

# Shape Curve Analysis Using Curvature

James Miller

*A Dissertation Submitted to the  
University of Glasgow  
for the degree of  
Doctor of Philosophy*

Department of Statistics

June 2009

© James Miller, June 2009

# Abstract

Statistical shape analysis is a field for which there is growing demand. One of the major drivers for this growth is the number of practical applications which can use statistical shape analysis to provide useful insight. An example of one of these practical applications is investigating and comparing facial shapes. An ever improving suite of digital imaging technology can capture data on the three-dimensional shape of facial features from standard images. A field for which this offers a large amount of potential analytical benefit is the reconstruction of the facial surface of children born with a cleft lip or a cleft lip and palate. This thesis will present two potential methods for analysing data on the facial shape of children who were born with a cleft lip and/or palate using data from two separate studies. One form of analysis will compare the facial shape of one year old children born with a cleft lip and/or palate with the facial shape of control children. The second form of analysis will look for relationships between facial shape and psychological score for ten year old children born with a cleft lip and/or palate. While many of the techniques in this thesis could be extended to different applications much of the work is carried out with the express intention of producing meaningful analysis of the cleft children studies.

Shape data can be defined as the information remaining to describe the shape of an object after removing the effects of location, rotation and scale. There are numerous techniques in the literature to remove the effects of location, rotation and scale and thereby define and compare the shapes of objects. A method which does not require the removal of the effects of location and rotation is to define the shape according to the bending of important shape curves. This method can naturally provide a technique for investigating facial shape. When considering a child's face there are a number of curves which outline the important features of the face. Describing these feature curves gives a large amount of information on the shape of the face.

This thesis looks to define the shape of children's faces using functions of

bending, called curvature functions, of important feature curves. These curvature functions are not only of use to define an object, they are apt for use in the comparison of two or more objects. Methods to produce curvature functions which provide an accurate description of the bending of face curves will be introduced in this thesis. Furthermore, methods to compare the facial shape of groups of children will be discussed. These methods will be used to compare the facial shape of children with a cleft lip and/or palate with control children.

There is much recent literature in the area of functional regression where a scalar response can be related to a functional predictor. A novel approach for relating shape to a scalar response using functional regression, with curvature functions as predictors, is discussed and illustrated by a study into the psychological state of ten year old children who were born with a cleft lip or a cleft lip and palate. The aim of this example is to investigate whether any relationship exists between the bending of facial features and the psychological score of the children, and where relationships exist to describe their nature.

The thesis consists of four parts. Chapters 1 and 2 introduce the data and give some background to the statistical techniques. Specifically, Chapter 1 briefly introduces the idea of shape and how the shape of objects can be defined using curvature. Furthermore, the two studies into facial shape are introduced which form the basis of the work in this thesis. Chapter 2 gives a broad overview of some standard shape analysis techniques, including Procrustes methods for alignment of objects, and gives further details of methods based on curvature. Functional data analysis techniques which are of use throughout the thesis are also discussed.

Part 2 consists of Chapters 3 to 5 which describe methods to find curvature functions that define the shape of important curves on the face and compare these functions to investigate differences between control children and children born with a cleft lip and/or palate. Chapter 3 considers the issues with finding and further analysing the curvature functions of a plane curve whilst Chapter 4 extends the methods to space curves. A method which projects a space curve onto two perpendicular planes and then uses the techniques of Chapter 3 to calculate curvature is introduced to facilitate anatomical interpretation. Whilst the midline profile of a control child is used to illustrate the methods in Chapters 3 and 4, Chapter 5 uses curvature functions to investigate differences between control children and children born with a cleft lip and/or palate in terms of the bending of their upper lips.

Part 3 consists of Chapters 6 and 7 which introduce functional regression

techniques and use these to investigate potential relationships between the psychological score and facial shape, defined by curvature functions, of cleft children. Methods to both display graphically and formally analyse the regression procedure are discussed in Chapter 6 whilst Chapter 7 uses these methods to provide a systematic analysis of any relationship between psychological score and facial shape.

The final part of the thesis presents conclusions discussing both the effectiveness of the methods and some brief anatomical/psychological findings. There are also suggestions of potential future work in the area.

# Acknowledgements

Many thanks must go to my supervisor Professor Adrian Bowman for his patience, guidance and support throughout my research and ultimately in the production of this thesis. Without his help and encouragement this thesis would not have been possible. I would like to thank all those involved in the Cleft-10 project, both for access to data and also for the chance to be part of informative project meetings. A big thank you to Sarah and Denise for saving me a lot of time and frustration with your help in collating and providing data, it was and is much appreciated. I am also extremely grateful to the Department of Statistics at the University of Glasgow for funding me throughout my research.

The support and friendship of all the staff and students will be my abiding memory of the Department of Statistics. I have genuinely enjoyed myself and that is solely down to the superb people I have met and friends I have made. Somehow the tough times don't seem so bad when there is a group of good people with whom to enjoy a trip to Ashton Lane, Sauchiehall Street or even to RSC!

To Claire, I am eternally grateful for your patience and understanding over the past few years. I very much appreciate your willingness to drop everything when I need your help or company. The fun times we have enjoyed are what I will always remember most from my time at Glasgow.

Finally, thank you to my family. The example I have been set by the most loyal, hard working and trustworthy people I know is in no small part responsible for giving me the tools to carry out this research. It is difficult to put into words my gratitude for the sacrifices you have made for me, but I can confidently say that this thesis would never have happened without the love, support and constant encouragement I have received from all of you over my life.

# Declaration

This thesis has been composed by myself and it has not been submitted in any previous application for a degree. The work reported within was executed by myself, unless otherwise stated.

June 2009

# Contents

|  |           |
|--|-----------|
| <b>Abstract</b>  | <b>i</b>  |
| <b>Acknowledgements</b>  | <b>iv</b> |
| <b>1 Introduction</b>  | <b>1</b>  |
| 1.1 Shape . . . . .  | 1         |
| 1.2 Curvature . . . . .  | 2         |
| 1.3 Cleft Lip and Palate Data . . . . .                            | 3         |
| 1.4 Psychological Questionnaire . . . . .                          | 5         |
| 1.5 Missing Data . . . . .   | 6         |
| 1.6 Overview of Thesis . . . . .                                   | 6         |
| <b>2 Review</b>  | <b>8</b>  |
| 2.1 Statistical Shape Analysis . . . . .                           | 8         |
| 2.1.1 Procrustes analysis . . . . .                                | 9         |
| 2.1.2 Thin-plate splines, deformations and warping . . . . .       | 12        |
| 2.1.3 Other shape analysis techniques . . . . .                    | 14        |
| 2.2 Curve Analysis . . . . .                                       | 15        |
| 2.3 Functional Data Analysis . . . . .                             | 17        |
| 2.3.1 Interpolating and smoothing splines . . . . .                | 18        |
| 2.3.2 Principal components analysis . . . . .                      | 24        |
| 2.3.3 Curve registration . . . . .                                 | 26        |
| 2.3.4 Selected recent literature on modelling with functional data | 27        |
| <b>3 Analysis of Plane Curves</b>                                  | <b>31</b> |
| 3.1 Characterising Plane Curves . . . . .                          | 31        |
| 3.1.1 Arc length . . . . .   | 31        |
| 3.1.2 Tangent and normal vectors . . . . .                         | 32        |
| 3.1.3 Curvature and the Frenet formulae . . . . .                  | 33        |

|          |  |           |
|----------|--|-----------|
| 3.1.4    | Calculating curvature in practice . . . . .  | 35        |
| 3.1.5    | Reconstructing a curve from its curvature . . . . .                                  | 37        |
| 3.2      | Curvature of a Plane Curve: Midline Profile Example . . . . .                        | 38        |
| 3.2.1    | Calculating curvature . . . . .  | 38        |
| 3.2.2    | Reconstructing the profile . . . . .   | 41        |
| 3.2.3    | Investigating a collection of curvature curves . . . . .                             | 42        |
| 3.3      | Warping of Plane Curves . . . . .  | 43        |
| 3.3.1    | Warping technique for functions . . . . .  | 45        |
| 3.3.2    | Warping of curvature functions: Midline profile example . . . . .                    | 47        |
| 3.4      | Calculating Curvature: Alternative Methods . . . . .                                 | 52        |
| 3.4.1    | Optimisation method . . . . .  | 52        |
| 3.4.2    | Frenet method . . . . .  | 55        |
| 3.5      | Concluding Remarks on Plane Curves . . . . .   | 57        |
| <b>4</b> | <b>Analysis of Space Curves</b>  | <b>59</b> |
| 4.1      | Characterising Space Curves . . . . .  | 59        |
| 4.1.1    | Arc length . . . . .   | 59        |
| 4.1.2    | Osculating plane . . . . .   | 60        |
| 4.1.3    | Tangent, normal and binormal vectors . . . . .                                       | 61        |
| 4.1.4    | Curvature, torsion and the Frenet formulae . . . . .                                 | 61        |
| 4.1.5    | Calculating curvature and torsion in practice . . . . .                              | 64        |
| 4.1.6    | Reconstructing a curve from curvature and torsion . . . . .                          | 65        |
| 4.2      | Curvature and Torsion of a Space Curve: Midline Profile Example . . . . .            | 68        |
| 4.2.1    | Derivative method . . . . .  | 69        |
| 4.2.2    | Optimisation method . . . . .  | 72        |
| 4.2.3    | Frenet method . . . . .  | 74        |
| 4.2.4    | Perpendicular plane method . . . . .   | 78        |
| 4.2.5    | Investigating a collection of yz and xy curvature functions . . . . .                | 84        |
| 4.3      | Warping in the Space Curve Setting . . . . .   | 86        |
| 4.3.1    | Position warping technique for yz and xy curvature curves . . . . .                  | 87        |
| 4.3.2    | Warping of yz and xy curvature functions: Midline profile example . . . . .          | 89        |
| 4.4      | Concluding Remarks on Space Curves . . . . .   | 95        |
| <b>5</b> | <b>Applications to the Cleft Lip Problem</b>   | <b>96</b> |
| 5.1      | Techniques for Highlighting Differences Between Control and Cleft Children . . . . . | 96        |

|          |   |            |
|----------|---|------------|
| 5.1.1    | Curvature functions and average reconstructed features . . .                  | 97         |
| 5.1.2    | Position and curvature of characteristic points . . . . .                     | 97         |
| 5.1.3    | Warping to the average feature . . . . .                                      | 98         |
| 5.1.4    | Principal components analysis . . . . .                                       | 99         |
| 5.2      | Case Study: Upper Lip . . . . .   | 100        |
| 5.2.1    | Curvature functions and average upper lip reconstructions                     | 100        |
| 5.2.2    | Investigating characteristic points . . . . .                                 | 105        |
| 5.2.3    | Position warping and amplitude adjustment . . . . .                           | 109        |
| 5.2.4    | Principal components analysis . . . . .                                       | 113        |
| <b>6</b> | <b>Regression with Functional Predictors</b>                                  | <b>118</b> |
| 6.1      | Scalar Response and a Single Functional Predictor . . . . .                   | 119        |
| 6.1.1    | Displaying the data . . . . .   | 120        |
| 6.1.2    | Regression on principal component scores . . . . .                            | 123        |
| 6.1.3    | Functional linear model . . . . .   | 126        |
| 6.1.4    | Nonparametric functional regression . . . . .                                 | 135        |
| 6.2      | Scalar Response and Multiple Functional Predictors . . . . .                  | 141        |
| 6.2.1    | Displaying the data . . . . .   | 142        |
| 6.2.2    | Regression on principal component scores . . . . .                            | 144        |
| 6.2.3    | Functional linear model . . . . .   | 147        |
| 6.2.4    | Nonparametric functional regression . . . . .                                 | 154        |
| 6.2.5    | Functional additive model . . . . .   | 157        |
| <b>7</b> | <b>Applied Functional Regression</b>  | <b>163</b> |
| 7.1      | Single Functional Predictor Analysis . . . . .                                | 165        |
| 7.1.1    | Functional linear model and investigating significant relationships . . . . . | 165        |
| 7.1.2    | Nonparametric regression . . . . .  | 172        |
| 7.2      | Multiple Functional Predictor Analysis . . . . .                              | 174        |
| 7.2.1    | Functional linear model and investigating significant relationships . . . . . | 175        |
| 7.2.2    | Nonparametric regression . . . . .  | 184        |
| 7.2.3    | Functional additive model . . . . .   | 187        |
| 7.3      | Concluding Remarks . . . . .  | 191        |
| <b>8</b> | <b>Conclusions</b>  | <b>192</b> |
| 8.1      | Curvature . . . . .   | 192        |

|          |                                      |            |
|----------|--------------------------------------|------------|
| 8.2      | Functional Regression . . . . .      | 197        |
| 8.3      | Potential for Further Work . . . . . | 201        |
|          | <b>Bibliography</b>                  | <b>204</b> |
| <b>A</b> | <b>Revised Rutter Questionnaire</b>  | <b>212</b> |

# List of Tables

|     |   |     |
|-----|---|-----|
| 4.1 | Time taken to carry out the optimisation method on 5 example profiles. . . . .  | 72  |
| 5.1 | Multiple comparisons of the arc length at the <i>cphL</i> landmark. . .   | 106 |
| 5.2 | Multiple comparisons of the arc length at the <i>cphR</i> landmark. . .   | 106 |
| 5.3 | Results of Kruskal-Wallis test for difference between the groups in terms of median population curvature at the characteristic points.                              | 108 |
| 5.4 | Multiple Mann Whitney tests of the xy curvature at both characteristic points. . . . .  | 108 |
| 5.5 | Multiple comparisons of the position warping required to produce the structural average control curvature function for the upper lip.                               | 111 |
| 5.6 | Multiple comparisons of the amplitude adjustment required to produce the structural average control curvature function for the upper lip. . . . .                   | 112 |
| 6.1 | Significance of each principal component score as a predictor of psychological score. . . . .   | 123 |
| 6.2 | Significance of each smooth function of principal component score as a predictor of psychological score. . . . .  | 126 |
| 6.3 | Significance of the first two principal component scores of the four curvature functions as combined linear predictors of psychological score. . . . .              | 145 |
| 6.4 | Significance of smooth functions of the first two principal component scores of the four curvature functions as combined predictors of psychological score. . . . . | 146 |
| 6.5 | Significance of each curvature function as a predictor of psychological score in addition to the other curvature functions. . . . .                                 | 154 |
| 6.6 | Significance of each function as a predictor of psychological score in addition to the other functions in a functional additive model.                              | 161 |

|      |   |     |
|------|---|-----|
| 7.1  | Description of the two curvature functions which describe each facial curve. . . . .  | 164 |
| 7.2  | Significance of each curvature function as a predictor of psychological score using the functional linear model. . . . .  | 166 |
| 7.3  | Significance of principal component scores for nasal base xy curvature functions as combined predictors of psychological score. . .   | 167 |
| 7.4  | Significance of each curvature function as a predictor of psychological score using nonparametric regression. . . . .   | 173 |
| 7.5  | Functional linear model backward stepwise selection. The p-value is for the removed predictor whilst the $R^2$ value is from the model with that predictor removed. . . . .                             | 175 |
| 7.6  | Significance of functional parameters in final functional linear model.   | 175 |
| 7.7  | Significance of first two principal component scores for nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature functions as combined predictors of psychological score. . . . . | 178 |
| 7.8  | Nonparametric regression backward stepwise selection. The p-value is for the removed predictor whilst the $R^2$ value is from the model with that predictor removed. . . . .                            | 185 |
| 7.9  | Significance of functional parameters in the final nonparametric regression model. . . . .  | 185 |
| 7.10 | Additive model backward stepwise selection. The p-value is for the removed predictor whilst the $R^2$ value is from the model with that predictor removed. . . . .                                      | 188 |
| 7.11 | Significance of functional parameters in the generalised additive model. . . . .  | 188 |

# List of Figures

|      |   |    |
|------|---|----|
| 1.1  | The names of selected landmarks and the facial curves placed on a one year control child. . . . .   | 4  |
| 3.1  | The positions of $\mathbf{t}(s)$ , $\mathbf{n}(s)$ and $\phi(s)$ relative to an arbitrary red point $r(s)$ on a quadratic curve . . . . .   | 33 |
| 3.2  | An example midline profile of a one year control child. . . . .   | 38 |
| 3.3  | Plot of $x$ and $y$ position against arc length for the example profile. . . . .  | 39 |
| 3.4  | Plot of curvature against arc length for the example profile using different degrees of freedom for smoothing. . . . .  | 40 |
| 3.5  | Reconstructed profile aligned to the original profile for curvature when different degrees of freedom for smoothing are used. . . . .   | 42 |
| 3.6  | Plot of curvature against arc length for 71 one year old control children midline profiles. . . . .   | 43 |
| 3.7  | Average curvature curve for the 71 profiles (left). Reconstructed average profile (right). . . . .  | 44 |
| 3.8  | Kernel probability density plots and histograms for the occurrence of both maximum (left) and minimum (right) turning points. . . . .   | 48 |
| 3.9  | Warped example curvature function (left) and the corresponding warping function (right). . . . .  | 49 |
| 3.10 | Aligned curvature functions for the midline profiles of 71 one year old control children. . . . .   | 49 |
| 3.11 | Warping functions (left) and warping function minus arc length (right), for the curvature functions for the midline profiles the 71 control children. . . . .   | 50 |
| 3.12 | Raw and structural average for the curvature functions for the midline profiles of 71 one year old control children (left) and the reconstructed profiles from these curvature functions (right). . . . . | 51 |

|      |  |    |
|------|--|----|
| 3.13 | Amplitude adjustment functions to produce the structural average curvature function from the individual curvature functions for the midline profiles of 71 one year old control children. . . . .  | 52 |
| 3.14 | Comparison of the curvature of the example profile calculated using the derivative and the optimisation method (left) and the reconstruction of the profile using the curvature calculated by the optimisation method (right). . . . .                   | 54 |
| 3.15 | Comparison of the curvature of the example profile calculated using the derivative and the smoothed optimisation method (left) and the reconstruction of the profile using the curvature calculated by the smoothed optimisation method (right). . . . . | 54 |
| 3.16 | Comparison of the curvature of the example profile calculated using the derivative and the Frenet method (left) and the reconstruction of the profile using the curvature calculated by the Frenet method (right). . . . .                               | 57 |
| 4.1  | Plot of the example profile from side on (left). Plot of example profile from straight on (right). . . . .   | 68 |
| 4.2  | Plot of curvature (left) and torsion (right) function for the example profile. . . . .   | 69 |
| 4.3  | Plot of $y = x^5$ and the comparison between the estimate and the actual values of the first, second and third derivatives. . . . .  | 70 |
| 4.4  | Plot of both the raw and smoothed curvature and torsion functions for the example profile calculated using the optimisation method. . . . .  | 73 |
| 4.5  | Plot of the reconstruction of the example profile using the raw and smooth curvature and torsion functions calculated using optimisation from side on (left) and front on (right). . . . .   | 74 |
| 4.6  | Plot of curvature (black) and torsion (red) functions for the example profile. . . . .   | 77 |
| 4.7  | Plot of the reconstruction of the example profile using the curvature and torsion functions calculated using the Frenet method from side on (left) and front on (right). . . . .   | 77 |
| 4.8  | The curves defining the face of a child projected onto the yz (top left), xy (top right) and xz(bottom left) planes. . . . .   | 79 |
| 4.9  | Plot of the yz (left) and xy (right) plane projections for the example profile. . . . .  | 81 |

|      |  |     |
|------|--|-----|
| 4.10 | Plot of yz and xy curvature functions for the example profile where $s$ is the arc length of the space curve. . . . .  | 82  |
| 4.11 | Plot of the reconstructed yz (left) and xy (right) plane projections for the example profile. . . . .  | 82  |
| 4.12 | Plot of the reconstructed profile matched to the original profile from side on (left) and front on (right). . . . .  | 84  |
| 4.13 | Plot of yz and xy curvature against arc length for all 71 profiles. . . . .  | 85  |
| 4.14 | Plot of average yz and xy curvature functions for all profiles (top left) and the corresponding reconstructed average profile from side on (top right) and front on (bottom left). . . . .   | 86  |
| 4.15 | Kernel probability density plots and histograms for the occurrence of both maximum and minimum turning points of the curvature curves. . . . .   | 90  |
| 4.16 | Plot of the actual and warped yz curvature functions (top left) and actual and warped xy curvature functions (top right) for the example one year old control children midline profile. The warping function (bottom left). . . . .  | 91  |
| 4.17 | Plot of the warped yz and xy curvature functions (top left). The warping functions (top right), to align these functions, and the warping function minus arc length (bottom left). . . . .   | 92  |
| 4.18 | Plot of raw and structural average yz curvature against arc length (left) and raw and structural average xy curvature against arc length (right). . . . .  | 93  |
| 4.19 | Reconstructed average profile using the raw and structural average from side on (left) and front on (right). . . . .   | 94  |
| 4.20 | Amplitude adjustment functions to produce the structural average yz curvature function from individual yz curvature functions of the control midline profiles (left). Amplitude adjustment functions to produce the structural average xy curvature function from individual xy curvature functions of the control midline profiles (right). . . . . | 94  |
| 5.1  | Plot of xz (left) and xy (right) curvature functions for all upper lips. . . . .   | 101 |
| 5.2  | Raw average xz curvature (dashed lines) and xy curvature (solid lines) functions for the upper lips of each group. . . . .   | 102 |

|      |   |     |
|------|---|-----|
| 5.3  | Plot of unaligned (left) and aligned (right) total curvature against arc length for all upper lips. . . . .   | 103 |
| 5.4  | Structural average xz (dashed lines) and xy (solid lines) curvature functions for the upper lips of each group. . . . .   | 104 |
| 5.5  | The reconstructed average upper lip from the structural averages of each group in the xz plane (left) and xy plane (right). . . . .   | 104 |
| 5.6  | Arc length at the start (left) of the Cupid's bow plotted against arc length at the end (right) of the Cupid's bow. . . . .   | 105 |
| 5.7  | Curvature at the start of the Cupid's bow against curvature at the end of the Cupid's bow for xz (left) and xy (right) curvature. . .   | 107 |
| 5.8  | Position warping functions (left), and position warping function minus arc length (right), to align to the position of the average characteristic points of the upper lip of control children. . . . .  | 109 |
| 5.9  | Amplitude adjustment functions to produce the structural average xz curvature function from individual xz curvature functions (left) and amplitude warping functions to produce the structural average xy curvature function from individual xy curvature functions (right). . . . .  | 110 |
| 5.10 | Boxplot of the sum of squares difference between the position warping function and the line of equality for the upper lips for each group (left) and boxplot of the sum of squares difference between the amplitude adjustment function and the $x = 0$ line for each group for both xz and xy curvature functions of upper lips (right). . . . . | 111 |
| 5.11 | The effect of the first two principal components on the average curve for the unaligned (top row) and aligned (bottom row) xz curvature functions. The solid line is the mean function, the dashed line is $mean + 3\sqrt{\lambda_i \mathbf{e}_i}$ while the dotted line is $mean - 3\sqrt{\lambda_i \mathbf{e}_i}$ . . .                         | 114 |
| 5.12 | The effect of the first two principal components on the average curve for the unaligned (top row) and aligned (bottom row) xy curvature functions. The solid line is the mean function, the dashed line is $mean + 3\sqrt{\lambda_i \mathbf{e}_i}$ while the dotted line is $mean - 3\sqrt{\lambda_i \mathbf{e}_i}$ . . .                         | 115 |
| 5.13 | Upper lip first and second principal component scores for the unaligned xz curvature functions (top left), aligned xz curvature functions (top right), unaligned xy curvature functions (bottom left) and the aligned xy curvature functions (bottom right). . . . .  | 116 |
| 6.1  | Boxplot of psychological scores for cleft children (left). Plot of yz curvature against arc length for each subject. (right) . . . . .  | 120 |

|      |  |     |
|------|--|-----|
| 6.2  | Psychological score against first principal component score. . . . .   | 121 |
| 6.3  | The <code>rpanel</code> showing; the nonparametric fit of the relationship between the first principal component score and psychological score (top left), the yz curvature function corresponding to that component score (top right) and the reconstructed midline profile using that curvature function (bottom right). The left <code>rpanel</code> shows a low component score whilst the right <code>rpanel</code> shows a high component score. . . . . | 122 |
| 6.4  | Principal component score for the first three principal components against psychological score. . . . .  | 124 |
| 6.5  | Additive model functions for the relationship between principal component score for the first three principal components and psychological score for all subjects. The solid lines are the fitted function of component score whilst the dashed lines are the standard errors. . . . .   | 125 |
| 6.6  | Cross-validation score function. . . . .   | 134 |
| 6.7  | Functional parameter for the functional linear model with psychological score as response and yz curvature function of the midline profile as predictor (the red lines indicate confidence limits for the parameter) and estimated psychological score using this model against the true psychological score with the line of equality for reference. . . . .  | 135 |
| 6.8  | Plot of the p-value from the test of no effect of the nonparametric regression model using various numbers of nearest neighbours (left) and estimated psychological score using nonparametric regression with 60 nearest neighbours against the true psychological score with the line of equality for reference (right). . . . .  | 141 |
| 6.9  | Curvature functions for the cleft subjects including; xy curvature of the midline profile (top left), xz curvature of the upper (top right) and xy curvature of the upper lip (bottom left). . . . .   | 142 |
| 6.10 | Matrix of scatterplots for psychological score and first principal component score of the four functional predictors. . . . .  | 143 |
| 6.11 | The <code>rpanel</code> and three-dimensional plot with best surface. Function 1 corresponds to xy curvature of the midline profile, function 2 corresponds to xy curvature of the upper lip and response corresponds to psychological score. . . . .  | 144 |

|      |   |     |
|------|---|-----|
| 6.12 | Additive model functions for first and second principal component score predictors of; yz curvature of the midline profile (function 1), xy curvature of the midline profile (function 2), xz curvature of the upper lip (function 3) and xy curvature of the upper lip (function 4). . . . .   | 147 |
| 6.13 | Cross-validation score function for the multiple predictor functional linear model. . . . .   | 151 |
| 6.14 | Functional parameters for the functional linear model with psychological score as response and; yz curvature function of the midline profile (top left), xy curvature function of the midline profile (top right), xz curvature function of the upper lip (bottom left), xy curvature function of the upper lip (bottom right) as predictors. The red lines indicate confidence limits for the parameter. . . . . | 152 |
| 6.15 | Estimated psychological score using the functional linear model with multiple functional predictors against the true psychological score with the line of equality for reference. . . . .   | 153 |
| 6.16 | Plot of the p-value from the test of no effect of the nonparametric regression model with four functional predictors using various numbers of nearest neighbours (left) and estimated psychological score using nonparametric regression with 20 nearest neighbours against the true psychological score (right). . . . .   | 157 |
| 6.17 | Plot of the p-value from the test of no effect of the additive model with four functional predictors using various numbers of nearest neighbours (left) and estimated psychological score using a generalised additive model with 25 nearest neighbours against the true psychological score (right). . . . .   | 160 |
| 6.18 | Plot of first principal component score against value returned by the additive model for predictors; yz (top left) and xy (top right) curvature of the midline profile and xz (bottom left) and xy (bottom right) curvature of the upper lip. . . . .   | 161 |
| 7.1  | Matrix plot of first principal component score for all functional predictors and scalar response. . . . .   | 164 |
| 7.2  | xy curvature functions for the nasal base and the effect (+/- two standard deviations of the principal component) of the first three principal components on the mean curvature function. . . . .   | 167 |

|      |   |     |
|------|---|-----|
| 7.3  | The functions of the principal components scores returned by the additive model. . . . .  | 168 |
| 7.4  | The <code>rpanel</code> showing; the nonparametric fit of the relationship between the second principal component score and psychological score (top left), the curvature function corresponding to that component score (top right) and the reconstructed nasal base using that curvature function (bottom right). The left <code>rpanel</code> shows a low component score whilst the right <code>rpanel</code> shows a high component score. . . . . | 169 |
| 7.5  | Cross-validation score function for the xy curvature of the nasal base predictor. . . . .   | 170 |
| 7.6  | Predicted psychological score, using the functional linear model, against true psychological score (left) and the functional parameter with red confidence bands (right). . . . .   | 170 |
| 7.7  | The <code>rpanel</code> showing; the nonparametric fit of the relationship between the first principal component score and psychological score (top left), the curvature function corresponding to that component score (top right) and the reconstructed nasal base using that curvature function (bottom right). The left <code>rpanel</code> shows a low component score whilst the right <code>rpanel</code> shows a high component score. . . . .  | 171 |
| 7.8  | Plot of the p-value from the test of no effect (left) and $R^2$ value (right) returned from a nonparametric regression using xy curvature of the nasal base and various numbers of nearest neighbours. . . . .  | 173 |
| 7.9  | Predicted psychological score, using nonparametric regression, against true psychological score. . . . .  | 174 |
| 7.10 | Curvature functions and effect (+/- two standard deviations) of the first two principal components for xy curvature of the nasal rim (left) and xz curvature of the nasal bridge (right). . . . .   | 177 |
| 7.11 | The functions of the principal components scores returned by the generalised additive model for nasal rim xy curvature (top), nasal base xy curvature (middle) and nasal bridge xz curvature (bottom). . . . .  | 179 |

|      |   |     |
|------|---|-----|
| 7.12 | The <code>rpanel</code> and three-dimensional plot with best surface giving a high psychological score (top) and a low psychological score (bottom). Function 1 corresponds to xy curvature of the nasal rim, function 2 corresponds to xz curvature of the nasal bridge and response corresponds to psychological score. . . . . | 180 |
| 7.13 | Cross-validation score function for the final functional linear model.  | 181 |
| 7.14 | The functional parameters returned by the functional linear model for xy curvature of the nasal rim (top left), xy curvature of the nasal base (top right), xz curvature of the nasal bridge (bottom left) and xy curvature of the nasal bridge (bottom right). . . . .   | 182 |
| 7.15 | The functional parameters returned by the functional linear model superimposed on the curvature functions for xy curvature of the nasal rim (top left), xy curvature of the nasal base (top right), xz curvature of the nasal bridge (bottom left) and xy curvature of the nasal bridge (bottom right). . . . .                   | 183 |
| 7.16 | Plot of the p-value from the test of no effect (left) and $R^2$ value (right) returned from a nonparametric regression using xy curvature of the midline profile, xy curvature of the nasal rim, xy curvature of the nasal base and xz curvature of the nasal bridge as predictors. . . . .                                       | 186 |
| 7.17 | Predicted psychological score against true psychological score using nonparametric regression (left) and functional linear model (right). . . . .   | 187 |
| 7.18 | Plot of first principal component score against value returned by the additive model function for predictors; xy curvature of nasal rim (left) and xy curvature of nasal base (right). . . . .  | 189 |
| 7.19 | Plot of the p-value from the test of no effect (left) and $R^2$ value (right) returned from a generalised additive model using xy curvature of the nasal rim and xy curvature of the nasal base. . . . .  | 190 |
| 7.20 | Estimated psychological score, using an additive model, against true psychological score. . . . .   | 190 |

# Chapter 1

## Introduction

### 1.1 Shape

Much of our initial perception of objects, creatures and fellow humans is dependent on their shape. The human brain is supremely efficient at analysing shape instantly. However, to allow shape analysis to be considered as a useful research area it is important to have systematic techniques to qualitatively describe and compare shapes. Techniques currently used to describe shapes include placing landmarks, curves or meshes on the shape. Landmarks can be placed on geometric or anatomically important points on shapes and these can be used to either define the shape or compare between shapes. One drawback with simply using landmarks to define the shapes is that information between the landmarks is lost. To overcome this loss of information, curves on the shape which may be of interest can be defined by connecting a set of pseudo-landmarks which lie on the curve at small increments between the anatomical landmarks. To define the whole shape it is possible to place a mesh over the complete surface of the shape. Often landmarks and curves are extracted from these surface meshes but there are also techniques to analyse the shape of the whole surface through these meshes. This thesis will solely analyse shape curves.

Much of the current shape analysis literature involves the alignment of shapes by location, rotation and in some instances scale. Dryden and Mardia (1998) describe a variety of methods for aligning shapes, with the popular Procrustes methods described in Chapter 2. The methods are in the main simple to use and make further analysis of shapes straightforward; however there are sometimes difficulties. If there is a large number of shapes defined in high definition, alignment methods can prove computationally expensive. Furthermore, care must be

taken when comparing aligned shapes to ensure that particular differences are in fact discrepancies between the shapes and cannot be attributed to the alignment process. Throughout this thesis shapes will be described by the bending of defined shape curves. An advantage of this method is that it removes the need for shape alignment.

## 1.2 Curvature

One technique for defining the shape of objects is to calculate the bending experienced by shape curves. For plane curves the bending of the curve at any point is given by a single scalar value, called curvature, whilst for space curves the bending at any point on the curve is defined by two scalar values, called curvature and torsion (or in some literature first and second curvature). A major reason for defining shapes in this way is that curvature is independent of the location or rotation of the curve meaning alignment techniques are not required.

There are a number of techniques which may be used to calculate the bending of a curve. A simple technique uses formulae which are dependent on the derivatives of the position of the curve on each axis with respect to the arc length of the curve. The bending of a plane curve (called curvature and denoted by  $\kappa$ ) is defined by the equation

$$\kappa(s) = \frac{x'(s)y''(s) - x''(s)y'(s)}{(x'(s)^2 + y'(s)^2)^{3/2}}$$

where  $x(s)$  and  $y(s)$  are the  $x$  and  $y$  positions of the curve at arc length  $s$  and the dash notation indicates derivatives with respect to the arc length of the curve. The bending of a space curve (called curvature and torsion which are denoted by  $\kappa$  and  $\tau$  respectively) is defined by the equations

$$\begin{aligned} \kappa(s) &= |r'(s) \times r''(s)| \\ \tau(s) &= \frac{((r'(s) \times r''(s)) \cdot r'''(s))}{|r'(s) \times r''(s)|^2} \end{aligned}$$

where  $r'(s) = [x'(s), y'(s), z'(s)]$  and  $\times$  denotes the cross-product.

For simple shape comparison it is useful to compare the functions of curvature (and torsion) against arc length of the various shapes. Also, by defining the bending of each feature curve as functions of arc length, a large array of functional data analysis techniques (Ramsay and Silverman (2006) give an overview

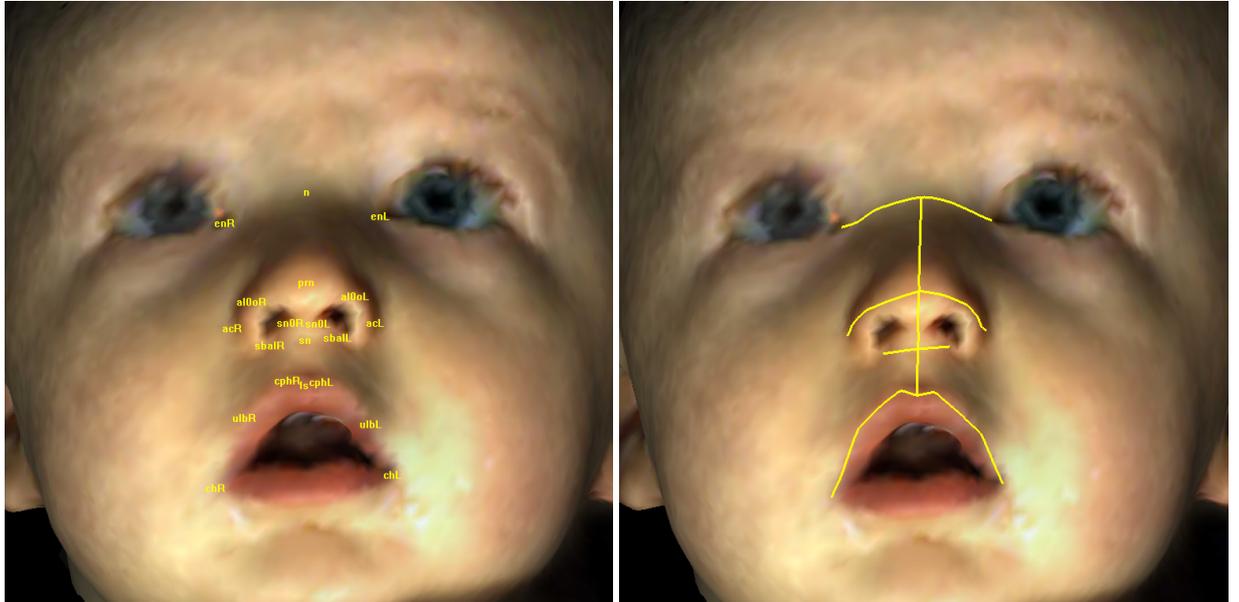
of a selection of techniques) are available for further analysis. One of the major interests in this thesis will be comparing the feature curves of different groups using curvature functions. Interestingly, both plane curves and space curves can be reconstructed up to location and rotation from the curvature (and torsion) functions with relative ease. This can be useful in several settings, for example, it enables ‘average’ feature curves to be produced from ‘average’ curvature functions which can assist in anatomical visualisation of the nature of differences in the curvature functions.

Curvature functions can be useful in extracting the position of anatomically or geometrically important points on the shape. Points of interest will often be associated with turning points of the curvature function. By considering the turning points of the curvature functions information is available on the position (in relation to the rest of the feature) and the amount of bending experienced at important points on the shape. This thesis will investigate a number of techniques which use the fact that curvature functions indicate anatomical points of interest, which complement and enhance methods of group comparison.

### 1.3 Cleft Lip and Palate Data

The data analysed in this thesis come from two studies comparing the facial shape of children in the Glasgow area with particular interest in comparing the facial shapes of children with a cleft lip and/or palate to that of control children. One study investigated children at 1 year of age while a separate study investigated children at 10 years. In both studies three-dimensional images were built using a sophisticated stereophotogrammetry system, which takes pictures from two different angles and uses these to build up the three-dimensional image much in the way two eyes build up the image in human sight. Validation studies were carried out by Ayoub et al. (2003) and Khambay et al. (2008) on the equipment used in the 1 year and 10 year studies respectively with discussion on the variability in the three-dimensional images produced and the variability in the manual placement of the landmarks. The three-dimensional image is transformed into a surface mesh onto which landmarks and curves can be placed. Figure 1.1 shows an example of a one year control child with selected landmarks and curves marked. There are data available on five facial curves; the midline profile, the top of the upper lip, the nasal rim, the nasal base and the nasal bridge. The curves are produced by marking important anatomical landmarks manually and

calculating pseudo-landmarks at many small increments between landmarks to give close to continuous curves.



**Figure 1.1:** The names of selected landmarks and the facial curves placed on a one year control child.

In the 1 year study there are curves available for 9 children born with a cleft lip and 13 children born with a cleft lip and palate as well as for 71 control children. All the cleft children underwent surgical repair between the age of 3 and 6 months. The major interest is in comparing the facial shapes of the three groups of children, in particular comparing between the control group and both cleft groups. It is clearly the aim of the surgery to attempt to, where possible, remove all shape effects of the cleft, therefore it is important to investigate systematic differences between the control group and either of the cleft groups.

In the 10 year study there are curves available for 44 children born with a cleft lip and 51 children born with a cleft lip and palate as well as 68 control children. Again all the cleft children underwent surgical repair between the age of 3 and 6 months. In addition to the digital imaging all cleft children, and their parents, were asked to fill in a set of questionnaires on their psychological state. There were numerous aims to the study but the aim discussed in this thesis is to investigate potential relationships between the facial shape of the cleft children, as described by curvature functions, and their psychological score.

In general the cleft will only occur on one side of the child's face with no evidence of a tendency to occur more regularly on the right or the left side.

While the cleft will affect both sides of the face, clearly the side on which the cleft occurs will be most affected. The side of the face on which the cleft occurs is of no interest in this study. In fact it would be beneficial to have the cleft occur on the same side of the face for each child. To facilitate this all cleft children who have the cleft occurring on the left side of the face have the landmarks and corresponding curves ‘reflected’ around the midline so that in the analysis the cleft is seen on the right side of the face.

## 1.4 Psychological Questionnaire

Although the children completed a number of questionnaires it is a questionnaire completed by the parents which will be used throughout this thesis. The reason for this is that the questionnaires completed by the parents contained much fewer missing responses allowing a larger data set to be analysed. The questionnaire used was the Revised Rutter Scale which is a revision of the original Rutter Parents’ and Teachers’ Scale (see Rutter (1967)). A copy of the questionnaire can be found in Appendix A.

The questionnaire contains 50 statements (for example ‘Very restless, has difficulty staying seated for long’ or ‘Not much liked by other children’) and the parents were asked to select whether the statement ‘does not apply’, ‘applies somewhat’ or ‘certainly applies’. Each statement is scored 0 for ‘does not apply’, 1 for ‘applies somewhat’ and 2 for ‘certainly applies’. By summing the scores for certain statements a total difficulties score out of 52 is obtained with higher scores suggesting more difficulties. Furthermore, scores for emotional difficulties, conduct difficulties, hyperactivity/inattention and prosocial behaviour can be obtained. Throughout this thesis the total difficulties score will be used as a measure of psychological score.

Although there has been little review of the updated version of the questionnaire it is clearly strongly linked to the original Rutter Parents’ and Teachers’ Scale and earlier adaptations which were reviewed by Elander and Rutter (1996). They concluded that ‘for reliability the picture is of generally positive results that are better for antisocial than emotional behaviours and better for teachers’ than parents’ ratings.’ They also show that the results from the Rutter Parents’ and Teachers’ Scales correlate well with several social competency scales and a variety of observational studies.

## 1.5 Missing Data

Both studies incur some levels of missingness for a variety of reasons. One problem is in accurately extracting curves, particularly for cleft children. Barry (2008) briefly discusses missing data for the 1 year old study. The major reason for missing data in this study was difficulty in extraction of the curves, particularly for the cleft group. It may be that curve extraction is more difficult for the most severe cases so data may be missing for these severe cases. This will not be considered throughout the thesis but should be kept in mind. The numbers quoted as part of the study in Section 1.3 are subjects for whom complete sets of curves were available. Due to difficulties in curve extraction there is data missing for 24 (26.4%) control children and 19 (47.5%) cleft children.

The 10 year study initially contained 95 cleft subjects. However, due to problems with curve extraction, 8 subjects were removed. Furthermore, only subjects for which there are completed questionnaires can be included in the analysis. When parents attended the data capture session it was requested that they completed the questionnaire. If they refused or were reticent no pressure was applied to ensure completion. There did not appear to be any systematic reasoning for non-completion of questionnaires so subjects with no psychological score available are simply removed from the analysis and no further consideration is given to this issue. Due to missing questionnaires a further 7 subjects are removed. This leaves 80 cleft subjects from which to investigate the relationship between psychological score and facial shape. The decision was made to simply analyse a single cleft group rather than the separate cleft lip and cleft lip and palate groups as the larger size of the group allows sufficient information from which conclusions can be drawn.

## 1.6 Overview of Thesis

Chapter 2 will outline some simple shape analysis techniques and discuss some recent developments in the field. Furthermore, there will be a discussion of some curve analysis methods along with an overview of standard functional data analysis techniques. Chapter 3 will introduce methods for calculating the curvature function of a plane curve. Techniques for producing average curvature functions, including aligning the functions in terms of important landmarks, will also be discussed. The methods for calculating the curvature of a plane curve are extended

to a space curve in Chapter 4 and furthermore a method to aid anatomical interpretation of the curvature functions is introduced. Chapter 5 uses the methods introduced in the previous chapters to carry out a comparison between the three groups of 1 year old children (controls, cleft lip, cleft lip and palate) concentrating specifically on the upper lip curves. Methods to examine a relationship, both graphically and formally, between facial shape defined using curvature functions and psychological score are outlined in Chapter 6 with a motivating example contained for illustration. Chapter 7 contains a systematic investigation of the relationships between facial shape and psychological score using the data of 10 year old children. Chapter 8 provides a summary, some conclusions and potential further work.

This thesis focuses on the practical application of analysing the cleft lip studies by carrying out analysis which is informative in this context. The major aims are:

- To introduce a curvature based technique to describe the facial shape of children through the bending of important facial features.
- To produce curvature functions for the facial features which are anatomically simple to understand.
- To compare the facial shape of one year old children born with a cleft lip and/or palate to control children using these curvature functions.
- To investigate relationships between facial shape, defined by curvature functions, and psychological score for ten year old children born with a cleft lip and/or palate.
- To introduce methods which may be useful in interpreting any significant relationship between facial shape and psychological score.

Many of the techniques chosen throughout are driven by the aims of the thesis.

# Chapter 2

## Review

This chapter will provide a brief overview of the statistical shape analysis techniques which will be used in this thesis. A full overview of statistical shape analysis can be found in Dryden and Mardia (1998) and this work is the basic reference for the first section of the chapter. Much of Section 2.1 will discuss popular shape analysis techniques when the shapes are defined by landmarks. Defining shapes solely by landmarks will not be the method adopted by this thesis; however, the methods are commonly used and useful to understand. Much of the work in this thesis will involve analysing shapes which are defined by curves. Section 2.2 will investigate how two- and three-dimensional curves have been analysed in the current literature with a particular interest in the use of curves to describe and analyse shapes. Section 2.3 will describe some commonly used techniques to analyse functional data as much of the data to be analysed in this thesis will be in functional form.

### 2.1 Statistical Shape Analysis

Dryden and Mardia (1998) define shape as ‘all the geometrical information which remains when location, scale and rotational effects are filtered out from an object’. It is usual for shapes to be defined by pre-specified landmarks which provide a correspondence between and within populations. These landmarks are either anatomical landmarks which have a specific biological meaning, such as the tip of the nose or the edge of the upper lip, mathematical landmarks, such as points of high curvature or turning points, or pseudo-landmarks which are used to connect other types of landmarks.

To define a set of landmarks on an object a  $k \times m$  matrix,  $X$  say, is produced

where  $k$  is the number of landmarks and  $m$  is the number of dimensions. This matrix is called the configuration matrix for that object.

When comparing shapes it is useful to scale each object to the same size. For this to be possible the size of each shape must be defined. Dryden and Mardia (1998) define a size measure  $g(X)$  as any positive real valued function of the configuration matrix such that

$$g(aX) = ag(X) \quad (2.1)$$

for any positive scalar  $a$ . One measure of size which satisfies (2.1) is the centroid size. The centroid size of a shape with configuration matrix  $X$  is given by

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_j)^2} \quad (2.2)$$

where  $\bar{X}_j = \frac{1}{k} \sum_{i=1}^k X_{ij}$ ,  $C = I_k - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T$  and  $\|X\| = \sqrt{\text{tr}(X^T X)}$ . Equation (2.2) effectively states that the centroid size is the square root of the sum of squared distances of each point from the mean point in each dimension.

### 2.1.1 Procrustes analysis

Procrustes methods are popular techniques used to remove the effects of location, rotation and scale for configurations with two or more dimensions. By removing these three effects all that remains is information on the shape given by the configuration. Procrustes analysis is the process which matches configurations by using least squares to minimise the Euclidean distance between them following centring (location adjustment), rotation and scaling. There are two major methods of Procrustes analysis. Full ordinary Procrustes analysis (OPA), which matches two configurations, and full generalised Procrustes analysis (GPA), which matches  $n$  configurations. Both methods will be outlined here.

To give an overview of full ordinary Procrustes analysis suppose there are two configurations  $X_1$  and  $X_2$  which contain information on  $k$  landmarks in  $m$  dimensions. The first stage of the process is to centre the configurations using

$$X_{iC} = CX_i \quad (2.3)$$

where  $C$  is as defined above. For simplicity  $X_i$  will denote  $X_{iC}$  after the configurations have been centred. Now the full OPA method involves minimising

$$D_{OPA}^2(X_1, X_2) = \| X_2 - \beta X_1 \Gamma - \mathbf{1}_k \gamma^T \|^2 \quad (2.4)$$

where  $\Gamma$  is an  $(m \times m)$  rotation matrix,  $\beta > 0$  is a scale parameter and  $\gamma$  is an  $(m \times 1)$  location vector. The minimum of (2.4) is the ordinary (Procrustes) sum of squares ( $OSS(X_1, X_2)$ ). The parameter values are given by  $(\hat{\gamma}, \hat{\beta}, \hat{\Gamma})$  where

$$\begin{aligned} \hat{\gamma} &= 0 \\ \hat{\Gamma} &= UV^T \\ \hat{\beta} &= \frac{\text{tr}(X_2^T X_1 \hat{\Gamma})}{\text{tr}(X_1^T X_1)} \end{aligned}$$

where  $U, V \in SO(m)$  and  $SO(m)$  is the set of  $(m \times m)$  orthogonal matrices,  $\Lambda$ , where  $\Lambda^T \Lambda = \Lambda \Lambda^T = I_m$ . The ordinary (Procrustes) sum of squares can be calculated as  $OSS(X_1, X_2) = \| X_2 \|^2 \sin^2 \rho(X_1, X_2)$  where  $\rho(X_1, X_2)$  is the Procrustes distance defined by Dryden and Mardia (1998). The full Procrustes fit of  $X_1$  onto  $X_2$  is then given by

$$X_1^P = \hat{\beta} X_1 \hat{\Gamma} + \mathbf{1}_k \hat{\gamma}^T \quad (2.5)$$

The residual matrix after the Procrustes matching can be defined as

$$R = X_2 - X_1^P$$

If the roles of  $X_1$  and  $X_2$  are reversed then the ordinary Procrustes superimposition will be different. Therefore the ordinary Procrustes fit is not reversible i.e.  $OSS(X_1, X_2) \neq OSS(X_2, X_1)$  unless the objects are both of the same size. Therefore,  $\sqrt{OSS(X_1, X_2)}$  cannot be used as a distance measure between the shapes. Instead normalising the shapes to unit size gives a useful measure of distance between the two shapes. This distance is given by  $\sqrt{OSS\left(\frac{X_1}{\|X_1\|}, \frac{X_2}{\|X_2\|}\right)}$ .

To give an overview of full generalised Procrustes analysis (GPA) suppose that there are  $n$  configurations  $X_1, \dots, X_n$  which each contain information on  $k$  landmarks in  $m$  dimensions. Once again assume that each configuration is centred. Full GPA can be thought of as a direct extension of full OPA such that

the full GPA minimises the generalised (Procrustes) sum of squares

$$G(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \| (\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T) - (\beta_j X_j \Gamma_j + \mathbf{1}_k \gamma_j^T) \|^2 \quad (2.6)$$

subject to the constraint  $S(\bar{X}) = 1$ . That is, the centroid size of the average configuration is 1 where the average configuration is  $\bar{X} = \frac{1}{n} \sum_{i=1}^n (\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T)$ . The generalised (Procrustes) sum of squares is proportional to the sum of squared norms of pairwise differences. Minimising the generalised (Procrustes) sum of squares involves translating, rotating and rescaling each object so that all objects are placed close to each other in a way which minimises the sum, over all pairs, of the squared Euclidean distances. This process can be defined by

$$\begin{aligned} G(X_1, \dots, X_n) &= \inf_{\beta_i, \Gamma_i, \gamma_i} \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \| (\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T) - (\beta_j X_j \Gamma_j + \mathbf{1}_k \gamma_j^T) \|^2 \\ &= \inf_{\beta_i, \Gamma_i, \gamma_i} \sum_{i=1}^n \| (\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T) - \frac{1}{n} \sum_{j=1}^n (\beta_j X_j \Gamma_j + \mathbf{1}_k \gamma_j^T) \|^2 \end{aligned}$$

The minimisation can alternatively be viewed from the perspective of estimation of the mean shape  $\mu$  so

$$\begin{aligned} G(X_1, \dots, X_n) &= \inf_{\mu: S(\mu)=1} \sum_{i=1}^n OSS(X_i, \mu) \\ &= \inf_{\mu: S(\mu)=1} \sum_{i=1}^n \sin^2 \rho(X_i, \mu) \end{aligned}$$

where  $\rho(X_i, \mu)$  is the Procrustes distance between  $X_i$  and  $\mu$  given by Dryden and Mardia (1998). The full Procrustes fit of each  $X_i$  is given by

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + \mathbf{1}_k \hat{\gamma}_i^T, \quad i = 1, \dots, n$$

Dryden and Mardia (1998) describe an algorithm for estimating the transformation parameters  $(\beta_i, \Gamma_i, \gamma_i)$ . The full Procrustes estimate of mean shape can be calculated by

$$\hat{\mu} = \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^n \sin^2 \rho(X_i, \mu)$$

$$= \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^n d_F^2(X_i, \mu)$$

where  $d_F^2(X_i, \mu)$  is the squared full procrustes distance, as defined by Dryden and Mardia (1998), between  $X_i$  and  $\mu$ . It is also possible to calculate the full Procrustes mean by calculating the arithmetic mean of each coordinate, across each configuration after full Procrustes matching. Therefore,  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i^P$  where  $X_i^P$  is the Procrustes coordinates for individual  $i$ .

### 2.1.2 Thin-plate splines, deformations and warping

A quantity such as Procrustes distance can give a numerical measure to compare two shapes. However, it is often the case that the interest is more in how shapes differ locally as opposed to simply by how much they differ. To investigate these local differences it can be informative to map one landmark configuration onto another. Suppose that there are two configurations,  $T$  and  $Y$ , both defined by  $k$  landmarks in  $m$  dimensions such that  $T = (t_1, \dots, t_k)^T$  and  $Y = (y_1, \dots, y_k)^T$ . The aim is to map  $T$  onto  $Y$  where  $t_j, y_j \in \mathbb{R}^m$ . This process is called a deformation and is defined by the transformation

$$y = \Phi(t) = (\Phi_1(t), \Phi_2(t), \dots, \Phi_m(t))^T$$

where the multivariate function  $\Phi(t)$  should, where possible, be continuous, smooth, bijective, not prone to large distortions, equivariant under the similarity transformations and an interpolant i.e.  $y_j = \Phi(t_j) \forall j = 1, \dots, k$ . In two dimensions, when  $m = 2$ , deformation can be carried out using a pair of thin-plate splines. Dryden and Mardia (1998) state that ‘a bijective thin-plate spline is analogous to a monotone cubic spline’. A pair of thin-plate splines can be given by the bivariate function

$$\begin{aligned} \Phi(t) &= (\Phi_1(t), \Phi_2(t))^T \\ &= c + At + W^T s(t) \end{aligned}$$

where  $t$  is a  $(2 \times 1)$  vector,  $s(t) = (\sigma(t - t_1), \dots, \sigma(t - t_k))^T$  is a  $(k \times 1)$  vector with

$$\sigma(h) = \begin{cases} \|h\|^2 \log(\|h\|), & \|h\| > 0 \\ 0, & \|h\| = 0 \end{cases}$$

Incorporating some necessary constraints this can be written in vector-matrix form as

$$\begin{bmatrix} S & \mathbf{1}_k & T \\ \mathbf{1}_k^T & 0 & 0 \\ T^T & 0 & 0 \end{bmatrix} \begin{bmatrix} W \\ c^T \\ A^T \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}$$

where  $S_{ij} = \sigma(t_i - t_j)$ . The inverse of the matrix  $\Gamma$  where

$$\Gamma = \begin{bmatrix} S & \mathbf{1}_k & T \\ \mathbf{1}_k^T & 0 & 0 \\ T^T & 0 & 0 \end{bmatrix}$$

can be written, since  $\Gamma$  is symmetric positive definite, as

$$\Gamma^{-1} = \begin{bmatrix} \Gamma^{11} & \Gamma^{12} \\ \Gamma^{21} & \Gamma^{22} \end{bmatrix}$$

where  $\Gamma^{11}$  is  $(k \times k)$ . The bending energy matrix,  $B_e$  is then defined as  $B_e = \Gamma^{11}$ .

One way to explain the components of a thin-plate spline deformation is to use principal and partial warps. These techniques were introduced by Bookstein (1989). The idea of principal and partial warps is somewhat analogous to principal components in a multivariate context in that each principal and partial warp explains a separate part of the overall deformation. Suppose that  $T$  and  $Y$  are  $(k \times 2)$  configuration matrices for different shapes and the thin-plate spline transformation which interpolates the  $k$  points of  $T$  to  $Y$  gives a  $(k \times k)$  bending energy matrix  $B_e$ . The principal warps, which construct an orthogonal basis for re-expressing the thin-plate spline transformations, can then be defined as

$$P_j(t) = \gamma_j^T s(t)$$

for  $j = 1, \dots, k - 3$  where  $\gamma_1, \dots, \gamma_{k-3}$  are the eigenvectors corresponding to the non-zero eigenvalues  $(\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{k-3})$  obtained from an eigen-decomposition of  $B_e$ . Further  $s(t) = (\sigma(t - t_1), \dots, \sigma(t - t_k))^T$ . Partial warps can now be defined as

$$R_j(t) = Y^T \lambda_j \gamma_j P_j(t)$$

The  $j$ th partial warp will largely show movement of landmarks which are heavily weighted in the  $j$ th principal warp. In general as the eigenvalue which corresponds to the warps increases the more local the deformation described by the

warp becomes i.e.  $P_1(t)$  and  $R_1(t)$  will typically explain an overall large scale deformation of the shape whilst  $P_{k-3}(t)$  and  $R_{k-3}(t)$  will explain a small scale and localised deformation often between the two closest landmarks. The partial warps are useful to give greater understanding of the deformation explained by the corresponding principal warps.

### 2.1.3 Other shape analysis techniques

Shape analysis is a large and progressive subject in Statistics which has seen many advances in recent years. Bowman and Bock (2006) discuss a number of techniques to explore three-dimensional shape, including graphical displays of longitudinal changes between groups and a permutation test to compare principal components of groups across time, with comparing the facial shapes of control children and children with a unilateral cleft lip and/or palate used as an example. Further, Bock and Bowman (2006) introduce a method to measure the asymmetry of the faces of children whilst Pazos et al. (2007) investigate the reliability of asymmetry measures of body trunks.

Often comparisons between shapes is simpler when the original shapes can be defined using a set of lines or elementary shapes. Shen et al. (1994) use predefined simple shapes of various sizes to attempt to classify tumours in mammograms while Guliato et al. (2008) find a polygonal model which best fits the outline of the tumor to aid classification. More generally Pavlidis and Horowitz (1974) outline a method for describing shapes using straight line segments and examine the importance of the position of the line joints. To better represent more complex shapes Chong et al. (1992) propose a similar method using B-spline segments whilst the method used by Cootes et al. (1995) depends on snake segments.

An alternative way to define and compare shapes is by finding a minimum spanning tree that fits the landmarks of the shape. An algorithm for calculating the minimum spanning tree is given by Fredman and Willard (1994) while Steele et al. (1987) discuss the asymptotics of the number of leaves of the minimum spanning tree. Minimum spanning tree measures can be useful in shape correspondence analysis. For example Munsell et al. (2008) investigate the performance of a number of tests of landmark based shape correspondence, including one based on minimum spanning trees.

Some standard methods of describing shapes make further analysis complex. To make further analysis simpler Kume et al. (2007) introduce shape-space smoothing splines to allow a smooth curve to be fitted to landmark data in

two-dimensions. Also Dryden et al. (2008) introduce a computationally simple framework for performing inference in three or more dimensions.

## 2.2 Curve Analysis

Shape analysis can be performed using landmarks, curves or surfaces defined on the shape. This thesis will concentrate on curves defining shapes. Defining shapes using curves can have many practical uses. For example Samir et al. (2006) show how defining facial shapes using curves can assist in facial recognition. This section will outline various methods for analysing shape curves.

One way to define curves is by examining the amount of bending shape curves exhibit at various points along the curve. For a plane curve the bending at any point on the curve can be represented using a single scalar value called curvature. In standard geometry curvature of a plane curve at the point  $a$  is defined as

$$\kappa(s) = \frac{d\phi(a)}{ds}$$

where  $\phi(a)$  is the angle between the tangent line at point  $a$  and the positive direction of the  $x$  axis and  $s$  is arc length. Alternatively, for computational simplicity, curvature of a plane curve can be defined as

$$\kappa(a) = \frac{x'(a)y''(a) - x''(a)y'(a)}{(x'(a)^2 + y'(a)^2)^{3/2}}$$

where  $x(a)$  and  $y(a)$  denote the  $x$  and  $y$  position of the curve and the dash notation indicates derivatives with respect to the arc length of the curve. Curvature of a plane curve is thoroughly examined in Chapter 3 and Gray (1998) is a good reference.

It is often computationally difficult to estimate curvature of a plane curve. The accuracy and precision of several methods are shown to have some inaccuracy by Worring and Smeulders (1992). Major problems are often found at extrema and a method to deal with these problems is proposed by Craizer et al. (2005). In spite of these difficulties curvature is a useful measure for use in shape analysis and there are many examples in the literature. A method to represent planar curves using curvature is given by Mokhtarian and Mackworth (1992). Small and Le (2002) introduce a model which can be used to describe the shape of plane curves using their curvature and propose a measurement of difference between two

curves using curvature. Curvature is used to enhance biological shape analysis in Costa et al. (2004) while Noriega et al. (2008) use curvature to find mutations in Arabidopsis roots.

The bending of a three-dimensional space curve at any point is represented by two scalar values called curvature and torsion. In standard geometry curvature ( $\kappa$ ) and torsion ( $\tau$ ) at the point  $a$  are defined as

$$\begin{aligned}\kappa(a) &= \left| \frac{d\mathbf{t}(a)}{ds} \right| \\ \tau(a) &= -\frac{d\mathbf{b}(a)/ds}{\mathbf{n}(a)}\end{aligned}$$

where  $\mathbf{t}(a)$ ,  $\mathbf{n}(a)$  and  $\mathbf{b}(a)$  are the tangent, normal and binormal vector of the curve at point  $a$  respectively. For computational simplicity the curvature and torsion can also be defined as

$$\begin{aligned}\kappa(a) &= |r'(a) \times r''(a)| \\ \tau(a) &= \frac{((r'(a) \times r''(a)) \cdot r'''(a))}{|r'(a) \times r''(a)|^2}\end{aligned}$$

where  $r(a) = [x(a), y(a), z(a)]$  and  $r'(a) = [x'(a), y'(a), z'(a)]$  where for example  $x'(a)$  is the first derivative of the  $x$  position of the curve with respect to arc length and  $\times$  denotes the cross-product. Curvature of a space curve will be comprehensively discussed in Chapter 4 while Gray (1998) is once again a good reference.

There are more difficulties when it comes to calculating curvature and torsion in space curves. There is literature on the subject including a recent attempt by Lewiner et al. (2005) to calculate curvature and torsion based on weighted least squares and local arc-length approximation. Further, Rieger and van Vliet (2002) propose a method using the gradient structure tensor which obtains the orientation field and a description of shape locally and then computes curvature in this tensor representation. One of the difficulties in estimating curvature and torsion is to control the sign of torsion. Karousos et al. (2008) address this issue by computing a domain which allows the space curve to have constant sign of torsion. A method to match shapes using space curves represented using a wavelet transformation is described by Tieng and Boles (1994). A potential application of curvature and torsion in space curves is found in Hausrath and Goriely (2007) where helical proteins are described in terms of their curvature and torsion.

Shape curves can be represented using various techniques which enhance the further analysis on the curves. Rosin and West (1995) represent shape curves using a set of superellipses whilst Alkhodre et al. (2001) represent curves using Fourier series. Srivastava and Lisle (2004) use Bézier curves to allow simple analysis of fold shapes. An interesting technique which allows curves to be matched is to represent each curve as a vector of turning angles and use some form of dynamic programming to calculate the distance between the two vectors. This technique is described by Niblack and Yin (1995) with discussion given to the problem of selecting a starting point. An alternative technique for matching curves is described by Pajdla and Van Gool (1995). This technique involves using semi-differentials to match the curves. The major issue is in finding reference points common to the curves being matched for which two techniques, one based on total curvature and the other based on arc-chord length ratios, are proposed. Aykroyd and Mardia (2003) propose a technique to describe the shape change of curves using a wavelet decomposition to construct a deformation function which is estimated using a Markov Chain Monte Carlo approach.

