data (Observations and approximations are overlapped).

4.7 The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 25$ (Hearing data).

4.8 The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 25$ (Hearing data).

4.9 Solid curves are observations and dashed curves are their approximations with $\lambda = 25$ (Hearing data).

6.1 Variance and Squared-Bias averaged over $t$ and $v$ for various values of smoothing parameters $\lambda$ and $\gamma$ ($\sigma = 0$).

6.2 Variance and Squared-Bias averaged over $t$ and $v$ for various values of smoothing parameters $\lambda$ and $\gamma$ ($\sigma = 0.15$).

6.3a Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0$, $v = 1.0$).

6.3b Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0$, $v = 1.0$).

6.4a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0$, $v = 1.0$).

6.4b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$, $v = 1.0$).

6.5a Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15$, $v = 1.0$).

6.5b Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15$, $v = 1.0$).
6.6a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15, v = 1.0$). .............................. 60
6.6b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, v = 1.0$). .............................................. 61
6.7a Simulated Pointwise Bias for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0$) ............................................ 62
6.7b Simulated Pointwise Bias for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15$). ............................................ 63
6.8a Simulated Pointwise Bias for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$). ............................................ 64
6.8b Simulated Pointwise Bias for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15$) ............................................. 65
6.9 Variance and Squared-Bias averaged over $t$ and $v$ for different values of $c$ ($\lambda=0.25, \sigma = 0$). ......................................................... 67
6.10 Variance and Squared-Bias averaged over $t$ and $v$ for different values of $c$ ($\lambda=250000, \sigma = 0.15$). ..................................................... 68

7.1 Age distribution of individuals with normal hearing .................... 69
7.2a Null space basis functions of normal hearing data for different covariate (age) values : Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=0.0025$). .................. 71
7.2b Null space basis functions of normal hearing data for different covariate (age) values : Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=0.25$). .................. 72
7.2c Null space basis functions of normal hearing data for different covariate (age) values : Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=25$). .................. 73
7.2d Null space basis functions of normal hearing data for different covariate (age) values : Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=2500$). .................. 74
7.3a Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\lambda=25$ and $\gamma=0.0025$). ......................................................... 75
7.3b Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\lambda=25$ and $\gamma=0.25$). 76
7.3c Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\lambda=25 \text{ and } \gamma=25) \). 77

7.3d Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\lambda=25 \text{ and } \gamma=2500) \). 78

7.4a Null space basis functions of normal hearing data for different covariate (age) values : Solid line for \( \hat{U}_1 \), Dashed line for \( \hat{U}_2 \) (\( \gamma=2.5 \text{ and } \lambda=0.0025 \)). 80

7.4b Null space basis functions of normal hearing data for different covariate (age) values : Solid line for \( \hat{U}_1 \), Dashed line for \( \hat{U}_2 \) (\( \gamma=2.5 \text{ and } \lambda=0.025 \)). 81

7.4c Null space basis functions of normal hearing data for different covariate (age) values : Solid line for \( \hat{U}_1 \), Dashed line for \( \hat{U}_2 \) (\( \gamma=2.5 \text{ and } \lambda=0.25 \)). 82

7.4d Null space basis functions of normal hearing data for different covariate (age) values : Solid line for \( \hat{U}_1 \), Dashed line for \( \hat{U}_2 \) (\( \gamma=2.5 \text{ and } \lambda=2.5 \)). 83

7.5a Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\gamma=2.5 \text{ and } \lambda=0.0025) \). 84

7.5b Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\gamma=2.5 \text{ and } \lambda=0.025) \). 85

7.5c Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\gamma=2.5 \text{ and } \lambda=0.25) \). 86

7.5d Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of \( L (\gamma=2.5 \text{ and } \lambda=2.5) \). 87

A.1a Simulated estimates of the basis functions for various values of \( \gamma \) with \( \lambda = 0.25 \) (\( \sigma = 0 \), \( v = 0.5 \)). 92

A.1b Simulated estimates of the basis functions for various values of \( \gamma \) with \( \lambda = 0.25 \) (\( \sigma = 0 \), \( v = 1.0 \)). 93
A.1c Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0, v = 1.5$) ................................................................. 94
A.2a Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, v = 0.5$) ................................................................. 95
A.2b Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, v = 1.0$) ................................................................. 96
A.2c Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, v = 1.5$) ................................................................. 97
A.3a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0, v = 0.5$) ................................................................. 98
A.3b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0, v = 1.0$) ................................................................. 99
A.3c Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0, v = 1.5$) ................................................................. 100
A.4a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0, v = 0.5$) ................................................................. 101
A.4b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0, v = 1.0$) ................................................................. 102
A.4c Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0, v = 1.5$) ................................................................. 103
A.5a Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15, v = 0.5$) ................................................................. 104
A.5b Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15, v = 1.0$) ................................................................. 105
A.5c Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15, v = 1.5$) ................................................................. 106
A.6a Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, v = 0.5$) ................................................................. 107
A.6b Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, v = 1.0$). ......................................................... 108

A.6c Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, v = 1.5$). ......................................................... 109

A.7a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15, v = 0.5$). ......................................................... 110

A.7b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15, v = 1.0$). ......................................................... 111

A.7c Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15, v = 1.5$). ......................................................... 112

A.8a Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, v = 0.5$). ......................................................... 113

A.8b Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, v = 1.0$). ......................................................... 114

A.8c Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, v = 1.5$). ......................................................... 115
Chapter 1

Introduction

In Principal Differential Analysis (PDA) a linear differential equation that captures features of a collection of curves is found in order to have a low dimensional approximation for functional observations. Functional observations are measured at distinct time intervals. Time intervals can be either equally spaced or unequally spaced, but here we consider only the equally spaced observations. Lip data obtained from Ramsay and Silverman (2005) [15] is an example for functional observations; Figure 1.1 contains 20 records of the movement of the center of the lower lip as a single speaker said the syllable “bob”. The data are measured at 51 sampling points and are shown in Figure 1.1 after preprocessing by smoothing.

In many practical situations, functional data collected from experimental units are described by covariates and hence the method of PDA was extended by Jin (2006) [6] to include covariates. It helps to have a low dimensional approximation for functional data that is described by covariates.

The second chapter of this thesis reviews smoothing splines and P-splines that are used throughout this thesis. The third chapter reviews Principal Component Analysis (PCA) for multivariate data and for curves. Chapter 4 reviews Principal Differential Analysis (PDA) and it is given as an alternative to the PCA. PDA with covariates is reviewed in Chapter 5 and Chapter 6 includes the results of the Monte-Carlo simulation study. Chapter 7 contains an analysis of data.
Figure 1.1: Twenty records of lip position
Chapter 2

Nonparametric Regression with P-Splines

2.1 Introduction to Nonparametric Regression

In parametric regression, the relationship between the \( i^{th} \) response and explanatory variables is given as

\[
x(t_j) = f(\beta, t_j) + \epsilon_j \quad j = 1, 2, \ldots, N,
\]

where \( f \) is a prespecified function, \( \beta \) is a vector of parameters that should be estimated and \( N \) is the number of observations obtained. The main objective here is estimating the parameters. It is assumed that the \( \epsilon_1, \ldots, \epsilon_N \) are independent with mean 0 and variance \( \sigma^2 \).

In nonparametric regression this relationship is written as

\[
x(t_j) = f(t_j) + \epsilon_j \quad j = 1, 2, \ldots, N, \tag{2.1}
\]

where \( f \) is a function that is not specified. The main objective here is estimating this function. There is no reduction of the data to a low dimensional vector of parameters like in parametric regression. Nonparametric regression assumes that the function \( f \) is smooth and continuous. There are many methods that are used in nonparametric regression. Smoothing splines, kernel estimates and local linear regression are some of them. Smoothing splines and P-splines are reviewed in this chapter.
2.2 Smoothing Splines

Suppose the response $X$ is measured at different time points as $x(t_1), x(t_2), \ldots, x(t_N)$, $0 < t_1 < t_2 < \ldots < t_N < T$. Using these points, the nonparametric regression model (2.1) is considered for a fixed, unknown, but smooth function $f$ that should be estimated.

**Definition 2.1**: The Sobolev space of order $m$ is defined as,

$$W_2^m [0, T] = \left\{ f : f^{(j)} \text{ where } j = 1, \ldots, m - 1 \text{ is absolutely continuous and } \int_0^T |f^{(m)}(t)|^2 dt < \infty \right\}.$$

Suppose the function $f \in W_2^2 [0, T]$ and we need an estimator for $f$ that fits the data well with some degree of smoothness. A measure of goodness of fit to the data is given by

$$\frac{1}{N} \sum_{j=1}^N (x(t_j) - g(t_j))^2$$

while the measure of smoothness of the function is

$$\int_0^T |g^{(2)}(t)|^2 dt.$$

Hence, the overall performance of the estimator is obtained by combining these two criteria as

$$(1 - q) \frac{1}{N} \sum_{j=1}^N (x(t_j) - g(t_j))^2 + q \int_0^T |g^{(2)}(t)|^2 dt,$$

for some $0 < q < 1$; the estimator is obtained by minimizing over $g \in W_2^2 [0, T]$ (Eubank (1988))[4]. When $\lambda$ is set to $\frac{q}{1-q}$, this is equivalent to estimating $f$ with the minimizer of

$$\frac{1}{N} \sum_{j=1}^N (x(t_j) - g(t_j))^2 + \lambda \int_0^T |g^{(2)}(t)|^2 dt > 0, \quad (2.2)$$

over $g \in W_2^2 [0, T]$. The solution is a natural spline that will be defined later in this chapter.

Large values of $\lambda$ give smooth curves that have a poor fidelity to data while small values of $\lambda$ give more wiggly curves that can lead to overfitting the data. In extremes, $\lambda = 0$ interpolates the data, while very large $\lambda$ gives the linear least squares fit to the data. Because of these properties of $\lambda$, it is called the smoothing parameter.
2.2.1 Spline of order p

Splines are piecewise polynomials. Piecewise polynomials are very useful because they are more flexible than polynomials. Because of this flexibility they can adjust more effectively to the data.

**Definition 2.2** (Eubank(1988)[4])

Let \( t_1 < t_2 < \ldots < t_N \) be a set of ordered points called knots, in some interval \([0, T]\). Then the spline of order \( p \) is any function, \( S(t) \) of the form

\[
S(t) = \sum_{i=0}^{p-1} \theta_i t^i + \sum_{i=1}^{N} \delta_i (t - t_i)^{p-1}, \quad t \in [0, T],
\]

(2.3)

where

\[
(t - t_i)_+ = \begin{cases} 
(t - t_i) & \text{for } t > t_i \\
0 & \text{otherwise}
\end{cases}
\]

and for some real valued coefficients \( \theta_1, \ldots, \theta_{p-1} \) and \( \delta_1, \ldots, \delta_N \). This has the following properties:

1. It is a piecewise polynomial of order \( p \) on \([t_i, t_{i+1}]\);

2. It has \((p - 2)\) continuous derivatives;

3. It has \((p - 1)\)st derivative with jumps at design points(knots) \( t_1, \ldots, t_N \).

A piecewise polynomial of order \( p \) with \( p - 1 \) continuous derivatives is a single polynomial of order \( p \). This implies that spline is the smoothest possible piecewise polynomial which retains a segmented nature.

**Definition 2.3** (Eubank(1988)[4])

A spline of order \( p = 2m \) with knots \( t_1, \ldots, t_N \) is called a natural spline if it is,
4. A polynomial of order \( m \) outside the interval \([t_1, t_N]\).

