This document is comprised of two parts: two chapters on the subject of Bayesian functional data analysis, and two chapters on the solvability of nonlinear equations.

Chapter 2 deals with the clustering of functional data. The use of exploratory methods is an important step in the understanding of data. When clustering functional data, most methods use traditional clustering techniques on a vector of estimated basis coefficients, assuming that the underlying signal functions live in the $L_2$-space. Bayesian methods use models which imply the belief that some observations are realizations from some signal plus noise models with identical underlying signal functions. The method we propose differs in this respect: we employ a model that does not assume that any of the signal functions are truly identical, but possibly share many of their local features, represented by coefficients in a multiresolution wavelet basis expansion. We cluster each wavelet coefficient of the signal functions using conditionally independent Dirichlet process priors, thus focusing on exact matching of local features. We then demonstrate the method using two datasets from different fields to show broad application potential.

Chapter 3 studies the area of principal components analysis (PCA), which has seen relatively few contributions from the Bayesian school of inference. In this chapter, we propose a Bayesian method for PCA in the case of functional data observed with error. We suggest modeling the covariance function by use of an approximate spectral decomposition, leading to easily interpretable parameters. We perform model selection, both over the number of principal components and the number of basis functions used in the approximation. We study in depth the choice of using the implied distributions arising from the inverse Wishart prior and prove a convergence theorem for the case of an exact finite dimensional representation. We also discuss computational issues, as well as the care needed in choosing hyperparameters. A simulation study is used to demonstrate competitive performance against a recent frequentist procedure, particularly in terms of the principal component estimation. Finally, we apply the method to a real dataset, where we also incorporate model selection on the dimension of the finite basis used for modeling.

In Chapter 5, we provide sufficient conditions for the existence of solutions to certain classes of second-order discrete and continuous systems. In particular, we examine problems that can be posed as nonlinear perturbations of Sturm-Liouville problems. We first provide a lemma on the
invertibility of a nonlinearily-perturbed invertible linear operator, and apply this result to extend previous work on these topics.

In Chapter 6, we provide sufficient conditions for the existence of solutions to nonlinear boundary value problems. We do so by applying a general abstract strategy for solving nonlinear equations with a linear component. We apply this to general systems by first isolating a linear periodic system and then using the general theory of periodic solutions to find conditions on the additional nonlinear components to guarantee solutions.
Bayesian Methods for Exploratory Functional Data Analysis and Existence Theorems for Solutions to Nonlinear Differential and Difference Equations

by
Adam Justin Suarez

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics and Mathematics

Raleigh, North Carolina

2016

APPROVED BY:

Dr. Subhashis Ghosal
Co-chair of Advisory Committee

Dr. Jesús Rodríguez
Co-chair of Advisory Committee

Dr. Stephen Schecter

Dr. Eric Stone

Dr. Justin Post
DEDICATION

To my parents.
BIOGRAPHY

I don’t remember where I was before I came to this world, and I’m still not quite sure where I am now, but I certainly hope I remember this place when I get to wherever I’m going next.
ACKNOWLEDGEMENTS

Thank you to my advisors for all of their support and encouragement. It has been a great honor to work so closely with both of them. Thank you also to my entire committee, each one of whom has had a major influence on my graduate career.

Thank you to Sarah Thomson, who not only supported me through my decision to apply to graduate school, but has been one of the major inspirations in my life, more than she could ever know. Thank you to all of the great friends that have supported me during my time in Raleigh, specifically David and Laura Vock, Sarah Hale, and Christopher Krut. Thank you to my amazing girlfriend, Jessalyn Crawford, whose love and support has made the most stressful time of my life also the most enjoyable. Thank you again to my parents to whom I owe everything.

Chapters 2 and 3 were partially supported by NSF grant number DMS-1106570. Chapter 2 has already been published as Suarez & Ghosal [1]. Chapter 3 has already been accepted for publication and will appear in the journal, Bayesian Analysis. The copyright for these chapters is retained by the authors, but full license was given to the publisher. Chapter 5 has already been published as Rodríguez & Suarez [2], and the copyright no longer belongs to the authors, but this chapter is reprinted with the permission of the publisher. Chapter 6 has been submitted for publication, and although the copyright currently belongs to the authors, it is anticipated that the copyright will be transferred to the publisher.
# TABLE OF CONTENTS

## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td></td>
</tr>
</tbody>
</table>

## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>viii</td>
<td></td>
</tr>
</tbody>
</table>

## Chapter 1  Bayesian Methods for Exploratory Functional Data Analysis  1

### 2.1 Introduction  3

### 2.2 The Model  5

### 2.3 Prior Distributions  7

### 2.4 The Similarity Matrix and Clustering  9

### 2.5 Interpretation of Prior Characteristics  10

### 2.6 Convergence Results  11

### 2.7 Computation  13

### 2.8 Applications to Data  14

#### 2.8.1 EEG Data  15

#### 2.8.2 Canadian Weather Data  16

### 2.9 Discussion  17

### 2.10 Proofs  30

## Chapter 2  Bayesian Clustering of Functional Data Using Local Features  3

### 2.1 Introduction  3

### 2.2 The Model  5

### 2.3 Prior Distributions  7

### 2.4 The Similarity Matrix and Clustering  9

### 2.5 Interpretation of Prior Characteristics  10

### 2.6 Convergence Results  11

### 2.7 Computation  13

### 2.8 Applications to Data  14

#### 2.8.1 EEG Data  15

#### 2.8.2 Canadian Weather Data  16

### 2.9 Discussion  17

### 2.10 Proofs  30

## Chapter 3  Bayesian Estimation of Principal Components for Functional Data  37

### 3.1 Introduction  37

### 3.2 Background  38

### 3.3 Model, Prior Specification, and Posterior Computation  40

#### 3.3.1 Model and Priors  41

#### 3.3.2 Low Rank Model  42

#### 3.3.3 Model Choice  43

#### 3.3.4 Choice of Hyperparameters  43

#### 3.3.5 Posterior Computation  44

#### 3.3.6 Alternative Posterior Computation  45

### 3.4 Asymptotic Results  47

### 3.5 Simulation Study  48

#### 3.5.1 Description of FACE  48

#### 3.5.2 Results  49

### 3.6 Canadian Weather Data  51

### 3.7 Discussion  54

### 3.8 Proofs  60

## Chapter 4  Solvability of Nonlinear Equations  64
Chapter 5  On Nonlinear Perturbations of Sturm-Liouville Problems in Discrete and Continuous Settings 66

5.1 Introduction ................................................................. 66
5.2 Generalities ............................................................... 67
5.3 Differential Equations .................................................. 69
5.4 Difference Equations .................................................... 79

Chapter 6  Existence of Solutions to Nonlinear Boundary Value Problems 85

6.1 Introduction ................................................................. 85
6.2 General Strategy .......................................................... 86
6.3 Preliminaries .............................................................. 89
6.4 Nonlinear Systems ....................................................... 89
6.5 Comparison to Previous Results ...................................... 93
6.6 Example ................................................................. 95
6.7 Discussion ............................................................... 96

BIBLIOGRAPHY ................................................................. 98
Table 3.1  **Simulation Study**: Comparison of MSEs. Rows in dark red text are experimental conditions where our method outperformed FACE. In all of these cases the difference is significant in a frequentist sense with the null hypothesis that there is equal probability of either method winning a trial. . . . . . . . . . . . 54
LIST OF FIGURES

Figure 2.1  **EEG Data**: Plot of the proportion of pairs of data points with distance less than a given value. 20

Figure 2.2  **EEG Data** - Top: Dendrogram generated using the dissimilarity matrix by Ward’s method. The stars on the margins represent observations from the fifth group (suspected seizure activity). Bottom: Groups formed when the dendrogram is cut to yield 2 groups. Dashed lines are pointwise posterior means for the observations, and solid blues lines are the pointwise group average. These figures correspond to \( M = 20 \). 21

Figure 2.3  **EEG Data**: Adjusted Rand index comparison between 3 hyperparameter choices and a default method. The clustering was evaluated compared to the “true” clustering defined by whether the observation was from known seizure activity. 22

Figure 2.4  **EEG Data**: Clustering the EEG data by cutting the tree at 12 groups. This corresponds to the last cut point that still maintains a relatively high adjusted Rand index. 23

Figure 2.5  **Weather Data** - Top: Dendrogram created from the dissimilarity matrix by Ward’s method. Bottom: Map of the cities coded by symbol in color to represent the groups formed when the dendrogram is cut to yield 6 groups. These plots correspond to \( M = 10 \). 24

Figure 2.6  **Weather Data** - Dendrogram and heatmap created from the dissimilarity matrix by Ward’s method. This plot corresponds to \( M = 10 \). 25

Figure 2.7  **Weather Data**: Boxplots of the number of distinct points in a given MCMC step for a range of coefficients and for different values of the mass parameter, \( M \). The coefficient numbers were chosen arbitrarily to show the clustering at different resolution levels. 26

Figure 2.8  **Weather Data**: Clustering formed by cutting at 6 groups for \( M = 1 \). For each group, the observed functional data are plotted on the same axes. 27

Figure 2.9  **Weather Data**: Clustering formed by cutting at 6 groups for \( M = 10 \). For each group, the observed functional data are plotted on the same axes. 28

Figure 2.10  **Weather Data**: Clustering formed by cutting at 6 groups for \( M = 20 \). For each group, the observed functional data are plotted on the same axes. 29

Figure 3.1  **Simulation Study**: Boxplots for the simulation results corresponding to \( \kappa_1 \) used for both the truth and the prior. Results are shown in pairs, with the left box (red color) representing the proposed Bayesian procedure, and the right (black) representing the FACE procedure. The value above each pair represents the proportion of data sets in which the Bayesian procedure performed superiorly. 50
Figure 3.2 **Simulation Study**: Boxplots for the simulation results corresponding to $\kappa_1$ used for the truth and $\kappa_2$ used for the prior. Results are shown in pairs, with the left box (red color) representing the proposed Bayesian procedure, and the right (black) representing the FACE procedure. The value above each pair represents the proportion of data sets in which the Bayesian procedure performed superiorly.

Figure 3.3 **Simulation Study**: Boxplots for the simulation results corresponding to $\kappa_2$ used for both the truth and the prior. Results are shown in pairs, with the left box (red color) representing the proposed Bayesian procedure, and the right (black) representing the FACE procedure. The value above each pair represents the proportion of data sets in which the Bayesian procedure performed superiorly.

Figure 3.4 **Simulation Study**: Boxplots for the simulation results corresponding to $\kappa_2$ used for the truth and $\kappa_1$ used for the prior. Results are shown in pairs, with the left box (red color) representing the proposed Bayesian procedure, and the right (black) representing the FACE procedure. The value above each pair represents the proportion of data sets in which the Bayesian procedure performed superiorly.

Figure 3.5 **Simulation Study**: Examples of the posterior distribution of models for each of the four experimental conditions. Row 1 corresponds to $\kappa_1$ as the prior, and column 1 corresponds to $\kappa_1$ as the truth. The other row and column correspond to $\kappa_2$.

Figure 3.6 **Canadian Weather Data**: Posterior probabilities for each model, $(J, K)$. Only models $1 \leq K \leq J \leq 25$ had prior mass.

Figure 3.7 **Canadian Weather Data**: Observations (dots) along with pointwise posterior means (lines). The posterior means are the mean of $H_jU_k\beta_{j,k}$ over $(J, K)$.

Figure 3.8 **Canadian Weather Data**: Full posterior mean of the covariance function over all $(J, K)$ pairs.

Figure 3.9 **Canadian Weather Data**: First four FPCAs corresponding to the full posterior mean of the covariance function (Figure 3.8) along with their associated proportion of variance explained.
Chapter 1

Bayesian Methods for Exploratory Functional Data Analysis

Exploratory methods in statistics become increasingly necessary when the dimensionality of the data increases. Because of this, they are a primary focus of multivariate data analysis. With the growing popularity of functional data, many of the methods from multivariate statistics have been adapted for the special structures that functional data present. When applying Bayesian methods to functional data, these particular features lead to important considerations in the specification of prior distributions.

Particular to functional data, there are typically two classes of models: models for perfectly observed functional data, and models for noisy functional data. In this document, we only consider models for functional data observed with error (noisy data). Related to this, the main result of Chapter 2 is an asymptotic result in terms of this error variance converging to 0.

In the next two chapters we consider two Bayesian methods for exploring functional data: hierarchical clustering and principal components analysis. Both of these methods seek to discover structure within the data.

Hierarchical clustering can be viewed as a sequence of strict clusterings (partitions), such that each partition is finer than the previous. It can give more insight into the relationship/similarity between observations than descriptions or estimates of distances. In fact, many hierarchical clustering techniques take as an input a matrix of pairwise similarities (or dissimilarities). In Chapter 2, our approach is to estimate a similarity matrix within a Bayesian framework, and
then use a deterministic hierarchical clustering technique with this matrix as its input. Particular to functional data, we model using a wavelet decomposition of the individual functions. This directly influences the prior choice since wavelet decompositions are known to be sparse in many situations. The intricacies of this model are discussed in detail in the next chapter.

Principal components analysis (PCA) attempts to find the major sources (directions) of variation present in the data. The result of the analysis is a set of orthogonal directions, along with estimates of the proportion of variation in the data along the corresponding direction. In an abstract sense, PCA simply rotates the natural axes of the data and interprets the new directions. Functional data differs from typical multivariate data because there are a priori beliefs about the relatedness of the components of the individual observations; because the functions are assumed to have a certain amount of smoothness, their evaluations at two points close to each other are subjectively correlated. This leads to special considerations for the prior distributions on the covariance function. PCA for functional data is discussed in Chapter 3.
Bayesian Clustering of Functional Data Using Local Features

2.1 Introduction

Exploratory analysis of new data is an important first step to understanding many scientific questions. Cluster analysis is a popular tool in exploratory analysis to try to discover underlying group structure present in the data. The idea behind cluster analysis is to define sets of data points which are similar within groups and dissimilar between groups. The main question that is raised is what concept to use to define “similar”. Many clustering techniques (especially hierarchical clustering) can be implemented based solely on a matrix of pairwise similarities (equivalently, dissimilarities). One obvious choice for a notion of similarity is that of distance, which is nearly always available since data are most commonly assumed to be elements of some metric space. However, distances are not the only possible choices, and their usage may miss some important qualitative features. A similarity index can be chosen to be any function of two arguments, as long as it represents our qualitative view of what makes two observations similar.

The interest of this chapter lies in functional data, where each subject under study gives rise to a noisy function observation. The clustering of functional data has applications in many scientific fields, such as clustering gene expression time series. Recently, functional data have received a lot of attention. In terms of clustering functional data, most of the work has been done from a frequentist perspective. One approach is to adapt clustering methods from multivariate analysis to
2.1. INTRODUCTION

CHAPTER 2. CLUSTERING FUNCTIONAL DATA

functional data. Tarpey & Kinateder [3] generalized the $k$-means clustering to the functional data setting, and proved results analogous to those of multivariate $k$-means. James & Sugar [4] used a random effects model to cluster sparsely sampled functional data, and dealt with the situation where the data are not observed on the same fixed time grid. One extension we do not pursue is heteroscedasticity, which was studied in Serban [5]. Functional data can often lead to viewing multiple time series in a new light. For a review of clustering of time series see Liao [6].

We shall not choose to define similarity in terms of distance; instead, we define a similarity function, for functions with a finite wavelet basis expansion, that strongly encourages exact matches of coefficients, meaning that the corresponding observations come from the same component of the population. Since the observed functional data include error, we use a model-based approach to estimate the true functions and use the posterior distribution of the basis coefficients to compute an estimated true similarity index. In a Bayesian setting, Ray & Mallick [7] used a truncated wavelet basis expansion and a Dirichlet process prior on their unknown joint distribution. Crandell & Dunson [8] extended this model to species sampling model priors, and also allowed the basis to be unknown. Both of these approaches cluster curves based on all of their basis coefficients jointly. By a well-known clustering property of the Dirichlet process, this implies the prior belief that some of the underlying functions have all wavelet coefficients identical, and hence, the functions themselves are identical. Often, this would not be an acceptable assumption, and this is one aspect in which the current chapter differs from most previous work. Petrone et al. [9] also dealt with this problem using "canonical curves," from which pieces of the observed functions are drawn. For example, Figure 2.1 shows, for the EEG data discussed later, how very few pairs of data points are within a distance that would be a reasonable estimate of the error standard deviation. If the procedure of Ray & Mallick [7] were used on this data, very few observations would have positive posterior probability of sharing underlying functions (see Section 2.8).

We assign priors on the wavelet coefficients independently, where each individual coefficient gets a Dirichlet process prior distribution. The Dirichlet process allows for exact coefficient matches between functions while allowing for new values to arise also. The strength of the model is in the Bayesian approach, where the underlying coefficients across subjects are seen as exchangeable but correlated, and hence, allow for shared learning among them.

In this chapter, we present a method for quantifying similarity of functional data that can be used in a hierarchical clustering scheme. Wavelet coefficient parameters are clustered separately, and we define a function that quantifies our preference for exact matching of coefficients. We
2.2 THE MODEL

present theoretical results relating to the interpretation of the center measure of our prior, and, additionally, provide asymptotic justification of the clustering performance of our method by analyzing the small variance performance. We then demonstrate the method’s use on real datasets, and show competitive performance on a dataset with a known true clustering.

2.2 The Model

There are two common models, closely related in the asymptotic sense, that can be used to describe functional data. First let \( \{ \phi_k : k \in \mathbb{Z} \} \cup \{ \psi_{jk} : j \in \mathbb{N}, k \in \mathbb{Z} \} \) be a given wavelet basis in the multiresolution framework. In particular, we consider the space \( L_2[0,1] \) and the family called wavelets on the interval [10]. The first model can be viewed as a problem of measurement error, where there is a true function, \( f_i \in L_2([0,1]) \), but when we measure it at a point \( t_j \in [0,1] \), we only see a noisy version, so that

\[
Y_i(t_j) = f_i(t_j) + \epsilon_{ij},
\]

where \( \epsilon_{ij} \) is normally distributed with mean 0 and variance \( \sigma^2 \), and independent across \( i \) and \( j \).

We shall assume that all functions are observed on the same fixed time grid, and that the total number of time points is a power of 2, \( n = 2^m \); this is done mainly for computational convenience since we will employ the discrete wavelet transform (DWT). Let \( Y_i = (Y_i(t_1), \ldots, Y_i(t_n))' \), \( f_i = (f_i(t_1), \ldots, f_i(t_n))' \), and \( \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in})' \). If we let \( W \) denote the \( n \times n \) orthogonal matrix corresponding to the DWT for a certain wavelet family, then our model can be transformed to

\[
W Y_i = W f_i + W \epsilon_i.
\]

One important property of the multivariate normal distribution is its rotational invariance, which implies that \( W \epsilon_i \overset{d}{=} \epsilon_i \), meaning equal in distribution. Throughout, we shall use \( \phi_\sigma \) to represent the Lebesgue density of the normal distribution with mean zero and variance \( \sigma^2 \).

The second model is the so-called Gaussian white noise model, given by

\[
d Y_i(t) = f_i(t) \, dt + \sigma \, dB_i(t),
\]
2.2. THE MODEL

CHAPTER 2. CLUSTERING FUNCTIONAL DATA

where $B_i(\cdot)$ are independent Wiener processes (Brownian motions) on $[0, 1]$. This corresponds to ideal observations of continuously sampled functions. Now let

$$a_k^{(i)} = \int_0^1 \phi_k(t) dY_i(t),$$

$$b_{jk}^{(i)} = \int_0^1 \psi_{jk}(t) dY_i(t),$$

$$\bar{e}_k^{(i)} = \sigma \int_0^1 \phi_k(t) dB_i(t),$$

$$e_{jk}^{(i)} = \sigma \int_0^1 \psi_{jk}(t) dB_i(t).$$

Due to the properties of stochastic integrals with respect to the Wiener process, all of the $\bar{e}_k^{(i)}$ and $e_{jk}^{(i)}$ are independent and normally distributed with mean 0 and variance $\sigma^2$. The model implied on the wavelet coefficients by (2.3) is then

$$a_k^{(i)} = a_k^{(i)} + \bar{e}_k^{(i)},$$

$$b_{jk}^{(i)} = b_{jk}^{(i)} + e_{jk}^{(i)}. \quad (2.4)$$

For the finite (measurement error) model, our decomposition is for $k = 0, \ldots, 2^j - 1$, and $j = 0, \ldots, m - 1$. Note that $m$ is not a parameter, but the assumed length (in terms of its base 2 logarithm) of each observation vector. In the infinite (random function) model, $j$ can range over the natural numbers. In both models, we only need $k = 0$ for the scaling coefficient, $a_0^{(i)}$. Our procedure focuses mainly on the detail coefficients, $\{\beta_{jk}^{(i)}\}$, so we shall rarely mention the scaling coefficient. On the detail coefficients, for both models, the observed coefficients can be represented as

$$b_{jk}^{(i)} \sim N(\beta_{jk}^{(i)}, \sigma^2). \quad (2.5)$$

In the results section of this chapter we shall also consider the case where the variance decreases to 0, which corresponds to the case of independent and identically distributed (i.i.d.) replications of the entire experiment. Because of this unifying framework, we shall not make much distinction between the finite and infinite model. We shall state results mainly for the infinite model; however, they can be shown to be true for the finite model using minor modifications.
2.3 Prior Distributions

One of the attractive aspects of using a wavelet expansion for modeling is that, for many functions, the coefficients are sparse. This knowledge is easily incorporated in the prior distributions placed on the wavelet coefficients. If the error is normally distributed, a conjugate prior on the coefficients is given by independent normal priors. Since we know that some of the coefficients are identically zero, we can incorporate a point mass at 0 into the prior. Specifically, the prior, for \( i = 1, \ldots, N \),

\[
\beta_{jk}^{(i)} \overset{\text{ind}}{\sim} \pi_j N(0, \tau_{jk}^2) + (1 - \pi_j) \delta_0, \quad (2.6a)
\]

\[
\sigma^2 \sim \text{IG}(a, b), \quad (2.6b)
\]

where \( \delta_0 \) is a point mass at 0 and IG stands for inverse gamma. The first coefficient, \( \alpha_0 \), is known as the scaling coefficient, and is usually modeled differently, with a vague prior. In our case, we simply assume that the observations have been detrended, so that the value of \( \alpha_0 \) is identically 0.

Abramovich et al. [11] showed that under certain conditions on the mother wavelet, choices of the hyperparameters in this model will guarantee that the corresponding random functions almost surely lie in specific Besov spaces (denoted by \( B^{s}_{p,q} \)). In particular,

\[
\tau_j^2 = \nu_1 \sigma^2 2^{-\gamma_1 j}, \quad \pi_j = \min(1, \nu_2 2^{-\gamma_2 j}), \quad (2.7)
\]

for \( j = 0, 1, \ldots, \) and constants \( \nu_1, \nu_2, \gamma_1, \gamma_2, \) provide the desired interpretation. To be specific, if \( \gamma_2 \geq 1 \) and \( \gamma_1 \geq 0 \), then the random function drawn from this prior, \( f_j \), is almost surely an element of \( B^{s}_{p,q} \) if and only if

\[
s + \frac{1}{2} - \frac{\gamma_2}{p} - \frac{\gamma_1}{2} < 0, \quad \text{if } q < \infty, \quad (2.8a)
\]

\[
s + \frac{1}{2} - \frac{\gamma_2}{p} - \frac{\gamma_1}{2} = 0, \quad \text{if } q = \infty. \quad (2.8b)
\]

The main result of Abramovich et al. [11] is given for any fixed value of the scaling coefficient.

Recently, Ray & Mallick [7] extended the prior of Abramovich et al. [11] to the setting of clustering functional data in the finite model case. Their approach was to assign a prior on the sequence of
vectors \( \beta_i = \{ \beta_{jk}^{(i)} \}_{jk} \) in the following manner:

\[
\beta_1, \ldots, \beta_N \overset{iid}{\sim} F, \quad F \sim \text{DP}(M, G_0),
\] (2.9)

where \( G_0 \) was the product of the priors from Abramovich et al. [11] over \( j \) and \( k \), and \( \text{DP}(M, G_0) \) stands for the Dirichlet process with center measure \( G_0 \) and concentration \( M \). This induces a posterior distribution over partitions of the data, but also implies the prior belief that some true functions are identically equal. When this is not a reasonable assumption, other choices should be made.