Whilst curves often represent a feature on a shape it is also possible to produce curves which show movement in space. Facial movements of cleft children are observed and analysed by Trotman et al. (2005). Also Faraway et al. (2007) use Bézier curves with geometrically important control points to track, analyse and predict hand motion. A similar but alternative technique using Bayesian methods to describe the mean and variability of human movement curves is described by Alshabani et al. (2007). Procrustes techniques which normalise stride patterns, in terms of time and magnitude of the stride, to allow gait patterns to be compared are outlined by Decker et al. (2007).

## 2.3 Functional Data Analysis

There are many sets of data where it is natural to think of the process as functional. Increased computing power in recent times has enabled the process of producing, recording and analysing functional data to be carried out without being overly computationally expensive. Ramsay and Silverman (1997) is a good reference for discussing basic functional data analysis techniques.

### 2.3.1 Interpolating and smoothing splines

Even when working with functional data it is unusual for the data to be available in a completely functional form. Often the function is defined by a large number of data points with a very small interval between neighbouring points. When this is the case it is important to be able to join the data points to produce a function in a smoother form than simply considering a straight line between neighbouring points. A technique for producing a function from the data points is cubic splines. An excellent overview of interpolating and smoothing using cubic splines is given by Green and Silverman (1994). Much of the description in this section comes from this work.

Suppose that there are a set of data pairs  $(s_i, y_i)$  on the closed interval  $[a, b]$  where  $i = 1, \dots, n$ . A simple way to describe the relationship between  $s$  and  $y$  is to fit the linear relationship

$$y = a + bs + \epsilon .$$

However it is often the case that fitting a linear relationship to the data is inappropriate. When this is the case a model of the form

$$y = g(s) + \epsilon , \tag{2.7}$$

where  $g$  is some function, is often a more appropriate model. The model in (2.7) could be fitted using least squares. However if there are no constraints on the function  $g$  it is clear that the residual sum of squares would have a minimum of zero when  $g$  is chosen as any function which interpolates the  $n$  points. Therefore a roughness penalty approach is taken which provides a good fit to the data but which avoids the fluctuations caused by interpolation.

The roughness penalty approach requires some measure of the roughness of a curve. There are numerous ways that this quantity can be measured. Green and Silverman (1994) suggest using the integrated squared second derivative of  $g$  i.e.  $\int_a^b [g''(s)]^2 ds$ . Two reasons why this choice of roughness measure is attractive are firstly that any measure of roughness should not be affected by a constant or a linear function and secondly if a thin piece of flexible wood was bent to the shape of  $g$  then the leading term of the strain energy is proportional to  $\int g''$ . The roughness penalty approach now states that the least squares estimator of  $g$  is

the function  $\hat{g}$  which minimises  $S(g)$  where

$$S(g) = \sum_{i=1}^n [y_i - g(s_i)]^2 + \alpha \int_a^b [g''(s)]^2 ds \quad (2.8)$$

The smoothing parameter  $\alpha$  is a positive scalar which controls the trade off between minimising the residual sum of squares and removing roughness from the model. As  $\alpha$  approaches zero  $\hat{g}$  will approach the interpolating function while the larger  $\alpha$  becomes the smoother the function  $g$  will become.

With this roughness penalty approach in mind the idea of cubic splines will now be discussed. Suppose  $s_1, \dots, s_n$  are real numbers in the interval  $[a, b]$  and that  $a < s_1 < s_2 < \dots < s_n < b$  then a cubic spline function  $g$  can be fitted on  $[a, b]$ . The conditions for a cubic spline are that  $g$  is a cubic polynomial in each interval  $(a, s_1), (s_1, s_2), \dots, (s_n, b)$  and at each point  $s_1, \dots, s_n$  the polynomials fit together in such a way that  $g, g'$  and  $g''$  are continuous on the whole interval  $[a, b]$ . The points  $s_i$  are called knots.

There are a number of ways that a cubic spline can be defined. Green and Silverman (1994) suggest that the cubic spline should be defined by its value and the value of its second derivative at each of the knots. Therefore the cubic spline  $g$  is defined by

$$\begin{aligned} g_i &= g(s_i) \\ \gamma_i &= g''(s_i) \end{aligned}$$

For  $g$  to be a natural cubic spline there is a further condition that  $\gamma_1 = \gamma_n = 0$ . Call  $\mathbf{g}$  the  $n$  length vector  $(g_1, \dots, g_n)^T$  and  $\gamma$  the  $n-2$  length vector  $(\gamma_2, \dots, \gamma_{n-1})$ . These vectors specify  $g$  completely although it turns out that not all possible  $\mathbf{g}$  and  $\gamma$  define a natural cubic spline. This requires a further condition depending on two band matrices, called  $Q$  and  $R$  say, which depend on the distances between the knots. Define the distance between knot  $i$  and knot  $i+1$  as  $h_i = s_{i+1} - s_i$  then let  $Q$  be the  $n \times (n-2)$  matrix with entries  $q_{ij}$ , where  $i = 1, \dots, n$  and  $j = 2, \dots, n-1$ , and

$$\begin{aligned} q_{j-1,j} &= h_{j-1}^{-1} \\ q_{jj} &= -h_{j-1}^{-1} - h_j^{-1} \\ q_{j+1,j} &= h_j^{-1} \\ q_{ij} &= 0 \quad \text{if } |i-j| \geq 2 \end{aligned}$$

for  $j = 2, \dots, n - 1$ . The columns of  $Q$  are numbered so that the top left entry of  $Q$  is  $q_{12}$  and the top right entry is  $q_{1,n-1}$ . Now let  $R$  be the  $(n - 2) \times (n - 2)$  symmetric matrix with entries  $r_{ij}$ , where  $i = 2, \dots, n - 1$  and  $j = 2, \dots, n - 1$ , which are given by

$$\begin{aligned} r_{ii} &= \frac{1}{3}(h_{i-1} + h_i) \\ r_{i,i+1} = r_{i+1,i} &= \frac{1}{6}h_i \\ r_{ij} &= 0 \quad \text{if } |i - j| \geq 2 \end{aligned}$$

Now Green and Silverman (1994) state that  $\mathbf{g}$  and  $\gamma$  specify a natural cubic spline  $g$  if and only if

$$Q^T \mathbf{g} = R\gamma \quad (2.9)$$

If (2.9) holds then the roughness penalty satisfies

$$\int_a^b g''(s)^2 ds = \gamma^T R\gamma = \mathbf{g}^T K \mathbf{g} \quad (2.10)$$

where  $K = QR^{-1}Q^T$ .

Although it is often the case that smooth cubic splines are required there are occasions where it is of interest to produce a smooth interpolating spline. Green and Silverman (1994) state that if  $n \geq 2$  and  $s_1 < \dots < s_n$  then given any values  $y_1, \dots, y_n$  there is a unique natural cubic spline  $g$  with knots at  $s_i$  which satisfies  $g(s_i) = y_i$  for  $i = 1, \dots, n$  i.e. there is one natural cubic spline which is an interpolant of the  $n$  points. To find this unique natural cubic spline (2.9) must be solved. Since  $Q$  is tridiagonal  $Q^T \mathbf{g}$  can be found from  $\mathbf{g}$  using a linear number of operations. Green and Silverman (1994) suggest that the premultiplication by  $Q^T$  can be achieved by

$$(Q^T \mathbf{g})_i = \frac{g_{i+1} - g_i}{h_i} - \frac{g_i - g_{i-1}}{h_{i-1}} \quad (2.11)$$

Now setting  $g_i = y_i$  and using (2.11),  $\gamma$  can be found from (2.9) since  $R$  is known and  $Q^T \mathbf{g}$  has been calculated numerically. The terms  $\gamma$  and  $\mathbf{g}$  now define the smooth curve which interpolates all data points and minimises  $\int g''(s)ds$ . The **spline** function in **R** uses this method to calculate the interpolating natural cubic spline.

It is more usual that a smooth function, which is close to the original data points but does not necessarily interpolate them, is required for analysis. Using

the roughness penalty approach the aim is to find the function,  $g$ , which minimises (2.8). The parameter  $\alpha$  controls the trade off between interpolating the points and removing all roughness from the function. Green and Silverman (1994) show that the function which minimises (2.8) must be a natural cubic spline and that, assuming  $n \geq 3$ , there is a unique natural cubic spline which results in the minimised  $S(g)$ . To show this, write the residual sum of squares as

$$\sum_{i=1}^n (y_i - g(t_i))^2 = (\mathbf{Y} - \mathbf{g})^T (\mathbf{Y} - \mathbf{g})$$

and using (2.10) show that (2.8) becomes

$$\begin{aligned} S(g) &= (\mathbf{Y} - \mathbf{g})^T (\mathbf{Y} - \mathbf{g}) + \alpha \mathbf{g}^T K \mathbf{g} \\ &= \mathbf{g}^T (I + \alpha K) \mathbf{g} - 2\mathbf{Y}^T \mathbf{g} + \mathbf{Y}^T \mathbf{Y} \end{aligned} \quad (2.12)$$

The unique minimum function,  $g$ , can then be found by setting

$$\mathbf{g} = (I + \alpha K)^{-1} \mathbf{Y} \quad (2.13)$$

The vector  $\mathbf{g}$  now defines the smooth spline function  $g$ . It is however inefficient to use (2.13) to calculate the spline function so an algorithm given by Reinsch (1967) is a preferable alternative.

The Reinsch algorithm for finding the smooth cubic spline function which minimises  $S(g)$  involves setting up a system of linear equations which can be solved to find the second derivatives at the knots ( $\gamma_i$ ). The values of the function at the knots ( $g_i$ ) can then be found from the second derivatives and the data values. The algorithm uses the idea of band matrices. Green and Silverman (1994) state that ‘a matrix is said to be a band matrix if all of its non-zero entries are concentrated on some small number of diagonals; the number of non-zero diagonals is called the bandwidth of the matrix.’ Before setting out the algorithm some useful formula must be derived. Firstly by replacing  $K$  with  $QR^{-1}Q^T$  and rearranging (2.13) becomes

$$(I + \alpha QR^{-1}Q^T) \mathbf{g} = \mathbf{Y} \quad (2.14)$$

This can be rearranged to give

$$\mathbf{g} = \mathbf{Y} - \alpha QR^{-1}Q^T \mathbf{g} \quad (2.15)$$

Equation (2.9) shows that  $Q^T \mathbf{g} = R\gamma$  so by replacement

$$\mathbf{g} = \mathbf{Y} - \alpha Q\gamma \quad (2.16)$$

Equation (2.16) gives a means to calculate the values of the function at the knots from the second derivatives and the data values. Now, using (2.9), this can be rewritten as

$$\begin{aligned} Q^{-T}R\gamma &= \mathbf{Y} - \alpha Q\gamma \\ R\gamma &= Q^T\mathbf{Y} - \alpha Q^TQ\gamma \\ R\gamma + \alpha Q^TQ\gamma &= Q^T\mathbf{Y} \\ (R + \alpha Q^TQ)\gamma &= Q^T\mathbf{Y} \end{aligned} \quad (2.17)$$

Since both  $Q$  and  $R$  are tridiagonal band matrices (with bandwidth 3), this equation can be solved using band matrix techniques. It is the major equation required for the Reinsch algorithm. It can further be shown that the matrix  $(R + \alpha Q^TQ)$  has bandwidth 5 and is symmetric and positive definite, so therefore has the Cholesky decomposition

$$R + \alpha Q^TQ = LDL^T \quad (2.18)$$

where  $D$  is a strictly positive diagonal matrix and  $L$  is a lower triangular band matrix with  $L_{ij} = 0$  when  $j < i - 2$  and  $j > i$  and where  $L_{ii} = 1$ . Since  $Q$  and  $R$  are band matrices the matrices  $L$  and  $D$  can be found using linear computation. Using these equations Green and Silverman (1994) describe the algorithm as the following four stage process.

1. Calculate  $Q^T\mathbf{Y}$  using (2.11).
2. Find the non-zero diagonals of the band matrix  $R + \alpha Q^TQ$  and from these calculate  $L$  and  $D$ .
3. Using replacement rewrite (2.17) as  $LDL^T\gamma = Q^T\mathbf{Y}$  and use forward and back substitution on this equation to find the second derivatives  $\gamma$  of the smooth function.
4. Once  $\gamma$  has been found use this in (2.16) to find the values of the smooth function at the knots  $\mathbf{g}$ .

Using this algorithm allows the smooth cubic spline function  $g$  to be calculated using the data points and the smoothing parameter.

So far little has been mentioned about the choice of the smoothing parameter  $\alpha$ . There are various ways in which the smoothing parameter can be chosen. Various values of the smoothing parameter may be investigated and the most appropriate value chosen subjectively. Alternatively the smoothing parameter can be chosen from the data using an automatic method such as cross-validation. In this thesis the method used to choose the smoothing parameter will involve selecting the equivalent degrees of freedom that the smooth function should have. To explain this recall from (2.13) that

$$\mathbf{g} = (I + \alpha K)^{-1} \mathbf{Y} = (I + \alpha Q R^{-1} Q^T)^{-1} \mathbf{Y}$$

since  $K = Q R^{-1} Q^T$ . Therefore

$$\mathbf{g} = A(\alpha) \mathbf{Y} \tag{2.19}$$

where  $A(\alpha)$ , which is called the hat matrix, is the matrix which maps the position of the observed values  $\mathbf{Y}_i$  onto their position on the smooth spline function  $\hat{g}(s_i)$  and is given by

$$A(\alpha) = (I + \alpha Q R^{-1} Q^T)^{-1} \tag{2.20}$$

To introduce a measure of the number of parameters fitted for a particular value of the smoothing parameter, suppose that the smooth curve is fitted using parametric regression where  $g(s) = \sum_{j=1}^k \theta_j g_j(s)$  with fixed functions  $g_j$  and  $k$  parameters  $\theta_j$ . The hat matrix is then a projection onto a  $k$  dimensional space. By analogy, the equivalent degrees of freedom (*EDF*) for a smoothing spline can be defined to be

$$EDF = tr\{A(\alpha)\}$$

Therefore the smoothing parameter of a curve can be calculated by setting a desired equivalent degrees of freedom for the curve. This is the technique used for specifying the smoothing parameter in this thesis. Note that, throughout, equivalent degrees of freedom will simply be referred to as degrees of freedom.

This is the technique used by the `smooth.spline` function in **R** which will be used throughout this thesis to fit smooth cubic splines to the data. The simplest way to roughly fit a smooth spline to three-dimensional data is equivalent to finding a smooth curve in the  $xy$  plane and a smooth curve in the  $xz$  plane and

combining these to give the three-dimensional smoothed fit.

### 2.3.2 Principal components analysis

In multivariate Statistics principal components analysis, see Jolliffe (1986) for details, is carried out by finding the eigenvalues and eigenvectors of either the covariance or correlation matrix. These ideas can be extended to the functional case. Suppose that there are  $n$  functional variables  $x_i(s)$  where  $i = 1, \dots, n$  and that  $\xi(s)$  is a weight function then the functional eigenequation can be written as

$$\int v(s, t)\xi(t)dt = \rho\xi(s) \quad (2.21)$$

where  $v$  is the covariance function which is given by

$$v(s, t) = \frac{1}{n} \sum_{i=1}^n x_i(s)x_i(t)$$

It is possible to think of  $\int v(s, t)\xi(t)dt$  as an integral transform,  $V$  say, of the weight function with the covariance function being a kernel of the transform such that

$$V\xi = \int v(s, t)\xi(t)dt$$

Now the eigenequation can be simply rewritten as

$$V\xi = \rho\xi$$

where  $\xi$  is an eigenfunction and  $\rho$  is an eigenvalue. Functional principal components analysis, see Ramsay and Silverman (1997) for details, can be thought of as the eigenanalysis of the covariance operator  $V$  which is defined by the covariance function.

To carry out functional principal components analysis it is preferable to reduce the functional eigenanalysis problem to a matrix eigenanalysis which is approximately equivalent. One such technique is to discretise the functions so that each function is defined by  $m$  equally spaced points along the  $s$  axis. This produces an  $(n \times m)$  data matrix from which the eigenanalysis can be carried out using standard multivariate techniques, with each function as the observation and each point on the  $s$  axis as the variable. Using this technique,  $n$  must be greater than  $m$ .

An alternative to discretising each function, which potentially retains more of

the functional structure of the variables, is to express each function as a linear combination of basis functions. To do this, suppose that each function has the basis expansion

$$x_i(s) = \sum_{j=1}^k c_{ij} \phi_j(s)$$

where  $\phi_j(s)$  is the basis function  $j$  and  $c_{ij}$  is the coefficient of basis function  $j$  for the  $i$ th original function. The basis expansion can be written in the vector-matrix form

$$x = C\phi$$

where  $x$  is the  $(n \times 1)$  matrix containing  $x_1(s), \dots, x_n(s)$ ,  $\phi$  is the  $(k \times 1)$  matrix containing  $\phi_1(s), \dots, \phi_k(s)$  and  $C$  is the  $(n \times k)$  matrix with row  $i$  containing the basis coefficients for original function  $i$ . The covariance function can now be written in matrix form as

$$v(s, t) = \frac{1}{n} \phi(s)^T C^T C \phi(t)$$

Now define an eigenfunction for the eigenequation (2.21) using the expansion

$$\xi(s) = \sum_{j=1}^k b_j \phi_j(s)$$

This can also be written in matrix form as

$$\xi(s) = \phi(s)^T b$$

The left side of the eigenequation (2.21) can then be rewritten in matrix form as

$$\begin{aligned} \int v(s, t) \xi(t) dt &= \int \frac{1}{n} \phi(s)^T C^T C \phi(t) \phi(t)^T b dt \\ &= \frac{1}{n} \phi(s)^T C^T C W b \end{aligned}$$

where  $W$  is the  $(k \times k)$  matrix containing the entries

$$W_{i,j} = \int \phi_i(t) \phi_j(t) dt$$

Therefore (2.21) can be expressed as

$$\frac{1}{n}\phi(s)^T C^T C W b = \rho \phi(s)^T b \quad (2.22)$$

and since this must hold for all  $s$  this implies the matrix equation

$$\frac{1}{n}C^T C W b = \rho b$$

Now in the basic theory of principal components analysis two constraints are introduced. Firstly a continuous analogue to the unit sum of squares constraint in the multivariate setting is introduced such that  $\|\xi\| = 1$ . This constraint implies that  $b^T W b = 1$ . The second constraint is that all eigenfunctions are orthogonal i.e.  $\int \xi_i(s) \cdot \xi_j(s) ds = 0$ . This implies that  $b_i^T W b_j = 0$ . Now to calculate the principal components define  $u = W^{\frac{1}{2}} b$  and solve the eigenvalue problem

$$\frac{1}{n}W^{\frac{1}{2}} C^T C W^{\frac{1}{2}} u = \rho u$$

and compute  $b = W^{-\frac{1}{2}} u$  for each eigenvector to give the coefficients of the eigenfunction.

### 2.3.3 Curve registration

It is often the case that a number of functions share a common pattern with the major variation being in the position of the function on the time (or  $s$ ) axis. There is a large literature on techniques available for reducing this  $s$  variation by aligning common features of the functions to the average  $s$  position. Note that the discussion here will be on  $s$  variation and alignment where  $s$  is any ordered variable which is often, but not necessarily, time. Suppose that there are  $n$  functions  $f_i(s)$  which are to be aligned then these can be aligned by applying warping functions  $g_i$ 's so that  $f_i(g_i(s))$  is the aligned function for function  $i$ . Gasser et al. (1990) and Kneip and Gasser (1992) suggest a method for aligning functions so that common individual characteristic points are aligned to the average position. Characteristic points, where the  $r$ th characteristic point for function  $i$  is denoted as  $\gamma_{ri}$ , can be defined as predetermined geometrically important points or as turning points of the functions. The  $g_i$ 's are calculated by setting  $g_i(\gamma_{ri}) = \bar{\gamma}_r$  so that  $f_i(g_i(\gamma_{ri})) = f_i(\bar{\gamma}_r)$  and then interpolating to give continuous warping functions. This method is relatively straightforward and simple to implement and will be used in this thesis. There are clearly numerous alternative methods

for aligning functions. Ramsay and Li (1998) introduce a method which finds warping functions by minimising a penalised squared error criterion which is dependent on a target function (usually a mean function) and a smoothing penalty based on the second derivative of the warping function. They also introduce a Procrustes fitting criterion to update estimates of the warping functions. Tang and Müller (2008) describe a method where pairwise warping functions are calculated to align function  $i$  to all other  $n - 1$  functions. The overall warping function for function  $i$  is then found as the average of the  $n - 1$  pairwise warping functions. A semiparametric approach to estimating the warping functions is provided by Gervini and Gasser (2004) where the warping functions are modelled using B-spline component functions multiplied by scores such that

$$g_i(s) = s + \sum_{j=1}^q k_{ij} \phi_j(s)$$

where  $\phi_j(s) = c_j^T \beta(s)$  for a set of B-spline basis functions  $\beta(s)$ . The B-spline component functions are estimated from the data so only the scores  $k_{ij}$  have to be individually estimated, hence the semiparametric nature of the process. The warping functions can also be estimated nonparametrically using nonparametric maximum likelihood methods as outlined by Rønn (2001) and Gervini and Gasser (2005). Another method for function alignment is based on the process of dynamic time warping developed in engineering originally for speech analysis. Both Wang and Gasser (1997) and Wang and Gasser (1999) describe how dynamic time warping can be used to align functions in the general setting. For a general overview Ramsay and Silverman (1997) contains a chapter on the registration of functional data which describes a variety of techniques for aligning functions.

Taking the piecewise average of the aligned function produces an average function which gives a good representation of the average pattern of the functions and the average value of the function at the characteristic points. Averaging of the aligned functions is called structural averaging. Comparing structural averages for various groups is one of the major reasons for aligning functions.

### 2.3.4 Selected recent literature on modelling with functional data

The scope of functional data analysis is too large to make a comprehensive overview feasible. However, since this thesis will be concerned with techniques

which involve modelling using functional data, this section will briefly introduce some recent literature from this area.

Ramsay and Silverman (1997) introduces functional linear models for functional responses. They suggest that a functional response  $y_i(t)$  can be modelled using a functional predictor  $x_i(s)$  as

$$y_i(t) = \alpha(t) + \int_s x_i(s)\beta(s, t)ds + \epsilon_i(t)$$

where the regression function  $\beta$  is a function of both  $s$  and  $t$ . Cuevas et al. (2002) investigate the convergence of the regression function and also the consistency of the results while Cardot et al. (1999) also discuss convergence of the estimator. A comparison of various methods to calculate the estimator of the regression function is given by James and Silverman (2005). Rice and Silverman (1991) treat the response as a stochastic process and attempt to estimate the mean and covariance structure simultaneously. The mean for each response is simply the expected value of the response whilst the kernel covariance function is dependent on the responses. Yao et al. (2005) and Shi et al. (2007) provide alternatives to this method with the difference in methods being that in the former the covariance structure is dependent on  $s$  and is also specifically designed for sparse data whereas in the latter the covariance structure is dependent on the  $x_i(s)$ . In Shi et al. (2007) the issue of heterogeneity is not discussed and this is addressed in Shi and Wang (2008). An F-test to investigate the relationship between a functional response and a functional predictor is described by Yang et al. (2007). There is also extensive literature on linear models with a functional response and scalar predictors. Regression with variables of this form is described with an example in Faraway (1997).

The functional modelling in this thesis will focus on the situation where there is a scalar response and functional predictors. Ramsay and Silverman (1997) introduce a functional linear model for a scalar response with a single functional predictor. This is an analogue of the case with a functional response such that

$$y_i = \alpha + \int_s x_i(s)\beta(s)ds + \epsilon$$

where a functional parameter is applied to each functional predictor. As stressed by Besse et al. (2005), estimating the functional parameter requires some form

of discretisation. This can be done either by discretising the functional predictors and making the problem a scalar multivariate problem or by representing the functional predictors using splines. The technique of using splines for the discretisation will be discussed further and implemented in Section 6.1.3. An extension of the technique employed by Ramsay and Silverman (1997), which also estimates the between curve variation as well as the within curve variation, is employed by James (2002) while a method which accounts for variation in the predictors when estimating the functional parameter is introduced by Cardot et al. (2007). Müller and Stadtmüller (2005) introduce a generalised functional linear model which is estimated using functional quasi-likelihood. In much of the work in this area the major interest is in interpreting the estimate of the functional parameter. Due to this Cardot and Sarda (2006) and Li and Hsing (2007) discuss the convergence of  $\hat{\beta}(s)$  to  $\beta(s)$  whilst Cai and Hall (2006) concentrate on optimising the prediction of  $\int_s x_i(s)\beta(s)$  or  $\alpha + \int_s x_i(s)\beta(s)$ . Further Apanasovich and Goldstein (2008) discuss how  $\beta(s)$  should be estimated to optimise the prediction of future responses. An extension to the functional linear model is the varying coefficient functional model proposed by Cardot and Sarda (2008) where the functional parameter can vary according to the other scalar inputs.

Nonparametric techniques can be used to estimate an unknown scalar response from a functional predictor using known response-functional predictor pairs. Ferraty and Vieu (2006) give a good overview of techniques using kernel functions and semi-metrics to predict the response. Estimating the response as the conditional expectation of the unknown response given the functional predictor will be discussed and implemented in Chapter 6. Bias, variance, mean square errors and asymptotic distributions for the predictor can be found in Ferraty et al. (2007). Methods estimating the unknown response as the median of the cumulative distribution function of the conditional density, or as the mode of the probability density function of the conditional density, are also described by Ferraty and Vieu (2006). An alternative to the kernel and semi-metric estimator is proposed by Baïllo and Grané (2007) and is based on local linear regression. Cardot and Sarda (2006) describe two tests, one which relies on the  $\chi^2$  distribution and the other which approximates the  $\chi^2$  distribution using a Gaussian distribution, to formally test whether the functional predictors are related to the response while Cardot et al. (2004) introduce a computational test of no effect using a permutation test. The regression operator found nonparametrically can also be found using functional principal components analysis. This idea was

proposed by Bosq (1991) and developed by Cardot et al. (1999).

# Chapter 3

## Analysis of Plane Curves

### 3.1 Characterising Plane Curves

A plane curve is a curve which is contained in a two-dimensional plane. This section will look to describe a plane curve as a function of how much the curve is bending at points along the curve. For a plane curve the amount of bending experienced at each point is a scalar value called curvature. A plane curve can be determined up to rigid transformations by its curvature.

Much of the current work on shape analysis (Dryden and Mardia (1998) give an excellent overview of popular techniques) uses a limited number of landmarks to describe the shape. Using only a limited number of landmarks may well result in a large amount of useful information being lost. Furthermore, if the shapes lie in different areas of space, Procrustes methods using landmarks (or pseudo-landmarks) are required to align the shapes. The technique of using curvature to analyse shapes offers an alternative to these approaches. Curvature can be calculated over the whole curve which limits the amount of information about the shape which is lost. Furthermore, since curvature is independent of the position of the shape in space then analysing shapes by curvature avoids the need for any use of Procrustes methods.

Before describing the calculation of curvature it is important to define various terms and the way that these relate to curvature.

#### 3.1.1 Arc length

Suppose that an arbitrarily parameterised curve  $r(a) = [x(a), y(a)]$  is represented by  $n$  points  $(a_1, \dots, a_n)$  say) on the curve. If the coordinates of any two points  $a_b$

and  $a_c$  are  $[x(a_b), y(a_b)]$  and  $[x(a_c), y(a_c)]$  respectively then the distance between these two points is

$$Dist(a_b, a_c) = \sqrt{[x(a_b) - x(a_c)]^2 + [y(a_b) - y(a_c)]^2}$$

Now approximate the curve by the lines which successively connect the points  $a_1, \dots, a_n$ . Then the length of the curve can be approximated as the length of the line segments which connect the points

$$L = \sum_{i=1}^{n-1} \sqrt{[x(a_{i+1}) - x(a_i)]^2 + [y(a_{i+1}) - y(a_i)]^2}$$

If the curve is rectifiable, as  $n$  increases  $L$  will approach the true length of the curve. The distance of a point along the curve, calculated in this way, is the arc length of the curve at this point. Therefore the arc length ( $s_q$  say) at any point  $a_q$  will be given by

$$s_q = \sum_{i=1}^{q-1} \sqrt{[x(a_{i+1}) - x(a_i)]^2 + [y(a_{i+1}) - y(a_i)]^2} \quad (3.1)$$

### 3.1.2 Tangent and normal vectors

Suppose there is a curve  $r(s) = [x(s), y(s)]$  which is parameterised by arc length  $s$  as opposed to the arbitrary parameter  $a$  in Section 3.1.1. Moving along the curve then means to move along the curve in the direction in which the arc length increases. At each point on the curve it is possible to define a vector, called the tangent vector, which has unit length and which points in the direction that the curve is travelling. To calculate the tangent vector at the point  $r(s_i)$  say, consider two points  $r(s_i) = [x(s_i), y(s_i)]$  and  $r(s_i + h_1) = [x(s_i + h_1), y(s_i + h_1)]$ . As  $h_1$  approaches zero the line which passes through both  $r(s_i)$  and  $r(s_i + h_1)$  approaches the tangent line at  $r(s_i)$ . The tangent vector at  $r(s_i)$  is a unit length vector in the direction of the tangent line. The tangent vector at the point  $r(s)$  will be denoted by  $\mathbf{t}(s)$ .

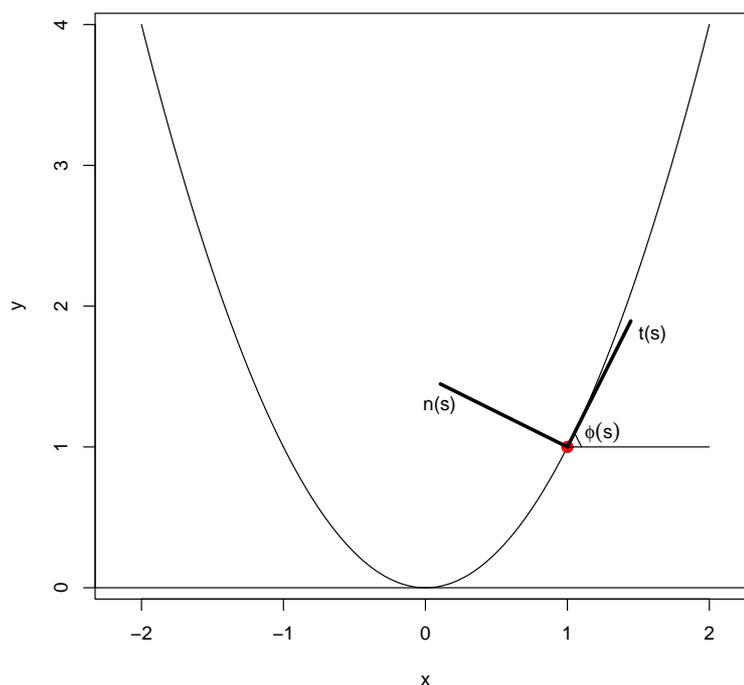
The normal vector is a unit vector which is perpendicular to the tangent vector in the anti-clockwise direction. The normal vector at the point  $r(s)$  will be denoted by  $\mathbf{n}(s)$ .

### 3.1.3 Curvature and the Frenet formulae

Curvature of a plane curve is a scalar measurement of the magnitude of the bending of the curve at a point as the point moves along the curve with constant velocity. The curvature at the point  $r(s)$  will be denoted by  $\kappa(s)$ . Consider the point  $r(s)$  with the tangent vector  $\mathbf{t}(s)$  and normal vector  $\mathbf{n}(s)$ . The angle that  $\mathbf{t}(s)$  makes with the positive direction of the  $x$ -axis is called the turning angle of the plane curve. The turning angle of the curve at the point  $r(s)$  will be denoted by  $\phi(s)$ . Gray (1998) shows that curvature can then be defined as the rate of change of the turning angle with respect to the arc length. So,

$$\kappa(s) = \frac{d\phi(s)}{ds} \quad (3.2)$$

Therefore it is clear that a straight line will have curvature of zero since the turning angle will not change. Figure 3.1 shows the positions of  $\mathbf{t}(s)$ ,  $\mathbf{n}(s)$  and  $\phi(s)$  relative to an arbitrary point on a quadratic curve.



**Figure 3.1:** The positions of  $\mathbf{t}(s)$ ,  $\mathbf{n}(s)$  and  $\phi(s)$  relative to an arbitrary red point  $r(s)$  on a quadratic curve

The tangent vector,  $\mathbf{t}(s)$ , is a unit length vector in the direction of the slope of the curve at the point  $r(s)$ . Therefore the direction of the tangent vector can be calculated by the first derivative of the curve with respect to arc length. To find the tangent vector, and ensure it has unit length, the vector given by the

first derivative of the curve with respect to arc length is normalised. However Pressley (2001) shows that the magnitude of this vector is one. Therefore the tangent vector at  $r(s)$  is given by

$$\mathbf{t}(s) = \frac{dr(s)}{ds} \quad (3.3)$$

Since the tangent vector has unit length  $\mathbf{t}(s) \cdot \mathbf{t}(s) = 1$ . Differentiating the scalar product with respect to arc length gives

$$0 = (\mathbf{t}(s) \cdot \mathbf{t}(s))' = \mathbf{t}'(s) \cdot \mathbf{t}(s) + \mathbf{t}(s) \cdot \mathbf{t}'(s) = 2(\mathbf{t}'(s) \cdot \mathbf{t}(s)) \quad (3.4)$$

Therefore  $\mathbf{t}'(s) \cdot \mathbf{t}(s) = 0$  hence  $\mathbf{t}'(s)$  and  $\mathbf{t}(s)$  are mutually perpendicular.

Now consider Figure 3.1 which shows that there is a right-angled triangle which has hypotenuse  $\mathbf{t}(s)$  and has the turning angle  $\phi(s)$  as one of the angles in the triangle. From, Pressley (2001), using basic trigonometry and the fact that the tangent vector has unit length it follows that

$$\begin{aligned} \mathbf{t}(s) &= (|\mathbf{t}(s)| \cos(\phi(s)), |\mathbf{t}(s)| \sin(\phi(s))) \\ \mathbf{t}(s) &= (\cos(\phi(s)), \sin(\phi(s))) \end{aligned} \quad (3.5)$$

From this differentiation gives

$$\frac{d\mathbf{t}(s)}{d\phi(s)} = (-\sin(\phi(s)), \cos(\phi(s)))$$

Once again refer to Figure 3.1. There is a right-angled triangle which has hypotenuse  $\mathbf{n}(s)$  and has  $\frac{\pi}{2} - \phi(s)$  as one of the angles in the triangle. Therefore using basic trigonometry rules

$$\begin{aligned} \mathbf{n}(s) &= (-|\mathbf{n}(s)| \times \cos(90 - \phi(s)), |\mathbf{n}(s)| \times \sin(90 - \phi(s))) \\ \mathbf{n}(s) &= (-\sin(\phi(s)), \cos(\phi(s))) \end{aligned} \quad (3.6)$$

Thus,

$$\frac{d\mathbf{t}(s)}{d\phi(s)} = (-\sin(\phi(s)), \cos(\phi(s))) = \mathbf{n}(s) \quad (3.7)$$

This relationship holds regardless of the orientation of  $\mathbf{t}(s)$  and  $\mathbf{n}(s)$ .

Now from (3.2) and (3.7) it follows that

$$\begin{aligned}\frac{d\mathbf{t}(s)}{ds} &= \frac{d\mathbf{t}(s)}{d\phi(s)} \cdot \frac{d\phi(s)}{ds} = \kappa(s)\mathbf{n}(s) \\ \frac{d\mathbf{n}(s)}{ds} &= \frac{d\mathbf{n}(s)}{d\phi(s)} \cdot \frac{d\phi(s)}{ds} = -\kappa(s)\mathbf{t}(s)\end{aligned}$$

These equations give us the famous two-dimensional Frenet formulae

$$\frac{d\mathbf{t}(s)}{ds} = \kappa(s)\mathbf{n}(s) \quad (3.8)$$

$$\frac{d\mathbf{n}(s)}{ds} = -\kappa(s)\mathbf{t}(s) \quad (3.9)$$

It is clear that curvature is defined by the positions of the tangent and normal vectors. Section 3.1.2 stated that the normal vector is perpendicular to the tangent vector in the anti-clockwise direction. The reason for the normal vector always being set in the anti-clockwise direction is that this makes it possible to determine the direction the curve is turning simply by the sign of the curvature. If the curvature is positive the curve is turning towards the normal vector (i.e. it is turning to the left). Alternatively, if the curvature is negative the curve is turning away from the normal vector (i.e. it is turning to the right).

### 3.1.4 Calculating curvature in practice

It has been shown that curvature can be calculated at any point on the curve from the normal and tangent vectors or from the turning angles along the curve. Using these techniques for many points on a curve would be time consuming. It is easier to use a computationally less complex method.

Consider once again an arbitrarily parameterised (i.e. not necessarily parameterised by arc length) curve  $r(a) = (x(a), y(a))$  with the arc length of each point  $s(a)$ . Recall from (3.2)

$$\begin{aligned}\kappa(a) &= \frac{d\phi(a)}{ds(a)} \\ &= \frac{d\phi(a)}{da} \cdot \frac{da}{ds(a)} \\ &= \frac{d\phi(a)/da}{ds(a)/da}\end{aligned} \quad (3.10)$$

The denominator in (3.10) is the rate of change of arc length  $s$  with respect to the position along the curve  $a$ . In other words this is the rate of change of the distance travelled along the curve with respect to the position along the curve. Now Gray (1998) shows that the arc length of a parameterised curve is

$$s = \int_a^b \sqrt{\left(\frac{dx(a)}{da}\right)^2 + \left(\frac{dy(a)}{da}\right)^2} da$$

Therefore,

$$\frac{ds(a)}{da} = \sqrt{\left(\frac{dx(a)}{da}\right)^2 + \left(\frac{dy(a)}{da}\right)^2}$$

For simplicity from now on  $\frac{dx(a)}{da}$  and  $\frac{dy(a)}{da}$  will be represented by  $x'$  and  $y'$  respectively. So,

$$\frac{ds(a)}{da} = \sqrt{x'^2 + y'^2} \quad (3.11)$$

When considering the numerator a look at Figure 3.1 shows that at any point  $\tan$  of the turning angle can be given by the rate of change of  $y$  with respect to  $x$ . Therefore,

$$\begin{aligned} \tan(\phi(a)) &= \frac{dy(a)}{dx(a)} = \frac{\frac{dy(a)}{da}}{\frac{dx(a)}{da}} \\ &= \frac{y'}{x'} \end{aligned} \quad (3.12)$$

Now using (3.12) and the quotient rule it becomes clear that,

$$\frac{d}{da} \tan(\phi(a)) = \frac{x'y'' - y'x''}{x'^2} \quad (3.13)$$

Also from the identity  $\frac{d}{dx} \tan(t) = \sec^2(t) \frac{dt}{dx}$  it can be seen that,

$$\frac{d}{da} \tan(\phi(a)) = \sec^2(\phi(a)) \frac{d\phi(a)}{da} \quad (3.14)$$

Using (3.12), (3.13) and (3.14) and the identity  $\sec^2(x) = 1 + \tan^2(x)$  it can be shown that,

$$\begin{aligned}
\frac{d\phi(a)}{da} &= \frac{1}{\sec^2(\phi(a))} \cdot \frac{d}{da}(\tan(\phi(a))) \\
&= \frac{1}{1 + \tan^2(\phi(a))} \cdot \frac{x'y'' - y'x''}{x'^2} \\
&= \frac{1}{1 + \frac{y'^2}{x'^2}} \cdot \frac{x'y'' - y'x''}{x'^2} \\
&= \frac{x'y'' - y'x''}{x'^2 + y'^2}
\end{aligned} \tag{3.15}$$

By substituting (3.11) and (3.15) into (3.10) it follows that,

$$\begin{aligned}
\kappa(a) &= \frac{\frac{x'y'' - y'x''}{x'^2 + y'^2}}{\sqrt{x'^2 + y'^2}} \\
&= \frac{x'y'' - y'x''}{x'^2 + y'^2} \cdot \frac{1}{\sqrt{x'^2 + y'^2}} \\
&= \frac{x'y'' - x''y'}{(x'^2 + y'^2)^{3/2}}
\end{aligned} \tag{3.16}$$

Equation(3.16) shows that to calculate the curvature of a plane curve all that is required is the first and second derivative of the  $x$  and  $y$  position of the curve with respect to the position along the curve. It is both convenient and computationally simple to reparameterise the curve using arc length.

### 3.1.5 Reconstructing a curve from its curvature

Gray (1998) shows that a plane curve can be reconstructed up to rigid transformations from its curvature. Equation (3.2) shows that

$$\phi(s) = \int_0^s \kappa(u) du + \phi(0) \tag{3.17}$$

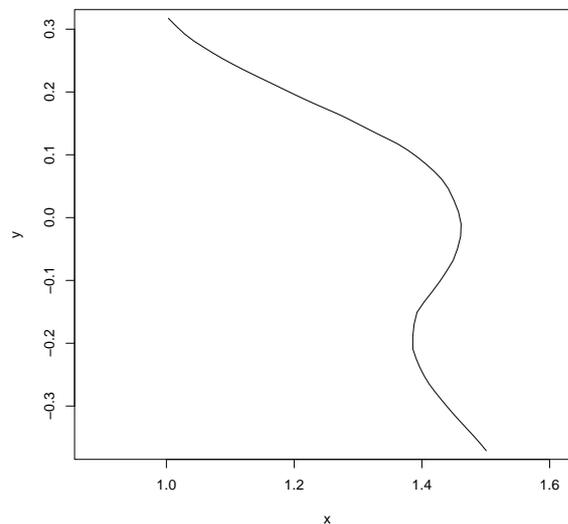
also (3.3) and (3.5) show that

$$\begin{aligned}
\frac{d\mathbf{r}(s)}{ds} &= \mathbf{t}(s) = (\cos(\phi(s)), \sin(\phi(s))) \\
\mathbf{r}(s) &= \left( \int_0^s \cos(\phi(u)) du + a, \int_0^s \sin(\phi(u)) du + b \right)
\end{aligned} \tag{3.18}$$

where  $a$  and  $b$  give the starting position of the reconstructed curve and  $\phi(0)$  gives the starting turning angle of the reconstructed curve. The curvature of the curve and the three parameters  $a$ ,  $b$  and  $\phi(0)$  give a rigid transformation of the curve.

## 3.2 Curvature of a Plane Curve: Midline Profile Example

To illustrate curvature in a plane curve an example, which considers the curvature of the midline profile of a one year old control child, is discussed. In practice most curves will be defined by a number of points which show the shape of the curve rather than a continuous function. Clearly the larger the number of points the more accurate the representation of the true curve. The profile is defined by 57 data points with the lower-most point being the top of the upper lip and the upper-most point being the point between the eyes. Movement in the  $y$  axis corresponds to movement up and down the face while movement in the  $x$  axis corresponds to movement into and out of the face. Figure 3.2 shows an example profile.



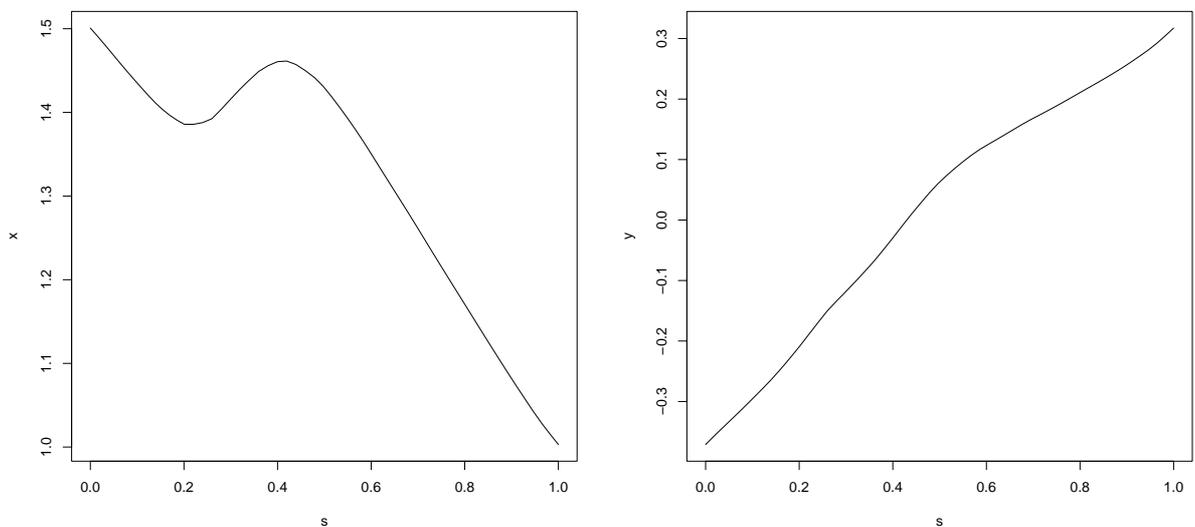
**Figure 3.2:** An example midline profile of a one year control child.

### 3.2.1 Calculating curvature

Equation (3.16) gives a formula which can be used to calculate the curvature at each point for which the value of the  $x$  and  $y$  coordinates are available. Before this

equation can be used the curve must be reparameterised. As has been mentioned, it is often simplest and computationally advantageous to reparameterise the curve in terms of arc length. The ‘starting point’ of the curve is set as the bottom of the profile (the point at the top of the upper lip) and using (3.1) the arc length is calculated such that it increases up the face.

It is important for further analysis that the arc length of each curve is the same. In practice this will not be the case therefore it is necessary to rescale each curve. For simplicity it makes sense to rescale to 1; however any value can be chosen. Of course this rescaling must be done without changing the shape of the curve. To do this, consider the set of points on the curve  $(x_1, y_1), \dots, (x_n, y_n)$  as  $(\alpha x_1, \alpha y_1), \dots, (\alpha x_n, \alpha y_n)$  where  $\alpha$  is an arbitrary scale parameter. Then to change the size of the curve without changing the shape all that is required is to change  $\alpha$ . Setting  $\alpha$  to the reciprocal of the arc length of the original shape will give an arc length for each curve which runs from  $(0, 1)$ . The example shown in Figure 3.2 has already been rescaled while Figure 3.3 shows plots of both the  $x$  and  $y$  position against arc length.

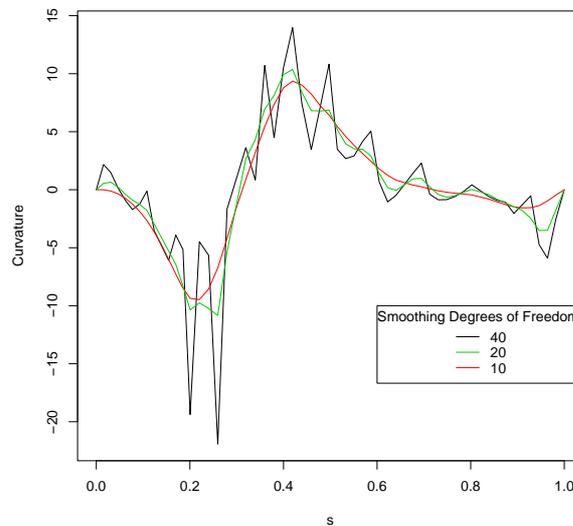


**Figure 3.3:** Plot of  $x$  and  $y$  position against arc length for the example profile.

Before using (3.16) to calculate the curvature at each point on the curve a numerical method is required to calculate the first and second derivative of both the  $x$  and the  $y$  position with respect to arc length. The method used throughout this thesis is to fit a cubic smoothing spline to the points (using the `smooth.spline` function in **R**, see Section 2.3.1 for details) and then use the **R** function `predict` to calculate the first and second derivatives at each point. The amount of smoothing applied by the spline is defined by the required degrees of

freedom. The lower the degrees of freedom the smoother the curves of  $x$  and  $y$  against  $s$  will be. Highly smoothed curves are often desirable for further analysis; however, too much smoothing may result in information from the original shape being lost. This trade-off must be made and will be discussed later.

A useful way to represent the curvature of a curve is by plotting curvature against arc length. This illustrates how the curvature function changes along the curve. Figure 3.4 shows the calculated curvature plotted against arc length for the example profile using 10, 20 and 40 degrees of freedom for smoothing.



**Figure 3.4:** Plot of curvature against arc length for the example profile using different degrees of freedom for smoothing.

All three curvature functions in Figure 3.4 show similar trends with an area of minimum curvature, which corresponds to the area where the profile bends at the base of the nose, and an area of maximum curvature, which corresponds to the tip of the nose. Figure 3.2 shows that the profile bends to the right at the base of the nose which generates negative curvature while the profile bends to the left at the tip of the nose which generates positive curvature. This will clearly only be the case when the profile is oriented in the manner shown in Figure 3.2. Other orientations (i.e. the profile being looked at from the left of the person) may result in a different sign of curvature although the magnitude of curvature will remain the same.

It is clear from Figure 3.4 that regardless of the smoothing used the curvature is zero at the end points of the curve. Recall from Section 2.3.1 that a condition of natural cubic splines is that the second derivative of the smooth function is equal to zero at the end points. Since the functions of the  $x$  and  $y$  position against

arc length are fitted using natural cubic splines, and considering the form of (3.16), it can be seen that using this derivative method to calculate curvature will always result in zero curvature at the end points. This is a somewhat undesirable property of the method. However, the facial curves considered in the context of this thesis are relatively stable at the end points. It is reasonable to assume that most curvature, and particularly the curvature of interest, will occur towards the middle of the curves and that zero curvature at the end points is not unreasonable and certainly not of major concern.

If it were felt that allowing non-zero curvature at the end points was of significant interest other types of splines could be used for smoothing. Eilers and Marx (1996) introduce the idea of p-splines which are an extension of smoothing splines. The major difference is that p-splines do not pre-specify the knots but allow a form of adaptive knot selection. This gives greater flexibility when producing a smooth fit of the data. The calculation of p-splines also does not require the condition that the second derivative is zero at the end of the curve. Therefore, using p-splines for the smoothing of the  $x$  and  $y$  position against arc length would allow for non-zero curvature at the ends of the facial curve. As mentioned this is not of primary interest in the cleft lip study so the method used for smoothing here will continue to be natural cubic smoothing splines.

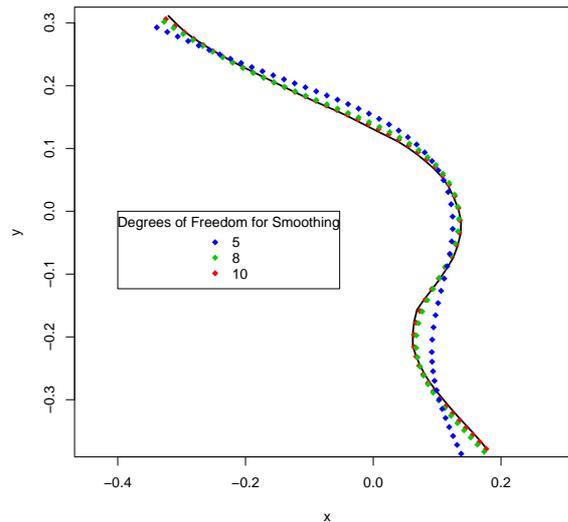
Figure 3.4 also shows the effect of changing the smoothing parameter. The roughness of the function produced when using 40 degrees of freedom makes it difficult to interpret and analyse further. The curves produced when using both 20 and particularly 10 degrees of freedom would be simpler to work with. However it is important to be confident that these smoother functions still approximate the original data well. To examine how well each of the curvature functions represent the data it is possible (see Section 3.1.5) to use these to reconstruct the original profile up to rigid transformations.

### 3.2.2 Reconstructing the profile

To check how well the curvature functions represent the original profile, a reconstruction of the profile using (3.17) and (3.18) is produced. Equation (3.17) uses the integral of the curvature function with respect to arc length to calculate the turning angle at each point. Equation (3.18) uses these turning angles to calculate an  $x$  and  $y$  position for each point of the reconstructed profile.

To complete the reconstruction the initial turning angle ( $\phi_0$ ) and the starting position ( $r(0)$ ) must be specified. Since these are arbitrarily chosen it is difficult

to compare the original and reconstructed profiles. To allow for a simple comparison it makes sense to rotate and translate the reconstructed profile as close as possible to the original profile. This can be done using ordinary Procrustes analysis without scaling. The matched profiles using 5, 8 and 10 degrees of freedom for smoothing can be found in Figure 3.5



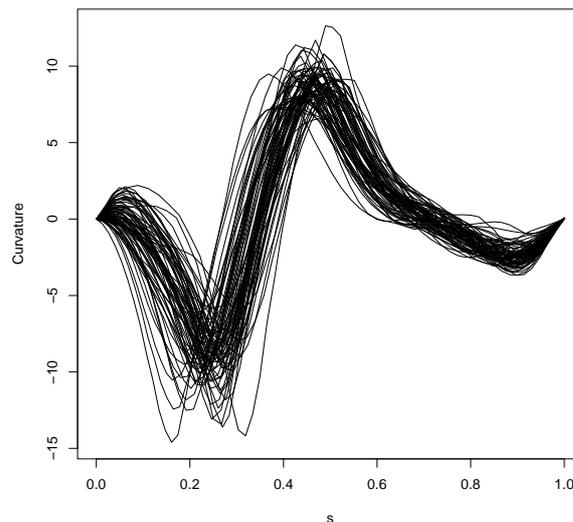
**Figure 3.5:** Reconstructed profile aligned to the original profile for curvature when different degrees of freedom for smoothing are used.

Figure 3.5 shows that, as expected, as the degrees of freedom decrease the reconstructed profile becomes less accurate. The reconstruction made with 5 degrees of freedom is a poor representation of the original profile and although the reconstruction made with 8 degrees of freedom is much closer to the original profile there are some inconsistencies, especially at the points where curvature has high magnitude and at the ends of the profile. For these reasons, and recalling that the curvature function using 10 degrees of freedom was relatively smooth, 10 degrees of freedom offers a suitable compromise between accuracy and simplicity.

### 3.2.3 Investigating a collection of curvature curves

It is often the case that interest lies not in a single plane curve but in a group of plane curves. In many situations a set of plane curves arises from measurement of a similar feature on different subjects and interest is in investigating this set of curves and identifying where any differences lie. The example considered here is an extension of the midline profile example discussed earlier in Section 3.2. Now the data considered is 71 midline profiles of one year old control children. Each profile is defined by 57 data points and is rescaled to have arc length 1.

Curvature is calculated for each profile separately using 10 degrees of freedom and the curvature functions are shown in Figure 3.6. These all appear to follow the same pattern. For each profile there is an area of minimum curvature at the base of the nose and an area of maximum curvature at the tip of the nose. The major differences between the curvature functions are the positions along the  $s$  axis of these areas of minimum and maximum curvature and the magnitude of the minimum and maximum curvature.

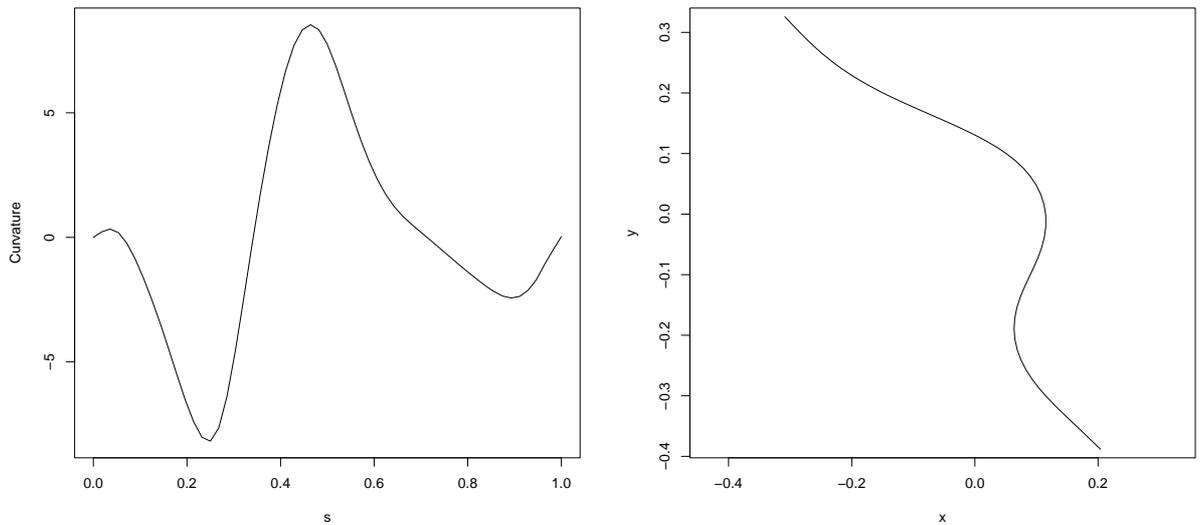


**Figure 3.6:** Plot of curvature against arc length for 71 one year old control children midline profiles.

It is of interest to produce an ‘average’ curvature function for this group of children and to use this curvature function to produce an average midline profile using the methods in Section 3.2.2. One simple way of doing this is to take the piecewise average of all 71 curvature functions at a number of points along the  $s$  axis. This is done here by finding the mean of the 71 curvature values at 57 regular points on the  $s$  axis and interpolating these mean points to give the average curvature function shown in Figure 3.7. The reconstructed average profile is also shown in Figure 3.7.

### 3.3 Warping of Plane Curves

When there is a set of functions with a common pattern any deviations from this pattern will either be caused by some between-subject variation or random noise. It can be useful to remove some of this between-subject variation and noise by



**Figure 3.7:** Average curvature curve for the 71 profiles (left). Reconstructed average profile (right).

aligning the curves so that features common in all (or the majority of) functions are shifted to a common point on the  $s$  axis.

When looking at shape data it is often of interest to shift the curvature functions so that anatomically or geometrically important landmarks are aligned. Aligning in this way has certain advantages. Firstly, plotting aligned curvature functions allows for a simple graphical display of the variation in curvature at the important points. Also it makes it possible to produce an average curvature function which shows the typical curvature irrespective of individual variations in position of these landmarks. There are numerous techniques for aligning functions according to corresponding important landmarks, with various advantages and disadvantages, which range in complexity. The one outlined here is relatively simple and uses the ideas of Gasser et al. (1990).

Once curvature functions are aligned to the average position of the important landmarks the major differences between the functions is in amplitude. It can be informative to investigate the amount of amplitude adjustment required to produce the average function from each of the individual functions. The technique which will be outlined to carry out this procedure here will involve adding an amplitude adjustment function to the individual aligned function.

### 3.3.1 Warping technique for functions

The technique of aligning functions according to important landmarks is called (position) warping. The anatomically or geometrically important landmarks to which all the functions are aligned are called characteristic points. These characteristic points can be defined using *a priori* knowledge of the shape of the functions. However, it is often safer and more informative to estimate these characteristic points from the data, typically as stationary points. In the curvature function example points of maximum or minimum curvature will clearly be potential points of interest.

Suppose the aim is to align  $n$  functions,  $f_i(s)$  where  $i = 1, \dots, n$ , to the characteristic points which are to be estimated from the data as points where the majority of functions have a maximum turning point. For each subject all local maxima are calculated such that

$$max_i = (m_{1i}, \dots, m_{p_i i})$$

contains the  $p_i$  local maxima for curve  $i$ . The local maxima for all the curves are combined so that

$$MAX = (m_{11}, \dots, m_{p_1 1}, m_{12}, \dots, m_{p_n n})$$

contains all  $\sum_{i=1}^n p_i$  local maxima in the set of functions.

The frequency of the occurrence of all local maxima across  $s$  is computed as a kernel probability density and is plotted. The plot of the kernel probability density will have modes where many of the functions have a maximum turning point at roughly the same position on the  $s$  axis and these modes indicate potential characteristic points. The width of the mode provides information on the scatter of the maximum turning point across the functions and the area indicates in what proportion of functions it has occurred. It is not necessarily the case that all modes indicate anatomically interesting characteristic points so it is possible a subset of the peaks is sufficient to align all interesting anatomical points. Clearly this process could be repeated using minima (or inflection points).

The average positions on the  $s$  axis (given by the position of the modes of the kernel probability density) of the  $k$  (say) characteristic points are denoted as

$$\bar{\gamma} = (\bar{\gamma}_1, \dots, \bar{\gamma}_k)$$

The positions on the  $s$  axis of the characteristic points of each individual function are calculated. The position of the characteristic points for subject  $i$  (say) are denoted by

$$\gamma_i = (\gamma_{1i}, \dots, \gamma_{ki})$$

For smooth functions with a common pattern it should be relatively simple to find the positions of all turning points. If it is not possible to calculate a certain characteristic point for a subject then this characteristic point is regarded as missing and the warping for that subject is carried out using the remaining  $k - 1$  characteristic points.

Once  $\bar{\gamma}$  and all  $\gamma_i$  have been calculated the warping is relatively simple. The idea is that the characteristic points on each individual curve are aligned to the position of the average characteristic points. A ‘position warping function’  $g_i(s)$  can be used to transform each curve individually. This is done by setting  $g_i(\gamma_{ri}) = \bar{\gamma}_r$  such that  $f_i(g_i(\gamma_{ri})) = f_i(\gamma_{ri})$  where  $r = 1, \dots, k$ , and interpolating between these points to give a continuous warping function. This results in the position of the characteristic points on individual curves being shifted to the position of the average characteristic points. For simplicity  $g$  will be used to denote the warped arc length  $g(s)$ .

Once position warping has removed much of the position variation, the major difference between the functions is in the amplitude of curvature. The average curvature function can be calculated, using piecewise averaging, from the aligned functions. Then it is possible to produce amplitude adjustment functions which match the individual aligned functions to the average function and give an indication of how far, in terms of amplitude, the individual function is from the average. Given the aligned curvature function  $f_i(g)$  and the average aligned function  $\hat{f}(g)$  then there is an amplitude adjustment function  $h_i(g)$  such that

$$\begin{aligned} f_i(g) + h_i(g) &= \hat{f}(g) \\ h_i(g) &= \hat{f}(g) - f_i(g) \end{aligned}$$

The amplitude adjustment function contains the difference in amplitude between the average function and the individual function and can be calculated by finding the piecewise difference at regular points on the  $s$  axis and interpolating between these differences. If the individual function is the same as the average function then  $h_i(g) = 0$  for all  $g$ . In the areas where the amplitude of the individual function is larger than the average  $h_i$  will be negative and in the areas where

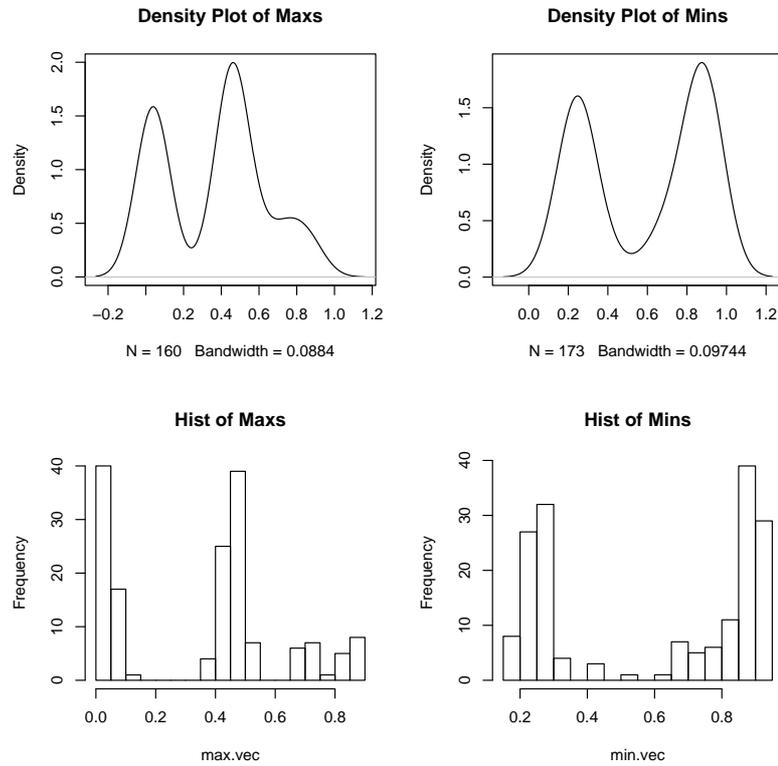
the amplitude of the individual function is smaller than the average  $h_i$  will be positive.