Let \( S^p(t_1, \ldots, t_N) \) denote the collection of all splines of order \( p \) with knots at \((t_1, \ldots, t_N)\). This definition of spline uses the truncated power basis and for the set of splines \( S^p(t_1, \ldots, t_N) \), the truncated power basis is given by 1, \( t, t^2, \ldots, t^{p-1} \) and \((t - t_1)^{p-1}, (t - t_2)^{p-1}, \ldots, (t - t_N)^{p-1}\).

Let \( NS^p(t_1, \ldots, t_N) \) with \( p = 2m \) denote the subspace of the vector space \( S^p(t_1, \ldots, t_N) \) obtained by placing additional linear restrictions on the derivatives of the spline that are needed to satisfy the fourth condition above. In particular for \( S \) to be a natural spline it is necessarily to have \( \theta_m = \ldots = \theta_{2m-1} = 0 \) in (2.3) since for \( t < t_1 \), \( S \) is a polynomial of order \( m \). \( NS^p(t_1, \ldots, t_N) \) is a vector space of dimension \( N \). There are nice properties of natural splines that can be used to show that the minimizer of (2.2) is a natural spline.

**Lemma 2.1** (Lyche and Schumaker(1973)[10])

Let \( \psi_1(t), \ldots, \psi_N(t) \) be a basis for \( NS^{2m}(t_1, \ldots, t_N) \). Then there exist coefficients \( \theta_{ij}, \ldots, \theta_{m-1j}, \delta_{1j}, \ldots, \delta_{Nj} \) such that,

\[
\psi_j(t) = \sum_{i=0}^{m-1} \theta_{ij} t^i + \sum_{i=1}^{N} \delta_{ij} (t - t_i)^{2m-1}.
\]

Then for any \( f \in W_2^{2m}[0, T] \) and \( S(t) = \sum_{j=1}^{N} \beta_j \psi_j(t) \),

\[
\int_0^T f^{(m)}(t)S^{(m)}(t)\,dt = (-1)^m (2m - 1)! \sum_{i=1}^{N} f(t_i) \sum_{j=1}^{N} \beta_j \delta_{ij}.
\]

**Theorem 2.1** (Lyche and Schumaker(1973)[10] as described by Eubank (1988)[4])

Let \( \psi_1(t), \ldots, \psi_N(t) \) be a basis for \( NS^{2m}(t_1, \ldots, t_N) \) and suppose that \( N \geq m \). For fixed \( 0 < \lambda < \infty \) there is a unique minimizer \( f_{\lambda} \in W_2^m [0, T] \) of (2.2). Also \( f_{\lambda} \in NS^{2m}(t_1, \ldots, t_N) \) and hence has the form \( \sum_{j=1}^{N} \beta_{\lambda j} \psi_j \). The coefficients \( \beta_{\lambda} = (\beta_{\lambda 1}, \ldots, \beta_{\lambda N})'s \) are the solution to

\[
(\Psi + N\lambda G)\beta_{\lambda} = X,
\]
where
\[ \Psi = \{ \psi_j(t_i) \}_{i,j=1,\ldots,N}, \]
\[ G = \{-1^m(2m-1)!\delta_{ij}\}_{i,j=1,\ldots,N}, \]
and \( \delta_{ij} \) are the coefficients that are defined in Lemma 2.1 for the basis \( \psi_1(t), \ldots, \psi_N(t) \).

### 2.3 B-Splines

The previous explanation used truncated power basis, but they can be ill-conditioned and make the computations difficult. Because of that reason, instead of using the truncated power basis for splines, the B-spline basis is used.

Let \( t_1, t_2, \ldots, t_N \) denote given knots in the interval \([0, T]\). When defining the B-spline basis for \( S^p(t_1, \ldots, t_N) \), \( 2p \) additional knots \( t_{-(p-1)}, \ldots, t_{-1}, t_0, t_{N+1}, \ldots, t_{N+p} \) are defined as
\[ t_{-(p-1)} = \ldots = t_0 = 0 \quad \text{and} \quad t_{N+1} = \ldots = t_{N+p} = T. \]

Then the B-spline of order \( p \) with knots \( t_1, \ldots, t_N \) is defined recursively:
\[ B_{i,p}(t) = \frac{t - t_i}{t_{i+p-1} - t_i} B_{i,p-1}(t) + \frac{t_{i+p} - t}{t_{i+p} - t_{i+1}} B_{i+1,p-1}(t) \]
for \( i = -(p-1), \ldots, N \). The recursion begins with,
\[ B_{i,1}(t) = \begin{cases} 
1 & t \in [t_i, t_{i+1}] \\
0 & \text{otherwise.} \end{cases} \]

B-splines have several important properties: the local support characteristic \( B_{i,p}(t) = 0 \) if \( t \notin [t_i, t_{i+p}] \) and the overlapping support from adjacent B-splines. Figure 2.1 shows linear, quadratic and cubic B-splines for \( t \in [0, 1] \) with interior knots placed at 0.2, 0.4, 0.6 and 0.8.
2.4 P-Splines

P-splines are the penalized splines. Eilers and Marx (1996)[3] proposed a new penalty for the smoothness because it is difficult to work with the penalty in (2.2) as it has integrals. The penalty proposed by them $\lambda \sum_{j=k+1}^{r}(\Delta^k \beta_j)^2$ uses differences. Here $r$ is the number of B-splines that are used to approximate the data. For the cubic splines this penalty is usually $\lambda \sum_{j}^{r}(\Delta^2 \beta_j)^2$, where

$$\Delta \beta_j = \beta_j - \beta_{j-1}$$

$$\Delta^2 \beta_j = \Delta \beta_j - \Delta \beta_{j-1}$$

$$= \beta_j - 2\beta_{j-1} + \beta_{j-2}.$$
This can be written as a quadratic form in $\beta$. The matrix that is needed for the quadratic form is

$$
D = \begin{bmatrix}
1 & -2 & 1 & 0 & 0 & \ldots & 0 \\
0 & 1 & -2 & 1 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 1 & -2 & 1 \\
\end{bmatrix}^{(r-2) \times r}.
$$

Then

$$(D\beta)^T(D\beta) = \sum_{j=3}^{r}(\Delta^2 \beta_j)^2,$$

where $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}_{r \times 1}$.

Now the minimizing problem can be written as

$$\min_{\beta}(X-B\beta)^T(X-B\beta) + \lambda[\beta^T(D^TD)\beta],$$

where

$$B = \begin{bmatrix}
B_1(t_1) & \ldots & B_r(t_1) \\
\vdots & \ddots & \vdots \\
B_1(t_N) & \ldots & B_r(t_N) \\
\end{bmatrix}_{N \times r}.$$

To find the minimizer this is differentiated with respect to $\beta$.

$$-2B^T(X-B\beta) + 2\lambda(D^TD)\beta = 0$$

$$(B^TB + \lambda(D^TD)\beta = B^TX.$$

Therefore, the estimated coefficients of the P-spline are given by

$$\hat{\beta} = (B^TB + \lambda D^TD)^{-1}B^TX.$$

There are several advantages of P-splines. Relative ease of computation and usefulness of fitting polynomial data exactly are two of them.
Chapter 3

Principal Component Analysis

In this chapter an introduction to Principal Component Analysis (PCA) is given with some examples. Section 1 of this chapter explains PCA for multivariate data and section 2 deals with curve data.

3.1 Principal Component Analysis for Multivariate Data

Principal Component Analysis (PCA) is a highly useful data analysis technique. This concept was first introduced by Pearson (1901)[11] and developed by Hotelling (1933)[5]. PCA is a well known technique for reduction of vector data to a minimal dimension representation. It is used to find a new set of variables, fewer in number than in the original set of variables that account for most of the variance in the original variables. This is done by a linear transformation of the original variables to a set of uncorrelated variables called principal components. PCA is also used as a method of extracting more information from the data which cannot be easily seen in the original variables (Johnson and Wichern 2007)[8].

Principal Components (PC) $Y_1, Y_2, \ldots, Y_p$ are a linear combination of the $p$ original random variables $X_1, X_2, \ldots, X_p$; the first PC gives the linear combination that has maximum variance. This means that the first PC represents the direction that spreads out the observations more than any other direction. The second PC gives the linear combination with second highest variance in the direction orthogonal to the first PC and so on.
3.1.1 Theory

Let the random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]'$ have the covariance matrix $\Sigma_{p \times p}$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. Consider linear combinations,

$$
Y_1 = \mathbf{a}_1' \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\
Y_2 = \mathbf{a}_2' \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\
\vdots \\
Y_p = \mathbf{a}_p' \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p,
$$

then

$$
\text{Var}(Y_i) = \mathbf{a}_i' \Sigma \mathbf{a}_i \quad i = 1, 2, \ldots, p \\
\text{Cov}(Y_i, Y_k) = \mathbf{a}_i' \Sigma \mathbf{a}_k \quad i, k = 1, 2, \ldots, p,
$$

where $\mathbf{a}_i' \mathbf{a}_i = 1$.

Since PCs are uncorrelated linear combinations, they are defined as follows. The first PC is the linear combination $\mathbf{a}_1' \mathbf{X}$ that maximizes $\text{Var}(\mathbf{a}_1' \mathbf{X})$ subject to $\mathbf{a}_1' \mathbf{a}_1 = 1$ and the second PC is the linear combination $\mathbf{a}_2' \mathbf{X}$ that maximizes $\text{Var}(\mathbf{a}_2' \mathbf{X})$ subject to $\mathbf{a}_2' \mathbf{a}_2 = 1$ and $\text{Cov}(\mathbf{a}_1' \mathbf{X}, \mathbf{a}_2' \mathbf{X}) = 0$. Generally the $i$th PC is the linear combination $\mathbf{a}_i' \mathbf{X}$ that maximizes $\text{Var}(\mathbf{a}_i' \mathbf{X})$ subject to $\mathbf{a}_i' \mathbf{a}_i = 1$ and $\text{Cov}(\mathbf{a}_i' \mathbf{X}, \mathbf{a}_k' \mathbf{X}) = 0$ for $k < i$.

**Theorem 3.1** (Johnson and Wichern 2007)([8])

Let $\mathbf{B}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then,

$$
\max_{\mathbf{a} \neq \mathbf{0}} \frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{a}} = \lambda_1 \quad \text{(attained when } \mathbf{a} = \mathbf{e}_1).$$

Also,

$$
\max_{\mathbf{a} \perp \mathbf{e}_1, \ldots, \mathbf{e}_k} \frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{a}} = \lambda_{k+1} \quad \text{(attained when } \mathbf{a} = \mathbf{e}_{k+1}, \ k = 1, 2, \ldots, p - 1).$$
As given in (Johnson and Wichern 2007)(see [8]), let $\Sigma$ be the covariance matrix that is positive definite and has the eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_p, e_p)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. When the above theorem is applied here,

\[
\max_{x \neq 0} \frac{x'\Sigma x}{x'x} = \lambda_1 \text{ (attained when } x = e_1) \]

$e'_1 e_1 = 1$, since the eigenvectors are normalized. Therefore,

\[
\max_{x \neq 0} \frac{x'\Sigma x}{x'x} = \frac{e'_1 \Sigma e_1}{e'_1 e_1} = e'_1 \Sigma e_1 = \text{Var}(Y_1).
\]

From the second part of the theorem 3.1,

\[
\max_{x \perp e_1, \ldots, e_k} \frac{x'\Sigma x}{x'x} = \lambda_{k+1} \text{ (attained when } x = e_{k+1}, k = 1, 2, \ldots, p - 1).
\]

For $x = e_{k+1}$ with $e'_{k+1} e_i = 0$ for $i = 1, 2, \ldots, k$ and $k = 1, 2, \ldots, p - 1$,

\[
\frac{e'_{k+1} \Sigma e_{k+1}}{e'_{k+1} e_{k+1}} = e'_{k+1} \Sigma e_{k+1} = \text{Var}(Y_{k+1}).
\]

But, $e'_{k+1} (\Sigma e_{k+1}) = \lambda_{k+1} e'_{k+1} e_{k+1} = \lambda_{k+1}$. Hence, $\text{Var}(Y_{k+1}) = \lambda_{k+1}$.