Our strategy for placing priors on the true wavelet coefficients is to do so independently for each coefficient using Dirichlet process priors, with center measures corresponding to the usual parametric models often used for modeling by wavelets. Thus, instead of modeling all the coefficients jointly, as in Ray & Mallick [7], they will be done so independently. We also consider \( \sigma^2 \) to be unknown, and for this reason, we scale the variance in the base measure in the traditional manner. The full model is therefore, \( \forall \ j, k, \)

\[
\begin{align*}
\beta_{jk}^{(i)} | \beta_{jk}, \sigma & \sim N(\beta_{jk}, \sigma^2), \\
\beta_{j1}, \ldots, \beta_{jN} | G_{jk} & \sim G_{jk}, \\
G_{jk} & \sim \text{DP}(M, G_{jk}^0), \\
G_{jk}^0 & = \pi_j N(0, \sigma_j^{-2}) + (1 - \pi_j) \delta_0 \\
\sigma^2 & \sim \text{IG}(a, b).
\end{align*}
\] (2.10)

Across levels of \((j, k)\), the random variables, \( \{G_{jk}\} \), are independent. If \( X_1, X_2, \ldots | F \overset{iid}{\sim} F \) where \( F \sim \text{DP}(M, G_0) \), then the predictive distribution of the sequence satisfies:

\[
P(X_{n+1} \in \cdot | X_1, \ldots, X_n) = \frac{M}{M+n} G_0 + \sum_{j=1}^{k(n)} \frac{m_{j,n}}{M+n} \delta_{X_j^*},
\] (2.11)

where \( X_1^*, \ldots, X_{k(n)}^* \) are the \( k(n) \) distinct points in the first \( n \) observations, and \( m_{j,n} = \# \{ i : X_i = X_j^* \} \), for \( j = 1, \ldots, k(n) \). This is the so-called Pólya urn representation of the Dirichlet process [12]. The Dirichlet model is particularly useful when it is known that there is a “typical” cluster with
some smaller “abnormal” groups.

This model tries to capture our belief that the functional data share local features that are expressed in their wavelet expansions. We also want to incorporate the knowledge of the possibility of an exactly zero wavelet coefficient, and we do that within the base measure of the Dirichlet process prior.

Remark 2.3.1. In Section 2.5 we show that all but finitely many coefficients are zero from any realization of the prior. This motivates a different approach to constructing a prior. For all \{j, k\} such that \( j > J^* \), let \((\beta_{jk}^{(1)}, \ldots, \beta_{jk}^{(N)}) = 0\). We then allow \( J^* \) to be random and have a Poisson distribution with parameter \( \lambda \). The decay of \( \pi_j \) and \( \tau_j \) are now not essential, since the resulting wavelet series is always convergent. As before, we have that the number of coefficients and levels are almost surely finite. This allows more freedom in the choice of \( \pi_j \) since the quick decay is no longer needed to give us this property. It is still useful, though, to keep the point mass at zero to account for reasonable prior beliefs about the wavelet expansion. This prior seems to be a much more natural choice, and even yields the later results more easily, but comes at the price of increased computational complexity.

2.4 The Similarity Matrix and Clustering

With the goal of comparing the similarity between functions, we have many choices. Recalling that there are \( n = 2^m \) sampled time points, and excluding one corresponding to the scaling coefficient, we choose to quantify the similarity between two functions using the similarity index

\[
S(i, i') = (2^m - 1)^{-1} \sum_{j=0}^{m-1} \sum_{k=0}^{2^j-1} 1(\beta_{jk}^{(i)} = \beta_{jk}^{(i')})
\]  

(2.12)

the average number of shared wavelet coefficients. This quantity is meaningful in our model since the Dirichlet process will give positive probability to this value being nonzero. The matrix is easily estimated using posterior samples from MCMC output.

Once an estimate is constructed, the obtained matrix can be used with any clustering method taking a (dis)similarity matrix as its input. In fact, given a clustering procedure, our method can be viewed as providing a posterior distribution on dendrograms (for example). Primarily, however,
we shall employ the posterior mean matrix to provide a single output from the chosen clustering algorithm.

2.5 Interpretation of Prior Characteristics

In this section we explore and review some of the properties of the previous model of Abramovich et al. [11] for a single function (the nonparametric regression setting). Instead of studying the model under a fixed value for the hyperparameter, \( \pi_j \), we consider the limiting case where \( \gamma_2 \to 1 \) from above, where \( \pi_j \) is also scaled by a factor. In the following, we let \( \gamma_2 = 1 + \delta \) and consider \( \delta \to 0 \), so we have

\[
\pi_j = \nu_2 \delta 2^{-(1+\delta)j}, \quad \text{where } \delta > 0, \ \delta \to 0.
\]  (2.13)

The reason behind this choice is to approximate the situation where \( \gamma_2 = 1 \) in the original hyperparameter choice, while keeping almost surely finiteness of the number of terms in the wavelet expansion. It is needed that \( \nu_2 \) be scaled by \( \delta \) so that, in the limit, the quantities of interest remain finite, else they would diverge without it to balance the growth.

The following proposition would be useful for prior elicitation in the case where the approach mentioned in Remark 2.3.1 was taken. It motivates and justifies the use of a Poisson prior on the number of nonzero coefficients and resolution levels, and provides an interpretation of their hyperparameters in this setting. For a proof, see Appendix 2.10.

**Proposition 2.5.1.** For the infinite product of the priors specified as above, the following hold:

1. The number of nonzero wavelet coefficients is a.s. finite, and this number converges in distribution to a Poisson random variable with mean \( \nu_2 / \log(2) \) as \( \delta \to 0 \).

2. The number of resolution levels with at least one nonzero coefficient is a.s. finite, and this number converges in distribution as \( \delta \to 0 \) to a Poisson random variable with mean

\[
\lim_{\delta \to 0} \sum_{j=0}^{\infty} \left\{ 1 - \nu_2 \delta 2^{-(1+\delta)j} \right\}^2 < \infty.
\]
2.6 Convergence Results

In the present situation, we first want to study what happens to our similarity matrix as the noise variance, $\sigma^2 \to 0$. This would be the situation where the noisy functional observations are approaching the true underlying functions, respectively. For the purposes of this section, we assume the continuous model of (2.3), with the full specification in terms of the coefficients being given in (2.10a).

The asymptotic regime $\sigma^2 \to 0$ can be understood as equivalent with averaging over $r$ i.i.d. replications of the observed scheme (2.10a)–(2.10d) with $r \to \infty$, thus replacing $\sigma^2$ by $\sigma^2_r = \sigma^2 / r$, with $\sigma^2$ known. Since $\sigma^2$ itself controls the asymptotics in the following, it is essential to treat $\sigma^2$ as given, or equivalently, $\sigma^2$ as known and $r \to \infty$. Although this setting contrasts with the methodology described, this has little effect when only learning about $f$ is the goal. More generally, it is easy to see that the arguments given below go through if $\sigma^2$ is unknown, but has a fixed upper bound. An upper-truncated inverse-gamma prior can still retain the computational conjugacy. If it is desirable to work in full generality without an upper bound for $\sigma^2$, we must fully observe all replications since the sample means are sufficient only when $\sigma^2$ is known. Below we forgo the full setting and treat $\sigma^2$ as known so that it is sufficient to observe the sample mean of $b_{jk}^{(i)}$ over $r$ replications and let $r \to \infty$.

We assume that $\alpha_0^{(i)} = 0$ for all $i = 1, \ldots, N$, and let

$$
\|f\|_2^2 = \sum_{i=1}^{N} \|f_i\|_2^2 = \sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{i-1} \sum_{k=0}^{\infty} |\beta_{jk}^{(i)}|^2,
$$

where $f = (f_1, \ldots, f_N)$. We also consider the Sobolev norm on the product space, defined by

$$
\|f\|_{\mathcal{H}_N^s}^2 = \sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{2js} \|\beta_{j.}^{(i)}\|_2^2.
$$

We shall refer to this space as the $N$-Sobolev space, $\mathcal{H}_N^s$. Note that since $N$ is fixed, we could have chosen to combine the $N$ Sobolev norms using any norm for $\mathbb{R}^N$. The parameter, $s$, relates to the number of weak derivatives possessed by the functions which themselves live in $L_2([0, 1])$. We use $D_r$ to be the set of all observations.

Before proceeding, we need some additional notation. The measure of similarity which holds
our interest is dependent on how we believe the data to be partitioned. We will thus be interested in knowing how our beliefs about the partition structure of the data change as $r \to \infty$. Let $\mathcal{P}$ be the set of all partitions of $\{1, \ldots, N\}$, and let a typical element be denoted by $p = \{A_0, \ldots, A_M\}$. For a given $j, k$, let $\mathcal{P}_{jk} = \{A_{0}^{jk}, A_{1}^{jk}, \ldots, A_{M}^{jk}\}$ be a random partition of $\{1, \ldots, N\}$, which is a function of $\beta_{jk}$ defined in the following way:

\begin{align}
\beta_{jk}^{(a)} = 0 & \iff a \in A_{0}^{jk}, \text{ and} \\
\beta_{jk}^{(a)} = \beta_{jk}^{(b)} \neq 0 & \iff a, b \in A_{i}^{jk} \text{ for some } i \in \{1, \ldots, M_{jk}\}. \tag{2.14}
\end{align}

By our prior specification, it is clear that any partition structure has positive probability a priori. Let $p_0$ represent the “true” partition generated by the true values of the parameters. By a compatible model, we mean a collection of all parameter values corresponding to a single partition which is finer than $p_0$. By an incompatible model, we mean any collection that is not a compatible model.

The following result on consistency of the posterior will be useful for studying the asymptotic properties of clustering. The techniques used in the proofs are both similar to, and certainly inspired by Lian [13]. For proofs of the following, see Appendix 2.10.

**Theorem 2.6.1.** Let $\gamma_1 > 2s + 1$, and assume that the true underlying functions satisfy $f_0 \in \mathcal{H}_{N}^s$. Then the posterior is norm-consistent, i.e., for any $\epsilon > 0$, $\Pi(\|f - f_0\| < \epsilon | D_r) \rightarrow P 1$ as $r \to \infty$.

**Lemma 2.6.1.** Assume that the true vector of functions lies in $\mathcal{H}_{N}^s$, $\gamma_1 > 2s + 1$, and let $p_{jk,0}$ be the true partition of the data for a given coefficient indexed by $j, k$. Then

\[
\Pi(\mathcal{P}_{jk} = p_{jk,0} | D_r) \rightarrow P 1 \text{ as } r \to \infty. \tag{2.16}
\]

Finally we consider neighborhoods of the true full model, that is $p_0 = \{p_{jk,0}\}_{jk}$, in the product topology. Each $p_{jk,0}$ lives in the space, $\mathcal{P}$, of all possible partitions of $\{1, \ldots, N\}$, which is finite and endowed with the discrete topology. Note that the entire model space is uncountable. When considering the product space, a basic neighborhood in the product topology consists of the product of finitely many singleton sets in $\mathcal{P}$ with infinitely many copies of $\mathcal{P}$. Because of this, we easily obtain the following theorem.

**Theorem 2.6.2.** Let $p_0$ be the true model. Then, for any neighborhood in the product topology, $N(p_0)$, we have that $\Pi(N(p_0) | D_r) \rightarrow 1$ in probability as $r \to \infty$. 

12
2.7 Computation

Computation is done using the urn representation of the Dirichlet process prior, and we follow the procedure of Navarrete et al. [14]. The main problem in posterior computation will be the fact that we employ an atomic base measure, and this means that in the urn representation, a value can be 0, either because it is tied to a previous value, or because it was drawn from the base measure. To simplify notation, we focus on updating one particular wavelet coefficient across observations, so we fix \( j, k \) and let \( \beta_i \), for \( i = 1, \ldots, N \), be the parameter for observation \( i \). Due to exchangeability in the Dirichlet process model, it suffices to describe the conditional posterior draws from \( \beta_N | \{ \beta_i \}_{i=1}^{N-1}, \{ b_i \}_{i=1}^{N}, \sigma^2 \) and \( \sigma^2 | \{ \beta_i \}_{i=1}^{N}, \{ b_i \}_{i=1}^{N} \). Similar to Section 2.2, let \( \beta_1^*, \ldots, \beta_{k(N-1)}^* \) be the \( k(N-1) \) unique values among the first \( N-1 \) parameters. Let \( m_{N-1} = \{ m_{1,N-1}, \ldots, m_{k(N-1),N-1} \} \), where \( m_{j,N-1} = \# \{ 1 \leq i \leq N : \beta_i = \beta_j^* \} \). When needed, we shall additionally subscript quantities to designate the \( (j,k) \) level. The Gibbs sampling algorithm executes the following steps:

- Set \( \beta_N \) equal to \( \beta_1^* \) with probability proportional to

\[
\frac{m_{1,N-1}}{M+N-1} \frac{\phi_\sigma(b_1 - \beta_1^*)}{\int \phi_\sigma(b_1 - \beta) dG_0(\beta)} = \frac{m_{1,N-1}}{M+N-1} (2\pi \sigma^2)^{-1/2} e \left\{ -\frac{1}{2\sigma^2} (b_1 - \beta_1^*)^2 \right\}.
\]

- With probability proportional to

\[
\frac{M}{M+N-1} \int \phi_\sigma(b_N - \beta) dG_0(\beta) = \\
\frac{M}{M+N-1} (2\pi \sigma^2)^{-1/2} e^{-1/(2\sigma^2)} \\
\times \left[ 1 - \pi_j + \pi_j \left( \frac{\tau_j^2}{1+\tau_j^2} \right)^{1/2} e \left\{ \frac{b_j^2}{2\sigma^2(1+\tau_j)} \right\} \right],
\]
sample $\beta_N$ from the following distribution
\[
\pi^* \delta_0 + (1 - \pi^*) N \left( \frac{b_N \tau^2_j}{1 + \tau^2_j}, \frac{\sigma^2 \tau^2_j}{1 + \tau^2_j} \right),
\]
where
\[
\pi^* = \left( 1 + \frac{\pi_j}{1 - \pi_j} \left( \frac{1}{1 + \tau^2_j} \right) \right)^{1/2} e^{\left( \frac{\tau^2_j b^2_{ij}}{\sigma^2 (1 + \tau^2_j)} \right)}.
\]

- Sample $\sigma^2$ from the following distribution:
\[
IG \left( a + Nn/2 + \sum_{j,k} k_{ij}(N)/2, b + \frac{1}{2} \sum_{i,j,k} (b_{ij}^{(i)} - \beta_{ijk}^{(i)})^2 + \sum_j \frac{1}{2 \tau^2_j} \sum_k \sum_{l=1}^{k_{ij}(N)} \beta_{ijkl}^{(i)} \right).
\]
- Finally, update any hyperparameters that have been added to the prior specification.

The main issue to notice is that both possibilities for $\beta_N$ can lead to a value of 0 (either being tied to an existing point which happens to be 0, or drawing a 0 from the base measure), and this needs to be taken into account when fitting the model, which simply requires careful bookkeeping. After each draw from these full conditionals, we need to update the unique points, along with $k(N-1)$, and $m_{N-1}$.

### 2.8 Applications to Data

In this section we present the usefulness of the above method by analyzing two different datasets. When presenting results, often it is convenient to display the similarity matrix after it has been used in a deterministic hierarchical clustering scheme. In particular, we apply Ward’s method of clustering [15] to the dissimilarity matrix defined by \( \{1/S^*(i, i')\}_{i, i'=1}^N \). This method is an agglomerative method, in which a single element is joined with an existing group, so that the sum of the variances of all groups is minimized. Other hierarchical clustering methods are also possible.

During the analysis, it was noted that the choice of $\nu_1$ in the prior was both difficult to make a priori and also strongly influenced the results. This situation was addressed by use of a hyperprior
for $\nu_1$ of a conditionally conjugate inverse gamma distribution. This did not cost much in terms of computation, and also provided more robust results.

In both examples, there are rational preconceived notions of how reasonable results should appear. This type of example was chosen to establish confidence that, when used for purely exploratory analysis, we have the potential to find meaningful relationships between observations. All three sets of data also fit well into the model.

Although we explored the use of the method of Ray & Mallick [7] on these data, we do not present the results of that analysis. To get a meaningful number of non-zero entries in the corresponding similarity matrix, or matrix of pairwise probabilities of shared group membership, either the Dirichlet process concentration parameter is required to be nearly zero, or the $a$ priori probability of a zero wavelet coefficient is required to be very large. Since the method of Ray & Mallick [7] is not intended for data for which the belief of identical true functions does not hold, we do not show the comparison in this section.

The method was coded in C and made use of the GNU Scientific Library [16]. It is available at the author's website (https://www.ajsuarez.com), in addition to the supplemental materials.

2.8.1 EEG Data

The first dataset is from a study by Andrzejak et al. [17], which is freely available online (http://epileptologie-bonn.de/cms/front_content.php?idcat=193). The data consist of 500 electroencephalography (EEG) time series, each of length 4096, corresponding to a sampling rate of 173.61Hz. Because of the periodic nature of the data and computational considerations, only the first 128 time points were used. For the MCMC algorithm, 10,000 steps were used, including 1,000 steps for burn-in. On a 3.6GHz AMD Bulldozer-powered desktop computer running single-threaded, this chain took approximately 6.5 hours to run.

The data are combined from 5 separate groups of data, which Andrzejak et al. [17] label as A–E. Sets A and B came from measurements on healthy individuals, while the rest are from patients who suffered from seizures, and who had later been treated with corrective surgery. The important fact about the data is that the observations from set E are all from known seizure activity.

Figure 2.2 shows some graphical representations of the results. The leaves of the dendrogram are marked with stars to denote observations coming from set E. The dendrogram clearly shows a group that seems quite different from the vast majority of the others. In general, this group corresponds
to the known seizure activity. We also display the results of obtaining a non-hierarchical clustering by cutting the dendrogram at a given height level. We chose to form 2 groups, and display the results also in Figure 2.2. Both groups are plotted on the same voltage range. The first group clearly has much lower voltage swings. Large voltage swings are characteristic of seizures [17]. Thus, the method has yielded a very interpretable result consistent with that known from neurobiology.

Although we apply a clustering algorithm to this data set, there is classification information available since we know which subset came from seizure activity. To evaluate the performance of our method, we use the following procedure: first, obtain a hierarchical clustering using the method described above. Subsequently, starting with 2 groups, cut the tree at various levels, and evaluate the strict clustering by computing the adjusted Rand index compared with the “true” clustering. For this comparison, we used 3 different values of $M$, and also compared this method to a default method. The default method was to use the same deterministic portion of our method, i.e. Ward’s method, but with a dissimilarity matrix defined by the Euclidean distance between observation vectors (an estimated $L_2$-distance). The results of this comparison are shown in Figure 2.3. As can be seen from the results, although our method depends on the choice of mass parameter, $M$, for a moderate number of groups, the results are very similar between choices. For two of the choices, $M = 10, 20$, our method outperforms the default method throughout most choices of cut point, except for the smallest number of groups. Since, in practice, many choices of groups are likely to be explored, this gives reason to believe that our method can certainly aid in this exploration.

Using the adjusted Rand index criteria just described, we present clustering corresponding to cutting the tree at 12 groups for the model $M = 20$. This is the point just before the drop-off in the adjusted Rand index seen in Figure 2.3. This clustering is shown in Figure 2.4.

### 2.8.2 Canadian Weather Data

The next analysis involves the very popular Canadian weather data, which was obtained via the fda package within R. These data consist of both average daily temperature and precipitation for 35 Canadian cities. Specifically, we analyzed the precipitation data for the first 256 days of the year. This was done to allow use of the DWT for our method. These data were chosen because of the known spatial and climate correlation for precipitation. As can be seen in Figure 2.5, the method certainly reproduces this correlation. This is clear from the map, and those familiar with Canadian geography can inspect the heatmap more closely. Although not presented in this chapter,
2.9. DISCUSSION

A naive approach based on $L_2$-distance between observations does not nearly show as much spatial clustering. This makes this dataset “harder” than the previous EEG data, in that a naive approach to the EEG data can yield a reasonable, but less clear, description of the data. Again, this example clearly demonstrates the ability of this method to find structure between observations.

Since there is no objectively true clustering for these data, we do not compare to any other methods; however, as in the previous example, we can still analyze the effect that the choice of $M$ has on the results. We focused on three different values, $M = 1, 10, 20$. We present two comparisons of the performance of the methods with these values: first, Figure 2.7 shows the approximate posterior distribution of number of distinct values for a range of coefficients in the model. As would certainly be expected, the number of groups increases as the mass parameter, $M$, increases because this controls the prior probability of an exact coefficient match.

To see how this affects the end result of forming a strict clustering, we show, in Figures 2.8–2.10, the end result of cutting the dendrogram obtained by Ward’s method at 6 groups. It can be seen from these plots that there is a subtle, but noticeable, effect on the results from different choices of $M$. Figure 2.5 shows that they generally correspond to physical proximity between the cities. This could be due to the fact that, although the cities’ climate differs in overall trends, local variations are shared, which is something our method was hoping to emphasize.

2.9 Discussion

There is an important point to note with respect to the ability of this procedure to generalize to priors other than the Dirichlet process used herewith. In the description of the predictive distribution corresponding to the Dirichlet process, (2.11), the form suggests the possibility of a generalization to other so-called species sampling models (SSMs). SSMs are random measures for which, under certain conditions, conditionally i.i.d. sequences have predictive distributions of the same general form as (2.11) [18]. However, as pointed out by the Associate Editor, the Dirichlet process is the only member of the class of SSMs for which (2.11) will be valid when using a base measure with atoms.

Generalization would therefore require some effort in one of two possible directions. First, predictive distributions could be derived for the case of a single atom in the base measure. The other option would be to remove the atom from the base measure and incorporate the probability
of a zero value elsewhere in the hierarchy. This second option would provide an advantage that the distribution of zeroes could be chosen arbitrarily instead of being implied by the choice of SSM (which is beta for the Dirichlet process).

We would like to mention a connection between the structure of the prior and the Indian Buffet Process in the special case of the Dirichlet process prior. Typically the most important aspect of a wavelet coefficient is whether it is zero or non-zero since this respectively indicates the absence or presence of the corresponding term, and thus it is an important indicator of the sparsity of the wavelet expansion. When the quantity of interest is only the indicator that a given coefficient is zero or not, for a given observation, this can be viewed as a binary string. The Section 2.5 shows that the distribution of the total number of ones in a given string is distributed as a Poisson random variable with some rate. Because of the use of the Dirichlet process prior, the successive observations are correlated, and, in particular, exchangeable. For our prior, the probability that the \((n+1)\)th value is 0, given that \(m_0\) out of the first \(n\) were zero, is

\[
\frac{m_0}{M+n} + (1-\pi_j)\frac{M}{M+n}.
\]

(2.17)

With \(m_1 = n - m_0\), the probability that the next value is nonzero is given by \((m_1 + \pi_j M)/(M+n)\).

If every level is considered separately, this probability coincides with that in a two-parameter Indian buffet process defined by Thibaux & Jordan [19].

However, there is a difference. The Indian buffet process is defined for equivalence classes which correspond to rearranging elements of the binary string called the left order. Since it is used mostly in latent feature models, the elements have no inherent meaning, unlike our situation, in which each element corresponds to a particular wavelet coefficient.

The assumption that the functional data are observed on an equally spaced grid of size that is a power of 2 was made to use fast DWT techniques for computation. However, this restriction is not essential for the proposed method. In the case where time points are not equally spaced, or the total of points is not a power of 2, instead of first transforming the data using the DWT, the model can be written as a linear model in which the elements of the design matrix are the point evaluations of the wavelet basis functions at the common grid points. This approach could also be used to handle missing data. When used as a purely exploratory method, since the computation is much quicker in the case of using the DWT, it may be desirable to use a subset of the data rather than increase the computational complexity.
Another extension is to let the hyperparameter $M$ also be given a prior. However, it is known that in Dirichlet mixture models the analysis can be quite sensitive to the choice of prior on $M$, and for exploratory purposes, running the model separately for various values of $M$ could give a deeper understanding of the data than the results from a marginalized (over $M$) analysis. With a nonatomic base measure for the Dirichlet process, the standard augmentation procedure of Escobar & West [20] can be extended to the case of conditionally independent Dirichlet processes with a common concentration parameter $M$. However, in the present case, because of the point mass at zero in the base measure, the conditional posterior distribution of $M$ depends also on $\pi_j$, the size of the mass at 0 in the base measure, so a slight modification of the Escobar & West [20] procedure will be needed. Alternatively, as the conditional posterior density of $M$ is completely explicit except for its normalizing constant and $M$ is only a one-dimensional parameter, standard sampling procedures can be applied.