### 3.3.2 Warping of curvature functions: Midline profile example

Figure 3.6 in Section 3.2.3 shows the curvature functions of 71 midline profiles belonging to one year old control children which all appear to follow a similar pattern with turning points at the base and tip of the nose. It is useful for further analysis, in particular to look at the variation in the magnitude of curvature at the turning points, to align the curvature functions so that points of interest are all at the average position. In this example potential characteristic points will be chosen to be points where the majority of functions have either a minimum or maximum turning point.

To calculate the position of the characteristic points all local maxima and minima on each of the 71 individual curvature functions are pooled together, separately for maxima and minima, and the kernel probability density of the occurrence of both maxima and minima calculated. Figure 3.8 shows the kernel probability density plots and histograms for the occurrence of both maxima and minima in the curvature functions.

The kernel probability density plots in Figure 3.8 show that there are two points where the majority of curvature functions have maximum turning points and two points where the majority of curvature functions have minimum turning points. The points where the majority of functions have minimum turning points correspond to the area at the base of the nose where the profile is turning to the right and the area between the eyes (where the profile is also turning to the right). The points where the majority of curvature functions have maximum turning points correspond to the point at the tip of the nose where the profile is turning to the left and a point at the start of the profile which is where the profile leaves the top of the upper lip. This implies that there are four potential characteristic points. However, by considering the curvature functions it is the minimum at the base of the nose and the maximum at the tip of the nose which have the largest magnitude of curvature out of the four turning points. It seems sensible then to align the curvature functions using just these two points as the characteristic points. It would, however, also be perfectly reasonable to use all four turning points.



**Figure 3.8:** Kernel probability density plots and histograms for the occurrence of both maximum (left) and minimum (right) turning points.

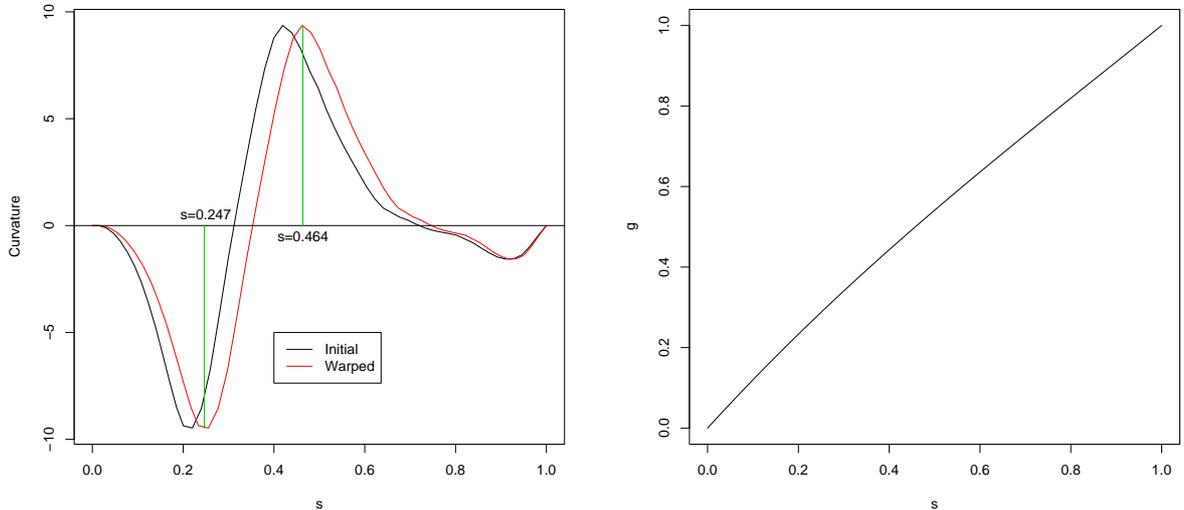
The average position of the characteristic points, given by the corresponding mode of the kernel probability density, are  $s = 0.247$  for the minimum curvature at the base of the nose and  $s = 0.464$  for the maximum curvature at the tip of the nose. The position of the individual characteristic points are found by extracting the appropriate local maxima and minima.

The warping functions can be produced by setting

$$\begin{aligned}
 g_i(0) &= 0 \\
 g_i(\gamma_1) &= \bar{\gamma}_1 \\
 g_i(\gamma_2) &= \bar{\gamma}_2 \\
 g_i(1) &= 1
 \end{aligned}$$

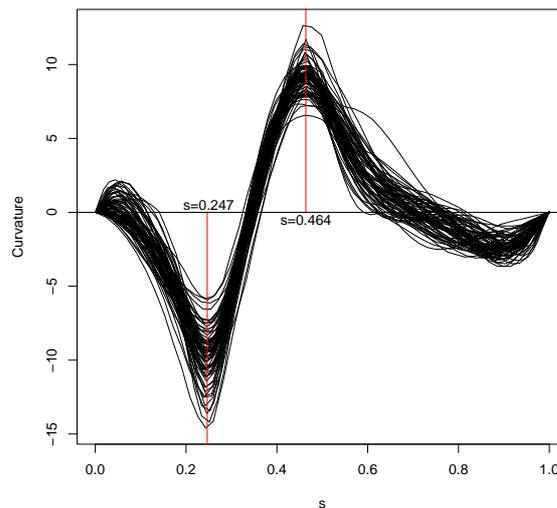
and using cubic spline interpolation between the four points to give a smooth warping function. In fact any form of interpolation can be used however cubic splines give a relatively smooth and accurate interpolation. Figure 3.9 shows the

effect of warping on the curvature function of the example profile used in Section 3.2 and the warping function used to produce the aligned curvature function.



**Figure 3.9:** Warped example curvature function (left) and the corresponding warping function (right).

Figure 3.9 shows that the warped curvature function is the initial curvature function with the characteristic points shifted to the position on the  $s$  axis of the average characteristic points. The main area of interest is to investigate the collection of warped curvature functions for all 71 cases. A plot of the warped curvature functions can be found in Figure 3.10.

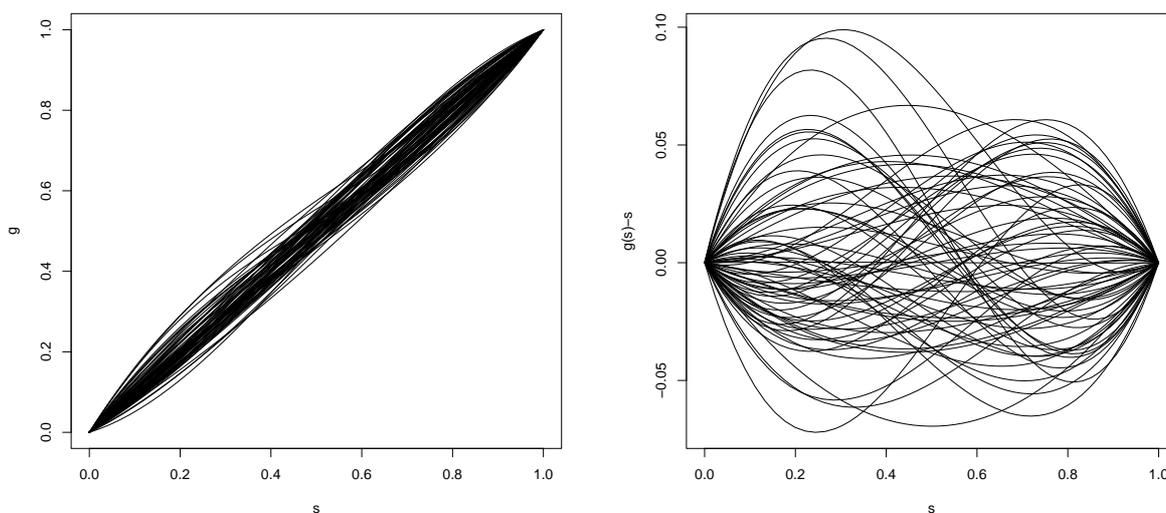


**Figure 3.10:** Aligned curvature functions for the midline profiles of 71 one year old control children.

Considering the aligned curvature functions in Figure 3.10 it seems that the

variation in the magnitude of the curvature at the base of the nose is larger than the variation of curvature at the tip of the nose.

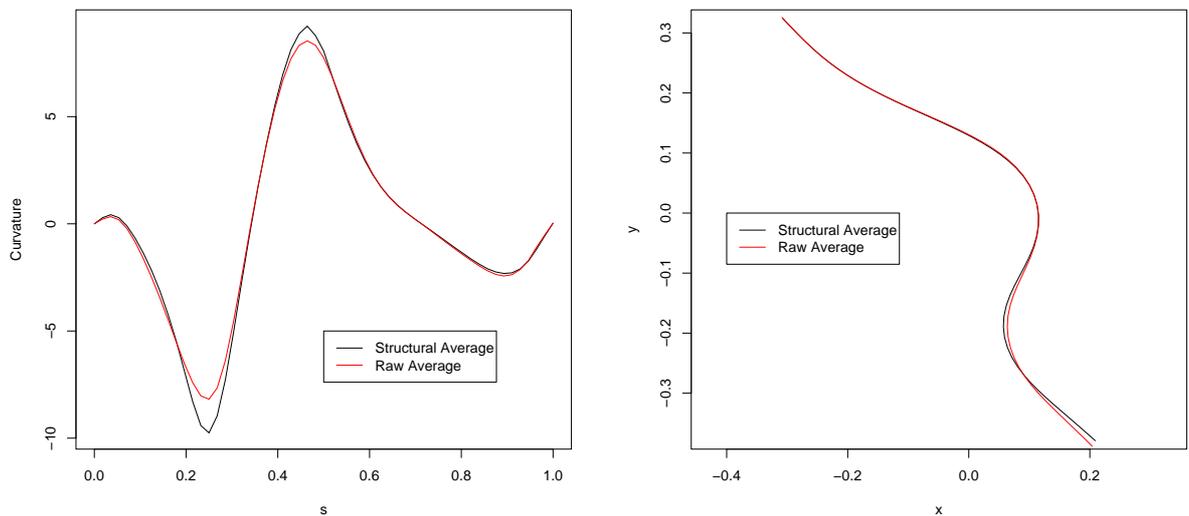
It can also be of interest to further investigate the warping functions that are used to produce the aligned functions. If no warping is required, the warping function is linear (i.e. it lies along the line of equality) and the more non-linear the warping function is the more warping required for aligning. The warping functions to align the curvature functions are shown in Figure 3.11 (left). It is difficult to interpret much from the warping functions in this situation apart from the expected observation that some curvature functions have required more warping than others. To aid interpretation  $s$  is subtracted from  $g(s)$  and these functions are plotted against arc length in Figure 3.11 (right). Deviation of  $g(s) - s$  from zero corresponds to deviation of  $g(s)$  from the line of equality. This perhaps shows more clearly the variation in the warping functions. Using more characteristic points may have resulted in more interesting warping functions.



**Figure 3.11:** Warping functions (left) and warping function minus arc length (right), for the curvature functions for the midline profiles the 71 control children.

Figure 3.7 in Section 3.2.3 showed the average curvature function and the average profile this curvature function produced when the average was calculated as a piecewise average of the original curvature functions. To reduce the information about the magnitude of curvature at the characteristic points which is lost due to piecewise averaging of the original curvature functions, the piecewise average of the aligned functions is taken. Averaging of this form is called structural averaging. Figure 3.12 shows a comparison between the structural average

and the raw average for the 71 curvature functions and a comparison between the average profiles reconstructed from these curvature functions. Note that the average profiles using the raw and structural average are Procrustes matched without scaling.

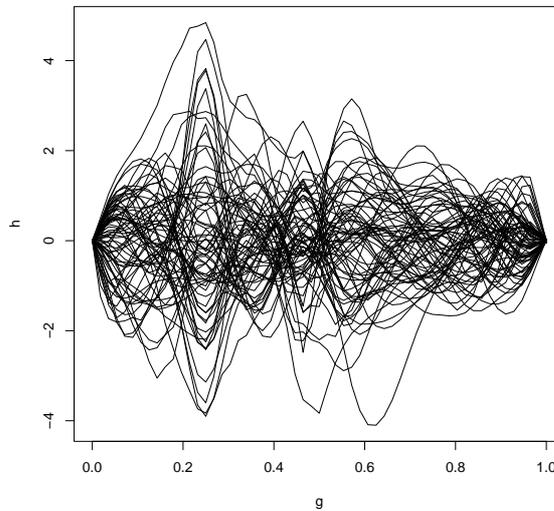


**Figure 3.12:** Raw and structural average for the curvature functions for the midline profiles of 71 one year old control children (left) and the reconstructed profiles from these curvature functions (right).

The curvature functions in Figure 3.12 show that when using the structural average the average function shows greater magnitude, and hence a better representation of average curvature, at the two characteristic points than when using the raw average. The average profiles show that although there were differences between the raw and the structural average at both the characteristic points it appears that the differences at the minimum characteristic point have the biggest effect on the reconstructed profiles with the structural average profile appearing more curved in the area from the bottom of the profile to the tip of the nose than the profile constructed using the raw average.

Finding the amplitude adjustment functions required to exactly produce the structural average curvature function from each of the individual functions gives an indication of the variation between the aligned functions in terms of amplitude. The amplitude adjustment functions for this example are shown in Figure 3.13.

Figure 3.13 shows little trend in the amplitude adjustment functions. There is however an indication that the amplitude adjustment functions have larger variation around the base of the nose than at the tip of the nose suggesting that there is more variability between the individuals in terms of curvature at the base



**Figure 3.13:** Amplitude adjustment functions to produce the structural average curvature function from the individual curvature functions for the midline profiles of 71 one year old control children.

of the nose.

## 3.4 Calculating Curvature: Alternative Methods

Section 3.1.4 explained how the curvature of a plane curve can be calculated using the first and second derivatives of the  $x$  and  $y$  position with respect to the arc length  $s$ . This technique is both relatively straightforward and effective for smooth plane curves. However the calculation of the second derivatives in  $\mathbf{R}$  is not always reliable for less smooth functions. Furthermore, when investigating space curves in Chapter 4 third derivatives, which are even less reliably calculated in  $\mathbf{R}$ , are required for the derivative calculation of curvature. This section will outline two possible alternative techniques which avoid the calculation of high order derivatives and use the midline profile example of Section 3.2 to illustrate them.

### 3.4.1 Optimisation method

Section 3.1.5 described how the original plane curve can be reconstructed up to location and rotation from the curvature function and the arc length of each point. It seems natural to suggest that a method for calculating the curvature

function is to find the curvature function which gives the optimum reconstruction of the original profile.

Consider the example profile. The arc length of each of the 57 points which represent the profile are available; therefore, all that is required to reconstruct the profile is the curvature at these 57 points  $(\kappa(s_1), \dots, \kappa(s_n))$ . Now think of each  $\kappa_i$  where  $i = 1, \dots, n$  as unknown parameters. If the starting position is set such that  $a = 0$  and  $b = 0$  and the starting angle is set such that  $\phi(0) = 0$  the  $\kappa_i$ 's can then be optimised so as to minimise the ordinary Procrustes sum of squares given by

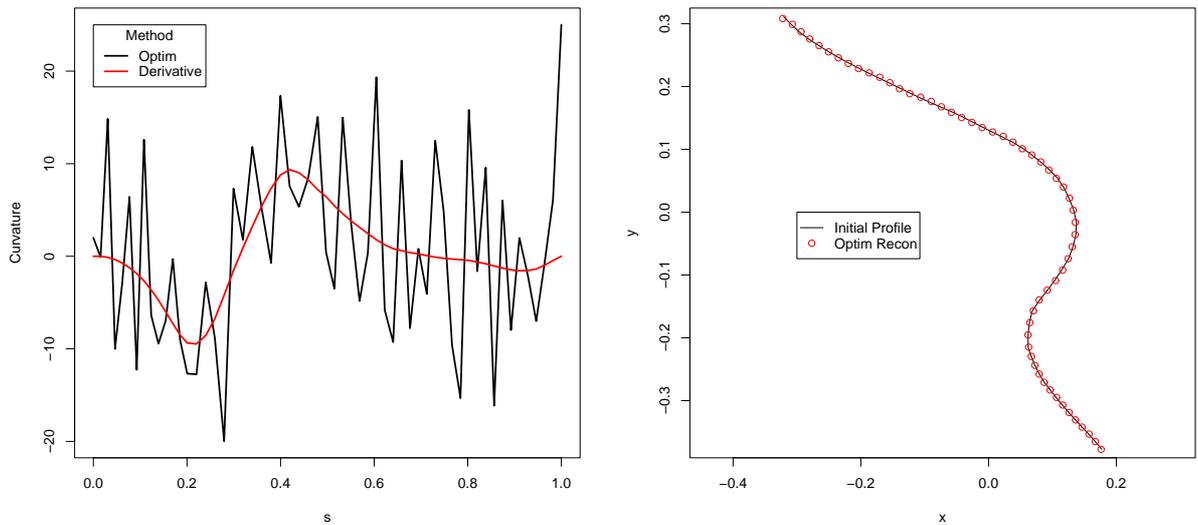
$$OSS = \sum_{i=1}^{57} \left[ (\hat{A}(i, x) - \hat{B}(i, x))^2 + (\hat{A}(i, y) - \hat{B}(i, y))^2 \right]$$

where  $\hat{A}$  is the centered configuration of the original profile and  $\hat{B}$  is the Procrustes registered configuration of the reconstructed profile.

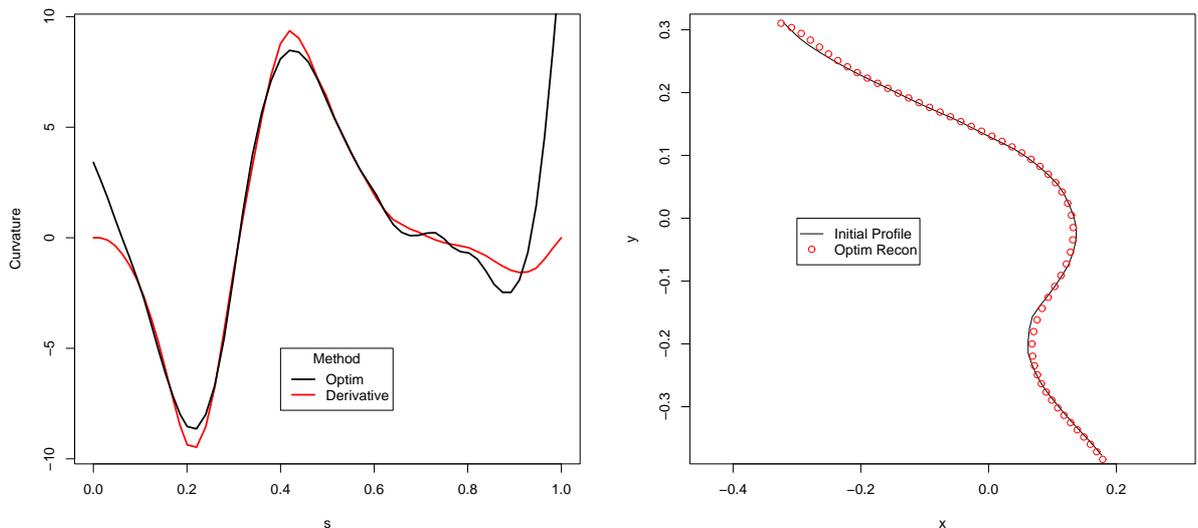
The optimisation is carried out using the `optimize` function in **R** which uses a combination of golden section search and successive parabolic interpolation to find the optimal parameters (see Brent (1973) for details). The disadvantage of this method is that it can be slow. This optimisation was carried out on the example profile and since the reconstruction function is relatively simple the optimisation stage was relatively quick. The curvature function returned by this method and the reconstruction calculated using this function can be found in Figure 3.14.

It can be seen from Figure 3.14 that although the curvature calculated using the optimisation method gives a near perfect reconstruction of the original profile the curvature function is extremely jagged and of little use for further analysis. To rectify this, the curvature function can be smoothed using smoothing splines; however it is important to check that the smoothed curvature function still produces an accurate reconstruction. The smoothed curvature function from the optimisation for the example profile and the corresponding reconstructed profile are shown in Figure 3.15. Note that 10 degrees of freedom are used to define the amount of smoothing.

The initial impression from Figure 3.15 is that when the curvature function for the optimisation is smoothed problems appear at the end of the curve (or top of the profile). Except for this issue, the curvature function of the smoothed optimisation method is very close to the curvature function from the derivative method apart from at the extrema where the optimisation method shows lower



**Figure 3.14:** Comparison of the curvature of the example profile calculated using the derivative and the optimisation method (left) and the reconstruction of the profile using the curvature calculated by the optimisation method (right).



**Figure 3.15:** Comparison of the curvature of the example profile calculated using the derivative and the smoothed optimisation method (left) and the reconstruction of the profile using the curvature calculated by the smoothed optimisation method (right).

magnitude of curvature. This may be due to the fact that 10 degrees of freedom results in over-smoothing of the curvature function. The degrees of freedom could be adjusted, however, it appears that there is sufficient evidence to suggest that the optimisation method provides a method for calculating curvature which results in values similar to the derivative method and produces an accurate

reconstruction of the original profile.

### 3.4.2 Frenet method

The Frenet formulae in Section 3.1.3 show how curvature can be calculated if the tangent and normal vectors are known, while Section 3.1.2 showed how the tangent and normal vectors are defined. The tangent vector can be calculated with relative simplicity while the normal vector is simply given by the derivative of the tangent vector with respect to arc length. Assuming the condition that the normal vector must be of unit length there are two possible normal vectors. The convention is that the normal is set anti-clockwise from the tangent.

The first stage of this method is to find the tangent vectors. Say that the curve  $f(s) = (x(s), y(s))$  parameterised by arc length is defined by  $n$  points. To calculate the direction of the tangent vector (for a point  $f(s_i)$ ) it is necessary to have a point  $f(s_i + h_i)$  where  $h_i$  is very small. To do this a spline can be used to define the curve so that the curve is defined by an extra  $l$  points between each of the  $n$  original points where  $l$  is a large number. The distance between each of the points between  $f(s_i)$  and  $f(s_{i+1})$  is given by  $d = \text{dist}(f(s_i), f(s_{i+1})) \times \frac{1}{l+2}$ . Now as  $d \rightarrow 0$ , which happens as  $l$  increases, the direction given when moving from  $f(s_i)$  to  $f(s_i + d)$  becomes the direction of the tangent vector. The tangent vector is then given by the unit length vector in this direction from  $f(s_i)$ . Clearly it is not possible to find the point  $f(s_n + h_n)$ . Therefore the direction of the tangent vector is set by the direction given when moving from  $f(s_{n-h_n})$  to  $f(s_n)$  and the tangent vector is given by the unit length vector in this direction from  $f(s_n)$ . Note that  $l$  is only the same between each  $f(s_i)$  and  $f(s_{i+1})$  if the original  $n$  points were regularised (i.e. the arc length between them is, or is close to, the same). If this is not the case the size of  $l$  will have to differ between different points to ensure that the distance between  $f(s_i)$  and  $f(s_{i+h_i})$  is the same for all  $i$ . The tangent vector at point  $f(s_i)$  will be denoted as  $\mathbf{t}(s_i)$ .

The next stage of the method is to calculate the normal vectors. To do this (at  $f(s_i)$  say) use the fact that, since  $\mathbf{t}(s_i)$  and the normal vector ( $\mathbf{n}(s_i)$ ) are perpendicular,  $\mathbf{t}(s_i) \cdot \mathbf{n}(s_i) = 0$  and also that  $|\mathbf{n}(s_i)| = 1$ . This gives the following equations

$$\begin{aligned} [x(\mathbf{t}(s_i)) \times x(\mathbf{n}(s_i))] + [y(\mathbf{t}(s_i)) \times y(\mathbf{n}(s_i))] &= 0 \\ x(\mathbf{n}(s_i))^2 + y(\mathbf{n}(s_i))^2 &= 1 \end{aligned}$$

where the only unknowns are the  $x$  and  $y$  position of the normal vector. Simple algebra leads to two possible positions for the normal vectors, one of which must be chosen.

$\kappa(s)$  is calculated at each point using (3.9) and by differentiating  $\mathbf{n}(s) \cdot \mathbf{t}(s) = 0$  with respect to  $s$  to give (as shown by Gray (1998))

$$\begin{aligned} \mathbf{n}'(s) \cdot \mathbf{t}(s) + \mathbf{n}(s) \cdot \mathbf{t}'(s) &= 0 \\ \mathbf{n}'(s) \cdot \mathbf{t}(s) &= -\mathbf{n}(s) \cdot \mathbf{t}'(s) \\ \mathbf{n}'(s) \cdot \mathbf{t}(s) &= -\mathbf{n}(s) \cdot \kappa(s)\mathbf{n}(s) \\ \mathbf{n}'(s) \cdot \mathbf{t}(s) &= -\kappa(s) \end{aligned} \tag{3.19}$$

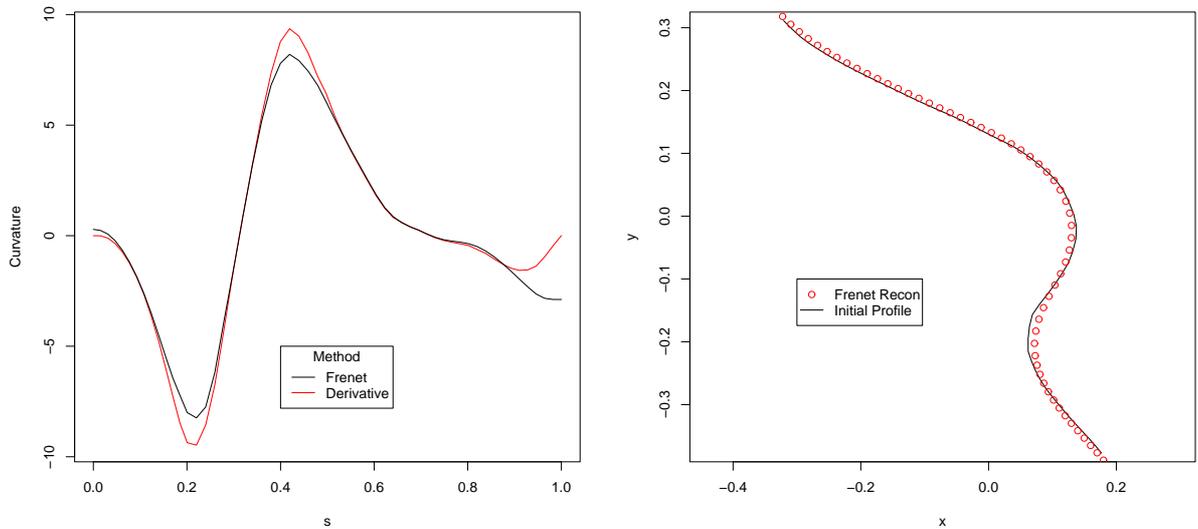
Since there are two choices for the normal vector (3.19) can also be written as

$$\mathbf{n}'(s) \cdot \mathbf{t}(s) = \kappa(s) \tag{3.20}$$

To illustrate this method the midline profile example is used. To calculate the direction of the tangent vector  $l$  is taken to be 9998 between each pair of neighbouring points since the original profile is approximately regularised. The normal vectors at each of the points are anti-clockwise perpendicular to the tangent vectors. The normal vectors will therefore be pointing into the face, meaning that if the profile is bending into the face curvature will be positive and if the profile is bending away from the face curvature will be negative.

Curvature is calculated at each of the 57 points using (3.20) with the first derivative of the tangent vector with respect to the arc length calculated using a smoothing spline with 10 degrees of freedom and the `predict` function. The curvature function and the reconstructed profile using this curvature function are shown in Figure 3.16.

The curvature functions in Figure 3.16 indicate that the major difference between the derivative method and the Frenet method for calculating curvature is in the magnitude of curvature at the extrema. More evidence of this is given by the fact that the reconstructed profile appears to be less curved than the original profile at the base of the nose and the tip of the nose. This is potentially due to 10 degrees of freedom resulting in over smoothing for the Frenet method. The degrees of freedom could be adjusted to produce a curvature function which provides a more accurate reconstruction of the original nose profile. However, it is clear from Figure 3.16 that the Frenet method produces curvature functions which



**Figure 3.16:** Comparison of the curvature of the example profile calculated using the derivative and the Frenet method (left) and the reconstruction of the profile using the curvature calculated by the Frenet method (right).

are similar to those from the derivative method and that provide a reasonable reconstruction of the original nose profile.

### 3.5 Concluding Remarks on Plane Curves

This chapter has shown that the shape of a plane curve can be represented by the curvature experienced by the curve as the arc length changes. Furthermore, it has shown that a plane curve can be reconstructed from the curvature function up to the similarity transformations. Three methods have been introduced to allow curvature to be calculated. Of the three methods it would appear that the derivative method is the simplest to use and the curvature values, even when the curvature functions have been smoothed to give functions useful for further analysis, can be used to reconstruct the original plane curve accurately. For most practical situations it seems reasonable to use the derivative method. However there may be situations where the other methods are preferable. For example if the interest was in finding the curvature function which allows the most accurate reconstruction of the original curve then the optimisation method before smoothing would be the best option. An interesting point to note is that the Frenet method only requires the calculation of first derivatives whereas the derivative method requires second derivatives. The numerical calculation of second derivatives is not always

straightforward or accurate especially when the function is not smooth. In the case of the midline profile this is not an issue as the functions of both the  $x$  and  $y$  position against arc length were smooth. However, for other features this may not always be the case and in these situations the Frenet method should be considered.

This chapter has also shown the advantages of warping the curvature functions so that characteristic points are aligned, particularly in terms of producing an average curvature function which more accurately represents the average of the group. The method outlined was simple; however thought should always go into which points are chosen as characteristic points. The decision should be based on both mathematical and anatomical reasoning to allow easily interpretable results.

# Chapter 4

## Analysis of Space Curves

### 4.1 Characterising Space Curves

A space curve is a curve which is contained in three-dimensional space. In this chapter space curves will be investigated with the aim of representing them using a measurement of how much the curve is bending at particular points on the curve. The bending in a space curve at a certain point can be measured using two scalar values called curvature and torsion. Curvature is a measure of how much the curve bends to one side in its current plane whereas torsion is a measure of how much the curve bends out of the plane it is currently in. Before calculating curvature and torsion for space curves some terms must be described in detail.

#### 4.1.1 Arc length

Calculating arc length in a space curve is simply an extension of the methods in Section 3.1.1. Suppose the arbitrarily parameterised curve  $r(a) = [x(a), y(a), z(a)]$  is defined by  $n$  points  $(a_1, \dots, a_n)$  say) on the curve. If the coordinates of any two points  $a_b$  and  $a_c$  are  $[x(a_b), y(a_b), z(a_b)]$  and  $[x(a_c), y(a_c), z(a_c)]$  respectively then the distance between these two points is

$$Dist(a_b, a_c) = \sqrt{[x(a_b) - x(a_c)]^2 + [y(a_b) - y(a_c)]^2 + [z(a_b) - z(a_c)]^2}$$

Now approximate the curve as the line which connects the points  $a_1, \dots, a_n$  then the length of the curve can be approximated as the total distance of the line

which connects the points

$$L = \sum_{i=1}^{n-1} \sqrt{[x(a_{i+1}) - x(a_i)]^2 + [y(a_{i+1}) - y(a_i)]^2 + [z(a_{i+1}) - z(a_i)]^2}$$

If the curve is rectifiable, as  $n$  increases the total distance of the line which connects the points will approach the true length of the curve. The total distance a point is along the curve, calculated in this way, is the arc length of the curve at this point. Therefore the arc length ( $s$  say) at any point  $q$  will be given by

$$s_q = \sum_{i=1}^{q-1} \sqrt{[x(a_{i+1}) - x(a_i)]^2 + [y(a_{i+1}) - y(a_i)]^2 + [z(a_{i+1}) - z(a_i)]^2} \quad (4.1)$$

### 4.1.2 Osculating plane

At each point on the curve there is a plane which ‘kisses’ the curve. This plane is called the osculating plane and is defined by Frost and Wolstenholme (2001) as ‘a plane passing through three adjacent points assuming a limiting position, when the three points are ultimately coincident.’

To calculate the osculating plane at point  $i$  (say) use the points  $r(a_i)$ ,  $r(a_{i+h})$  and  $r(a_{i+2h})$  where  $h$  is an arbitrary scalar value. Now let  $\mathbf{p} \cdot \mathbf{q} = b$  denote the plane which passes through the points  $r(a_i)$ ,  $r(a_{i+h})$  and  $r(a_{i+2h})$  where  $\mathbf{p}$  is a generic point on the plane,  $\mathbf{q}$  is a vector orthogonal to the plane and  $b$  is a constant. Then the function

$$f(a) = r(a) \cdot \mathbf{q} - b$$

has zeros at  $a = a_i$ ,  $a = a_{i+h}$  and  $a = a_{i+2h}$ . Now using Rolle’s theorem since  $f(a_i) = 0$ ,  $f(a_{i+h}) = 0$  and  $f(a_{i+2h}) = 0$  then

$$\begin{aligned} f'(l) &= 0 & l &\in (a_i, a_{i+h}) \\ f'(m) &= 0 & m &\in (a_{i+h}, a_{i+2h}) \end{aligned}$$

and further

$$f''(u) = 0 \quad u \in (l, m)$$

Therefore as  $h \rightarrow 0$  then  $l$ ,  $m$  and  $u$  all tend to  $i$  so the limiting values for  $\mathbf{q}$  and

$b$  can be obtained by

$$\begin{aligned} f(a_i) &= r(a_i) \cdot \mathbf{q} - b = 0 \\ f'(a_i) &= r'(a_i) \cdot \mathbf{q} = 0 \\ f''(a_i) &= r''(a_i) \cdot \mathbf{q} = 0 \end{aligned}$$

Thus the vectors  $\mathbf{p} - r(a_i)$ ,  $r'(a_i)$  and  $r''(a_i)$  belong to the same plane. Therefore the osculating plane is defined by

$$\begin{vmatrix} p_x - x(a_i) & p_y - y(a_i) & p_z - z(a_i) \\ x'(a_i) & y'(a_i) & z'(a_i) \\ x''(a_i) & y''(a_i) & z''(a_i) \end{vmatrix} = 0 \quad (4.2)$$

### 4.1.3 Tangent, normal and binormal vectors

Suppose now that there is a curve  $r(s) = [x(s), y(s), z(s)]$  which is parameterised by arc length  $s$ . In a direct extension of the planar method in Section 3.1.2 the tangent vector  $\mathbf{t}(l)$  is defined as the unit length vector starting at  $r(l) = [x(l), y(l), z(l)]$  and passing through  $r(l + h_1) = [x(l + h_1), y(l + h_1), z(l + h_1)]$  where  $h_1$  is very small.

The (principal) normal vector  $\mathbf{n}(s)$  is a unit vector which is perpendicular to the tangent vector in the osculating plane. Clearly there are two potential normal vectors which lie in the osculating plane. By convention the normal vector is set so that the curve is turning towards the direction of the normal vector at each point although this need not always be the case.

The binormal vector is a unit vector which is perpendicular to both the tangent vector and the normal vector i.e. it is perpendicular to the osculating plane. The binormal vector at the point  $r(s)$  will be denoted by  $\mathbf{b}(s)$ .

It is useful to think of the tangent vector, the normal vector and the binormal vector making up a moving trihedron of the curve. From this trihedron the curvature measures the rate at which the tangent vector turns and torsion measures the rate at which the binormal vector turns.

### 4.1.4 Curvature, torsion and the Frenet formulae

Curvature of a space curve  $\kappa(s)$  is a scalar measurement of the magnitude of the bending of the curve within the osculating plane at a point as the point moves along the curve. Torsion of a space curve  $\tau(s)$  is a scalar measurement of the

amount that the curve bends out of the osculating plane at a point as the point moves along the curve. Alternatively torsion can be thought of as the amount that the osculating plane changes as the point moves along the curve.

Suppose that there is a curve  $r(s) = [x(s), y(s), z(s)]$  which is parameterised by arc length  $s$ . Gray (1998) shows that the tangent vector  $\mathbf{t}(s)$  measures the rate of change of the position of the curve as the point moves along the curve (i.e. the rate of change of the position of the curve with respect to arc length). Therefore

$$\mathbf{t}(s) = \frac{dr(s)}{ds} \quad (4.3)$$

The magnitude of the rate of change of the position of the curve with respect to arc length will always be 1 since the change in the position of the curve and the arc length are equivalent. Therefore,  $\mathbf{t}(s)$  is a vector of unit length.

Section 4.1.3 has already stated that curvature measures the rate at which the tangent vector turns i.e the rate of change of the tangent vector with respect to the distance travelled round the curve. Therefore

$$\kappa(s) = \left| \frac{d\mathbf{t}(s)}{ds} \right| \quad (4.4)$$

Returning to (3.4) it can be seen that  $\mathbf{t}'(s)$  is perpendicular to  $\mathbf{t}(s)$  therefore from Pressley (2001)

$$\mathbf{n}(s) = \frac{d\mathbf{t}(s)}{ds}$$

gives the principal normal vector (i.e. vector perpendicular to the tangent vector in the osculating plane). However since (4.4) has shown that  $\left| \frac{d\mathbf{t}(s)}{ds} \right|$  is equal to  $\kappa(s)$  to ensure that  $\mathbf{n}(s)$  is unit length it is required that

$$\mathbf{n}(s) = \frac{1}{\kappa(s)} \frac{d\mathbf{t}(s)}{ds} \quad (4.5)$$

Now since  $\mathbf{t}(s)$  and  $\mathbf{n}(s)$  are perpendicular unit vectors in the osculating plane, the binormal vector  $\mathbf{b}(s)$  can be defined as

$$\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{n}(s)$$

$\{\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)\}$  give the orthonormal basis of the curve.

Now again Pressley (2001) shows that  $\frac{d\mathbf{b}(s)}{ds}$  is parallel to  $\mathbf{n}(s)$ .

$$\begin{aligned}\frac{d\mathbf{b}(s)}{ds} &= \frac{d(\mathbf{t}(s) \times \mathbf{n}(s))}{ds} \\ &= \frac{d\mathbf{t}(s)}{ds} \times \mathbf{n}(s) + \mathbf{t}(s) \times \frac{d\mathbf{n}(s)}{ds}\end{aligned}$$

Using (4.5) it is clear that

$$\frac{d\mathbf{t}(s)}{ds} \times \mathbf{n}(s) = \kappa(s) [\mathbf{n}(s) \times \mathbf{n}(s)] = 0$$

Therefore

$$\frac{d\mathbf{b}(s)}{ds} = \mathbf{t}(s) \times \frac{d\mathbf{n}(s)}{ds} \quad (4.6)$$

Equation 4.6 shows that  $\frac{d\mathbf{b}(s)}{ds}$  is perpendicular to  $\mathbf{t}(s)$ . It is also by definition perpendicular to  $\mathbf{b}(s)$ . A vector which is perpendicular to both  $\mathbf{t}(s)$  and  $\mathbf{b}(s)$  must be parallel to  $\mathbf{n}(s)$ . Therefore,

$$\frac{d\mathbf{b}(s)}{ds} = -\tau(s)\mathbf{n}(s) \quad (4.7)$$

for a scalar  $\tau(s)$ . This  $\tau(s)$  is the torsion at point  $s$ . Note that the minus sign is just convention and that torsion can only be defined if curvature is non-zero. If torsion is zero for the whole curve then the curve is a plane curve.

Finally consider the rate of change of the normal vector with respect to the arc length i.e.  $\frac{d\mathbf{n}(s)}{ds}$ . Pressley (2001) shows that

$$\begin{aligned}\frac{d\mathbf{n}(s)}{ds} &= \frac{d(\mathbf{b}(s) \times \mathbf{t}(s))}{ds} \\ &= \frac{d\mathbf{b}(s)}{ds} \times \mathbf{t}(s) + \mathbf{b}(s) \times \frac{d\mathbf{t}(s)}{ds} \\ &= -\tau(s) [\mathbf{n}(s) \times \mathbf{t}(s)] + \kappa(s) [\mathbf{b}(s) \times \mathbf{n}(s)] \\ &= -\kappa(s)\mathbf{t}(s) + \tau(s)\mathbf{b}(s)\end{aligned} \quad (4.8)$$

since  $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$  this allows for the change of sign. Combining (4.5), (4.7) and (4.8) leads to the three dimensional Frenet equations

$$\frac{d\mathbf{t}(s)}{ds} = \kappa(s)\mathbf{n}(s) \quad (4.9)$$

$$\frac{d\mathbf{n}(s)}{ds} = -\kappa(s)\mathbf{t}(s) + \tau(s)\mathbf{b}(s) \quad (4.10)$$

$$\frac{d\mathbf{b}(s)}{ds} = -\tau(s)\mathbf{n}(s) \quad (4.11)$$

### 4.1.5 Calculating curvature and torsion in practice

Section 4.1.4 has shown that at each point the values of curvature and torsion are affected by the positions of the tangent, normal and binormal vectors. In practice it is often computationally difficult to manually set the Frenet frame at each point and so it is useful to produce a simpler method for calculating the curvature and torsion which uses the derivatives of  $x$ ,  $y$  and  $z$  against  $s$ . Gray (1998) explains how this is done for an arbitrarily parameterised curve. However in this section, and also throughout the thesis, this will be simplified to consider the case where the curve is parameterised by arc length.

For simplicity the first, second and third derivatives of the position of the curve with respect to  $s$  (i.e.  $\frac{dr(s)}{ds}$ ,  $\frac{d^2r(s)}{ds^2}$  and  $\frac{d^3r(s)}{ds^3}$ ) will be denoted by  $r'(s)$ ,  $r''(s)$  and  $r'''(s)$  respectively where  $r(s) = [x(s), y(s), z(s)]$  and  $r'(s) = [x'(s), y'(s), z'(s)]$ . Section 4.1.4 shows that

$$\begin{aligned} r'(s) &= \mathbf{t}(s) \\ r''(s) &= \frac{d\mathbf{t}(s)}{ds} = \kappa(s)\mathbf{n}(s) \end{aligned}$$

Now consider the cross product of  $r'(s)$  and  $r''(s)$ . This gives

$$\begin{aligned} r'(s) \times r''(s) &= \mathbf{t}(s) \times \kappa(s)\mathbf{n}(s) \\ &= \kappa(s)(\mathbf{t}(s) \times \mathbf{n}(s)) \\ r'(s) \times r''(s) &= \kappa(s)\mathbf{b}(s) \end{aligned} \quad (4.12)$$

Taking the magnitude of both sides of (4.12) shows that

$$\begin{aligned} |r'(s) \times r''(s)| &= |\kappa(s)\mathbf{b}(s)| \\ &= \kappa(s) |\mathbf{b}(s)| \\ \kappa(s) &= |r'(s) \times r''(s)| \end{aligned} \quad (4.13)$$

Therefore the curvature of a space curve parameterised by arc length can be calculated as the magnitude of the cross product of the first and second derivative of the position of the curve with respect to arc length.

Now consider the third derivative of the position of the curve with respect to

arc length.

$$r'''(s) = (\kappa(s)\mathbf{n}(s))' = \kappa'(s)\mathbf{n}(s) + \kappa(s)\mathbf{n}'(s)$$

Using (4.10) and the fact that  $\mathbf{t}(s)$ ,  $\mathbf{n}(s)$  and  $\mathbf{b}(s)$  are mutually perpendicular it follows that

$$\begin{aligned} \mathbf{b}(s) \cdot r'''(s) &= \mathbf{b}(s) \cdot [\kappa'(s)\mathbf{n}(s) + \kappa(s)\mathbf{n}'(s)] \\ &= \kappa'(s)[\mathbf{b}(s) \cdot \mathbf{n}(s)] + \kappa(s)[\mathbf{b}(s) \cdot \mathbf{n}'(s)] \\ &= 0 + \kappa(s)[\mathbf{b}(s) \cdot \{-\kappa(s)\mathbf{t}(s) + \tau(s)\mathbf{b}(s)\}] \\ &= \kappa(s)[-\kappa(s)\{\mathbf{b}(s) \cdot \mathbf{t}(s)\} + \tau(s)\{\mathbf{b}(s) \cdot \mathbf{b}(s)\}] \\ &= \kappa(s)[0 + \tau(s)] \\ \mathbf{b}(s) \cdot r'''(s) &= \kappa(s)\tau(s) \end{aligned} \tag{4.14}$$

Using (4.12) and (4.13) this becomes

$$\begin{aligned} \left( \frac{r'(s) \times r''(s)}{\kappa(s)} \right) \cdot r'''(s) &= \kappa(s)\tau(s) \\ \frac{1}{\kappa(s)^2} ((r'(s) \times r''(s)) \cdot r'''(s)) &= \tau(s) \\ \tau(s) &= \frac{((r'(s) \times r''(s)) \cdot r'''(s))}{|r'(s) \times r''(s)|^2} \end{aligned} \tag{4.15}$$

Therefore the torsion of a space curve parameterised by arc length can be calculated from an equation (4.15) which uses the first, second and third derivative of the position of the curve with respect to arc length.

### 4.1.6 Reconstructing a curve from curvature and torsion

Gray (1998) shows that a space curve can be reconstructed up to rigid transformations from its curvature and torsion however the process is more complex than for plane curves. Suppose that the Frenet formulae ((4.9), (4.10) and (4.11)) and the first derivative of the position of the curve with respect to the arc length (4.3) are rewritten so that

$$\begin{aligned} r'_i(s) &= \mathbf{t}_i(s) \\ \mathbf{t}'_i(s) &= \kappa(s)\mathbf{n}_i(s) \\ \mathbf{n}'_i(s) &= -\kappa(s)\mathbf{t}_i(s) + \tau(s)\mathbf{b}_i(s) \\ \mathbf{b}'_i(s) &= -\tau(s)\mathbf{n}_i(s) \end{aligned}$$

where  $i = (1, 2, 3)$  represents the  $x$ ,  $y$  and  $z$  values of each vector. This gives a system of 12 differential equations. The initial conditions of this system of differential equations are

$$\begin{aligned} r_i(s_1) &= l_i \\ \mathbf{t}_i(s_1) &= m_i \\ \mathbf{n}_i(s_1) &= n_i \\ \mathbf{b}_1(s_1) &= m_2 n_3 - m_3 n_2 \\ \mathbf{b}_2(s_1) &= m_3 n_1 - m_1 n_3 \\ \mathbf{b}_3(s_1) &= m_1 n_2 - m_2 n_1 \end{aligned}$$

where  $\sum_{i=1}^3 m_i^2 = \sum_{i=1}^3 n_i^2 = 1$  and  $\sum_{i=1}^3 m_i \cdot n_i = 0$ . The initial conditions choose an arbitrary starting point and ensure that  $\mathbf{t}(s)$ ,  $\mathbf{n}(s)$  and  $\mathbf{b}(s)$  are mutually perpendicular vectors of unit length. Since there is a system of ordinary differential equations with initial conditions it is possible to find a unique solution. Boyce and DiPrima (1992) show that if curvature and torsion are constant it is possible to solve this system of differential equations. However, since curvature and torsion change along the curve the differential equations must be solved numerically. The Runge-Kutta method is both a ‘relatively simple and sufficiently accurate’ (Boyce and DiPrima (1992)) technique to solve a set of differential equations of this kind. The Runge-Kutta formula assumes that the values of the series of differential equations are known at a point  $s_n$  and calculates the values at  $s_{n+1}$  using a weighted average of values taken at different points in the interval  $s_n \leq s \leq s_{n+1}$ . To express the method rewrite the system of ordinary differential equations so that

$$\mathbf{Y}'(s_i)^T = \beta(s_i)\mathbf{Y}(s_i)^T \quad (4.16)$$

where

$$\mathbf{Y}(s_i) = [\mathbf{t}_1(s_i), \mathbf{n}_1(s_i), \mathbf{b}_1(s_i), \mathbf{t}_2(s_i), \mathbf{n}_2(s_i), \mathbf{b}_2(s_i), \mathbf{t}_3(s_i), \mathbf{n}_3(s_i), \mathbf{b}_3(s_i), r_1(s_i), r_2(s_i), r_3(s_i)]$$

and

$$\beta(s_i) = \begin{bmatrix} 0 & \kappa(s_i) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\kappa(s_i) & 0 & \tau(s_i) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\tau(s_i) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa(s_i) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\kappa(s_i) & 0 & \tau(s_i) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\tau(s_i) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \kappa(s_i) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\kappa(s_i) & 0 & \tau(s_i) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\tau(s_i) & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general Runge-Kutta formula can be found in Boyce and DiPrima (1992). More specifically, for the reconstruction of the curve problem it becomes

$$\mathbf{Y}(s_{n+1})^T = \mathbf{Y}(s_n)^T + \frac{h}{6}(k_{n1} + 2k_{n2} + 2k_{n3} + k_{n4}) \quad (4.17)$$

where

$$\begin{aligned} k_{n1} &= \beta(s_n)\mathbf{Y}(s_n)^T \\ k_{n2} &= \beta(s_n + \frac{1}{2}h)\mathbf{Y}(s_n + \frac{1}{2}hk_{n1})^T \\ k_{n3} &= \beta(s_n + \frac{1}{2}h)\mathbf{Y}(s_n + \frac{1}{2}hk_{n2})^T \\ k_{n4} &= \beta(s_n + h)\mathbf{Y}(s_n + hk_{n3})^T \end{aligned}$$

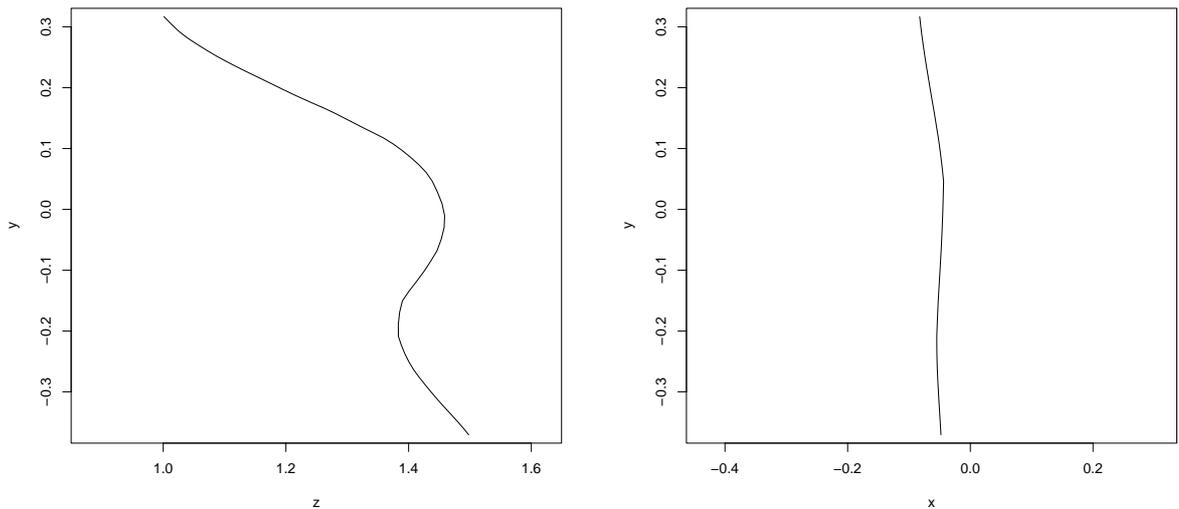
and  $h$  is the distance between point  $n$  and point  $n + 1$ . The sum  $\frac{1}{6}(k_{n1} + 2k_{n2} + 2k_{n3} + k_{n4})$  can be interpreted as the average movement in the curve with  $k_{n1}$  accounting for the movement at  $s_n$ ,  $k_{n2}$  and  $k_{n3}$  accounting for the movement at the midpoint between  $s_n$  and  $s_{n+1}$  and  $k_{n4}$  accounting for the movement at  $s_{n+1}$ .

Using the curvature values ( $\kappa(s) = [\kappa(s_1), \dots, \kappa(s_n)]$ ), the torsion values ( $\tau(s) = [\tau(s_1), \dots, \tau(s_n)]$ ) and the arc lengths ( $s_1, \dots, s_n$ ) it is possible to reconstruct a curve up to rigid transformations. The initial values of the system of differential equations are  $\mathbf{Y}(s_1)$ . Using the Runge-Kutta method it is possible to calculate  $\mathbf{Y}(s_2)$  and then repeat until  $\mathbf{Y}(s_n)$  has been calculated. This gives the

position along with the tangent, normal and binormal vectors at all  $n$  points on the curve.

## 4.2 Curvature and Torsion of a Space Curve: Midline Profile Example

To investigate curvature and torsion in a space curve the midline profile example from Section 3.2 will be extended from two-dimensions to three-dimensions. The profile is defined by 57 data points with the  $x$  axis showing movement across the face, the  $y$  axis showing movement up and down the face and the  $z$  axis showing movement in and out of the face. To illustrate the methods an example profile will be used. This is the same profile which was used to illustrate the methods in Section 3.2 and is plotted in Figure 4.1 from side on and from straight on.



**Figure 4.1:** Plot of the example profile from side on (left). Plot of example profile from straight on (right).

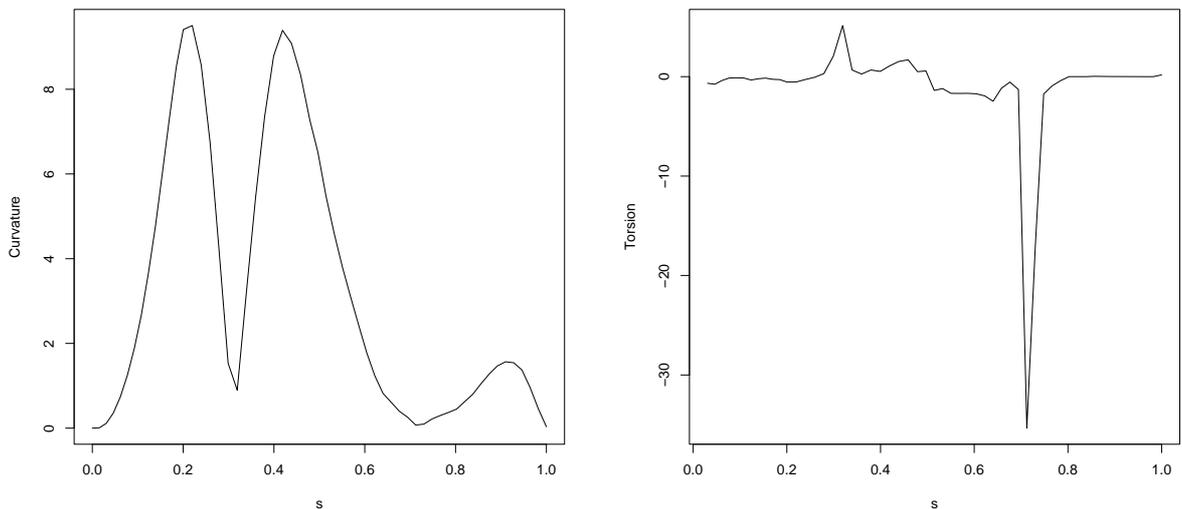
The start of the curve is taken as the lowest point on the  $y$  axis (i.e. the point at the top of the upper lip) and arc length increases as the curve moves towards the top of the face. All profiles are rescaled so that the three-dimensional arc length is equal to 1. The method for doing this is a simple extension of the rescaling method from Section 3.2.1 with the  $z$  coordinate also included.

Chapter 3 described three methods for calculating the curvature of a plane curve. These methods will be extended to space curves to allow calculation of curvature and torsion. Furthermore a new method involving two perpendicular

planes will be introduced.

### 4.2.1 Derivative method

Section 4.1.5 defined two formulas, (4.13) and (4.15), to calculate the curvature and torsion of a space curve using the first, second and third derivatives of the  $x$ ,  $y$  and  $z$  position with respect to arc length. Taking each direction individually this can be done numerically by using smoothing splines. The curvature and torsion of the example profile are calculated using (4.13) and (4.15) and smoothing splines with 10 degrees of freedom. The curvature and torsion functions are shown in Figure 4.2. Note that the torsion value calculated for the first point was greater than 600 so this is removed to aid the representation.



**Figure 4.2:** Plot of curvature (left) and torsion (right) function for the example profile.

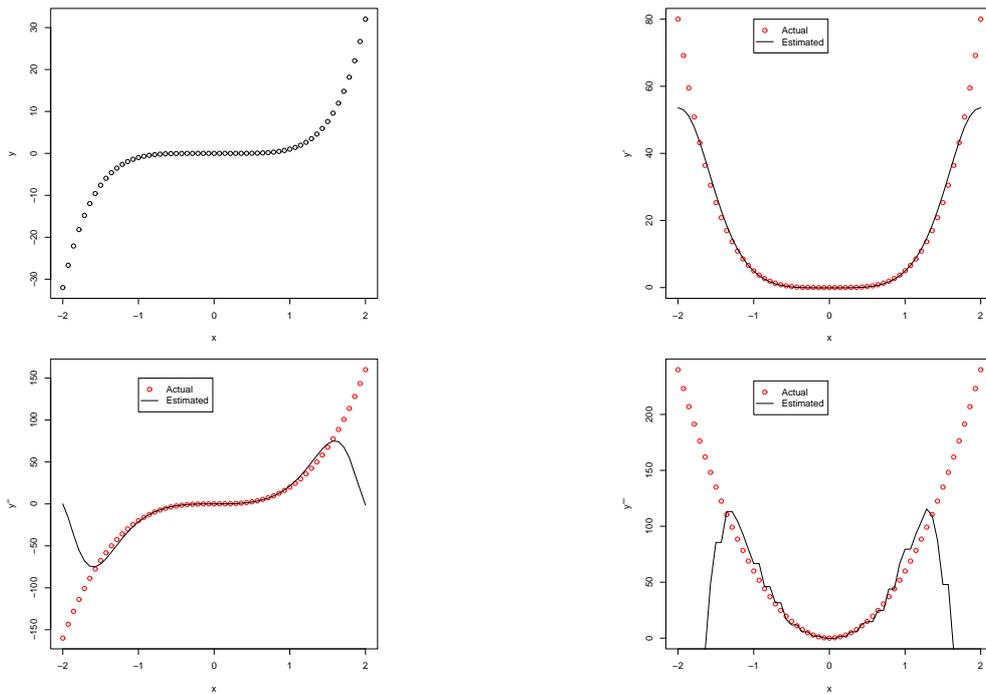
From an initial inspection the curvature function from Figure 4.2 is close to what would be expected with two areas of large bending corresponding to the base of the nose and the tip of the nose. The torsion function however was extremely large at the start of the profile and further shows a small area of very large magnitude torsion which is unexpected considering the smoothness of the midline profile. To check the accuracy of the curvature and torsion functions they are used to reconstruct the profile using the methods of Section 4.1.6. The initial values chosen (as they will be throughout the thesis) are

$$r(s_1) = (0, 0, 0)$$

$$\begin{aligned} \mathbf{t}(s_1) &= (1, 0, 0) \\ \mathbf{n}(s_1) &= (0, 1, 0) \\ \mathbf{b}(s_1) &= (0, 0, 1) \end{aligned}$$

This results in a reconstructed space curve which is not at all like the three-dimensional midline profile.

It may be that the torsion function is incorrect due to higher order derivatives not being accurately estimated using the smoothing spline technique in  $\mathbf{R}$ . To investigate this consider the simple curve  $y = x^5$  with  $-2 \leq x \leq 2$  defined by 57 points. Clearly  $y' = 5x^4$ ,  $y'' = 20x^3$  and  $y''' = 60x^2$ . Figure 4.3 shows the estimated first, second and third derivatives of  $y$  with respect to  $x$  using 10 degrees of freedom and compares the estimates to the actual values of the derivatives.



**Figure 4.3:** Plot of  $y = x^5$  and the comparison between the estimate and the actual values of the first, second and third derivatives.

Figure 4.3 shows that, excluding a few serious problems at the start and end of the curve, the smoothing spline technique is fairly accurate at estimating the first and second derivatives of the simple curve. However the technique seems to have serious problems estimating the third derivative. The difficulties experienced here occurred when estimating the derivatives of a simple curve. In practice the curves will be more complex. This suggests that the torsion values calculated

using this method may be questionable. The derivative technique for calculating curvature and torsion of a space curve runs into difficulties due to the issue of the inaccuracy of estimating higher order derivatives using the smoothing spline method.

There are a number of adjustments which could be made to attempt to improve the process of calculating the higher order derivatives. Recall that the smoothing of the  $x$ ,  $y$  and  $z$  position against arc length has been carried out using natural cubic splines. Considering that for torsion derivatives as high as order three are required it would perhaps be beneficial to increase the order of the splines. The estimate of the third derivative in Figure 4.3 illustrates the problem. Since the order of the spline is the same as the order of the derivative the estimate of the third derivative is constant in many areas. This illustrates the fact that the order of the spline should at least be larger than the order of derivative to be calculated.

In the example the same degrees of smoothing have been used when calculating the first, second and third derivatives. The calculation of the higher order derivatives however, is less robust than the calculation of lower order derivatives. Therefore, it would appear reasonable to apply more smoothing when calculating the higher order derivatives. Adjusting the smoothing for the different order of derivatives would potentially stabilise the process of estimating the high order derivatives and allow a more stable estimate of the curvature and torsion functions.

Section 3.2.1 briefly introduced the idea of p-splines proposed by Eilers and Marx (1996). P-splines allow adaptive knot selection which offers greater flexibility in the smoothing process. Smoothing using p-splines is known to give a straightforward and reliable method for estimating the derivatives of the smooth function. Replacing natural cubic splines with p-splines may improve the estimation of higher order derivatives and therefore the estimated curvature and torsion functions.

It is clear that a number of adjustments could be made to improve the estimation of the higher order derivatives. However, even allowing for these adjustments, high order derivatives are difficult to calculate accurately and robustly. Therefore, the two alternative methods introduced for plane curves in Section 3.4 are extended to space curves to remove the need for calculation of higher order derivatives.

## 4.2.2 Optimisation method

Section 4.1.6 outlined a method which can be used to reconstruct a curve from the curvature and torsion functions. It seems natural to attempt to calculate the curvature and torsion functions to allow the reconstruction of the original curve with the greatest possible accuracy. Consider the example midline profile where the curve is defined by 57 points. Since the arc length at each point is available all that is required to produce a reconstruction, using the Runge-Kutta method from Section 4.1.6, of the original profile is the curvature  $\kappa(s_1), \dots, \kappa(s_n)$  and torsion  $\tau(s_1), \dots, \tau(s_n)$  at each point. That means the aim is to find the  $\kappa_i$ 's and  $\tau_i$ 's which minimise the ordinary Procrustes sum of squares given by

$$OSS = \sum_{i=1}^{57} \left[ (\hat{A}(i, x) - \hat{B}(i, x))^2 + (\hat{A}(i, y) - \hat{B}(i, y))^2 + (\hat{A}(i, z) - \hat{B}(i, z))^2 \right]$$

where  $\hat{A}$  is the centered configuration of the original profile and  $\hat{B}$  is the Procrustes registered configuration of the reconstructed profile and the sum is over all 57 points which define the midline profile curve. To simplify the problem the  $\kappa_i$ 's are not forced to be positive i.e. the normal vectors do not have to be set so that the curve is bending towards them.