Also for any two eigenvectors $e_i$ and $e_k$, they are orthogonal if all eigenvalues are distinct. If they are not distinct, eigenvectors corresponding to common eigenvalues can be chosen to be orthogonal. Hence, $e'_i e_k = 0$ for $i \neq k$. Therefore, for any $i \neq k$,

\[
\text{Cov}(Y_i, Y_k) = e'_i \Sigma e_k = e'_i \lambda_k e_k = \lambda_k e'_i e_k = 0.
\]

According to the results above it can be seen the $i^{th}$ PC is given by

\[
Y_i = e'_i X = e_{i1} X_1 + e_{i2} X_2 + \ldots + e_{ip} X_p,
\]

with,

\[
\text{Var}(Y_i) = e'_i \Sigma e_i = \lambda_i \quad i = 1, 2, \ldots, p
\]

\[
\text{Cov}(Y_i, Y_k) = e'_i \Sigma e_k = 0 \quad i \neq k.
\]
Definition 3.1 (Johnson and Wichern 2007) (see [8])

Let $A$ be a $k \times k$ positive definite matrix with the spectral decomposition $A = \sum_{i=1}^{k} \lambda_i e_i e_i'$. Let the normalized eigenvectors be the columns of another matrix $P = [e_1, e_2, \ldots, e_k]$ and $\Lambda$ be the diagonal matrix of eigenvalues. Then,

$$A_{(k \times k)} = \sum_{i=1}^{k} \lambda_i e_i e_i' e_i'_{(k \times 1)} = P_{(k \times k)} \Lambda_{(k \times k)} P'_{(k \times k)}.$$ 

Letting $A = \Sigma$, it can be written that $\Sigma = P \Lambda P'$ where $P = [e_1, e_2, \ldots, e_k]$ so that $PP' = P'P = I$. Since for the symmetric matrices $A$ and $B$, $\text{tr}(AB) = \text{tr}(BA)$,

$$\text{tr}(\Sigma) = \text{tr}(P \Lambda P') = \text{tr}(\Lambda P' P) = \text{tr}(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_p$$

Therefore,

$$\text{tr}(\Sigma) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \text{Var}(X_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \text{Var}(Y_i).$$

Then the proportion of the total variation explained by the $k^{th}$ PC is given by

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \quad k = 1, 2, \ldots, p.$$

The number of PCs to retain is decided by considering the amount of total sample variance explained, the relative sizes of the eigenvalues (the variance of the sample components), and also depending on the problem. A useful visual aid for determining an appropriate number of PCs is a scree plot. In the scree plot the magnitude of the ordered eigenvalues are plotted against the eigenvalue number ($\lambda_i$ vs $i$). The changes in the shape of the scree plot are considered when deciding the number of PCs to retain.

After the principal components are obtained each observation can be expressed as a linear combination of the principal component directions as follows:
\[ \mathbf{X}_1 = (\mathbf{X}'_1 \mathbf{e}_1) \mathbf{e}_1 + (\mathbf{X}'_1 \mathbf{e}_2) \mathbf{e}_2 + \ldots + (\mathbf{X}'_1 \mathbf{e}_p) \mathbf{e}_p \]
\[ \mathbf{X}_2 = (\mathbf{X}'_2 \mathbf{e}_1) \mathbf{e}_1 + (\mathbf{X}'_2 \mathbf{e}_2) \mathbf{e}_2 + \ldots + (\mathbf{X}'_2 \mathbf{e}_p) \mathbf{e}_p \]
\[ \vdots \]
\[ \mathbf{X}_n = (\mathbf{X}'_n \mathbf{e}_1) \mathbf{e}_1 + (\mathbf{X}'_n \mathbf{e}_2) \mathbf{e}_2 + \ldots + (\mathbf{X}'_n \mathbf{e}_p) \mathbf{e}_p. \]

Since \( Y_i = \mathbf{e}'_i \mathbf{X} \), the \( j^{th} \) observation can be written as
\[ \mathbf{X}_j = Y_{j1} \mathbf{e}_1 + Y_{j2} \mathbf{e}_2 + \ldots + Y_{jp} \mathbf{e}_p. \]

We can write this using matrix notation as
\[ \mathbf{X} = \sum_{i=1}^{p} Y_i \mathbf{e}_i. \]

As in Castro et al. (1986)[1], for the centralized data, the \( k \)-dimensional linear model \((k < p)\) can be written as
\[ \mathbf{Z} = \mathbf{M} + \sum_{i=1}^{k} Y'_i \mathbf{a}_i, \]  
where the \( \mathbf{M} \) is the vector of means, the \( \mathbf{a}_i \)'s are normalized vectors and the \( Y'_i \)'s are the centered principal components.

The mean squared error
\[ S_k^2 = \min_{\mathbf{a}_i} \mathbb{E}(\mathbf{X} - \mathbf{Z})^2, \]
is minimum when \( \mathbf{a}_i = \mathbf{e}_i \), the normalized eigenvectors corresponding to \( i^{th} \) largest eigenvalue \( \lambda_i \) of the process variance-covariance matrix \( \mathbf{\Sigma} \), where
\[ (\mathbf{\Sigma})_{ij} = \sigma_{ij} = \text{Cov}(X_i, X_j) \quad i, j = 1, 2, \ldots , p. \]

The minimum \( S_k^2 \) is given by
\[ S_k^2 = S^2 - \sum_{i=1}^{k} \lambda_i, \]
where $S^2$ is the total process variance which can be computed as the trace of $\Sigma$.

The centered principal components $Y'_1,\ldots,Y'_k$ in (3.1) are determined by

$$Y'_i = \sum_{j=1}^{n} (X_j - M_j) e_{ji},$$

where $e_{ji}$ is the $j^{th}$ coordinate of the $i^{th}$ eigenvector of $\Sigma$.

Throughout this section, the theory of principal component analysis for multivariate data was discussed. When we have a data set, first the sample variance-covariance matrix should be calculated using

$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T,$$

where $\bar{X}$ is the sample mean vector. The estimates for the eigenvectors, $\hat{e}_1,\ldots,\hat{e}_p$ are obtained using this sample variance-covariance matrix. The centered principal component scores are determined using (3.2) replacing $e$ by $\hat{e}$ and $M_j$ by the $j^{th}$ element of sample mean vector. After the number of principal components to be retained is decided using the scree plot, we can approximate the data using (3.1). The following example illustrates this technique more clearly.

Example: Data on Ozone levels

Data of the minimum of daily one-hour-average ozone levels for 7 days of 47 weeks is analyzed using PCA as in Jones and Rice(1992)[9]. Figure 3.1 gives the ozone measurements for 47 weeks. It is not easy to find patterns in this plot. So PCA is used to make it interpretable.
Figure 3.1: Daily ozone values for 47 weeks.

Figure 3.2 gives the scree plot corresponding to this data. From the scree plot it can be decided that the first two PCs would be enough to explain the variation in the data. Also using the eigenvalues it can be said that the first two principal components account for 61.9% and for 18.3% of the total variation, respectively. So those two together explain more than 90% of the total variation. The first two PCs are shown in Figure 3.3. The first principal component(dashed line) alone can be used to differentiate many weeks as it gives the average ozone level of the week. The second principal component accounts for the change in the ozone level throughout the week. Figure 3.4 shows two observations approximated by using only two principal components. The solid line shows the original data and the dashed line shows its approximation. It can be seen that a linear combination using only the first two PCs has approximated the data reasonably well. Clear seasonal variation can be observed when the first PC-Score is drawn versus the week number. This is shown in Figure 3.5.
Figure 3.2: Scree Plot for the ozone data.

Figure 3.3: First two PCs for ozone data: Dashed line is the first PC, Solid line is the second PC.
Figure 3.4: Data approximation using the first two PCs for ozone data.

Figure 3.5: First PC Score for ozone data.
3.2 Best linear model for continuous processes

We can apply the same method discussed in the previous section for a random process indexed by a continuous parameter $t$. This method is explained for equally spaced observation points $t_i$, but it can be adapted to unequally spaced sample points. This is explained by (Castro et al.(1986)[1]).

Here we consider a continuous process $y(t), 0 \leq t \leq T$ with covariance function,

$$C(s, t) = \text{Cov}[y(s), y(t)] \quad 0 \leq s, t \leq T,$$

and the mean function

$$m(t) = E[y(t)] \quad 0 \leq t \leq T.$$

$C(s, t)$ is symmetric and positive definite. Therefore, the eigenvalues $\lambda_1, \ldots, \lambda_n$ and eigenfunctions $\varphi_1, \ldots, \varphi_n \ (k < n)$ satisfy

$$\int_0^T C(s, t) \varphi_i(t) dt = \lambda_i \varphi_i(s) \quad 0 \leq s, t \leq T$$

and

$$\int_0^T \varphi_i(t) \varphi_j(t) dt = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

As in (Castro et al.(1986)[1]), a $k$-dimensional linear model for this process is written as

$$z(t) = m(t) + \sum_{i=1}^k \alpha_i f_i(t), \quad (3.3)$$

where $f_1, f_2, \ldots, f_k$ are $k$ linearly independent functions and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are $k$ scalar variates. The residual squared error of the linear model (3.3) is given by

$$\Psi_k^2 = E \left[ \int_0^T [y(t) - z(t)]^2 dt \right].$$

The best $k$-dimensional linear model for $y(t)$ that minimizes $\Psi_k^2$ over all $k$ linearly independent functions $f_1, f_2, \ldots, f_k$ and all coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$ is given by $f_i = \varphi_i$, the
normalized eigenfunctions corresponding to the $i^{th}$ largest eigenvalue of the covariance function $C(s,t)$. The coefficients $\alpha_i$ for the best $k$-dimensional linear model are called principal component scores that are given by

$$\alpha_i = \int_0^T [y(t) - m(t)] \varphi_i(t) dt.$$ 

As in Castro et al. (1986) [1], the best linear model can be written using the $\varphi_i$'s as

$$z(t) = m(t) + \sum_{i=1}^k \alpha_i \varphi_i(t).$$

The total process variance can be written as

$$\sigma^2 = E \left[ \int_0^T |y(t) - m(t)|^2 dt \right].$$

The minimum mean squared error over all $k$-dimensional linear models of $y(t)$ is given by

$$\sigma_k^2 = E \left[ \int_0^T |y(t) - z(t)|^2 dt \right].$$

This can be written using eigenvalues as

$$\sigma_k^2 = \sigma^2 - \sum_{i=1}^k \lambda_i,$$

where $\sum_{i=1}^k \lambda_i$ is the amount of variation explained by the $k$-dimensional linear model.