To summarize, we have presented a model for functional data and priors which allow for the expression of beliefs related to common shared features (basis coefficients). We have proposed a measure of similarity distinct from the usual metrics, and have shown by theory and application that it yields useful results. In the model of Ray & Mallick [7], the similarity between functions can be quantified by calculating the probability that two observations are in the same cluster. However, the approach based on separate modeling of the tying pattern of the wavelet coefficients appears to be more appealing.
Figure 2.1 EEG Data: Plot of the proportion of pairs of data points with distance less than a given value.
Figure 2.2 EEG Data - Top: Dendrogram generated using the dissimilarity matrix by Ward’s method. The stars on the margins represent observations from the fifth group (suspected seizure activity). Bottom: Groups formed when the dendrogram is cut to yield 2 groups. Dashed lines are pointwise posterior means for the observations, and solid blues lines are the pointwise group average. These figures correspond to $M = 20$. 
Figure 2.3 EEG Data: Adjusted Rand index comparison between 3 hyperparameter choices and a default method. The clustering was evaluated compared to the “true” clustering defined by whether the observation was from known seizure activity.
Figure 2.4 EEG Data: Clustering the EEG data by cutting the tree at 12 groups. This corresponds to the last cut point that still maintains a relatively high adjusted Rand index.
Figure 2.5 Weather Data - Top: Dendrogram created from the dissimilarity matrix by Ward’s method. Bottom: Map of the cities coded by symbol in color to represent the groups formed when the dendrogram is cut to yield 6 groups. These plots correspond to $M = 10$. 
Figure 2.6 Weather Data - Dendrogram and heatmap created from the dissimilarity matrix by Ward’s method. This plot corresponds to $M = 10$. 

2.9. DISCUSSION

CHAPTER 2. CLUSTERING FUNCTIONAL DATA
The coefficient numbers were chosen arbitrarily to show the clustering at different resolution levels.
Figure 2.8 Weather Data: Clustering formed by cutting at 6 groups for $M = 1$. For each group, the observed functional data are plotted on the same axes.
Figure 2.9 Weather Data: Clustering formed by cutting at 6 groups for $M = 10$. For each group, the observed functional data are plotted on the same axes.
Figure 2.10 Weather Data: Clustering formed by cutting at 6 groups for $M = 20$. For each group, the observed functional data are plotted on the same axes.
2.10 Proofs

Proof of Proposition 2.5.1. First we show that the expected number of nonzero wavelet coefficients is finite a priori. Let $A_{j,k} = \{\beta_{j,k} \neq 0\}$, $k = 0, 1, \ldots, 2^j - 1$ and $j = 0, 1, \ldots$. Then, as $\delta \to 0$,

$$
\sum_{j=0}^{\infty} \sum_{k=0}^{2j-1} P(A_{j,k}) = \sum_{j=0}^{\infty} \sum_{k=0}^{2j-1} v_2 \delta 2^{-(1+\delta)j} = v_2 \delta \sum_{j=0}^{\infty} 2^j 2^{-(1+\delta)j}
$$

$$
= v_2 \delta \sum_{j=0}^{\infty} 2^{-j\delta} = \frac{v_2 \delta}{1-2^{-\delta}} < \infty,
$$

Thus, for any $\delta > 0$, by the Borel-Cantelli lemma, the number of nonzero wavelet coefficients is almost surely finite. Note that, as $\delta \to 0$, the expression, $v_2 \delta (1-2^{-\delta})^{-1}$, converges to $v_2 (\log 2)^{-1}$ by L'Hôpital’s rule.

Similarly, for the events $B_j = \bigcup_{k=0}^{2j-1} A_{j,k}$, using

$$
P(B_j^c) = \prod_{k=0}^{2j-1} P(A_{j,k}^c) = (1 - v_2 \delta 2^{1-\delta} j)^{2j},
$$

we get that

$$
\sum_{j=0}^{\infty} P(B_j) = \sum_{j=0}^{\infty} \left\{1 - (1 - v_2 \delta 2^{-(1+\delta)j})^{2j}\right\}
$$

$$
\leq \sum_{j=0}^{\infty} 2^j v_2 \delta 2^{-j - \delta j} = v_2 \delta \sum_{j=0}^{\infty} 2^{-\delta j} < \infty \quad (2.18)
$$

so that the number of levels with at least one nonzero coefficient is also almost surely finite.

In order to derive the Poisson limits, we apply Theorem 2 of Le Cam [21]. If $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables with success probabilities $p_1, p_2, \ldots$, respectively, then the total variation distance between the distribution of the sum, $Z_n = \sum_{j=1}^{n} X_j$ and a Poisson random variable is bounded by $\sum_{j=1}^{n} p_j^2$. Specifically if $Q_n$ is the measure on $\mathbb{N}$ induced by $\sum X_j$, and $Q_n^*$ is the Poisson measure with rate $\lambda_n = \sum_{j=1}^{n} p_j$, with $Y_n \sim Q_n^*$. Let $\|Q_n - Q_n^*\| = \sup_{|f| \leq 1} |E_{Q_n} f(Z_n) - E_{Q_n^*} f(Y_n)|$ and $|f| = \sup_{x \in \mathbb{N}} |f(x)|$. Then $\|Q_n - Q_n^*\| \leq \sum_{j=1}^{n} p_j^2$. Therefore, it suffices to bound $\sum_{j,k} \{P(A_{j,k})\}^2$ and $\sum_j \{P(B_j)\}^2$. 

30
Now,
\[ \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \{ P(A_{jk}) \}^2 = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \nu_j^2 \delta^2 2^{-(1+\delta)j} = \nu_j^2 \delta^2 \sum_{j=0}^{\infty} 2^{-(1+2\delta)j} = \frac{\nu_j^2 \delta^2}{1 - 2^{-(1+2\delta)}} \to 0, \]
as \delta \to 0.

Since the priors were specified independently across coefficients, the number of levels for which there is at least one nonzero coefficient (also the number of nonzero coefficients) follows the Poisson-binomial distribution, that is, the distribution of the sum of independent Bernoulli trials, but with varying parameters. Consider the sum of the squared success probabilities
\[ \sum_{j=0}^{\infty} \{ P(B_j) \}^2 = \sum_{j=0}^{\infty} \left\{ 1 - \left( 1 - \nu_j \delta 2^{-(1+\delta)j} \right)^2 \right\}^2 \leq \sum_{j=0}^{\infty} \left\{ 2^j \nu_j \delta 2^{-j-\delta} \right\}^2 = \nu_j^2 \sum_{j=0}^{\infty} \delta^2 2^{-2j\delta} = \frac{\nu_j^2 \delta^2}{2^{2\delta} - 1}. \] (2.19)

Notice that both the numerator and denominator of (2.19) converge to 0 as \( \delta \to 0 \), so by L'Hôpital's rule the limit of the expression in (2.19) is equal to the limit of \( \nu_j^2 \delta^2 (4^\delta \log 4)^{-1} \), which is 0.

**Proof of Theorem 2.6.1.** Since the true vector of functions lies in \( \mathcal{H}_N^s \), it is in a ball of radius \( B \) for sufficiently large \( B > 0 \). Let \( \epsilon > 0 \) and let \( J \) be the smallest integer satisfying both
\[ \sum_{i=1}^{N} \sum_{j>i}^{\infty} \sum_{k=0}^{2^j-1} |\beta^{(i)}_{j,k,0}|^2 < \epsilon^2 / 8. \] (2.20a)

Notice that \( J \) exists and is finite since \( ||f_0|| < \infty \). With this chosen we have, using \( (a + b)^2 \leq \)
2(a^2 + b^2), that

\[
\Pi \left( \sum_{j, i, k} |\beta_{j,k}^{(i)} - \beta_{j,k,0}^{(i)}|^2 < \epsilon^2 / 2 \right) \geq \Pi \left( \sum_{j > i, i, k} |\beta_{j,k}^{(i)}|^2 < \epsilon^2 / 8 \right) \\
\geq 1 - \frac{8}{\epsilon^2} \sum_{j > i, i, k} \mathbb{E}_{\Pi}(|\beta_{j,k}^{(i)}|^2) \\
= 1 - \frac{8}{\epsilon^2} N \nu_1 \nu_2 \frac{2(1-\gamma_2-\gamma_1)}{2(1-\gamma_2-\gamma_1)-1} > 0
\]

since \( \gamma_1, \gamma_2 > 1 \).

Next, notice that

\[
\Pi \left( \sum_{j \leq i, i, k} |\beta_{j,k}^{(i)} - \beta_{j,k,0}^{(i)}|^2 < \epsilon^2 / 2 \right) \geq \eta \prod_{i=1}^{N} \Pi \left( \sum_{j \leq i, k} |\beta_{j,k}^{(i)} - \beta_{j,k,0}^{(i)}|^2 < \epsilon^2 / 2N \right),
\]

where \( \eta > 0 \) is a constant representing the probability that for each \( k = 0, \ldots, 2^l - 1, j = 0, \ldots, J - 1 \), the prior makes all \( \beta_{j,k}^{(i)}, i = 1, \ldots, N \), distinct. Then,

\[
\eta \geq \pi^{2^l-1} \prod_{j \leq l, k} \prod_{i=1}^{N-1} \frac{M}{M+i}.
\]

This expression represents the probability that a new unique value must be drawn from the base measure and also must not be assigned to 0, which ensures unique values (it is a lower bound since it does not include the probability of having a single zero value). Now by the positivity of the prior on each \( \beta_{j,k}^{(i)} \), we have

\[
\eta \prod_{i=1}^{N} \Pi \left( \sum_{j \leq i, k} |\beta_{j,k}^{(i)} - \beta_{j,k,0}^{(i)}|^2 < \epsilon^2 / 2N \right) > 0.
\]

A quantitative estimate of the probability is given in Lian [13, Theorem 1], but we do not need that
2.10. PROOFS

Combining these results we have that for \( \varepsilon > 0 \),

\[
\Pi \left( \sum_{j, l, k} |\beta_{j k}^{(i)} - \beta_{j k, 0}^{(i)}|^2 < \varepsilon^2 \right) > 0,
\]

which shows the positivity of any Kullback-Leibler neighborhood since we have a Gaussian likelihood and the Kullback-Leibler divergence between two normal distributions, \( N(\mu_1, \sigma^2) \) and \( N(\mu_2, \sigma^2) \), is given by \( (\mu_1 - \mu_2)^2 / 2\sigma^2 \). Similarly, the Hellinger distance is equivalent to the \( L_2 \)-distance.

Together with the fact that \( H_{s, N}(B) \) has finite metric entropy by Belitser & Ghosal [22], we have that, using Theorem 2 from Ghosal et al. [23], for any given \( B > 0 \),

\[
\Pi \left( f \in H_{s, N}^i(B) : ||f - f_0|| > \varepsilon |D_r| \right) \rightarrow 0 \text{ in probability.}
\]

Now, it suffices to show that \( \lim_{B \rightarrow \infty} \sup_{r > 0} E_0 \Pi(H_{s, N}^i(B)^c | D_r) = 0 \). By Markov’s inequality and the monotone convergence theorem,

\[
\Pi(H_{s, N}^i(B)^c | D_r) \leq B^{-2} \sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{2js} \sum_{k=0}^{2j-1} E \left[ |\beta_{j k}^{(i)}|^2 | D_r \right]. \tag{2.21}
\]

We bound the needed expectations as

\[
E \left[ |\beta_{j k}^{(i)}|^2 | D_r \right] = \sum_{p \in \Psi} E \left[ |\beta_{j k}^{(i)}|^2 | \mathcal{P}_{j k} = p, D_r \right] \Pi(\mathcal{P}_{j k} = p | D_r)
\]

\[ \leq \max_{p \in \Psi} E \left[ |\beta_{j k}^{(i)}|^2 | \mathcal{P}_{j k} = p, D_r \right]. \]

Now, given \( \mathcal{P}_{j k} = p \), \( i \in A_{l}^{j k} \) for some \( l \in \{0, 1, \ldots, M_{j k} \} \) (for definition, see (2.14)), and the posterior is of conjugate form, so that the expectation is

\[
E \left[ |\beta_{j k}^{(i)}|^2 | \mathcal{P}_{j k} = p, D_r \right] = \frac{\tau_j^2 \sigma^2}{1 + \tau_j^2 r (\#A_l)} + \left( \frac{\tau_j^2}{1/r + \tau_j^2 (\#A_l)} \right)^2 \left( \sum_{m \in A_l} b_{j k}^{(m)} \right)^2.
\]

\[ \leq \frac{\tau_j^2 \sigma^2}{1 + \tau_j^2 r} + \left( \frac{\tau_j^2}{1/r + \tau_j^2} \right)^2 \left( \sum_{m=1}^{N} b_{j k}^{(m)} \right)^2, \]
where \( \#A \) is the cardinality of the set \( A \). We now bound the true expectation of this term by

\[
\tau_j^2 \sigma^2 + \left( \frac{\tau_j^2}{1/r + \tau_j^2} \right)^2 \left[ \frac{N \sigma^2}{r} + \left( \sum_{m=1}^{N} \beta_{jk,0}^{(m)} \right)^2 \right] \leq \tau_j^2 \sigma^2 + \left( \frac{r N \sigma^2}{(\tau_j^2 + r)^2} + \left( \sum_{m=1}^{N} |\beta_{jk,0}^{(m)}|^2 \right) \right) \\
\leq \tau_j^2 \sigma^2 + \left( \frac{N \sigma^2}{(\tau_j^2 + 1)^2} + \left( \sum_{m=1}^{N} |\beta_{jk,0}^{(m)}|^2 \right) \right).
\]

Now observe that under the assumption that \( \gamma_1 > 2s + 1 \), we have

\[
\sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{2js} \sum_{k=0}^{2^j-1} \tau_j^2 = N \nu_1 \sum_{j=0}^{\infty} 2^{j(2s+1-\gamma_1)} < \infty,
\]

\[
\sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{2js} \sum_{k=0}^{2^j-1} \frac{N \sigma^2}{(\tau_j^2 + 1)^2} = N \nu_1 \nu_0 \sum_{j=0}^{\infty} 2^{j(2s+1-\gamma_1)} < \infty,
\]

\[
\sum_{i=1}^{N} \sum_{j=0}^{\infty} 2^{2js} \sum_{k=0}^{2^j-1} \sum_{m=1}^{N} |\beta_{jk,0}^{(m)}|^2 = N ||f_0||_{X_N'}^2 < \infty.
\]

Hence, the right hand side of (2.21) tends to zero as \( r \to \infty \), completing the proof. \( \square \)

**Proof of Lemma 2.6.1.** Theorem 2.6.1 implies that for any neighborhood of the true value, \( \beta_{jk,0} \), its posterior probability tends to 1. By projection onto the coordinates, the posterior probability of \( \beta_{jk}^{(i)} \) lying in any neighborhood of \( \beta_{jk,0}^{(i)} \) also tends to 1.

Let \( \mathcal{P}_{jk} = p \) be an incompatible model, as defined in Section 3.4. Then there is an incorrect assignment for some \( i \), that is, either \( \beta_{jk}^{(i)} = 0 \) when \( \beta_{jk,0}^{(i)} \neq 0 \), or \( \beta_{jk}^{(i)} = \beta_{jk}^{(l)} \) for some \( l \) when this is not the case. For instance, consider the situation that the partition assigns \( \beta_{jk}^{(i)} = 0 \) when \( \beta_{jk,0}^{(i)} \neq 0 \). The inequality implies that there is a positive distance between the two points so that there is a neighborhood of \( \beta_{jk,0}^{(i)} \) which does not contain 0. The probability of the complement of this neighborhood must tend to 0 as \( r \to \infty \), and since the posterior probability of \( \mathcal{P}_{jk} = p \) must be less than or equal to this value, its posterior probability must also tend to 0.

Thus, we can focus our attention only on compatible models. Now let \( \mathcal{P}_{jk} = p = \{A_0, A_1, \ldots, A_m\} \) be a compatible model (we drop the \( j, k \) notation for the sets in the partition for this section).
Letting $H$ be the distribution function for the inverse-gamma prior, we then have that

$$f(b_{jk}|\mathcal{P}_{jk} = p) = \left(\prod_{i \in A_0} \phi_{\sigma}(b_i)\right) \prod_{i=1}^{m} \left[ \int \left( \prod_{i \in A_i} \phi_{\sigma}(b_i - \beta) \right) \phi_{\tau_j}(\beta) d\beta \right]$$

$$= \left(\frac{2\pi}{r}\right)^{-N/2} \left(\frac{\tau_j^2}{r}\right)^{-m/2} \left[ \prod_{i=1}^{m} \left( r c_i + \frac{1}{\tau_j} \right) \right]^{-1/2} (\sigma^2)^{(a+b-\gamma-1)}$$

$$\times e^{\left[ -\frac{1}{2} \left( b + \frac{\tau_j}{2} \right) \sum_{i=1}^{m} b_i^2 - \sum_{i=1}^{m} \left( r c_i + \frac{1}{\tau_j} \right) \left( \sum_{i \in A_i} b_i \right)^2 \right]}.$$ 

where we have dropped the $j,k$ subscripts, and let $c_i = #A_i$ and $b_i = b_{ij}^{(l)}$.

Now let $\mathcal{P}_{jk,0} = p_{jk,0} = \{A_0^0, A_1^0, \ldots, A_{m_0}^0\}$ be the true model. Since $\mathcal{P}_{jk} = p$ is compatible (finer), assume that $m > m_0$, and that $A_i \subset A_i^0$ for $i = 0, \ldots, m_0$. Now,

$$\frac{f(b_{jk}|\mathcal{P}_{jk} = p)}{f(b_{jk}|\mathcal{P}_{jk} = p_0)}$$

$$= \left(\frac{\tau_j^2}{r}\right)^{m_0-m} \left[ \prod_{i=1}^{m} \left( r c_i + \frac{1}{\tau_j} \right) \right]^{-m_0/m} \left[ \prod_{i=1}^{m} \left( r c_i + \frac{1}{\tau_j} \right) \right]^{-1/2}$$

$$\times e^{\left[ -\frac{1}{2} \left( \sum_{i=1}^{m} \frac{\tau_j^2}{r \sum_{i=1}^{m} c_i} b_i \right) - \sum_{i=1}^{m_0} \left( \frac{\tau_j^2}{r \sum_{i=1}^{m} c_i} \right) \left( \sum_{i \in A_i} b_i \right)^2 \right]}.$$ 

To examine the behavior of this expression note that $m > m_0$, and $c_i < c_i^0$ for $i = 1, \ldots, m_0$. Now, the expression under the square root converges to 0 as $r \to \infty$, since $m > m_0$ implies that the denominator is of a higher order.

Consider the term inside the square brackets of the exponential:

$$\sum_{i=1}^{m} \left( \frac{\tau_j^2}{1/r + c_i \tau_j^2} \right) \left( \sum_{i \in A_i} b_i \right)^2 - \sum_{i=1}^{m_0} \left( \frac{\tau_j^2}{1/r + c_i^0 \tau_j^2} \right) \left( \sum_{i \in A_i^0} b_i \right)^2.$$ 

It is clear that this expression can be written as a quadratic form, $b^t A b$, in $b$, for some matrix, $A$. It is clear that $A$ is $O_{r}(1)$ as $r \to \infty$. The expectation is then $O_r(1)$, and, since the dispersion matrix of $b$ is $\frac{\sigma^2}{r} I_N$, the variance of $b^t A b$ is $O_r(1/r)$ as $r \to \infty$. Thus, the exponential term is
bounded in probability as $r \to \infty$.

Thus, we have shown that the probability of any incompatible model goes to zero. Along with the fact that for any compatible model the marginal likelihood ratio tends to 0 implies that the only model that can possibly retain positive probability is the truth. Since, for any fixed $j, k$, there are only finitely many models for $\{\beta_{j,k}^{(i)}\}_{i=1}^N$, the probability of the true model must tend to 1. $\square$
Chapter 3

Bayesian Estimation of Principal Components for Functional Data

3.1 Introduction

In the rapidly expanding area of functional data analysis, data compression has become an oft-employed strategy. Principal component analysis (PCA) has become a widespread tool in the area of functional data, where the high dimensionality of the data can quickly become unmanageable. Principal components can be used to reconstruct a process approximately, using relatively few random variables. At its heart, PCA is an exploratory tool used to gain insight into the structure of the data. It is also used in less scrupulous endeavors, such as preprocessing for a regression analysis. For a textbook length treatment of classical multivariate PCA, see Jolliffe [24].

PCA for functional data has likewise become a very popular technique. Ramsay & Silverman [25] certainly helped make functional PCA (FPCA) a standard first step when dealing with functional data. For another textbook account of FPCA, see Horváth & Kokoszka [26].

Bayesian methods for multivariate PCA have been relatively absent from the literature. Tipping & Bishop [27] showed how the traditional method of PCA can be viewed as the solution to a maximum likelihood procedure; this likelihood was then used for a Bayesian treatment in Bishop [28]. For functional data, Behseta et al. [29] proposed a Bayesian method for FPCA, and Linde [30] proposed an approximate Bayesian method using variational calculations. In non-Bayesian methods, PCA is commonly used as the first step in multi-step procedures; one reason for the lack of the
subjective perspective in the PCA literature is certainly due to the fact that Bayesian procedures are, by their nature, not performed stepwise. The extension of Bayesian procedures to more complicated situations usually comes from hierarchically integrating the simpler model into a larger one.

In this chapter, we investigate a potential model for the covariance structure of functional data observed with noise. The model jointly smooths the observations and estimates the principal components. As a Bayesian procedure, model selection on the chosen number of basis functions is conceptually straightforward, and we demonstrate this on a real data set.

Section 3.2 provides background material. Section 3.3 describes the motivation for the model used for the data, along with the priors used throughout. Section 3.3 also discusses some issues in choosing hyperparameters for the typical priors. Section 3.3.6 describes the method used for model comparison in terms of the number of basis functions used for approximation. In Section 3.4, we prove a convergence theorem for the case of an inverse Wishart prior. Sections 3.5 and 3.6 present a simulation study and applied data example, respectively.

### 3.2 Background

Let \( \{X(t) : t \in [0,1]\} \) be a stochastic process such that the sample paths of \( X(t) \) are square integrable. By adopting a change of location and scale, if necessary, any bounded interval can be reduced to \([0,1]\), which we shall abbreviate by \( I \). Usually the index represents time, and for functional observations the boundedness of the domain is most natural because data can be observed only over a limited time. Let \( \mu(t) := \mathbb{E}[X(t)] \) for all \( t \in I \), and let the covariance function of the process be given by \( \kappa(s, t) = \text{Cov}(X(s), X(t)) \) for all \( s, t \in I \). Note that \( \kappa \) is symmetric in its arguments and is a positive definite function, i.e., for any \( t_1, \ldots, t_k \in I \), the matrix \( (\kappa(t_i, t_j)) \) is positive definite. We assume that \( \kappa \) is continuous on \( I \times I \) and let \( M := \sup\{|\kappa(s, t)| : s, t \in I\} < \infty \). On the space \( L_2 = L_2(I) \) of square integrable functions, we use the standard inner product \( \langle f, g \rangle = \int_I f(t)g(t)dt \), and the norm which this implies.

The covariance function then defines an integral operator, \( T_\kappa : L_2 \to L_2 \), given by

\[
T_\kappa f = \int_I \kappa(\cdot, t)f(t)dt,
\]

for all \( f \in L_2 \). It is well known that \( T_\kappa \) is a Hilbert-Schmidt operator, and, in particular, \( T_\kappa \) is a

\[38\]
compact linear operator, which necessarily has a countable spectrum \( \{ \lambda_1, \lambda_2, \ldots \} \) with 0 as the only accumulation point [31]. Therefore, the important quantities for us are the eigenfunctions and eigenvalues of the covariance operator, i.e., functions \( \{ \phi_1, \phi_2, \ldots \} \) and nonnegative real numbers \( \{ \lambda_1, \lambda_2, \ldots \} \) (assumed to be in nonincreasing order) such that

\[
T_k \phi_i = \lambda_i \phi_i. \tag{3.2}
\]

Mercer’s theorem [31, Theorem 2.4] states that the covariance function can be represented, for \( s, t \in I \), as

\[
\kappa(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(t), \tag{3.3}
\]

where the convergence is uniform on \( I \times I \).