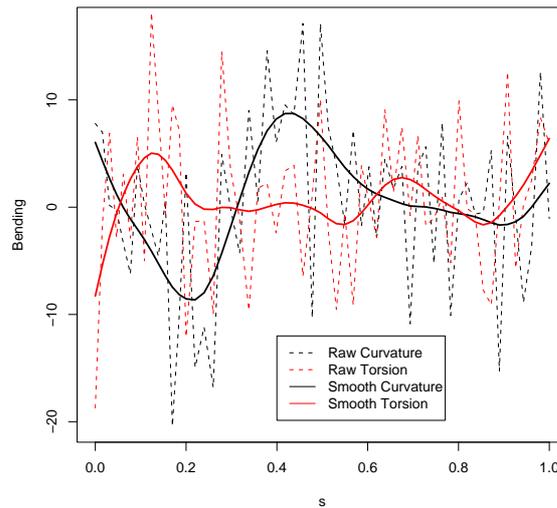
Once again the `optimize` function can be used to carry out this optimisation. However optimising over 114 parameters (57 curvature values and 57 torsion values) requires a lot of computing power and will require many iterations of the function. This process was carried out on 5 control profiles. The results for the example profile will be shown below but Table 4.1 shows the length of time taken to carry out the optimisation procedure for each of the 5 control profiles.

| Profile | Time Taken (hours) |
|---------|--------------------|
| 1       | 3.06               |
| 2       | 2.71               |
| 3       | 2.95               |
| 4       | 3.06               |
| 5       | 2.85               |
| Mean    | 2.93               |

**Table 4.1:** Time taken to carry out the optimisation method on 5 example profiles.

Table 4.1 shows that on average carrying out this optimisation method on a

control midline profile takes nearly 3 hours. Considering that in a study there could be numerous subjects the considerable computing time required is a potential drawback of this method. However, it is still of interest to investigate how the optimisation method performs on the example profile. Figure 4.4 shows curvature and torsion functions calculated using the optimisation method.

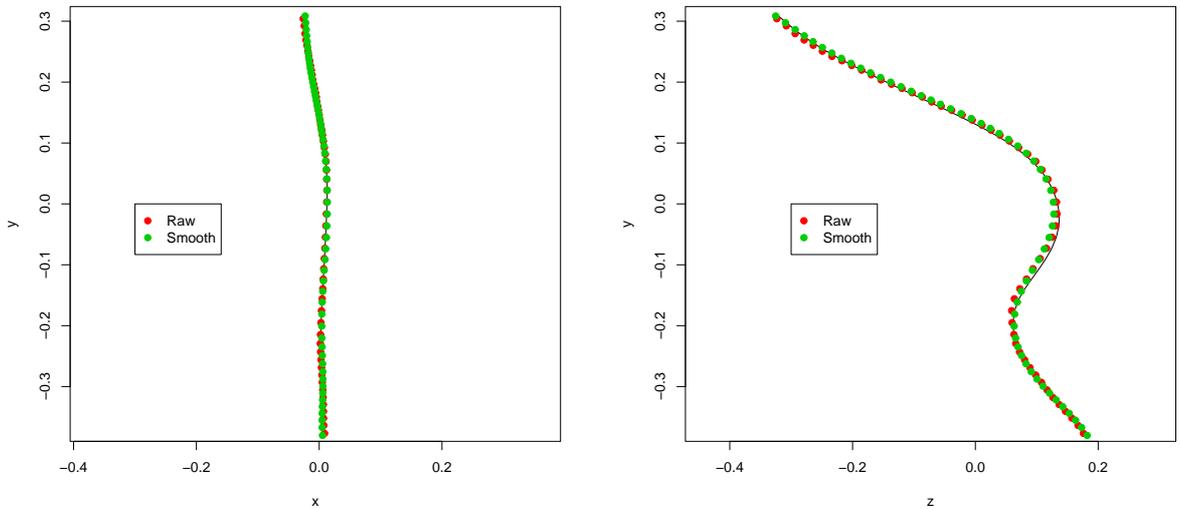


**Figure 4.4:** Plot of both the raw and smoothed curvature and torsion functions for the example profile calculated using the optimisation method.

The raw curvature and torsion functions in Figure 4.4 are extremely rough and are difficult to analyse. Therefore smoothing is applied to the functions with 10 degrees of freedom to produce functions which are much easier to interpret. The smoothed curvature function seems reasonable with high magnitude curvature at the base and tip of the nose. The smoothed torsion functions seem unrealistically high at the tails but apart from that show only one area of high torsion just before the base of the nose which is not unreasonable.

Using the Runge-Kutta method the original profile is reconstructed using both the raw and smoothed curvature and torsion functions with the reconstructions matched to the original profile, using ordinary Procrustes analysis, and shown in Figure 4.5.

Using both the raw and the smoothed functions it appears that the original profile is reconstructed fairly accurately. It does seem however that the reconstructed profiles come away from the original profile at the area between the base of the nose and the tip of the nose. Somewhat surprisingly this occurs for the reconstruction using the raw functions as well as the smoothed functions. A



**Figure 4.5:** Plot of the reconstruction of the example profile using the raw and smooth curvature and torsion functions calculated using optimisation from side on (left) and front on (right).

positive aspect, however, is that the smoothed functions do not seem to perform particularly worse in the reconstruction than the raw functions.

Calculating curvature and torsion using this optimisation method has the clear advantage of ensuring the curvature and torsion values allow an accurate reconstruction of the original curve. However there are two major disadvantages of this method. Firstly it is computationally expensive and the more points that define the curve the longer the process will take. Secondly the optimisation is simply a numerical method to ensure an accurate reconstruction meaning the anatomical interpretation of curvature and torsion is often unclear.

### 4.2.3 Frenet method

Section 4.1.5 described how the derivative formulae for curvature and torsion came from the directions of the tangent, normal and binormal vectors (the Frenet frame) and further recalling (4.9), (4.10) and (4.11) it is clear that if the Frenet frame is defined at each point of the curve it is simple to calculate curvature and torsion. The process for doing this is not as simple as using the derivative method; however it only requires the first derivative of the normal vector and so it should be more robust.

To calculate the Frenet frame at the point  $r(s_i) = [x(s_i), y(s_i), z(s_i)]$  the osculating plane must be defined. As shown in Section 4.1.2 to calculate the osculating plane the points  $r(s_i + h_1)$  and  $r(s_i + h_2)$  must be found. To do this

a spline can be used to interpolate the curve so that the curve is defined by an extra  $l$  points between each of the  $n$  original points where  $l$  is a large number. If  $l$  is sufficiently large then the osculating plane can be defined by the point  $r(s_i)$  and the next two points given by the spline. For the final point in the curve the previous two points are used. It is then simple to calculate the osculating plane at each point as the plane which contains the point and the two neighbouring points.

Now that there is an equation for the osculating plane it is relatively simple to define the tangent, normal and binormal vectors using the ideas of Section 4.1.3. The first stage is to calculate the tangent vector. The tangent vector for the point at  $s_i$  is the unit length vector which starts at  $r(s_i)$  and passes through  $r(s_i + h_1)$ . For the final point in the curve the tangent vector is a unit length vector which starts at  $r(s_i)$  and travels in the direction defined by the vector which starts at  $r(s_i - h_1)$  and passes through  $r(s_i)$ . The tangent vector, denoted as  $\mathbf{t}(s_i)$ , will by definition lie in the osculating plane.

Recalling that the binormal vector, denoted as  $\mathbf{b}(s_i)$ , is perpendicular to the osculating plane it is clear that the binormal vector is either the unit length normal vector to the osculating plane or the negative of the normal vector to the osculating plane depending on the desired interpretation of torsion.

Finally by recalling that the normal vector, denoted as  $\mathbf{n}(s_i)$ , is perpendicular to both the tangent vector and the binormal vector then the normal vector can be calculated by the cross product of the tangent and binormal vectors

$$\mathbf{n}(s_i) = \mathbf{t}(s_i) \times \mathbf{b}(s_i)$$

or

$$\mathbf{n}(s_i) = \mathbf{b}(s_i) \times \mathbf{t}(s_i)$$

since the cross product of two vectors gives a vector which is mutually perpendicular to both original vectors. Also since both  $\mathbf{t}(s_i)$  and  $\mathbf{b}(s_i)$  are of unit length  $\mathbf{n}(s_i)$  will be of unit length. The choice of the normal vectors is dependent on the desired interpretation of curvature.

Once the tangent, normal and binormal vectors have been calculated at each point the curvature and torsion can be calculated using (4.9) and (4.11). Section 3.4.2 equation (3.19) showed that

$$\kappa(s_i) = -\mathbf{n}'(s_i) \cdot \mathbf{t}(s_i) \tag{4.18}$$

where  $\mathbf{n}'(s_i)$  represents  $\frac{d\mathbf{n}(s)}{ds}$ . Similarly by differentiating  $\mathbf{n}(s_i) \cdot \mathbf{b}(s_i) = 0$  with respect to  $s$  it can be shown that

$$\begin{aligned}\mathbf{n}'(s_i) \cdot \mathbf{b}(s_i) + \mathbf{n}(s_i) \cdot \mathbf{b}'(s_i) &= 0 \\ \mathbf{n}'(s_i) \cdot \mathbf{b}(s_i) &= -\mathbf{n}(s_i) \cdot \mathbf{b}'(s_i) \\ \mathbf{n}'(s_i) \cdot \mathbf{b}(s_i) &= -\mathbf{n}(s_i) \cdot (-\tau(s_i)\mathbf{n}(s_i)) \\ \mathbf{n}'(s_i) \cdot \mathbf{b}(s_i) &= \tau(s_i)\end{aligned}\tag{4.19}$$

Therefore all that is required to calculate the curvature and torsion at each point of the profile is to find the tangent vector, normal vector, binormal vector and the first derivative of the normal vector with respect to arc length.

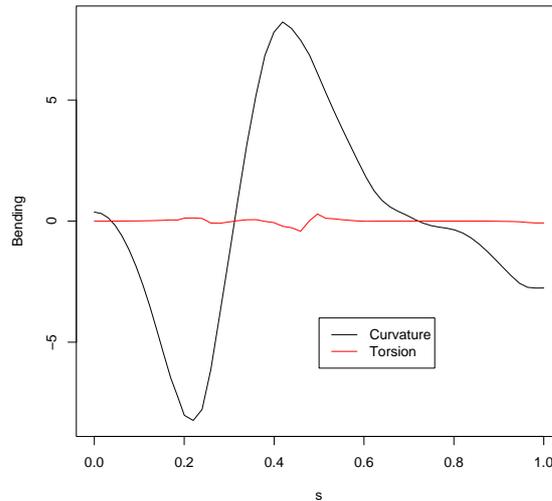
This process was used to calculate the curvature and torsion of the example profile. The osculating planes tended to cut vertically through the face and the binormal vectors were set to point to the left of the face whilst the normal vectors were set to point into the face. This means that the majority of the bending explained by the curvature value is made up of bending into and out of the face whilst the majority of the bending explained by the torsion value is bending across the face. However since the osculating plane twists whilst it moves along the midline profile curvature and torsion do not exclusively have this interpretation.

Figure 4.6 shows the curvature and torsion functions for the example profile with the first derivative of the normal vector calculated using the `predict` function on a smoothing spline with 10 degrees of freedom.

Figure 4.6 shows that the curvature values are in general a lot larger than the torsion values. Since the osculating planes were mainly cutting through the middle of the midline profile the low torsion values correspond to the fact that there is not a lot of movement across the face and that most of the movement into and out of the face can be accounted for by curvature which has a large magnitude at the base and tip of the nose.

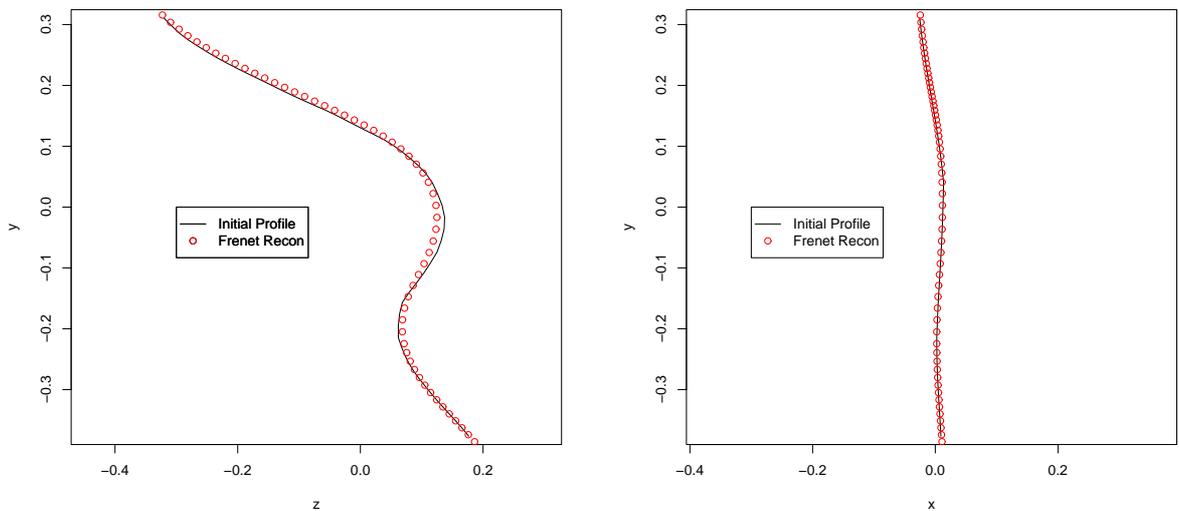
Using the Runge-Kutta method the original profile is reconstructed using the curvature and torsion functions calculated using this Frenet method. The reconstructed profile is matched to the original profile with the result shown in Figure 4.7.

The reconstruction of the original profile using these curvature and torsion functions is relatively accurate. There does however seem to be a slight error in that when looking from side on the reconstruction shows less bending at the base



**Figure 4.6:** Plot of curvature (black) and torsion (red) functions for the example profile.

and tip of the nose. This may be due to the smoothing involved in the process.



**Figure 4.7:** Plot of the reconstruction of the example profile using the curvature and torsion functions calculated using the Frenet method from side on (left) and front on (right).

The Frenet method has produced curvature and torsion functions which are relatively smooth and can be used to give a fairly accurate reconstruction of the original profile. However, the major disadvantage with the method would appear to be that it is difficult to produce a clear anatomical interpretation of curvature and torsion. For the profile it has been suggested that curvature accounts for most of the bending into and out of the face whilst torsion accounts for most of

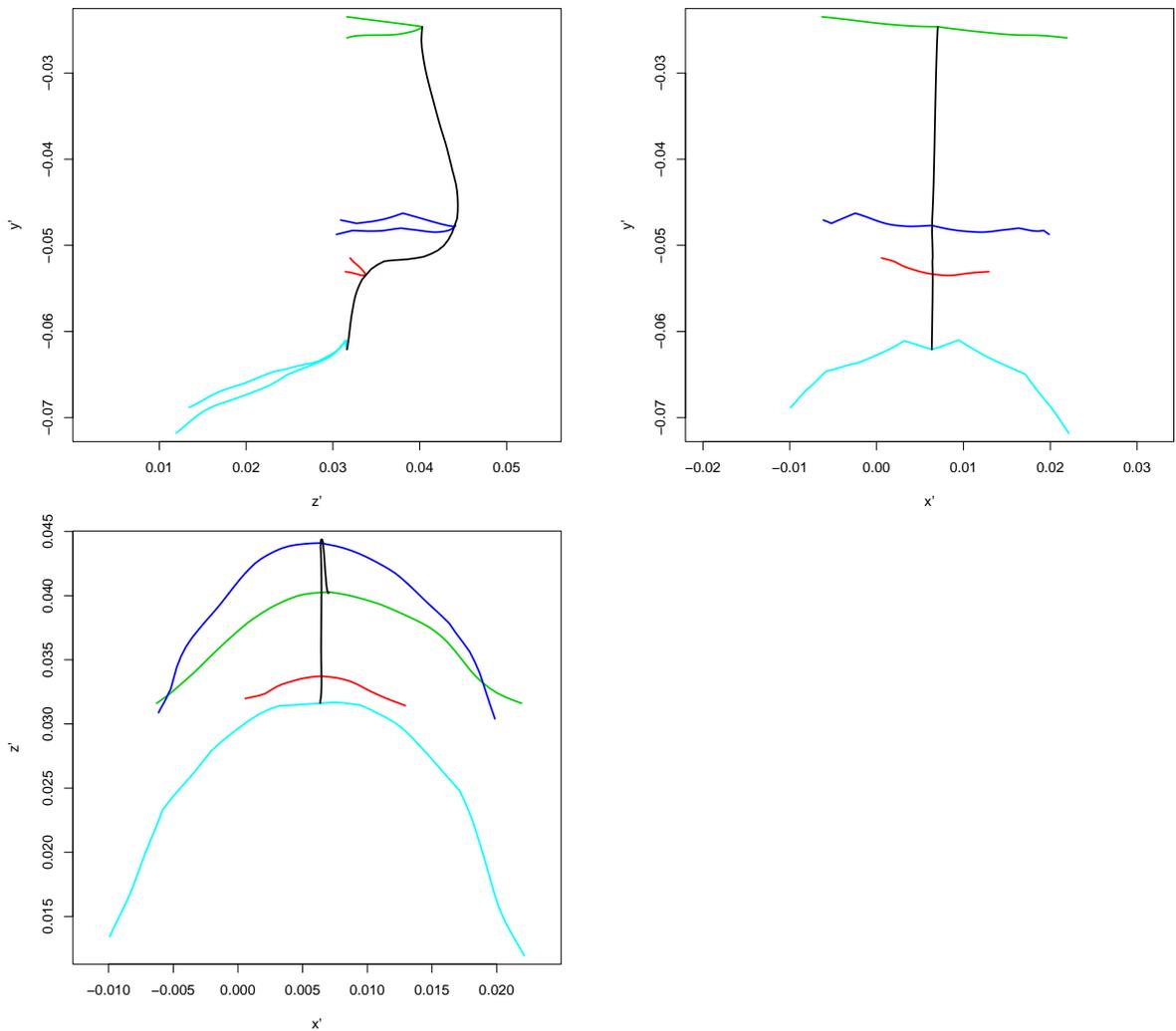
the bending across the face but because the Frenet frame twists as the position moves up the midline profile this anatomical interpretation is not definitive.

#### 4.2.4 Perpendicular plane method

This chapter has already shown some of the difficulties of calculating curvature and torsion in terms of requiring a large amount of computing power and also the difficulty in practically interpreting curvature and torsion. In this section a method is introduced to describe the amount of bending the curve experiences using two easily interpretable curvature values.

Consider the midline profile example. It would be preferable if there was one measure which represented the amount the curve bends into and out of the face as the profile moves up the face and another measure which represents the amount the curves bends across the face (i.e. from side to side) as the profile moves up the face. In *Traité de Géométrie* (1799), Gaspard Monge introduced the chief ideas of descriptive geometry; namely that a three-dimensional object or curve can be represented by projecting onto two perpendicular planes. Therefore it seems reasonable to project the points of the space curve onto two perpendicular planes, one (the  $yz$  plane) which cuts through the face vertically and another (the  $xy$  plane) which lies across the face, and treat the problem as two plane curves. Then bending in the  $yz$  plane is curvature into and out of the face, called  $yz$  curvature, and bending in the  $xy$  plane is curvature traversing the face, called  $xy$  curvature. The two plane curvature values at each point are sufficient to explain the bending experienced by the space curve. Note that to analyse some facial features, such as the upper lip, it is necessary to define a plane (the  $xz$  plane) which is perpendicular to the other two planes and cuts through the face horizontally. Bending in this plane, called  $xz$  curvature, also measures curvature into and out of the face and can be used as an alternative to  $yz$  curvature. The curves of an example child's face projected onto all three planes are shown in Figure 4.8.

The major difficulty is in setting up the perpendicular planes so that the bending in each plane is equivalent in all the profiles. One method for doing this would be to use principal components analysis to set the planes. Consider the plane defined by the first two principal components after carrying out a principal components analysis on the coordinates of the profile. The first two principal components almost completely explain the variation caused by movement up and down the face (component one) and movement in and out of the face (component



**Figure 4.8:** The curves defining the face of a child projected onto the  $yz$  (top left),  $xy$  (top right) and  $xz$  (bottom left) planes.

two). The third principal component, which is by definition perpendicular to the first two components, explains variation caused by movement across the face. Then the plane defined by the first two principal components is a plane which cuts through the face vertically and becomes the  $yz$  plane. The plane defined by the first and third principal components traverses the face and becomes the  $xy$  plane while the plane defined by the second and third principal components cuts through the face horizontally and becomes the  $xz$  plane. A potential problem with this method however is that the planes are not set using the orientation of the face but simply calculated using principal components therefore there may be questions over the consistency between different subjects.

A preferable alternative to either of these methods would be to set up the

perpendicular planes using landmarks on the face. To do this a plane which traverses the face can be calculated using three landmarks on the face. The landmarks chosen are the corner of the left eye, the corner of the right eye and the middle of the upper lip i.e. landmarks  $enL$ ,  $enR$  and  $ls$  shown in Figure 1.1. Since the three dimensional co-ordinates of the three landmarks are available it is simple to calculate the plane. This plane is called the  $xy$  plane.

The  $yz$  plane can then be calculated using the normal vector to the  $xy$  plane, to ensure that the planes are perpendicular, and a further landmark on the face. To ensure that the  $yz$  plane cuts through the centre of the face the landmark at the tip of the nose ( $prn$ ) is chosen. The  $xz$  plane is then the plane which is mutually perpendicular to the  $yz$  and  $xy$  plane.

The next stage of the process is to project the profile onto these three planes. To project a point ( $R = (R_x, R_y, R_z)$  say) from the profile onto the  $yz$  plane the distance between  $R$  and the closest point on the plane ( $Q$  say) must be calculated. The vector from  $A$ , which is a point in the plane, to  $R$  is  $AR (= (R_x - A_x, R_y - A_y, R_z - A_z))$ . Now clearly the direction which will give the shortest distance between  $R$  and  $Q$  is the direction given by the normal vector,  $\mathbf{n}$ . Therefore the distance between  $R$  and  $Q$  is given by  $-\mathbf{n} \cdot -AR = \mathbf{n} \cdot AR$ . From the definition of the dot product this is the projection of  $AR$  onto the direction defined by  $\mathbf{n}$  and therefore the shortest distance between  $R$  and  $Q$ . Now the vector  $RQ$  is a vector of length  $\mathbf{n} \cdot AR$  in the direction of  $-\mathbf{n}$  therefore

$$\begin{aligned} RQ &= (\mathbf{n} \cdot AR) \times -\mathbf{n} \\ RQ &= -(\mathbf{n} \cdot AR) \times \mathbf{n} \end{aligned}$$

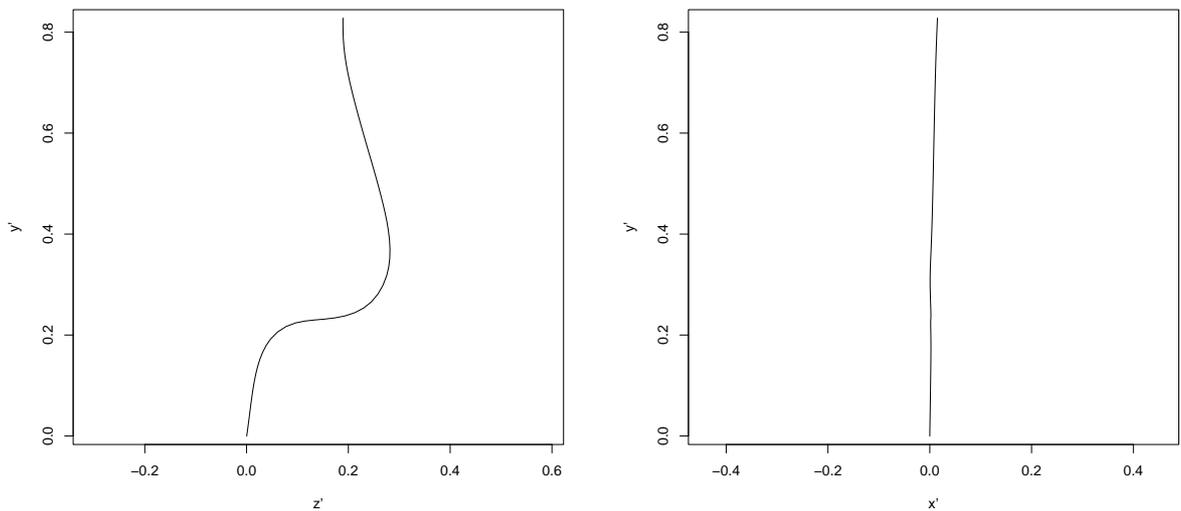
Therefore the position of the projected point is

$$Q = R - ((\mathbf{n} \cdot AR) \times \mathbf{n})$$

If this is done for all points on the profile a projection of the profile onto the  $yz$  plane is produced. The same process can be carried out for the  $xy$  plane and  $xz$  plane.

Now, all that is required to produce the plane projections is to define the axes in the three dimensional space. The origin for the new set of axes is chosen as the point at the middle of the upper lip ( $ls$ ) then the axes of the  $xy$  plane are chosen as the direction of the normal of the  $yz$  plane (call this  $x'$ ) and the direction of the cross product of the normal from the  $yz$  plane and the normal of the  $xy$  plane

(call this  $y'$ ). The final axis is defined by the normal of the  $xy$  plane (call this  $z'$ ). To calculate the position of a point on each axis all that is required is to find the dot product between the position and the unit length vector lying on the axis. A plot of the example profile on both the  $yz$  and  $xy$  planes is shown in Figure 4.9.

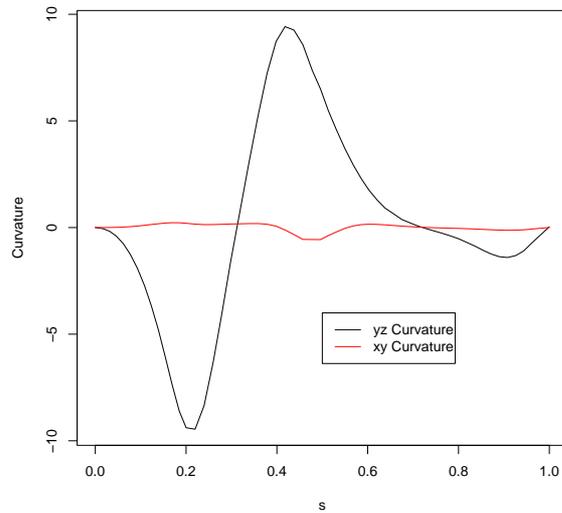


**Figure 4.9:** Plot of the  $yz$  (left) and  $xy$  (right) plane projections for the example profile.

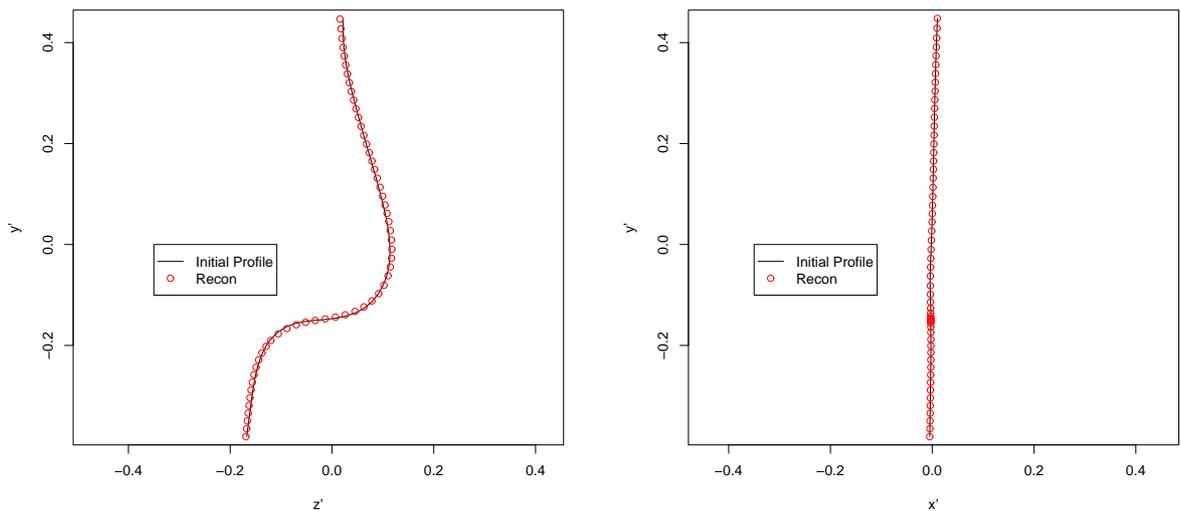
The curvature in both planes can be calculated using the techniques of Section 3.1.4 for plane curves. This allows the curvature both traversing the face and into and out of the face to be calculated. The  $yz$  and  $xy$  curvature functions are calculated using 10 degrees of freedom for smoothing and are shown in Figure 4.10. Note that the  $s$  axis is the fixed total arc length of the three-dimensional curve and not the variable arc length of the plane curves to allow comparison between cases.

The curvature functions in Figure 4.10 unsurprisingly show that the amount of bending into and out of the face is much larger than the bending traversing the face. Further the  $yz$  curvature function shows two areas of high magnitude curvature at the base and tip of the nose. There is perhaps an indication of an area of slightly higher  $xy$  curvature at the tip of the nose.

Section 3.1.5 outlined a technique for reconstructing a plane curve from the curvature function and arc length. Figure 4.11 shows the curves in the perpendicular planes reconstructed from the  $yz$  and  $xy$  curvature functions and aligned to the projected original profile. Note that the arc length used for the reconstruction is the arc length of the original curve in the perpendicular plane and not the total arc length of the space curve.



**Figure 4.10:** Plot of  $yz$  and  $xy$  curvature functions for the example profile where  $s$  is the arc length of the space curve.



**Figure 4.11:** Plot of the reconstructed  $yz$  (left) and  $xy$  (right) plane projections for the example profile.

The reconstruction of the plane projections of the profile using the  $yz$  and  $xy$  curvature functions are accurate. The difficulty here is in how to combine the two plane reconstructions to give an accurate space reconstruction. If the positions of the reconstructed points on the  $y'$  axis are the same for both plane reconstructions then the space reconstruction is simply given using the position of the reconstructed points on the  $x'$ ,  $y'$  and  $z'$  axes. Since the initial directions of the plane reconstructions are given by an initial turning angle the problem of matching the  $y'$  positions for each plane reconstruction, assuming the chosen

initial position is  $(0, 0, 0)$ , becomes the problem of finding the two appropriate initial turning angles. In theory this problem could be solved using the system of equations

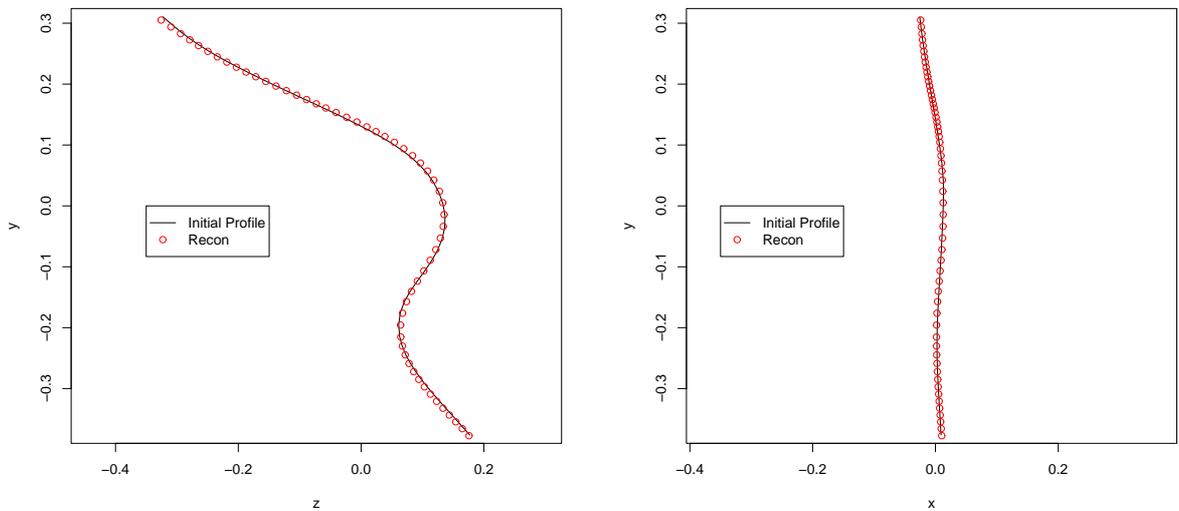
$$\begin{aligned} x(s_i) &= \int_{s=0}^{s_i} \cos\left(\int \kappa(s)ds + \phi_1\right)ds \\ y(s_i) &= \int_{s=0}^{s_i} \sin\left(\int \kappa(s)ds + \phi_1\right)ds = \int_{s=0}^{s_i} \sin\left(\int \kappa(s)ds + \phi_2\right)ds \\ z(s_i) &= \int_{s=0}^{s_i} \cos\left(\int \kappa(s)ds + \phi_2\right)ds \end{aligned}$$

where  $i = 1, \dots, n$  are the  $n$  positions which define the curve and  $\phi_1, \phi_2$  are the initial turning angle of the  $yz$  and  $xy$  plane reconstruction respectively. However, in this set of equations there are  $4n$  equations and  $3(n - 1) + 2$  unknowns (corresponding to all the reconstructed positions excluding the defined starting position and the two initial turning angles) therefore in general there are more equations than unknowns and the system is over-determined.

To avoid having to solve the over-determined system a grid search can be carried out over all possible (i.e. from 0 to  $2\pi$ ) combinations of the two starting angles to find the combination which minimises the sum of squared difference between the  $y'$  positions of the two reconstructions. The `optimize` function can be used to efficiently perform this grid search and find the optimum starting angles. For the example profile the function returns optimum starting angles of  $84.53^\circ$  and  $89.95^\circ$  with the sum of squared difference between the  $y'$  values of  $9.93 \times 10^{-5}$ . Using these starting angles the reconstructed space curve is given by the  $z'$  and  $y'$  values from the  $yz$  plane reconstruction and the  $x'$  values from the  $xy$  plane reconstruction. Figure 4.12 shows the reconstructed curve matched to the original profile on the cartesian axes using ordinary Procrustes analysis.

The reconstruction of the original profile in Figure 4.12 shows that combining the plane reconstructions made using the  $yz$  and  $xy$  curvature functions calculated by perpendicular planes produces an accurate representation of the original profile with only minimal evidence of problems in accuracy at the base of the nose.

This perpendicular plane method produces curvature functions which appear to give an accurate reconstruction of a space curve. The major advantage of this method is that both  $yz$  and  $xy$  curvature have very clear anatomical meaning which is useful for presenting results as well as further analysis.



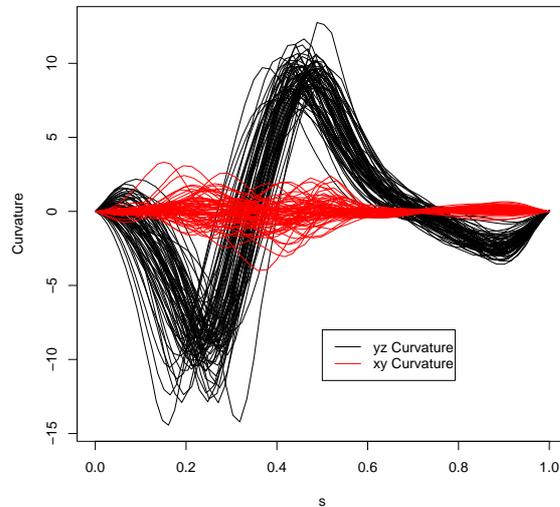
**Figure 4.12:** Plot of the reconstructed profile matched to the original profile from side on (left) and front on (right).

### 4.2.5 Investigating a collection of $yz$ and $xy$ curvature functions

In general, interest lies not in examining the three-dimensional bending of a single space curve but in comparing the three-dimensional bending of a collection of space curves. The fact that the amount of bending of the shape is explained by two scalar values at each point as opposed to the single scalar value in the planar case presents interesting issues for comparing a collection of curves. The example which will be considered is an extension of the midline profile example from Section 4.2 with data from 71 midline profiles of one year old control children with each three-dimensional profile defined by 57 points and rescaled to have arc length 1.

The perpendicular plane method (outlined in Section 4.2.4) will be used to calculate the  $yz$  and  $xy$  curvature for all 71 profile curves separately. Positive  $yz$  curvature indicates that the profile is bending towards the face whilst negative  $yz$  curvature indicates that the profile is bending away from the face. Positive  $xy$  curvature indicates that the profile is bending to the left when viewed from front on while negative  $xy$  curvature indicates that the profile is bending to the right. Figure 4.13 shows both curvatures functions, calculated using 10 degrees of freedom for smoothing, for all profiles.

It is clear from the curvature functions in Figure 4.13 that the range of  $yz$  curvature values is much greater than the range of  $xy$  curvature values. This

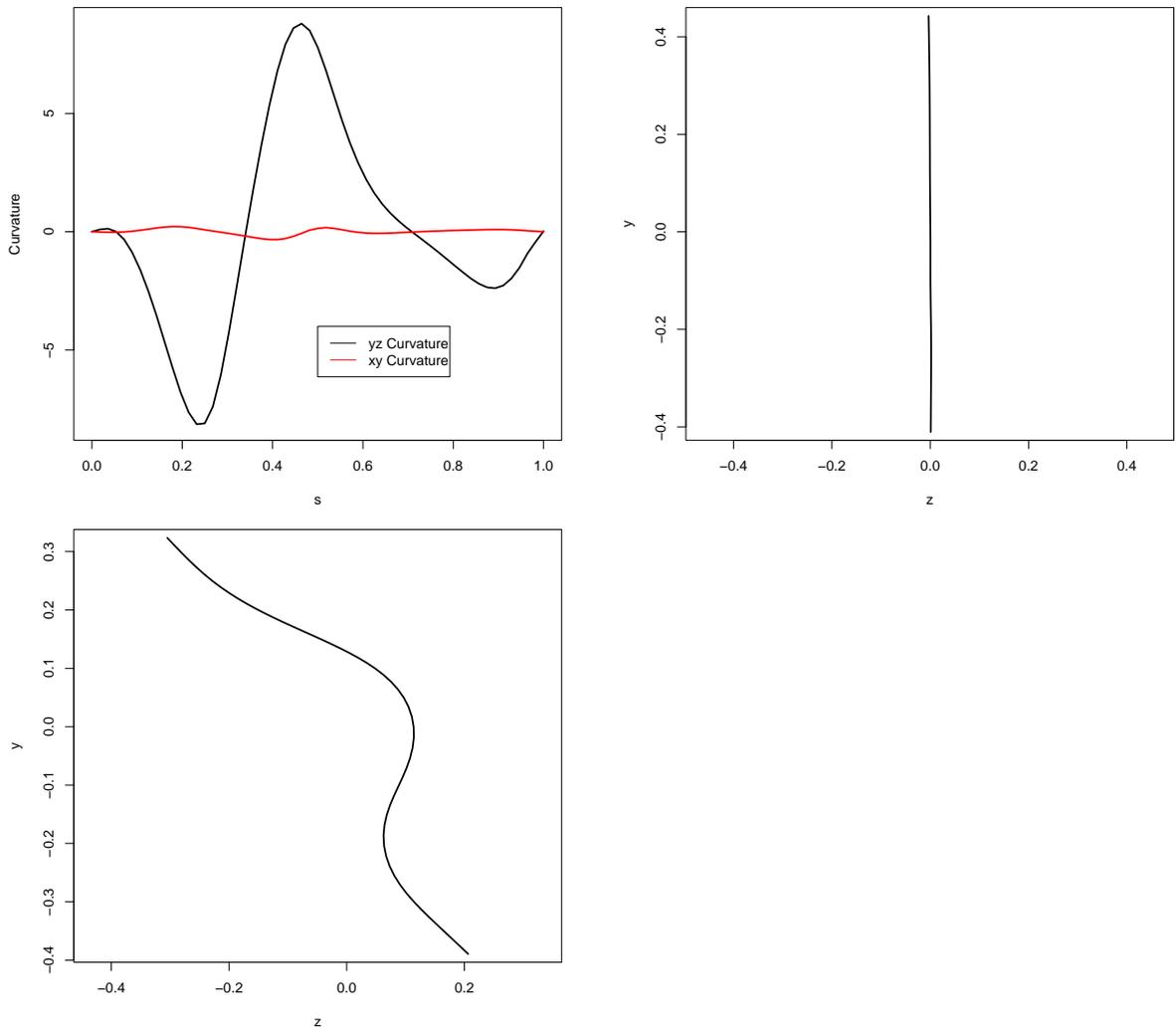


**Figure 4.13:** Plot of yz and xy curvature against arc length for all 71 profiles.

is expected since it seems reasonable that in general midline profiles bend considerably more into and out of the face than across the face. The yz curvature functions all appear to have one area of minimum curvature which represents the bending at the base of the nose and an area of maximum curvature which represents the bending at the tip of the nose. The xy curvature functions do not exhibit a general pattern, which is not unexpected since there would not appear to be any anatomical reason for a midline profile to bend to the left or the right at certain points. However any bending traversing the face appears to occur in the area which spans from just before the base of the nose to just after the tip of the nose.

The ‘average’ curvature functions for this group of children can be calculated by taking the piecewise average of all functions at regular points on the  $s$  axis and interpolating between the average curvature values. These average functions can then be used to reconstruct the ‘average’ midline profile for this group of children using the methods detailed in Section 4.2.4. However, the arc lengths of the original plane curves are required to produce the reconstruction. Although the arc lengths of the original space curves are rescaled to be of length 1, the plane curves produced from these space curves are not the same length. To obtain an average arc length for each of the plane curves, the mean of the arc length at each of the 57 points which define the profile is calculated across all the individual curves. This gives the arc lengths for each of the average plane curves which combined with the average curvature functions can be used to reconstruct the average feature. Figure 4.14 shows the average curvature functions for this

group of children as well as the reconstructed profile using these functions.



**Figure 4.14:** Plot of average yz and xy curvature functions for all profiles (top left) and the corresponding reconstructed average profile from side on (top right) and front on (bottom left).

### 4.3 Warping in the Space Curve Setting

Section 3.3 used the ideas of Gasser et al. (1990) to outline methods and issues associated with aligning a collection of curvature functions calculated from the midline profile defined as a plane curve. To extend these methods to align the data produced from the midline profile defined as a space curve some issues must be addressed. The difficulty here is that the bending of each profile is defined by both the yz and xy curvature functions. To ensure consistency between the yz

and xy curvature functions the same position warping function must be applied to both functions. Therefore, the process is similar to the plane curve process with the major difference being in defining characteristic points using two curvature functions. Section 4.3.1 will outline a method to define the characteristic points. Note that this technique could be easily altered to analyse curvature and torsion curves from standard methodology.

Since the magnitude of the yz and xy curvature values, are to a large degree, independent it seems reasonable to produce two amplitude adjustment functions (for both yz and xy curvature) for each midline profile. This can be achieved using the techniques of Section 3.3.1 without adjustment.

### 4.3.1 Position warping technique for yz and xy curvature curves

The major difficulty in the warping process using two curvature functions is how best to select characteristic points. In the plane curve setting potential characteristic points were calculated as positions where many of the individual curvature functions had extrema. A natural extension of this method would be to use the two curvature functions individually to produce potential characteristic points.

Suppose that the characteristic points are to be defined as positions where many individual functions have maximum turning points then the local maxima in both curvature functions can be calculated for each individual such that

$$\begin{aligned} \max_{\kappa i}^{yz} &= (c_{1i}, \dots, c_{p_i i}) \\ \max_{\kappa i}^{xy} &= (d_{1i}, \dots, d_{q_i i}) \end{aligned}$$

where  $\max_{\kappa i}^{yz}$  and  $\max_{\kappa i}^{xy}$  contain the  $p_i$  and  $q_i$  local maxima for the yz and xy curvature functions of subject  $i$  respectively. Now all local maxima for each curvature function can be combined across all subjects such that

$$\begin{aligned} \text{MAX}_{\kappa}^{yz} &= (c_{11}, \dots, c_{p_1 1}, c_{12}, \dots, c_{p_n n}) \\ \text{MAX}_{\kappa}^{xy} &= (d_{11}, \dots, d_{q_1 1}, d_{12}, \dots, d_{q_n n}) \end{aligned}$$

The position of the modes of the kernel probability density of  $\text{MAX}_{\kappa}^{yz}$  and  $\text{MAX}_{\kappa}^{xy}$  give the position of potential characteristic points calculated using the yz and xy curvature functions respectively. Suppose that the yz curvature functions

return  $d$  characteristic points with the average positions  $\bar{\gamma}_\kappa^{yz} = (\bar{\gamma}_{\kappa 1}^{yz}, \dots, \bar{\gamma}_{\kappa d}^{yz})$  and the xy curvature functions return  $e$  characteristic points with the average positions  $\bar{\gamma}_\kappa^{xy} = (\bar{\gamma}_{\kappa 1}^{xy}, \dots, \bar{\gamma}_{\kappa e}^{xy})$  then the combined set of characteristic points is the full ordered  $d + e$  length set  $\bar{\gamma} = (\bar{\gamma}_1, \dots, \bar{\gamma}_{d+e})$  where  $\bar{\gamma}_j < \bar{\gamma}_{j+1}$ .

The position warping process is now similar to the plane setting. The characteristic points for each individual,  $\gamma_i$ , which correspond to the average characteristic points can be found from the set of local maxima. The warping function  $g_i$  is then produced by setting  $g_i(\gamma_{ji}) = \bar{\gamma}_j$  such that  $f_i(g_i(\gamma_{ji})) = f_i(\gamma_{ji})$  where  $j = 1, \dots, d + e$  and interpolating to give a smooth warping function. If any of the characteristic points are not present for an individual this point is considered missing and the aligning is carried out on the remaining  $d + e - 1$  characteristic points. The fact that characteristic points calculated using both curvature functions produce this single warping function ensures consistency between the aligned functions.

One potential problem with this method occurs if a characteristic point is estimated by both the yz and xy curvature functions with slight differences in the average position. In this situation it is worth considering which of the two curvature functions is the more natural estimator of the position of interest and defining the position of the characteristic point using this curvature function. Alternatively simply taking an average of the position estimated by each curvature function will be sufficiently accurate.

A technique which would avoid this problem and has other appealing properties is to combine yz and xy curvature to give a global measure of bending at each point. Korn and Korn (1968) define total curvature of a space curve as

$$\Omega = \sqrt{\kappa_{yz}^2 + \kappa_{xy}^2} \quad (4.20)$$

If the amount of bending the curve experiences at each point is defined by the total curvature then warping could be carried out on one function as opposed to two and the warping process would be the same as the plane curve case in Section 3.3. Further analysis could then be carried out on the warped total curvature functions or the warped  $s$  axis could be applied to the yz and xy curvature functions and further analysis carried out on these two warped functions. A potential drawback of this method is that if the local maxima are evenly spread some peaks which define characteristic points may be missed unless the bandwidth for calculating the kernel density is reduced.

Clearly the position warping can be calculated using yz curvature functions,

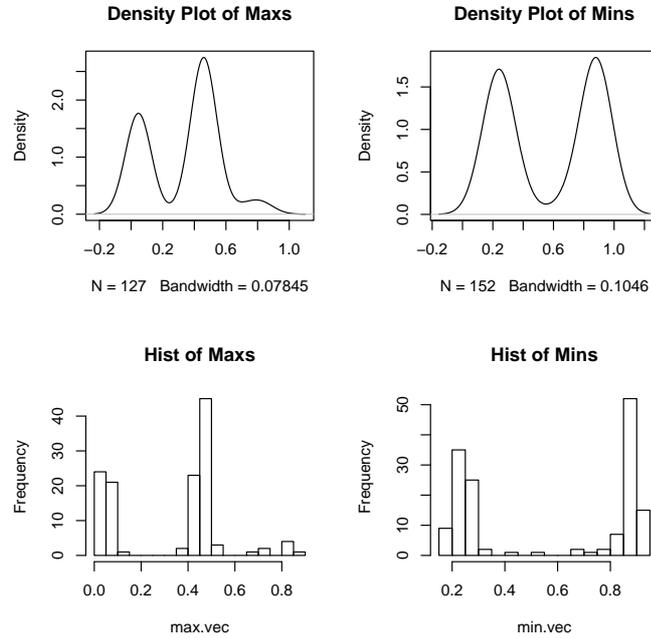
xy curvature functions, a combination of these or total curvature functions. The choice of which method to use is solely dependent on which is most appropriate for the data available.

### 4.3.2 Warping of yz and xy curvature functions: Midline profile example

Figure 4.13 in Section 4.2.5 shows the yz and xy curvature functions of 71 midline profiles belonging to one year old control children. The two major turning points of the yz curvature functions at around  $s = 0.2$  and  $s = 0.4$  indicate the base of the nose and the tip of the nose. These would be the primary points of interest for aligning the curvature functions. Thinking anatomically it is clear that both these features and in fact any features of interest on the midline profile will be found using yz curvature rather than xy curvature. Therefore the warping functions in this case will be calculated using just the yz curvature functions and then applied to both functions. The process is exactly equivalent to the plane curve warping.

Potential characteristic points will be estimated as points where the majority of functions have either a maximum or minimum turning point. Figure 4.15 shows the kernel probability density plots and histograms for the occurrence of both maximum and minimum turning points in the yz curvature functions.

As for the similar example on plane curves the kernel probability density plots in Figure 4.15 show two areas, top of the upper lip and tip of the nose, where the majority of curvature functions have maximum turning points and two areas, the base and bridge of the nose, where the majority of curvature functions have minimum turning points. This means that there are four potential characteristic points to use in the warping process. However when looking at the curvature functions (Figure 4.13) it is clear that, due to the magnitude of curvature and for anatomical reasons, the base of the nose and the tip of the nose are the major points of interest. Due to this it seems reasonable to carry out the warping process by aligning the functions to the average position of these features. The average position on the  $s$  axis of these characteristic points, calculated by the corresponding mode of the kernel probability density, is 0.242 for the base of the nose and 0.459 for the tip of the nose.



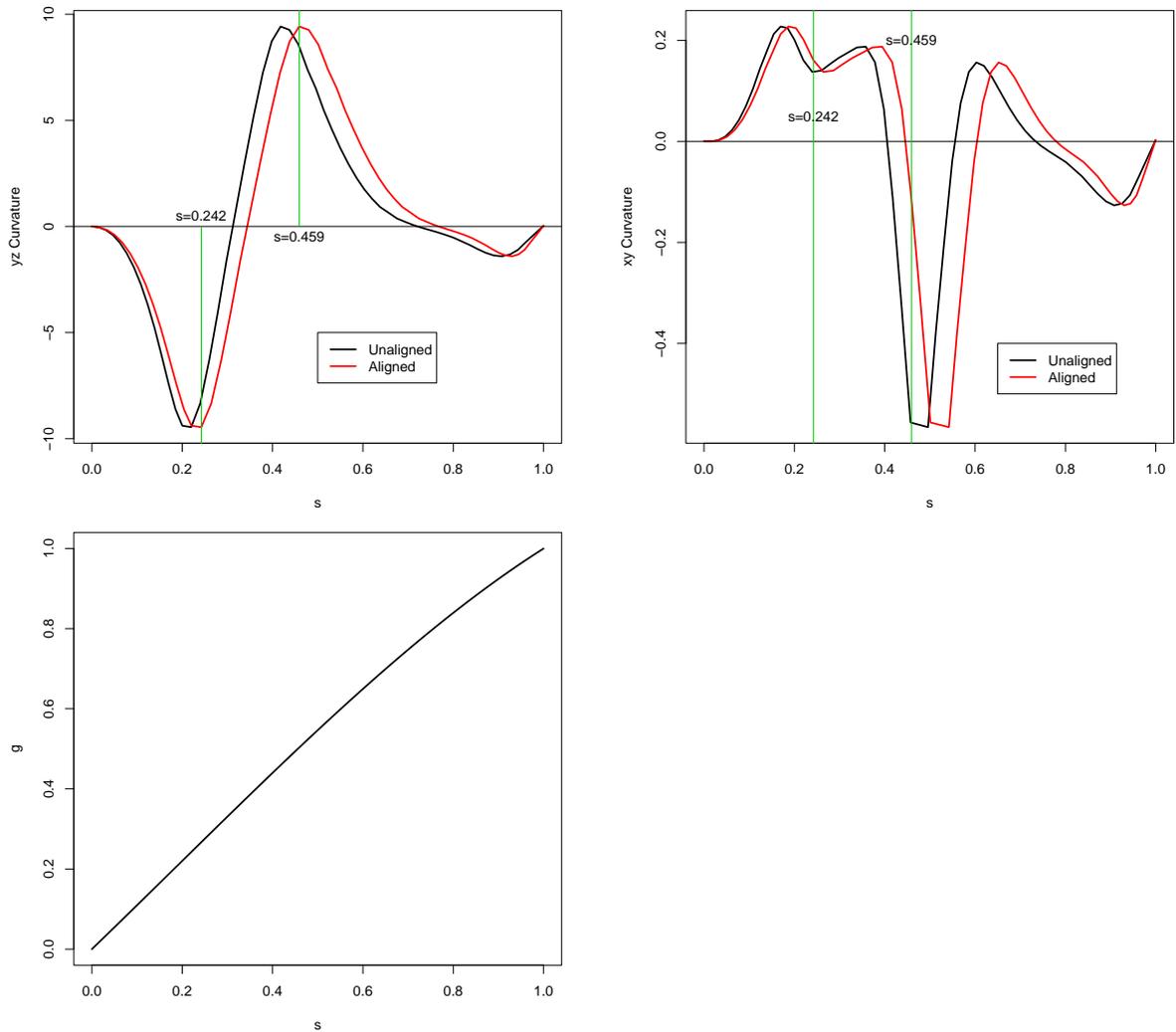
**Figure 4.15:** Kernel probability density plots and histograms for the occurrence of both maximum and minimum turning points of the curvature curves.

The warping functions can be produced by setting

$$\begin{aligned}
 g_i(0) &= 0 \\
 g_i(\gamma_1) &= \bar{\gamma}_1 \\
 g_i(\gamma_2) &= \bar{\gamma}_2 \\
 g_i(1) &= 1
 \end{aligned}$$

and using cubic spline interpolation between the four points to give a smooth warping function. The warping function is applied to both the yz and xy curvature functions so that the position of each point on the  $s$  axis is the same for both curvature functions to allow them to be directly comparable. The effect of applying the warping function to the yz and xy curvature functions from the example profile and the warping function required to align the functions are shown in Figure 4.16.

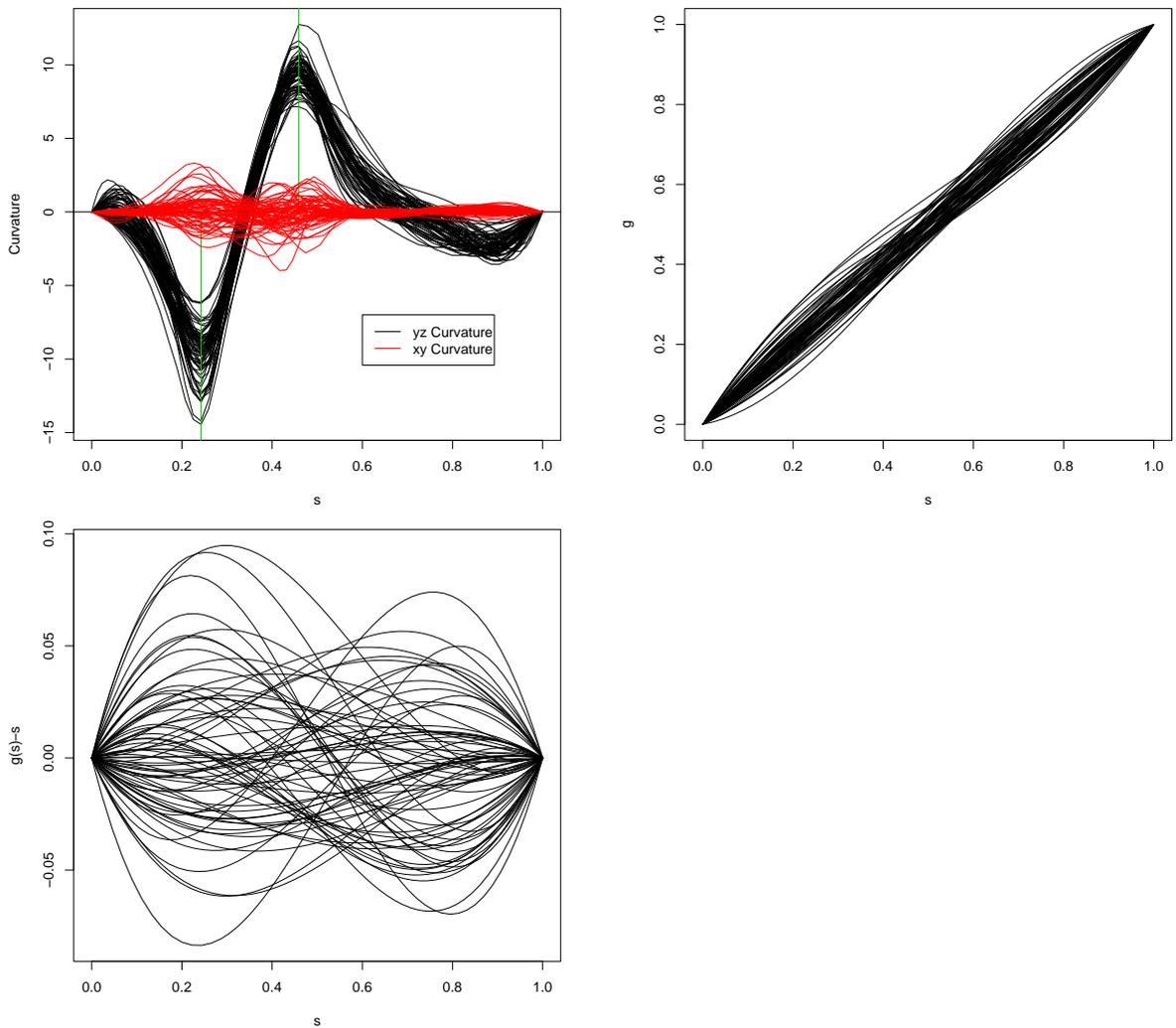
Figure 4.16 shows that the effect of warping on the yz curvature function is that the maximum and minimum points of yz curvature (i.e. the tip and base of the nose) have been aligned with the average position. None of the turning



**Figure 4.16:** Plot of the actual and warped yz curvature functions (top left) and actual and warped xy curvature functions (top right) for the example one year old control children midline profile. The warping function (bottom left).

points on the xy curvature function have been aligned to the characteristic points. This implies that in this example the areas of high xy curvature were not at the base of the nose or the tip of the nose. To consider the effect of warping a plot of the aligned curvature functions and the corresponding warping functions, and warping function minus arc length, are shown in Figure 4.17.

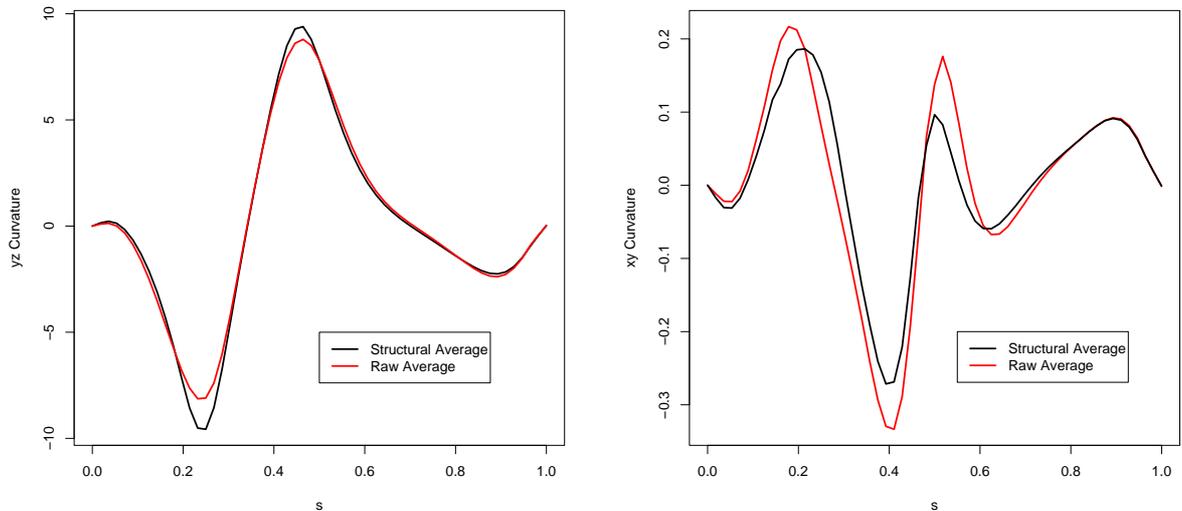
Considering the aligned functions in Figure 4.17 it seems that the variation in the magnitude of curvature at the base of the nose is larger than the variation in the magnitude of curvature at the tip of the nose. Looking at the aligned xy curvature functions it appears that in general the profiles have highest magnitude of xy curvature around the base of the nose and the tip of the nose. It seems



**Figure 4.17:** Plot of the warped yz and xy curvature functions (top left). The warping functions (top right), to align these functions, and the warping function minus arc length (bottom left).

that the profiles only appear to bend to the left or the right around the area from the base of the nose to the tip of the nose. Little can be said about the warping functions except that the amount of warping required by different subjects is, as expected, variable.

The structural average of the yz and xy curvature functions is calculated as the piecewise average of these aligned functions and compared to the raw average in Figure 4.18.

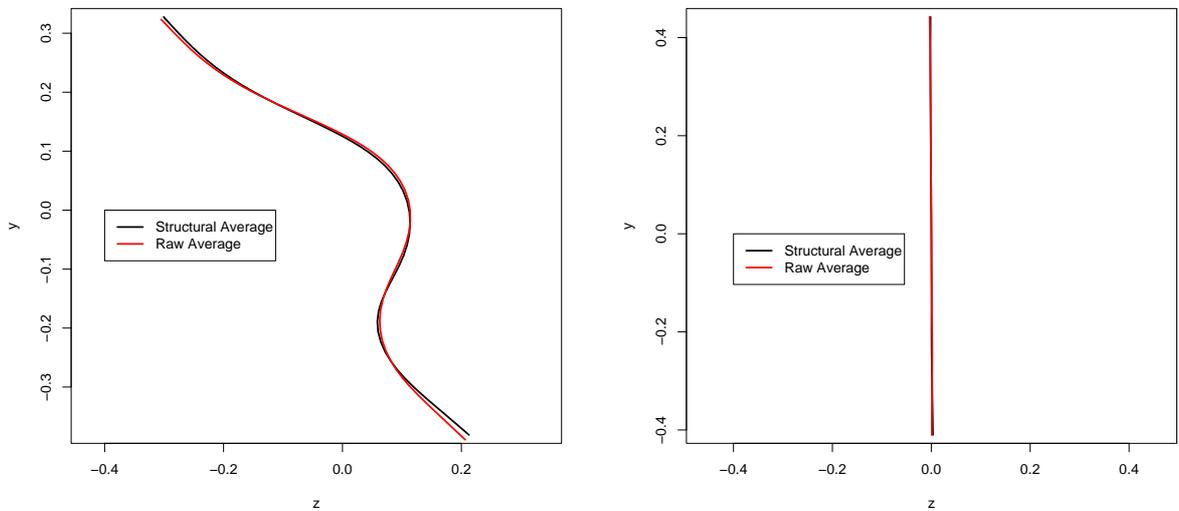


**Figure 4.18:** Plot of raw and structural average  $yz$  curvature against arc length (left) and raw and structural average  $xy$  curvature against arc length (right).