Now we will discuss how the approximations are made using data. It is assumed that $y(t)$ is continuous in $t$ and $C(s,t)$ is continuous in $(s,t)$ pairs. According to Castro et al. (1986) [1] the integral $\int_0^T C(s,t) f(t) dt$ is replaced by finite sum $\sum_{j=1}^p C(s, t_j) f(t_j) w_j$ using a quadrature rule. Then the eigenvalues and eigenvectors are estimated using

$$\sum_{j=1}^p C(t_i, t_j)f(t_j) w_j = \lambda f(t_i), \quad (3.4)$$

where the covariance kernel $C(s,t)$ is estimated from the sample of curves $y(t)$. The $w_j$ in (3.4) is the weight that is used when we have unequal grid of $t$'s, but here we consider only
the equal grid of \( t_i \) and hence \( w_j = (t_i - t_{i-1}) \) for \( j = 1, \ldots, p \). Once the eigenfunctions \( \varphi_i \) are estimated we can approximate the curves \( y(t_i) \) by substituting

\[
\hat{y}(t) = m(t) + \sum_{v=1}^{p} \hat{\alpha}_v \hat{\varphi}_v(t),
\]

where \( \hat{\alpha}_v \) are the principal component scores and the \( \hat{\varphi}_v \) are the principal component functions estimates using B-splines introduced in Chapter 2. First, the eigenfunction \( \varphi_1 \) is written using a linear combination of B-spline functions \( B_j \) as \( \varphi_1 = \sum_{j=1}^{l} \beta_j B_j(t) \), where \( l \) is the number of B-spline functions and \( \beta_j \) are the coefficients of the basis expansion. Then \( \varphi_1 \) is estimated by maximizing the expression for the variance of the first PC score subject to the constraint \( \int_0^T [\varphi_1(t)]^2 dt + \lambda \int_0^T [\varphi_1^{(2)}(t)]^2 dt = 1 \), where \( \lambda \) is the smoothing parameter. For \( j \geq 2 \) the estimate of the \( j^{th} \) eigenfunction is defined to maximize the variance of the PC scores subject to \( \int_0^T [\varphi_j(t)]^2 dt + \lambda \int_0^T [\varphi_j^{(2)}(t)]^2 dt = 1 \) and the additional constraint \( \int_0^T \varphi_i(t) \varphi_j(t) dt + \lambda \int_0^T \varphi_i^{(2)}(t) \varphi_j^{(2)}(t) dt = 0 \) for all \( i < j \) (Ramsay and Silverman(2002))[14].

The following examples illustrate this method. The data were analyzed using “R” software. Ramsay’s programs can be found in “ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns/R/” were used for analysis. In his programs Ramsay has implemented regularized PCA that applies smoothing into the functional PCA. So it has two smoothing parameters one for smoothing the curve data in a preprocessing step and one for smoothing the eigenfunctions in functional PCA.

Examples :

1) First, lets take the lip data introduced in Chapter 1. Figure 3.6 shows the data after preprocessing by smoothing with smoothing parameter set to \( 1 \times e^{-12} \). The results of the functional principal component analysis with smoothing parameter \( \lambda = 1 \times e^{-6} \) are shown in Figure 3.7. As can be seen, both PC functions have nonzero values at the beginning of the time interval. They become zero while achieving the peak at the middle. The peaks at the boundary and the middle have different signs. The first two principal components
together explain close to 85% of the total variation in the data. Approximation of the data using the first two PC functions is given in Figure 3.8. It can be seen that we have obtained a good approximation to the data using the first two PC functions.

Figure 3.6: Twenty records of lip position after smoothing with smoothing parameter value $1 \times e^{-12}$.
Figure 3.7: First two PC functions for lip data.

Figure 3.8: Data approximation using the first two PC functions for lip data: Solid curves are the observations and dashed curves are their approximations.
2) Now let’s use some simulated data for illustration. Data were generated using

\[ x_i(t) = e^{-2t}(a_{1,i} \cos(6t) + a_{2,i} \sin(6t)), \]

where

\[ \begin{pmatrix} a_{1,i} \\ a_{2,i} \end{pmatrix} \sim N_2 \left( \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \right) \]

for \( i = 1, \ldots, 50 \) and \( t \in [0, 2] \) with \( t = \{t_j\}_{j=1}^{200} \), where \( t_j = \frac{j-1}{100} \) for \( j = 1, \ldots, 200 \). The generated data were smoothed in a preprocessing step using the smoothing parameter value \( 1 \times e^{-3} \); the results of the smoothing are shown in Figure 3.9. Functional PCA was carried out with \( \lambda = 1 \times e^{-6} \). The first two PC functions corresponding to the smoothed curves are shown in Figure 3.10. Both PC functions have some variation at the beginning and converge to zero after half of the time interval. We can see that the first two PC functions explain the total variation of the data. When the data curves were drawn with their approximation (Figure 3.11), it can be seen that we have obtained a perfect approximation for the data curves using the first two PC functions.

Figure 3.9: Simulated data after smoothing with smoothing parameter value \( 1 \times e^{-3} \).
Figure 3.10: First two PC functions for simulated data.

Figure 3.11: Data approximation using the first two PC functions for simulated data (Observations and approximations are overlapped).
3) Finally, the hearing data described in Wood(2007)[16] is analyzed. The hearing data contains the data of the evoked potential of the individuals with normal hearing and cochlear implanted individuals when they hear the syllable “ba”. The evoked potential was measured -100ms before they hear the syllable until 600ms after they hear the syllable. There are data of 92 individuals with normal hearing and 45 cochlear implant individuals. Only the individuals whose age is less than 7 years with normal hearing is analyzed here. Figure 3.12 shows the evoked potential of those 74 individuals.

The data were preprocessed by fitting cubic regression splines of dimension 13. Functional principal component analysis was carried out with $\lambda = 25$; the resulting first two principal component functions are shown in Figure 3.13. The first principal component function has the shape of the data curves. The second principal component function adjusts the location of the peak. The first two PC functions together account for only 57% of total variation of the data. Figure 3.14 shows that the approximation of the data is not good especially for the second curve indicating that only the first two principal components are not sufficient to explain the variation in the data well.
Figure 3.12: Evoked potential of the individuals below 7 years of age with normal hearing.

Figure 3.13: First two PC functions for hearing data.
Figure 3.14: Data approximation using the first two PC functions for hearing data: Solid curves are the observations and dashed curves are their approximations.
Chapter 4

Principal Differential Analysis

In the previous chapter, Principal Component Analysis that is a highly useful data analysis technique was introduced. In this chapter another important data analysis technique is discussed with examples. That is Principal Differential Analysis (PDA).

4.1 Linear Differential Operator

Differential equations are capable of capturing the features of a single curve or variation of the features across a collection of curves. Because of that, differential equations are highly useful when data are given as curves. As Ramsay and Silverman(2005)[15] mentioned, the objective of PDA is to find a linear differential equation that captures features of a collection of curves(observations). When $n$ functional observations $x_1, \ldots, x_n$ are presented, the goal is identifying the coefficients $w_0(t), \ldots, w_{m-1}(t)$ of the linear differential operator

$$L = w_0 I + w_1 D + \ldots + w_{m-1} D^{m-1} + w_m D^m, \quad (4.1)$$

which annihilates the observations as best as possible. Coddington and Levinson (1955)[2] assures the existence of the annihilating Linear Differential Operator(LDO) of order $m$ for a collection of $m$ linearly independent curves. Theorem 4.1 justifies the use of B-splines for estimation of the coefficients $w_0, \ldots, w_m$.

**Theorem 4.1** (Jin et al.(2011)[7])

Let $f_1, \ldots, f_m \in W^2_2 [0, T]$ be $m$ linearly independent functions, and consider $\mathcal{M} = \text{span} \{f_1, \ldots, f_m\}$. Then $\mathcal{M}$ has an $m^{th}$ order annihilating LDO of the form given by
\[ L = w_0 I + w_1 D + \ldots + w_{m-1} D^{m-1} + w_m D^m \] with coefficients in \( W^m_2 [0,T] \).

Ramsay(1996)[13] obtained the LDO \( L = w_0 I + w_1 D + \ldots + w_{m-1} D^{m-1} + D^m \) by dividing all the coefficient by \( w_m \). But it does not guarantee that the new coefficients are in \( W^m_2 \) as the Sobolev space is not closed under division. However, we proceed with the estimation of the coefficients \( w_0(t), \ldots, w_{m-1}(t) \) using B-spline basis approximation as in Ramsay and Silverman(2005)(see [15]).

The operator \( L \) should annihilate the sample curves. Therefore, the function \( Lx_i \) is considered as the residual from the application of \( L \). Ramsay and Silverman(2005) use the sum of squared norms of the residuals,

\[
\sum_{i=1}^{n} \int_0^T (Lx_i(t))^2 dt,
\]

as the fitting criterion. To estimate the linear differential operator that annihilates the curve data, (4.2) is minimized over \( L \). Since \( L \) consists of \( m \) coefficient functions, this amounts to estimating the \( w_j, j = 0, 1, \ldots, m - 1 \) in the Sobolev space.

The null space of the linear differential operator is defined as \( \mathcal{N}(L) = \{ f \in W^m_2 [0,T] : Lf = 0 \} \). There are \( m \) linearly independent functions \( u_1, \ldots, u_m \) that can be used to span any function that satisfies \( Lx = 0 \). Hence \( u_1, \ldots, u_m \) is called the basis functions of the null space of the LDO of order \( m \). A low dimensional approximation of the observations \( x_i \)'s can be obtained by expanding them by a linear combination of \( u_1, \ldots, u_m \). This reminds us of the principal component functions that were used in the previous chapter, which also provided a method for approximation of the observations. As showed in the previous chapter, in functional PCA each observation can be written using basis functions called eigenfunctions. The first \( m \) eigenfunctions that were used to project the observations play the same role in approximation as the basis functions \( u_1, \ldots, u_m \) for the null space of \( L \) used in PDA.
Ramsay and Silverman (2005) [15] added a penalty term to the sum of squared norms to obtain a regularized version of (4.2) as

\[ \sum_{i=1}^{n} \int_{0}^{T} (Lx_i(t))^2 dt + \lambda \sum_{j=1}^{m} \int_{0}^{T} (w_j(t))^2 dt, \lambda > 0. \]

We assume that \( w_j \in W_2^2[0,T] \) (\( j = 0, \ldots, m - 1 \)). Then we can approximate \( w_j \)'s using \( p^{th}(p \geq 3) \) order B-spline basis functions. We take \( w_j \approx \sum_{i=1}^{n} \beta_{ji} B_i \), where \( \beta_{ji} \) coefficients should be approximated from data. For \( p^{th} \) order basis functions \( B_1, \ldots, B_n \), let \( B = (B_1, \ldots, B_n)^T \) and \( \beta = (\beta_0^T, \ldots, \beta_{m-1}^T)^T \), where \( \beta_j = (\beta_{j1}, \ldots, \beta_{jn})^T \). Jin (2006) [6] used the smoothness penalty \( \sum_{j=0}^{m-1} \lambda_j \| D^2 \beta_j \|^2 \) introduced by Eilers and Marx (1996) [3] and proceeded as follows. The new fitting criterion was written as

\[ f(\beta) = \sum_{i=1}^{n} \int_{0}^{T} (Lx_i(t))^2 dt + \sum_{j=0}^{m-1} \lambda_j \| D^2 \beta_j \|^2. \]  

(4.3)

After substituting \( w_j \approx \sum_{i=1}^{n} \beta_{ji} B_i \) into (4.1), \( Lx_i(t) \) was written using matrix notation:

\[ Lx_i(t) = \left( \begin{array}{c} Bx_i \\ BDx_i \\ \vdots \\ ND^{m-1}x_i \end{array} \right) + D^m x_i. \]

Substituting this into (4.3),

\[ f(\beta) = \sum_{i=1}^{n} \int_{0}^{T} (L_{m-1}x_i(t))(L_{m-1}x_i(t))^T dt + 2 \sum_{i=1}^{n} \int_{0}^{T} (L_{m-1}x_i(t))D^m x_i(t) dt \]

\[ + \sum_{i=1}^{n} \int_{0}^{T} (D^m x_i(t))^2 dt + \sum_{j=0}^{m-1} \lambda_j (D^2 \beta_j)^T (D^2 \beta_j) \]

\[ = \beta^T D \beta + 2\beta^T d + \sum_{i=1}^{n} \int_{0}^{T} (D^m x_i(t))^2 dt + \beta^T H^T H \beta. \]