The eigenfunctions, \( \{ \phi_1, \phi_2, \ldots \} \) are called the principal components for reasons we will now discuss. The map \( f \mapsto \text{Var}(f, X) = \text{Var} \left( \int_I f(t)X(t) \, dt \right) \) obtains its maximum on the unit sphere at \( \phi_1 \). The set \( \{ \phi_1, \phi_2, \ldots \} \) satisfies

\[
\phi_i = \arg \max_{f \in L_2} \{ \text{Var}(f, X) : \|f\| = 1, \langle f, \phi_k \rangle = 0 \text{ for } k = 1, \ldots, i-1 \}. \tag{3.4}
\]

In this way, the eigenfunctions are the principal directions of variation for the process. The other reason for the term, “principal components,” is due to the Karhunen-Loève expansion theorem, which states that the process, \( X \), as a random element of \( L_2 \), can be represented as

\[
X(t) = \mu(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} Z_i \phi_i(t), \tag{3.5}
\]

where \( \{ Z_1, Z_2, \ldots \} \) are uncorrelated random variables with unit variance and the convergence is in mean square, uniformly on \( I \). In this way, the eigenfunctions can be seen as a way to decompose the process into orthogonal (uncorrelated) components. An important special case to note is that (3.5) holds pointwise on \( I \) almost surely if \( X \) is a Gaussian process, and then each \( Z_i \) is also a Gaussian random variable, and hence they are all independent.
3.3 Model, Prior Specification, and Posterior Computation

Let $X_1, X_2, \ldots$ be independent and identically distributed observations from a Gaussian process, $\text{GP}(\mu, \kappa)$, on $I = [0, 1]$, where $\mu$ is the mean function, and $\kappa$ is the covariance function. We will assume, however, that our observations have been contaminated with additional noise, i.e., we observe noisy data on some grid of points $\{t_1, \ldots, t_T\}$. We assume that all of the data is observed on the same grid for simplicity of formulas. Let $X_i = (X_i(t_1), \ldots, X_i(t_T))'$ be the $i$th underlying discretized function, and let $Y_i = X_i + \epsilon_i$ be the $i$th observation, where $\epsilon_i \sim \text{N}(0, \sigma^2 I_T)$. Our goal will be to estimate $\kappa$ and the principal components it induces from data, and we will do so by placing priors on all the parameters.

To put a prior on $\mu$, we shall use a Gaussian process, which reduces to the multivariate normal distribution on the discretized observations. In putting a prior for $\kappa$, we shall construct our prior based on an approximate spectral representation by truncating the series in (3.3), but allowing a prior on the number of terms to ensure full support. We shall induce a prior on the eigenvalues and eigenfunctions indirectly from that on the covariance matrix on the finite grid of time points, which will be chosen as the inverse Wishart distribution.

Consider a given basis, $\{h_1, h_2, \ldots\}$, for $L_2$. Since any eigenfunction, $\phi_i$, of $T_\kappa$ can be expanded as

$$\phi_i = \sum_{j=1}^{\infty} \alpha_{ij} h_j,$$

a particularly convenient method of putting a prior on $\{\phi_1, \phi_2, \ldots\}$ is by truncating (3.6) at some level, $J \in \mathbb{N}$. Let $h_J = (h_1, \ldots, h_J)'$. We will also truncate the expansion of $\kappa$ in (3.3) at some level, $K \in \mathbb{N}$, putting a prior on the resulting coefficients and also a prior on $J$. Since the finitely truncated series converges to (3.6) as $J \to \infty$, this procedure ensures that the resulting objects, $\{\phi_1, \phi_2, \ldots\}$, get a fully supported prior if the coefficients get such a prior for each value of $J$. Furthermore, a prior on $\kappa$ is induced by truncating (3.3) at level $K$, i.e. $\kappa(s, t) = \sum_{i=1}^{K} \lambda_i \phi_i(s) \phi_i(t)$, and imposing a prior on $K$. Let $A_{KJ} = (\alpha_{ij})$ be the $K \times J$ matrix of coefficients and $\phi_K = (\phi_1, \ldots, \phi_K)'$ be given by

$$\phi_K = A_{KJ} h_J.$$
Let $\Lambda_K = \text{diag}(\lambda_1, \ldots, \lambda_K)$. Then the prior on $\kappa$ can be induced by the relation
\[
\kappa(s, t) = h_J'(s)A'_{KJ}\Lambda_K A_K h_J(t),
\] (3.8)
and priors on $K$ and $J$. However, as mentioned above, instead of directly putting priors on $A_{KJ}$ and $\Lambda_K$, we proceed in the reverse order and induce priors on them through a convenient prior on $\Sigma = A'_{KJ}\Lambda_K A_{KJ}$. Details of the specification are explained below.

### 3.3.1 Model and Priors

In the following, if $K = J$, it leads to substantial simplification, although such a choice may lead to overfitting of the covariance function; it is reasonable to believe that more basis functions would be needed than the number of principal components needed to reconstruct the covariance function. We first describe the model for the case of $K = J$ to then motivate the low rank model ($K < J$). Conditional on $K = J$, the model and the prior distribution can be described by the following hierarchical scheme for $i = 1, \ldots, n$,

\[
Y_i \sim \text{ind} N_J(H_J \beta_{i,J}, \sigma^2 I) 
\]
(3.9a)
\[
\beta_{i,J} \sim \text{iid} N_J(\theta, \Sigma) 
\]
(3.9b)
\[
\theta \sim N_J(\theta_0, \tau \Sigma) 
\]
(3.9c)
\[
\sigma^2 \sim \text{inv-Gamma}(a, b) 
\]
(3.9d)
\[
\Sigma^{-1} \sim \text{Wishart}(v, \Xi^{-1}) 
\]
(3.9e)
\[
J \sim \text{Poisson}(j_0), \text{ truncated to } \{1, \ldots, T - 1\}, 
\]
(3.9f)

where $H_J$ is a $T \times J$ matrix whose columns consist of the basis functions evaluated at all grid points. When convenient, we may drop the subscript from certain expressions.

The functions $h(t)\beta_i, i = 1, \ldots, n$ correspond to the underlying (unobserved) noise-free functional observations. The function $h(t)\theta$ corresponds to the overall population mean for the functional observations, $h(s)\Sigma h(t)$ is the covariance function of interest, and $Ah(t)$ is the vector of functional principal components. When estimating the full model, that is, averaging across the posterior distribution of $K$, it is important to restrict attention to parameters whose dimension does not depend on $K$. For example, the covariance function evaluated at the observed grid points,
3.3. MODEL, PRIOR SPECIFICATION, AND POSTERIOR COMPUTATION
CHAPTER 3. BAYESIAN FPCA

\( \Sigma H'H \), has dimension \( T \times T \), and has meaning across values of \( K \), whereas \( \beta_i \) only has meaning for a given value of \( K \).

3.3.2 Low Rank Model

Because the Wishart distribution gives probability 1 to nonsingular matrices, if we wish to allow the number of principal components, \( K \), to be less than the number of basis functions used in approximations, \( J \), we should allow for singular Wishart matrices. The most straightforward approach is to use a singular center matrix, of rank \( K \), in the prior specification. First, let 

\[ \Xi = U L U' \]

where \( U \) is orthogonal and \( L \) is the diagonal matrix of ordered eigenvalues. Choose 

\[ \Xi_K = U_K L_K (U_K)' \]

where \( U_K \) is the \( J \times K \) matrix formed by the first \( K \) columns of \( U \), and \( L_K \) is the \( K \times K \) matrix formed by first \( K \) rows and \( K \) columns of \( L \). This implies that although \( \Xi_K \) is \( J \times J \), it has rank \( K \). Let \( A^+ \) denote the Moore-Penrose inverse of a matrix \( A \). Then, if 

\[ \Omega_1 \sim \text{Wishart}(\nu, \Xi_K^+) \]

\[ \Omega_2 \sim \text{Wishart}(\nu, L_{-1}K) \]

\[ \Omega_1 \overset{d}{=} U_K \Omega_2 U_K' \]

\[ \Omega_1^+ = U_K \Omega_2^{-1} U_K' \]

This gives the motivation for our full, low-rank model:

\[ Y_i \overset{ind}{\sim} N_T(H_j U_K \beta_{i,K}, \sigma^2 I) \]  \hspace{1cm} (3.10a)
\[ \beta_{i,K} \overset{iid}{\sim} N_K(\theta, \Sigma) \]  \hspace{1cm} (3.10b)
\[ \theta \sim N_K(\theta_0, \tau \Sigma) \]  \hspace{1cm} (3.10c)
\[ \sigma^2 \sim \text{inv-Gamma}(a, b) \]  \hspace{1cm} (3.10d)
\[ \Sigma^{-1} \sim \text{Wishart}(\nu, L_K^{-1}) \]  \hspace{1cm} (3.10e)
\[ J \sim \text{Poisson}(j_0), \text{ truncated to } \{1, \ldots, T-1\}, \]  \hspace{1cm} (3.10f)
\[ K \sim \text{Poisson}(k_0), \text{ truncated to } \{1, \ldots, J\}. \]  \hspace{1cm} (3.10g)

This model implies that \( U_K \Sigma^{-1} U_K' \sim \text{Wishart}(\nu, \Xi_K^+) \) that provides the singularity we desire in our modeling. Because our prior puts mass on all \((J,K)\) pairs with \(1 \leq K \leq J\), we still have a marginal full-rank model, but our posterior will be mixed over rank-deficient models.
3.3.3 Model Choice

There is a vast literature on Bayesian model choice, and many of the possible procedures could be used for our purposes. Reversible jump MCMC [32] has become a widely used method, but presents a major hurdle in our case with the need for a proposal distribution when using the model that includes the parameter $\Sigma$. The main challenge is respecting the positive definiteness of the matrix when proposing a jump to a higher dimension. This can be overcome by marginalizing out $\Sigma$, obtaining posterior samples of the $\beta_i$’s, and, finally, generating samples for $H\Sigma H'$, which would then all live in the same dimension, even though $K$ might change between steps.

The approach we take is to estimate the posterior distribution of $K$ through approximations to the marginal likelihood for each model. We employ independent Gibbs samplers for each value of $K$, and compute estimates of their marginal likelihood using results from Chib [33]. The obvious disadvantage of this strategy is the need to run MCMC chains for each possible model. The likely faster convergence of the Gibbs sampler, however, can partially offset the cost along with the ability to implement the chains in parallel. In the application considered below, the relatively small number of sampled time points lends itself well to this method, and gives much more confidence in the convergence than other approaches not based on Gibbs sampling.

3.3.4 Choice of Hyperparameters

Especially in the case of the inverse Wishart prior, selection of the hyperparameters is an important issue. Because of the link between principal components and covariance estimation, prior elicitation can be done in either domain. However, some problems can arise with what appears to be a good default choice. Specifically, because of the required smoothness conditions on the covariance function, it is not possible to choose the identity matrix as the center matrix for the inverse Wishart prior.

A sensible choice of $\Xi$ (whose size depends on $J$) is given below. Using the prior covariance function, construct the covariance matrix corresponding to the grid being used; call this $\Sigma^*$. We then propose using $\Xi^{-1} = \left( H' H \right)^{-1} H' \Sigma^* H \left( H' H \right)^{-1}$ as the choice of hyperparameter, which can be seen as a least-squares projection. This matrix can be shown to be invertible using the facts that $H$ has full column rank because it is comprised of function evaluation from an orthogonal set of functions, and $\Sigma^*$ is invertible because it is derived from a valid covariance function of a
vector without linear restrictions. To complete the specification of the Wishart prior, for model 
\((J, K)\), a reasonable default choice for the degrees of freedom is \(\nu = 2K\), which, in the Gibbs step described later, implies the mean of the inverse Wishart distribution that resembles a typical covariance estimate with denominator \(n + k - 1\). This choice is implemented in the empirical comparisons below, and performs very competitively.

Finally, for the inverse gamma prior, the choice of \((a, b)\) is very important since it controls the amount of smoothing performed on the data. It corresponds to the prior beliefs on the amount of sampling noise present in the data. It turns out that the choice is sensitive, and, in practice, several values should be tried when performing the analysis. There is an empirical choice available for these hyperparameters. Let \(\hat{\beta}_i = (U_K' H_j' H_j U_K')^{-1} U_K' H_j' Y_i\) be the ordinary least-squares estimate. It is well known that, conditional on \(\sigma^2\) and \(\{\beta_i, K\}_{i=1}^n\), the quantity,

\[
\sigma^{-2} \text{SSE}_i = \sigma^{-2} \|Y_i - H_j U_K \hat{\beta}_i\|^2 \sim \chi^2_{n-K}.
\]

It can be shown that the marginal distribution of \(\text{SSE}_i\), with \(\sigma^2\) integrated out, is proportional to

\[
\frac{b^a}{\Gamma(a)} \left( b + \frac{1}{2} \sum_{i=1}^n \text{SSE}_i \right)^{-\frac{n(n-K)}{2} - n + a + 1} \Gamma \left( \frac{n(n-K)}{2} - n + a + 1 \right).
\]  

Using this, \(a\) and \(b\) can be chosen to be the values which maximize this quantity evaluated at the data.

### 3.3.5 Posterior Computation

Posterior computation is done independently for each pair \((J, K)\), \(1 \leq K \leq J \leq J_{\text{max}}\). The primary advantage is the ability to implement a Gibbs sampler for each model. Since the models are computed separately, they can be run in parallel to offset the computation cost of running all models. Having independent samples for each model also allows for the adjustment of the priors on \(J\) and \(K\) without the need to rerun the MCMC chains. Finally, it provides full information for each individual model in the case that a single model is desired, instead of a fully Bayesian, model-averaged posterior.

For a fixed pair \((J, K)\), the Gibbs sampler following from the low rank model (3.10) is as follows:
• For $i = 1, \ldots, n$, sample $\beta_{i,K}$ from a $K$-dimensional Gaussian distribution with mean

$$\left(\sigma^{-2} U'_K H'_J H J U_K + \Sigma^{-1}\right)^{-1} \left(\sigma^{-1} U'_K H'_J Y_i + \Sigma^{-1} \theta\right),$$

and variance

$$\left(\sigma^{-2} U'_K H'_J H J U_K + \Sigma^{-1}\right)^{-1}.$$

• Sample $\theta$ from a $K$-dimensional Gaussian distribution,

$$N\left(\left(n + \tau^{-1}\right)^{-1} \left(\sum_{i=1}^{n} \beta_{i,K} + \tau^{-1} \theta_0\right), \left(n + \tau^{-1}\right)^{-1} \Sigma\right).$$

• Sample $\Sigma^{-1}$ from a Wishart distribution with degrees of freedom $\nu + n + 1$, and center matrix

$$\left(L_K + \tau^{-1} (\theta - \theta_0)(\theta - \theta_0)' + \sum_{i=1}^{n} (\beta_{i,K} - \theta)(\beta_{i,K} - \theta)'\right)^{-1}.$$

• Finally, sample $\sigma^{-2}$ from a gamma distribution with shape parameter $a + n/2$ and rate parameter

$$b + \sum_{i=1}^{n} \left\|Y_i - H_J U_K \beta_{i,K}\right\|^2$$

3.3.6 Alternative Posterior Computation

Conditional on $(J, K)$, the posterior distribution can be obtained using a Gibbs sampling scheme. When computation is performed within this chapter, this was the implemented approach. Model averaging over $(J, K)$ can be obtained by using independent MCMC chains for each value. We do wish to point out some equivalent models that have potential computational benefits. Marginalizing
out $\beta_i$ gives

\[
X_i \overset{iid}{\sim} N_T(HU\theta, \sigma^2 I + HU\Sigma U'H')
\]

\[
\theta \sim N_K(\theta_0, \tau\Sigma)
\]

\[
\sigma^2 \sim \text{inv-Gamma}(a, b)
\]

\[
\Sigma^{-1} \sim \text{Wishart}(K, L^{-1})
\]

\[
K \sim \text{Poisson}(k_0), \text{ truncated to } \{1, \ldots, T-1\},
\]

for the top level. In this case, we see that, for the covariance to be an identifiable parameter, we require that $K < T$, so that any prior that we use should have probability one of meeting this restriction.

We can also marginalize out $\theta$ and $\Sigma$ to obtain an equivalent prior for the $\beta_i$'s. Specifically, if we let $B_K = [\beta_{1,K}, \beta_{2,K}, \ldots, \beta_{n,K}]$, and $\Theta_0 = [\theta_0, \ldots, \theta_0]$, then the marginal prior on $B_K$ is the so-called matrix t-distribution [34] with density

\[
\left[ \prod_{i=1}^{K} \frac{\Gamma\left(\frac{v+n+1-i}{2}\right)}{\Gamma\left(\frac{v+1-i}{2}\right)} \right] \left( \det \Xi \right)^{v/2} \left( \det \left( I - (n + \tau)^{-1}1 \right) \right)^{k/2} \times \left( \det \left( L_K + (B_K - \Theta_0) \left[ I - (n + \tau)^{-1}1 \right] (B_K - \Theta_0)' \right) \right)^{-(v+n)/2}.
\]

Since we have then lost conjugacy, we can also integrate out $\sigma^2$ to represent the model in terms of $B_K$ only. The joint density of $(Y_1, \ldots, Y_n)$ is proportional to

\[
\left( 1 + \frac{1}{2b} \sum_{i=1}^{n} ||Y_i - H_j U_k \beta_{i,K}||^2 \right)^{-(a+nT/2)}.
\]

This formulation can be useful for implementing a reversible jump MCMC scheme, so that proposals are not needed for the covariance parameter, which would require positive definiteness constraints. It can also be used to obtain the posterior mode for $\beta_i, i = 1, \ldots, n$. This can potentially be much faster than full MCMC, especially when we would like to compare many different models. In either of these cases, we can use the following conditional posterior distributions, and expectations
3.4. Asymptotic Results

We now study the posterior rate of contraction, \( \epsilon_n \), such that the posterior probability of the \( M_n \epsilon_n \)-ball around the true parameter given \( n \) observations tends to 1 as \( n \) increases to infinity for any \( M_n \to \infty \).

Let \( \Pi \) denote the prior measure on \( \mathcal{P} \), the parameter space regarded as a subset of all probability measures, with a typical element \( P \) having density \( p \), and within which exists the true measure, \( P_0 \). For \( P, Q \in \mathcal{P} \), let

\[
K(p, q) = P \log \frac{p}{q}, \quad V_+(p, q) = P \log_+ \frac{p}{q},
\]

where \( \log_+ x = \max(\log x, 0) \). We will also make use of the Hellinger distance, defined as \( d_H(P, Q) = (\int \sqrt{p} - \sqrt{q})^2 \), and the Frobenius norm on matrices, which, for \( A = (a_{ij}) \), is defined to be \( \|A\|_F^2 = \sum_{i,j} a_{ij}^2 \).

Let \( P_0 \) stand for the true distribution with density \( p_0 \). In order to obtain the posterior rate of convergence, \( \epsilon_n \), we apply Theorem 2.1 of Ghosal et al. [35]. Thus, we need to verify that for a constant \( C > 0 \), and a sequence \( \{\mathcal{P}_n\} \) of subsets of the parameter space,

\[
\log N(\epsilon_n/2, \mathcal{P}_n, d) \leq n\epsilon_n^2, \tag{3.18}
\]

\[
\Pi(\mathcal{P}_n^\epsilon) \leq e^{(-n\epsilon_n^2(C+4))}, \tag{3.19}
\]

\[
\Pi\left( P : K(p_0, p) \leq \epsilon_n^2, V_+(p_0, p) \leq \epsilon_n^2 \right) \geq e^{(-n\epsilon_n^2C)}, \tag{3.20}
\]

where \( N(\epsilon/2, \mathcal{P}_n, d) \) is the covering number, i.e., the minimum number of \( d \)-balls of size \( \epsilon_n/2 \) needed to cover \( \mathcal{P}_n \).
We assume given values of $J$, $K$, and $\sigma$. We also make the simplifying assumption that the functional observations have already been detrended, so that $\mu(\cdot) \equiv 0$ and the prior mean, $\theta_0$ is also taken to be 0. The theorem is stated in terms of the Frobenius norm on matrices. Now we state the main theorem of the chapter on the rate of convergence of our posterior.

**Theorem 3.4.1.** Let $Y_i \overset{iid}{\sim} N_T(0, \sigma^2 I + H U \Sigma_0 U' H')$, $i = 1, 2, \ldots$ for a known $\sigma^2$ and $K = K_0 < T$ fixed. Using the inverse Wishart prior from above,

$$\Pi\left( \Sigma : ||\Sigma_0 - \Sigma||_F \geq M_n n^{-1/2} \log n | Y_1, \ldots, Y_n \right) \rightarrow 0,$$

as well as

$$\Pi\left( \Sigma : ||\Sigma_0^{-1} - \Sigma^{-1}||_F \geq M_n n^{-1/2} \log n | Y_1, \ldots, Y_n \right) \rightarrow 0,$$

3.5 Simulation Study

To assess the finite-sample performance of the proposed method, we present the results of a simulation study comparing the approach advocated herewithin to a recent frequentist approach, FACE [36], implemented using the `refund` package in R.

3.5.1 Description of FACE

The frequentist method to which we compare our method is the “Fast Covariance Estimation” (FACE) method of Xiao et al. [36]. It is a very common frequentist method for the analysis of functional data, and has been made popular by the `refund` package available in R. The FACE estimator, $\tilde{F}$, is simply a sandwich-smoothed sample covariance matrix, that is

$$\tilde{F} = S \hat{F} S,$$

where $\hat{F}$ is the sample covariance matrix and $S$ is a symmetric “smoothing” matrix, which is constructed using penalized B-splines. The form allows fast computation of the estimator. Thus, the comparison to our Bayesian method can be seen as demonstrating the potential benefits of a more complex method when the available computation allows for a fully Bayesian method.
3.5.2 Results

Each data set consists of 20 noisy observations on an evenly spaced grid of 50 time points in the interval $[-1, 1]$. The true underlying functional observations all have a true mean of $\mu(t) = \sin(2\pi t)$, and a covariance of either $\kappa_1(s, t) = e^{-3(t-s)^2}$ or $\kappa_2(s, t) = \min\{s + 1, t + 1\}$, depending on the experimental conditions. Independent sampling noise is then added in the form of independent $N(0, 0.3)$ random variables (other values for the variance of the noise were considered, but the results remained qualitatively very similar). The two methods are compared in the following realms: estimation of the mean function, estimation of the covariance function, estimation of the principal components, and reconstruction of a new set of underlying functional observations (generated according to the same model). Function estimation is evaluated using the supremum, $L_1$ and $L_2$ metrics, and principal component estimation is evaluated using the angle between the estimate and the true function (this was chosen instead of using squared distance, to take advantage of the Hilbert space structure). Observations are always reconstructed using the first 4 principal components.

For each choice of $\kappa_1$ and $\kappa_2$ as the true covariance function, we ran two simulations corresponding to choosing $\kappa_1$ or $\kappa_2$ as the prior covariance, yielding four total experimental conditions. Each scenario is repeated 100 times.

The proposed Bayesian procedure is computed using a Gibbs sampler with 5000 burn-in iterations, and 5000 iterations to estimate posterior means. The concentration parameter for the Wishart distribution, $\nu$, was chosen to be $2K$; this empirically seemed to be a reasonable default choice. Legendre polynomials were used as the orthonormal basis, a Poisson prior with mean 7 was placed on the number of basis functions, and a Poisson prior with mean 1 was placed on the number of principal components. However, for the computation, only models with fewer than 20 basis functions were run, which would correspond to a truncated prior. This practically had no effect on the results since models outside that range had negligible posterior mass. See Figure 3.5 for examples of the posterior distribution of $(J, K)$. The principal components were estimated using the decomposition of the estimated posterior mean of the covariance matrix. The results can be seen in Figures 3.1–3.4.

As can be seen from the results, the Bayesian procedure performs consistently well across the conditions in the estimation of the principal components themselves when measured by the angle from the truth. The practical importance of prior information can be seen in the improvements
in reconstruction when the true covariance is used to construct the prior. In the functional data setting, the smoothness of the underlying true observations is usually well understood scientifically in an applied context, and should be incorporated into the analysis. The overall picture that these results show is that the proposed Bayesian method has the ability to perform competitively with the most modern frequentist procedures when judged by repeated sampling criteria.
3.6 CANADIAN WEATHER DATA

To illustrate our method on real data, we analyzed the popular Canadian weather data, which is freely available in the \texttt{fda} package in R. The data was made popular by Ramsay & Silverman \cite{25}, and our analysis is consistent with theirs. These data consist of 35 functional observations.
observed on a common grid of 365 time points. They correspond to the average daily temperature of 35 Canadian cities. We employ Legendre polynomials as the basis, with an unknown number of basis functions. We use a modified Poisson distribution on $K$ with mean 7, and truncated above at 30. The parameter $\theta_0$ is taken to be zero, and the prior covariance function that is approximated was $\kappa(s, t) = e^{-3(s-t)^2}$. For each model, 180,000 MCMC iterations were used for estimation after
20,000 burn-in iterations. Estimates are only calculated at the sampled time points; if there are other time points of interest, ideally, they should be treated as missing data and incorporated into the MCMC approximations.