The comparison of the raw and structural average for the  $yz$  curvature functions, shown in Figure 4.18 shows that the magnitude of the curvature at the characteristic points was greater for the structural average than the raw average. The structural average gives a better estimate of the average  $yz$  curvature at the characteristic points across all the profiles. The comparison of the raw and structural average for the  $xy$  curvature functions is interesting. The structural average follows a similar pattern to the raw average; however, the raw average has a larger magnitude of  $xy$  curvature. Presumably this is due to the fact that at the characteristic points the  $xy$  curvature functions had either a maximum or minimum of  $xy$  curvature. Therefore when these maxima and minima were aligned the positive and negative curvature values cancelled each other out in the averaging.

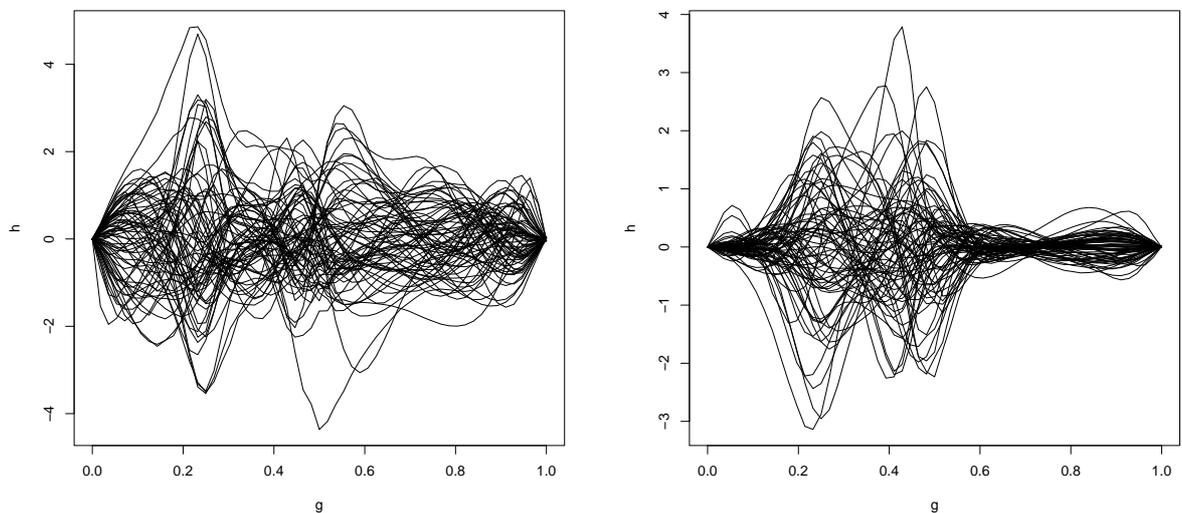
The space curve representation of the midline profile is reconstructed using the structural average curvature functions and Figure 4.19 shows the comparison between the average profiles constructed using the raw and structural averages.

The comparison between the reconstructed profiles using both raw and structural averaging, illustrated in Figure 4.19, shows very little difference across the face. There does appear to be a slight difference in the reconstructed profiles into and out of the face. It appears that the structural average profile shows more bending at the characteristic points (particularly the base of the nose) than the raw average profile. This is due to the difference in magnitude of curvature at the characteristic points between the raw and structural averages.



**Figure 4.19:** Reconstructed average profile using the raw and structural average from side on (left) and front on (right).

It is now interesting to investigate how much amplitude adjustment is required to exactly produce the structural average  $yz$  and  $xy$  curvature functions from the aligned individual functions. This is done for  $yz$  and  $xy$  curvature individually with the amplitude adjustment functions shown in Figure 4.20.



**Figure 4.20:** Amplitude adjustment functions to produce the structural average  $yz$  curvature function from individual  $yz$  curvature functions of the control midline profiles (left). Amplitude adjustment functions to produce the structural average  $xy$  curvature function from individual  $xy$  curvature functions of the control midline profiles (right).

Figure 4.20 shows a slight difference in the amplitude adjustment functions

between yz and xy curvature. Whilst there is an indication that slightly more amplitude adjustment is required around the base and tip of the nose than the rest of the nose for yz curvature, this phenomenon is clearer for xy curvature. In fact, for xy curvature, there is almost no amplitude adjustment required away from the area between the base and tip of the nose. This is due to the fact there is almost no xy curvature away from the base and tip of the nose.

## 4.4 Concluding Remarks on Space Curves

This chapter has shown that describing a space curve by the amount of bending the curve experiences requires two functions. In standard methodology these two functions are a function of curvature and a function of torsion where curvature describes the amount of bending at each point within the osculating plane whilst torsion describes the amount the curve bends out of the osculating plane. Three potential methods to calculate curvature and torsion were described; however, the derivative method ran into problems due to the difficulties in calculating third derivatives. The optimisation method successfully calculated curvature and torsion; however, the computing time required was large. Another major drawback to both the methods, as well as the Frenet method, was the difficulty in anatomically explaining the curvature and torsion values calculated. To attempt to remedy this problem a perpendicular plane method was introduced where the space curve was projected onto two perpendicular planes to describe the movement of the curve. In the midline profile example these two planes were a plane which cut through the face vertically and a plane which traversed the face. In general if anatomical explanation is required, or at least is of interest, from the functions of curvature then this perpendicular plane method would appear preferable.

Also described in this chapter were issues of warping a space curve. Clearly since the bending of the curve is explained by two functions, as opposed to one in the plane curve setting, this introduced some issues. However, the techniques used were a direct extension of the techniques used for plane curves. Also introduced was the idea of describing a space curve using the idea of total curvature. This is simply a measure of the complete bending in all directions experienced by a space curve and is given by the square root of the sum of squares of the two bending values (either yz and xy curvature or curvature and torsion depending on the method).

## Chapter 5

# Applications to the Cleft Lip Problem

The work in Chapters 3 and 4 was concerned with calculating the amount of bending experienced by the midline profiles of control children in two and three dimensions respectively. Data are available on 9 children with a cleft lip and 13 children with a cleft lip and palate as well as the 71 control children with the major interest in comparison between the three groups. The children are one year old and from the Glasgow area. There are data on other anatomically important curves as well as the midline profile. The curves which, along with the midline profile, are marked for potential analysis are the nasal base, the nasal rim, the upper lip and the nasal bridge. Figure 4.8 showed these curves from a control child projected onto three perpendicular planes.

The major interest of this chapter is in investigating what differences are apparent between control children and children with a cleft lip and/or palate. Techniques are required to firstly determine whether there appear to be differences between control and cleft children and secondly to describe these differences. A number of techniques will be described and then illustrated by a case study on the upper lip.

### 5.1 Techniques for Highlighting Differences Between Control and Cleft Children

The techniques used to highlight and investigate differences in shape between the facial curves of control and cleft children will concentrate on working with the

curvature functions of the three-dimensional features. Since the perpendicular plane method (Section 4.2.4) calculates two curvature functions with anatomically interpretable meaning this method will be used. The remainder of this section will outline the techniques which will be used to analyse the curvature functions to give comparisons between the control children and the children with a cleft lip and/or palate.

### 5.1.1 Curvature functions and average reconstructed features

The simplest way to compare the control and cleft children is to look for visually obvious differences between the curvature functions for each group. If there are clear differences this may be seen by plotting the curvature functions for each subject; however, it may be more informative to examine the average curvature functions for each group. The average curvature function can either be calculated as the piecewise average of the raw curvature functions or alternatively, to give a better indication of average curvature at anatomically or geometrically important points, as the average of the aligned curves. Aligning each group to the average position of the characteristic points for that group only, as opposed to the global average, provides structural averages which give a good indication of differences in the average position on the  $s$  axis and average curvature between the groups.

Section 3.3.2 illustrated how the curvature functions in the perpendicular plane method can be used to accurately reconstruct the original space curve. This can be done for the average curvature functions to give an average reconstructed feature for each group. Section 4.2.5 described how the average arc lengths are sufficient along with the curvature functions to produce the reconstruction. Reconstructions for each of the groups allows for a simple visual comparison between the shape of the average feature.

### 5.1.2 Position and curvature of characteristic points

Section 3.3.1 outlined the warping technique used in this thesis. The technique involves selecting the position (in terms of arc length) of characteristic points using turning points in the curvature functions. It is of interest to investigate any differences between the groups in terms of the position of the characteristic point and the magnitude of curvature at these characteristic point. This can be done informally using plots or more formally using multiple comparisons.

### 5.1.3 Warping to the average feature

Section 3.3 outlined a method to produce the average curvature function from an individual function by position warping followed by amplitude adjustment. It is of interest to investigate whether the amount of warping (both position and amplitude) required to warp to the average function is the same for all groups. It seems reasonable to use the structural average of the control functions as the average reference function. This is because the aim of surgery on cleft children is to reconstruct the face so that the shape is similar to that of an ‘average’ control child. To investigate whether there is a difference between the groups in terms of the amount of warping required it is necessary to quantify the warping applied to each function.

When carrying out the warping procedure if a function does not need to be shifted to align its characteristic points to the average position then the warping function will lie on the line of equality. The greater the departure from the line of equality the more warping has been required. It is of interest then to quantify this departure from the line of equality.

The method for doing this is similar to that for calculating the sum of squares in a regression context. Suppose there is a warping function  $g(s)$  where  $g(0) = 0$ ,  $g(1) = 1$  and  $s \in (0, 1)$ . The sum of the squared distance between the warping function and the straight line ( $s = g$  from  $s = 0$  to  $s = 1$ ) gives a measure of how close to linear the warping function is. This can be calculated as

$$D_p = \sum_{j=1}^n (g(s_j) - s_j)^2 \quad j = 1, \dots, n \quad (5.1)$$

where the  $s_j$ 's are  $n$  regular points along the  $s$  axis such that  $s_j - s_{j-1} = \frac{1}{n}$ .

A similar method can be used to determine the amount of amplitude adjustment required to produce the average control curvature function from the aligned individual functions. If the amplitude of the individual function is the same as the average function then the amplitude adjustment function  $h(g(s))$  will be  $h = 0$  at all  $g(s)$ . Therefore to quantify the amount of amplitude adjustment required the sum of squared distance between the amplitude adjustment function and the line  $h = 0$  is calculated as

$$D_a = \sum_{j=1}^n (h(g(s_j)))^2 \quad j = 1, \dots, n \quad (5.2)$$

where the  $g(s_j)$ 's are  $n$  regular points along the  $g$  axis.

The amount of warping (both position and amplitude) required to produce the average control curvature function from the original function can be compared between the groups to investigate whether the curvature of the faces of cleft children are on average further away from the average face of a control child than the curvature of the faces of control children. This can be done informally using plots or more formally using multiple comparisons.

### 5.1.4 Principal components analysis

When a set of functions,  $f_i(s) = (x_i(s), y_i(s))$  where  $s \in (0, 1)$ , are available one of the initial interests is how the functions vary and in general where the areas of greatest variation are. This can be investigated by carrying out a principal components analysis on the functions. One method for this is to reduce the problem of carrying out principal components analysis on functional data to a standard multivariate problem which can be analysed as shown by Jolliffe (1986). Suppose there are  $p$  functions, then this can be done by calculating the curvature at  $n$  regular points on the  $s$  axis so that each function is described as

$$f_i(s) = (\{x_i(s_1), y_i(s_1)\}, \dots, \{x_i(s_n), y_i(s_n)\})$$

where  $s_j - s_{j-1} = \frac{1}{n}$ . The multivariate case can then be set up where the variables are the  $n$  regular points on the  $s$  axis and the items are the  $i$  individual curves. Note that  $p$  must be greater than  $n$ . This discretisation method is used here; however, it would be equally appropriate to use the functional principal components analysis method described in Section 2.3.2.

To informatively display the results of the principal components analysis a plot of the effect on the mean curve ( $\bar{f}(s)$  say) of adding and subtracting a multiple of each of the first three principal components will be produced. For example, for the first principal component, if there is an eigenvector  $\mathbf{e}_1$  and standard deviation  $\sqrt{\lambda_1}$  then to display the results a plot of the mean curve and also the curves given by  $\bar{f}(s) \pm 3\sqrt{\lambda_1}\mathbf{e}_1$  will be produced.

This process can be carried out on the curvature functions from individual groups or from all groups combined to give a visual representation of where most of the variation between the functions lies. If the principal components analysis is carried out on all groups combined there is specific interest in the component scores. If there appears to be clustering of groups in terms of their scores for

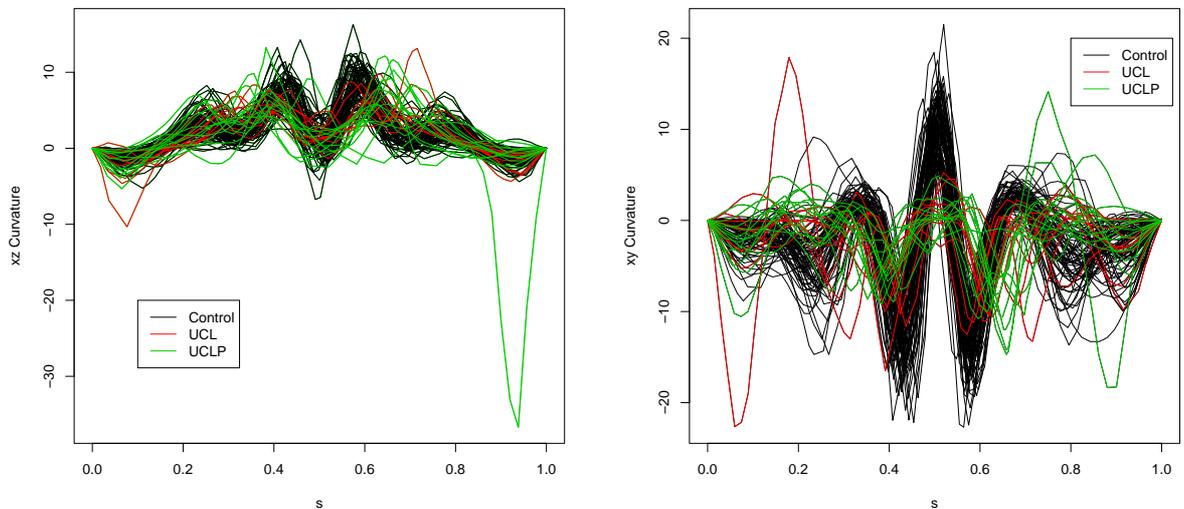
a particular component it may be that a certain group carries a large amount of the weight in the variation explained by that principal component. Finding principal components for which either of the cleft groups have particularly large scores may show areas on the feature where that cleft group have a large variation in curvature values but the control group do not show a large variation. Plots of the principal component scores with different symbols for the groups may give an indication of components which are dominated by a certain group.

## 5.2 Case Study: Upper Lip

Section 5.1 has outlined various techniques available to give a comparison between control children and children with a cleft lip and/or palate in terms of the curvature of certain features. This section will use these techniques to analyse the upper lip for differences in bending between the control and cleft children. The analysis of the upper lip curvature functions has been chosen over the other facial curves as this produced the most interesting results. Note that from here children with a cleft lip only will be referred to as UCL children whilst children with a cleft lip and palate will be referred to as UCLP children.

### 5.2.1 Curvature functions and average upper lip reconstructions

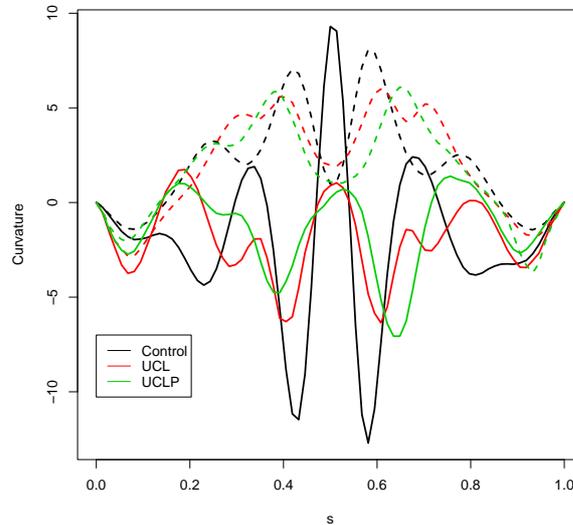
The curvature of the three-dimensional upper lip can be calculated both into and out of the face, by projecting onto the  $xz$  plane, and up and down the face, by projecting onto the  $xy$  plane, as the upper lip travels from the left of the face to the right of the face. Note that positive  $xz$  curvature indicates that the upper lip is bending towards the face whilst negative curvature indicates that the upper lip is bending away from the face. Also note that positive  $xy$  curvature indicates that the upper lip is bending towards the top of the face whilst negative  $xz$  curvature indicates that the upper lip is bending towards the bottom of the face. The curvature functions for all cases are shown in Figure 5.1.



**Figure 5.1:** Plot of  $xz$  (left) and  $xy$  (right) curvature functions for all upper lips.

The curvature functions in Figure 5.1 appear much rougher than the curvature functions for the midline profiles. This is because 20 degrees of freedom have been used in the smoothing. It was necessary to use large degrees of freedom to detect the sharp turn in curvature near the middle of the upper lip. Looking at the  $xz$  curvature functions it is clear that there is an area between  $s = 0.4$  and  $s = 0.6$  where the upper lips bend towards the face, then either flatten off (zero curvature) or bend away from the face (negative curvature) and then bend towards the face again. There is also an indication that, for the control functions in particular, there are two further areas around  $s = 0.2$  and  $s = 0.8$  where the curvature functions show a maximum turning point of curvature. This corresponds to bending towards the face between the ends and the middle of the upper lip. The  $xy$  curvature functions show a similar but more pronounced pattern. In the area between  $s = 0.4$  and  $s = 0.6$  almost all functions have an area where the upper lip bends towards the bottom of the face, followed by an area where the upper lip bends towards the top of the face (this is much more pronounced for the control cases) and an area where the curve bends towards the bottom of the face. There is also the indication of some minimum turning point in the curvature functions between the ends and the middle of the upper lip at  $s = 0.2$  and  $s = 0.8$ . The area between  $s = 0.4$  and  $s = 0.6$  is called the Cupid's bow which contains the landmarks  $cphR$ ,  $ls$  and  $cphL$  shown in Figure 1.1. The medical belief is that while control children in general experience three pronounced turning points in the Cupid's bow cleft children even after surgery tend not to experience the

upwards turn in the middle of the Cupid's bow at landmark  $ls$ . The piecewise average curvature functions for each group are shown in Figure 5.2.



**Figure 5.2:** Raw average xz curvature (dashed lines) and xy curvature (solid lines) functions for the upper lips of each group.

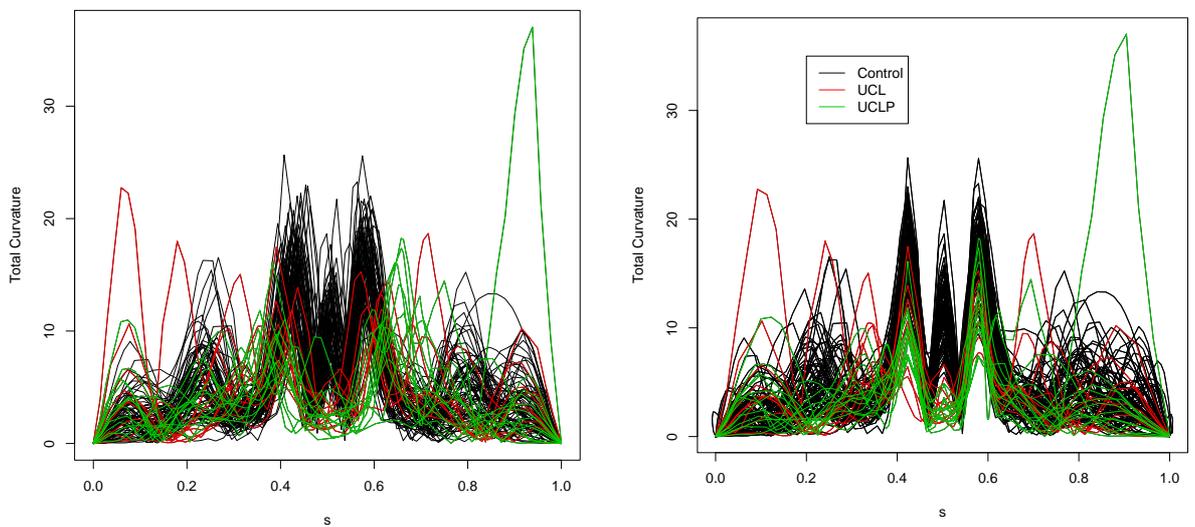
Considering the average xz curvature functions from Figure 5.2 there is further evidence that for all groups there are two areas of high magnitude positive curvature between  $s = 0.4$  and  $s = 0.6$  while on average the curvature approaches zero around  $s = 0.5$  for all groups but does not become negative. This suggests that on average the upper lip does not turn away from the face at landmark  $ls$ . Since the average curvature values at landmarks  $cphL$  and  $cphR$  are larger for control children this suggests that on average control children have a more pronounced turn towards the face at the ends of the Cupid's bow than UCL and UCLP children. It also appears that UCLP children on average have a wider Cupid's bow than control and UCL children. Looking at the xy curvature functions the Cupid's bow phenomenon is again shown in the controls by two areas of low curvature around  $s = 0.4$  and  $s = 0.6$  and an area of high curvature around  $s = 0.5$ . Interestingly the average UCL and UCLP functions do not have an area of high positive curvature at  $s = 0.5$  but have zero curvature at this point indicating that as opposed to turning up the face at landmark  $ls$  cleft children tend to have a flat area. Both curvature functions, in particular for controls, show potential points of interest at around  $s = 0.2$  and  $s = 0.8$  although discussion of these points will not be of primary interest and instead the Cupid's bow area will be the focus of much of the analysis.

It is again of interest to produce aligned curvature functions. In this example

turning points in both  $xy$  and  $xz$  curvature functions indicate the  $cphR$ ,  $ls$  and  $cphL$  landmarks therefore these three points will be taken as the characteristic points for aligning. To allow both curvature functions to be used in the calculation of the position of the characteristic points the total curvature ( $\kappa_T$ ) will be defined as

$$\kappa_T = \sqrt{\kappa_{xz}^2 + \kappa_{xy}^2}$$

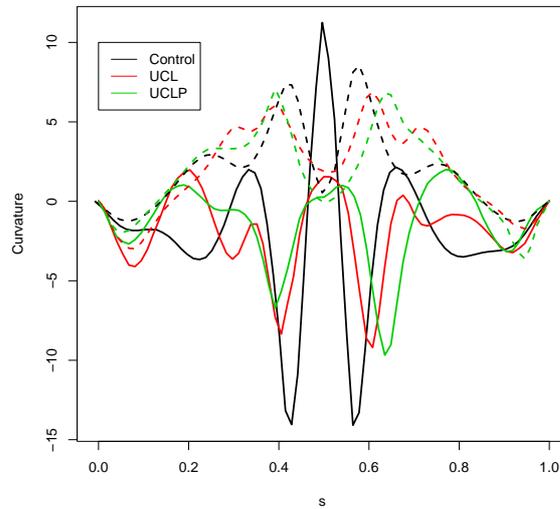
where  $\kappa_{xz}$  and  $\kappa_{xy}$  are the  $xz$  and  $xy$  curvature functions respectively. A plot of both the raw and aligned total curvature functions is shown in Figure 5.3.



**Figure 5.3:** Plot of unaligned (left) and aligned (right) total curvature against arc length for all upper lips.

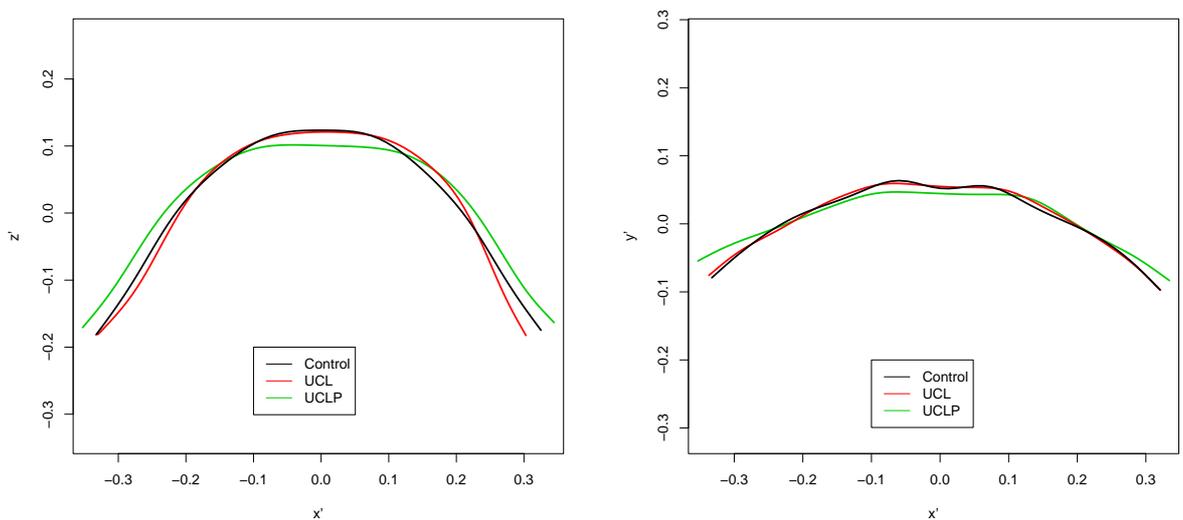
For many of the UCL and UCLP cases there is no indication of the upper lips experiencing a turning point of the curvature function at landmark  $ls$  while one UCL case has no clear  $cphL$  landmark. In these cases the characteristic point which cannot be found is regarded as missing and the aligning is carried out using the remaining characteristic points.

The warping is carried out for each group individually and the aligned  $s$  axis applied to both the  $xz$  and  $xy$  curvature functions and the structural averages calculated as the piecewise average of these aligned curvature functions. These structural average functions are shown in Figure 5.4.



**Figure 5.4:** Structural average xz (dashed lines) and xy (solid lines) curvature functions for the upper lips of each group.

The interpretation of the structural average functions is very similar to that of the raw average functions with the same differences between the groups evident. As expected the magnitude of curvature at the characteristic points is larger in the structural average. Using the method described in Section 5.1.1 the average upper lip for each group is calculated from these structural average functions and the reconstruction is plotted on the xz and xy plane in Figure 5.5.



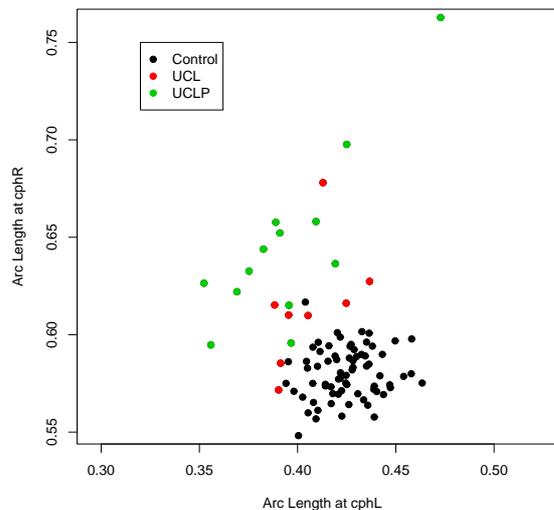
**Figure 5.5:** The reconstructed average upper lip from the structural averages of each group in the xz plane (left) and xy plane (right).

The reconstructed UCLP average upper lip shows clear differences to the other groups in both planes. The ends of the Cupid's bow are less rounded and

further apart than for the average UCL or control and in the xy plane there is no evidence of an upturn at the  $ls$  landmark. On the other hand the average UCL upper lip, while showing slightly less rounded ends of the Cupid's bow, is relatively similar to the average control upper lip. However, the average UCL in the xy plane gives little indication of any upturn at the  $ls$  landmark.

### 5.2.2 Investigating characteristic points

It appears that there are potential differences between all three groups in terms of the arc length and the amount of bending experienced at the characteristic points. Since few UCL or UCLP children had a distinguishable  $ls$  landmark it seems reasonable to only compare the arc length and curvature of the characteristic points at the start ( $cphL$ ) and end ( $cphR$ ) of the Cupid's bow. Note that since the  $cphL$  landmark was missing for one of the UCL cases this case is not included in this analysis. The arc length at the start of the Cupid's bow is plotted against the arc length at the end of the Cupid's bow for all children in Figure 5.6.



**Figure 5.6:** Arc length at the start (left) of the Cupid's bow plotted against arc length at the end (right) of the Cupid's bow.

It is interesting to note, from the plot of arc length at the characteristic points shown in Figure 5.6, that there is a relatively tight cluster of the control points whilst the points for the UCL and to a large extent the UCLP group are well spread. Therefore while the control group shows little variability in the position of the characteristic points there is a large amount of variability for the UCL and UCLP groups. It appears that the largest difference between the groups is that

the position of the *cphR* landmark occurs much further along the upper lip for the UCL and UCLP children than for control children. Any difference in the position of the *cphL* landmark is not so clear, although it does seem that the point occurs earlier for the UCL and UCLP children than the control children. The fact that the differences between the groups are clearer around the *cphR* landmark than the *cphL* landmark is presumably due to all clefts occurring on the right side of the face after reflection.

For formal testing the data for each group is approximately normally distributed; however, there are issues with an assumption of similar spread with the control values having a much smaller spread than the UCL or UCLP values. Therefore the nonparametric Kruskal-Wallis rank sum test is used to test for any evidence of a difference between the groups. The null hypothesis is that there is no evidence of a difference between the groups in terms of population median arc length. The p-values for both the characteristic points are much less than 0.001 indicating strong evidence at each characteristic point of some difference between the groups. Multiple two sample Mann Whitney tests with a Bonferroni correction are carried out to investigate between which groups these differences lie with the results shown in Table 5.1 (for the *cphL* landmark) and Table 5.2 (for the *cphR* landmark).

| Comparison     | Interval           | p-value |
|----------------|--------------------|---------|
| UCL - Control  | (-0.0364, -0.0028) | 0.023   |
| UCLP - Control | (-0.0502, -0.0153) | <0.001  |
| UCLP - UCL     | (-0.0400, 0.0180)  | 0.903   |

**Table 5.1:** Multiple comparisons of the arc length at the *cphL* landmark.

| Comparison     | Interval          | p-value |
|----------------|-------------------|---------|
| UCL - Control  | (0.0134, 0.0488)  | 0.003   |
| UCLP - Control | (0.0402, 0.0762)  | <0.001  |
| UCLP - UCL     | (-0.0143, 0.0723) | 0.191   |

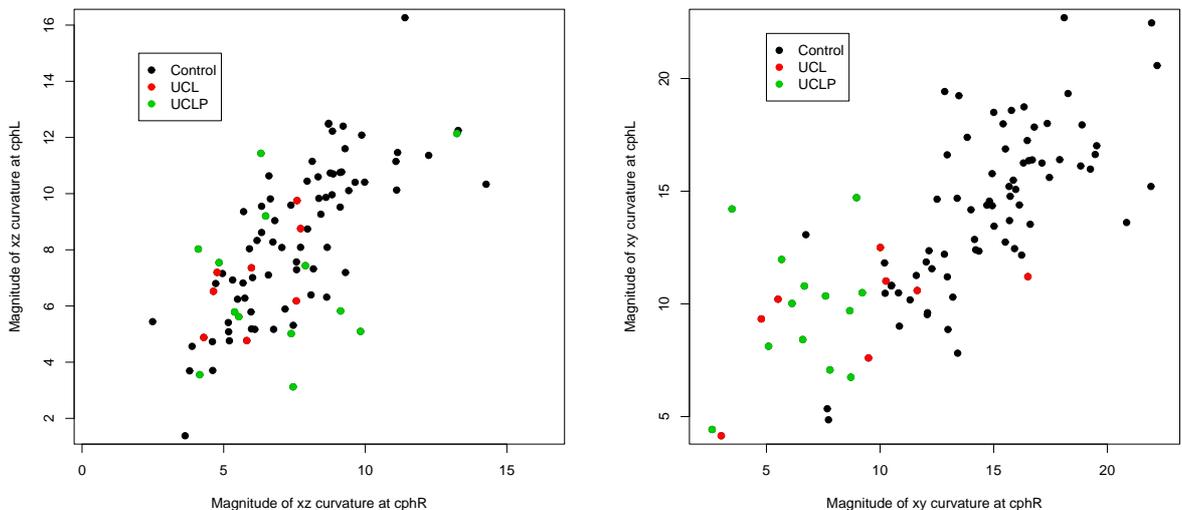
**Table 5.2:** Multiple comparisons of the arc length at the *cphR* landmark.

Table 5.1 shows that it is highly likely that in the wider population the start of the Cupid's bow will in general occur further along the upper lip for control

children than for UCL or UCLP children. There is no evidence of a difference in the population median arc length at the start of the Cupid's bow between the UCL and UCLP children.

Table 5.2 shows that it is highly likely that in the wider population the end of the Cupid's bow will in general occur earlier on the upper lip for control children than for UCL or UCLP children. There is no evidence of a difference in the population median arc length at the end of the Cupid's bow between the UCL and UCLP children.

To investigate the differences in the curvature at the characteristic points Figure 5.7 shows the plots of both xz and xy curvature at the start of the Cupid's bow against curvature at the end of the Cupid's bow. Note that the absolute value of curvature is taken as the interest is in the difference between the magnitude of curvature at each characteristic point. The only result of taking the absolute values is that the negative xy curvatures at the characteristic points become positive.



**Figure 5.7:** Curvature at the start of the Cupid's bow against curvature at the end of the Cupid's bow for xz (left) and xy (right) curvature.

From Figure 5.7 it is clear that there is much greater separation between the groups in terms of xy curvature as opposed to xz curvature. In general control children have larger xy curvature values at both the start and end of the Cupid's bow than UCL or UCLP children. This implies that control children in general have upper lips which bend more towards the bottom of the face at the start and end of the Cupid's bow. There does not appear to be much evidence of a difference between control children and UCL and UCLP children in terms of xz

curvature at the characteristic points.

For formal analysis the data for each group is approximately normally distributed; however, there are issues with an assumption of similar spread with the control values experiencing a much larger spread than the UCL or UCLP values. Therefore the nonparametric Kruskal-Wallis rank sum test is used to test for any evidence of a difference between the groups in the wider population. The p-values from the Kruskal-Wallis test can be found in Table 5.3.

| Curvature | Characteristic Point | p-value |
|-----------|----------------------|---------|
| xz        | <i>cphL</i>          | 0.095   |
| xz        | <i>cphR</i>          | 0.053   |
| xy        | <i>cphL</i>          | <0.001  |
| xy        | <i>cphR</i>          | <0.001  |

**Table 5.3:** Results of Kruskal-Wallis test for difference between the groups in terms of median population curvature at the characteristic points.

Table 5.3 shows that there is insufficient evidence of a difference between the groups in the wider population in terms of median xz curvature at either characteristic point. However, there is evidence of some difference between the groups in the wider population in terms of median xy curvature both at the start and the end of the Cupid's bow. To investigate between which groups these differences occur multiple two sample Mann Whitney tests with a bonferroni correction are carried out. The results are shown in Table 5.4.

| Characteristic Point | Comparison     | Interval          | p-value |
|----------------------|----------------|-------------------|---------|
| <i>cphL</i>          | UCL - Control  | (-10.001, -2.270) | 0.002   |
| <i>cphL</i>          | UCLP - Control | (-10.113, -6.144) | <0.001  |
| <i>cphL</i>          | UCLP - UCL     | (-6.542, 2.658)   | 0.554   |
| <i>cphR</i>          | UCL - Control  | (-7.67, -1.609)   | 0.002   |
| <i>cphR</i>          | UCLP - Control | (-7.195, -2.059)  | <0.001  |
| <i>cphR</i>          | UCLP - UCL     | (-3.089, 3.617)   | >0.999  |

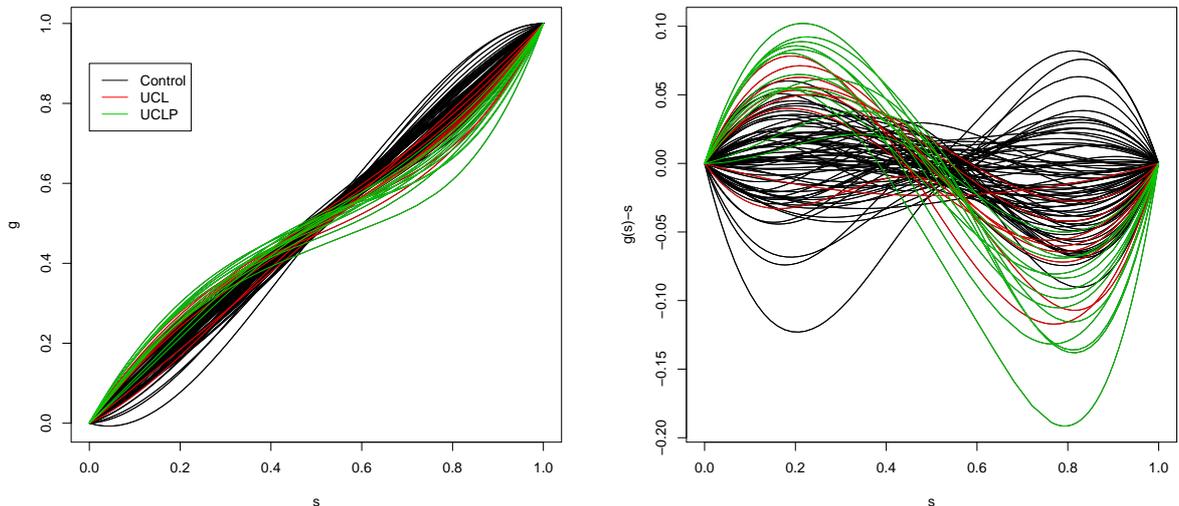
**Table 5.4:** Multiple Mann Whitney tests of the xy curvature at both characteristic points.

Table 5.4 shows that it is highly likely that the population median xy curvature for control children is greater than for UCL and UCLP children at both

characteristic points. There is no evidence of a difference between UCL and UCLP children at either characteristic point. This confirms the fact that control children in general have upper lips which bend more towards the bottom of the face at the start and end of the Cupid's bow.

### 5.2.3 Position warping and amplitude adjustment

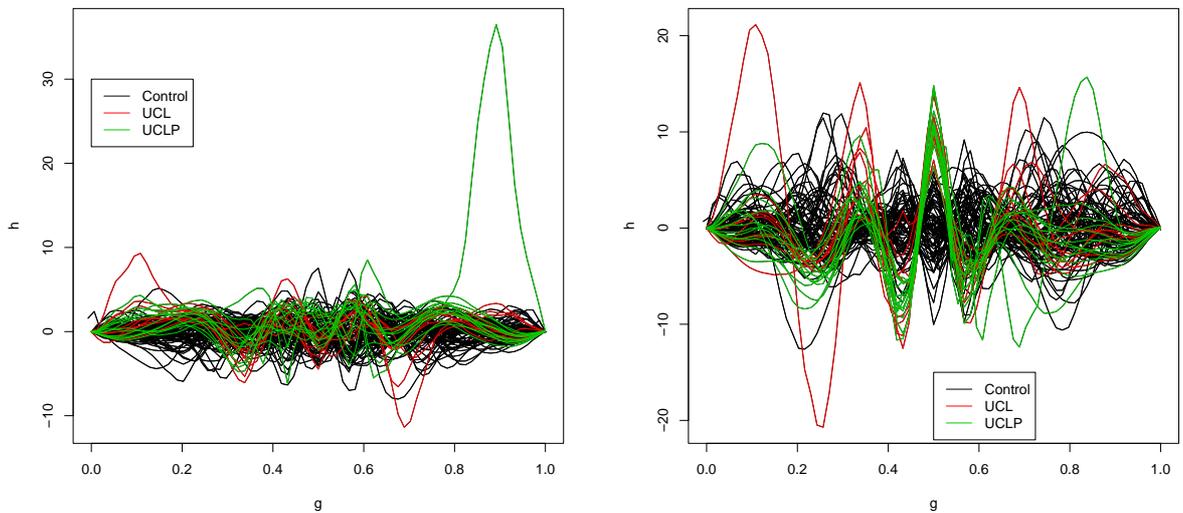
The amount of warping (both position and amplitude) required to produce the structural average of the control functions from the individual functions can give a useful indicator of inconsistencies remaining between the control children and the UCL and UCLP children even after surgery. Figure 5.8 shows the position warping functions for each individual curve.



**Figure 5.8:** Position warping functions (left), and position warping function minus arc length (right), to align to the position of the average characteristic points of the upper lip of control children.

The most notable aspect of the position warping functions, and position warping minus arc length functions, (Figure 5.8) is that the UCLP group appear to require considerably more position warping than either the UCL or control group. The fact that the UCLP functions cross the line of equality indicates that in general UCLP have an earlier start of the Cupid's bow and a later end of the Cupid's bow than the average control child. There is also perhaps an indication that UCL children require more position warping than control children although the distinction is not as clear as for the UCLP children.

Figure 5.9 shows the amplitude adjustment functions for both the  $xy$  and  $xz$  curvature functions for each individual.

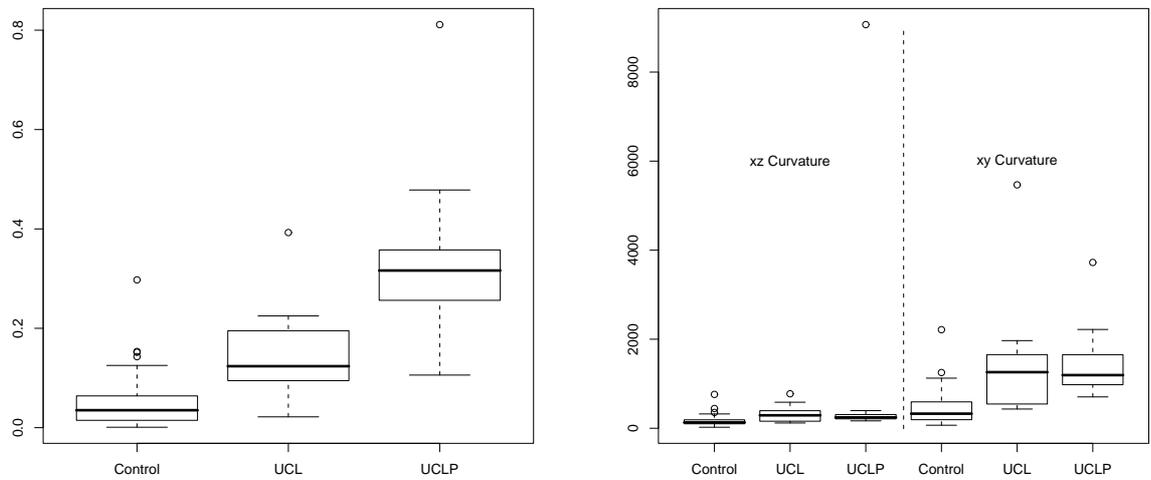


**Figure 5.9:** Amplitude adjustment functions to produce the structural average  $xz$  curvature function from individual  $xz$  curvature functions (left) and amplitude warping functions to produce the structural average  $xy$  curvature function from individual  $xy$  curvature functions (right).

The immediately noticeable aspect of the amplitude adjustment of the  $xz$  curvature functions is the extremely large amount of amplitude adjustment required by a UCLP case at the end of the upper lip. This is the amplitude adjustment required by a UCLP case which had severe disfigurements at the end of the upper lip. Away from this extreme case there is not a great deal of evidence to suggest much difference between the groups in terms of amplitude adjustment. Considering the amplitude adjustment of the  $xy$  curvature functions there is clearly one UCL case which requires a large amount of amplitude adjustment (both up and down the face) at the start of the upper lip. Furthermore, it appears that in general both UCL and, to a greater extent, UCLP children require more amplitude adjustment of the  $xy$  curvature functions than control children particularly around the area of the Cupid's bow.

The amount of both position warping and amplitude adjustment required to produce the structural average control function is plotted as a boxplot in Figure 5.10.

In terms of position warping it appears, from the boxplots in Figure 5.10, that the UCLP group in general require more position warping than the UCL group which in turn requires more position warping than the control group. One UCLP case requires an extremely large amount of position warping. Formal testing is required to investigate the evidence of a difference between the groups in terms of



**Figure 5.10:** Boxplot of the sum of squares difference between the position warping function and the line of equality for the upper lips for each group (left) and boxplot of the sum of squares difference between the amplitude adjustment function and the  $x = 0$  line for each group for both  $xz$  and  $xy$  curvature functions of upper lips (right).

the amount of position warping required. Since there appears to be a difference in the variation of the groups, and there is evidence that the control group is not normally distributed, the non-parametric Kruskal-Wallis test is used to test the null hypothesis that there is no evidence of a difference between the groups in terms of population median position warping. The p-value of the test is less than 0.001 suggesting strong evidence of a difference between the groups in terms of the amount of position warping required. To investigate between which groups differences lie Mann-Whitney intervals of the population median differences with Bonferroni correction for multiple comparisons are calculated with the results shown in Table 5.5.

| Comparison     | Interval         | p-value |
|----------------|------------------|---------|
| UCL - Control  | (0.0326, 0.1592) | 0.001   |
| UCLP - Control | (0.2045, 0.3214) | <0.001  |
| UCLP - UCL     | (0.0314, 0.2900) | 0.021   |

**Table 5.5:** Multiple comparisons of the position warping required to produce the structural average control curvature function for the upper lip.

Table 5.5 shows that there is evidence of a difference between the population

median position warping of each of the groups. It is highly likely that the population median of the UCLP group will be larger than the population median of the UCL which in turn will be larger than the population median of the control group. This suggests that in general, to warp the characteristic points to the average position of control children, UCLP children will require the most position warping whilst UCL children will also require more position warping than control children.

Figure 5.10 shows that the amplitude adjustment required for the xy curvature functions was in general larger than the amplitude adjustment required for the xz curvature functions. This suggests that in general there was more variation in the curvature up and down the face than into the face. The point with very large amplitude adjustment of xz curvature is due to the UCLP case which had severe disfigurement at the end of the upper lip. It seems that in general both the UCL and UCLP group require more amplitude adjustment of the xy curvature functions on average than the control group while there does not appear to be a great deal of difference between the groups in terms of the amount of amplitude adjustment required for the xz curvature functions. Again the Kruskal-Wallis test is used to test the null hypothesis that there is no evidence of a difference between the groups in terms of the population median amplitude adjustment required. For the amplitude adjustment of both the xy and xz curvature functions the p-value from the Kruskal-Wallis test is less than 0.001 suggesting clear evidence of a difference between the groups. To investigate where these differences lie Mann-Whitney intervals of the population median differences with Bonferroni correction for multiple comparisons are calculated for the xy and xz curvature functions separately with the results shown in Table 5.6.

| Curvature | Comparison     | Interval          | p-value |
|-----------|----------------|-------------------|---------|
| xz        | UCL - Control  | (19.68, 289.40)   | 0.015   |
| xz        | UCLP - Control | (61.23, 182.30)   | <0.001  |
| xz        | UCLP - UCL     | (-228.49, 136.46) | >0.99   |
| xy        | UCL - Control  | (238.59, 1331.71) | 0.001   |
| xy        | UCLP - Control | (552.12, 1128.01) | <0.001  |
| xy        | UCLP - UCL     | (-775.23, 838.61) | >0.99   |

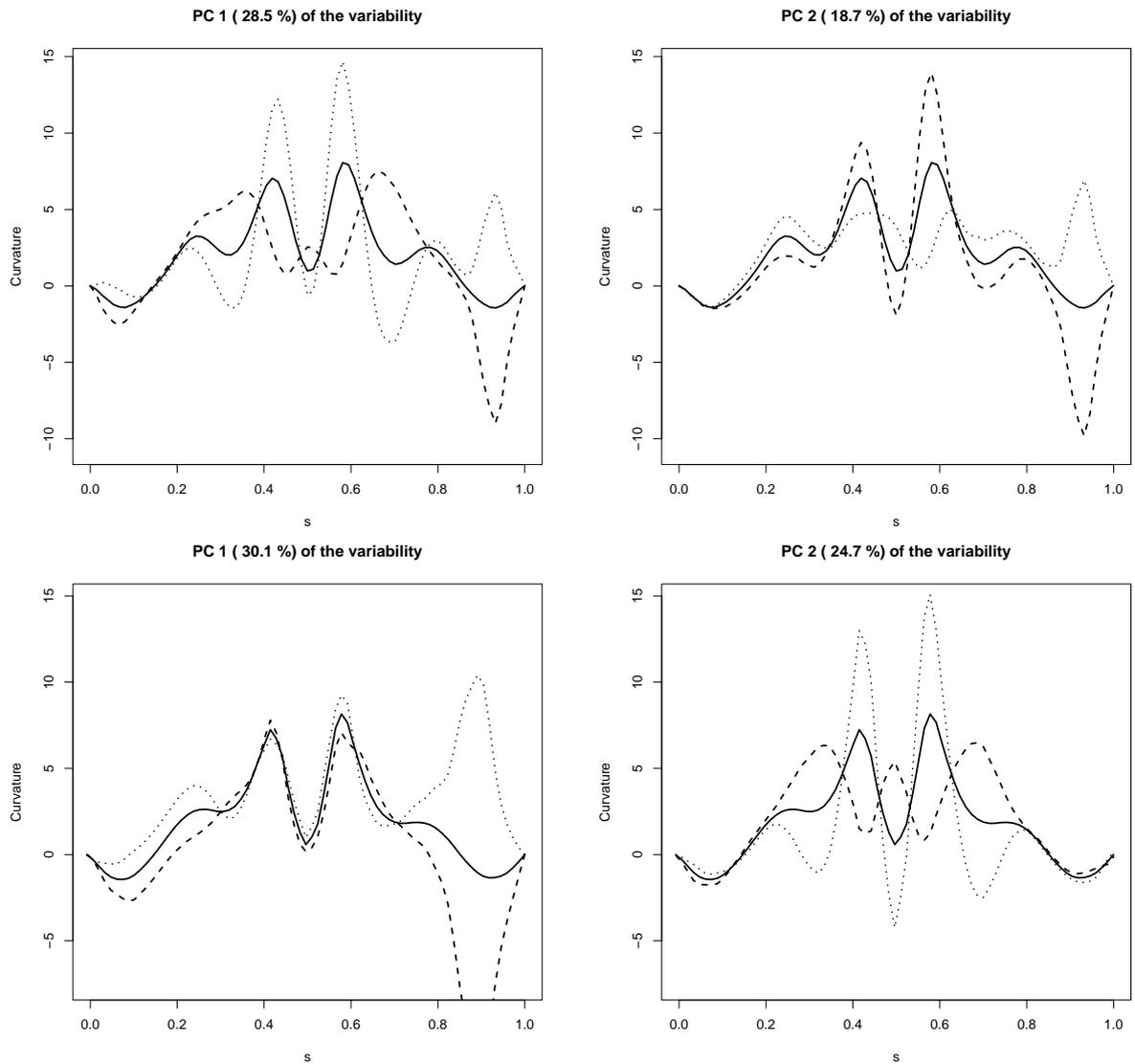
**Table 5.6:** Multiple comparisons of the amplitude adjustment required to produce the structural average control curvature function for the upper lip.

The results from Table 5.6 show that it is highly likely that, for both  $xy$  and  $xz$  curvature, the population median amplitude adjustment is larger for UCL and UCLP children than control children. This indicates that in general the amount of bending, both into and out of the face and up and down the face, experienced along the upper lip is different for UCL and UCLP children than for control children. These results suggest that the operation to correct the cleft lip and palate has not ensured the upper lip of the UCL and UCLP children are in general the same as control children in terms of three dimensional bending. There is no evidence of a difference between the UCL and UCLP children in terms of amplitude adjustment for either the  $xy$  or  $xz$  curvature functions.

#### 5.2.4 Principal components analysis

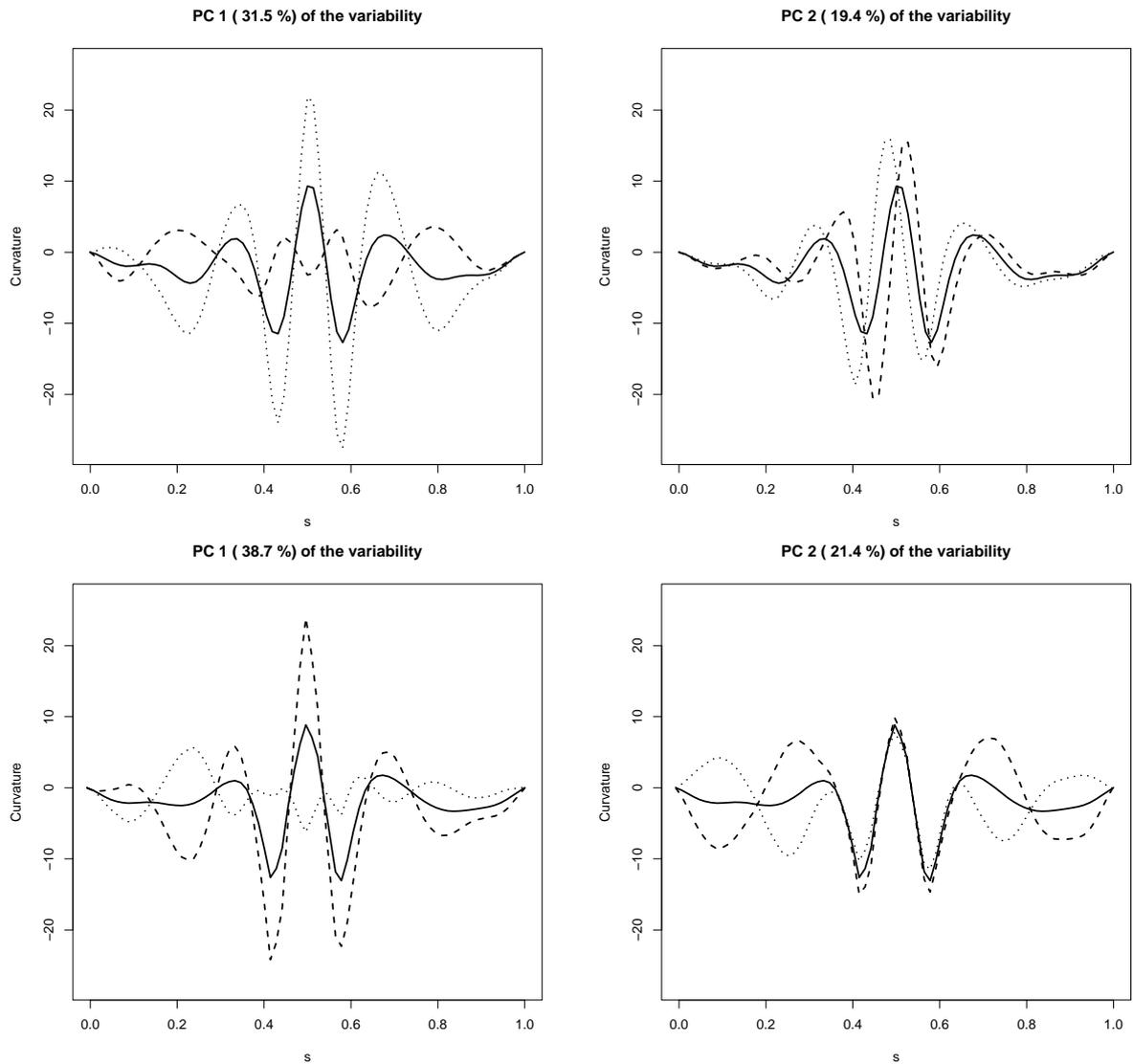
To investigate how the two sets of curvature functions vary and how the variation is affected by the different groups of children principal components analysis is carried out on the curvature functions as outlined in Section 5.1.4. Plots showing the effect of the first two principal components on the mean function are found in Figure 5.11, for  $xz$  curvature unaligned and aligned, and in Figure 5.12, for  $xy$  curvature unaligned and aligned.

The principal components analysis of the raw  $xz$  curvature functions in Figure 5.11 shows that the major source of variation, which is explained by principal component one, appears to be the variation in the position of the start and the end of the Cupid's bow and the variation caused at the end of the function due to the unusual UCLP case. The second component appears to explain the variation in the magnitude of the curvature at the ends of the Cupid's bow along with further variation caused by the unusual case. After aligning the curvature functions the variation explained by the first principal component appears to be in part variation in the curvature values just before and after the Cupid's bow but is dominated by the variation due to the unusual case. The second principal component mainly explains the variation in curvature around the Cupid's bow.



**Figure 5.11:** The effect of the first two principal components on the average curve for the unaligned (top row) and aligned (bottom row) xz curvature functions. The solid line is the mean function, the dashed line is  $mean + 3\sqrt{\lambda_i}e_i$  while the dotted line is  $mean - 3\sqrt{\lambda_i}e_i$ .

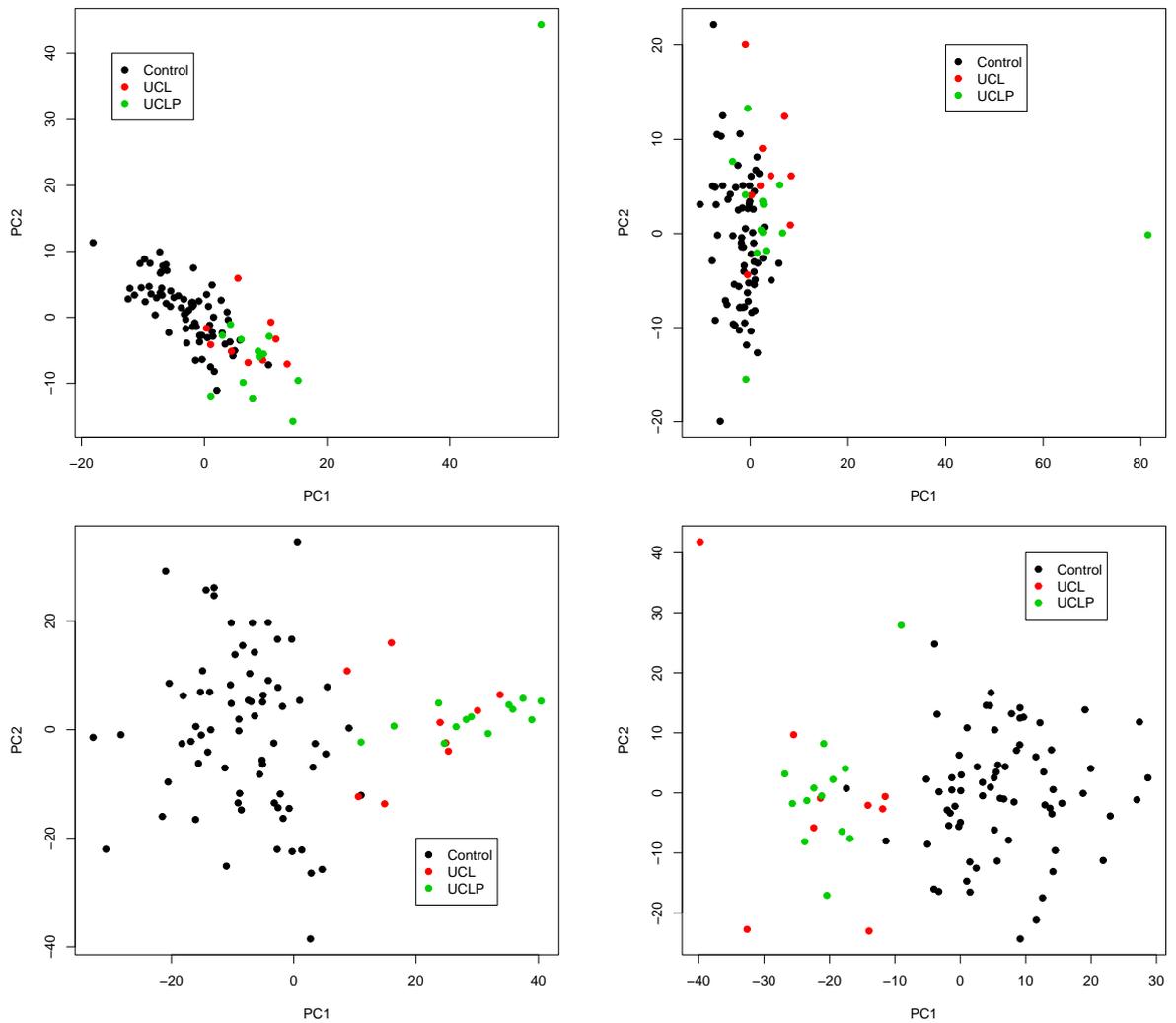
The principal components analysis of the unaligned xy curvature functions in Figure 5.12 shows that the first principal component describes a global variation between the functions along the whole upper lip. The second principal component describes the variation in curvature, and potentially position, around the Cupid's bow region. Once the functions are aligned the first principal component still shows global variation but concentrates on the magnitude of curvature at the characteristic points. Now however, the second principal component shows only variation away from the Cupid's bow region.



**Figure 5.12:** The effect of the first two principal components on the average curve for the unaligned (top row) and aligned (bottom row)  $xz$  curvature functions. The solid line is the mean function, the dashed line is  $mean + 3\sqrt{\lambda_i}\mathbf{e}_i$  while the dotted line is  $mean - 3\sqrt{\lambda_i}\mathbf{e}_i$ .

Now that the effects of each of the principal components have been explained it is of interest to examine the scores for each of the first two principal components to see which, if any, of the groups dominate the variation in that component. To do this the first and second component scores for each principal components analysis are plotted in Figure 5.13.

Considering firstly the principal component scores for the  $xz$  curvature functions it is clear that one UCLP case dominates the variation explained by both the first and second principal components for the unaligned curvature functions.



**Figure 5.13:** Upper lip first and second principal component scores for the unaligned xz curvature functions (top left), aligned xz curvature functions (top right), unaligned xy curvature functions (bottom left) and the aligned xy curvature functions (bottom right).

In the rest of the data it appears that UCL and UCLP cases in general have higher first principal component scores and lower second principal component scores than control children. These groupings would appear to be due to the difference between control children and UCL and UCLP children in position (first principal component) and magnitude of curvature (second principal component) of the characteristic points. The first principal component of the aligned xz curvature functions is dominated by the unusual UCLP case while the second principal component shows no clustering of points due to group.

The principal component scores for the xy curvature functions show that for

both unaligned and aligned functions there is a clear separation between the control group and the UCL and UCLP group in the first principal component. The first principal component shows a general variation along the curvature function. It would appear that a number of the UCL and UCLP cases have a large effect on this principal component. This indicates that a number of the UCL and UCLP curves are very different from the usual control pattern. There is little separation between the groups in terms of the second principal component.

The shape of the upper lip for UCL and UCLP children has been compared to the shape of the upper lip of control children using various techniques involving curvature. From this analysis there is evidence that even after surgery UCL and UCLP children in general have a start of a Cupid's bow which occurs earlier on the upper lip and an end of the Cupid's bow which occurs later on the upper lip than control children. There is evidence that the upper lips for UCL and UCLP children are less rounded than control children especially up and down the face and also that UCL and particularly UCLP children do not have an upturn in the middle of the Cupid's bow that almost all control children have.