(4.4)
\( \mathbf{D} \) in (4.4) is a symmetric matrix consisting of \( m \times m \) array of \( n \times n \) submatrices \( \mathbf{D}_{jl} \) of the form

\[
\mathbf{D}_{jl} = \int_a^b \mathbf{B}(t) \mathbf{B}(t)^T \sum_{i=1}^n (D^{j-1}x_i(t)D^{l-1}x_i(t))dt
\]  

(4.5)

for \( j, l = 1, \ldots, m \). The vector \( \mathbf{d} \) consists of \( m \) subvectors \( \mathbf{d}_j \), where

\[
\mathbf{d}_j = \int_a^b \mathbf{B}(t) \sum_{i=1}^n (D^{j-1}x_i(t)D^m x_i(t))dt
\]  

(4.6)

for \( j = 1, \ldots, m \). Similarly \( \mathbf{H} \) consists of an \( m \times m \) array of \((n-2) \times n\) submatrices \( \mathbf{H}_{jl} \), where

\[
\mathbf{H}_{jl} = \begin{cases} 
\sqrt{\lambda_{j-1}} \mathbf{H}_0 & \text{if } j = l \\
0 & \text{otherwise}
\end{cases}
\]

for \( j, l = 1, \ldots, m \) and \( \mathbf{H}_0 \) is a \((n-2) \times n\) matrix with the elements,

\[
H_{h,i} = \begin{cases} 
1 & \text{if } i = h, h+2 \\
-2 & \text{if } i = h+1 \\
0 & \text{otherwise}
\end{cases}
\]

for \( h = 1, 2, \ldots, n-2 \) and \( i = 1, 2, \ldots, n \). We can estimate \( \beta \) by differentiating (4.4) with respect to \( \beta \) and setting the derivative equal to 0 as follows,

\[
2\mathbf{D}\beta + 2\mathbf{d} + 2\mathbf{H}^T \mathbf{H} \beta = 0
\]

\[
-\mathbf{d}(\mathbf{D} + \mathbf{H}^T \mathbf{H})^{-1} = \hat{\beta}.
\]

Once we have an estimate for \( \beta \), the \( w_j \)'s in (4.1) can be estimated. Using the \( \hat{w}_j \)'s we can estimate the linear differential operator \( L \). In computation, \( x_i(t) \) in (4.5) and (4.6) are substituted by a linear combination of B-spline basis functions and the integrals are replaced by a Reimann sum.
4.2 Low Dimensional Approximation for Functional Data

To have a low dimensional approximation of the data curves, we should find the null space basis functions $u_1, \ldots, u_m$. Once the $u_1, \ldots, u_m$ are found, we can approximate each curve $x_i(t)$ using a linear combination of the null space basis functions $\hat{x}_i(t) = \sum_{j=1}^{m} a_{ij} u_j$, for $i = 1, \ldots, n$. The Runge-Kutta method (Press et al(1992))(see [12]) is used to compute the basis functions iteratively. The initial conditions for the iteration are taken as $u(t)|_{t=0} = I$. Here $I$ is the identity matrix and $u(t)$ is a $m \times m$ matrix containing the basis functions $u_1, \ldots, u_m$ in the first row and their successive derivatives in subsequent rows.

The low dimensional approximation for the data curves is obtained by projecting data curves onto the null space using the Sobolev inner product $<.,.>_2$, where $< f,g >_{r,2} = \int_0^T \sum_{k=0}^{r} (D^k f)(D^k g)dt$. Ramsay has implemented this method for the square integral functions $L_2(0,T)$ using the inner product $< f,g >_2 = \int_0^T f(t)g(t)dt$, but this does not consider the fidelity to derivatives. Ramsay’s programs found in his web site “http://ego.psych.mcgill.ca/misc/fda/index.html” were used to obtain the basis functions. Projecting the data curves onto the null space was done using the Sobolev inner product.

The following examples illustrate PDA.

Examples:

1) The data that were used in the previous chapter are used here to illustrate and to compare and contrast PCA and PDA. First, the lip data was smoothed in a preprocessing step with smoothing parameter $1 \times e^{-12}$ and was analyzed with $\lambda = 1 \times e^{-6}$. Since the differential equation is of order two, we have to obtain two weight functions $w_0(t)$ and $w_1(t)$. These weight functions are shown in Figure 4.1. It can be seen that $w_0(t)$ is positive for almost all the time values, while $w_1(t)$ is closer to zero for many time values.
Figure 4.2 shows the basis functions satisfying the boundary conditions for the differential equation $Lx = 0$. When this is compared with the first two eigenfunctions corresponding to PCA (Figure 3.7) it can be seen that they are approximately the same. Since we have obtained the basis functions, we can approximate the observation curves using them. Figure 4.3 gives two observation curves and their approximations. It can be seen that we have obtained a good approximation using two basis functions.

Figure 4.1: The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 1 \times e^{-6}$ (lip data).
Figure 4.2: The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 1 \times e^{-6}$ (lip data).

Figure 4.3: Solid curves are observations and dashed curves are their approximations with $\lambda = 1 \times e^{-6}$ (lip data).
2) Our next example is simulated data discussed in the previous chapter. After smoothing data with smoothing parameter value $1 \times e^{-3}$, PDA was carried out with $\lambda = 1 \times e^{-6}$. Figure 4.4 gives the two weight functions corresponding to this. It can be seen that the weight functions are constant except in the boundary values of $t$. Figure 4.5 gives the two basis functions $u_1(t)$ and $u_2(t)$. When this is compared with the PC functions (Figure 3.10) we can say that these are approximately the same. Figure 4.6 shows that we have obtained a very good approximation to the data.

Figure 4.4: The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 1 \times e^{-6}$ (simulated data).
Figure 4.5: The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 1 \times e^{-6}$ (simulated data).

Figure 4.6: Observations and their approximations with $\lambda = 1 \times e^{-6}$ for simulated data (Observations and approximations are overlapped).
3) The final example is the normal hearing data analyzed in the previous chapter. The data were preprocessed by fitting cubic regression splines of dimension 13. PDA was carried out with $\lambda = 25$ and Figure 4.7 shows the weight functions corresponding to this data. Figure 4.8 shows the two basis functions corresponding to this data. Both of them follow a sinusoidal shape. When these are compared with the PC functions in Figure 3.13 it can be seen that they are not the same. This is an example to show that the results of PCA and PDA are not always the same. Two observations are approximated using the basis functions and are shown in Figure 4.9. It can be seen that we have obtained a somewhat closer approximation.

![Figure 4.7: The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 25$ (Hearing data).](image)

Figure 4.7: The two weight functions $w_0(t)$ and $w_1(t)$ with $\lambda = 25$ (Hearing data).
Figure 4.8: The two basis functions $u_1(t)$ and $u_2(t)$ with $\lambda = 25$ (Hearing data).

Figure 4.9: Solid curves are observations and dashed curves are their approximations with $\lambda = 25$ (Hearing data).
Chapter 5

Principal Differential Analysis with Covariates

5.1 Linear Differential Operator with a Covariate

In many practical situations functional data are collected from experimental units that are described by covariates. So the idea explained in the previous chapter is expanded in this chapter to include a single continuous covariate. PDA with covariates was developed by Jin(2006)[6] and are given here. Given $n$ functional observations $x_1, \ldots, x_n$ with covariates $v_1, \ldots, v_n$, the linear differential operator was defined as

$$L_v = w_0(. , v)I + w_1(. , v)D + \ldots + w_{m-1}(., v)D^{m-1} + D^m,$$

(5.1)

where $w_j(., v)(j = 0, \ldots, m-1)$ is a function of $t$ and $v$. The fitting criterion remains the same as before

$$\sum_{i=1}^{n} \int_{0}^{T} (L_i x_i(t))^2 dt,$$

where $L_i = L_{v_i} = v_i$ for $i = (1, \ldots, n)$. The coefficients $w_j(t, v)$ were assumed to be a product of a function of $t$ and a function of $v$. Therefore $w_j(t, v)$ can be approximated as

$$w_j(t, v) \approx (\sum_{i=1}^{I} \alpha_{ji} N_{i}(t))(\sum_{l=1}^{L} \beta_{jl} B_{l}(v)),$$

(5.2)

where $N_1(t), \ldots, N_I(t)$ are B-spline functions on the support $[0, T]$, and $B_1(v), \ldots, B_L(v)$ are B-Spline basis functions on the support of the covariate $v$. Let $\theta_{jil} = \alpha_{ji} \beta_{jl}$, then $\theta_{jil}$ are identifiable even though $\alpha_{ji}$ and $\beta_{jl}$ are not. Also let $N = (N_1(t), \ldots, N_I(t))^T, B = (B_1(v), \ldots, B_L(v))^T$.
The following penalized criterion is minimized with respect to $\beta$

$\left( B_1(v), \ldots, B_L(v) \right)^T, N_t \otimes B_v = \begin{bmatrix} (N_1(t)B_v) \\ \vdots \\ (N_L(t)B_v) \end{bmatrix}$. Let $\alpha_j = (\alpha_{1j}, \ldots, \alpha_{lj})^T, \beta_j = (\beta_{1j}, \ldots, \beta_{lj})^T$ ($j = 0, \ldots, m - 1$) and $\alpha = (\alpha_0^T, \ldots, \alpha_{m-1}^T)^T$, $\beta = (\beta_0^T, \ldots, \beta_{m-1}^T)^T$. Set $\alpha \times \beta = (\alpha_0 \otimes \beta_0)^T, \ldots, (\alpha_{m-1} \otimes \beta_{m-1})^T)^T$. Smoothness is imposed on $w_j$ ($j=0, \ldots, m-1$) by adding two penalty terms. The usual penalty term for $\alpha$ is $\sum_j D^2 \alpha_j$ and the usual penalty term for $\beta$ is $\sum_j D^2 \beta_j$. Therefore, the obvious choice for the penalized fitting criterion is

$$
\sum_{i=1}^n \int_0^T (L_i x_i(t))^2 dt + \sum_{j=0}^{m-1} \gamma_j \| D^2 \beta_j \|^2 + \sum_{j=0}^{m-1} \lambda_j \| D^2 \alpha_j \|^2
$$

$$
= (\alpha \times \beta)^T D(\alpha \times \beta) + 2(\alpha \times \beta)^T d + \alpha^T H_1^T H_1 \alpha + \beta^T H_2^T H_2 \beta + C. \quad (5.3)
$$

Here $D, H_1, H_2$ are matrices and $d$ is a vector and those will be defined later. The $\alpha$ and $\beta$ appear in (5.3) as $\alpha$ only terms, $\beta$ only terms, and $\alpha \times \beta$ terms, so that working with this criterion is not easy. Hence it is chosen to work with the penalty $\sum_{j=0}^{m-1} \gamma_j \| D^2 (\alpha_j \otimes \beta_j) \|^2$ that differs the $\beta_j$ and $\sum_{j=0}^{m-1} \lambda_j \| D^2 (\beta_j \otimes \alpha_j) \|^2$ that differs the $\alpha_j$. So the following penalized criterion is minimized with respect to $\alpha \times \beta$:

$$
f(\alpha, \beta) = \sum_{i=1}^n \int_0^T (L_i x_i(t))^2 dt + \sum_{j=0}^{m-1} \gamma_j \| D^2 (\alpha_j \otimes \beta_j) \|^2 + \sum_{j=0}^{m-1} \lambda_j \| D^2 (\beta_j \otimes \alpha_j) \|^2. \quad (5.4)
$$