In the posterior, almost all the mass lies on the model with $J = 12, K = 12$ (96.1%), and a small amount on $J = 16, K = 15$ (3.9%); a plot of the marginal likelihoods for each model can

---

**Figure 3.4 Simulation Study**: Boxplots for the simulation results corresponding to $\kappa_2$ used for the truth and $\kappa_1$ used for the prior. Results are shown in pairs, with the left box (red color) representing the proposed Bayesian procedure, and the right (black) representing the FACE procedure. The value above each pair represents the proportion of data sets in which the Bayesian procedure performed superiorly.
Table 3.1 Simulation Study: Comparison of MSEs. Rows in dark red text are experimental conditions where our method outperformed FACE. In all of these cases the difference is significant in a frequentist sense with the null hypothesis that there is equal probability of either method winning a trial.

<table>
<thead>
<tr>
<th>Prior</th>
<th>Truth</th>
<th>MSE(Bayes)</th>
<th>MSE(FACE)</th>
<th>P(MSE(Bayes) &lt; MSE(FACE))</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>$\kappa_1$</td>
<td>0.066</td>
<td>0.075</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.020</td>
<td>0.016</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.011</td>
<td>0.017</td>
<td>0.89</td>
</tr>
<tr>
<td>PC2</td>
<td>$\kappa_1$</td>
<td>0.073</td>
<td>0.088</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.048</td>
<td>0.069</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.076</td>
<td>0.086</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.025</td>
<td>0.071</td>
<td>0.98</td>
</tr>
<tr>
<td>PC3</td>
<td>$\kappa_1$</td>
<td>0.045</td>
<td>0.065</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.074</td>
<td>0.101</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.078</td>
<td>0.064</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>$\kappa_2$</td>
<td>0.031</td>
<td>0.104</td>
<td>0.98</td>
</tr>
</tbody>
</table>

be seen in Figure 3.6. Posterior estimates shown are full model estimates, although they will be extremely close to conditioning on the maximum *a posteriori* model. The observations along with their smoothed estimates can be seen in Figure 3.7. The estimated covariance function can be seen in Figure 3.8, along with the implied principal components in Figure 3.9.

The first principal component represents the overall temperature of the city throughout the year; it differentiates between generally “mild” and “cold” cities. The second principal component seems to quantify the relative difference in temperature between summer and winter months, and differentiates between cities that have a more flat temperature function, compared to those with extremely cold winters. The higher order principal components represent more complicated phenomena.

### 3.7 Discussion

There are two points that warrant further discussion beyond what has already been presented: modifications to the model under special circumstances, and the computational difficulties of fitting the model.
3.7. **DISCUSSION**

**CHAPTER 3. BAYESIAN FPCA**

**Figure 3.5 Simulation Study**: Examples of the posterior distribution of models for each of the four experimental conditions. Row 1 corresponds to $\kappa_1$ as the prior, and column 1 corresponds to $\kappa_1$ as the truth. The other row and column correspond to $\kappa_2$.

Although our method allows for the possibility that the number of basis functions used for approximation and the number of principal components require different values, in the form we present, the number of basis functions is the same for both mean and covariance modeling. One possible concern is that, in the situation where the mean function requires many more basis functions to be well-approximated compared with the covariance function, the chosen value of $K$ will be forced too high. Specifically, to deal with this, we can allow for an extra overall mean term,
3.7. DISCUSSION

Figure 3.6 Canadian Weather Data: Posterior probabilities for each model, \((J, K)\). Only models \(1 \leq K \leq J \leq 25\) had prior mass.

Figure 3.7 Canadian Weather Data: Observations (dots) along with pointwise posterior means (lines). The posterior means are the mean of \(H_j U_k \beta_{i,k}\) over \((J, K)\).
3.7. DISCUSSION

Figure 3.8 Canadian Weather Data: Full posterior mean of the covariance function over all \((J, K)\) pairs.

\[ \eta \text{ in (3.10) that can capture the excess roughness present in the mean:} \]

\[ Y_i \overset{iid}{\sim} N_{J}(H_J U_K \beta_{i,K} + \eta, \sigma^2 I) \]  
\[ \eta \sim N_{J}(\eta_0, \Psi) \]  

(3.21a) (3.21b)

Allowing \(\Psi\) to be full rank would be sure to provide great flexibility, but may lead to over-fitting. A further modification could be to express \(\eta\) in the same basis domain as the rest of the model, and let \(\eta = H_M \xi, M > K\), with a prior on \(\xi\). This is similar to the approach in generalized additive models of allowing differing number of basis functions for each component, and the addition of this term causes an identifiability issue with the collection \(\{\beta_{i,K}\}\). This issue can be handled in a similar fashion to Lang & Brezger [37]. A step-wise frequentist approach could be to de-mean the data before processing, possibly including smoothing, and then apply our method with \(\theta\) removed from the model, that is, \(\beta_{i,K} \overset{iid}{\sim} N_{K}(0, \Sigma)\) in (3.10).

The final issue is the computational difficulties in fitting the full model. The approach we have taken is to run independent MCMC chains for each pair of values \((J, K)\), \(1 \leq K \leq J \leq J_{\text{max}}\), up to some predefined value, \(J_{\text{max}}\). The software approach the we have implemented will be
Figure 3.9 Canadian Weather Data: First four FPCAs corresponding to the full posterior mean of the covariance function (Figure 3.8) along with their associated proportion of variance explained.
made available on one of the authors’ website (https://www.ajsuarez.com). It uses R’s C interface to be able to take advantage of using OpenMP for parallelization of the MCMC chains. Specifically, for each value of $J$, the models, $1 \leq K \leq J$, are run in parallel batches. For example of computation time, on a data set containing 50 functional observations at 100 time points, fitting all models up to $J_{\text{max}} = 25$ takes approximately 3 seconds per 1000 MCMC steps on a 6-core Intel Haswell CPU running at 4 GHz. Computation increases on the order of $J_{\text{max}}^2$, and this dominates compared with the time points. For an MCMC procedure that provides full information on each individual model, we believe this approach to be worth the computational time.
3.8 Proofs

Lemma 3.8.1 (Kullback-Leibler divergence under information loss). Let $X$ and $Y$ be random variables with densities $p$ and $q$, respectively, and $U$ be a random variable with uniform distribution on the unit interval, independent of $X$ and $Y$. Let $\tilde{p}$ and $\tilde{q}$ be the densities of $T(X,U)$ and $T(Y,U)$, respectively, for a measurable function $T$. Then $K(\tilde{p}, \tilde{q}) \leq K(p,q)$ and $V_+(\tilde{p}, \tilde{q}) \leq V_+(p,q)$.

This lemma can be proved by considering the conditional distributions of $X$ given $T(X,U)$ and $Y$ given $T(Y,U)$, and using the convexity of the maps $(u,v) \mapsto u \log \frac{u}{v}$ and $(u,v) \mapsto u \log^2 \frac{u}{v}$ on the set $u > v > 0$. A complete proof may be found in Ghosal & Vaart [74, Appendix, Lemma B.12].

Next, we need a lemma relating the Hellinger distance to the Frobenius norm induced on our parameter of interest. Recall that the Frobenius norm on matrices is defined to be, for $A = (a_{ij})$, $\|A\|_F^2 = \sum_{i,j} a_{ij}^2 = \text{tr}(A'A)$. This next lemma will be used in the entropy calculation.

Lemma 3.8.2. Let $d$ be the metric induced by the Hellinger distance on the $k$-dimensional centered multivariate Gaussian family. Then,

$$d(\Sigma_1, \Sigma_2) = d_H(N(0, \Sigma_1), N(0, \Sigma_2)) \leq \left\| \Sigma_1^{-1/2} (\Sigma_2 - \Sigma_1) \Sigma_1^{-1/2} \right\|_F. \quad (3.22)$$

Furthermore, there exists $\delta > 0$ and constant $C > 0$, depending on $\Sigma_1$, such that, if $d(\Sigma_1, \Sigma_2) < \delta$, then

$$\left\| \Sigma_1^{-1/2} (\Sigma_2 - \Sigma_1) \Sigma_1^{-1/2} \right\|_F \leq C d(\Sigma_1, \Sigma_2). \quad (3.23)$$

Proof. Let $\{\lambda_j\}$ be the eigenvalues of $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$, which is a symmetric, positive-definite matrix. Then for some orthogonal matrix $P$, Then, $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} = P \Lambda P'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$. 
Then,

\[
d_H(N(0, \Sigma_1), N(0, \Sigma_2))^2 = 1 - \frac{\det(\Sigma_1)^{1/4} \det(\Sigma_2)^{1/4}}{\det(\frac{1}{2}(\Sigma_1 + \Sigma_2))^{1/2}} \\
= 1 - \frac{\det(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})^{1/4}}{\det(\frac{1}{2}(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} + I))^{1/2}} \\
= 1 - \frac{\prod_{j=1}^{k} \gamma_j^{1/4}}{\left(\prod_{j=1}^{k} \frac{1}{2}(\lambda_j + 1)\right)^{1/2}}. \tag{3.24}
\]

We now show that (3.24) is less than or equal to \(\sum_{j=1}^{k} (\lambda_j - 1)^2\) by induction.

For \(k = 1\), define a function \(\xi(\lambda) = (\lambda - 1)^2 + \lambda^{1/4}(\frac{1}{2}(1 + \lambda))^{-1/2}\). We have that \(\xi(0) = \xi(1) = 1\), and we claim that \(\xi(\lambda) \geq 1\) for \(\lambda > 0\). Now,

\[
\frac{d}{d\lambda} \xi(\lambda) = (\lambda - 1)(8\lambda^{3/4}(\lambda + 1)^{3/2} - \sqrt{2}) \tag{3.25}
\]

The two positive real roots of the numerator are 1 and 0.403. It can then be seen that \(\frac{d}{d\lambda} \xi(\lambda) \geq 0\) for \(\lambda \leq 0.403\) or \(\lambda \geq 1\), and that \(\frac{d}{d\lambda} \xi(\lambda) \leq 0\) in between these values. Taking this together, this implies that \(\xi(\lambda) \geq 1\) for all \(\lambda \geq 0\), which is equivalent to the induction hypothesis for \(k = 1\).

For the induction step, we define \(\gamma(\lambda) = \lambda^{1/4}(\frac{1}{2}(1 + \lambda))^{-1/2}\) and claim that \(\gamma(\lambda) \leq 1\) for \(\lambda \geq 0\). This can be seen by the fact that \(\gamma(1) = 1\) and

\[
\frac{d}{d\lambda} \gamma(\lambda) = -(\lambda - 1)2^{-3/2} \lambda^{-3/4}(\lambda + 1)^{-3/2}
\]

satisfies \(\frac{d}{d\lambda} \gamma(\lambda) \geq 0\) for \(0 \leq \lambda \leq 1\) and \(\frac{d}{d\lambda} \gamma(\lambda) \leq 0\) for \(\lambda \geq 1\). So, by the induction hypothesis and the case \(k = 1\), we have that

\[
1 - \frac{\prod_{j=1}^{k} \lambda_j^{1/4}}{\left(\prod_{j=1}^{k} \frac{1}{2}(\lambda_j + 1)\right)^{1/2}} \leq 1 - \frac{\lambda_k^{1/4}}{(\frac{1}{2}(1 + \lambda_k))^{1/2}} + \frac{\lambda_k^{1/4}}{(\frac{1}{2}(1 + \lambda_k))^{1/2}} \sum_{j=1}^{k-1} (\lambda_j - 1)^2 \\
\leq \sum_{j=1}^{k} (\lambda_j - 1)^2. \tag{3.26}
\]
3.8. PROOFS

Chapter 3. Bayesian FPCA

The proof is now complete by noting that \[ \| \Sigma_1^{-1/2} (\Sigma_2 - \Sigma_1) \Sigma_1^{-1/2} \|_F = \text{tr} (P (\Lambda - I) P') = \sum_{j=1}^{k} (\lambda_j - 1)^2. \]

The reverse inequality can be established following the arguments given in Banerjee & Ghosal [75, Lemma A.1].

**Proof of Theorem 3.4.1.** Using the earlier notation, \( p_0 = N(0, \sigma^2 I + H \Sigma_0 U' H') \) and \( p = N(0, \sigma^2 I + H \Sigma U' H') \). Let \( q_0 = N(0, \Sigma_0) \), \( q = N(0, \Sigma) \), and \( \Sigma^* = \Sigma_0^{-1/2} \Sigma \Sigma_0^{-1/2} \). Hence, by Lemma 3.8.1, we obtain that

\[
\begin{align*}
\Pi \left( P_0 \log \frac{P_0}{p} \leq \epsilon_n^2, P_0 (\log_+ \frac{P_0}{p})^2 \leq \epsilon_n^2 \right) \\
\geq \Pi \left( \frac{Q_0}{q} \log \frac{Q_0}{q} \leq \epsilon_n^2, Q_0 (\log_+ \frac{Q_0}{q})^2 \leq \epsilon_n^2 \right) \\
= \Pi \left( \text{tr}(\Sigma^*) - k - \log \det(\Sigma^*) < 2\epsilon_n^2, (\text{tr}(\Sigma^*) - k)^2 \leq 4\epsilon_n^2 \right) \\
= \Pi \left( \sum_{j=1}^{k} (\lambda_j - 1 - \log \lambda_j) < 2\epsilon_n^2, \left( \sum_{j=1}^{k} (\lambda_j - 1)^2 \right)^2 \leq 4\epsilon_n^2 \right),
\end{align*}
\]

where \( \{\lambda_j\}_{j=1}^{k} \) are the eigenvalues of the matrix \( \Sigma^* \). Now, \( \Sigma^{-1} \sim \text{Wishart}(v, \Xi) \) implies \( \Sigma^* \sim \text{Wishart}(v, \Sigma_0^{1/2} \Xi \Sigma_0^{-1/2}) \). Now, for large enough \( n \),

\[
\begin{align*}
\Pi \left( \sum_{j=1}^{k} (\lambda_j - 1 - \log \lambda_j) < 2\epsilon_n^2, \left( \sum_{j=1}^{k} (\lambda_j - 1)^2 \right)^2 \leq 4\epsilon_n^2 \right) \\
\geq \Pi \left( \sum_{j=1}^{k} (\lambda_j - 1)^2 < 2\epsilon_n, \lambda_j \geq 1, j = 1, \ldots, k \right) \\
\geq \Pi \left( 1 \leq \lambda_j \leq 1 + k^{-1/2} \sqrt{2\epsilon_n^{1/2}} \right) \\
\geq C_k \epsilon_n^{k^2/4 + k/4} \geq e^{(-n \epsilon_n^2 C)}
\end{align*}
\]

for \( \epsilon_n = n^{-1/2} \log n \), and some constants, \( C \) and \( C_k \). The first inequality of (3.28) is obtained by applying Lemma 1 of Shen et al. [76]; see Lemma 9.23 of Ghosal & Vaart [74] for a complete proof, including the case \( \Xi \neq I \).

Now, let \( \mathcal{P}_n = \{ \Sigma : \| \Sigma^{-1} \|_F \leq n^{1/8} \} \). Let \( d \) be the distance on the space of symmetric, positive definite matrices induced by the Hellinger metric on the multivariate Gaussian family. For
\[ \Sigma_1, \Sigma_2 \in \mathcal{P}_n, \text{ we have that} \]
\[ d(\Sigma_1, \Sigma_2) \leq \left\| \Sigma_1^{-1/2}(\Sigma_2 - \Sigma_1)\Sigma_1^{-1/2} \right\| F \leq \|\Sigma_1^{-1}\|_F \|\Sigma_2 - \Sigma_1\|_F \leq n^{1/8}\|\Sigma_2 - \Sigma_1\|_F, \tag{3.29} \]

which implies that
\[ \log N(\epsilon_n/2, \mathcal{P}_n, d) \leq \log N(n^{-1/8}\epsilon_n/2, \mathcal{P}_n, \|\cdot\|_F) \leq k^2 \log \left( \frac{6n^{1/4}}{\epsilon_n} \right) \lesssim n\epsilon_n^2 \tag{3.30} \]

for the chosen \( \epsilon_n \). The second inequality on the preceding line is due to the fact that, using the Frobenius norm, the space of positive definite \( k \times k \) matrices can be viewed as a subset of \( \mathbb{R}^{k^2} \) (see Pollard [77, Section 4] for the entropy calculation in Euclidean space).

Finally,
\[ \Pi \left( \|\Sigma^{-1}\|_F > n^{1/8} \right) \leq \Pi \left( \text{tr}(\Sigma^{-1}) > n^{1/8} \right) \leq e^{-n^{1/8}} \det(I - 2\Xi)^{-\nu/2} \]
\[ \lesssim e^{-n\epsilon_n^2(C+4)} \tag{3.31} \]

for the chosen \( \epsilon_n \). The first inequality is true because of the relationship between the Frobenius norm and the trace of a positive definite matrix. The second inequality follows from Markov’s inequality using the moment generating function of \( \text{tr}(\Sigma^{-1}) \) evaluated at 1 [78].

This establishing the consistency result in terms of the Hellinger distance, specifically, as \( n \to \infty \),
\[ \Pi \left( \Sigma : d(\Sigma, \Sigma_0) \geq M_n n^{-1/2} \log n | Y_1, \ldots, Y_n \right) \to 0. \]

Now, the reverse inequality of Lemma 3.8.2 implies the desired result in terms of the Frobenius norm:
\[ \Pi \left( \Sigma : \left\| \Sigma_0^{-1/2} (\Sigma - \Sigma_0) \Sigma_0^{-1/2} \right\|_F \geq M_n n^{-1/2} \log n | Y_1, \ldots, Y_n \right) \to 0. \]

Finally, since \( \Sigma_0 \) is a fixed matrix, the first stated result follows from the relation \( \|AB\|_F \leq \|A\|_F \|B\|_F \), for two matrices \( A \) and \( B \). To derive the second assertion, write \( \Sigma_0^{-1} - \Sigma^{-1} \) as \( \Sigma_0^{-1} (\Sigma - \Sigma_0) \Sigma^{-1} \) and apply the norm inequality repeatedly. \( \square \)
Chapter 4

Solvability of Nonlinear Equations

Much of the area of nonlinear analysis focuses on the problem of solving the equation

\[ Lx = F(x), \quad (4.1) \]

where \( L \) is a linear operator and \( F \) is a nonlinear operator defined on appropriate spaces. Many problems in differential and difference equations are approached in this manner; there are many classes of linear differential and difference problem which are well-understood and whose solutions have well-studied properties. The solvability of these linear problems, if uniquely solvable, correspond to the operator \( L \) having an inverse. The typical first-attempt strategy if \( L \) is invertible is to solve the equivalent problem,

\[ x = L^{-1}F(x). \quad (4.2) \]

This can be using any of the many fixed-point theorems that are prevalent in the nonlinear literature applied to the operator \( L^{-1} \circ F \). These theorems yield either unique solutions (invertible operators) through theorems such as the contraction mapping principle or non-unique solutions through theorems such as Schauder’s fixed point theorem.

The general approach adopted in the next two chapters is slightly different: consider (4.1) with the right-hand side decomposed into two parts

\[ Lx = F_1(x) + F_2(x). \quad (4.3) \]
Now the solvability of (4.1) can be potentially established in two steps

1. Prove that the operator $L - F_1$ is invertible, that is, solve the problem

$$Lx - F_1(x) = y$$  \hspace{1cm} (4.4)

for any $y$.

2. Solve the full problem, (4.3), by considering the equivalent problem

$$x = (L - F_1)^{-1} F_2(x)$$  \hspace{1cm} (4.5)

by applying a fixed point theorem to the operator $(L - F_1)^{-1} \circ F_2$.

The first step in this strategy leads us to study the invertibility of nonlinearly-perturbed invertible linear operators; this is dealt with in Lemma 5.2.1.

We apply this strategy in two situations where the properties of the linear structures are well-understood: Sturm-Liouville problems and periodic systems. In both cases, we employ basis expansions for the inverse of the linear operator to begin our study. In the first case, Chapter 5, this expansion is in terms of the eigenfunctions defined by the linear operator, and in the second case, Chapter 6, we use Fourier series to express the inverse.
On Nonlinear Perturbations of Sturm-Liouville Problems in Discrete and Continuous Settings

5.1 Introduction

The purpose of this chapter is to provide sufficient conditions for the existence of solutions to nonlinearly-perturbed Sturm–Liouville problems, both in the differential equations and difference equations settings. The perturbations that we consider also include modifications to the boundary conditions. We extend the results of Rodríguez & Abernathy [38, 39] by removing differentiability requirements in both cases, and, in the differential case, we also work in a different normed space, which leads to benefits in a subset of problems. For previous work in the discrete case, see also Rodríguez [40].

The main lemma in Section 5.2 can be seen as extending the so-called global inverse function theorems of Brown [41] and Brown & Lin [42], where the differentiability conditions are removed. The relationship between the eigenvalues of the original linear problem and the allowable nonlinearities is similar to Dolph [43], which studies closely related Hammerstein integral equations.

There has been much work done on similar problems or using similar approaches to the current chapter. Graef and Kong [44] study multiple solutions to boundary value problems including nonlinear Sturm–Liouville problems. For analyses of nonlinear discrete systems with linear boundary conditions see Etheridge & Rodríguez [45], Etheridge & Rodríguez [46], Rodríguez [47, 48], Rodríguez & Taylor [49, 50], and Agarwal [51]. For the differential equations setting, also with linear boundary
conditions, see Rodríguez & Taylor [52]. For the use of projection methods in more general nonlinear problems, see Rodríguez & Sweet [53]. For the use of Galerkin methods in differential problems, see Rodríguez [54, 55]. In Landesman, Lazer, et al. [56], nonlinear partial differential boundary value problems are considered. Three point boundary conditions in the context of nonlinear second-order differential equations are studied in Boucherif [57], and Lazer & Leach [58] considered similar nonlinear equations under two-point boundary conditions. Periodic perturbations of second order equations were considered in Leach [59], and then generalized to functions on $\mathbb{R}^n$ in Lazer, Sánchez, et al. [60]. Integral boundary conditions are studied in Benchohra et al. [61], Feng et al. [62], Kiguradze & Šremr [63], and Väth [64]. Fixed point theorems applied to fractional differential equations can be found in Ahmad & Nieto [65] and Chang & Nieto [66]. Problems in elliptic boundary values problems can be found in Behrndt [67]. For boundary conditions that can be expressed as continuous linear functionals, see Zhao & Liang [68].

In Section 5.2 we cover the unified framework for the problems considered later. Section 5.3 deals with the continuous setting, and Section 5.4 examines the discrete setting.

5.2 Generalities

In this chapter we consider two related problems, both of which fall into the general framework of searching for solutions to equations of the form

$$Lx = H(x), \quad (5.1)$$

where $L$ is a linear operator, and $H$ is a nonlinear operator, both defined on some Banach space. Suppose that $L$ is known to have an inverse, and that $H = \Psi + G$. The strategy we shall employ is then to first find conditions under which $L - \Psi$ is guaranteed to also have an inverse. This will uniquely solve the related equation

$$Lx - \Psi(x) = y, \quad (5.2)$$

for any point $y$ in the Banach space. Given a result of this type, we then study conditions under which (5.1) has a (possibly non-unique) solution by studying the operator $(L - \Psi)^{-1}G$. For this part we shall rely upon Schauder’s fixed point theorem and degree theory arguments. To study
5.2 GENERALITIES

CHAPTER 5. NONLINEAR STURM-LIOUVILLE PROBLEMS

conditions for which (5.2) can be solved for any \( y \), we will make use of the following lemma, which considers when a nonlinear perturbation to an invertible linear operator preserves the invertibility.

**Lemma 5.2.1.** Let \( X \) be a Banach space, \( Y \) be a normed linear space, \( D \subseteq X \) be any subspace, \( T : X \to Y \), and \( L : D \subseteq X \to Y \), where \( L \) is linear, \( L^{-1} \) exists and is Lipschitz continuous with constant \( K_1 \), \( L^{-1} \circ T \) is Lipschitz continuous with constant \( K_2 < 1 \). Then, the map \( L-T : D \subseteq X \to Y \) is invertible; furthermore, \((L-T)^{-1} : Y \to D \) is Lipschitz continuous with constant \( K_1(1-K_2)^{-1} \). In addition, if \( L^{-1} \) is compact and \( T \) is continuous, then \((L-T)^{-1} \) is also compact.