## Chapter 6

# Regression with Functional Predictors

The work carried out in Chapters 3, 4 and 5 involved determining the amount of bending experienced by various facial features and using this measure of facial shape to compare control children with children born with a cleft lip and/or palate. In the study of one year old children the facial shapes of children born with a cleft lip and/or palate do appear to differ from control children. It is possible to imagine that having facial shapes which are different from control children may potentially have a detrimental effect on the psychological state of a child. As part of a study into ten year old children born with a cleft lip and/or palate, facial image data, similar to the data available for the one year old children, were produced for 68 control children, 44 children born with a unilateral cleft lip (UCL) and 51 children born with a unilateral cleft lip and palate (UCLP). The images were taken when the child's mouth was closed. Furthermore, a psychological questionnaire was administered to the parents of the children born with a cleft. The questionnaire was the 'Revised Rutter Scales' which is a revision of the original 'Rutter Parents' and Teachers' Scales' outlined by Rutter (1967) and gives a total emotional/behavioural difficulties score ranging from 0 to 56. A low score indicates that the child has few psychological difficulties whilst a large score indicates that the child has more psychological difficulties.

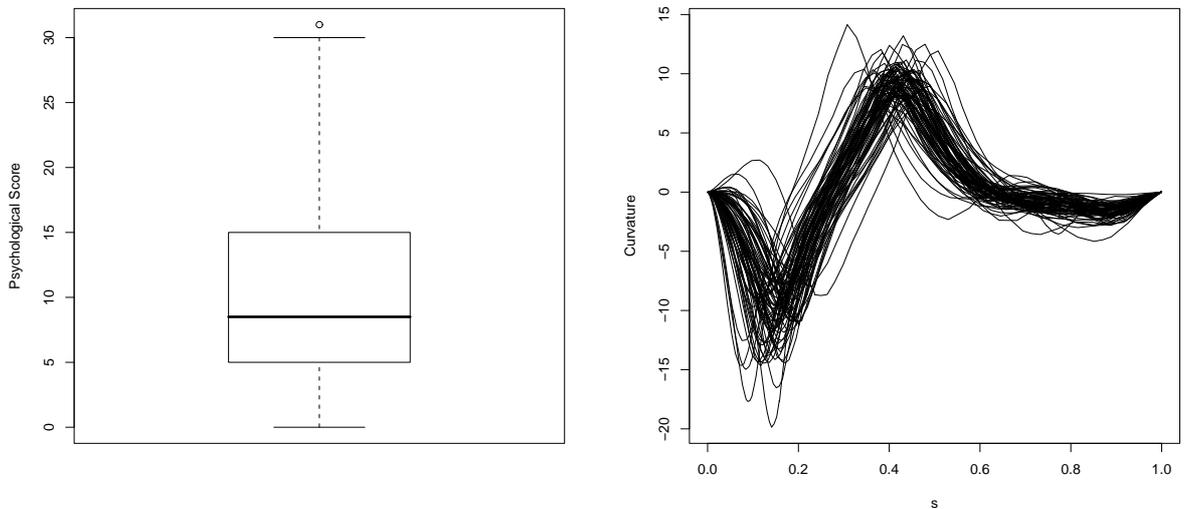
There is interest in investigating any relationship between psychological state and facial shape for the cleft children. The response in this setting is the scalar value which characterises the psychological state of the child. Using the work from Chapters 3 and 4 facial shape will be defined by curvature functions of various features. There are curves available on five facial features (midline profile, upper

lip, nasal base, nasal rim and nasal bridge) with the perpendicular planes method for calculating curvature describing each feature using two curvature functions. This results in ten potential functional predictors of the scalar response. Children born with a unilateral cleft lip and children born with a unilateral cleft lip and palate are combined into one group giving 95 respondents. Due to a number of difficulties in data collection, including extracting curves from the images and refusal/inability of participants to fill in questionnaires, only 80 subjects have a psychological score and a full set of facial curves. Analysis will be carried out on these 80 respondents.

The difficulty in working with functional predictors is two-fold. Firstly, it is clear that standard regression techniques will have to be extended to deal with the functional nature of the predictors. Secondly, displaying the data and the results can be more complex than with scalar predictors. This chapter aims to deal with both these issues. Section 6.1 will describe some techniques for investigating the relationship between a scalar response and a single functional predictor, with the relationship between the psychological score and the yz curvature (bending into and out of the face) function of the midline profile used to illustrate the methods. Section 6.2 will describe some techniques for investigating the relationship between a scalar response and multiple functional predictors. These methods will be illustrated by investigating the relationship between psychological score and four functional predictors: yz curvature of the midline profile, xy curvature (across the face) of the midline profile, xz curvature of the nasal bridge and xy curvature (up and down the face) of the nasal bridge. Chapter 7 will provide a more in-depth investigation into the relationship between psychological score and facial shape.

## 6.1 Scalar Response and a Single Functional Predictor

This section will describe techniques used to investigate the relationship between a scalar response and a functional predictor. Figure 6.1 shows a boxplot of the psychological scores and the yz curvature functions of the midline profiles of each of the subjects.



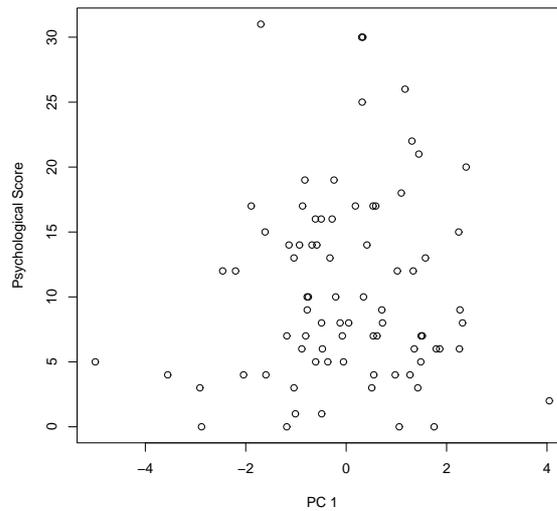
**Figure 6.1:** Boxplot of psychological scores for cleft children (left). Plot of yz curvature against arc length for each subject. (right)

The boxplot of psychological scores in Figure 6.1 shows scores ranging from 0 to 30 with a median score of about 8. The plot of yz curvature functions shows a similar pattern to earlier plots of yz curvature of the midline profile, with all functions showing an area of minimum curvature corresponding to the base of the nose and an area of maximum curvature corresponding to the tip of the nose. The major source of variation between the functions is in the position of, and magnitude of, these areas of minimum and maximum curvature. The interest throughout this section will be in introducing techniques to investigate any potential relationship between a functional predictor and a scalar response. The data introduced here are simply for illustration and only brief interpretation of results will be given.

### 6.1.1 Displaying the data

The first problem with using functional predictors is considering how best to display the data to give an indication of any relationship between the response and predictor. Clearly in the standard regression context a plot of response against predictor will give an indication of any relationship. However, with functional predictors this is not possible. A major problem is that it is difficult to naturally order the functions. However, if functional principal component analysis (details given in Section 2.3.2) is carried out on the predictors, then each functional predictor has a component score for each principal component. Suppose

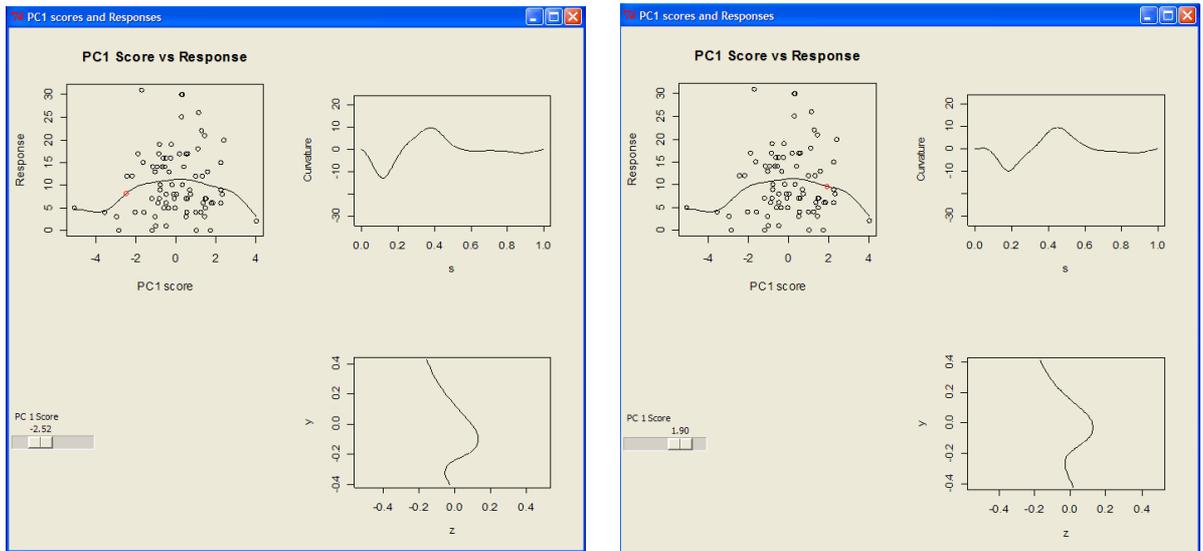
only the first principal component is considered then the response can be plotted against the component score and any relationship between the response and first principal component score may give an indication of potential relationships between the response and the functional predictor. This introduces an ordering of the predictors, allowing simpler graphical displays. Figure 6.2 shows a plot of psychological score against the first principal score of the yz curvature functions.



**Figure 6.2:** Psychological score against first principal component score.

It seems from Figure 6.2 that there may be some slight suggestion of a positive relationship between first principal component score and psychological score suggesting there may be an indication of a relationship between psychological score and yz curvature of the midline profile for UCLP children.

A function of the best fit of response against principal component score can look more formally at the potential relationship. For various component scores, it is interesting to show the estimated response from the best fit function and the function produced by adding the effect of the principal component score to the mean predictor. The best fit can be found using numerous techniques including standard regression, smoothing splines or nonparametric regression. The results can be displayed both easily and neatly in **R** using `rpanel` where a slider can be used to control the setting of the principal component score. Two `rpanel` displays are shown in Figure 6.3 for the cleft data, one for a low component score and one for a high component score.



**Figure 6.3:** The `rpanel` showing; the nonparametric fit of the relationship between the first principal component score and psychological score (top left), the  $yz$  curvature function corresponding to that component score (top right) and the reconstructed midline profile using that curvature function (bottom right). The left `rpanel` shows a low component score whilst the right `rpanel` shows a high component score.

The `rpanel` is very useful in investigating how both the response and the functional predictor changes as the principal component score changes. This can give an insight into how the functional predictor may be related to the response. However, in the example shown in Figure 6.3, for the majority of the data there does not appear to be much relationship between first component score and psychological score. This is perhaps an indication that there is little relationship between psychological score and  $yz$  curvature of the midline profile. The `rpanel` plot is still interesting as it shows that the effect of the component score decreasing is that the base of the nose and the tip of the nose occur lower on the midline profile.

The component scores from any of the principal components can be used to illustrate the data. However, since the first principal component explains the largest variation between the functional predictors it is the relationship between the response and this component which is most likely to point to a potential relationship between the response and the functional predictor.

### 6.1.2 Regression on principal component scores

As an extension of simply plotting the response against the first principal component score it is worthwhile to look more formally at the relationship between the response and the first principal component score and investigate whether there is evidence of a significant relationship. Furthermore, there is no reason to limit the predictor to just the first principal component. It makes sense to include all the principal components which explain a reasonably large amount of the variation in the functional predictors.

Suppose that the scalar response is denoted by  $y_i$  and the functional predictor denoted by  $x_i(s_i)$ , where  $s \in (0, 1)$ , for the  $n$  response-predictor pairs. Further denote the  $j$ th principal component score of the  $i$ th predictor as  $a_{ij}$ . Then a standard linear regression can be produced such that

$$y = \alpha + \sum_j \beta_j a_j + \epsilon \tag{6.1}$$

Finding the parameters  $\alpha$  and  $\beta_j$  which minimise the residual sum of squares gives the best linear relationship between the response and the principal component scores.

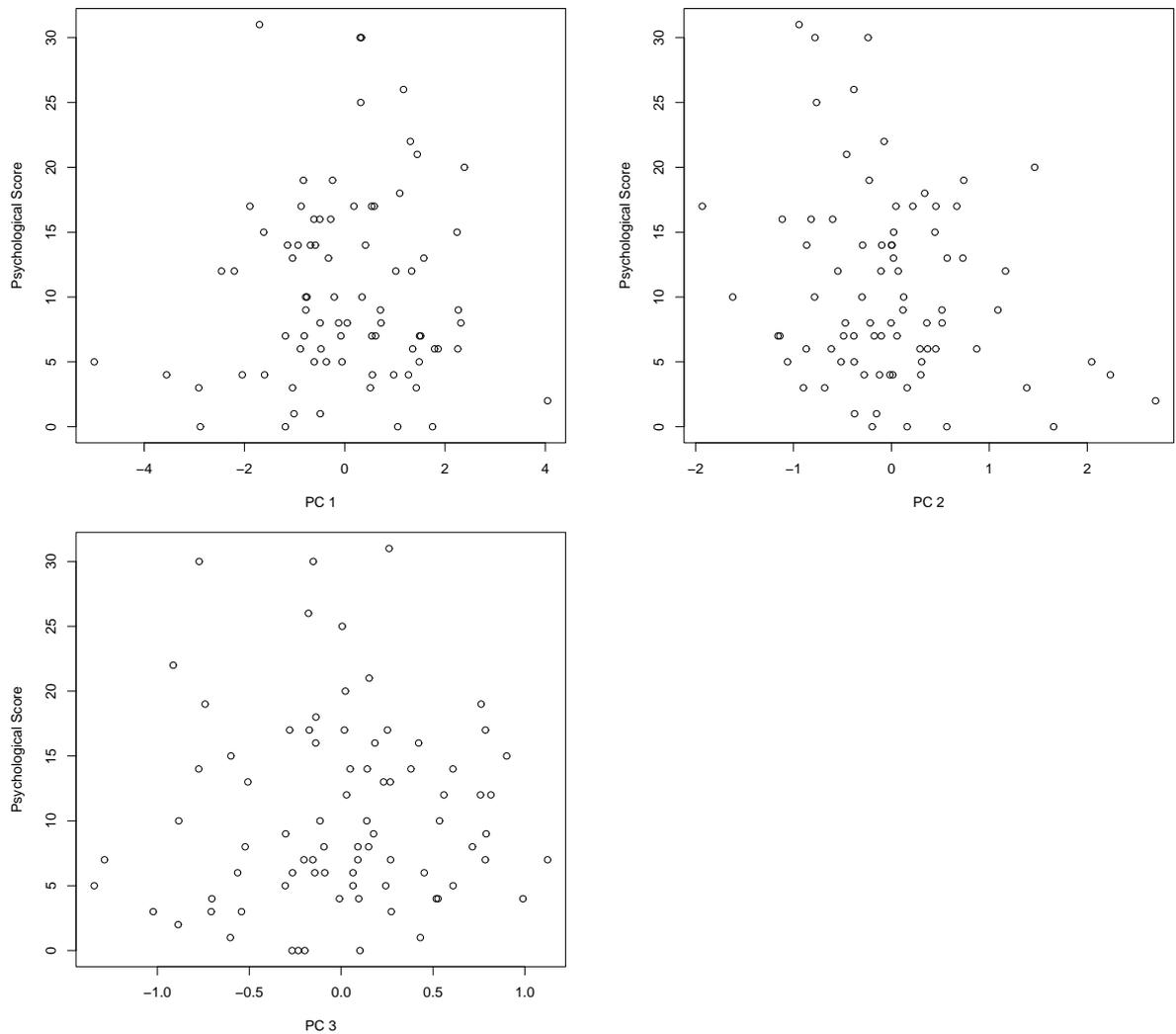
This linear regression process is carried out on the cleft data set using the first three principal component scores as scalar predictors. Plots of each component score against psychological score are given in Figure 6.4

The p-values corresponding to the evidence of a linear relationship between the component score predictors and the psychological score response when all three component scores are included in the model are shown in Table 6.1.

| Predictor                         | Coefficient | p-value |
|-----------------------------------|-------------|---------|
| First principal component scores  | 0.25        | 0.648   |
| Second principal component scores | -2.06       | 0.044   |
| Third principal component scores  | 0.501       | 0.741   |

**Table 6.1:** Significance of each principal component score as a predictor of psychological score.

The p-values in Table 6.1 suggest that the scores for the second principal component have a significant linear relationship with psychological score. In general it would appear that as the second principal component score increases the psychological score decreases. The fact that the second component score has



**Figure 6.4:** Principal component score for the first three principal components against psychological score.

a relationship with psychological score gives some indication that there may be a relationship between the yz curvature functions and psychological score.

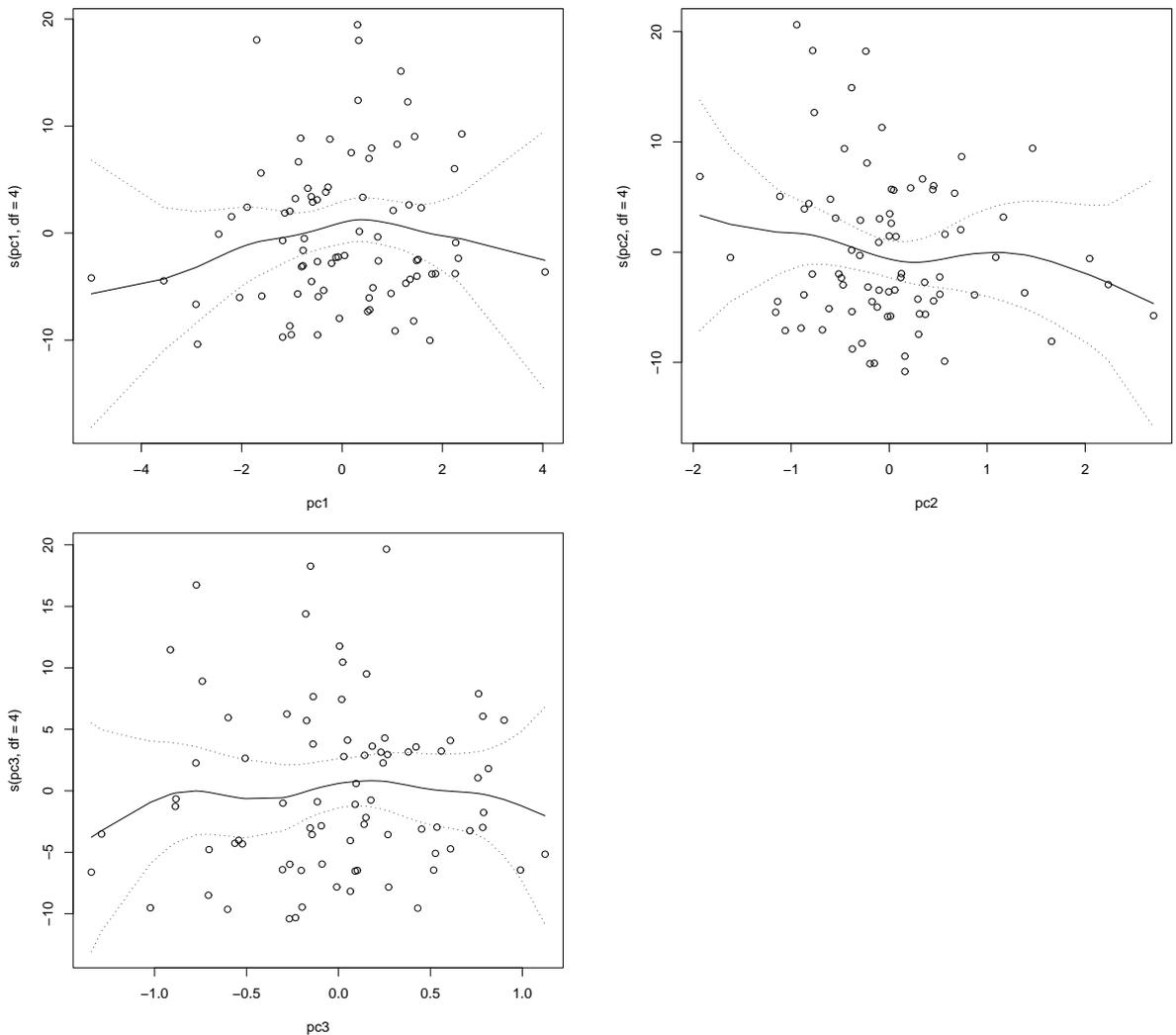
It is often the case that the relationship between the scalar response and scalar predictors is not necessarily linear. Hastie and Tibshirani (1990) suggest that the linear terms  $\beta_j x_j$  in standard linear regression can be replaced by smooth functions to give additive models. When looking for the relationship between the response and principal component scores an additive model can be produced by extending (6.1) to

$$y = \alpha + \sum_j f_j(a_j) + \epsilon \quad (6.2)$$

The  $f_j$ s are arbitrary univariate functions, one for each principal component,

which are assumed to be smooth. The  $f_j$ 's can be estimated using the backfitting algorithm outlined by Hastie and Tibshirani (1990).

An additive model is calculated for the cleft data using the first three principal component scores as predictors. The smoother used in the algorithm is a smoothing spline with 4 degrees of freedom. Figure 6.5 shows the fitted functions for each set of principal component scores.



**Figure 6.5:** Additive model functions for the relationship between principal component score for the first three principal components and psychological score for all subjects. The solid lines are the fitted function of component score whilst the dashed lines are the standard errors.

The p-values corresponding to the evidence of a relationship between the component score predictors and the psychological score response when all three sets of component scores are included in the additive model are shown in Table 6.2.

| Predictor                         | p-value |
|-----------------------------------|---------|
| First principal component scores  | 0.7     |
| Second principal component scores | 0.7     |
| Third principal component scores  | 0.6     |

**Table 6.2:** Significance of each smooth function of principal component score as a predictor of psychological score.

The p-values in Table 6.2 suggest there is no evidence of a significant relationship between any of the sets of principal component scores and psychological score when the additive model contains all three principal component scores as predictors.

### 6.1.3 Functional linear model

Ramsay and Silverman (1997) suggest that standard linear regression can be extended to allow a functional predictor by replacing the usual scalar parameter  $\beta$  with a functional parameter  $\beta(s)$  such that

$$y = \alpha + \int_0^T \beta(s)x(s)ds + \epsilon \quad (6.3)$$

where the errors  $\epsilon$  have mean zero and are independent from the predictors.  $T$  is the length of interval over which the functional predictors are measured.

Clearly the difficulty with the functional linear model (6.3) is how best to estimate the functional parameter. In standard linear regression, the least squares method is used and this can be extended to estimate the parameters in (6.3) by minimising

$$SSE = \sum_{i=1}^n [y_i - (\alpha + \int_0^T \beta(s)x_i(s))]^2 \quad (6.4)$$

It can be shown that for any response and any functional predictor there is a set of  $\alpha$  and  $\beta(s)$  which results in a residual sum of squares of zero. Clearly it is not the case that any functional predictor can predict any response perfectly. The reason for the residual sum of squares being equal to zero is that, in a sense, there is an infinite number of parameters, since  $\beta(s)$  is defined continuously along  $s$ , while there are only  $n$  equations to solve. This means that, regardless of the sample size, using least squares to find an estimate of the functional parameter is not appropriate. Therefore the function must be regularised.

One regularisation approach is to evaluate the predictors at specific points and consider a multiple linear regression using these values as the predictors. Each functional predictor  $x_i(s_i)$  can be evaluated at  $p$  points  $(x_{i1}, \dots, x_{ip})$  at  $(s_1, \dots, s_p)$  where  $s_{j+1} - s_j = \frac{1}{p-1}$  and the multiple linear regression can be written as

$$y = \alpha + \sum_{j=1}^p \beta_j x(s_j) + \epsilon \tag{6.5}$$

Clearly the choice of  $p$  is important. If  $p$  is too small then  $(x_{i1}, \dots, x_{ip})$  will not give a good representation of the functional predictor. However, it is important that the number of parameters  $p + 1$  is less than the number of observations to allow least squares to be used to fit the model. This illustrates one of the major draw-backs of this regularisation method, namely the fact that when the sample size is low only a small number of points can be chosen to define each function. It would be preferable if the whole function could be used as the predictor rather than a set of discrete points.

An approach which regularises the functional predictor but retains the functional structure of the predictor is to express the functional predictors in terms of basis functions. The description here will be in terms of B-spline bases but any bases can be used to suit the problem. Note that the solution for Fourier bases is slightly simpler (see Ramsay and Silverman (1997) for details). Suppose that each predictor is described using  $m$  B-spline basis functions and denote these as  $\phi_1, \dots, \phi_m$ . Each predictor can then be written as the expansion

$$x_i(s_i) = \sum_{v=1}^m c_{iv} \phi_v = c_i^T \phi \tag{6.6}$$

where  $\phi = \phi_1, \dots, \phi_m$  is the set of basis functions and  $c_{iv}$  is the coefficient of basis function  $v$  for predictor  $i$ . The functional parameter can also be expressed in terms of these basis functions as

$$\beta(s) = \sum_{v=1}^m b_v \phi_v = b^T \phi \tag{6.7}$$

where  $b_v$  is the coefficient of basis function  $v$  for the parameter. Now the aim is to express  $\int_0^T \beta(s)x_i(s)$  in terms of the coefficients of the basis functions. To allow this, since B-splines are not orthonormal, the  $m \times m$  matrix  $J = \int \phi \phi'$  must be

defined where

$$J_{jk} = \int \phi_j(s)\phi_k(s)ds \tag{6.8}$$

Now using an extension of the Parseval identity it follows that

$$\begin{aligned} \int_0^T \beta(s)x_i(s)ds &= \sum_{j=1}^m \sum_{v=1}^m c_{ij}b_v \int_0^T \phi_j(s)\phi_v(s)ds \\ \int_0^T \beta(s)x_i(s)ds &= \sum_{j=1}^m \sum_{v=1}^m c_{ij}J_{jv}b_v \end{aligned} \tag{6.9}$$

It is perhaps simpler to think of the process in matrix form. Denote the  $(n \times 1)$  matrix of the responses as  $Y = (y_1, \dots, y_n)^T$ , the  $((m + 1) \times 1)$  matrix of the parameters as  $\zeta = (\alpha, b_1, \dots, b_m)^T$  and the  $(n \times (m + 1))$  coefficient matrix of the predictors as  $Z = [1 \ CJ]$  where  $C$  is the  $(n \times m)$  coefficient matrix with row  $i$  containing the B-spline coefficients for predictor  $i$ . Now the model can be defined by the equation

$$Y = Z\zeta + \epsilon \tag{6.10}$$

where  $\epsilon = (\epsilon_1, \dots, \epsilon_n)$  and the  $\epsilon_i$ s have mean zero. It follows then that

$$\hat{Y} = Z\hat{\zeta} \tag{6.11}$$

Therefore the least squares estimate of the parameter vector  $\zeta$  is given by

$$\begin{aligned} Z^T Z\hat{\zeta} &= Z^T Y \\ \hat{\zeta} &= (Z^T Z)^{-1} Z^T Y \end{aligned} \tag{6.12}$$

It is clear that the major interest from this analysis is the shape of the functional parameter  $\beta(s)$  to investigate the effect of the functional predictor on the response at different points along the  $s$  axis. A potential disadvantage of using (6.12) to calculate  $\beta(s)$  is that there is no smoothness constraint on the functional parameter and a rough functional parameter may be difficult to interpret. One way to ensure the functional parameter is smooth is to use a low number of basis functions to define the functional predictors and in turn the functional parameter. There are two major draw-backs to simply lowering the number of basis functions; firstly a low number of basis functions may result in a poor representation of the original predictors and secondly important features in the functional parameter may be missed. To avoid reducing the number of basis functions a roughness

penalty can be applied to (6.4). The roughness penalty suggested by Ramsay and Silverman (1997) is to use the second derivative of the functional parameter squared to penalise the residual sum of squares. Using this method the penalised sum of square error becomes

$$PENSSE = \sum_{i=1}^n [y_i - (\alpha + \int_0^T \beta(s)x_i(s))]^2 + \lambda \int_0^T \beta''(s)^2 ds \quad (6.13)$$

The integral of the squared second derivative gives a measure of the roughness of the functional parameter whilst the smoothing parameter  $\lambda$  controls the trade off between the roughness of the functional parameter and the accuracy of  $\hat{y}$  as a predictor of  $y$ .

To write the penalised sum of squares error in vector-matrix form define the matrix  $K$  with entries

$$K_{jk} = \int_0^T \phi_j''(s)\phi_k''(s)ds \quad (6.14)$$

Now the penalised sum of square error can be defined as

$$\begin{aligned} PENSSE &= \sum_{i=1}^n (y_i - \alpha - \sum_{v=1}^m c_{iv}b_v\phi_v)^2 + \lambda \int_0^T (\sum_{v=1}^m b_v\phi_v''(s))^2 ds \\ PENSSE &= \|y - \alpha - Cb\|^2 + \lambda b^T K b \end{aligned} \quad (6.15)$$

Once again define the  $(n \times 1)$  matrix of the responses as  $Y = (y_1, \dots, y_n)^T$ , the  $((m + 1) \times 1)$  matrix of the parameters as  $\zeta = (\alpha, b_1, \dots, b_m)^T$  and the  $(n \times (m + 1))$  coefficient matrix of the predictors as  $Z = [1 \ CJ]$  where  $C$  is the  $(n \times m)$  coefficient matrix with row  $i$  containing the B-spline coefficients for predictor  $i$ . Now augment the penalty matrix  $K$  with a leading row and column of  $m + 1$  zeros and call this augmented matrix  $K_0$  i.e.

$$K_0 = \begin{bmatrix} 0 & 0 \\ 0 & K \end{bmatrix} \quad (6.16)$$

Now (6.15) is further simplified to

$$PENSSE = \|y - Z\zeta\|^2 + \lambda \zeta^T K_0 \zeta \quad (6.17)$$

and the minimising  $\hat{\zeta}$  is given by

$$(Z^T Z + \lambda K_0)\hat{\zeta} = Z^T y$$

$$\hat{\zeta} = (Z^T Z + \lambda K_0)^{-1} Z^T y \tag{6.18}$$

Ramsay and Silverman (2006) state that it is possible to calculate confidence limits for the parameters  $\alpha$  and  $\beta$  using the vector-matrix equation

$$Var(\hat{\zeta}) = \hat{\sigma}_e^2 (Z^T Z + \lambda K_0)^{-1} Z^T Z (Z^T Z + \lambda K_0)^{-1} \tag{6.19}$$

where  $\hat{\sigma}_e^2$  is the mean of the squared residuals.

Clearly the choice of smoothing parameter is very important. It can either be chosen subjectively or can be estimated using cross-validation. There may be occasions where manually setting  $\lambda$  is useful as a certain smoothness of the functional parameter is necessary for simple interpretation. However, the advantage of the cross-validation method is that it gives an automatic way to find the smoothing parameter which gives the ‘best’ trade off between roughness and accuracy according to the observed data. The cross-validation score for a smoothing parameter  $\lambda$  can be define as

$$CV(\lambda) = \sum_{j=1}^n [y_j - (\hat{\alpha}_{-j} + \int_0^T x_j(s)\beta_{-j}(s)ds)]^2 \tag{6.20}$$

where  $\hat{\alpha}_{-j}$  and  $\hat{\beta}_{-j}(s)$  are the estimates of the parameters calculated by minimising the penalised sum of square errors from (6.17) using all the data except that from subject  $j$ .

Using (6.20) would be computationally expensive since the parameters have to be calculated  $n$  times for each  $\lambda$  in the cross-validation procedure. However Green and Silverman (1994) show that the cross-validation score can be calculated from

$$CV(\lambda) = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - S_{ii}} \right)^2 \tag{6.21}$$

where  $S$  is the hat-matrix for the smoothing process which maps the observed values to their predict values for a given  $\lambda$  such that  $\hat{y} = Sy$ . Now Ramsay and Silverman (2006) state that if there is a large number of data points and at most there is a moderate number of basis functions then the diagonal elements of  $S$  can be estimated from

$$S = Z(Z^T Z + \lambda K_0)^{-1} Z^T \tag{6.22}$$

This method for calculating the cross-validation score is much less computationally expensive.

It is of interest to formulate a test to analyse whether there is evidence that the scalar response is related to the functional predictor. Bowman and Azzalini (1997) suggest a test for no effect which is a test with a null hypothesis that variability in the response is due to natural variability against an alternative hypothesis that variability in the response can be explained to some degree by the predictor. The work looks at the relationship between a scalar predictor and a scalar response but this can be naturally extended to incorporate the functional structure of the predictor.

The two models can be written as

$$\begin{aligned} H_0 & : \mathbb{E}(y_i) = \mu \\ H_1 & : \mathbb{E}(y_i) = \alpha + \int_0^T \beta(s)x_i(s)ds \end{aligned}$$

As in standard model comparison the residual sum of squares is an appropriate way of describing how much of the variability in the data is explained by each model. The residual sum of squares can be defined for each model as

$$\begin{aligned} RSS_0 & = \sum_{i=1}^n [y_i - \bar{y}]^2 \\ RSS_1 & = \sum_{i=1}^n [y_i - (\alpha + \int_0^T \hat{\beta}(s)x_i(s)ds)]^2 \end{aligned}$$

To quantify the difference between these residual sums of squares Bowman and Azzalini (1997) suggest a pseudo-likelihood ratio statistic given by

$$F^* = \frac{RSS_0 - RSS_1}{RSS_1} \tag{6.23}$$

$F^*$  is proportional to the usual  $F$  statistic, although no mention is made of the degrees of freedom, with the ratio effect scaling out the error variance. Note that a major reason for using this pseudo-likelihood statistic is that the alternative hypothesis model is not parametric and therefore not fitted using maximum likelihood.

Now the distribution of  $F^*$  under the null hypothesis must be found. To facilitate this it is helpful to express  $F^*$  in terms of quadratic forms. So the

residual sum of squares for each of the models can be written as

$$\begin{aligned} RSS_0 &= y^T(I - L)^T(I - L)y = y^T(I - L)y \\ RSS_1 &= y^T(I - W)^T(I - W)y \end{aligned}$$

where  $L$  is an  $(n \times n)$  matrix filled with  $\frac{1}{n}$ .  $W$  is the  $(n \times n)$  weight matrix which for the functional linear model is given by the hat matrix  $S$ . Now  $F^*$  can be expressed as

$$F^* = \frac{y^T B y}{y^T A y} \tag{6.24}$$

where  $A = (I - W)^T(I - W)$  and  $B = I - L - A$ .

The major interest here is to calculate the significance of the  $F^*$  statistic. This can be found from the p-value

$$p = \mathbb{P}(F^* > F_{obs}^*)$$

where  $F_{obs}^*$  is the  $F^*$  statistic calculated from the observed data. This can be rewritten as

$$\begin{aligned} p &= \mathbb{P}\left(\frac{y^T B y}{y^T A y} > F_{obs}^*\right) \\ &= \mathbb{P}(y^T C y > 0) \end{aligned} \tag{6.25}$$

where  $C = B - F_{obs}^* A$ . Johnson and Kotz (1972) summarise general results about the distribution of a quadratic form in normal variables for any symmetric matrix (e.g.  $C$ ). These results are simplest to apply when the normal variables have mean zero. Here,  $y_i$  has mean  $\mu$ . It is simple to see though that due to the differencing involved in the residual sum of squares that  $\mu$  is eliminated. To show this, consider rewriting the residual sum of squares for model  $H_0$  as

$$\begin{aligned} RSS_0 &= \sum_{i=1}^n [(\mu + \epsilon_i) - (\mu + \bar{\epsilon})]^2 \\ &= \sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2 \end{aligned}$$

Therefore the quadratic form  $y^T C y$  is equivalent to the quadratic form  $Q = \epsilon^T C \epsilon$  and clearly the  $\epsilon_i$ 's have mean zero. The work of Johnson and Kotz (1972) now allow the probability to be calculated in numerical form. However, when  $n$  is

large the calculations become awkward. Since an exact p-value is not essential it is sufficient to provide an approximation. This can be done by replacing  $Q$  with a more computationally convenient distribution with the same first three moments. This works well here since the cumulants of  $Q$  can be computed as

$$k_j = 2^{j-1}(j-1)!tr\{(IC)^j\}$$

where  $C$  is as above. Johnson and Kotz (1972) state that a shifted and scaled chi-squared distribution often gives a good approximation of a quadratic form. So matching the moments of a  $a\chi + b$  distribution to the moments of  $Q$  gives

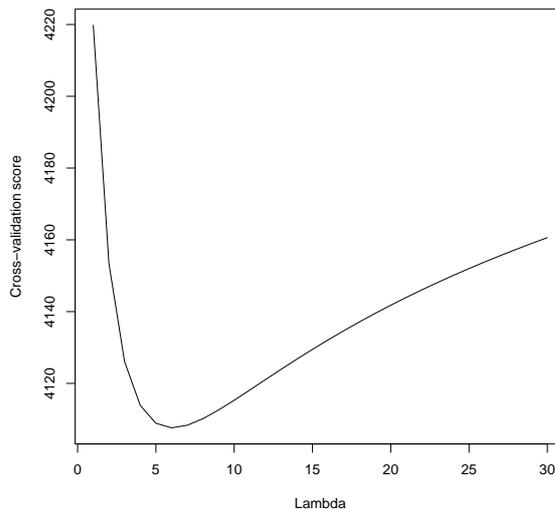
$$\begin{aligned} a &= \frac{|k_3|}{4k_2} \\ b &= \frac{8k_2^3}{k_3^2} \\ c &= k_1 - ab \end{aligned}$$

With a chi-squared distribution with  $b$  degrees of freedom, the p-value can be estimated as  $1 - q$  where  $q$  is given by the probability of lying below the point  $\frac{-c}{a}$ .

A simple simulation study was carried out to validate this pseudo-likelihood ratio test. In the study there were 50 responses  $(y_1, \dots, y_{50})$  taken as a random draw from a sequence, of length 10000, from 0 to 100 in each simulation. There were four functional predictors tested. Two functional predictors were chosen such that a relationship between response and predictor would be expected. The first functional predictor ( $x_1$  say) simply scaled a cubic in the range  $-1 \leq s \leq 1$  dependent on the response such that  $x_{1i} = y_i \times s^3$ . The second of these ( $x_2$ ) added a value drawn from a random normal, with mean dependent on the response, to the line of equality at 100 regular points along the  $s$  axis then interpolated these 100 points i.e.  $x_{2ij} = s_j + \delta_{ij}$  where  $\delta_{ij}$  is a random draw from a  $N(y_i, 1)$  distribution. Two functional predictors were chosen such that no relationship between the response and predictor would be expected. The first of these ( $x_3$ ) added a value drawn from a random normal with mean zero to the line of equality at 100 regular points along the  $s$  axis then interpolated these 100 points while  $x_4$  added a value from a random uniform at 100 regular points then interpolated. In technical form  $x_{3ij} = s_j + \delta_{ij}$  where  $\delta_{ij}$  is a random draw from a  $N(0, 10)$  distribution while  $x_{4ij} = s_j + \delta_{ij}$  where  $\delta_{ij}$  is a random draw from a  $Un(-1, 1)$  distribution.

The simulation was run 100 times for each functional predictor with functional regression used to estimate the response. At a significance level of 95% both  $x_1$  and  $x_2$  showed a significant relationship with the response on all 100 occasions. At the same significance level  $x_3$  incorrectly showed a significant relationship on 1 occasion and  $x_4$  a significant relationship on four occasions. These results suggest that the pseudo-likelihood ratio test is a valid test of a relationship between a functional predictor and a scalar response.

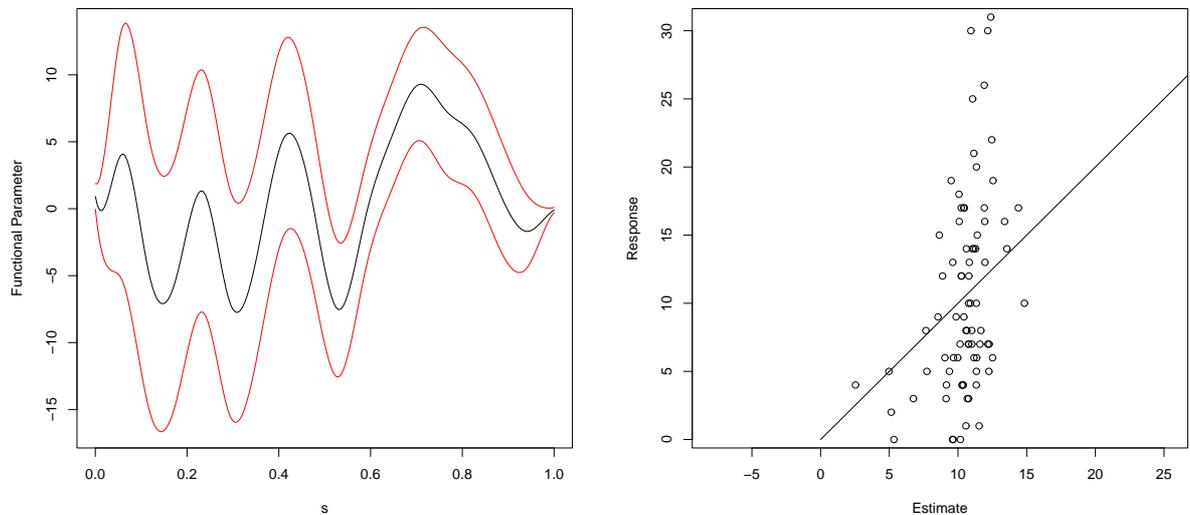
A functional linear model was fitted to the cleft data. The functional predictors and the functional parameter were defined by B-splines with 20 basis functions. To find the smoothing parameter a search across a variety of potential smoothing parameters was carried out with their CV score estimated using (6.21). Since the functional predictors are smooth,  $\lambda$  does not need to be particularly large. Figure 6.6 shows the CV score plotted against  $\lambda$ .



**Figure 6.6:** Cross-validation score function.

The grid search shows that the smoothing parameter which minimises the cross-validation score function is close to 6 so  $\lambda$  is set to 6 and the functional linear model is calculated. Figure 6.7 shows the estimate of the functional parameter and the estimated psychological score using the functional linear model against the true psychological score.

The functional parameter in Figure 6.7 is somewhat difficult to interpret despite the smoothing. To aid interpretation the parameter could be further smoothed by increasing the smoothing parameter or decreasing the number of basis functions. It is however possible to see that the area where the functional parameter is clearly non-zero, i.e. zero is outwith the confidence bands, is around



**Figure 6.7:** Functional parameter for the functional linear model with psychological score as response and yz curvature function of the midline profile as predictor (the red lines indicate confidence limits for the parameter) and estimated psychological score using this model against the true psychological score with the line of equality for reference.

$s = 0.8$ . It is difficult to see why curvature towards the face in this area has an effect on psychological score. Considering the curvature functions in Figure 6.1 there is a very small range of curvature values at  $s = 0.8$  so it is unlikely that the non-zero functional parameter will offer much in terms of predicting psychological score. In fact considering the plot of true against predicted psychological score it is clear that this functional linear model does not appear to be particularly useful in predicting psychological score. To confirm this view a test of no effect is carried out and returns a p-value of 0.126 suggesting that yz curvature of the midline profile is not a useful predictor of psychological score. The  $R^2$  value of the model is calculated using the formula

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

and is found to be 0.14.

#### 6.1.4 Nonparametric functional regression

If there are  $n$  subjects each with a scalar response  $y_i$  and a functional predictor  $x_i(s)$  it is possible to use nonparametric statistics to estimate the response  $y^*$  of

a new subject with a known functional predictor  $x^*(s)$ . Ferraty and Vieu (2006) give a good overview of nonparametric methods for analysing functional data and can be taken as a reference for this section. Note that throughout this section  $x_i(s)$  and  $x^*(s)$  will often simply be denoted by  $x_i$  and  $x^*$ .

The first stage of the method is to define how close the functions  $x_i$  and  $x^*$  are. This can be done using a metric function. The properties of a metric function,  $d$  say, are

1.  $d(x_i, x^*) \geq 0$
2.  $d(x_i, x^*) = 0 \Leftrightarrow x_i = x^*$
3.  $d(x_i, x^*) = d(x^*, x_i)$
4.  $d(x_i, x_j) \leq d(x_i, x^*) + d(x^*, x_j)$

for any functions  $x_i$  and  $x^*$ . A common example of a metric is the  $L_2$  metric e.g.

$$d(x_i(s), x^*(s)) = \sqrt{\int (x_i(s) - x^*(s))^2 ds} \quad (6.26)$$

where  $x_i(s)$  and  $x^*(s)$  are functions parameterised by arc length. However, in some cases it is overly simple to define closeness between functions by the area between them. It may be the case that although the magnitudes of the functions are close the path that they follow may be different. To define closeness between functions in terms of shape rather than location a semi-metric can be used. The semi-metric is a metric with the condition that  $d(x_i, x^*) = 0 \Leftrightarrow x_i = x^*$  relaxed. A commonly used semi-metric function is a function which calculates the area between the second derivatives of each function i.e.

$$d(x_i(s), x^*(s)) = \sqrt{\int (x_i''(s) - x^{*''}(s))^2 ds} \quad (6.27)$$

where  $x_i''(s)$  and  $x^{*''}(s)$  are the second derivatives with respect to  $s$  of the  $x_i$  and  $x^*$  functions respectively. In fact for order greater than zero the integral of the squared difference between the derivatives of the functions gives a semi-metric. Clearly for order zero we have the  $L_2$  metric. In practice any sensible measure of closeness between functions can be chosen when carrying out nonparametric regression. In fact, in the examples shown throughout this thesis, it is often useful to retain the condition  $d(x_i, x^*) = 0 \Leftrightarrow x_i = x^*$  and use the  $L_2$  metric as the measure of closeness.

After producing a measure of closeness between  $x^*$  and the  $x_i$ 's it is necessary to weight the predictive power of the  $y_i$ 's. In short the aim is to give a larger weight to responses for which the corresponding functional predictor is close to the functional predictor of the unknown response. Kernel local weighting is a popular nonparametric tool to give local weights. The weighting is dependent on a density function and a smoothing parameter (called the bandwidth).

To explain the transition from kernel weighting in the scalar case to the functional case, a brief discussion of kernel weighting for multivariate data is given. Suppose now that there is a fixed known vector  $\mathbf{a}$  in  $\mathbb{R}^p$  and weights are required for  $n$  random vectors  $\mathbf{b}_1, \dots, \mathbf{b}_n$ . The previous methods for a single scalar value can be extended relatively simply to the multivariate case. One technique is to produce a multivariate kernel which is a scalar combination of the kernels in each of the  $p$  dimensions. A natural way of doing this is to define the multivariate kernel,  $K^*$  say, as the product of kernel functions in each dimension i.e. if  $\mathbf{x} = (x_1, \dots, x_p)$  then

$$K^*(\mathbf{x}) = K(x_1) \times K(x_2) \times \dots \times K(x_p) \tag{6.28}$$

where  $K(x_i)$  is the value of the kernel function for the scalar value in dimension  $i$  and calculation of  $K^*(\mathbf{x})$  is trivial using kernel weighting of scalar values.

Now the weights  $\omega_1, \dots, \omega_n$  for each of the random vectors  $\mathbf{b}_1, \dots, \mathbf{b}_n$  are given by

$$\omega_i = \frac{1}{h^p} K^* \left( \frac{\mathbf{a} - \mathbf{b}_i}{h} \right) \tag{6.29}$$

Again this weighting function assigns a weight to each  $\mathbf{b}_i$  dependent on how close it is to  $\mathbf{a}$ . The normalisation ( $\frac{1}{h^p}$ ) is given by the product of the normalisations,  $\frac{1}{h}$ , in each dimension.

The problem of assigning kernel weights to functions can now be thought of as an extension of the multivariate method. Suppose there is a fixed known function  $x^*$  and  $n$  random functions  $x_1, \dots, x_n$  all in a set  $\mathcal{E}$ . The direct extension of the multivariate weighting would be to set the weight for function  $i$  as

$$\frac{1}{V(h)} K \left( \frac{d(x^*, x_i)}{h} \right)$$

where  $d(\cdot, \cdot)$  is a suitable semi-metric,  $K$  is an asymmetrical kernel and  $V(h)$  is proportional to the volume of the set in which the  $x_i$ 's have non-zero weight.

$V(h)$  can be defined as the volume of

$$B(x^*, h) = \left\{ x^{*'} \in \mathcal{E}, d(x^*, x^{*'}) \leq h \right\}$$

which is a ball centred at  $x^*$  with radius  $h$ . However, it is not possible to calculate this quantity  $V(h)$  since there is no available reference measure for  $\mathcal{E}$ . Therefore the normalisation must be carried out using another technique. Ferraty and Vieu (2006) suggest that a way of carrying out the normalisation is by using the probability distribution of the random functions. So the weights  $\omega_1, \dots, \omega_n$  for each of the random functions  $x_1, \dots, x_n$  can be given by

$$\omega_i = \frac{K\left(\frac{d(x^*, x_i)}{h}\right)}{\mathbb{E}\left(K\left(\frac{d(x^*, x_i)}{h}\right)\right)} \quad (6.30)$$

Using both (semi-)metrics and kernel functions Ferraty and Vieu (2006) describe a method to estimate  $y^*$  given its functional predictor  $x^*$  and the known response-functional predictor pairs  $(y_i, x_i)$ . Since the functional predictor is known it seems reasonable to use a method based on the conditional distribution of  $y^*$  given  $x^*$ . In fact the method which will be used here and throughout the thesis will be based on the conditional expectation. It is also reasonable to use the conditional median or conditional mode with further details given in Ferraty and Vieu (2006). Define the regression operator  $r$  of  $y^*$  given  $x^*$  as

$$r(x^*) = \mathbb{E}(y|x = x^*) \quad (6.31)$$

Clearly the conditional expectation can be defined as the best estimate of the regression operator and therefore

$$\hat{y}^* = \hat{r}(x^*) \quad (6.32)$$

Therefore the interest is in using nonparametric ideas to calculate a best estimate of the conditional expectation. This can be done by defining the kernel regression estimator as

$$\hat{r}(x^*) = \sum_{i=1}^n \omega_i y_i$$

$$\begin{aligned}
 &= \sum_{i=1}^n \frac{K\left(\frac{d(x^*, x_i)}{h}\right)}{\mathbb{E}\left(K\left(\frac{d(x^*, x_i)}{h}\right)\right)} y_i \\
 &= \sum_{i=1}^n \frac{K\left(\frac{d(x^*, x_i)}{h}\right)}{\frac{1}{n} \sum_{i=1}^n K\left(\frac{d(x^*, x_i)}{h}\right)} y_i \\
 \hat{r}(x^*) &= \frac{\sum_{i=1}^n y_i K(h^{-1}d(x^*, x_i))}{\sum_{i=1}^n K(h^{-1}d(x^*, x_i))} \tag{6.33}
 \end{aligned}$$

where  $K$  is a suitable asymmetric kernel,  $d$  is a suitable (semi-)metric and  $h$  is real and positive. It is relatively simple to show that this is an intuitively sensible estimator. By setting

$$w_i(x^*) = \frac{K(h^{-1}d(x^*, x_i))}{\sum_{i=1}^n K(h^{-1}d(x^*, x_i))}$$

it is possible to rewrite (6.33) as

$$\hat{r}(x^*) = \sum_{i=1}^n w_i(x^*) y_i \tag{6.34}$$

Since  $\sum_{i=1}^n w_i(x^*)$  is clearly 1 this indicates that the best estimate of  $\mathbb{E}(y^* | x = x^*)$  is a weighted average of the known responses with the weight given to  $y_i$  dependent on how close  $x_i$  is to  $x^*$ .

The bandwidth  $h$  has been discussed throughout this section with little mention of its selection. Ferraty and Vieu (2006) outline various methods for selecting the bandwidth when calculating the conditional regression predictor. The method which will be used throughout this thesis involves a fixed choice of the number of nearest neighbours. Specifically this means that to predict each unknown response only  $k$  (say) responses, corresponding to the  $k$  functional predictors closest to the predictor of the unknown response, are given non-zero weight. Recall that there is information on the response predictor pairs  $(y_i, x_i)_{i=1, \dots, n}$ . Suppose that a particular response predictor pair, indexed by  $a$  say, is removed from the data and the remainder of the data is used to predict the response  $y_a$  given  $x_a$ . The bandwidth  $h_{ka}$  is set such that there are  $k$  functional predictors with  $d(x_a, x_i) < h_{ka}$ . The estimated response is then

$$\hat{y}_a = \frac{\sum_{i(-a)} y_i K(h_{ka}^{-1}d(x_a, x_i))}{\sum_{i(-a)} K(h_{ka}^{-1}d(x_a, x_i))}$$

The test of no effect discussed in Section 6.1.3 can be used to analyse the relationship between functional predictor and psychological score. Here the two models can be written as

$$\begin{aligned} H_0 & : \mathbb{E}(y_i) = \mu \\ H_1 & : \mathbb{E}(y_i) = r(x_i(s)) \end{aligned}$$

where  $r$  is a functional regression operator. The residual sum of squares can be defined for each model as

$$\begin{aligned} RSS_0 & = \sum_{i=1}^n [y_i - \bar{y}]^2 \\ RSS_1 & = \sum_{i=1}^n [y_i - \hat{r}(x_i(s))]^2 \end{aligned}$$

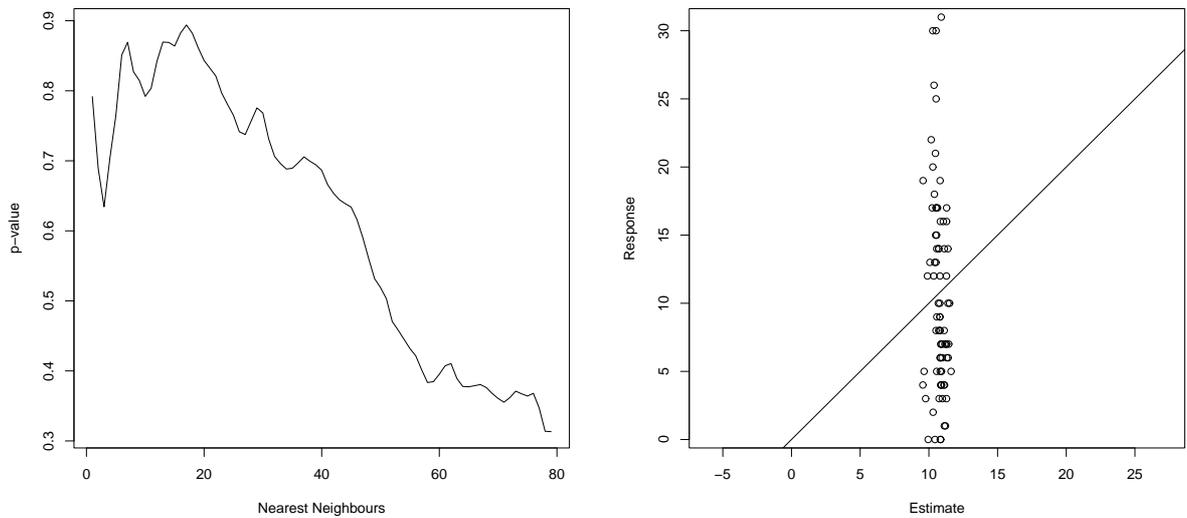
and in quadratic form as

$$\begin{aligned} RSS_0 & = y^T (I - L)^T (I - L) y = y^T (I - L) y \\ RSS_1 & = y^T (I - W)^T (I - W) y \end{aligned}$$

where  $W$  is the  $(n \times n)$  weight matrix which defines the process of the regression operator such that  $\hat{y} = Wy$  where  $\hat{y}_i = \hat{r}(x_i(s))$ . The regression operator here will typically be the operator defined in (6.33) so  $W$  will contain the weights given by the kernel function.

Nonparametric regression is carried out on the cleft data by estimating the mean responses using the neighbouring response-functional predictor pairs. The local number of neighbours method is used to give the bandwidth for each estimate with the measure of closeness calculated using the  $L_2$  metric. Figure 6.8 shows the p-value from the test of no effect for various numbers of nearest neighbours and a plot of the estimated psychological score using nonparametric regression with 60 nearest neighbours against the true psychological score.

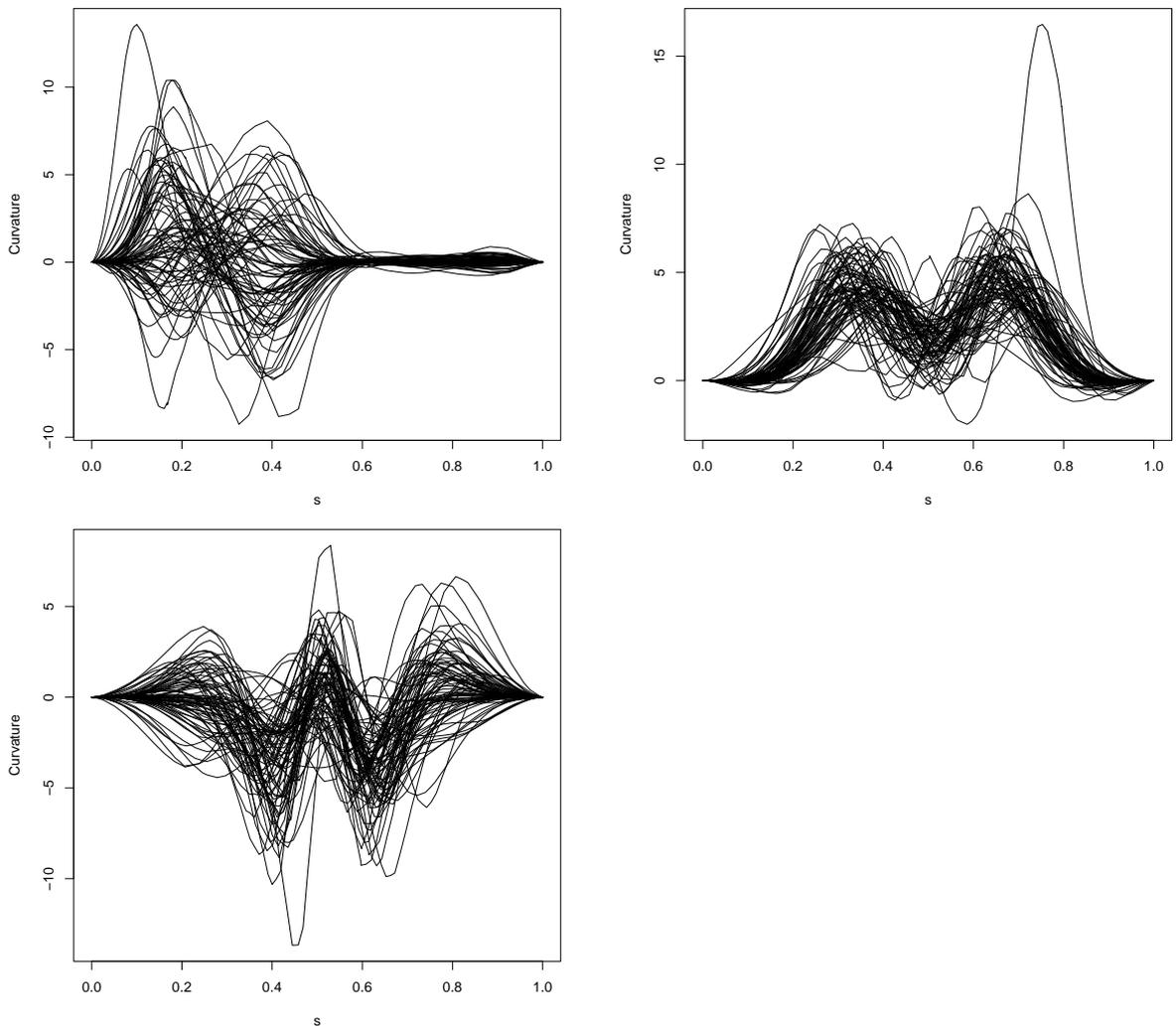
Figure 6.8 shows that, regardless of the number of nearest neighbours used, yz curvature of the midline profile is not a significant predictor of psychological score in a nonparametric regression model. The plot of predicted against true psychological score shows that the yz curvature function has had very little effect on the predicted psychological score.



**Figure 6.8:** Plot of the p-value from the test of no effect of the nonparametric regression model using various numbers of nearest neighbours (left) and estimated psychological score using nonparametric regression with 60 nearest neighbours against the true psychological score with the line of equality for reference (right).

## 6.2 Scalar Response and Multiple Functional Predictors

This section will describe techniques used to investigate the relationship between a scalar response and multiple functional predictors. Many of the techniques will be simple extensions of the techniques in Section 6.1. In general the techniques will be explained for use with an arbitrary number ( $p$  say) of functional predictors. However, the data used to illustrate the techniques will come from the 10 year old cleft children with the psychological score as the response and the yz curvature function of the midline profile, the xy curvature function of the midline profile, the xz curvature function of the upper lip and the xy curvature function of the upper lip as the functional predictors. A boxplot of the psychological scores and a plot of the yz curvature functions was shown in Figure 6.1. Figure 6.9 shows the remaining curvature functions.

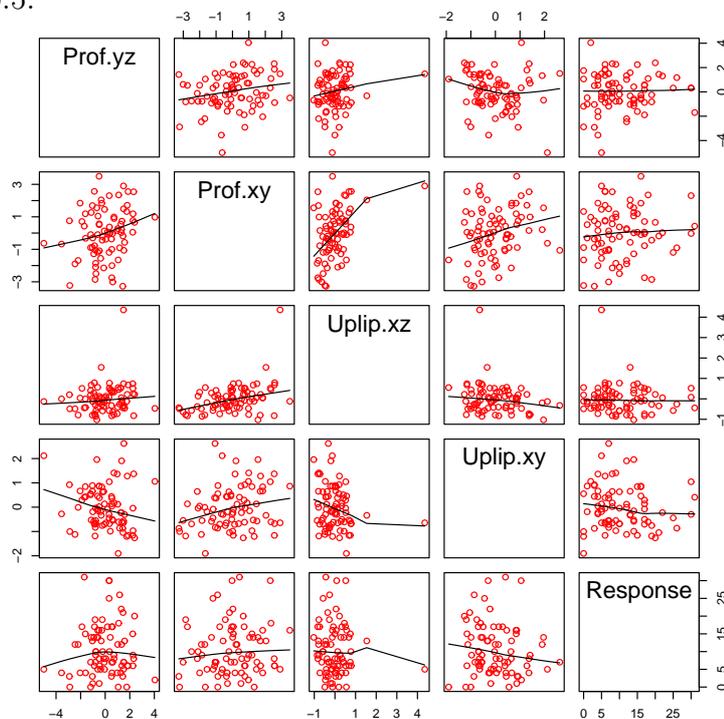


**Figure 6.9:** Curvature functions for the cleft subjects including; xy curvature of the midline profile (top left), xz curvature of the upper (top right) and xy curvature of the upper lip (bottom left).

### 6.2.1 Displaying the data

Displaying the data to subjectively investigate relationships between the functional predictors and response becomes even more complex for multiple predictors. As in Section 6.1.1, to make graphical display simpler, the functional predictors are represented by their component scores from the first principal component. This means that each subject has a response and a principal component score for each of the  $p$  predictors. It is then straightforward to produce a matrix of scatterplots to investigate the relationship between the response and the component score of each functional predictor. Adding a smooth best fitting line to the points is helpful in investigating relationships between the variables.

Figure 6.10 shows a matrix of scatterplots for the cleft data to investigate the relationship between psychological score and the first principal component score for each of the functional predictors. The smooth best fitting line is a lowess smoother (see Cleveland (1979) and Cleveland (1981) for details) with a bandwidth of 0.5.