After substituting (5.2) into (5.1), $L_i x_i(t)$ can be written using matrix notation as

$$
L_i x_i(t) = \begin{pmatrix}
\alpha_0^T \\
\vdots \\
\alpha_{m-1}^T 
\end{pmatrix}
\begin{pmatrix}
(\sum_{i=1}^L \beta_0 N_1(t) B_i(v_i)) \cdot x_i(t) \\
\vdots \\
(\sum_{i=1}^L \beta_{m-1} N_1(t) B_i(v_i)) \cdot D^{m-1} x_i(t) 
\end{pmatrix}
+ D^m x_i(t)
$$

41
\[
\begin{align*}
\left( (\alpha_0 \otimes \beta_0)^T, \ldots, \left( \alpha_{m-1} \otimes \beta_{m-1} \right)^T \right) & \left( \begin{array}{c}
(N_t \otimes B_{v_1}) \cdot x_1(t) \\
\vdots \\
(N_t \otimes B_{v_1}) \cdot D^{m-1}x_1(t)
\end{array} \right) + D^m x_1(t) \\
= L_{i,m-1} x_1(t) + D^m x_1(t).
\end{align*}
\]

Substituting this into (5.4),

\[
f(\alpha, \beta) = \sum_{i=1}^{n} \int_{0}^{T} (L_{i,m-1} x_1(t))(L_{i,m-1} x_1(t))^T dt + 2 \sum_{i=1}^{n} \int_{0}^{T} (L_{i,m-1} x_1(t))D^m x_1(t) dt
+ \sum_{i=1}^{n} \int_{0}^{T} (D^m x_1(t))^2 dt + (\alpha \times \beta)^T H_1^T H_1 (\alpha \times \beta) + (\alpha \times \beta)^T H_2^T H_2 (\alpha \times \beta).
\]

This can be written using matrix notation as

\[
f(\alpha, \beta) = (\alpha \times \beta)^T D(\alpha \times \beta) + 2(\alpha \times \beta)^T d + \sum_{i=1}^{n} \int_{0}^{T} (D^m x_1(t))^2 dt
+ (\alpha \times \beta)^T H_1^T H_1 (\alpha \times \beta) + (\alpha \times \beta)^T H_2^T H_2 (\alpha \times \beta). \tag{5.5}
\]

Minimizing (5.5) with respect to \( \alpha \times \beta \), is easier than minimizing (5.3).

\( D \) in (5.5) is a symmetric matrix consisting of an \( m \times m \) array of \( IL \times IL \) submatrices \( D_{jl} \) of the form

\[
D_{jl} = \int_{0}^{T} \sum_{i=1}^{n} (D^{j-1} x_1(t))(D^{l-1} x_1(t))(N_t \otimes B_{v_1})(N_t \otimes B_{v_1})^T dt \quad (j, l = 1, \ldots, m).
\]

Similarly, \( d \) in (5.5) is a vector that consists of \( m \) subvectors \( d_j \) of length \( IL \) of the form

\[
d_j = \int_{0}^{T} \sum_{i=1}^{n} (D^{j-1} x_1(t))(D^m x_1(t))(N_t \otimes B_{v_1}) dt \quad (j = 1, \ldots, m).
\]
\( \mathbf{H}_1 \) in (5.5) is a \( m(IL - 2) \times mIL \) matrix that consists of an \( m \times m \) array of \( (IL - 2) \times IL \) submatrices \( \mathbf{H}_{jl}^{(1)} \) of the form

\[
\mathbf{H}_{jl}^{(1)} = \begin{cases} 
\sqrt{\gamma_{j-1}} \mathbf{H}_0^{(1)} & \text{if } j = l \\
0 & \text{otherwise}
\end{cases}
\]

for \( j, l = 1, \ldots, m \), and \( \mathbf{H}_0^{(1)} = (H_{h,k}) \) is a \( (IL - 2) \times IL \) matrix of the form

\[
H_{h,k} = \begin{cases} 
1 & \text{if } k = h, h + 2 \\
-2 & \text{if } k = h + 1 \\
0 & \text{otherwise}
\end{cases}
\]

for \( h = 1, 2, \ldots, IL - 2, k = 1, 2, \ldots, IL \). The matrix \( \mathbf{H}_2 \) in (5.5) is a \( m(IL - 2) \times mIL \) matrix, and consists of an \( m \times m \) array of \( (IL - 2) \times IL \) submatrices \( \mathbf{H}_{jl}^{(2)} \) of the form

\[
\mathbf{H}_{jl}^{(2)} = \begin{cases} 
\sqrt{\lambda_{j-1}} \mathbf{H}_0^{(2)} & \text{if } j = l \\
0 & \text{otherwise}
\end{cases}
\]

for \( j, l = 1, \ldots, m \), and \( \mathbf{H}_0^{(2)} = \mathbf{H}_0^{(1)} \mathbf{P} \) is a \( (IL - 2) \times IL \) matrix, where \( \mathbf{P} \) is an orthogonal matrix of order \( IL \) satisfying \( (\beta_j \otimes \alpha_j) = \mathbf{P}(\alpha_j \otimes \beta_j) \).

The solution for the minimization problem (5.5) satisfies

\[
(\mathbf{D} + \mathbf{H}_1^T \mathbf{H}_1 + \mathbf{H}_2^T \mathbf{H}_2)(\alpha \times \beta) = -\mathbf{d},
\]

therefore, the estimates \( \hat{\alpha}, \hat{\beta} \) are found by solving this linear equation. Once we have estimates for \( \alpha \) and \( \beta \), \( w_0(t, v), \ldots, w_{m-1}(t, v) \), the coefficients of the equation (5.1) can be estimated. Having estimates for \( w_0(t, v), \ldots, w_{m-1}(t, v) \) means we have an estimate for the linear differential operator \( L_v \) in (5.1). When \( v \) is fixed, we can use the same methods we discussed in the previous chapter to compute the basis functions to be used for a low dimensional approximation of the data curves.
Chapter 6
Simulations

A Monte-Carlo simulation study was carried out to understand the effect of the smoothing parameters $\lambda, \gamma$ in the penalized fitting criterion and the results of the simulation study are discussed in this chapter. 50 simulated data sets were generated and the same 50 data sets were analyzed with each pair of $\lambda, \gamma$ values in a grid with $\gamma = (0.0025, 0.25, 2.5, 25, 250, 2500)$ and $\lambda = (0.0025, 0.0125, 0.025, 0.25, 25, 2500, 25000, 250000, 25000000)$. The pointwise mean and variance of the estimators of the basis functions $U_1(t, v)$ and $U_2(t, v)$ were calculated using the 50 estimates. The coefficients $w_0(t, v)$ and $w_1(t, v)$ of the second order differential operator ($m = 2$ in (5.1)) were fixed as,

$$w_0(t, v) = 10v^2, w_1(t, v) = 2v$$

for $t, v \in [0, 2]$.

This second order differential equation annihilates the curves $e^{-vt}(a_1\cos(3vt) + a_2\sin(3vt))$. The null space basis functions that satisfy the initial condition $U_1(0, v) = 1, U_2(0, v) = 0, U_1'(0, v) = 0, U_2'(0, v) = 1$ are:

$$U_1(t, v) = e^{-vt}\cos(3vt) + \frac{1}{3}e^{-vt}\sin(3vt)$$
$$U_2(t, v) = \frac{1}{3v}e^{-vt}\sin(3vt).$$

The functional data were randomly generated using

$$x_i(t, v) = e^{-vt}(a_{1,i}\cos(3vt) + a_{2,i}\sin(3vt)) + \sigma\epsilon_i(t),$$

(6.1)

where

$$\begin{pmatrix} a_{1,i} \\ a_{2,i} \end{pmatrix} \sim N_2 \left( \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \right)$$

and $\epsilon_i \sim N(0, 1)$ for $i = 1, \ldots, 50$ were independent of $(a_{1,i}, a_{2,i})^T$. Here the covariate values were set to $v_i = \frac{2i}{50}, i = 1, \ldots, 50$ and the
curves were sampled on the regular grid \( t_j = \frac{j-1}{100}, j = 1, \ldots, 200. \)

The simulations were run without noise (\( \sigma = 0 \) in (6.1)) and with noise (\( \sigma = 0.15 \) in (6.1)) with the same seed. The simulated sample curves were preprocessed using smoothing. The value of the smoothing parameter in the preprocessing step was set to 0.00001 when \( \sigma = 0 \) and increased to 10 when \( \sigma = 0.15 \). The methods described in Chapter 5 were used to estimate the coefficients \( \hat{w}_0(t,v), \hat{w}_1(t,v) \) of the second order linear differential operator. R-software was used for analysis using a cubic B-spline bases of dimension \( I = L = 20 \) for regularly spaced knots in \( t \) and \( v \). The analysis was carried out for the quartiles of the covariate \( v = (0.5, 1.0, 1.5) \). Dr. Ramsay’s algorithms available at “http://ego.psych.mcgill.ca/misc/fda” were used to obtain the estimates for the null space basis functions \( U_1(t,v) \) and \( U_2(t,v) \) for \( v = 0.5, 1.0 \) and 1.5.

Tables 6.1a and 6.1b show the values of variance and squared bias of \( \hat{U}_1(t,v) \) and \( \hat{U}_2(t,v) \) averaged over both \( t \in \{0.00, \ldots, 1.99\} \) and \( v \in \{0.5, 1.0, 1.5\} \) for \( \sigma = 0 \) and a grid of values of smoothing parameters \( \lambda \) and \( \gamma \). Tables 6.2a and 6.2b show the corresponding values for \( \sigma = 0.15 \). Tables 6.1a, 6.2a show the average variance and average squared bias of \( \hat{U}_1(t,v) \) and Tables 6.1b, 6.2b show the corresponding values of \( \hat{U}_2(t,v) \). To identify relationships clearly, the average variance and the average squared bias were plotted for chosen values of \( \lambda \) versus \( \log \gamma \). Those are shown in Figure 6.1-without noise(\( \sigma = 0 \)) and 6.2-with noise(\( \sigma = 0.15 \)).