**Proof.** Consider the map \( S : X \times Y \to X \) defined by \( S(x, y) = L^{-1}(T(x) + y) \), and define the families of maps, \( S_x : y \mapsto S(x, y) \) and \( S_y : x \mapsto S(x, y) \). Let \( x_1, x_2 \in X \) and \( y \in Y \). Then,

\[
\|S_y(x_1) - S_y(x_2)\| = \|L^{-1}T(x_1) - L^{-1}T(x_2)\| \leq K_2 \|x_1 - x_2\|.
\]

So, \( S_y \) is a uniform (over \( Y \)) contraction on \( X \). Now, let \( y_1, y_2 \in Y \) and \( x \in X \). Then,

\[
\|S_x(y_1) - S_x(y_2)\| = \|L^{-1}y_1 - L^{-1}y_2\| \leq K_1 \|y_1 - y_2\|.
\]

So, \( S_x \) is uniformly (over \( X \)) Lipschitz continuous with constant \( K_1 \). By the contraction mapping theorem with parameters [69, Corollary 2.3.2], there exists a Lipschitz continuous function, \( g : Y \to X \), with constant \( K_1(1-K_2)^{-1} \), such that

\[
S(x, y) = x \iff x = g(y).
\]

Since \( L^{-1}(Y) = D \), \( g : Y \to D \). For \( x \in D \),

\[
S(x, y) = x \iff Lx - T(x) = y,
\]

which means that \( g = (L-T)^{-1} \).

To see that \( g \) is compact if \( L^{-1} \) is compact and \( T \) is continuous, note that the fixed point relations means that

\[
g = L^{-1} \circ ((T \circ g) + I),
\]

which is the composition of a compact operator with a continuous operator, since \( g \) is continuous.
We consider modifications to classical Sturm-Liouville problems, both in the differential equations setting and in the difference equation setting. Due to the infinite dimensional nature of the spaces considered when studying the differential equations setting, more issues arise that require special attention. For the difference equations that are considered, the finite dimensionality simplifies many of the problems encountered in the following section.

5.3 Differential Equations

We consider differential equations on the interval $[0, 1]$ of the form

$$(p(t)x'(t))' + q(t)x(t) + \psi(x(t)) = G(x(t)), \quad (5.3)$$

subject to boundary conditions of the form

$$\alpha x(0) + \beta x'(0) + \eta_1(x) = \phi_1(x) \quad (5.4a)$$
$$\gamma x(1) + \delta x'(1) + \eta_2(x) = \phi_2(x). \quad (5.4b)$$

In the above equations, $\psi$ and $G$ are function-valued operators, and $\eta_1, \eta_2, \phi_1,$ and $\phi_2$ are real-valued functions. These represent the nonlinear perturbations of the classical linear problem, but conditions on the right-hand and left-hand side perturbations that guarantee the existence of a solution will be qualitatively different. We make the usual assumptions on the linear portions of the problem, namely that $p, p', q$ are continuous, $p > 0,$ and $\alpha^2 + \beta^2 \neq 0, \gamma^2 + \delta^2 \neq 0.$

The strategy for analyzing (5.3)-(5.4) will be to first determine conditions under which we can uniquely solve the following problem:

$$(p(t)x'(t))' + q(t)x(t) + \psi(x(t)) = h(t), \quad (5.5)$$

subject to

$$\alpha x(0) + \beta x'(0) + \eta_1(x) = v_1 \quad (5.6a)$$
$$\gamma x(1) + \delta x'(1) + \eta_2(x) = v_2. \quad (5.6b)$$
This problem corresponds to finding an inverse for the operator representing the left-hand side of (5.5)-(5.6), and is dealt with in Theorem 5.3.1.

In general, we use capital Roman letters to denote sets, and capital script letters to denote the corresponding spaces equipped with a specified norm. On any Euclidean space, $\mathbb{R}^n$, we denote the usual norm by $|\cdot|$, regardless of dimension. Let $L^2$ denote the set of square Lebesgue integrable functions on $[0,1]$, and let $\mathcal{L}^2 = (L^2, \|\cdot\|_2)$, where $\|\cdot\|_2$ is the usual norm defined by the inner product $(x, y) = \int_0^1 x(t)y(t)\,dt$ for $x, y \in L^2$. Let $D = \{x \in L^2 | x''(\text{weak}) \in L^2\}$, and let $\mathcal{D}_2 = (D, \|\cdot\|_2)$. On this space we define the operators representing the linear part of the problem, $A : D \to L^2$ and $B : D \to \mathbb{R}^2$, where

$$A(x) \equiv (px')' + qx,$$

$$B(x) \equiv \begin{pmatrix} \alpha x(0) + \beta x'(0) \\ \gamma x(1) + \delta x'(1) \end{pmatrix},$$

and let

$$\mathbb{L}(x) \equiv \begin{pmatrix} A(x) \\ B(x) \end{pmatrix}.$$

For the nonlinear portion, let $\eta = (\eta_1, \eta_2)$ and $\phi = (\phi_1, \phi_2)$. Then define $\Psi, \mathcal{G} : D \to L^2 \times \mathbb{R}^2$ as

$$\Psi(x) \equiv \begin{pmatrix} -\psi(x) \\ -\eta(x) \end{pmatrix}, \mathcal{G}(x) \equiv \begin{pmatrix} G(x) \\ \phi(x) \end{pmatrix}.$$

On $L^2 \times \mathbb{R}^2$ we will use the sum of the the two usual component norms, and denote this space $\mathcal{L}^2 \times \mathbb{R}^2$. The properties of these components guaranteeing the existence of solutions will be the content of the main results of this chapter. For now, simply view them as nonlinear maps. With this notation, we rewrite (5.3)-(5.4) as

$$\mathbb{L}x - \Psi(x) = \mathcal{G}(x),$$

for $x \in D$. The norm with which we choose to pair $D$ will have a major impact on the conditions for existence of solutions.
5.3. DIFFERENTIAL EQUATIONS

CHAPTER 5. NONLINEAR STURM-LIOUVILLE PROBLEMS

Classical Sturm-Liouville theory deals with equations of the form $Ax = h$ for $x \in B^{-1}(\{0\})$, and we will require well-known properties from this case (see, for example, Al-Gwaiz [70]). First, on $B^{-1}(\{0\}) \cap \mathcal{D}_2$, $A$ has countably many simple eigenvalues, $\{-\lambda_k\}_{k=1}^\infty$, which we can assume are in decreasing order, such that $\lambda_k \to \infty$ as $k \to \infty$. For each $-\lambda_k$, there is an eigenfunction, $u_k$, with unit norm, which spans the corresponding eigenspace. In general, $A$ might fail to be injective. Because of this, we take $\mu \in \mathbb{R}$, which is not an eigenvalue, and let

$$L_\mu x \equiv \begin{pmatrix} Ax + \mu x \\ Bx \end{pmatrix}, \Psi_\mu(x) \equiv \begin{pmatrix} -\psi(x) + \mu x \\ -\eta(x) \end{pmatrix}.$$ 

This implies that on the set $B^{-1}(\{0\}) \cap \mathcal{D}_2$, the operator $A + \mu I$ has eigenvalues $\{\mu - \lambda_k\}_{k=1}^\infty$ corresponding to the same eigenfunctions, but where it is now guaranteed that 0 is not an eigenvalue.

From the theory of second order differential equations, it is known that $(A + \mu I)x = 0$ has a two dimensional solution space. We can choose a basis, $\{w_1, w_2\}$, for this space such that $\|w_1\|_2 + \|w_2\|_2 \leq 1$. Then let $w = (w_1, w_2)^t$ and define the $2 \times 2$ matrix $B = [Bw_1 | Bw_2]$. With these definitions we can now give the form of the inverse of $L_\mu$. The following result is cited from Rodríguez & Abernathy [38], but we provide a short proof.

**Lemma 5.3.1** (from Rodríguez & Abernathy [38]). $L_\mu : D \to L^2 \times \mathbb{R}^2$ is bijective and

$$L_\mu^{-1}(h, v) = \sum_{k=1}^{\infty} \frac{\langle h, u_k \rangle}{\mu - \lambda_k} u_k + w^t B^{-1} v,$$

(5.7)

where the limit is in the sense of uniform convergence.

**Proof.** It is well known that the map, $h \mapsto \sum_{k=1}^{\infty} u_k \langle h, u_k \rangle / (\mu - \lambda_k)$ is the eigenfunction expansion of the integral operator

$$\mathcal{G} : h \mapsto \int_0^1 g(\cdot, s)h(s)\,ds,$$
where \( g \) is the so-called Green’s function for the problem

\[
(A + \mu I)x(t) = (p(t)x'(t))' + q(t)x(t) = h(t)
\]

\[
\mathbb{B} x = \begin{pmatrix} \alpha x(0) + \beta x'(0) \\ \gamma x(1) + \delta x'(1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

The eigenfunction expansion is known to converge in \( L^2 \), and furthermore this convergence can also be shown to be uniform. Furthermore, the kernel of \( G \) is known to be \( \{0\} \). This fact, along with the definitions of \( w \) and \( B \), imply that

\[
(A + \mu I) \mathbb{L}^{-1}_\mu (h, v) = h + (A + \mu I) w^t B^{-1} v = h,
\]

\[
\mathbb{B} \mathbb{L}^{-1}_\mu (h, v) = 0 + B B^{-1} v = v,
\]

which proves the result.

**Lemma 5.3.2.** \( \mathbb{L}^{-1}_\mu \) is a compact operator from \( L^2 \times \mathbb{R}^2 \) onto \( \mathcal{D}_2 \).

**Proof.** The map \( h \mapsto \int_0^1 g(t, s) h(s) ds \) is known to be a compact map from \( L^2 \) to \( L^2 \), hence it is compact on the product domain. The map \( v \mapsto w^t B^{-1} v \) is also compact on the domain since its range is finite dimensional. Thus, the sum is compact.

The next corollary provides estimates on norms of interest related to \( \mathbb{L}^{-1}_\mu \). The operator norm is given the same notation as the norm on the corresponding domain of the operator. The definition of the graph norm, \( \| \cdot \|_{gr} \), can be found in proposition 5.3.1.

**Corollary 5.3.1.** Let \( A_0 = \sup_k |\mu - \lambda_k|^{-1} \), \( B_0 = \|B^{-1}\| \), \( C_0 = \sup_k |\lambda_k/(\mu - \lambda_k)| \), and \( D_0 = \max\{\|\mu\|, \|B^{-1}\|\} \).

\[
\|\mathbb{L}^{-1}_\mu (h, v)\|_2 \leq A_0 \|h\|_2 + B_0 \|v\|, \\
\|\mathbb{L}^{-1}_\mu \|_2 \leq \max\{A_0, B_0\} \equiv K^*, \\
\|\mathbb{L}^{-1}_\mu (h, v)\|_{gr} \leq (A_0 + C_0) \|h\|_2 + (B_0 + D_0) \|v\|, \\
\|\mathbb{L}^{-1}_\mu \|_{gr} \leq \max\{A_0 + C_0, B_0 + D_0\} \equiv N^*.
\]
Proof.

\[ \|L^{-1}_\mu(h, v)\|_2 \leq \sum_{k=1}^{\infty} \left| \frac{\langle h, u_k \rangle}{\mu - \lambda_k} \right| + \|w^t B^{-1} v\|_2 \leq A_0 \sum_{k=1}^{\infty} |\langle h, u_k \rangle|^2 + |B^{-1} v| \]

\[ \leq A_0 \|h\|_2 + B_0 |v|, \]

\[ \|A L^{-1}_\mu(h, v)\|_2 = \left\| \frac{\sum_{k=1}^{\infty} \mu \langle h, u_k \rangle}{\mu - \lambda_k} u_k - \mu w^t B^{-1} v \right\|_2 \]

\[ \leq C_0 \|h\|_2 + |\mu| \|B^{-1} v\|. \]

So we have that

\[ |B L^{-1}_\mu(h, v)| = |v|. \]

The bounds on operator norms easily follow from these inequalities. □

**Proposition 5.3.1.** \( L : \mathcal{D}_2 \to \mathcal{L}^2 \times \mathcal{R}^2 \) is a closed operator. Hence, \( \mathcal{D}_{gr} \equiv (D, \|\cdot\|_{gr}) \) is a Banach space, where the norm is the graph norm defined by \( \|x\|_{gr} = \|x\|_2 + \max\{|\mathcal{A} x\|_2, |\mathcal{B} x\|\} \).

**Proof.** \( L_\mu \) is a closed operator since it has a continuous inverse. The operator \( \mu I \) is clearly a bounded linear operator. Thus \( L = L_\mu - \mu I \) is a closed operator. It follows that \( \mathcal{D}_{gr} \) is a Banach space by the closed graph theorem. □

**Remark 5.3.1.** The fact that \( L \) is closed can be used to show that the convergence of the sum in (5.7) is in the stronger sense of the graph norm.

As opposed to relying on a global inverse function theorem, as in Rodríguez & Abernathy [38], we instead use Lemma 5.2.1. One significant benefit of this approach is the removal of any differentiability requirements, which are then replaced by requirements of Lipschitz continuity. As we shall see later on, another advantage is the ability to work on an incomplete space, which allows certain constraints on constants to be relaxed.
Remark 5.3.2. In our case, \( D \) is incomplete, but we can view it as a subset of its completion, \( L^2 \). To apply the lemma, we will then need to only consider differences with operators defined on all of \( L^2 \). Since \( D_{gr} \) is already a Banach space, we do not need to consider operators defined on any larger set.

**Theorem 5.3.1.**

(i) Assume \( \psi - \mu I : L^2 \rightarrow L^2 \) is Lipschitz continuous with constant \( K_1 \), and \( \eta : L^2 \rightarrow \mathbb{R}^2 \) is Lipschitz continuous with constant \( K_2 \). Then, if \( A_0 K_1 + B_0 K_2 < 1 \), \( L - \Psi : L^2 \rightarrow L^2 \times \mathbb{R}^2 \) is invertible; moreover \( (L - \Psi)^{-1} \) is compact and Lipschitz continuous with constant \( K \equiv K^*(1 - A_0 K_1 - B_0 K_2)^{-1} \).

(ii) Assume \( \psi - \mu I : D_{gr} \rightarrow L^2 \) is Lipschitz continuous with constant \( K_1 \), and \( \eta : D_{gr} \rightarrow \mathbb{R}^2 \) is Lipschitz continuous with constant \( K_2 \). Then, if \( (A_0 + C_0)K_1 + (B_0 + D_0)K_2 < 1 \), \( L - \Psi : D_{gr} \rightarrow L^2 \times \mathbb{R}^2 \) is invertible; moreover \( (L - \Psi)^{-1} \) is Lipschitz continuous with constant \( N \equiv N^*(1 - (A_0 + C_0)K_1 - (B_0 + D_0)K_2)^{-1} \).

**Proof.** First, notice that \( L - \Psi = L_{\mu} - \Psi_{\mu} \).

(i) We check the conditions of lemma (5.2.1). Let \( x_1, x_2 \in D \). Then, using corollary (5.3.1),

\[
\|L^{-1}_{\mu} \Psi_{\mu}(x_1) - L^{-1}_{\mu} \Psi_{\mu}(x_2)\| \quad \leq A_0\|\psi - \mu I(x_2) - (\psi - \mu I)(x_1)\|_2 + B_0\|\eta(x_2) - \eta(x_1)\|_2
\]

\[
\leq (A_0 K_1 + B_0 K_2)\|x_2 - x_1\|.
\]

Lemma (5.2.1) now implies the desire result since \( L^{-1}_{\mu} \) is compact and Lipschitz continuous with constant \( K^* \) by corollary (5.3.1).

(ii) We check the conditions of lemma (5.2.1). Let \( x_1, x_2 \in D \). Then, using corollary (5.3.1),

\[
\|L^{-1}_{\mu} \Psi_{\mu}(x_1) - L^{-1}_{\mu} \Psi_{\mu}(x_2)\|_{gr} \quad \leq (A_0 + C_0)\|\psi - \mu I(x_2) - (\psi - \mu I)(x_1)\|_2
\]

\[
+ (B_0 + D_0)\|\eta(x_2) - \eta(x_1)\|_2
\]

\[
\leq ((A_0 + C_0)K_1 + (B_0 + D_0)K_2)\|x_2 - x_1\|.
\]

Lemma (5.2.1) now implies the desired result since \( L^{-1}_{\mu} \) is Lipschitz continuous with constant \( N^* \) by corollary (5.3.1).
In part (2), it can be assumed that $\psi - \mu I$ and $\eta$ are only defined on $D$, since $D_{gr}$ is a Banach space whereas $D_2$ is not. This conclusion was a result from Rodríguez & Abernathy [38]; however, it was proved under different conditions, notably, the assumption of differentiability replaced our Lipschitz conditions. It should be noted that the important constant for the result to hold is strictly larger in part (2) than in part (1). This shows that there is a trade-off between the size of the domain of the functions of interest and the restrictions of their corresponding Lipschitz constants. For example, consider the operator $\eta_1$ in (5.4). Theorem 5.3.1 part (2) can be used for the case of $\eta_1(x) = \sum_{i=1}^{n} f(x(t_i)),$

where $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, and $t_i \in [0, 1]$. This is a so-called multipoint boundary condition. Theorem 5.3.1 part (1) does not cover this case. However, the case of $\eta_1(x) = \int_0^1 f(x(t)) dt,$

where $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, can be handled by part (1), and the constraints on the constants are less stringent than part (2) would require.

Before stating the next theorem, we must recall some basic facts about the spaces of interest. In Rodríguez & Abernathy [38], it was shown that, on $D$, $\| \cdot \|_{g_r}$ is equivalent to the Sobolev norm, $\| \cdot \|_{2,2}$. It is well known that the Sobolev space $\mathcal{H}^2$ has compact embeddings into both $(C^1[0, 1], \| \cdot \|_{C^1})$ and $\mathcal{L}^2$, and hence, so does $D_{gr}$. Here we have used $C^1[0, 1]$ to denote the functions on $[0, 1]$ which have a continuous first derivative, and for $x \in C^1[0, 1]$, $\| x \|_{C^1} = \sup |x(t)| + \sup |x'(t)|$. Denote these embeddings by $j_1 : D_{gr} \hookrightarrow C^1[0, 1]$ and $j_2 : D_{gr} \hookrightarrow \mathcal{L}^2$, and their embedding constants by $C_1$ and $C_2$, respectively.

**Theorem 5.3.2.** Assume the conditions of Theorem 5.3.1 (I) for part 1 below; assume the conditions of Theorem 5.3.1 (2) for parts 2-3 below.

(i) Let $\mathcal{G} : \mathcal{L}^2 \to \mathcal{L}^2 \times \mathbb{R}^2$ be continuous and assume that there exists an $M \in \mathbb{N}$ such that, for $\| x \|_2 \leq M$, $\| \mathcal{G}(x) \|_2 \leq K^{-1}(M - \| (L-\Psi)^{-1}0 \|_2)$. Then there exists at least one point, $x_0 \in D$ such that $\mathcal{L}x_0 - \Psi(x_0) = \mathcal{G}(x_0)$.

(ii) Let $\mathcal{G} : D_{gr} \to \mathcal{L}^2 \times \mathbb{R}^2$ be compact and assume that there exists an $M \in \mathbb{N}$ such that, for
5.3. DIFFERENTIAL EQUATIONS

\[ \|x\|_{\text{gr}} \leq M, \quad \|\mathcal{G}(x)\|_2 \leq N^{-1}(M - \|\mathcal{L} - \Psi^{-1}(0)\|_{\text{gr}}). \] Then there exists at least one point, \( x_0 \in D \) such that \( Lx_0 - \Psi(x_0) = \mathcal{G}(x_0) \).

(iii) Let \( \mathcal{G} : C^1[0,1] \to \mathcal{L}^2 \times \mathbb{R}^2 \) be continuous, and map bounded sets into bounded sets. Assume that there exists an \( M \in \mathbb{N} \) such that, for \( \|x\|_C \leq M, \quad \|\mathcal{G}(x)\|_2 \leq N^{-1}(C^{-1}M - \|\mathcal{L} - \Psi^{-1}(0)\|_{\text{gr}}). \] Then there exists at least one point, \( x_0 \in D \) such that \( Lx_0 - \Psi(x_0) = \mathcal{G}(x_0) \).

Proof. (i) \( H \equiv (\mathcal{L} - \Psi)^{-1}\mathcal{G} : \mathcal{L}^2 \to \mathcal{L}^2 \) is compact since it is the composition of a compact operator with a continuous operator. Now, let \( B = \{z \in L^2 : \|z\|_2 \leq M\} \). Let \( x \in B \), then

\[
\begin{align*}
\|(\mathcal{L} - \Psi)^{-1}\mathcal{G}(x)\|_2 & \leq \|(\mathcal{L} - \Psi)^{-1}\mathcal{G}(x) - (\mathcal{L} - \Psi)^{-1}(0)\|_2 + \|(\mathcal{L} - \Psi)^{-1}(0)\|_2 \\
& \leq KK^{-1}(M - \|(\mathcal{L} - \Psi)^{-1}(0)\|_2) + \|(\mathcal{L} - \Psi)^{-1}(0)\|_2 \\
& \leq M.
\end{align*}
\]

Therefore, \( H(B) \subseteq B \), which is clearly closed, bounded, and convex, so by Schauder’s fixed point theorem [71, Theorem 8.8], there exists an \( x_0 \in L^2 \) such that \( (\mathcal{L} - \Psi)^{-1}\mathcal{G}(x_0) = x_0 \). Since \( (\mathcal{L} - \Psi)^{-1}(L^2 \times \mathbb{R}^2) = D, \ x_0 \in D \), the result is proven.

(ii) \( H \equiv (\mathcal{L} - \Psi)^{-1}\mathcal{G} : \mathcal{D}_{\text{gr}} \to \mathcal{D}_{\text{gr}} \) is compact since it is the composition of a continuous operator with a compact operator. Now, let \( B = \{z \in D : \|z\|_{\text{gr}} \leq M\} \). Let \( x \in B \), then

\[
\begin{align*}
\|(\mathcal{L} - \Psi)^{-1}\mathcal{G}(x)\|_{\text{gr}} & \leq \|(\mathcal{L} - \Psi)^{-1}\mathcal{G}(x) - (\mathcal{L} - \Psi)^{-1}(0)\|_{\text{gr}} + \|(\mathcal{L} - \Psi)^{-1}(0)\|_{\text{gr}} \\
& \leq NN^{-1}(M - \|(\mathcal{L} - \Psi)^{-1}(0)\|_{\text{gr}}) + \|(\mathcal{L} - \Psi)^{-1}(0)\|_{\text{gr}} \\
& \leq M.
\end{align*}
\]

Therefore, \( H(B) \subseteq B \), which is clearly closed, bounded, and convex, so by Schauder’s fixed point theorem, the result is proven.

(iii) \( H \equiv (\mathcal{L} - \Psi)^{-1}\mathcal{G} \circ j_1 : \mathcal{D}_{\text{gr}} \to \mathcal{D}_{\text{gr}} \) is compact since it is the composition of two continuous operators with a compact operator. Now, let \( B = \{z \in D : \|z\|_{\text{gr}} \leq C^{-1}M\} \), and notice that
\[ \|z\|_{gr} \leq C^{-1}_1 M \implies \|z\|_{C^1} \leq M. \] Let \( x \in B \), then

\[
\| (L - \Psi)^{-1} g(j, x) \|_{gr} \\
\leq \| (L - \Psi)^{-1} g(j, x) - (L - \Psi)^{-1}(0) \|_{gr} + \| (L - \Psi)^{-1}(0) \|_{gr} \\
\leq N N^{-1}(M - \| (L - \Psi)^{-1}(0) \|_{gr}) + \| (L - \Psi)^{-1}(0) \|_{gr} \\
\leq M.
\]

Therefore, \( H(B) \subseteq B \), which is clearly closed, bounded, and convex, so by Schauder’s fixed point theorem, the result is proven.

**Remark 5.3.3.** In the cases that require the compactness of \( \mathcal{G} \), this requirement can be replaced by the assumption that \( \mathcal{G} \) is \( \alpha \)-Lipschitz with constant less than \( K^{-1} \), where \( \alpha \) is a measure of noncompactness. The generalized form of Schauder’s theorem would then be used.

If the norm conditions in Theorem 5.3.2 can be shown not to be sharp, in the sense that for \( \|x\| = M \), the bound for the norm of \( \|\mathcal{G}\| \) can be reduced by some \( \epsilon > 0 \), then the result can be extended slightly. We state the extension only for the first case; however, it applies analogously to the others.