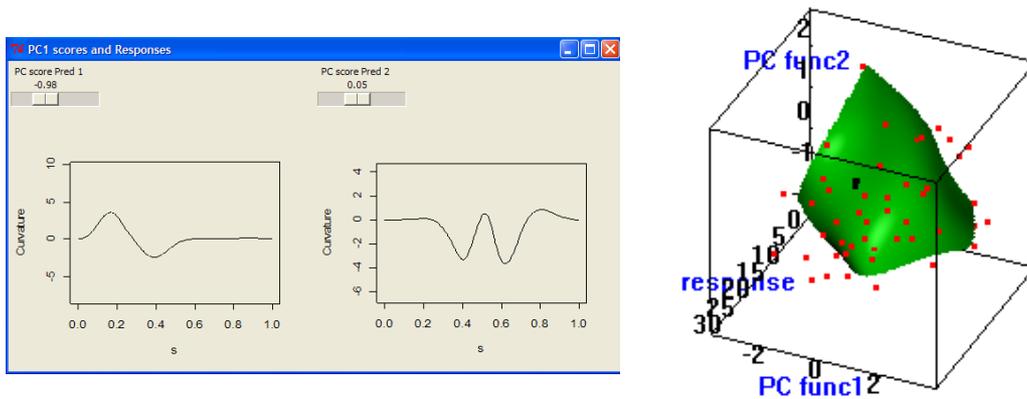
**Figure 6.10:** Matrix of scatterplots for psychological score and first principal component score of the four functional predictors.

In Figure 6.10 it appears that the first principal component scores of xy curvature of the upper lip show most relationship with psychological score. While this does not necessarily indicate that this functional predictor has a significant relationship with the response, it is a useful guide.

It may be that there is interest in interactions between two of the functional predictors and the response. When this is the case it is possible to use nonparametric statistics (as shown by Bowman and Azzalini (1997)) to produce a surface which gives a best fit of the relationship between the response and the first principal component scores for the two functional predictors. For various combinations of component scores, it is interesting to view the best estimate of the response given by these component scores whilst also viewing the affect of each score on the corresponding mean functional predictor. The results can be displayed both easily and neatly in **R** using `rpanel` where two sliders can be set (one for each

functional predictor) to slide along the range of principal component scores.

Figure 6.11 shows the surface which gives the best fit of the relationship between the response and the first principal component scores for xy curvature of the midline profile and xy curvature of the upper lip with a point indicating the position on the surface corresponding to the chosen component scores. Also shown is the function produced by adding the effect of the component score to the mean function for each predictor.



**Figure 6.11:** The `rpanel` and three-dimensional plot with best surface. Function 1 corresponds to xy curvature of the midline profile, function 2 corresponds to xy curvature of the upper lip and response corresponds to psychological score.

Although it is difficult to see the relationship in the static representation in Figure 6.11, by rotating the three-dimensional plot this is a useful tool for investigating the relationship between the combined effect of the component scores of two functional predictors and the response.

As explained in Section 6.1.1 it need not be the component scores from the first principal component which are used to look for relationships in the data. However, since the first principal component explains the largest variability in the data it is sensible to at least start by using the first principal component scores.

## 6.2.2 Regression on principal component scores

Section 6.1.2 introduced both simple linear regression and additive models for investigating the relationship between the response and the  $j$  principal component

scores for a single functional predictor. It is relatively simple to extend these methods to multiple functional predictors.

Suppose that the scalar response is denoted by  $y_i$  and the  $p$  functional predictors are denoted by  $x_{ki}(s_{ki})$ , where  $s \in (0, 1)$  and  $i = 1, \dots, n$ . Further denote the  $j$ th principal component score of the  $k$ th predictor from the  $i$ th subject as  $a_{kij}$ . Then the standard linear regression can be extended from (6.1) so that

$$y = \alpha + \sum_k \sum_j \beta_{kj} a_{kj} + \epsilon \tag{6.35}$$

Finding the parameters  $\alpha$  and  $\beta_{kj}$  which minimise the residual sum of squares gives the best linear relationship between the response and the principal component scores of the various functional predictors.

A simple linear model is fitted with the psychological score as the response and the first two principal component scores for each of the four functional predictors as scalar predictors i.e. there are eight scalar predictors. The p-values found from this full model are shown in Table 6.3.

| Functional Predictor         | Component | Coefficient | p-value |
|------------------------------|-----------|-------------|---------|
| yz curvature midline profile | 1         | 0.167       | 0.768   |
| yz curvature midline profile | 2         | -2.57       | 0.015   |
| xy curvature midline profile | 1         | 1.45        | 0.047   |
| xy curvature midline profile | 2         | -0.281      | 0.804   |
| xz curvature upper lip       | 1         | -2.56       | 0.082   |
| xz curvature upper lip       | 2         | 1.70        | 0.372   |
| xy curvature upper lip       | 1         | -1.23       | 0.273   |
| xy curvature upper lip       | 2         | 1.28        | 0.257   |

**Table 6.3:** Significance of the first two principal component scores of the four curvature functions as combined linear predictors of psychological score.

Only the second principal component score for the yz curvature of the midline profile and first principal component score for the xy curvature of the midline profile show a significant relationship with psychological score when all eight scalar predictors are combined. This gives an indication that curvature of the midline profile may be a better predictor of psychological score than curvature of the upper lip.

As in Section 6.1.2 it may be more informative to replace the scalar parameters

$\beta_{kj}$  with a function. This can be done with an additive model, using component score multiple predictors by simply extending (6.2) so that

$$y = \alpha + \sum_k \sum_j f_{kj}(a_{kj}) + \epsilon \tag{6.36}$$

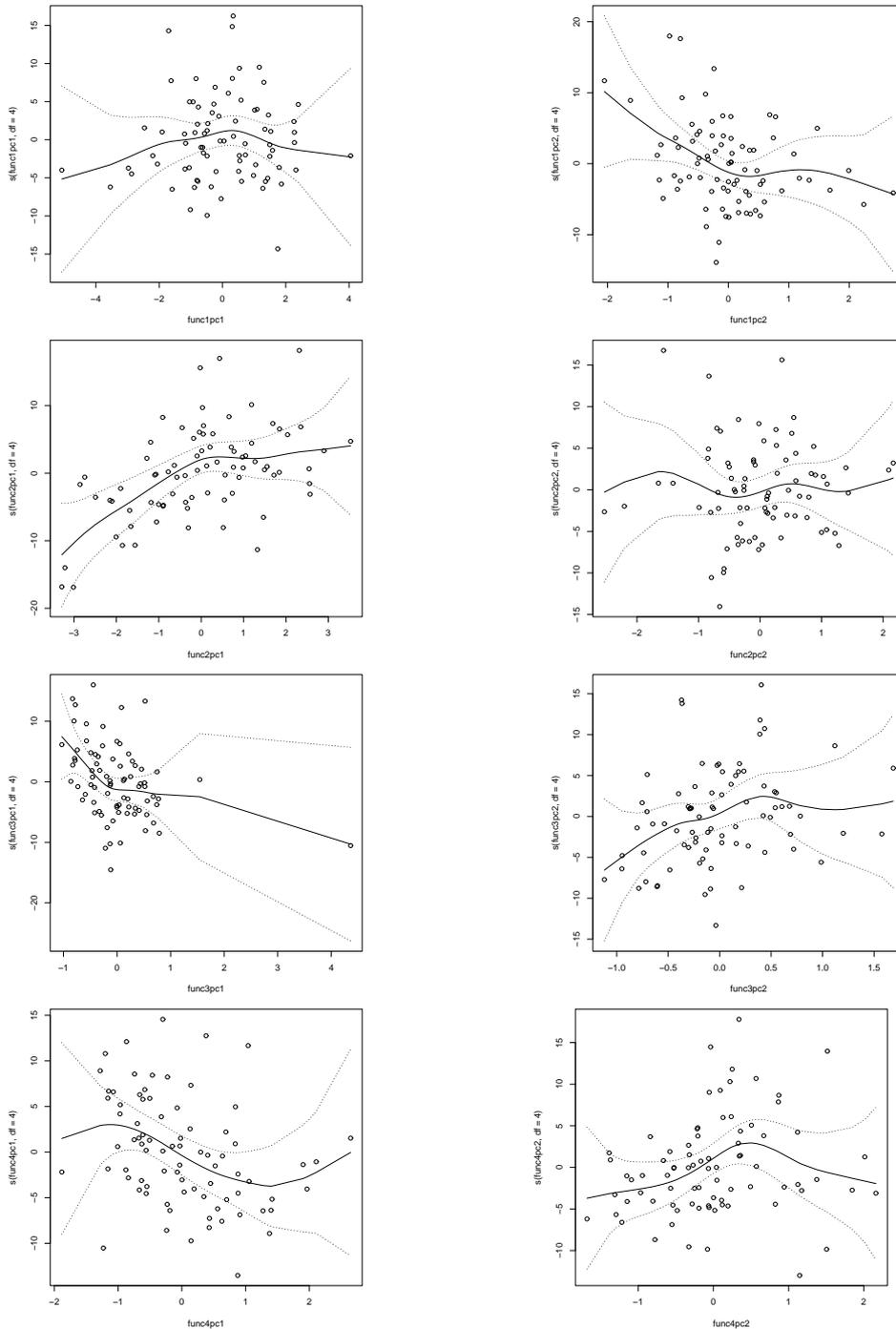
with the backfitting algorithm used to find the  $f_{kj}$ 's.

An additive model was fitted to the cleft data with the same predictors and response as the linear model above. The fitted functions of the predictors can be found in Figure 6.12 while the p-values for the function of each predictor from the full model can be found in Table 6.4. The smooth fitted functions are calculated using smoothing splines with 4 degrees of freedom.

| Functional Predictor         | Component | p-value |
|------------------------------|-----------|---------|
| yz curvature midline profile | 1         | 0.79    |
| yz curvature midline profile | 2         | 0.24    |
| xy curvature midline profile | 1         | 0.04    |
| xy curvature midline profile | 2         | 0.57    |
| xz curvature upper lip       | 1         | 0.18    |
| xz curvature upper lip       | 2         | 0.34    |
| xy curvature upper lip       | 1         | 0.17    |
| xy curvature upper lip       | 2         | 0.25    |

**Table 6.4:** Significance of smooth functions of the first two principal component scores of the four curvature functions as combined predictors of psychological score.

Figure 6.12 seems to suggest that the principal component score of some functions may have an effect on the value given to the prediction of psychological score by the additive model. However, the p-values in Table 6.4 show that when all eight functions of principal component score are used in the model only one predictor, first principal component score of xy curvature of the midline profile, has a significant effect. Clearly by dropping the least significant terms it would be possible to investigate further any relationship between principal component scores and psychological score. However, as this is primarily an explanatory exercise this is not necessary and referring to the plots of the additive model functions is of more interest to assess whether the combined effect of the functions may provide a useful predictor of psychological score.



**Figure 6.12:** Additive model functions for first and second principal component score predictors of;  $yz$  curvature of the midline profile (function 1),  $xy$  curvature of the midline profile (function 2),  $xz$  curvature of the upper lip (function 3) and  $xy$  curvature of the upper lip (function 4).

### 6.2.3 Functional linear model

Section 6.1.3 outlined a technique of using B-spline representation to regularise a functional linear model to allow estimation of a functional parameter  $\beta(s)$ .

This section will extend the technique to allow for multiple ( $p$  say) functional predictors. The functional linear model for multiple predictors is the simple extension of (6.3) such that

$$y = \alpha + \sum_k \int_0^T \beta_k(s)x_k(s)ds + \epsilon \tag{6.37}$$

where  $x_k(s)$  ( $k = 1, \dots, p$ ) is the  $k$ th functional predictor and  $\beta_k(s)$  is the corresponding functional parameter. The errors are once again independent from the predictors with mean zero.

The first stage in estimating the functional parameters is to represent the functional predictors using B-spline basis functions. Since all predictors are defined on the same interval  $s \in (0, 1)$  the same set of basis functions can be used to describe each predictor. Here  $m$  B-spline basis functions will be used and are denoted as  $\phi_1, \dots, \phi_m$ . Each predictor can then be written as the expansion

$$x_{ik}(s) = \sum_{v=1}^m c_{ikv}\phi_v = c_{ik}^T\phi \tag{6.38}$$

where  $x_{ik}(s)$  is the  $k$ th functional predictor for the  $i$ th subject and  $c_{ikv}$  is the coefficient of basis function  $v$  for this predictor-subject combination. The functional parameters are also expressed in terms of these basis functions as

$$\beta_k(s) = \sum_{v=1}^m b_{kv}\phi_v = b_k^T\phi \tag{6.39}$$

where  $b_{kv}$  is the coefficient of the  $v$ th basis functional for the parameter of the  $k$ th predictor. Now the aim is to express  $\sum_k \int_0^T \beta_k(s)x_{ik}(s)ds$  in terms of the spline coefficients. For simplicity, since B-spline basis functions are not orthonormal, recall  $J$  from (6.8) then using an extension of the Parseval identity it can be seen that

$$\begin{aligned} \sum_k \int_0^T \beta_k(s)x_{ik}(s)ds &= \sum_{k=1}^p \left[ \sum_{j=1}^m \sum_{v=1}^m c_{ijk}b_{kv} \int_0^T \phi_j(s)\phi_v(s)ds \right] \\ \sum_k \int_0^T \beta_k(s)x_{ik}(s)ds &= \sum_{k=1}^p \left[ \sum_{j=1}^m \sum_{v=1}^m c_{ijk}J_{jv}b_{kv} \right] \end{aligned} \tag{6.40}$$

It is perhaps once again simpler to think of the process in matrix form. Denote the  $(n \times 1)$  matrix of the responses as  $Y = (y_1, \dots, y_n)^T$ . The matrix of the parameters

is the  $((p \cdot m) + 1) \times 1$  matrix  $\zeta = (\alpha, b_{11}, b_{12}, \dots, b_{1m}, b_{21}, \dots, b_{2m}, \dots, b_{pm})^T$ . The coefficient matrix of the predictors is the  $(n \times [(p \cdot m) + 1])$  matrix  $Z = [\mathbf{1} \ C_1 J \ C_2 J \ \dots \ C_p J]$  where  $\mathbf{1}$  is a column of  $n$  1s and  $C_k$  is the  $(n \times m)$  coefficient matrix for predictor  $k$  with row  $i$  containing the B-spline coefficients for subject  $i$ . Now it is clear that the model can be defined by the equation

$$Y = Z\zeta + \epsilon \tag{6.41}$$

where  $\epsilon = (\epsilon_1, \dots, \epsilon_n)$  and the  $\epsilon_i$ s have mean zero and are independent from the predictors. It now follows that

$$\hat{Y} = Z\hat{\zeta} \tag{6.42}$$

Therefore the least squares solution of the parameter vector with multiple functional predictors  $\zeta$  is given by

$$\begin{aligned} Z^T Z \hat{\zeta} &= Z^T Y \\ \hat{\zeta} &= (Z^T Z)^{-1} Z^T Y \end{aligned} \tag{6.43}$$

Section 6.1.3 described the advantages of penalising the sum of squares error using a roughness penalty to force a degree of smoothness on the functional parameter. This can be extended for use with multiple functional predictors by extending the penalised sum of square errors shown by (6.13) to

$$PENSSE = \sum_{i=1}^n \left[ y_i - \left( \alpha + \sum_{k=1}^p \int_0^T \beta_k(s) x_{ik}(s) \right) \right]^2 + \sum_{k=1}^p \lambda_k \int_0^T \beta_k''(s)^2 ds \tag{6.44}$$

Note that each functional parameter can be assigned its own smoothness parameter meaning that the parameters for some predictors can be forced to be smoother than those for others. However, unless there is a good reason for this it seems sensible to have a common smoothness parameter i.e.  $\lambda = \lambda_1, \dots, \lambda_p$ .

To write this penalised sum of square errors for the multiple predictor model in vector-matrix form, recall the matrix  $K$  defined in (6.14). Now the penalised sum of square errors can be defined as

$$\begin{aligned} PENSSE &= \sum_{i=1}^n (y_i - \alpha - \sum_{k=1}^p \sum_{v=1}^m c_{ivk} b_{vk} \phi_v)^2 + \sum_{k=1}^p \lambda_k \int (\sum_{v=1}^m b_{vk} \phi_v''(s))^2 ds \\ PENSSE &= \| y - \alpha - \sum_{k=1}^p C_k b_k \|^2 + \sum_{k=1}^p \lambda_k b_k^T K b_k \end{aligned} \tag{6.45}$$

Again define the  $(n \times 1)$  matrix of the responses as  $Y = (y_1, \dots, y_n)^T$ , the  $((p \cdot m) + 1 \times 1)$  matrix of parameters as  $\zeta = (\alpha, b_{11}, b_{12}, \dots, b_{1m}, b_{21}, \dots, b_{2m}, \dots, b_{pm})^T$  and the  $(n \times (p \cdot m) + 1)$  coefficient matrix as  $Z = [\mathbf{1} \ C_1 J \ C_2 J \ \dots \ C_p J]$ . Further recall from (6.16) the augmented penalty matrix  $K_0$ . It is now possible to simplify (6.45) such that

$$PENSSE = \|y - Z\zeta\|^2 + \sum_{k=1}^p \lambda_k \zeta^T K_0 \zeta \quad (6.46)$$

and the minimising  $\hat{\zeta}$  is given by

$$\begin{aligned} (Z^T Z + \sum_{k=1}^p \lambda_k K_0) \hat{\zeta} &= Z^T y \\ \hat{\zeta} &= (Z^T Z + \sum_{k=1}^p \lambda_k K_0)^{-1} Z^T y \end{aligned} \quad (6.47)$$

Section 6.1.3 outlined a cross-validation procedure for calculating the ‘best’ smoothing parameter. This can be extended to the multiple functional prediction setting by defining the hat matrix  $S$  which is required to calculate the CV score using (6.21) as

$$S = Z(Z^T Z + \sum_{k=1}^p \lambda_k K_0)^{-1} Z^T \quad (6.48)$$

This calculates the CV score for the smoothing vector  $\lambda = (\lambda_1, \dots, \lambda_p)$ . Clearly there are a huge number of permutations of the  $\lambda_k$ 's so this cross-validation method is only of practical use if the smoothing parameters are forced to be the same for each functional parameter i.e.  $\lambda = \lambda_1 = \lambda_2 = \dots = \lambda_p$ .

Section 6.1.3 outlined a pseudo-likelihood ratio test of no effect for a functional predictor. Here a natural extension of this test is described to examine whether there is significant effect of the functional predictor  $r$  in addition to the other functional predictors. The two models for this test can be written as

$$\begin{aligned} H_0 &: y = \alpha + \sum_{k \neq r} \int_0^T \beta_k(s) x_k(s) ds + \epsilon \\ H_1 &: y = \alpha + \sum_k \int_0^T \beta_k(s) x_k(s) ds + \epsilon \end{aligned}$$

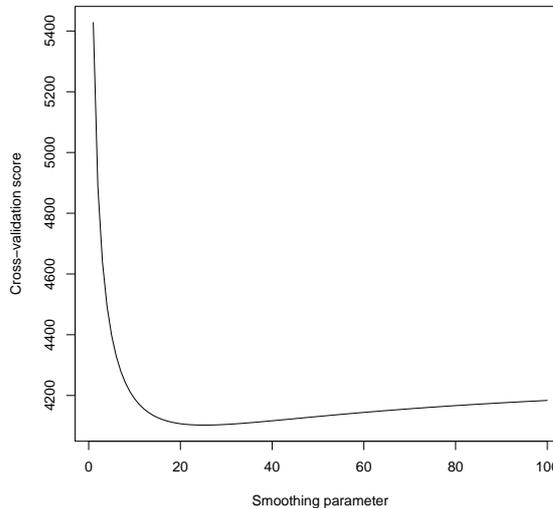
Now call  $W_{-k}$  the weight function given by the hat matrix,  $S$ , from the model

$H_0$  and  $W_f$  the weight function given by the hat matrix from the full model  $H_1$ . Then the residual sum of squares for each model can be written as

$$\begin{aligned} RSS_0 &= y^T(I - W_{-k})^T(I - W_{-k})y \\ RSS_1 &= y^T(I - W_f)^T(I - W_f)y \end{aligned}$$

The pseudo-likelihood ratio test statistic of (6.23) can then be expressed in the form of (6.24) where  $A = (I - W_f)^T(I - W_f)$  and  $B = (I - W_{-k})^T(I - W_{-k}) - A$ . The remainder of the test follows the same process as the test in Section 6.1.4. The only difference between the tests is replacing the  $(n \times n)$  matrix filled with  $1/n$  by the weight matrix from the model  $H_0$  which changes the test from a test of no effect to a comparison of models where  $H_0$  is nested within  $H_1$ .

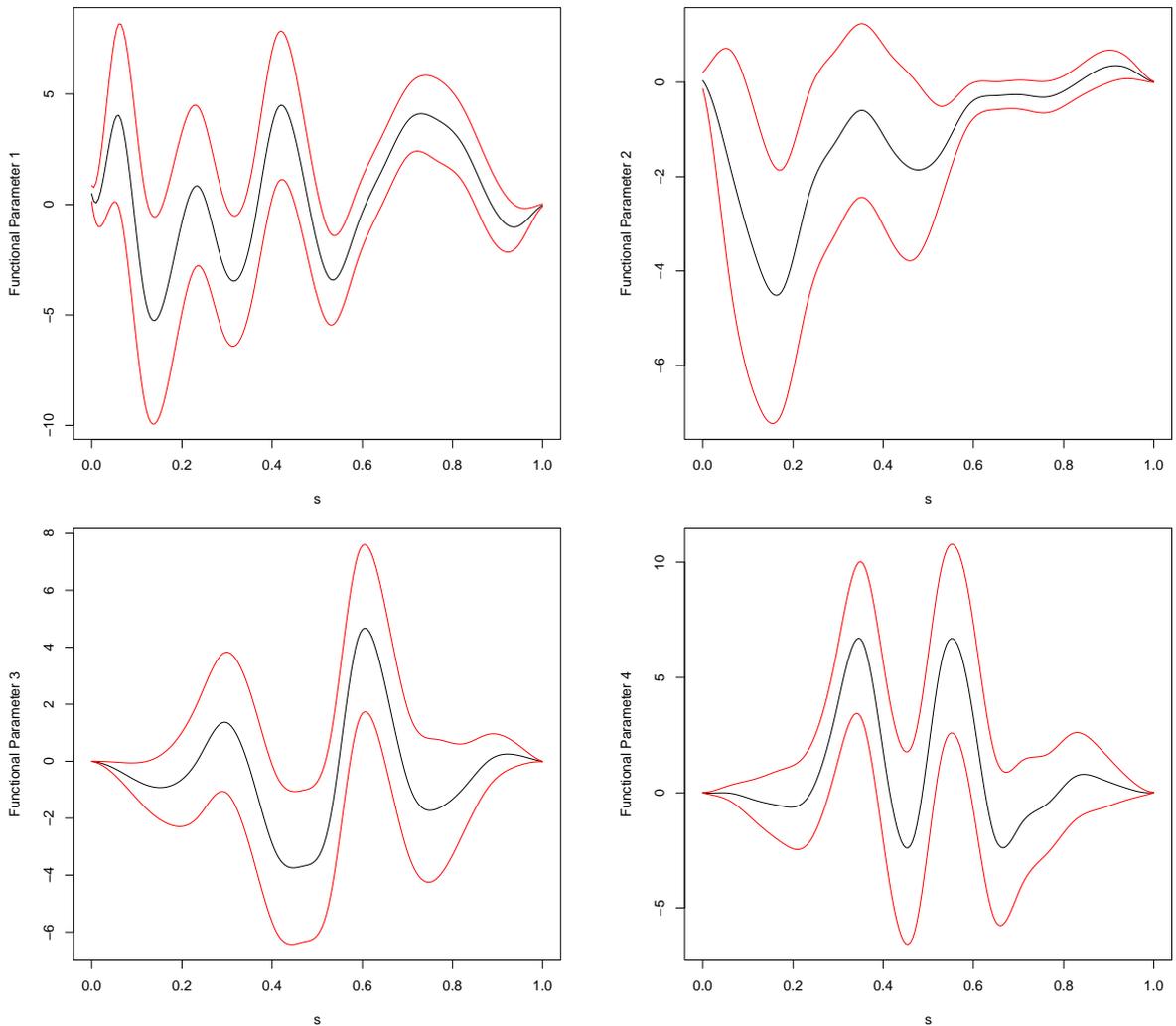
A functional linear model is calculated on the cleft data. The functional predictors and the functional parameters are defined using B-splines with 20 basis functions. To select a smoothing parameter a grid search is carried out on numerous potential smoothing parameters to select the value which minimises the cross-validation score defined in (6.21). Note that this method is limited to selecting the same smoothing parameter for each functional parameter. Figure 6.13 shows the cross-validation function.



**Figure 6.13:** Cross-validation score function for the multiple predictor functional linear model.

The grid search across numerous smoothing parameters showed that the value which minimised the cross-validation function was close to 25 so  $\lambda$  is set to 25 for all four functional predictors. Figure 6.14 now shows the functional parameter

for each of the functional predictors in the model.

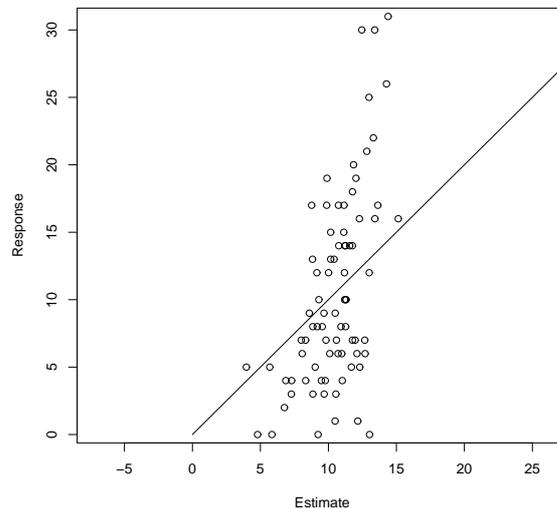


**Figure 6.14:** Functional parameters for the functional linear model with psychological score as response and; yz curvature function of the midline profile (top left), xy curvature function of the midline profile (top right), xz curvature function of the upper lip (bottom left), xy curvature function of the upper lip (bottom right) as predictors. The red lines indicate confidence limits for the parameter.

The interpretation of the functional parameters shown in Figure 6.14 is not entirely straightforward although it can be very useful in defining a relationship between the functional predictors and the psychological score. The curvature functions must be consulted during the interpretation. The parameter for yz curvature of the midline profile is similar to that of the parameter in the single predictor case so interpretation is similar. Looking at the xy curvature function

of the midline profile, the functional parameter is negative around  $s = 0.2$ . This corresponds to bending at the tip of the nose and it appears that high magnitude of negative curvature (bending to the right) indicates higher psychological scores. Psychological and anatomical reasons for this are unclear. The functional parameter for the xz curvature of the upper lip is negative at about  $s = 0.45$  and positive at about  $s = 0.6$ . The point at  $s = 0.6$  is the end of the Cupid's bow where the upper lip turns back towards the face with positive curvature. Therefore it seems that the larger the bending towards the face at the end of the Cupid's bow the larger the psychological score. The negative value at  $s = 0.45$  is difficult to interpret but it may be influenced by a large spread of positions for the start of the Cupid's bow from  $s = 0.35$  to  $s = 0.45$ . The functional parameter for xy curvature of the upper lip is positive at  $s = 0.35$  and  $s = 0.55$ . These are roughly the ends of the Cupid's bow where the upper lip turns down the face with negative curvature. This suggests that the higher the magnitude of curvature at these points the lower the psychological score. This may be due to subjects with more pronounced Cupid's bows looking more like control subjects and therefore having lower psychological scores.

Figure 6.15 shows the psychological score estimated using the functional linear model with the four functional predictors against the true psychological score.



**Figure 6.15:** Estimated psychological score using the functional linear model with multiple functional predictors against the true psychological score with the line of equality for reference.

There does appear to be a positive relationship between the true psychological score and the estimated score using the full functional linear model. This suggests

that there is a relationship between the combined effect of the four functional predictors and the psychological score. To test the individual significance of each of the functional predictors in addition to the other predictors a set of tests comparing the full model to the model with each predictor removed are carried out and the p-values are reported in Table 6.5.

| Functional Predictor         | p-value |
|------------------------------|---------|
| yz curvature midline profile | 0.169   |
| xy curvature midline profile | 0.140   |
| xz curvature nasal bridge    | 0.371   |
| xy curvature nasal bridge    | 0.122   |

**Table 6.5:** Significance of each curvature function as a predictor of psychological score in addition to the other curvature functions.

Table 6.5 shows that none of the curvature functions are useful predictors of psychological score in addition to the other predictors in the model. It would be desirable to reduce the model until all predictors are significant. However, as this example is simply for illustration model reduction will not be covered here. It is interesting to note that the  $R^2$  value of the model is 0.251.

### 6.2.4 Nonparametric functional regression

Section 6.1.4 outlined a method for predicting the scalar response for a subject with a given functional predictor by a weighted average of known responses. The weight given to each known response was dependent on the ‘closeness’ between the functional predictor for the known response and the functional predictor of the unknown response. This section will extend this method to the case where there are multiple ( $p$  say) functional predictors for each response.

To motivate kernel weighting of the functional predictors Section 6.1.4 briefly introduced kernel weighting of multivariate data. Equation (6.28) showed that the multivariate kernel can be thought of as the product of the kernels in each dimension whilst (6.29) showed how each multivariate observation can be weighted using the multivariate kernel. Now suppose that instead of the multivariate observations being a collection of scalar values it is in fact a collection of functions  $x_{ik}(s)$  where  $i = 1, \dots, n$  denotes the subject and  $k = 1, \dots, p$  denotes the functions. Now

$$K^*(\mathbf{d}_i) = K(d_{i1}) \times \dots \times K(d_{ip}) \tag{6.49}$$

where  $K(d_{ik})$  is the functional kernel weighting calculated using a semi-metric  $d$  and an asymmetric kernel as in Section 6.1.4. Now the weight given to subject  $i$  can be taken as a simple extension of (6.30), which gives weight to single variate functional kernels, so that the weight given to the multivariate functional variable  $\mathbf{x}_i$  according to its closeness to  $\mathbf{x}^*$  is given by

$$\omega_i = \frac{K^* \left( \frac{d(\mathbf{x}^*, \mathbf{x}_i)}{h} \right)}{\mathbb{E} \left( K^* \left( \frac{d(\mathbf{x}^*, \mathbf{x}_i)}{h} \right) \right)} \quad (6.50)$$

where

$$K^* \left( \frac{d(\mathbf{x}^*, \mathbf{x}_i)}{h} \right) = K \left( \frac{d(x_1^*, x_{i1})}{h} \right) \times K \left( \frac{d(x_2^*, x_{i2})}{h} \right) \times \dots \times K \left( \frac{d(x_p^*, x_{ip})}{h} \right)$$

The normalisation in (6.50) is carried out using the probability distribution of the multivariate functional random variable. To remove the effect of the predictors being on different scales, in particular to avoid functions with large measures of closeness dominating the process, the values returned by the (semi-)metric are normalised before the kernel weighting such that

$$d(x_1^*, x_{i1}) = \frac{d_{orig}(x_1^*, x_{i1})}{\max_n(d_{orig}(x_1^*, x_{n1}))}$$

Therefore these adjusted (semi-)metric values lie between 0 and 1.

Now suppose that the aim is to find an estimate of the unknown scalar response  $y^*$  which corresponds to the multivariate functional variable  $\mathbf{x}^*$ , where  $\mathbf{x}^* = (x_1^*(s), \dots, x_p^*(s))$ , using the information from  $n$  known response-multivariate functional variable combinations, where  $y_i$  denotes the response for subject  $i$  whilst  $\mathbf{x}_i = (x_{i1}(s), \dots, x_{ip}(s))$  denotes the multivariate functional variable for subject  $i$ . Using (6.33) the expected value of  $y^*$  given  $\mathbf{x}^*$  can be estimated by the regression operator such that

$$\begin{aligned} \hat{y}^* &= \hat{r}(\mathbf{x}^*) \\ \hat{y}^* &= \frac{\sum_{i=1}^n y_i K^*(h^{-1}d(\mathbf{x}^*, \mathbf{x}_i))}{\sum_{i=1}^n K^*(h^{-1}d(\mathbf{x}^*, \mathbf{x}_i))} \end{aligned} \quad (6.51)$$

Clearly the choice of bandwidth is again extremely important. Section 6.1.4 described a method where the number of nearest neighbours (i.e. response-predictor pairs given non-zero weight) is pre-specified and the bandwidth is set

accordingly. In this multivariate setting, if the response for subject  $a$  is to be estimated using the remaining  $n - 1$  response-functional predictor pairs, for the weight  $\omega_i$  for subject  $i$  to be non-zero  $d(x_{ak}, x_{ik}) < h, \forall k, \text{ where } k = 1, \dots, p$ . Therefore the order of neighbours from closest to  $\mathbf{x}_a$  to furthest from  $\mathbf{x}_a$  is given by the order of the maximum  $d(x_{ak}, x_{ik})$  for each subject. The bandwidth  $h_{ja}$  which would allow the chosen number of nearest neighbours ( $j$  say) to be given non-zero weight is set as the bandwidth. The estimated response is then

$$\hat{y}_a = \frac{\sum_{i=1}^{n-1} y_i K_a(h_{ja}^{-1}d(\mathbf{x}_a, \mathbf{x}_i))}{\sum_{i=1}^{n-1} K_a(h_{ja}^{-1}d(\mathbf{x}_a, \mathbf{x}_i))}$$

where  $h_{ja}$  is the bandwidth which gives  $j$  nearest neighbours non-zero weight.

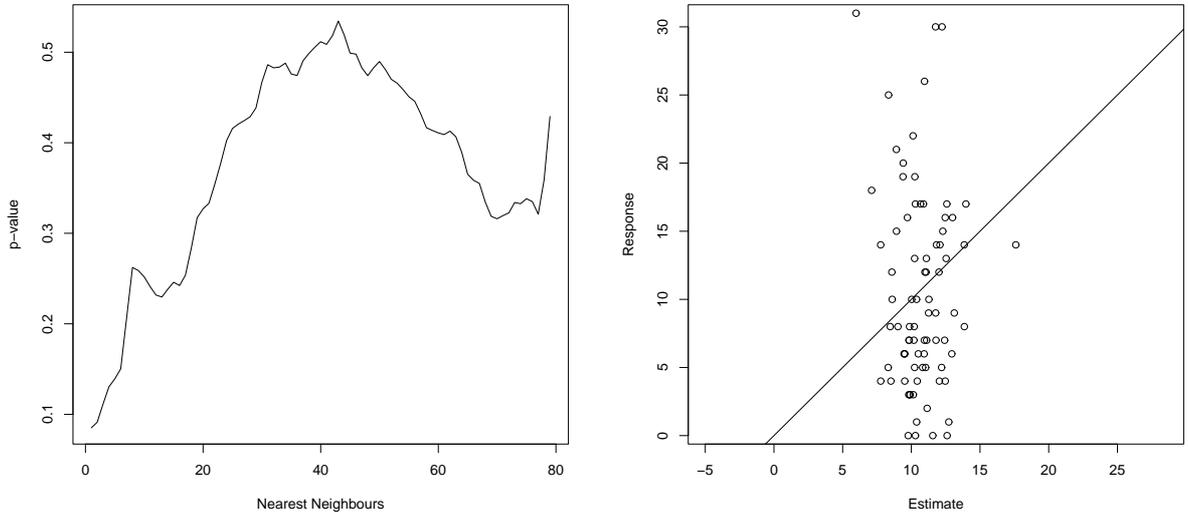
Section 6.2.3 outlined the extension of the pseudo-likelihood ratio test of no effect for multiple functional predictors. This can be used to test the effect of each predictor in the nonparametric regression. The two models for this test can be written as

$$\begin{aligned} H_0 & : \mathbb{E}(y_i) = r_{-k}(x_i(s)) \\ H_1 & : \mathbb{E}(y_i) = r_f(x_i(s)) \end{aligned}$$

where  $r_{-k}$  is the regression operator calculated using all functional predictors except predictor  $k$  and  $r_f$  is the regression operator calculated using all functional predictors. Now  $W_{-k}$  is the weight function such that  $\hat{r}_{-k}(x(s)) = W_{-k}y$  and  $W_f$  the weight function such that  $\hat{r}_f(x(s)) = W_f y$ . The weights are returned by the kernel function when calculating the regression operator under each model. The method is then the same as in Section 6.2.3. It is also trivial to carry out a test of no effect of the full non-parametric regression model. The test is the same as outlined in Section 6.1.4 with the weight matrix  $W$  the weights returned by the kernel function when calculating the regression operator under the non-parametric regression model being tested.

Nonparametric regression is carried out on the cleft data by estimating the known psychological scores using the other response-multivariate functional predictor pairs with the measure of closeness calculated using the  $L_2$  metric. The local number of neighbours method is used to set the bandwidths. Figure 6.16 shows the p-value from the test of no effect for the full nonparametric regression model for various numbers of nearest neighbours and a plot of the estimated psychological score using nonparametric regression with 20 nearest neighbours

against the true psychological score.



**Figure 6.16:** Plot of the p-value from the test of no effect of the nonparametric regression model with four functional predictors using various numbers of nearest neighbours (left) and estimated psychological score using nonparametric regression with 20 nearest neighbours against the true psychological score (right).

Figure 6.16 shows that regardless of the number of nearest neighbours chosen a nonparametric regression with functional predictors midline profile yz curvature, midline profile xy curvature, upper lip xz curvature and upper lip xy curvature, does not provide a significant predictor of psychological score. The plot of estimated against true psychological score shows that the nonparametric regression model does not appear to be useful in estimating the true psychological score. It would be possible to test the effect of each predictor and remove the least significant predictor until a significant model is found but since this example is for illustration this is not necessary here.

### 6.2.5 Functional additive model

Section 6.2.2 described how the relationship between the response and the functional predictors could be investigated using the principal component scores of the functional predictors. Additive models gave a model which best described the relationship between the response and the principal component scores using a function of each of the component scores. Hastie and Tibshirani (1990) give a good overview of generalised additive models. In this section a variation of the standard generalised additive model will be suggested where the standard scalar

predictors are replaced by functional predictors.

Recall that the equation of a generalised linear model is

$$y = \alpha + \sum_j f_j(x_j) + \epsilon$$

where  $x_j$  is a scalar predictor. The proposed functional additive model will have the form

$$y = \alpha + \sum_j f_j(x_j(s)) + \epsilon \tag{6.52}$$

where the errors have mean zero and are independent from the predictors. The basic aim of the functional additive model is to find a set of functions (the  $f_j$ 's) which, when the functional predictors are passed to them, returns the best estimates of the mean of  $y$ . One of the major difficulties here is the problem of visualising the process, in particular visualisation of the  $f_j$ 's is difficult. This is in contrast with standard additive models. The major difference is that scalar predictors are simple to order and only differ over one-dimension whereas functional predictors do not have a natural ordering and can differ in many ways. However, it is still possible to define the  $f_j$ 's for the functional generalised linear model using the back-fitting algorithm.

Suppose that a response  $y_i$  and  $p$  functional predictors  $x_{ij}(s)$  (where  $j = 1, \dots, p$ ) are available for  $n$  subjects. Then the back-fitting algorithm will proceed as follows:

1. Initialise the system so that  $\alpha = \bar{y}_i$  and set the initial functions to zero i.e.  $f_j = 0$  therefore  $y = \alpha$ . So before the system starts the best estimate of each response is simply the mean of all the responses.
2. Cycling  $k = 1, \dots, p$ . Calculate the partial residual for each subject where the partial residual is the error remaining after removing the effects of  $\alpha$  and  $\sum_{j \neq k} f_j(x_{ij}(s))$  from the known response i.e.

$$PR_{ik} = y_i - \alpha - \sum_{j \neq k} f_j(x_{ij}(s))$$

Now a fit of the partial residuals  $PR_{ik}$  against the functional predictors  $x_{ik}(s)$  must be produced. This is a similar problem as carrying out a regression on a scalar response using a functional predictor so can be found using a nonparametric functional regression (Section 6.1.4) with the scalar

response replaced by the partial residuals. The estimated responses given are the values returned by  $f_k(x_k(s))$  for each subject.

3. Repeat (2) until the  $f_j$ 's do not change.

The bandwidths in the nonparametric regression are chosen using the number of nearest neighbours method.

As mentioned one difficulty with interpreting the results from this additive model is how best to illustrate the functions of the functional predictors. This is due to the fact that functional predictors have no natural ordering. It may be interesting however to plot the first principal component score for each functional predictor against the value returned by the additive model function for that predictor. This is a relatively exploratory procedure; however it may show some interesting relationships between component score and value of the additive model function.

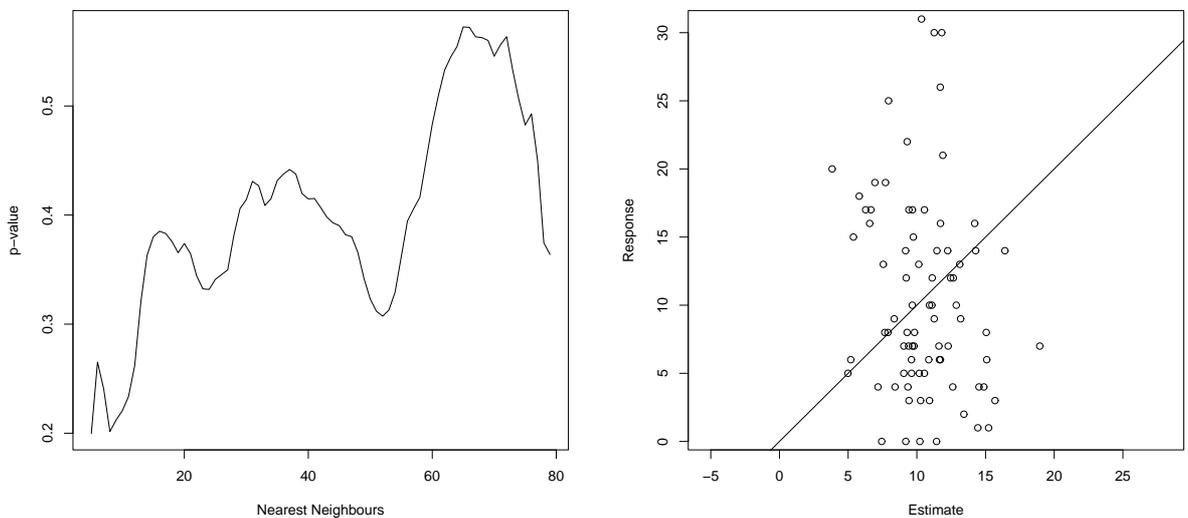
It is possible to test the significance of each functional predictor, in addition to the other predictors, in showing a relationship with the response. A pseudo-likelihood ratio test as outlined by Section 6.2.3 is an appropriate test. The hypotheses can be written as

$$\begin{aligned}
 H_0 & : y = \alpha + \sum_{j \neq k} f_j(x_j(s)) + \epsilon \\
 H_1 & : y = \alpha + \sum_j f_j(x_j(s)) + \epsilon
 \end{aligned}$$

The only difficulty in carrying out the test is in finding the weight matrices  $W_{-k}$  and  $W_f$ . It is possible to keep track of these matrices (called projection matrices as they project the responses to their fitted values) throughout the back-fitting algorithm. Denote the projection matrix which produces the estimate of the partial residual  $PR_k$  as  $P_k$  so that  $\hat{PR}_k = P_k y$ . It is then possible to describe the back-fitting algorithm as the process which iteratively finds  $P_k = (I_n - P_0)S_k(I_n - \sum_{j \neq k} P_j)$  cycling through  $k = 1, \dots, p, 1, \dots, p, 1, \dots$ .  $S_k$  is the smoothing matrix (which can be found from the nonparametric regression) such that  $\hat{PR}_k = S_k PR_k$ ,  $P_0$  is an  $(n \times n)$  matrix filled with  $\frac{1}{n}$  and  $I_n$  is the order  $n$  identity matrix. After the process converges  $P_1, \dots, P_p$  give the weight matrices such that  $\hat{PR}_k = P_k y$  and the fitted values given by the generalised additive model are given by  $\hat{y} = P y$  where  $P = \sum_{j=1}^p P_j$ . The weight matrix  $W_{-k}$  is the projection matrix for an additive model calculated without functional predictor  $k$  whilst the weight matrix  $W_f$  is the projection matrix for the full additive model.

With these weight matrices the pseudo-likelihood ratio test can be carried out as in Section 6.2.4. It is also possible to carry out a test of no effect of the full additive model using the test outlined in Section 6.1.4 with the weight matrix  $W$  given by the projection matrix  $P$  from the full additive model being tested.

An additive model is fitted to the cleft data using the back-fitting algorithm. To examine the effect of the choice of the number of nearest neighbours Figure 6.17 shows p-values returned by the test of no effect for various numbers of nearest neighbours and also shows a plot of predicted psychological score using a generalised additive model with 25 nearest neighbours.



**Figure 6.17:** Plot of the p-value from the test of no effect of the additive model with four functional predictors using various numbers of nearest neighbours (left) and estimated psychological score using a generalised additive model with 25 nearest neighbours against the true psychological score (right).

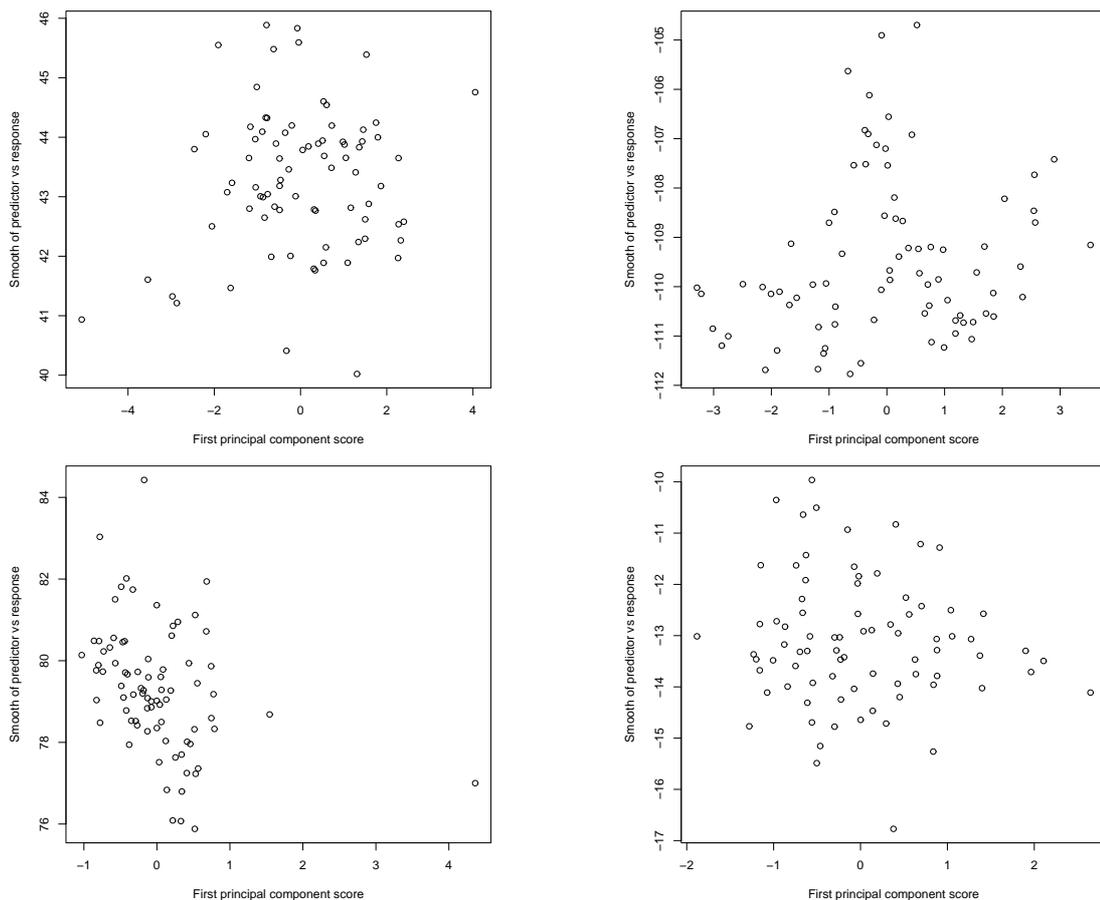
Figure 6.17 shows that regardless of the number of neighbours chosen the additive model is not a significant predictor of psychological score. The plot of estimated against true psychological score shows that the nonparametric regression model does not appear to be related to the mean psychological score. Table 6.6 shows the results of pseudo-likelihood ratio tests of no effect to examine the significance of each functional predictor in addition to the other predictors in an additive model with 25 nearest neighbours used in the modelling.

None of the p-values in Table 6.6 are close to significant. It may be that by removing the least significant functional predictor the model would become significant. Since this section is for illustrating methods this will not be done here. It may however be interesting, especially for a significant model, to attempt to

| Functional Predictor         | p-value |
|------------------------------|---------|
| yz curvature midline profile | 0.618   |
| xy curvature midline profile | 0.136   |
| xz curvature upper lip       | 0.246   |
| xy curvature upper lip       | 0.499   |

**Table 6.6:** Significance of each function as a predictor of psychological score in addition to the other functions in a functional additive model.

interpret the function of the predictor returned by the additive model. To do this a plot of the first principal component score for each functional predictor against the value returned by the additive model function for that predictor is shown in Figure 6.18.



**Figure 6.18:** Plot of first principal component score against value returned by the additive model for predictors; yz (top left) and xy (top right) curvature of the midline profile and xz (bottom left) and xy (bottom right) curvature of the upper lip.

From Figure 6.18 it appears that there may be some form of negative relationship between the first principal component scores of xz curvature of the upper lip and the value returned by the additive model for that predictor. By investigating the effect of increasing the principal component it may be possible to draw some inferences about any small effect of xz curvature of the upper lip on psychological score.

## Chapter 7

# Applied Functional Regression

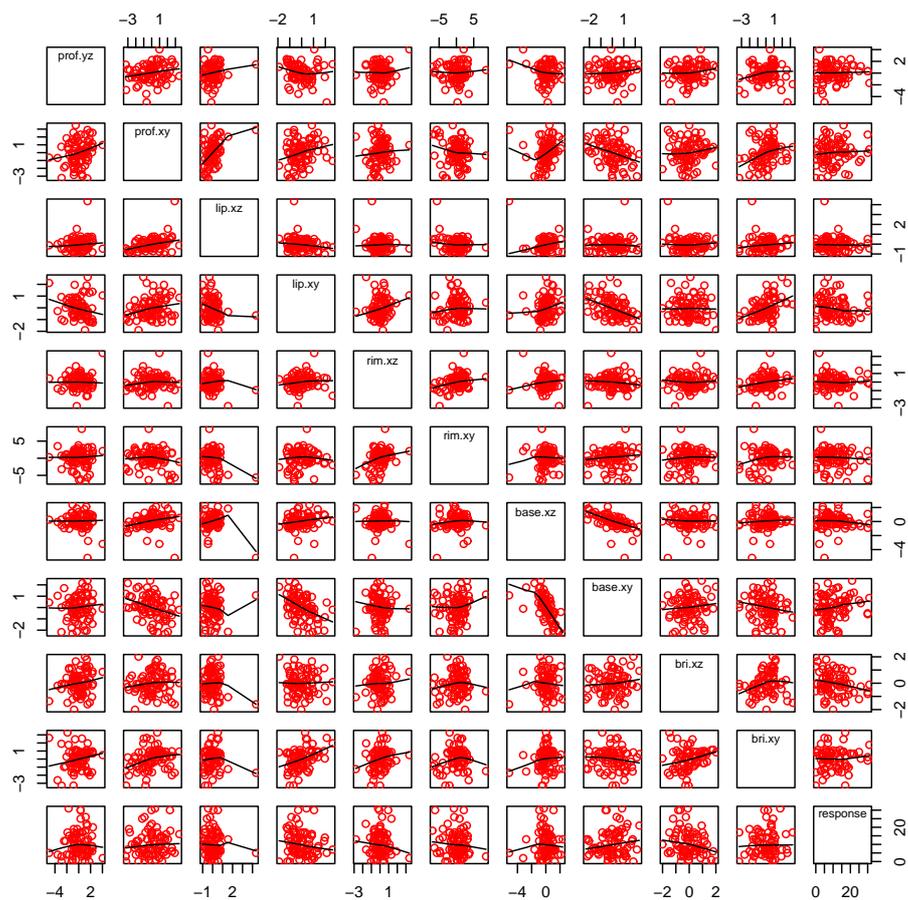
Chapter 6 outlined various techniques to investigate the relationship between a scalar response and functional predictors. Using data from the study on ten year old cleft children these techniques can be used to systematically analyse potential relationships between psychological score and the shape of facial features. There are five three-dimensional curves defining facial features for which data are available. These curves run down the midline profile, along the top of the upper lip, along the rim of the nose, along the base of the nose and along the bridge of the nose. The shape of each curve will be defined by two curvature functions calculated using the perpendicular plane method outlined in Section 4.2.4. Therefore there are ten potential functional predictors which may be related to the psychological score. The name of the curvature ( $yz$ ,  $xy$  or  $xz$ ) is dependent on which planes the curve is projected on to. For reference Table 7.1 shows which curves are explained by each curvature and an anatomical description of what that curvature shows.

To give a brief insight into potential relationships both between the predictors and the response, and between predictors, a matrix scatterplot using the first principal component score of each functional predictor and the scalar response is produced as described in Section 6.2.1 and plotted in Figure 7.1.

Although some of the scatterplots on the bottom row of Figure 7.1 appear to show some relationship between the functional predictor (defined by the first principal component) and the response none appear strong enough to confidently expect a significant relationship. There also appears to be some relationship between functional predictors so there is potential for concurvity, the nonparametric analogue of multicollinearity, in the modelling. This will not be discussed throughout the chapter but is worth noting.

| Predictor       | yz Curvature                     | xy Curvature                 | xz Curvature                     |
|-----------------|----------------------------------|------------------------------|----------------------------------|
| Midline Profile | Bending into and out of the face | Bending across the face      | N/A                              |
| Upper Lip       | N/A                              | Bending up and down the face | Bending into and out of the face |
| Nasal Rim       | N/A                              | Bending up and down the face | Bending into and out of the face |
| Nasal Base      | N/A                              | Bending up and down the face | Bending into and out of the face |
| Nasal Bridge    | N/A                              | Bending up and down the face | Bending into and out of the face |

**Table 7.1:** Description of the two curvature functions which describe each facial curve.



**Figure 7.1:** Matrix plot of first principal component score for all functional predictors and scalar response.

Section 7.1 will use the techniques of Section 6.1 to investigate evidence of a significant relationship between the psychological score and each curvature function and describe the nature of this relationship. Section 7.2 will use the techniques of Section 6.2 to find the combination of curvature functions which give the strongest relationship with the psychological score and give some insight into this relationship.

## 7.1 Single Functional Predictor Analysis

Although none of the plots in Figure 7.1 suggested strong relationships between the first principal component scores of the curvature functions and psychological score, it may be that relationships are simply not apparent when only considering component scores as a predictor as opposed to whole functions. This section will use the ideas of Section 6.1 to look for significant relationships and investigate these relationships further.

### 7.1.1 Functional linear model and investigating significant relationships

Since there are ten potential functional predictors it seems reasonable to use functional regression techniques to find the curvature function which is the strongest predictor of psychological score and then investigate the relationship further. Section 6.1.3 described how to fit and test a functional linear model. The functional predictors and parameters are defined by B-splines with 20 basis functions. For each predictor cross validation was carried out to find a smoothing parameter which offered a trade-off between model accuracy and smoothness of the functional parameter. The smoothing parameters returned by the cross-validation range from 6 to 200 with a mean of around 70. To allow direct comparability the mean smoothing parameter of 70 is used for all curvature functions. Table 7.2 shows the p-value of the test of no effect for each curvature function predictor along with the  $R^2$  value for the significant predictors.

A functional regression, carried out using the functional linear model with single functional predictors, suggests that it is xy curvature of the nasal base and xz curvature of the nasal bridge which show a significant relationship with psychological score, while xz curvature of the nasal base is close to being a significant predictor. The  $R^2$  values for both significant curvature function predictors

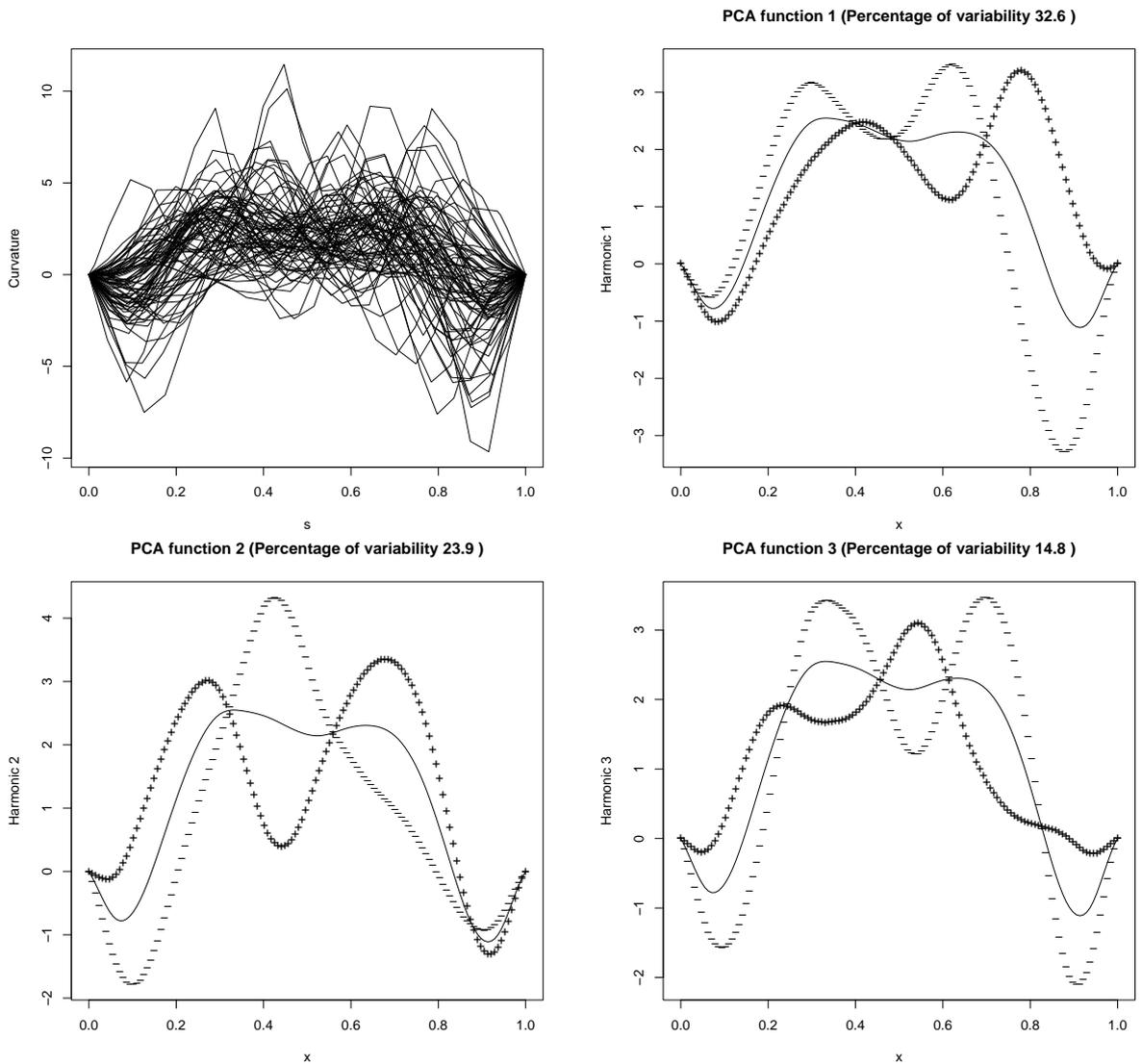
| Predictor                        | p-value      | $R^2$       |
|----------------------------------|--------------|-------------|
| Profile yz curvature             | 0.173        |             |
| Profile xy curvature             | 0.184        |             |
| Upper lip xz curvature           | 0.733        |             |
| Upper lip xy curvature           | 0.252        |             |
| Nasal rim xz curvature           | 0.768        |             |
| Nasal rim xy curvature           | 0.109        |             |
| Nasal base xz curvature          | 0.057        |             |
| <b>Nasal base xy curvature</b>   | <b>0.019</b> | <b>0.12</b> |
| <b>Nasal bridge xz curvature</b> | <b>0.034</b> | <b>0.07</b> |
| Nasal bridge xy curvature        | 0.071        |             |

**Table 7.2:** Significance of each curvature function as a predictor of psychological score using the functional linear model.

are low showing that none of the curvature functions individually explain a great deal of the variation in psychological score. Since xy curvature of the nasal base is the individual curvature function which has the strongest relationship with psychological score this relationship will be investigated further.

For reference, Figure 7.2 shows the xy curvature functions of the nasal base and the effect of the first three principal components on the mean function. Note that  $s = 0$  is the right end of the nasal base while  $s = 1$  is the left end of the nasal base.

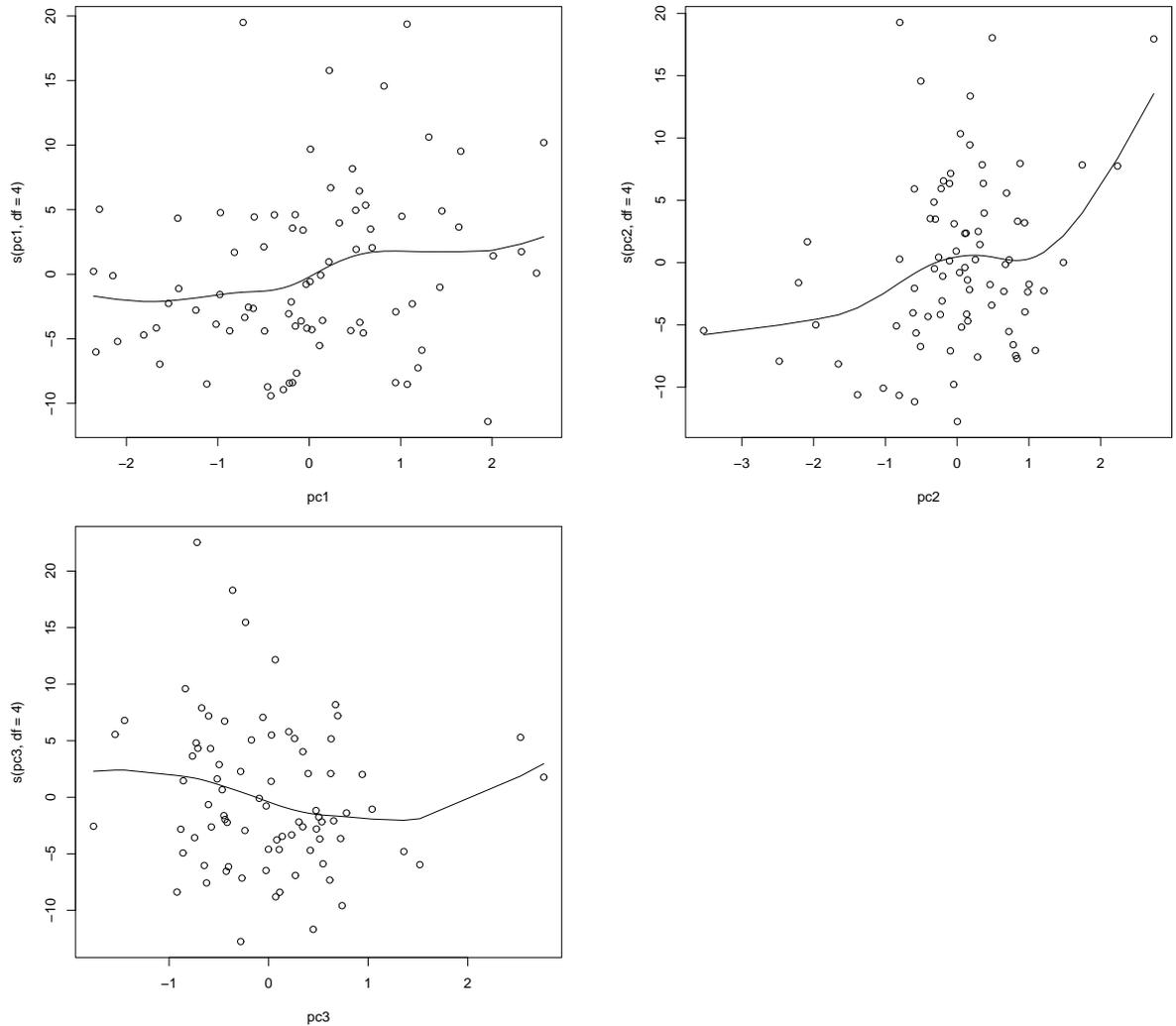
The first principal component appears to mainly explain variation in the magnitude of curvature at the end of the nasal base whilst the second principal component mainly explains variation in curvature at the start and middle of the nasal base. The third principal component appears to contain a global variation between the curvature functions. To attempt to explain the relationship between psychological score and the xy curvature function for the nasal base a linear model and an additive model are fitted to examine the relationship between the combined effect of the first three principal component scores and the psychological score. Table 7.3 shows the p-values for each term in the linear and additive models whilst Figure 7.3 shows the function given by the additive model for each principal score.