According to Figure 6.1 it can be said that for the data curves that are not contaminated with noise, the average variance decreases with \( \lambda \) and increases with \( \gamma \). The average squared bias increases with \( \gamma \) and decreases with \( \lambda \) except for large \( \lambda \) values that lead to oversmoothing. Figure 6.2 shows that for the data curves that are contaminated with noise, the average variance decreases with \( \lambda \), but decreases with \( \gamma \) for small values of \( \lambda \) and increases with \( \gamma \) for large values of \( \lambda \). The average squared bias increases with \( \gamma \) and decreases with \( \lambda \) except for large \( \lambda \) values that lead to oversmoothing.
Table 6.1a: $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_1$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0$)

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>0.0025</th>
<th>0.0125</th>
<th>0.025</th>
<th>0.25</th>
<th>25</th>
<th>2500</th>
<th>$2.5 \times 10^5$</th>
<th>$2.5 \times 10^7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0025</td>
<td>10^4 Variance</td>
<td>49</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>33</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>28</td>
<td>14</td>
<td>13</td>
<td>17</td>
<td>18</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>42</td>
<td>33</td>
<td>35</td>
<td>49</td>
<td>40</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>2500</td>
<td></td>
<td>58</td>
<td>36</td>
<td>37</td>
<td>53</td>
<td>45</td>
<td>40</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>$10^3$ Bias$^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0025</td>
<td></td>
<td>110</td>
<td>16</td>
<td>6</td>
<td>0.6</td>
<td>0.2</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>116</td>
<td>20</td>
<td>8</td>
<td>1</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>136</td>
<td>32</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>90</td>
<td>39</td>
<td>21</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>87</td>
<td>64</td>
<td>55</td>
<td>55</td>
<td>50</td>
<td>39</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>2500</td>
<td></td>
<td>133</td>
<td>98</td>
<td>93</td>
<td>101</td>
<td>91</td>
<td>69</td>
<td>71</td>
<td>71</td>
</tr>
</tbody>
</table>
Table 6.1b: $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_2$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0$)

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>0.0025</th>
<th>0.0125</th>
<th>0.025</th>
<th>0.25</th>
<th>25</th>
<th>2500</th>
<th>2.5×10^5</th>
<th>2.5×10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>42</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>0.25</td>
<td>19</td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>0.8</td>
<td>0.4</td>
<td>0.7</td>
<td>3</td>
</tr>
<tr>
<td>2.5</td>
<td>34</td>
<td>23</td>
<td>16</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>41</td>
<td>38</td>
<td>28</td>
<td>11</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>250</td>
<td>79</td>
<td>56</td>
<td>39</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2500</td>
<td>122</td>
<td>56</td>
<td>37</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>0.0025</th>
<th>0.0125</th>
<th>0.025</th>
<th>0.25</th>
<th>25</th>
<th>2500</th>
<th>2.5×10^5</th>
<th>2.5×10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^3$Bias^2</td>
<td>27</td>
<td>10</td>
<td>6</td>
<td>0.8</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.25</td>
<td>25</td>
<td>10</td>
<td>5</td>
<td>0.7</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>2.5</td>
<td>48</td>
<td>26</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>39</td>
<td>26</td>
<td>16</td>
<td>2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>250</td>
<td>39</td>
<td>24</td>
<td>18</td>
<td>8</td>
<td>9</td>
<td>14</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>2500</td>
<td>42</td>
<td>29</td>
<td>25</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 6.2a: \(10^4\times\) Average Variance and \(10^3\times\) Average Squared-Bias of \(\hat{U}_1\) for grid of \(\lambda\) and \(\gamma\) values (\(\sigma = 0.15\))

<table>
<thead>
<tr>
<th>(\gamma)</th>
<th>(\lambda)</th>
<th>0.25</th>
<th>25</th>
<th>2500</th>
<th>25000</th>
<th>2.5(\times)10^5</th>
<th>2.5(\times)10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0025</td>
<td>10^4Variance</td>
<td>210</td>
<td>66</td>
<td>61</td>
<td>45</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>72</td>
<td>29</td>
<td>38</td>
<td>40</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>43</td>
<td>18</td>
<td>28</td>
<td>31</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>33</td>
<td>21</td>
<td>30</td>
<td>33</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>37</td>
<td>40</td>
<td>44</td>
<td>48</td>
<td>49</td>
<td>54</td>
</tr>
<tr>
<td>2500</td>
<td></td>
<td>33</td>
<td>44</td>
<td>62</td>
<td>68</td>
<td>69</td>
<td>70</td>
</tr>
<tr>
<td>0.0025</td>
<td>10^3Bias^2</td>
<td>54</td>
<td>7</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>44</td>
<td>5</td>
<td>15</td>
<td>17</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td>52</td>
<td>8</td>
<td>23</td>
<td>25</td>
<td>29</td>
<td>67</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>56</td>
<td>8</td>
<td>21</td>
<td>24</td>
<td>25</td>
<td>82</td>
</tr>
<tr>
<td>250</td>
<td></td>
<td>69</td>
<td>53</td>
<td>87</td>
<td>96</td>
<td>97</td>
<td>116</td>
</tr>
<tr>
<td>2500</td>
<td></td>
<td>94</td>
<td>96</td>
<td>138</td>
<td>152</td>
<td>155</td>
<td>155</td>
</tr>
</tbody>
</table>
Table 6.2b: $10^4 \times$ Average Variance and $10^3 \times$ Average Squared-Bias of $\hat{U}_1$ for grid of $\lambda$ and $\gamma$ values ($\sigma = 0.15$)

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>0.25</th>
<th>25</th>
<th>2500</th>
<th>25000</th>
<th>$2.5 \times 10^5$</th>
<th>$2.5 \times 10^7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>0.0025</td>
<td>426</td>
<td>44</td>
<td>32</td>
<td>23</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>151</td>
<td>18</td>
<td>23</td>
<td>23</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>$10^4$ Variance</td>
<td>2.5</td>
<td>133</td>
<td>19</td>
<td>17</td>
<td>18</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>163</td>
<td>28</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>199</td>
<td>30</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2500</td>
<td>240</td>
<td>45</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0025</td>
<td>125</td>
<td>20</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>155</td>
<td>27</td>
<td>16</td>
<td>17</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>$10^3$ Bias$^2$</td>
<td>2.5</td>
<td>231</td>
<td>50</td>
<td>39</td>
<td>39</td>
<td>43</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>278</td>
<td>39</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>276</td>
<td>43</td>
<td>21</td>
<td>20</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>2500</td>
<td>332</td>
<td>69</td>
<td>34</td>
<td>33</td>
<td>33</td>
<td>34</td>
</tr>
</tbody>
</table>
Figure 6.1: Variance and Squared-Bias averaged over $t$ and $v$ for various values of smoothing parameters $\lambda$ and $\gamma$ ($\sigma = 0$).
Figure 6.2: Variance and Squared-Bias averaged over $t$ and $v$ for various values of smoothing parameters $\lambda$ and $\gamma$ ($\sigma = 0.15$).

To confirm the behavior of the average variance as a function of $\gamma$, the 50 realizations of the estimators $\hat{U}_1$ and $\hat{U}_2$ for the data not contaminated with noise were drawn in Figures 6.3a and 6.3b with $v=1$; the figures with $v=0.5, 1, 1.5$ are given in appendix-A. Figure 6.3a shows the estimates for fixed small $\lambda$ ($\lambda = 0.25$) for undersmoothing in $t$. Figure 6.3b shows the estimates for fixed large $\lambda$ ($\lambda = 25000000$) for oversmoothing in $t$. These figures
confirm the behavior that was seen in the first column of Figure 6.1; for the data that is not contaminated with noise, the average variance of the estimators is increasing in $\gamma$. To see the effect of the smoothing parameter $\lambda$ for the data that is not contaminated with noise, Figure 6.4a was drawn for fixed small $\gamma$ ($\gamma=0.0025$) and Figure 6.4b for large $\gamma$ ($\gamma=25$).

Here, only the covariate value $v=1$ is shown and plots corresponding to the other values of $v$ are given in appendix-A. Figures 6.4a and 6.4b show that the average variance decreases with increasing $\lambda$ except for very large $\lambda$($\lambda=25000000$) that leads to oversmoothing.

According to the Figures A.1a($v = 0.5, \lambda = 0.25$), A.1c($v = 1.5, \lambda = 0.25$), A.2a($v = 0.5, \lambda = 25000000$) and A.2c($v = 1.5, \lambda = 25000000$) it can be seen that for the boundary values of $v$ the average variance is increasing with increasing in $\gamma$. According to the Figures A.3a($v = 0.5, \gamma = 0.0025$) and A.3c($v = 1.5, \gamma = 0.0025$), A.4a($v = 0.5, \gamma = 25$) and A.4c($v = 1.5, \gamma = 25$) it can be seen that for the left boundary undersmoothing in $v$ and for the right boundary oversmoothing in $v$, average variance is first decreasing and then increasing with smoothing in $t$. For the right boundary undersmoothing in $v$ and left boundary oversmoothing in $v$ average variance is decreasing with smoothing in $t$. These results can be summarized as the boundary points $v = 0.5, 1.5$ are behaving the same as the interior $v = 1.0$ for smoothing in $v$, but for smoothing in $t$ they are not behaving the same.
Figure 6.3a: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0$, $\nu = 1.0$).
Figure 6.3b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, \nu = 1.0$).
Figure 6.4a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0$, $v = 1.0$).
Figure 6.4b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$, $\nu = 1.0$).

To confirm the behavior of the average variance when the data is contaminated with noise, 50 realizations of the estimators $\hat{U}_1$ and $\hat{U}_2$ were drawn with $\nu = 1$; plots corresponding to the other values of the covariate $\nu$ are given in appendix-A. Figure 6.5a shows the estimates for small $\lambda$ ($\lambda = 0.25$) and Figure 6.5b shows the estimates for large $\lambda$ ($\lambda = 25000000$). Figure 6.5a shows that the variance of the estimators $\hat{U}_1$ and $\hat{U}_2$ decreasing in $\gamma$ and then
changing to increasing in $\gamma$ for undersmoothing in $t$. Figure 6.5b shows that for oversmoothing in $t$ the variance is increasing with $\gamma$. Focusing on the interior, these results can be summarized as the average variance of the estimators is increasing with smoothing over the covariate $v$, except where undersmoothing in $t$ on the data contaminated with noise. To see the effect of smoothing in $t$ for the data contaminated with noise, Figure 6.6a was drawn for small $\gamma(\gamma=0.0025)$ and Figure 6.6b was drawn for large $\gamma(\gamma=25)$. Only the figures corresponding to $v=1$ are shown here; the figures for $v = 0.5,1,1.5$ are given in appendix-A. Figures 6.6a and 6.6b confirm the behavior that we observed in the first column of the Figure 6.2 namely that the average variance decreases with the value of the smoothing parameter $\lambda$ that smooths in $t$.

According to the Figures A.5a($v = 0.5, \lambda = 0.25$), A.5c($v = 1.5, \lambda = 0.25$), A.6a($v = 0.5, \lambda = 25000000$) and A.6c($v = 1.5, \lambda = 25000000$) it can be seen that for the boundary values of $v$ the average variance is decreasing with smoothing in $v$ for undersmoothing in $t$ and increasing with smoothing in $v$ for oversmoothing in $t$. Figures A.7a($v = 0.5, \gamma = 0.0025$), A.7c($v = 1.5, \gamma = 0.0025$), A.8a($v = 0.5, \gamma = 25$) and A.8c($v = 1.5, \gamma = 25$) show that for the boundary values of $v$, the average variance is decreasing with smoothing in $t$. These results can be summarized as the boundary points $v = 0.5,1.5$ are behaving the same as the interior $v = 1.0$ for smoothing in $t$. For smoothing in $v$ the boundary points are behaving the same except for undersmoothing in $t$. 
Figure 6.5a: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15, v = 1.0$).
Figure 6.5b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15$, $v = 1.0$).
Figure 6.6a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15$, $v = 1.0$).
Figure 6.6b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, v = 1.0$).

Now we are interested in examining the behavior of the pointwise bias of $\hat{U}_1$ and $\hat{U}_2$ as a function of $\gamma$ for fixed $\lambda=0.25$ (Figure 6.7a with $\sigma=0$; Figure 6.7b with $\sigma=0.15$) and as a function of $\lambda$ for fixed $\gamma=25$ (Figure 6.8a with $\sigma=0$; Figure 6.8b with $\sigma=0.15$). It can be seen that the bias is smaller for the interior value $v=1.0$ than the boundary values $v=0.5$ and 1.5. For the interior value of the covariate $v=1.0$, bias is increasing with smoothing.
in \( v \) and is decreasing up to some point and then increasing with smoothing in \( t \). For the boundary values of \( v \), the bias is generally increasing with smoothing in \( v \) and first decreasing and then increasing with smoothing in \( t \).