**Corollary 5.3.2.** Let \( \mathcal{G} : \mathcal{L}^2 \to \mathcal{L}^2 \times \mathbb{R}^2 \) be continuous and assume that there exists an \( M \in \mathbb{N} \) such that, for \( \|x\|_2 \leq M, \|\mathcal{G}(x)\|_2 \leq K^{-1}(M - \| (L - \Psi)^{-1}(0) \|_2) - \delta \), where \( \delta > 0 \). Consider \( \mathcal{G}_\epsilon \equiv \mathcal{G} + \epsilon \mathcal{F} \), where \( \mathcal{F} : \mathcal{L}^2 \to \mathcal{L}^2 \times \mathbb{R}^2 \) is continuous such that \( \sup_{x \in B} \|\mathcal{F}(x)\| = F < \infty \). Then, for every \( \epsilon < K\delta / F \), there exists at least one point, \( x_0 \in D \) such that \( Lx_0 - \Psi(x_0) = \mathcal{G}_\epsilon(x_0) \).

**Proof.** In Theorem 5.3.2, it is shown that the operator \( H \equiv (L - \Psi)^{-1} \mathcal{G} : \mathcal{L}^2 \to \mathcal{L}^2 \) has a fixed point on the set \( B = \{ z \in \mathcal{L}^2 : \|z\|_2 \leq M \} \), which means that \( d(I - H, 0, 0) \neq 0 \), where \( d \) is the Leray-Schauder degree. By the assumed condition, \( \rho(0, (I - H)(\partial B)) \geq K\delta \) since for \( \|x\|_2 = M, \|H(x)\| \leq M - K\delta \), where \( \rho \) is the \( \mathcal{L}^2 \) distance. By a well-known property of the degree [71, Theorem 8.2], any compact operator, \( H_2 \) within an \( K\delta \)-ball (sup norm) of \( H \) will satisfy \( d(I - H_1, 0, 0) = d(I - H_2, 0, 0) \). The operator \( H_2 \equiv (L - \Psi)^{-1}(\mathcal{G} + \epsilon \mathcal{F}) \) clearly satisfies this with the given assumption on \( \mathcal{F} \).

We now provide examples for each of the four cases of Theorem 5.3.2. We focus only on the first component of \( \mathcal{G}, G \), which maps into \( \mathcal{L}^2 \). Assume \( k : \mathbb{R} \times [0, 1] \to \mathbb{R} \) is continuous.
Example 5.3.1. The operator \(G(x) = k(x(\cdot), \cdot)\) is continuous from \(L^2\).

Example 5.3.2. The operator \(G(x) = \int_0^1 k(s, \cdot)x''(s)ds\) is compact from \(D_{gr}\).

Example 5.3.3. The operator \(G(x) = k(x'(\cdot), \cdot)\) is continuous from \(C^1[0,1]\).

The next corollary is an immediate consequence of Theorem 5.3.2, and will apply to all parts of Theorem 5.3.2. The norms will be left ambiguous, but correspond to those in the respective parts of the theorem.

Corollary 5.3.3. If \(G\) is sublinear in the sense that \(\|G(x)\| \leq a + b\|x\|^\zeta\), with \(\zeta \in [0,1)\), then Theorem 5.3.2 holds.

As an application of Theorem 5.3.2, we state the following two corollaries, where we use the notation from above for the linear part of the differential equation.

Corollary 5.3.4. Consider the multipoint boundary value problem on the interval \([0,1]\)

\[(p(t)x'(t))' + q(t)x(t) + \psi(x(t)) = G(x(t)),\]  

subject to the boundary conditions

\[\alpha x(0) + \beta x'(0) + \sum_{i=1}^{N} h_{1,i}(x(t_i)) = 0\]  

\[\gamma x(1) + \delta x'(1) + \sum_{i=1}^{M} h_{2,i}(x(s_i)) = 0,\]  

where \(\psi, G, h_{j,i} : \mathbb{R} \to \mathbb{R}\) are Lipschitz continuous functions. Let \(t \mapsto \psi(t) - \mu t\) be Lipschitz with constant \(K_1\), and let \(h_{j,i}\) be Lipschitz with constant \(k_{j,i}\). Let \(K_2 = \sum_{i=1}^{N} k_{1,i} + \sum_{i=1}^{M} k_{2,i}\). Then if \(\exists a, b, \zeta \in [0,1)\) such that \(|G(t)| \leq a + b|t|^\zeta\) for all \(t \in \mathbb{R}\), and, in addition, \((A_0 + C_0)K_1 + (B_0 + D_0)K_2 < 1\), then there exists a solution to the above problem. If \(G = 0\), then this solution is unique.

Corollary 5.3.5. Consider the boundary value problem on the interval \([0,1]\)

\[(p(t)x'(t))' + q(t)x(t) + \psi(x(t)) = G(x(t)),\]  

where \(\psi, G, h_{j,i} : \mathbb{R} \to \mathbb{R}\) are Lipschitz continuous functions. Let \(t \mapsto \psi(t) - \mu t\) be Lipschitz with constant \(K_1\), and let \(h_{j,i}\) be Lipschitz with constant \(k_{j,i}\). Let \(K_2 = \sum_{i=1}^{N} k_{1,i} + \sum_{i=1}^{M} k_{2,i}\). Then if \(\exists a, b, \zeta \in [0,1)\) such that \(|G(t)| \leq a + b|t|^\zeta\) for all \(t \in \mathbb{R}\), and, in addition, \((A_0 + C_0)K_1 + (B_0 + D_0)K_2 < 1\), then there exists a solution to the above problem. If \(G = 0\), then this solution is unique.
subject to the boundary conditions

\[ \alpha x(0) + \beta x'(0) + \int_0^1 h_1(x(t)) dt = 0 \]  
(5.9b)

\[ \gamma x(1) + \delta x'(1) + \int_0^1 h_2(x(t)) dt = 0, \]  
(5.9c)

where \( \psi, G, h_j : \mathbb{R} \rightarrow \mathbb{R} \) are continuous functions. Let \( t \mapsto \psi(t) - \mu t \) be Lipschitz with constant \( K_1 \), and let \( h_j \) be Lipschitz with constant \( k_j \). Let \( K_2 = k_1 + k_2 \). Then if \( \exists a, b \in \mathbb{R}, \zeta \in [0, 1) \) such that \( |G(t)| \leq a + b|t|^\zeta \) for all \( t \in \mathbb{R} \), and, in addition, \( A_0 K_1 + B_0 K_2 < 1 \), then there exists a solution to the above problem. If \( G = 0 \), then this solution is unique.

This last corollary shows the trade-off that using the weaker norm in the domain has on a typical result. If the problem only contains operators which can be defined on all of \( L^2 \), then the condition on the Lipschitz constants can be weakened. The case of point evaluations can, however, be approximated using this last corollary, by taking the integrals in the boundary conditions to be the convolution with a mollifier. The level of approximation to the point evaluation will also determine the Lipschitz constant, which will increase along with the accuracy.

### 5.4 Difference Equations

We consider difference equations of the form

\[ \Delta(p(t-1)\Delta x(t-1)) + q(t)x(t) + \psi(x)(t) = G(x)(t) \]  
(5.10)

for \( t = a + 1, \ldots, b + 1 \), subject to boundary conditions of the form

\[ \alpha x(a) + \beta \Delta x(a) + \eta_1(x) = \phi_1(x) \]  
(5.11a)

\[ \gamma x(b+1) + \delta \Delta x(b+1) + \eta_2(x) = \phi_2(x). \]  
(5.11b)

Since we will work over the integers, we will drop the intersection in the notation and simply view \( [a, b] \) as \( [a, b] \cap \mathbb{Z} \). We assume that \( x \in \mathbb{R}^{[a,b+2]}, p \in \mathbb{R}^{[a,b+1]}, \) and \( q \in \mathbb{R}^{[a+1,b+1]} \). Analogous to the previous section, \( \psi \) and \( G \) are function-valued operators, and \( \eta_1, \eta_2, \phi_1, \) and \( \phi_2 \) are...
real-valued functions. These represent the nonlinear perturbations of the classical linear problem, but conditions on the right-hand and left-hand side perturbations that guarantee the existence of a solution will be qualitatively different. We make the usual assumptions on the linear portions of the problem, namely that $p, p', q$ are continuous, $p > 0$, $\alpha^2 + \beta^2 \neq 0$, $\gamma^2 + \delta^2 \neq 0$, $\alpha \neq \beta$, and $\gamma \neq \delta$.

The strategy for analyzing (5.10)-(5.11) will be to first determine conditions under which we can uniquely solve the following problem:

$$
\Delta(p(t-1)\Delta x(t-1)) + q(t)x(t) + \psi(x)(t) = h(t),
$$

subject to

$$
\begin{align*}
\alpha x(a) + \beta \Delta x(a) + \eta_1(x) &= v_1 \\
\gamma x(b+1) + \delta \Delta x(b+1) + \eta_2(x) &= v_2,
\end{align*}
$$

where $h \in \mathbb{R}[a+1,b+1]$ and $v_1, v_2 \in \mathbb{R}$. This problem corresponds to finding an inverse for the operator representing the left-hand side of (5.12)-(5.13), and is dealt with in Theorem 5.4.1.

Let $X = \mathbb{R}^{[a,b+2]}$, $Y = \mathbb{R}^{[a+1,b+1]}$, and $N = b-a+1$. On any Euclidean space, $\mathbb{R}^n$, we denote the usual norm by $|\cdot|$, regardless of dimension. On $X$ and $Y$ we use the inner product $(x, y) = \sum_t x(t)y(t)$, where the sum ranges over the domain of the functions. Along with this inner product, we use the typical norm of $\|x\| = (x, x)$.

With this notation, we define the operators representing the linear part of the problem, $A : X \to Y$ and $B : X \to \mathbb{R}^2$, where

$$
A(x) \equiv \Delta(p(t-1)\Delta x(t-1)) + q(t)x(t),
$$

$$
B(x) \equiv \begin{pmatrix}
\alpha x(a) + \Delta \beta x(a) \\
\gamma x(b+1) + \delta \Delta x(b+1)
\end{pmatrix},
$$

and let

$$
\mathbb{L}(x) \equiv \begin{pmatrix}
A(x) \\
B(x)
\end{pmatrix}.
$$
For the nonlinear portion, let \( \eta = (\eta_1, \eta_2) \) and \( \phi = (\phi_1, \phi_2) \). Then define \( \Psi, \mathcal{G} : X \to Y \times \mathbb{R}^2 \) as

\[
\Psi(x) \equiv \begin{pmatrix} -\psi(x) \\ -\eta(x) \end{pmatrix}, \quad \mathcal{G}(x) \equiv \begin{pmatrix} G(x) \\ \phi(x) \end{pmatrix}.
\]

On \( Y \times \mathbb{R}^2 \) we will use the sum of the two usual component norms. The properties of these components guaranteeing the existence of solutions will be the content of the main results of this chapter. For now, simply view them as nonlinear maps. With this notation, we rewrite (5.10)-(5.11) as

\[
Lx - \Psi(x) = \mathcal{G}(x),
\]

for \( x \in X \).

Classical Sturm-Liouville theory deals with equations of the form \( A x = h \) for \( x \in \mathbb{B}^{-1}(\{0\}) \), and we will require well-known properties from this case. First, \( A \) has \( N \) simple eigenvalues, \( \{-\lambda_k\}_{k=1}^N \), which we can assume are in decreasing order. For each \( -\lambda_k \), there is an eigenfunction, \( u_k \), with unit norm, which spans the corresponding eigenspace. It is known that for \( k \neq j \), \( \langle u_k, u_j \rangle = 0 \). In general, \( A \) might fail to be injective. Because of this we take \( \mu \in \mathbb{R} \), which is not an eigenvalue, and let

\[
L_\mu x \equiv \begin{pmatrix} Ax + \mu x \\ \mathbb{B} x \end{pmatrix}, \quad \Psi_\mu(x) \equiv \begin{pmatrix} -\psi(x) + \mu x \\ -\eta(x) \end{pmatrix}.
\]

This implies that the operator \( A + \mu I \) has eigenvalues \( \{\mu - \lambda_k\}_{k=1}^N \) corresponding to the same eigenfunctions, but where it is now guaranteed that \( 0 \) is not an eigenvalue.

From the theory of second order difference equations, it is known that \( (A + \mu I)x = 0 \) has a two dimensional solution space. We can choose a basis, \( \{w_1, w_2\} \), for this space such that \( \|w_1\| + \|w_2\| \leq 1 \). Then let \( w = (w_1, w_2)^t \) and define the \( 2 \times 2 \) matrix \( \mathbb{B} = [\mathbb{B} w_1 | \mathbb{B} w_2] \). With these definitions we can now give the form of the inverse of \( L_\mu \). The following result is cited from Rodríguez & Abernathy [39].
Lemma 5.4.1 (from Rodríguez & Abernathy [39]). \( L_{\mu} : X \rightarrow Y \times \mathbb{R}^2 \) is bijective and

\[
L_{\mu}^{-1}(h, v) = \sum_{k=1}^{N} \frac{\langle h, u_k \rangle}{\mu - \lambda_k} u_k + w^t B^{-1} v.
\]

The next corollary provides estimates on norms of interest related to \( L_{\mu}^{-1} \). The operator norm is given the same notation as the norm on the corresponding domain of the operator.

Corollary 5.4.1. Let \( A_0 = \sup_k |\mu - \lambda_k|^{-1}, B_0 = \|B^{-1}\| \).

\[
\|L_{\mu}^{-1}(h, v)\|_2 \leq A_0 \|h\|_2 + B_0 |v|,
\]

\[
\|L_{\mu}^{-1}\|_2 \leq \max\{A_0, B_0\} \equiv K^*.
\]

Proof.

\[
\|L_{\mu}^{-1}(h, v)\|_2 \leq \sum_{k=1}^{N} \left| \frac{\langle h, u_k \rangle}{\mu - \lambda_k} \right| + \|w^t B^{-1} v\|_2 \leq A_0 \sum_{k=1}^{N} |\langle h, u_k \rangle|^2 + |B^{-1} v|
\]

\[
\leq A_0 \|h\|_2 + B_0 |v|.
\]

\(\square\)

As opposed to relying on a global inverse function theorem, as in Rodríguez & Abernathy [39], we instead employ Lemma 5.2.1. As in the previous section, the significant benefit of this approach is the removal of any differentiability requirements, which are then replaced by requirements of Lipschitz continuity.

Theorem 5.4.1. Assume \( \psi - \mu I : X \rightarrow X \) is Lipschitz continuous with constant \( K_1 \), and \( \eta : X \rightarrow \mathbb{R}^2 \) is Lipschitz continuous with constant \( K_2 \). Then, if \( A_0 K_1 + B_0 K_2 < 1 \), \( L - \Psi : X \rightarrow Y \times \mathbb{R}^2 \) is invertible; moreover \( (L - \Psi)^{-1} \) is Lipschitz continuous with constant \( K \equiv K^*(1 - A_0 K_1 - B_0 K_2)^{-1} \).

Proof. First, notice that \( L - \Psi = L_{\mu} - \Psi_{\mu} \). We check the conditions of lemma (5.2.1). Let \( x_1, x_2 \in X \).
Then, using corollary (5.3.1),

$$||L_\mu \Psi(x_1) - L_\mu \Psi(x_2)||_2 \leq A_0 \|(\psi - \mu I)(x_2) - (\psi - \mu I)(x_1)||_2 + B_0 \|\eta(x_2) - \eta(x_1)||_2$$

$$\leq (A_0 K_1 + B_0 K_2)\|x_2 - x_1\|.$$ 

Lemma (5.2.1) now implies the desired result since $L_\mu^{-1}$ is Lipschitz continuous with constant $K^*$ by Corollary (5.4.1).

The conclusion was a result from Rodríguez & Abernathy [39]; however, it was proved under different conditions, notably the assumption of differentiability replaced our Lipschitz conditions.

**Theorem 5.4.2.** Assume the conditions of Theorem 5.3.1. Let $\mathcal{G} : X \to Y \times \mathbb{R}^2$ be continuous and assume that there exists an $M \in \mathbb{N}$ such that, for $\|x\| \leq M$, $\|\mathcal{G}(x)\| \leq K^{-1}(M - ||(L - \Psi)^{-1}(0)||)$. Then there exists at least one point, $x_0 \in X$ such that $Lx_0 - \Psi(x_0) = \mathcal{G}(x_0)$.

**Proof.** $H \equiv (L - \Psi)^{-1}\mathcal{G} : X \to X$ is continuous since it is the composition of two continuous operators. Now, let $B = \{z \in X ||z||_2 \leq M\}$, and let $x \in B'$, then

$$||(L - \Psi)^{-1}\mathcal{G}(x)||_2 \leq||(L - \Psi)^{-1}\mathcal{G}(x) - (L - \Psi)^{-1}(0)||_2 +||(L - \Psi)^{-1}(0)||_2 \leq K K^{-1}(M - ||(L - \Psi)^{-1}(0)||_2) + ||(L - \Psi)^{-1}(0)||_2 \leq M.$$ 

Therefore, $H(B) \subseteq B$, which is clearly closed, bounded, and convex, so by Brouwer’s fixed point theorem [71, Theorem 3.2], there exists an $x_0 \in X$ such that $(L - \Psi)^{-1}\mathcal{G}(x_0) = x_0$. 

**Corollary 5.4.2.** Let $\mathcal{G} : X \to Y \times \mathbb{R}^2$ be continuous and assume that there exists an $M \in \mathbb{N}$ such that, for $\|x\| \leq M$, $||\mathcal{G}(x)||_2 \leq K^{-1}(M - \|\mathcal{G}(x)\|_2)$. Assume that for any $x \in X$ such that $Lx - \Psi(x) = \mathcal{G}(x)$, $\|x\| < M$, and consider $\mathcal{G}_\epsilon \equiv \mathcal{G} + \epsilon \mathcal{F}$, where $\mathcal{F} : X \to Y \times \mathbb{R}^2$ is continuous. Then, for every $\epsilon$ small enough, there exists at least one point, $x_0 \in X$ such that $Lx_0 - \Psi(x_0) = \mathcal{G}_\epsilon(x_0)$.

**Proof.** In Theorem 5.4.2, it is shown that the operator $H \equiv (L - \Psi)^{-1}\mathcal{G} : X \to X$ has a fixed point on the set $B = \{z \in X ||z|| \leq M\}$, which means that $d(I - H, \hat{B}, 0) \neq 0$, where $d$ is the Brouwer degree. Since $\partial B$ is compact, so is $(I - H)(\partial B)$, and hence $\rho(0,(I - H)(\partial B)) = \delta$, for some
\[ \delta > 0, \text{ where } \rho \text{ is the Euclidean distance. This is because, by assumption, } I - H \text{ has no zeroes on the boundary of } B. \text{ By a well-known property of the degree [71, Theorem 3.1], any continuous operator, } H_2, \text{ within an } \delta \text{-ball (sup norm) of } H \text{ will satisfy } d(I - H, \hat{B}, 0) = d(I - H_2, \hat{B}, 0). \text{ Let } \sup_{x \in B} \| F(x) \| = F. \text{ Then, the operator } H_2 \equiv (L - \Psi)^{-1}(G + \epsilon F) \text{ clearly satisfies this whenever } \epsilon < K \delta / F. \]

The next corollary is an immediate consequence of Theorem 5.4.2.

**Corollary 5.4.3.** If \( G \) is sublinear in the sense that \( \| G(x) \| \leq a + b \| x \| ^{\zeta}, \) with \( \zeta \in [0, 1) \), then Theorem 5.4.2 holds.

As an application of Theorem 5.4.2, we state the following two corollaries, where we assume the notation as above for the linear part of the differential equation.

**Corollary 5.4.4.** Consider the multipoint boundary value problem on the interval \([0, 1]\),

\[
\Delta(p(t - 1) \Delta x(t - 1)) + q(t)x(t) + \psi(x(t)) = G(x(t)), \tag{5.14a}
\]

subject to the boundary conditions

\[
\alpha x(a) + \beta \Delta x(a) + \sum_{i=1}^{M_1} h_{1,i}(x(t_i)) = 0 \tag{5.14b}
\]

\[
\gamma x(b + 1) + \delta \Delta x(b + 1) + \sum_{i=1}^{M_2} h_{2,i}(x(s_i)) = 0, \tag{5.14c}
\]

where \( \psi, G, h_{j,i} : \mathbb{R} \rightarrow \mathbb{R} \) are Lipschitz continuous functions. Let \( t \rightarrow \psi(t) - \mu t \) be Lipschitz with constant \( K_1 \), and let \( h_{j,i} \) be Lipschitz with constant \( k_{j,i} \). Let \( K_2 = \sum_{i=1}^{M_1} k_{1,i} + \sum_{i=1}^{M_2} k_{2,i} \). Then if \( \exists a, b \in \mathbb{R}, \zeta \in [0, 1) \) such that \( |G(t)| \leq a + b |t|^{\zeta} \) for all \( t \in \mathbb{R} \), and, in addition, \( A_0 K_1 + B_0 K_2 < 1 \), then there exists a solution to the above problem. If \( G = 0 \), then this solution is unique.
Chapter 6

Existence of Solutions to Nonlinear Boundary Value Problems

6.1 Introduction

In this chapter we study general nonlinear systems of differential equations with boundary conditions. The goal is to provide sufficient conditions under which solutions exist. We do so by applying an abstract strategy previously used by the authors. We first consider the linear portion of the problem, together with periodic boundary conditions, and then analyze the requirements on the remaining nonlinearities to guarantee a solution.

In Section 6.2, we layout the basic strategy for proving the existence of solutions to general nonlinearly perturbed equations. This strategy has been used previously in the context of Sturm-Liouville problems [2, 38, 39]. This approach provides a type of global inverse function theorem, which can be seen as extending results from Brown [41] and Brown & Lin [42].

For a textbook study of multiple solutions to boundary value problems, see Graef & Kong [44]. Rodríguez & Taylor [52] studies similar boundary value problems to the present chapter, but for multipoint boundary conditions. Rodríguez & Sweet [53] provides conditions for when nonlinear boundary value problems, such as those considered here, can be reduced to finite-dimensional alternative problems. Rodríguez [55] uses Galerkin’s method to study nonlinear boundary value problems. For nonlinear boundary value problems with linear, Stieltjes boundary conditions, see Rodríguez [54]. Lazer & Leach [58] studies nonlinear two-point boundary value problems, and
Landesman, Lazer, et al. [56] covers the analagous case for partial differential equations. For the case of second order differential equations with three point boundary conditions, see Boucherif [57]. For second order equations with boundary conditions that can be represented as continuous linear functionals, see Zhao & Liang [68]. Elliptic boundary value problems are considered by Behrndt [67]. Boundary value problems with nonlocal, integral boundary conditions are studied in Benchohra et al. [61], Feng et al. [62], Kiguradze & Šremr [63], and Väth [64].

Related applications of fixed point theorems in the setting of fractional differential equations are dealt with in Ahmad & Nieto [65] and Chang & Nieto [66]. Dolph [43] studies nonlinear Hammerstein integral equations with applications to elliptical partial differential equations.

Section 6.3 describes the necessary background and defines the spaces on which we work. Section 6.4 applies the general theory to the case of nonlinear systems of differential equations. Section 6.5 shows how the previous results on Sturm-Liouville problems can be seen in the present context, and Section 6.6 provides an example of the application of the main theorem of the present chapter to a more general problem. Finally, in Section 6.7, we discuss alternative choices of normed spaces and how these choices would affect the main results.

### 6.2 General Strategy

In this section, we present a general strategy for proving the solvability of nonlinear equations when there is a linear component that is well-understood. This formalizes the abstract results from Rodríguez & Suarez [2].

Our first lemma describes the form of the inverse for a linear operator with two components under special conditions. For a linear operator, \( L \), let the kernel be denoted by \( \ker(L) = L^{-1}(\{0\}) \).