Figure 6.7a: Simulated Pointwise Bias for various values of \( \gamma \) with \( \lambda = 0.25 \) (\( \sigma = 0 \)).
Figure 6.7b: Simulated Pointwise Bias for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15$).
Figure 6.8a: Simulated Pointwise Bias for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$).
Finally, we are interested in the effect of changing variances of the coefficients $a_{1,i}$ and $a_{2,i}$ in (6.1). The covariance of $a_{1,i}$ and $a_{2,i}$ was taken as $c = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and the value of $c$ was $c = 0.5, 1, 2, 3$ and 4. As before, 50 simulated data sets were generated without noise ($\sigma = 0$) and with noise ($\sigma = 0.15$) for each $c$ using the same seed. Data were analyzed with each pair of $\lambda$ and $\gamma$ with $\gamma = (0.0025, 0.25, 2.5, 25, 250, 2500)$ and $\lambda = 0.25$ when $\sigma = 0$ and $\lambda = 250000$ when $\sigma = 0.15$. Here $\lambda = 0.25$ was used for the data that is not contaminated.
with noise because it is the value that has the smallest average variance and the smallest average bias across $\gamma$; $\lambda = 250000$ was used for the data that is contaminated with noise because it is the first value displaying the average variance increasing with $\gamma$. Variance and squared bias of $\hat{U}_1(t, v)$ and $\hat{U}_2(t, v)$ averaged over both $t$ and $v$ were obtained as earlier and were plotted against $\log \gamma$ with changing $c$ in Figure 6.9 ($\sigma = 0$) and Figure 6.10 ($\sigma = 0.15$). As expected, the average variance is increasing with increasing $c$, but average squared bias is decreasing with increasing $c$ (with increasing the variance of $a_{1,i}$ and $a_{2,i}$). At this point we do not have an explanation for this unexpected behavior of the bias.
Figure 6.9: Variance and Squared-Bias averaged over $t$ and $v$ for different values of $c$ ($\lambda=0.25$, $\sigma = 0$).
Figure 6.10: Variance and Squared-Bias averaged over $t$ and $v$ for different values of $c$ ($\lambda=250000, \sigma = 0.15$).
Chapter 7

Analysis of Data

The hearing data introduced in Chapter 3 are analyzed in this chapter using Principal Differential Analysis with a covariate. Figure 7.1 shows the age distribution of the 92 individuals with normal hearing. It can be seen that the majority of the individuals’ ages are less than 6 years. There are very few individuals within the ages 7-12 years.

Figure 7.1: Age distribution of individuals with normal hearing
PDA with a covariate method is first applied to the data of the individuals with normal hearing using the smoothing parameter $\lambda=25$ and 4 different values of the smoothing parameter $\gamma=0.0025, 0.25, 25$ and 2500. The basis functions $\hat{U}_1$ and $\hat{U}_2$ corresponding to the covariate (age) values 0.5, 1, 2, 3, 4, 5, 13, 16, 19 are shown in Figures 7.2a ($\gamma=0.0025$), 7.2b ($\gamma=0.25$), 7.2c ($\gamma=25$) and 7.2d ($\gamma=2500$). It can be seen that both the basis functions $\hat{U}_1$ and $\hat{U}_2$ have a sinusoidal shape. Changing $\gamma$ does not have much effect on the shape of the estimated basis functions. The coefficients that give a low dimensional approximation for the data curves using the null space basis functions, are given in Figures 7.3a ($\gamma=0.0025$), 7.3b ($\gamma=0.25$), 7.3c ($\gamma=25$) and 7.3d ($\gamma=2500$). The line in the coefficient figures represents the loess smoothing (R software local polynomial regression fitting) curve for the smoothing value 0.75. Increasing $\gamma$ shrinks the coefficient values. It can be seen that the coefficients approach a constant with increasing the value of the covariate (age).
Figure 7.2a: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=0.0025$).
Figure 7.2b: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=0.25$).
Figure 7.2c: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=25$).
Figure 7.2d: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\lambda=25$ and $\gamma=2500$).
Figure 7.3a: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\lambda=25$ and $\gamma=0.0025$).
Figure 7.3b: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\lambda=25$ and $\gamma=0.25$).
Figure 7.3c: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L (\lambda=25$ and $\gamma=25)$. 
Figure 7.3d: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\lambda=25$ and $\gamma=2500$).

Now, PDA with a covariate method is applied to the normal hearing data using the smoothing parameter $\gamma=2.5$ and 4 different values of the smoothing parameter $\lambda=0.0025$, 0.025, 0.25 and 2.5. The estimated basis functions $\hat{U}_1$ and $\hat{U}_2$ corresponding to the covariate(age) values 0.5, 1, 2, 3, 4, 5, 13, 16, 19 are shown in Figures 7.4a ($\lambda=0.0025$), 7.4b ($\lambda=0.025$), 7.4c ($\lambda=0.25$) and 7.4d ($\lambda=2.5$). It can be seen that both the basis functions
$\tilde{U}_1$ and $\tilde{U}_2$ have the sinusoidal shape same as earlier. Changing $\lambda$ has more of an effect on the shape of the estimated basis functions than changing $\gamma$. The coefficients that give a low dimensional approximation for the data curves using the null space basis functions, are given in Figures 7.5a ($\lambda=0.0025$), 7.5b ($\lambda=0.025$), 7.5c ($\lambda=0.25$) and 7.5d ($\lambda=2.5$). The line in the coefficient figures represents the loess smoothing curve for the smoothing value 0.75. Increasing $\lambda$ shrinks the coefficient values, but less than $\gamma$ did. It can be seen that the coefficients approach a constant with increasing value of the covariate (age). Both changing $\gamma$ and changing $\lambda$ have affected the estimates that are in the right boundary (higher age values) where the data on the covariate age are sparse. In summary for this data analysis, $\gamma$ seems to have a greater impact on the magnitude of the coefficients than $\lambda$. But $\lambda$ has a greater impact on the estimated basis functions than $\gamma$. 
Figure 7.4a: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\gamma=2.5$ and $\lambda=0.0025$).
Figure 7.4b: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\gamma=2.5$ and $\lambda=0.025$).
Figure 7.4c: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\gamma=2.5$ and $\lambda=0.25$).
Figure 7.4d: Null space basis functions of normal hearing data for different covariate (age) values: Solid line for $\hat{U}_1$, Dashed line for $\hat{U}_2$ ($\gamma=2.5$ and $\lambda=2.5$).
Figure 7.5a: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\gamma=2.5$ and $\lambda=0.0025$).
Figure 7.5b: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\gamma=2.5$ and $\lambda=0.025$).
Figure 7.5c: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L$ ($\gamma=2.5$ and $\lambda=0.25$).
Figure 7.5d: Coefficients that give a low dimensional approximation for the normal hearing data curves using the null space of basis functions of $L (\gamma=2.5$ and $\lambda=2.5)$. 
Chapter 8

Conclusion

The main purpose of this thesis was to study the effect of the smoothing parameters in PDA with a covariate. This was done using a simulation study. The variance and the squared bias of the estimates of the basis functions, averaged over the covariate $v$ and time $t$ were obtained. It was found that for the data curves that are not contaminated with noise, the average variance decreases with smoothing in $t$ and increases with smoothing in $v$. The average squared bias increases with smoothing in $v$ and decreases with smoothing in $t$ except for oversmoothing in $t$. For the data curves that are contaminated with noise, the average variance decreases with smoothing in $t$, but decreases with smoothing in $v$ for undersmoothing in $t$ and increases with smoothing in $v$ for oversmoothing in $t$. The average squared bias increases with smoothing in $v$ and decreases with smoothing in $t$ except for oversmoothing in $t$. The effect of changing the variance of the coefficients in the generated curves was also analyzed. It was found that the average variance is increasing with increasing the variance of the coefficients and surprisingly the average bias is decreasing with increasing the variance of the coefficients.

The hearing data for normal subjects was analyzed using PDA with a covariate. It was found that smoothing in the covariate(age) has greater impact on the magnitude of the coefficients than smoothing in time. But smoothing in time has greater impact on the estimated basis functions than smoothing in the covariate. The basis functions do not seem to be changing with age. The coefficients were converging to a constant as age increases. PDA was better than PCA at capturing the shape of the curves with a low-dimensional approximation using $m = 2$. 
References


Appendix A

More Simulation Results

More simulation results are given here. 50 realizations of the estimators \( \hat{U}_1 \) and \( \hat{U}_2 \) are first drawn with varying \( \gamma \) for a fixed small \( \lambda \) (\( \lambda = 0.25 \)) and for a fixed large \( \lambda \) (\( \lambda = 25000000 \)). Then, they were drawn with varying \( \lambda \) for a fixed small \( \gamma \) (\( \gamma = 0.0025 \)) and for a fixed large \( \gamma \) (\( \gamma = 25 \)). All the graphs are drawn for the data not contaminated with noise (\( \sigma = 0 \)) and for the data contaminated with noise (\( \sigma = 0.15 \)) for the covariate values 0.5, 1.0 and 1.5.
A.1 Without Noise

Figure A.1a: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0$, $v = 0.5$).
Figure A.1b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0, v = 1.0$).
Figure A.1c: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0$, $v = 1.5$).
Figure A.2a: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, \nu = 0.5$).
Figure A.2b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000 \ (\sigma = 0, \nu = 1.0)$. 
Figure A.2c: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0, v = 1.5$).
Figure A.3a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0, \nu = 0.5$).
Figure A.3b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0, \nu = 1.0$).
Figure A.3c: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0$, $v = 1.5$).
Figure A.4a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0, \nu = 0.5$).
Figure A.4b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$, $\nu = 1.0$).
Figure A.4c: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0$, $\nu = 1.5$).
A.2 With Noise

\[ \gamma = 0.0025 \]  
\[ \gamma = 2.5 \]  
\[ \gamma = 2500 \]

**Figure A.5a:** Simulated estimates of the basis functions for various values of \( \gamma \) with \( \lambda = 0.25 \) (\( \sigma = 0.15 \), \( \nu = 0.5 \)).
Figure A.5b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15$, $v = 1.0$).
Figure A.5c: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 0.25$ ($\sigma = 0.15$, $v = 1.5$).
Figure A.6a: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, \nu = 0.5$).
Figure A.6b: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, \nu = 1.0$).
Figure A.6c: Simulated estimates of the basis functions for various values of $\gamma$ with $\lambda = 25000000$ ($\sigma = 0.15, \nu = 1.5$).

\[ \gamma = 0.0025 \quad \gamma = 2.5 \quad \gamma = 2500 \]

\[ \hat{u}(t, \nu) \]

\[ U^1(t, \nu) \]

\[ U^2(t, \nu) \]
Figure A.7a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15, \nu = 0.5$).
Figure A.7b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 0.0025$ ($\sigma = 0.15$, $v = 1.0$).
Figure A.7c: Simulated estimates of the basis functions for various values of λ with γ = 0.0025 (σ = 0.15, v = 1.5).
Figure A.8a: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15, \nu = 0.5$).
Figure A.8b: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15$, $v = 1.0$).
Figure A.8c: Simulated estimates of the basis functions for various values of $\lambda$ with $\gamma = 25$ ($\sigma = 0.15$, $v = 1.5$).
Curriculum Vitae


In the fall of 2009, he entered the Graduate School of The University of Texas at El Paso. While pursuing a master’s degree in Statistics, he worked as a Teaching Assistant until December of 2010. He is currently working as a Research Technician III under the supervision of Dr. Joan Staniswalis, conducting research in functional data analysis, since February 2011. He is planning to continue studying for a PhD after completing his master’s degree.

Present address: 5325 Hanawalt Dr.
El Paso, Texas 79903