**Lemma 6.2.1.** Let \( D, X, \) and \( Y \) be vector spaces. Assume \( L : D \to Y \) and \( B : D \to Z \) are linear operators, and define \( \mathcal{L} = \begin{pmatrix} L \\ B \end{pmatrix} \). If \( L \upharpoonright_{\ker(B)} \) and \( B \upharpoonright_{\ker(L)} \) are both bijections, then

\[
\mathcal{L}^{-1}(y, z) \equiv \begin{pmatrix} L \\ B \end{pmatrix}^{-1}(y, z) = (L \upharpoonright_{\ker(B)})^{-1}(y) + (B \upharpoonright_{\ker(L)})^{-1}(z). \tag{6.1}
\]

**Proof.** It is clear that \( \ker(\mathcal{L}) = \{0\} \), which shows that \( \mathcal{L} \) is injective. Applying \( L \) and \( B \) to the right hand side of (6.1) shows that it is surjective, and that this is the inverse of \( \mathcal{L} \). \( \square \)
The general strategy pursued herewithin can be summarized in the following theorem:

**Theorem 6.2.1.** Let \( X \) be a Banach space, \( Y \) and \( Z \) be normed linear spaces, and \( D \subseteq X \) be a subspace. Assume \( L : D \to Y \) and \( B : D \to Z \) are linear operators. Assume \( F_1 : X \to Y \) and \( F_2 : X \to Z \) are Lipschitz continuous with constants \( K_1 \) and \( K_2 \), respectively. Assume \( L \upharpoonright \text{Ker}(B) \) and \( B \upharpoonright \text{Ker}(L) \) are both bijections, their inverses are bounded linear operators, and

\[
K^* = K_1 \left\| (L \upharpoonright \text{Ker}(B))^{-1} \right\| + K_2 \left\| (B \upharpoonright \text{Ker}(L))^{-1} \right\| < 1.
\]

If \( \mathcal{F} = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} \), then

\[
\mathcal{L} - \mathcal{F} = \begin{pmatrix} L \\ B \end{pmatrix} - \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}
\]

is invertible, and, in particular, \((\mathcal{L} - \mathcal{F})^{-1}\) is Lipschitz with constant

\[
K = (1 - K^*)^{-1} \max \left\{ \left\| (L \upharpoonright \text{Kern}(B))^{-1} \right\|, \left\| (B \upharpoonright \text{Kern}(L))^{-1} \right\| \right\}.
\]

If \( \mathcal{L}^{-1} \) is compact, then so is \((\mathcal{L} - \mathcal{F})^{-1}\).

Furthermore, if \( \mathcal{G} : X \to Y \times Z \) is a compact operator that satisfies: \( \exists M \in \mathbb{N}, \) such that for \( \|x\| \leq M, \|\mathcal{G}(x)\| \leq K^* M \) \( -(\mathcal{L} - \mathcal{F})^{-1}(0)\)), then the equation

\[
\mathcal{L}(x) - \mathcal{F}(x) = \mathcal{G}(x)
\]

has a solution in \( D \). If \( \mathcal{L}^{-1} \) is compact, then \( \mathcal{G} \) need only be continuous.

**Proof.** \( \mathcal{L} \equiv \begin{pmatrix} L \\ B \end{pmatrix} \) is invertible, and its inverse is given by Lemma 6.2.1. From this, it is clear that

\[
\left\| \mathcal{L}^{-1} \right\| \leq \max \left\{ \left\| (L \upharpoonright \text{Kern}(B))^{-1} \right\|, \left\| (B \upharpoonright \text{Kern}(L))^{-1} \right\| \right\}.
\]
6.2. GENERAL STRATEGY

We also have that

\[ L^{-1} F (x) = (L \upharpoonright_{\text{Ker}(B)})^{-1} F_1 (x) + B \upharpoonright_{\text{Ker}(L)}^{-1} F_2 (x). \]

It is clear from this that \( L^{-1} F \) is Lipschitz with constant \( K^* \). Since, by assumption, \( K^* < 1 \), this satisfies the conditions of Rodríguez & Suarez [2, Lemma 1].

Define \( H \equiv (L - F)^{-1} \circ G \). If both operators are continuous, and either is compact, then \( H \) is compact. Let \( B = \{ z \in D \| z \| \leq M \} \). Then,

\[
\| (L - F)^{-1} G (x) \| \leq \| (L - F)^{-1} G (x) - (L - F)^{-1} (0) \| + \| (L - F)^{-1} (0) \| \\
\leq K K^{-1} (M - \| (L - F)^{-1} (0) \|) + \| (L - F)^{-1} (0) \| \\
\leq M.
\]

Therefore, \( H(B) \subseteq B \), and, since \( B \) is clearly closed, bounded, and convex, by Schauder’s fixed point theorem, \( H \) has at least one fixed point.

\[ \square \]

**Remark 6.2.1.** If the operator, \( G \), satisfies a sublinear growth condition, that is

\[ \| G(x) \| \leq a + b \| x \|^\epsilon, \]

for some \( a, b \in \mathbb{R} \) and \( \epsilon < 1 \), then the growth condition in the previous theorem is automatically satisfied.

The second part of the previous theorem can be extended in the case that the growth condition is made stricter. The proof of the following can be found in Rodríguez & Suarez [2, Corollary 2].

**Corollary 6.2.1.** Assume the conditions of Theorem 6.2.1, but assume that for some \( \delta > 0 \), \( \exists M \in \mathbb{N} \), such that for \( \| x \| \leq M \), \( \| G(x) \| \leq K^{-1} (M - \| (L - F)^{-1} (0) \|) - \delta \). Consider \( G_\epsilon \equiv G + \epsilon H \), where \( H : X \to Y \) is continuous, such that \( \sup_{\| x \| \leq M} \| H(x) \| = H < \infty \). Then for every \( \epsilon \leq \frac{K \delta}{M} \), there exists at least one point, \( x_0 \in D \), such that \( L(x_0) + F(x_0) = G_\epsilon(x_0) \).
6.3 Preliminaries

We start by defining the spaces that we will consider. First, let $C[0, 1]$ be the set of continuous functions from $[0, 1]$ into $\mathbb{R}^n$, and let $C^1[0, 1]$ be those that are continuously differentiable. Let

$$C(1) = \{ f : [0, 1] \to \mathbb{R}^n | f(0) = f(1) \}.$$  \hspace{1cm} (6.2)

$$C^1(1) = \{ f : [0, 1] \to \mathbb{R}^n | f(0) = f(1) \}.$$  \hspace{1cm} (6.3)

Let $\|x\|_\infty = \sup_{t \in [0, 1]} |x(t)|$, where $|\cdot|$ is the usual Euclidean norm on $\mathbb{R}^n$. Let $\mathcal{C} = (C[0, 1], \|\cdot\|_\infty)$. We will also consider the following norm on $C^1[0, 1]$:

$$\|x\|_{C^1} = \sup_{t \in [0, 1]} |x(t)| + \sup_{t \in [0, 1]} |x'(t)|.$$  \hspace{1cm}

Then let

$$\mathcal{C}^1_1 = (C^1[0, 1], \|\cdot\|_\infty), \text{ and } \mathcal{C}^1 = (C^1[0, 1], \|\cdot\|_{C^1}).$$  

Note that the $\|\cdot\|_\infty$-completion of $C^1(1)$ is $C(1)$, and the $\|\cdot\|_\infty$-completion of $C^1[0, 1]$ is $C[0, 1]$.

6.4 Nonlinear Systems

In this chapter, we consider systems of ordinary differential equations on $[0, 1]$ of the form

$$\dot{x}(t) - f(x)(t) = g(x)(t)$$  \hspace{1cm} (6.4)

$$x(0) - x(1) - \eta(x) = \phi(x),$$  \hspace{1cm} (6.5)

where $\eta$ and $\phi$ are operators from $C[0, 1]$ into $\mathbb{R}^n$ and $f$ and $g$ are operators from $C[0, 1]$ to $C[0, 1]$. We look for solutions to (6.4) in $C^1[0, 1]$. We first shift the differential operator by a constant matrix, so that

$$\dot{x} - Ax - (f(x) - Ax) = g(x)$$  \hspace{1cm} (6.6)

$$x(0) - x(1) - \eta(x) = \phi(x).$$  \hspace{1cm} (6.7)
where \( A \) is a constant \( n \times n \) matrix. Let \( |\lambda_1| \leq \ldots \leq |\lambda_J| \) be the eigenvalues of \( A \). We assume that for \( j = 1, \ldots, J \), \( \lambda_j \neq 2\pi i k \) for any \( k \in \mathbb{Z} \).

Let \( L, L_A : \mathcal{C}_1 \subset \mathcal{C} \rightarrow \mathcal{C} \) be defined by
\[
Lz(t) = \dot{z}(t) \quad \text{and} \quad L_Az(t) = \dot{z}(t) - Az(t).
\]
Now, let \( B : \mathcal{C}_1 \subset \mathcal{C} \rightarrow \mathbb{R}^n \) be defined by
\[
Bz = z(0) - z(1),
\]
and consider the operators \( \mathcal{L}, \mathcal{L}_A : \mathcal{C}_1 \subset \mathcal{C} \rightarrow \mathcal{C} \times \mathbb{R}^n \), where
\[
\mathcal{L} \equiv -L \quad \text{and} \quad \mathcal{L}_A \equiv -L_A B.
\]

Then we have the following:

**Lemma 6.4.1.** Let \( h \in C[0,1] \) have the Fourier series
\[
h(t) \sim \sum_{k=-\infty}^{\infty} \alpha_k e^{2\pi i k t} \quad \text{for} \quad j = 1, \ldots, n.
\]
Then, \( L_A |_{\text{Ker}(B)} : \text{Ker}(B) \subset \mathcal{C}_1 \rightarrow \mathcal{C} \) is an invertible map, whose inverse is given by
\[
\left( L_A |_{\text{Ker}(B)} \right)^{-1} h(t) = \sum_{k=-\infty}^{\infty} (2\pi i k I - A)^{-1} \alpha_k e^{2\pi i k t},
\]
where the convergence is uniform.

**Proof.** If \( h \in C[0,1] \), \( h = L_A z \), with Fourier series,
\[
h(t) \sim \sum_k \alpha_k e^{2\pi i k t}.
\]
All solutions to \( h = L_A z \) are given by the variations of constants formula. With the restriction that \( z(0) = z(1) \), there is a unique solution for any \( h \in C[0,1] \) obtained by solving for the initial condition. Thus, \( L_A z = h \) has a solution in \( C^1(1) \) given by
\[
z(t) = e^{At} (I - e^{At})^{-1} e^A \int_0^1 e^{-As} h(s) ds + e^{At} \int_0^t e^{-As} h(s) ds,
\]
which has a uniformly convergent Fourier series since it is a member of \( C^1(1) \). Let \( z \) be in the domain of \( L_A \) with Fourier series, \( \sum_k \beta_k e^{2\pi i k t} \). It is well known that the Fourier series of an absolutely continuous function can be differentiated termwise [72, p. 2.3.4]. Also, multiplication by a constant matrix can clearly be moved inside the infinite sum. Therefore, the Fourier series (in the \( L^2 \) sense) of \( L_A z \) is given by
\[
L_A z(t) = \sum_k (2\pi i k I - A) \beta_k e^{2\pi i k t}.
\]
Equating coefficients, we see the given form for the inverse.

The following lemma provides estimates that will help to bound the operator norm of this inverse.

**Lemma 6.4.2.**

\[
\left\| \left( L_A \mid_{\text{Ker}(B)} \right)^{-1} h \right\|_\infty \leq \left( \sum_k \left\| (2\pi ikI - A)^{-1} \right\|_2^2 \right)^{1/2} \|h\|_\infty \equiv a_1 \|h\|_\infty \tag{6.11}
\]

**Proof.** To show that \( \sum_k \left\| (2\pi ikI - A)^{-1} \right\|_2^2 \) is summable, note that

\[
\left\| (2\pi ikI - A)^{-1} \right\|_2^2 = \max_j \sigma_j ((2\pi ikI - A)^{-1})^2,
\]

where \( \sigma_j \) is the \( j \)-th singular value. Then we have

\[
\max_j \sigma_j (2\pi ikI - A)^{-1} = \min_j \sigma_j (2\pi ikI - A)^{-1} \leq \left( \min_j \sigma_j (A) - 2\pi k \right)^{-1}.
\]

The inequality comes from the fact [73, Theorem 3.3.16] that

\[
\sigma_j (2\pi ikI - A) \geq \sigma_j (A) - 2\pi k.
\]

Now, for the inequality,

\[
\left| \sum_k (2\pi ikI - A)^{-1} a_k e^{2\pi i k t} \right| \leq \sum_k \left\| (2\pi ikI - A)^{-1} \right\|_2 |a_k| \leq \sum_k \left\| (2\pi ikI - A)^{-1} \right\|_2^2 \sum_k |a_k|^2 \leq \sum_k \left\| (2\pi ikI - A)^{-1} \right\|_2^2 \|h\|_\infty^2. \tag{6.14}
\]

**Lemma 6.4.3.**

\[
\left( \left( L_A \mid_{\text{Ker}(B)} \right)^{-1} v \right)(t) = e^{At} \left( I - e^A \right)^{-1} v \tag{6.15}
\]
Proof. The kernel of $L_A$ is $\{e^{At}w | w \in \mathbb{R}^n\}$. On this set, $Bx = v$ if and only if $x(t) = e^{At}(I - e^{A})^{-1}v$. 

The following immediately follows from the form of the inverse.

**Lemma 6.4.4.**

$$\left\| (B \restriction_{\text{ker}(L)})^{-1} v \right\|_{\infty} \leq \sup_{t \in [0,1]} \left\| e^{At}(I - e^{A})^{-1} \right\| \|v\| \equiv a_2\|v\|$$  \hspace{1cm} (6.16)

We now consider the left-hand side nonlinearities. Since we have shifted the linear portion by a constant matrix, we shift the nonlinear operator correspondingly. Define $F, F_A,$ and $G$ as

$$F \equiv \begin{bmatrix} f \\ \eta \end{bmatrix}, F_A \equiv \begin{bmatrix} f - A \\ \eta \end{bmatrix}, \text{and } G \equiv \begin{bmatrix} g \\ \phi \end{bmatrix}. \hspace{1cm} (6.17)$$

The original problem, (6.4), can now be written as

$$L - F = L_A - F_A = G$$  \hspace{1cm} (6.18)

The following provides conditions on $F_A$ so that $L - F$ is invertible; it is a direct consequence of Theorem 6.2.1.

**Theorem 6.4.1.** Assume that $f - A$ and $\eta$ are Lipschitz continuous, with constants $K_1$ and $K_2$, respectively.

Let $F_A: \mathcal{C} \to \mathcal{C} \times \mathbb{R}^2$ and $L_A: \mathcal{C}_1 \subset \mathcal{C} \to \mathcal{C}$, as above. If

$$K^* = a_1K_1 + a_2K_2 < 1,$$

Then for any $h \in C[0,1]$ and $v \in \mathbb{R}^n$, there exists a solution, $x_0 \in C^1[0,1]$, such that $L(x_0) - F(x_0) = (h, v)$.

To give an idea of the types of problems covered by the previous theorem, we give examples of Lipschitz functions that could appear in the boundary conditions.

**Example 6.4.1.** Let $\{t_i\}_{i=1}^n \subseteq [0,1]$ and $h: \mathbb{R} \to \mathbb{R}^n$ Lipschitz.
6.5. COMPARISON TO PREVIOUS RESULTS

1. \( \eta : \mathcal{C} \to \mathbb{R}^2 \), \( \eta(x) = \sum_{j=1}^{n} h(x(t_j)) \)

2. \( \eta : \mathcal{C} \to \mathbb{R}^2 \), \( \eta(x) = \int_{0}^{1} h(x(t)) dt \)

The importance of this example is to notice that both multipoint evaluations and global boundary conditions can be considered in this framework. The following is a direct consequence of the last portion of Theorem 6.2.1.

**Theorem 6.4.2.** Assume the conditions of Theorem 6.4.1. Let \( K = (1 - K^*)^{-1} \max\{a_1, a_2\} \). If \( \mathcal{G} : \mathcal{C} \to \mathcal{C} \times \mathbb{R}^2 \) is such that there exists an \( M \in \mathbb{N} \) such that for \( \|x\| \leq M \), \( \|\mathcal{G}(x)\| \leq K^{-1}(M - \|(\mathcal{L} - \mathcal{F})^{-1}(0)\|) \), then there exists at least one point, \( x_0 \in \mathcal{C}^1[0, 1] \) such that

\[
\mathcal{L}(x_0) - \mathcal{F}(x_0) = \mathcal{G}(x_0).
\]

Correspondingly, the following is a direct consequence of Corollary 6.2.1. It allows for further small perturbations.

**Theorem 6.4.3.** Assume the conditions of Theorem 6.4.1. Let \( K = (1 - K^*)^{-1} \max\{a_1, a_2\}, \delta > 0 \), and \( \mathcal{K} : \mathcal{C} \to \mathcal{C} \times \mathbb{R}^2 \) be continuous such that \( \sup_{\|x\| \leq M} \|\mathcal{K}(x)\| = H < \infty \). If \( \mathcal{G} : \mathcal{C} \to \mathcal{C} \times \mathbb{R}^2 \) is such that there exists an \( M \in \mathbb{N} \) such that for \( \|x\| \leq M \),

\[
\|\mathcal{G}(x)\| \leq K^{-1}(M - \|(\mathcal{L} - \mathcal{F})^{-1}(0)\|) - \delta,
\]

then, if \( \varepsilon < \frac{K\delta}{H} \) there exists at least one point, \( x_0 \in \mathcal{C}^1 \) such that

\[
\mathcal{L}(x_0) - \mathcal{F}(x_0) = \mathcal{G}(x_0) + \varepsilon \mathcal{K}(x_0).
\]

### 6.5 Comparison to Previous Results

The results presented in this chapter allow us to establish sufficient conditions for the solvability of systems of differential equations that do not fall within the framework of Rodríguez & Suarez [2] and Rodríguez & Abernathy [38]. Conversely, the examples of Sturm-Liouville problems that appear in Rodríguez & Suarez [2] and Rodríguez & Abernathy [38] can be studied within the framework of the present results, and this is what we present in this section. It is important to note, however,
that the sufficient conditions from those previous papers neither imply, nor are consequences of, the sufficient conditions in the present chapter. Results in the present chapter allow us to consider more general problems, but do not take advantage of special properties that exist in more specific contexts. In the previous paper, differential equations on $[0, 1]$ that were considered were of the form

$$\left(p(t)x'(t)\right)' + q(t)x(t) + \psi(x)(t) = G(x)(t),$$

subject to boundary conditions of the form

$$\alpha x(0) + \beta x'(0) + \eta_1(x) = \phi_1(x) \quad (6.20a)$$
$$\gamma x(1) + \delta x'(1) + \eta_2(x) = \phi_2(x). \quad (6.20b)$$

Clearly this can be rewritten as the following two dimensional problem:

$$x_1'(t) - x_2(t) = 0$$
$$x_2'(t) + \frac{1}{p(t)}(p'(t)x_2(t) + q(t)x_1(t) + \psi(x)(t)) = \frac{1}{p(t)}G(x)(t)$$

subject to boundary conditions of the form

$$x_1(0) - x_1(1) + (\alpha - 1)x_1(0) + \beta x_2(0) + \eta_1(x) = \phi_1(x) \quad (6.23a)$$
$$x_2(0) - x_2(1) + \gamma x_1(1) + (\delta - 1)x_2(1) + \eta_2(x) = \phi_2(x). \quad (6.23b)$$

Thus, a choice can be made as to which theorems to apply. The main practical consideration would typically come down to whether the eigenvalues for the linear Sturm-Liouville problem are easily calculated compared with the Lipschitz constants and growth conditions for the transformed problem.

Again, it should be noted that by applying the present results to these types of problems ignores the special structure of Sturm-Liouville problems.
6.6 Example

In this section we establish the solvability of a system of integro-differential equations that could not be established using the results of Rodríguez & Suarez [2] and Rodríguez & Abernathy [38]. Since the theorems herewithin can cover more than 2-dimensional problems, we provide a more concrete example of a 3-dimensional system to which our main theorems apply.

**Corollary 6.6.1.** Let \( \{s_{j,i}\} \bigcup \{t_{j,i}\} \bigcup \{u_{j,i}\} \subseteq [0,1] \). Consider the multipoint boundary value problem on the interval \([0,1]\)

\[
\begin{align*}
x'(t) + x(t) + f_1(x(t), y(t), z(t)) &= \int_0^1 \psi(y(s), t) \, ds \\
y'(t) + y(t) + f_2(x(t), y(t), z(t)) &= g_1(x(t), y(t), z(t)) \\
z'(t) + z(t) + f_3(x(t), y(t), z(t)) &= g_2(x(t), y(t), z(t))
\end{align*}
\]

subject to the boundary conditions

\[
\begin{align*}
x(0) - x(1) + \sum_{i=1}^{N} h_{1,i}(x(s_{1,i}), y(t_{1,i}), z(u_{1,i})) &= \phi_1(x, y, z) \\
y(0) - y(1) + \sum_{i=1}^{N} h_{2,i}(x(s_{2,i}), y(t_{2,i}), z(u_{2,i})) &= \phi_2(x, y, z) \\
z(0) - z(1) + \sum_{i=1}^{N} h_{3,i}(x(s_{3,i}), y(t_{3,i}), z(u_{3,i})) &= \phi_3(x, y, z)
\end{align*}
\]

where \( \{h_{j,i}\} : \mathbb{R}^3 \to \mathbb{R} \) are Lipschitz continuous functions with constants \( \{k_{j,i}\} \). Let \( \psi : \mathbb{R}^n \to \mathbb{R} \) be bounded, and \( \{f_j\} : \mathbb{R}^3 \to \mathbb{R} \) are Lipschitz continuous with constants \( \{c_i\} \). Let \( C_1 = \sum_{j=1}^{3} c_i \) and \( C_2 = \sum_{j=1}^{3} \sum_{i=1}^{N} k_{j,i} \). Assume \( \exists a_i, b_i \in \mathbb{R}, \xi \in [0,1) \), such that \( |g_j(x, y, z)| \leq a_i + b_i \|(x, y, z)\|^\xi \) for all \( t \in \mathbb{R} \) and \( i = 1,2 \). Let \( \phi_i : \mathcal{C} \to \mathbb{R}^3 \) be continuous and additionally assume that \( \exists A_i, B_i \in \mathbb{R}, \xi \in [0,1) \), such that \( |\phi_i(x, y, z)| \leq A_i + B_i \|(x, y, z)\|^\xi \) for all \( (x, y, z) \in \mathcal{C} \) and \( i = 1,2,3 \). Let \( K_1 = .5 \coth(.5) \), \( K_2 = \frac{e}{1-e} \), and assume that, \( C_1 K_1 + C_2 K_2 < 1 \). Then there exists at least one solution to the above problem.

This example particularly shows the flexibility of Theorem 6.4.2. It can handle relatively
large nonlinearities along with multipoint boundary conditions, and even in the context of integro-differential problems.

6.7 Discussion

In the language of Theorem 6.2.1, we have been working in the following spaces:

\[ D = \mathcal{C}^1, X = \mathcal{C}, Y = \mathcal{C}, Z = \mathbb{R}^n. \]

In this case, \( D \) is not complete, but can be seen as a subset of its completion, \( X \). One possible modification would be to consider the following spaces:

\[ D = \mathcal{C}^1, X = \mathcal{C}^1, Y = \mathcal{C}^1, Z = \mathbb{R}^n. \]

This has the effect of allowing more flexibility in the nonlinearities that can be considered since they would only need to be defined on a smaller set. For example, functions of the derivative could be considered. This comes at the expense of larger constants to consider when formulating an analogue to Theorem 6.4.1. This type of trade-off is discussed in detail in Rodríguez & Suarez [2].

Two other sets of spaces should be mentioned as well, and also correspond closely with the choices in Rodríguez & Suarez [2]. Let \( \| x \|_2 = \left( \int_{[0,1]} |x|^2 \right)^{1/2} \), and consider

\[
\| x \|_{1,2} = \left( \int_{[0,1]} |x|^2 \right)^{1/2} + \left( \int_{[0,1]} |x'|^2 \right)^{1/2}.
\]

We use the notation, \( x' \), to denote the weak derivative. Let \( H^1 \) denote the Sobolev space of square integrable functions whose weak derivatives are also members of \( L^2 \). Then let

\[ \mathcal{H}^1_2 = (H^1, \| \cdot \|_2) \quad \text{and} \quad \mathcal{H}^1_1 = (H^1, \| \cdot \|_{1,2}). \]
Analogously to the previous paragraph, the following pair of choices can be considered:

\[ D = H^1, X = L^2, Y = L^2, Z = \mathbb{R}^n \]
\[ D = H^1, X = H^1, Y = H^1, Z = \mathbb{R}^n. \]

Again, there is a trade-off between the types of nonlinearities that can be covered by the corresponding version of the major theorems and the constants that become part of the sufficient conditions for the existence of solutions.
BIBLIOGRAPHY


BIBLIOGRAPHY