**Figure 7.2:** xy curvature functions for the nasal base and the effect (+/- two standard deviations of the principal component) of the first three principal components on the mean curvature function.

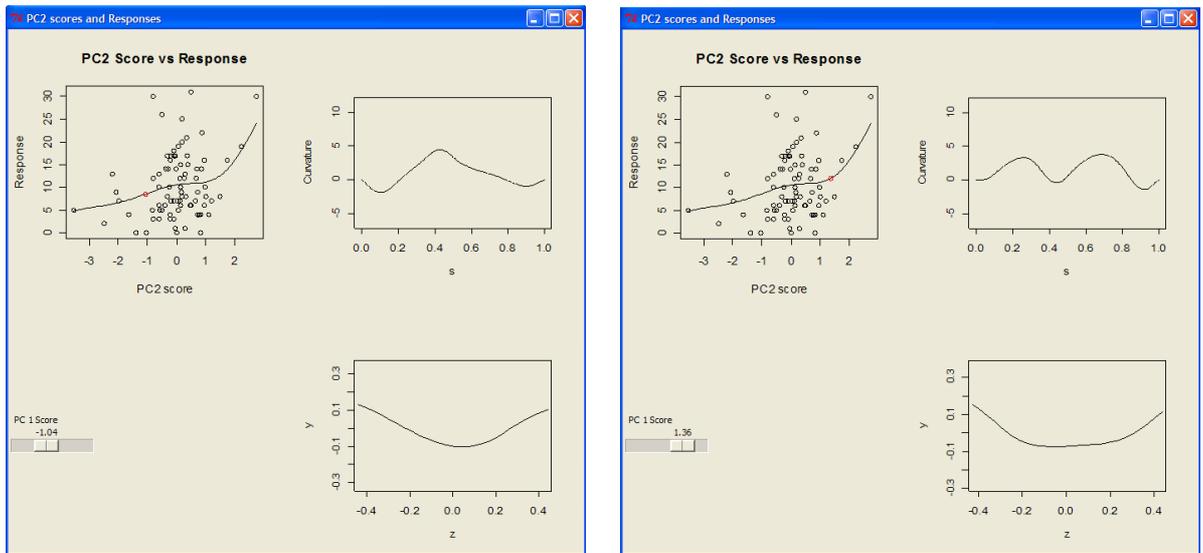
| Principal Component | Linear model coefficient | Linear model p-value | AM p-value |
|---------------------|--------------------------|----------------------|------------|
| First               | 1.23                     | 0.078                | 0.3        |
| Second              | 1.96                     | 0.017                | 0.014      |
| Third               | -1.15                    | 0.263                | 0.4        |

**Table 7.3:** Significance of principal component scores for nasal base xy curvature functions as combined predictors of psychological score.



**Figure 7.3:** The functions of the principal components scores returned by the additive model.

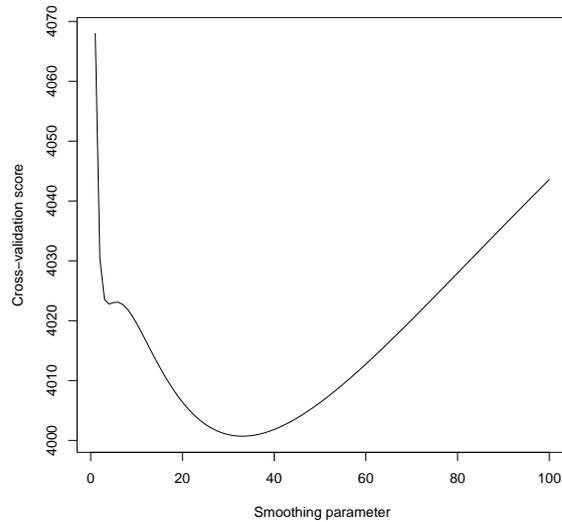
From both the p-values in Table 7.3 and the plots in Figure 7.3 it appears that the relationship with the second principal component score may be able to give an insight into the relationship between the curvature function and psychological score. Interestingly, as already mentioned, the second principal component describes more general variation in the functions than the first principal component score and only describes about 9% less variability. Considering both the linear model coefficient and the additive model function it does appear that, in general, as second principal component score of the curvature function increases psychological score increases. To graphically investigate the relationship and consider how a change in second component score affects a reconstructed nasal base an `rpanel` plot of the type described in Section 6.1.1 can be produced. The `rpanel` is shown for low and high component scores in Figure 7.4.



**Figure 7.4:** The `rpanel` showing; the nonparametric fit of the relationship between the second principal component score and psychological score (top left), the curvature function corresponding to that component score (top right) and the reconstructed nasal base using that curvature function (bottom right). The left `rpanel` shows a low component score whilst the right `rpanel` shows a high component score.

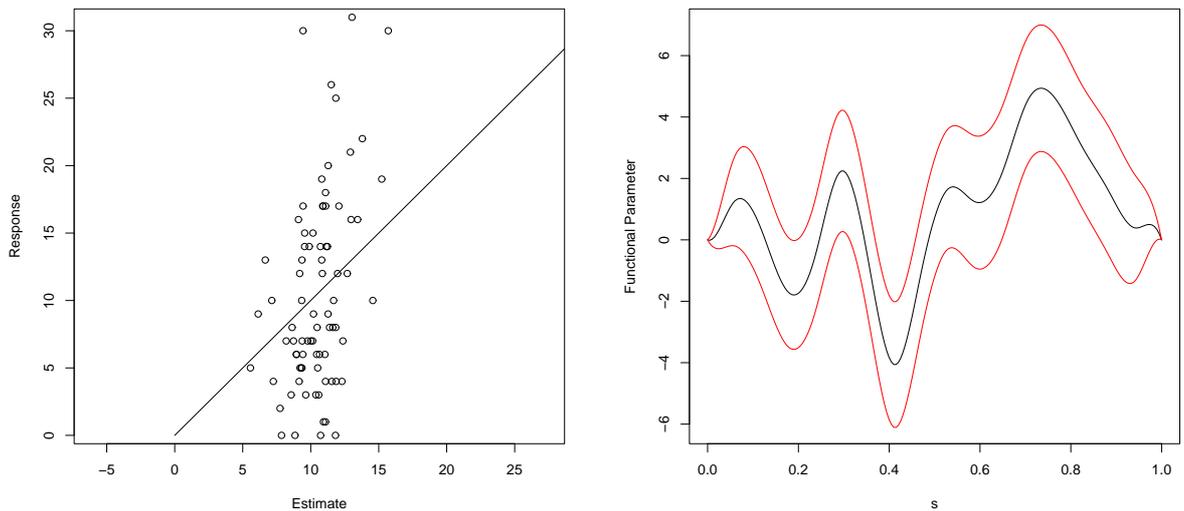
The `rpanel` snapshots from Figure 7.4 indicate that as the second principal component increases the psychological score tends to increase. Furthermore the effect of increasing the second principal component score is that  $xy$  curvature in the middle of the nasal base tends towards zero while curvature towards the ends of the nasal base increases. This results in a flattening out of the nasal base in the middle with the bending occurring towards the ends. The indication is that children with flatter areas in the middle of the nasal base may have higher psychological scores than children with a more rounded nasal base.

To formally investigate the relationship between  $xy$  curvature of the nasal base and psychological score the functional linear model and nonparametric model will be investigated in more detail. Earlier it was noted that there was a large spread in the smoothing parameters, returned by the cross-validation method, for the functional linear model using different curvature functions. To allow the results from all curvature functions to be directly comparable, the mean of the smoothing parameters was taken. To analyse the model using  $xy$  curvature of the nasal base in more detail the cross-validation process is repeated with the cross-validation score function shown in Figure 7.5.



**Figure 7.5:** Cross-validation score function for the xy curvature of the nasal base predictor.

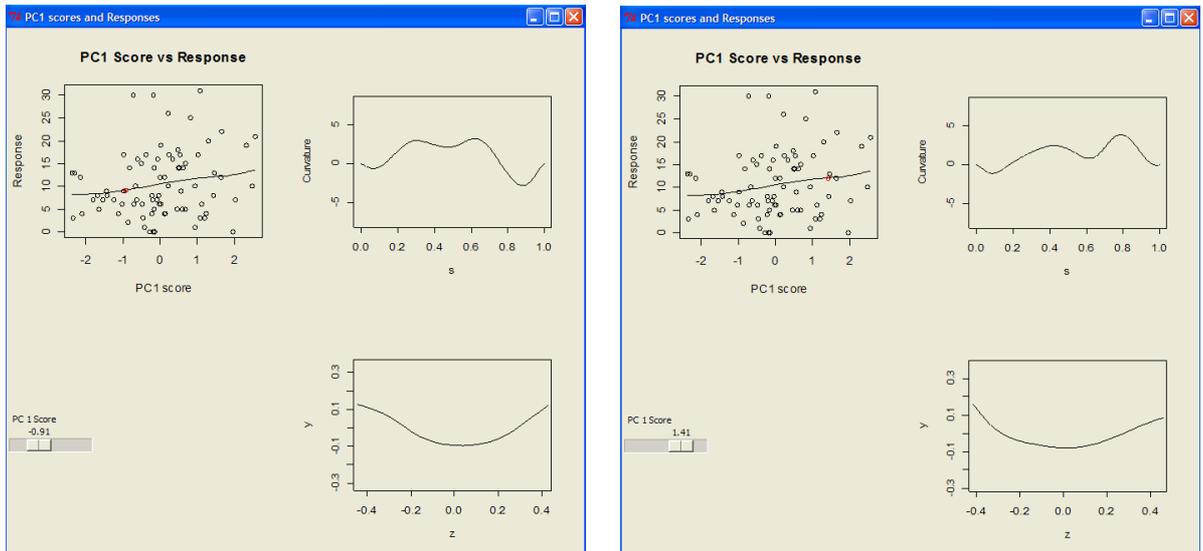
The cross-validation function returned a minimum cross-validation score when the smoothing parameter was 33 therefore the functional linear model is produced using this smoothing parameter. The p-value for a test of no effect of the model using this smoothing parameter is 0.022 whilst the  $R^2$  value is 0.15. Figure 7.6 shows a plot of the psychological score estimated using this method against the true psychological score and the functional parameter for this model.



**Figure 7.6:** Predicted psychological score, using the functional linear model, against true psychological score (left) and the functional parameter with red confidence bands (right).

The major interest in Figure 7.6 is the functional parameter. It appears that there are two areas where the functional parameter is clearly non-zero. Near  $s =$

0.4 the functional parameter is negative and since the  $xy$  curvature functions are generally positive at this point it appears that higher curvature values at  $s = 0.4$  indicate lower psychological score. This is further evidence of the suggestion that subjects with nasal bases which are more rounded in the middle may have lower psychological scores. The functional parameter is positive at around  $s = 0.8$  and since the curvature function is spread between positive and negative values of curvature at this point it appears that children with negative curvature at this point in general have lower psychological scores than children with positive curvature. Since the first principal component was primarily concerned with variation in this area of the curvature functions an `rpanel` plot may be useful to give greater insight into this effect. Figure 7.7 shows the `rpanel` for a low and high component score.



**Figure 7.7:** The `rpanel` showing; the nonparametric fit of the relationship between the first principal component score and psychological score (top left), the curvature function corresponding to that component score (top right) and the reconstructed nasal base using that curvature function (bottom right). The left `rpanel` shows a low component score whilst the right `rpanel` shows a high component score.

The `rpanel` snapshots show a slight trend of psychological score increasing as first principal component score increases. The effect of increasing first principal components score appears to be a skewing of the nasal base to the right so that the turning point of the nasal base which would usually indicate landmark  $sn$ , shown in Figure 1.1, occurs towards the right of the nasal base rather than at the centre of the nasal base. Negative curvature near  $s = 1$  suggests that the nasal

base is turning back up the face towards the end of the nasal base after turning down the face after landmark  $sn$  indicating that the  $sn$  landmark occurred near the centre of the nasal base. Positive curvature near  $s = 1$  suggests that the nasal base is still turning down the face towards the end of the nasal base as landmark  $sn$  occurred near the right end of the nasal base. Therefore it appears that the positive values of the functional parameter around  $s = 0.8$  are due to the fact that positive curvature at this point indicates a skewed nasal base which may result in increased psychological score whilst negative curvature values suggest a centred nasal base resulting in lower psychological scores.

### 7.1.2 Nonparametric regression

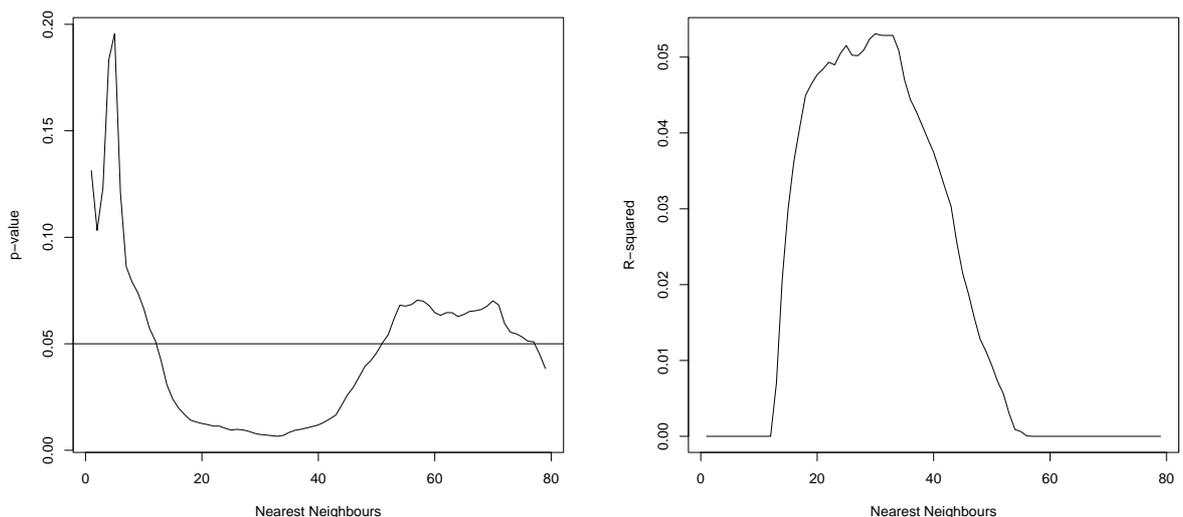
Nonparametric regression offers an alternative form of functional regression to the functional linear model. Section 6.1.4 described how nonparametric techniques are used to estimate the response using the functional predictors and how to test the effect of functional predictors. The semi-metric (in fact in this case metric) used was the  $L_2$  metric. The reason for using a metric rather than a semi-metric is that the condition appropriate for semi-metrics, namely  $d(x_i, x^*) = 0 \Leftrightarrow x_i = x^*$ , may result in curves with very different magnitudes of curvature being categorised as ‘close’. For each response estimated the bandwidth is chosen such that 40 nearest neighbours (half the subjects) are given non-zero weight in the prediction. Table 7.4 shows the p-value of the test of no effect for each curvature function predictor using this nonparametric regression technique along with the  $R^2$  value for the significant predictors.

The nonparametric regression suggests that xy curvature of the nasal base is the individual curvature function which has the strongest relationship with psychological score. This is consistent with the results of the functional linear model. However, whereas the functional linear model suggests that xz curvature of the nasal bridge is the only other curvature function which has a significant relationship with psychological score, nonparametric regression suggests that xy curvature of the nasal rim has a significant relationship with psychological score but xz curvature of the nasal bridge does not. This shows that the methods are not equivalent; although, it is reassuring that, despite the fact the  $R^2$  values are low, both show that xy curvature of the nasal base has the strongest relationship with psychological score.

| Predictor                      | p-value      | $R^2$       |
|--------------------------------|--------------|-------------|
| Profile yz curvature           | 0.687        |             |
| Profile xy curvature           | 0.126        |             |
| Upper lip xz curvature         | 0.185        |             |
| Upper lip xy curvature         | 0.985        |             |
| Nasal rim xz curvature         | 0.845        |             |
| <b>Nasal rim xy curvature</b>  | <b>0.032</b> | <b>0.02</b> |
| Nasal base xz curvature        | 0.061        |             |
| <b>Nasal base xy curvature</b> | <b>0.012</b> | <b>0.04</b> |
| Nasal bridge xz curvature      | 0.368        |             |
| Nasal bridge xy curvature      | 0.178        |             |

**Table 7.4:** Significance of each curvature function as a predictor of psychological score using nonparametric regression.

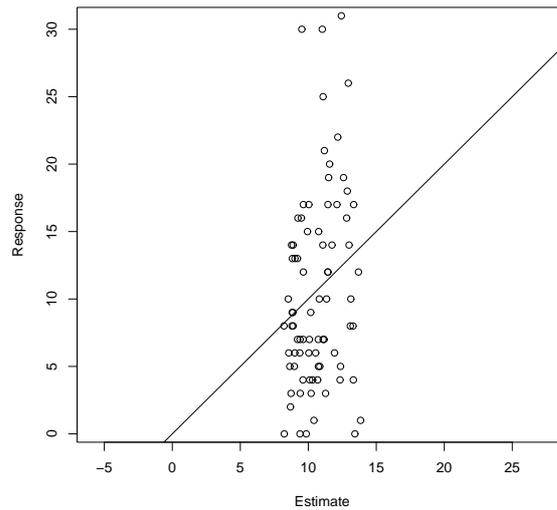
When using the nonparametric regression method to investigate relationships between the curvature functions and psychological score a rather arbitrary number of nearest neighbours were chosen. To investigate the effect of changing the number of nearest neighbours on the significance of the relationship between xy curvature of the nasal base and psychological score the p-value from the test of no effect and the  $R^2$  value returned by the nonparametric regression for all numbers of nearest neighbours from 1 to 79 are calculated and shown in Figure 7.8.



**Figure 7.8:** Plot of the p-value from the test of no effect (left) and  $R^2$  value (right) returned from a nonparametric regression using xy curvature of the nasal base and various numbers of nearest neighbours.

It appears that using a moderate number of nearest neighbours, anywhere

between 15 and 50, shows a significant relationship between xy curvature of the nasal base and psychological score. The  $R^2$  value although small for all numbers of nearest neighbours is largest, at 0.05, when using 30 nearest neighbours. Figure 7.9 shows a plot of known psychological score against the psychological score estimated using a nonparametric regression with 30 nearest neighbours.



**Figure 7.9:** Predicted psychological score, using nonparametric regression, against true psychological score.

Considering the  $R^2$  values for both the functional linear model and the nonparametric regression and the plots of predicted against true psychological score in Figures 7.6 and 7.9 it is clear that the relationship although significant is not particularly strong. Perhaps a model with multiple curvature functions would be a better predictor of psychological score.

## 7.2 Multiple Functional Predictor Analysis

Whilst using xy curvature of the nasal base as a single functional predictor of psychological score showed that the curvature function was a significant predictor, the relationship was not strong. A model using more than one of the curvature functions may be more useful as a predictor of psychological score. Section 6.2 outlines how many of the methods used for a single functional predictor can be extended to multiple functional predictors and many of these techniques will be used throughout this section.

### 7.2.1 Functional linear model and investigating significant relationships

Section 6.2.3 outlined how to fit and test a functional linear model with multiple predictors. Backward stepwise methods where the least significant predictor, in addition to the other predictors in the model, is removed until all variables have a p-value of less than 0.05 is used to select the combination of curvature functions which has the best significant relationship with psychological score. The results of the stepwise process are shown in Table 7.5 with a smoothing parameter of 70 for each functional predictor.

| Predictor removed       | p-value | $R^2$ |
|-------------------------|---------|-------|
| Full model              | -       | 0.467 |
| Nasal rim xz curvature  | 0.762   | 0.457 |
| Upper lip xz curvature  | 0.385   | 0.446 |
| Upper lip xy curvature  | 0.294   | 0.421 |
| Profile yz curvature    | 0.282   | 0.397 |
| Profile xy curvature    | 0.186   | 0.374 |
| Nasal base xz curvature | 0.176   | 0.352 |

**Table 7.5:** Functional linear model backward stepwise selection. The p-value is for the removed predictor whilst the  $R^2$  value is from the model with that predictor removed.

The final model given by the functional linear model contains the predictors nasal rim xy curvature, nasal base xy curvature, nasal bridge xz curvature and nasal bridge xy curvature with the model having an  $R^2$  value of 0.352. The significance of each predictor in addition to the other predictors in the model is shown in Table 7.6.

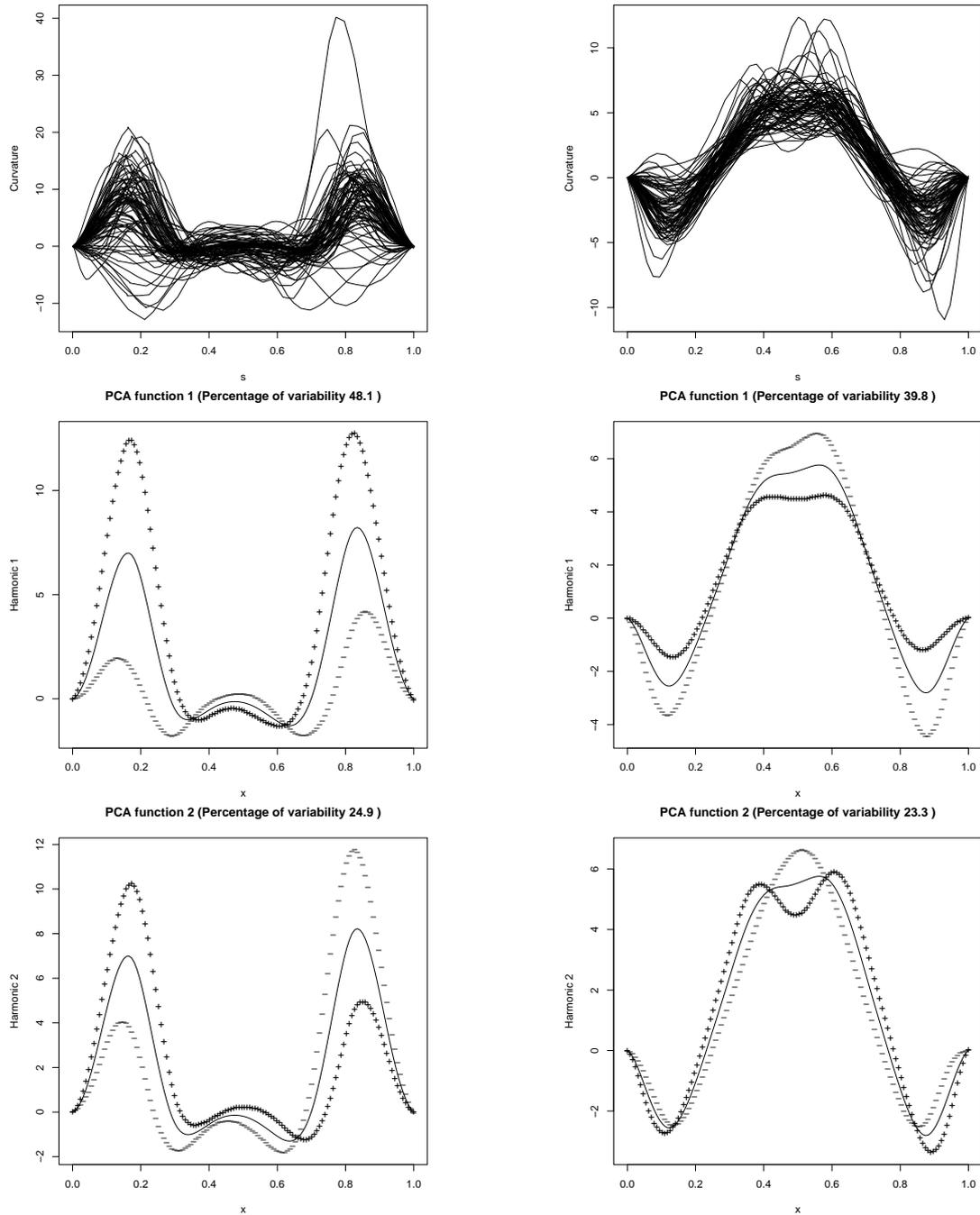
| Predictor                 | p-value |
|---------------------------|---------|
| Nasal rim xy curvature    | 0.018   |
| Nasal base xy curvature   | <0.001  |
| Nasal bridge xz curvature | 0.007   |
| Nasal bridge xy curvature | 0.046   |

**Table 7.6:** Significance of functional parameters in final functional linear model.

Nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature

are all highly significant predictors while the p-value for nasal bridge xy curvature is close to the significance level. Figure 7.2 has already shown the xy curvature functions of the nasal base and the effect of the first three principal components on the mean curvature function. Figure 7.10 shows both the xy curvature functions of the nasal rim and the xz curvature functions of the nasal bridge and the effects of the first two principal components on each mean curvature function. Note that, for simplicity, nasal bridge xy curvature is not included when informally investigating the nature of the relationships. This seems justified as the p-value is very close to the significance level.

The xy curvature functions of the nasal rim appear to have two areas of positive curvature where the nasal rim turns down towards the bottom of the face just before the nasal rim connects to the face. These turning points correspond to landmarks *alo0R* and *alo0L* shown in Figure 1.1. One function has extremely large curvature around  $s = 0.8$  this is due to that child having problems where the nasal rim connects to the face due to their cleft. The first principal component appears to explain variation in the magnitude of curvature at these points. The second principal component appears to explain differences in position of these points as well as further variation in magnitude of curvature. The xz curvature functions of the nasal bridge have an area of positive curvature around  $s = 0.5$  which indicates the area that the nasal bridge turns back towards the face between the eyes, at landmark *n*. There are also areas of negative curvature around  $s = 0.15$  and  $s = 0.85$  which indicate where the nasal bridge turns away from the face just after leaving/approaching the corner of the eye. The first principal component appears to explain variation in the magnitude of curvature at these three turning points whilst the second principal component primarily explains variation in curvature at the middle of the nasal bridge.



**Figure 7.10:** Curvature functions and effect (+/- two standard deviations) of the first two principal components for xy curvature of the nasal rim (left) and xz curvature of the nasal bridge (right).

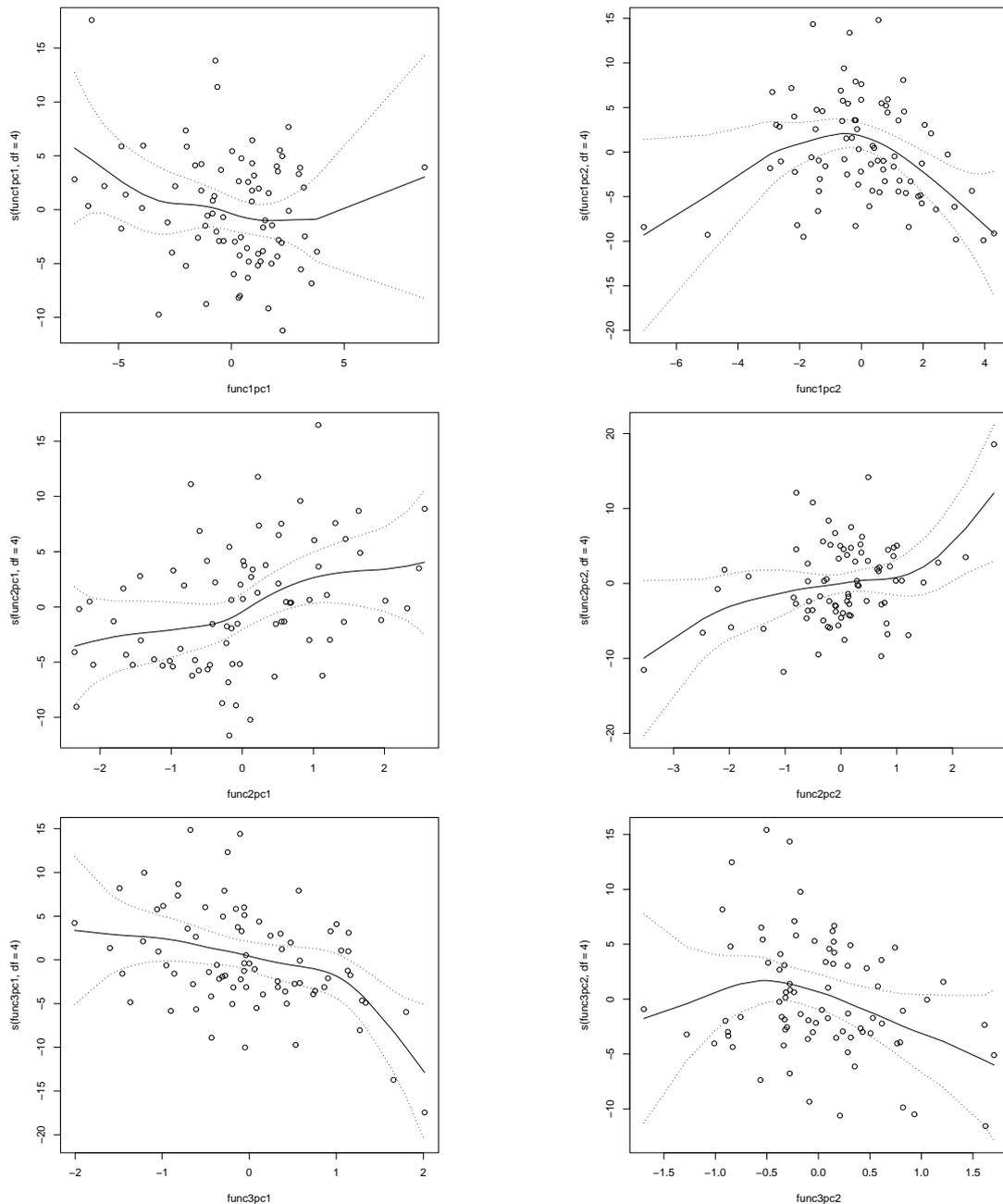
To attempt to give insight into the relationship between psychological score and the combined effect of nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature, a linear model and an additive model are fitted to investigate the combined effects of the first two principal components of each

functional predictor. Table 7.7 shows the p-values for each term in the linear and additive models whilst Figure 7.11 shows the function from the additive model for each principal component.

| Functional predictor      | Principal component | Linear model coefficient | Linear model p-value | AM p-value |
|---------------------------|---------------------|--------------------------|----------------------|------------|
| Nasal rim xy curvature    | First               | -0.48                    | 0.088                | 0.34       |
| Nasal rim xy curvature    | Second              | -0.62                    | 0.112                | 0.03       |
| Nasal base xy curvature   | First               | 1.59                     | 0.016                | 0.04       |
| Nasal base xy curvature   | Second              | 2.31                     | 0.003                | 0.02       |
| Nasal bridge xz curvature | First               | -2.32                    | 0.010                | 0.01       |
| Nasal bridge xz curvature | Second              | -2.31                    | 0.048                | 0.29       |

**Table 7.7:** Significance of first two principal component scores for nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature functions as combined predictors of psychological score.

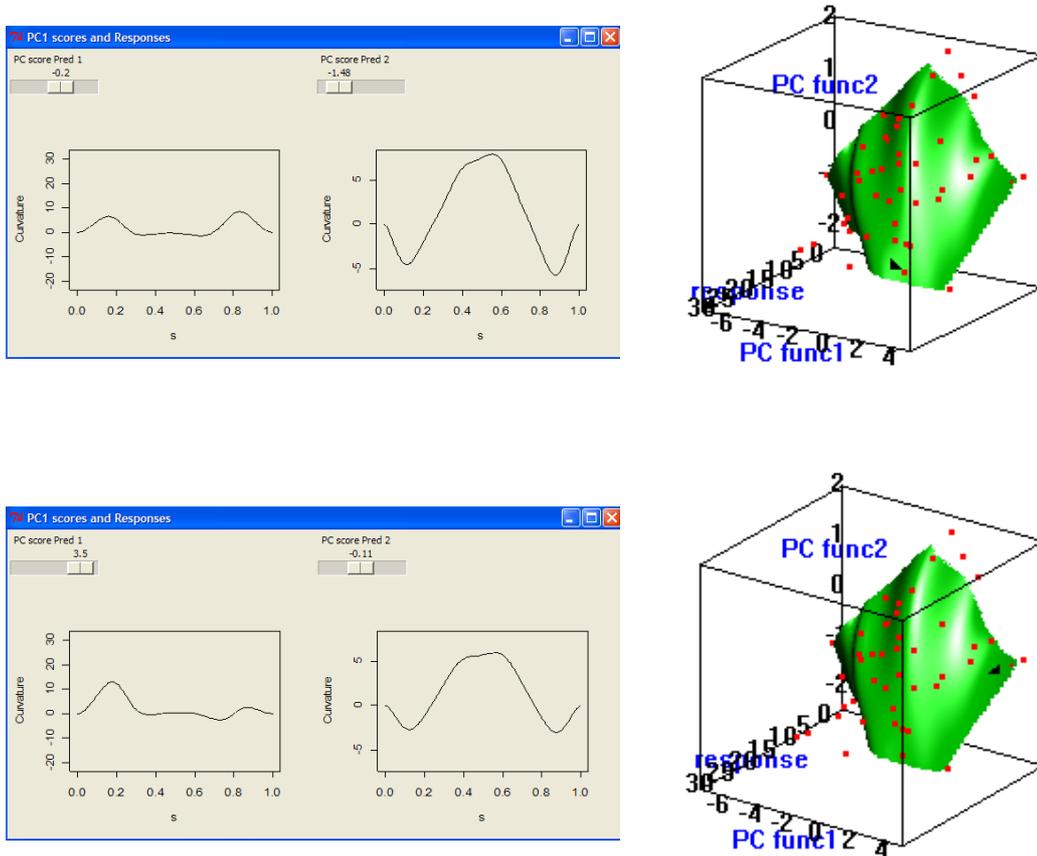
Interestingly when looking at the p-values for the linear model of the principal components it appears that neither principal component scores from the xy curvature of the nasal rim are useful predictors in addition to the other component scores. However, by looking at the scatterplots in Figure 7.11 it is clear that for the nasal rim the first principal component scores, and to a greater extent the second principal component scores, appear to have a quadratic relationship with psychological score. Evidence that this quadratic relationship may be useful as a predictor of psychological score is given by the fact that the second principal component of the nasal rim xy curvature functions is a significant term in the additive model. Further it seems that as both the first and second principal component scores of nasal bridge xz curvature functions increase the psychological score appears to decrease. Since Figures 7.4 and 7.7 investigated the relationship between the principal component scores for nasal base and psychological score the relationship between the joint effect of the second principal component for the nasal rim functions and the first principal component for the nasal bridge functions and psychological score will be examined.



**Figure 7.11:** The functions of the principal components scores returned by the generalised additive model for nasal rim xy curvature (top), nasal base xy curvature (middle) and nasal bridge xz curvature (bottom).

Section 6.2.1 discussed a method which involved finding a nonparametric fit of a surface showing the relationship between the two sets of principal component scores and the response then using an `rpanel` adjusting the values of the principal components to visualise the effect on the response and on the individual mean

curvature functions. Figure 7.12 shows the `rpanel` and corresponding position on the fitted function for a high and a low response.

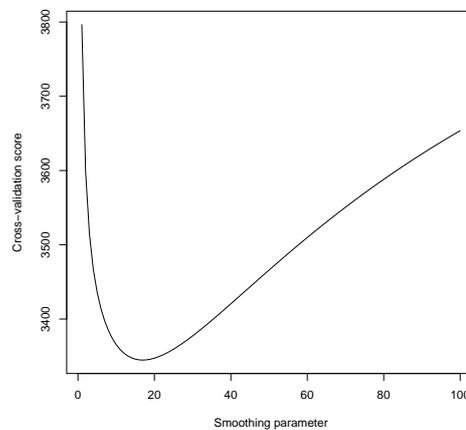


**Figure 7.12:** The `rpanel` and three-dimensional plot with best surface giving a high psychological score (top) and a low psychological score (bottom). Function 1 corresponds to  $xy$  curvature of the nasal rim, function 2 corresponds to  $xz$  curvature of the nasal bridge and response corresponds to psychological score.

According to the exploratory analysis in Figure 7.12 a subject with only two areas of bending down the face at the ends of the nasal rim and with large amount of bending in the nasal bridge is likely to have a large psychological score. Interestingly it appears that a subject who only has a larger magnitude of bending on one end of the nasal rim than the other and has a low amount of bending on the nasal bridge is likely to have a low psychological score. It does not seem unreasonable that a subject with a flatter, less ‘bumpy’ nasal bridge should have lower psychological score. The results from the nasal rim seem less intuitively sensible. As shown in Figure 7.11 the second principal component of

xy curvature of the nasal rim is dominated by a small number of subjects who have low psychological scores. These subjects appear to have a large amount of variability in curvature at the ends of the nasal rim so any departure from the mean curve, whether it be an increase or decrease in curvature, appears to show a decrease in estimated psychological score. This exploratory analysis has given an indication of some of the relationships between psychological score and the curvature functions. To further explain the relationships it is useful to investigate the functional parameters from the functional linear model.

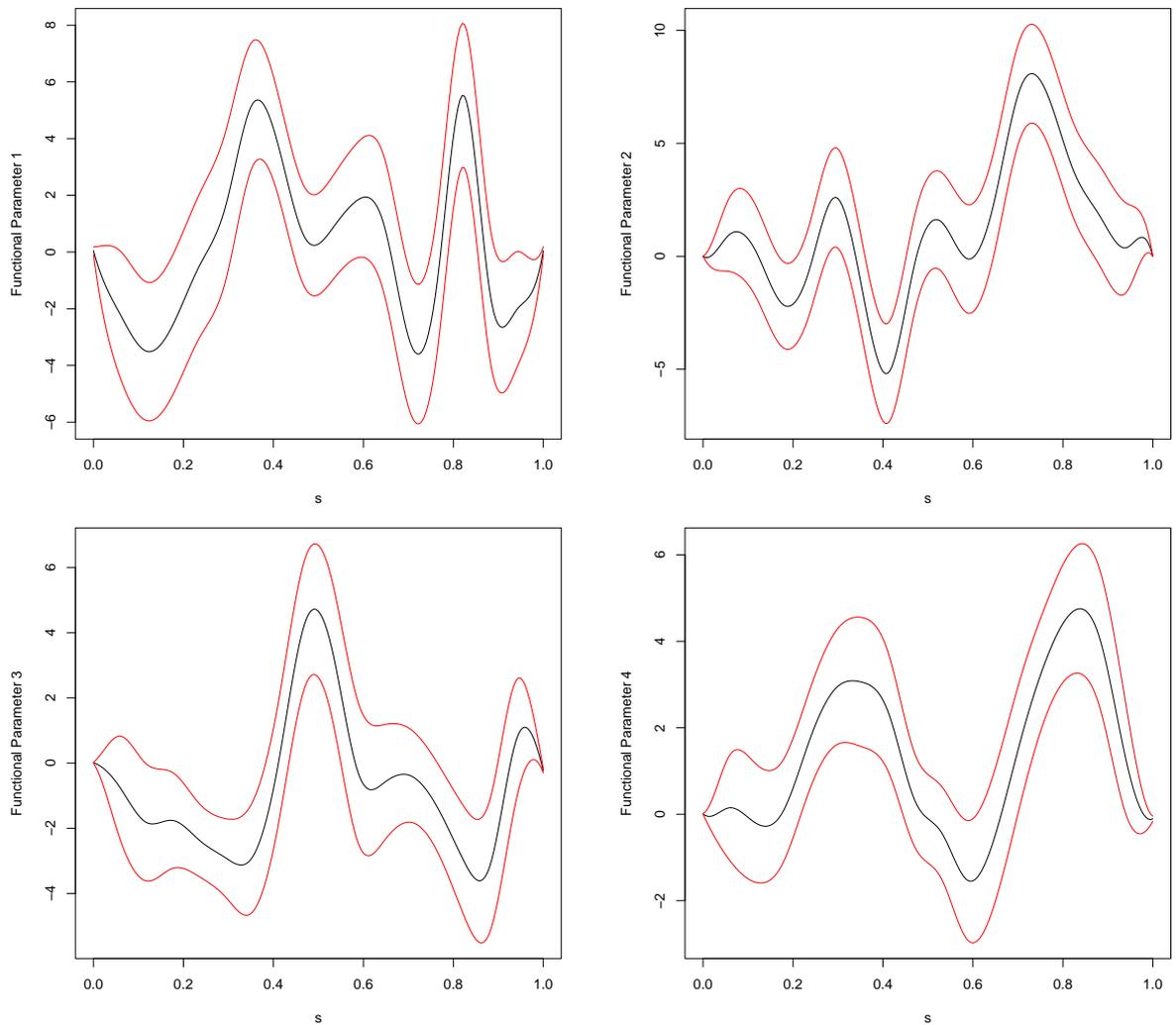
In the stepwise selection considered earlier in this section the smoothing parameter for each functional predictor in the multiple predictor model was set as the mean of the optimal smoothing parameter from the functional linear model between psychological score and each of the individual functional predictors. Section 6.2.3 outlined a cross-validation procedure to estimate the optimal smoothing parameter given that the smoothing parameter for each predictor is equal. Figure 7.13 shows the cross-validation function for the functional linear model produced by the backward stepwise selection with functional predictors nasal rim xy curvature, nasal base xy curvature, nasal bridge xz curvature and nasal bridge xy curvature.



**Figure 7.13:** Cross-validation score function for the final functional linear model.

The cross-validation procedure returned an optimal smoothing parameter of 17. The functional linear model is produced with this smoothing parameter used for each of the functional predictors. This model has an  $R^2$  value of 0.511 and the functional parameters from the model are shown in Figure 7.14.

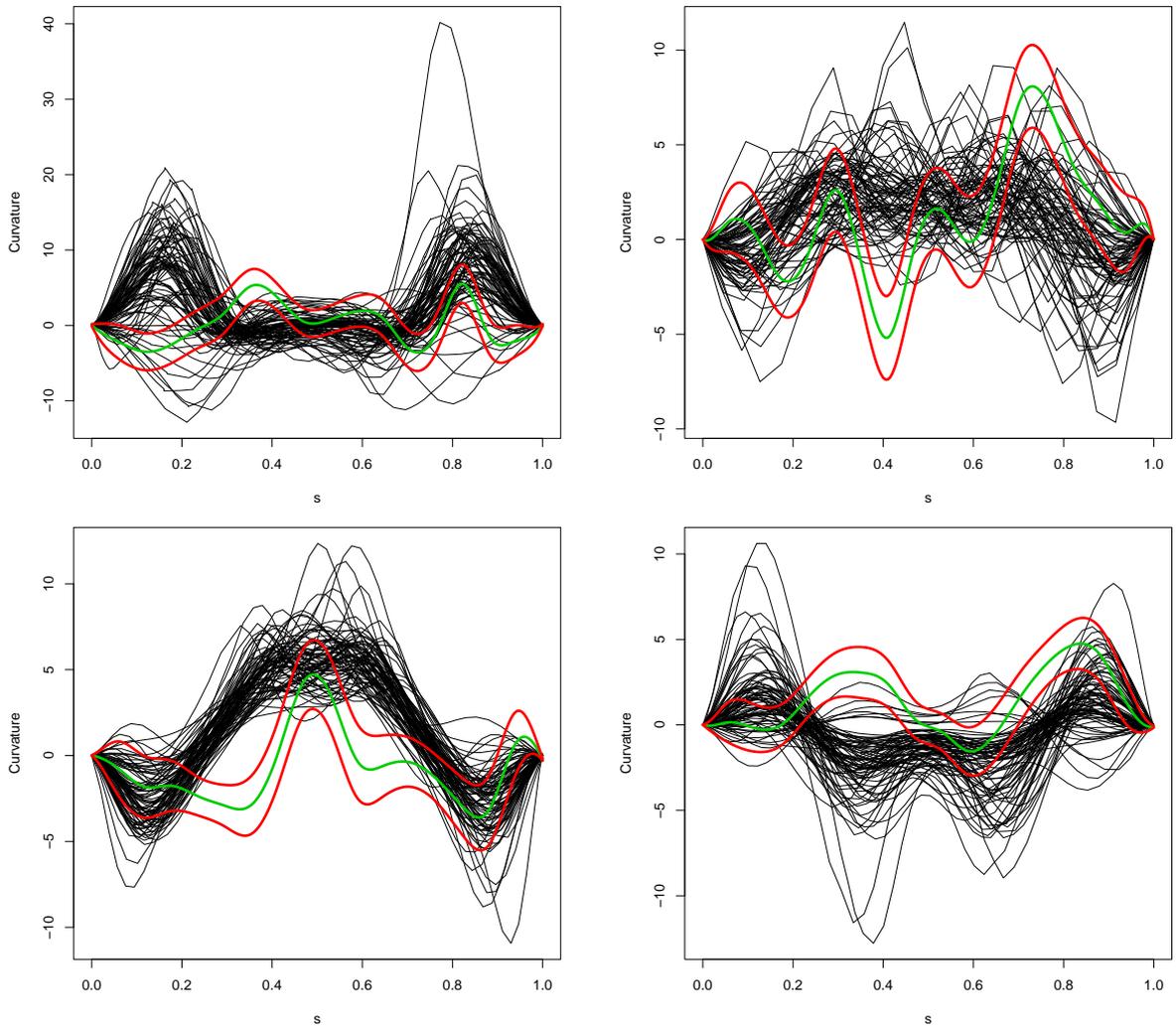
Anatomical interpretation of the functional parameters can become complicated. To aid this interpretation the four sets of curvature functions used in the



**Figure 7.14:** The functional parameters returned by the functional linear model for xy curvature of the nasal rim (top left), xy curvature of the nasal base (top right), xz curvature of the nasal bridge (bottom left) and xy curvature of the nasal bridge (bottom right).

functional linear model are shown in Figure 7.15 with the functional parameters and their confidence limits superimposed.

Considering firstly the functional parameter for xy curvature of the nasal rim, there are two areas where the parameter is clearly greater than zero at around  $s = 0.4$  and  $s = 0.8$ . At  $s = 0.4$  the curvature functions are generally negative though close to zero therefore the positive functional parameter suggests that those with higher magnitude of negative curvature at this point will have lower psychological scores. It is difficult to suggest any anatomical/psychological reason for this. At  $s = 0.8$  the curvature functions are positive therefore the



**Figure 7.15:** The functional parameters returned by the functional linear model superimposed on the curvature functions for xy curvature of the nasal rim (top left), xy curvature of the nasal base (top right), xz curvature of the nasal bridge (bottom left) and xy curvature of the nasal bridge (bottom right).

positive functional parameter suggests that a large amount of bending at this point indicates larger psychological scores. This is due to the fact that large bending up the face at this point may result in the shape of the nasal rim being distorted which suggests larger psychological scores.

The functional parameter for xy curvature of the nasal base is very similar to the functional parameter from the functional linear model in Section 7.1 containing only this functional predictor therefore interpretation is similar to that given in that section.

The functional parameter for xz curvature of the nasal bridge has an area

around  $s = 0.5$  where the parameter is clearly positive. The curvature functions are also positive at this area which is the point at the centre of the nasal bridge where it turns towards the face. The fact that both the parameter and curvature functions are positive suggest that subjects with larger curvature in the middle of the bridge of the nose will have larger psychological score. This is perhaps due to larger curvature towards the face indicating a more prominent nasal bridge which may result in larger psychological scores.

The functional parameter for  $xy$  curvature of the nasal bridge is clearly positive around  $s = 0.35$  and around  $s = 0.8$ . Since the curvature functions are negative around  $s = 0.35$  this suggests that a larger magnitude of negative curvature in this area suggests lower psychological score. As shown by the positive curvature around  $s = 0.15$  the nasal bridge turns up the face after leaving the corner of the eye but turns back down the face when it gets to the ridge of the nose at around 0.35. It is difficult to interpret why higher magnitude of curvature in this area suggests lower psychological scores. The area around  $s = 0.8$  where the functional parameter is positive is the transition between negative and positive curvature where the nasal bridge turns back up the face as it approaches the corner of the eye. The positive functional parameter suggest that subjects whose curvature function remains negative until later on the  $s$  axis will have a lower psychological score.

## 7.2.2 Nonparametric regression

As in the single predictor setting, nonparametric regression offers an alternative form of multiple functional regression to the functional linear model. Section 6.2.4 outlined how to carry out and test a nonparametric regression with multiple predictors. A backward stepwise procedure, with 40 nearest neighbours and the  $L_2$  metric used in the nonparametric regression, is used to select the combination of curvature functions which has the best significant relationship with psychological score with the results shown in Table 7.8.

| Predictor removed         | p-value | $R^2$ |
|---------------------------|---------|-------|
| Full model                | -       | 0.015 |
| Profile yz curvature      | 0.972   | 0.046 |
| Nasal rim xz curvature    | 0.792   | 0.071 |
| Nasal bridge xy curvature | 0.536   | 0.076 |
| Upper lip xy curvature    | 0.703   | 0.088 |
| Upper lip xz curvature    | 0.594   | 0.096 |
| Nasal base xz curvature   | 0.483   | 0.098 |

**Table 7.8:** Nonparametric regression backward stepwise selection. The p-value is for the removed predictor whilst the  $R^2$  value is from the model with that predictor removed.

The final model given by the nonparametric regression contains the predictors profile xy curvature, nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature with the model having an  $R^2$  value of 0.098. The significance of each predictor in addition to the other predictors in the model is shown in Table 7.9.

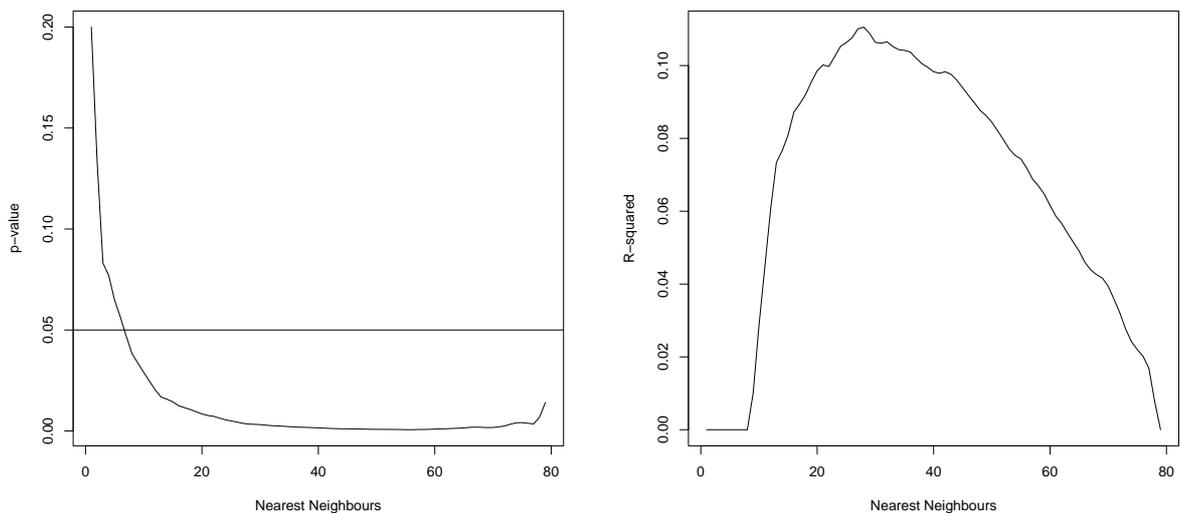
| Predictor                 | p-value |
|---------------------------|---------|
| Profile xy curvature      | 0.026   |
| Nasal rim xy curvature    | 0.007   |
| Nasal base xy curvature   | 0.041   |
| Nasal bridge xz curvature | 0.005   |

**Table 7.9:** Significance of functional parameters in the final nonparametric regression model.

Interestingly, three predictors, nasal rim xy curvature, nasal base xy curvature and nasal bridge xz curvature, are common to both the final functional linear model and the final nonparametric model. This gives a stronger indication of a significant relationship between the combined effect of these three predictors and psychological score. Although the variables included in the final model are similar for both the functional linear model and the nonparametric regression there is a major difference in the  $R^2$  values of the models using the different methods. The  $R^2$  values using the functional linear model are much larger than those for the nonparametric regression. This suggests that the functional linear model provides a better fit of the psychological scores when using multiple

predictors than nonparametric regression. As non-significant terms are removed in the functional linear model the  $R^2$  value decreases; however, interestingly, as non-significant terms are removed in the nonparametric regression the  $R^2$  value increases. This is due to non-significant functional predictors adversely affecting which subjects are chosen as nearest neighbours.

To investigate the effect, particularly on the  $R^2$  value, of changing the number of nearest neighbours a nonparametric regression using the variables returned by the stepwise selection, namely xy curvature of the midline profile, xy curvature of the nasal rim, xy curvature of the nasal base and xz curvature of the nasal bridge, is carried out for all possible numbers of nearest neighbours. The significance of the model, as tested by a test of no effect, and the  $R^2$  value of the model are shown in Figure 7.16.

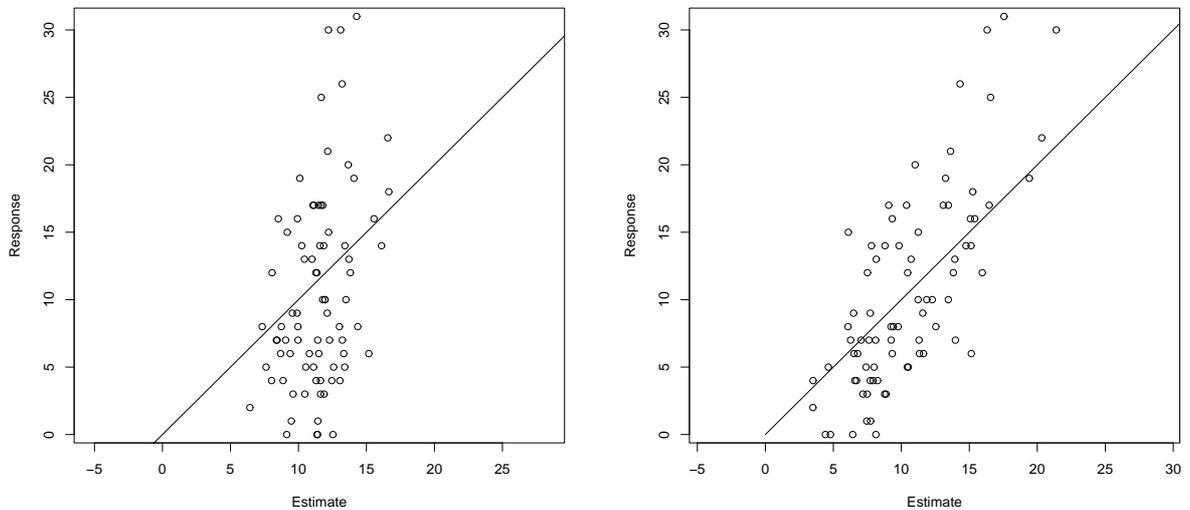


**Figure 7.16:** Plot of the p-value from the test of no effect (left) and  $R^2$  value (right) returned from a nonparametric regression using xy curvature of the midline profile, xy curvature of the nasal rim, xy curvature of the nasal base and xz curvature of the nasal bridge as predictors.

When more than 6 nearest neighbours are used in the nonparametric regression the model with the four functional variables is a significant predictor of psychological score. The optimal  $R^2$  value is 0.111 and is found using 28 nearest neighbours.

Recalling that the  $R^2$  value for the final functional linear model was 0.511 it is clear that the multiple functional linear model is a more useful predictor of psychological score than the nonparametric regression. To examine this a plot of predicted psychological score against true psychological score is shown for both

methods in Figure 7.17.



**Figure 7.17:** Predicted psychological score against true psychological score using nonparametric regression (left) and functional linear model (right).

It is clear from Figure 7.17 that the functional linear model gives a more accurate estimate of the psychological score than the nonparametric regression. It seems that whilst the difference in the predictive power of the two methods was not particularly large when only single functional predictors were used, there is a large difference in the predictive power of the methods when multiple functional predictors are used. It is however reassuring that both methods produce similar final models.

### 7.2.3 Functional additive model

A functional additive model also provides an alternative to the functional linear model when there are multiple functional predictors. Section 6.2.5 outlined how an additive model could be constructed using functional predictors. The method required nonparametric regression with the number of nearest neighbours selected. Here a backward stepwise selection method is used to select the best functional additive model with the number of nearest neighbours set to 40. Table 7.10 shows the result of the stepwise selection procedure.

The  $R^2$  values are extremely low for the models with a large number of predictors but increase slightly as predictors are removed from the model. It appears that this is again due to non-significant functional predictors adversely affecting which subjects are chosen as nearest neighbours in the nonparametric regression

| Predictor removed         | p-value | $R^2$  |
|---------------------------|---------|--------|
| Full model                | -       | <0.001 |
| Upper lip xy curvature    | 0.686   | <0.001 |
| Profile yz curvature      | 0.228   | <0.001 |
| Nasal rim xz curvature    | 0.218   | <0.001 |
| Upper lip xz curvature    | 0.224   | <0.001 |
| Profile xy curvature      | 0.149   | <0.001 |
| Nasal base xz curvature   | 0.105   | 0.030  |
| Nasal bridge xz curvature | 0.093   | 0.043  |
| Nasal bridge xy curvature | 0.110   | 0.059  |

**Table 7.10:** Additive model backward stepwise selection. The p-value is for the removed predictor whilst the  $R^2$  value is from the model with that predictor removed.

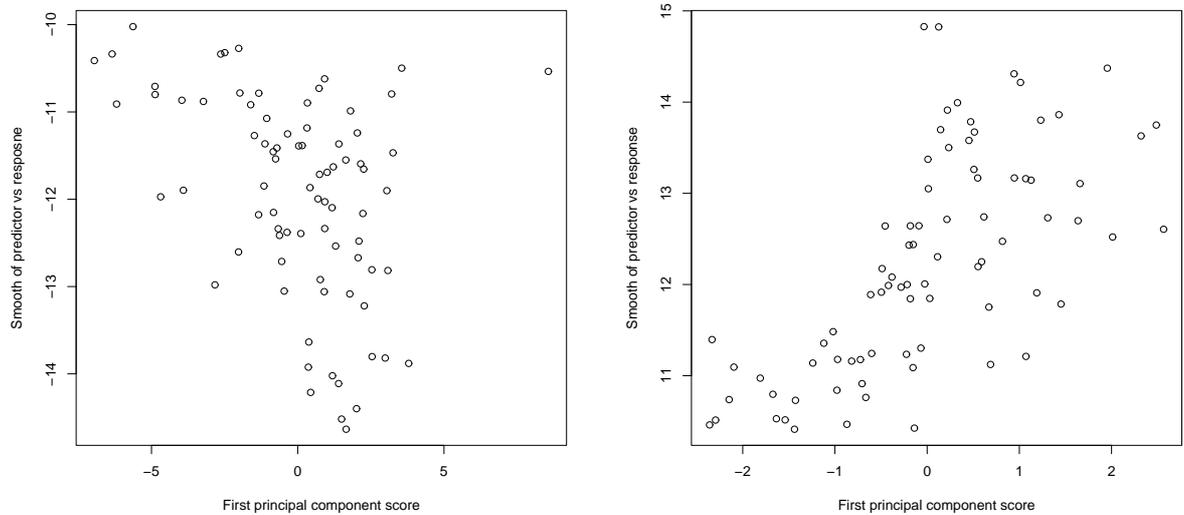
stage. The final model given by the additive model is one containing the xy curvature of the nasal base and xy curvature of the nasal rim and this model has an  $R^2$  value of 0.059. The significance of each predictor in addition to the other predictor is shown in Table 7.11

| Predictor               | p-value |
|-------------------------|---------|
| Nasal rim xy curvature  | 0.038   |
| Nasal base xy curvature | 0.013   |

**Table 7.11:** Significance of functional parameters in the generalised additive model.

To attempt to interpret the effect of the function of each predictor returned by the model, a plot of the first principal component of each predictor against the value returned by the model for that predictor is shown in Figure 7.18.

There does seem to be a relationship between the first component score and the value returned by the additive model for both predictors. For xy curvature of the nasal rim it appears that as principal component score increases the value returned by that function decreases. The effect of increasing the first principal component score is an increase in the amount of bending experienced at the ends of the nasal rim therefore it appears that subjects with a lower amount of bending at the ends of the nasal rim may have a higher psychological score. For xy curvature of the nasal base it appears that as the first principal component score

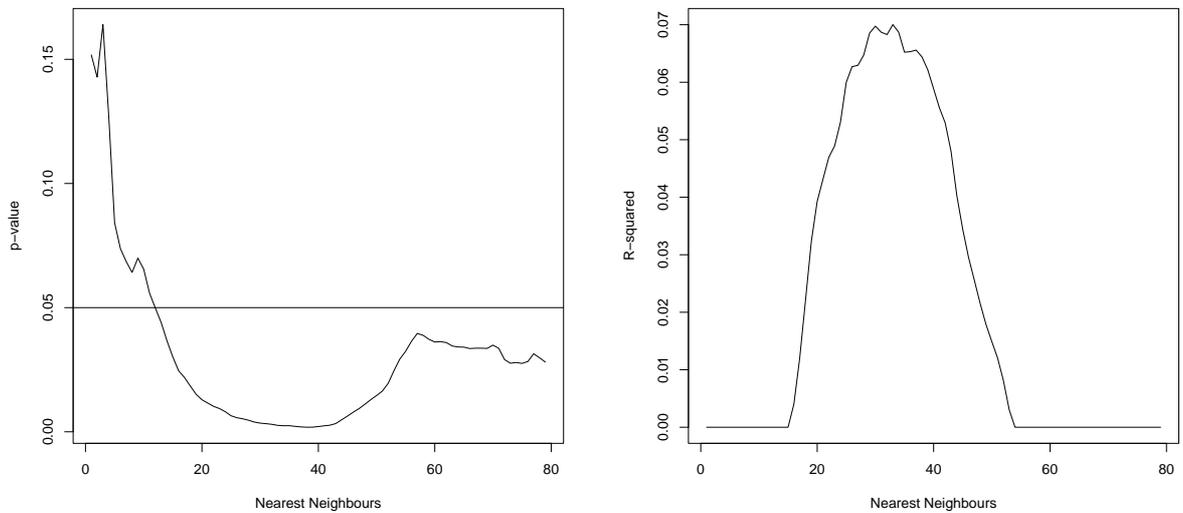


**Figure 7.18:** Plot of first principal component score against value returned by the additive model function for predictors; xy curvature of nasal rim (left) and xy curvature of nasal base (right).

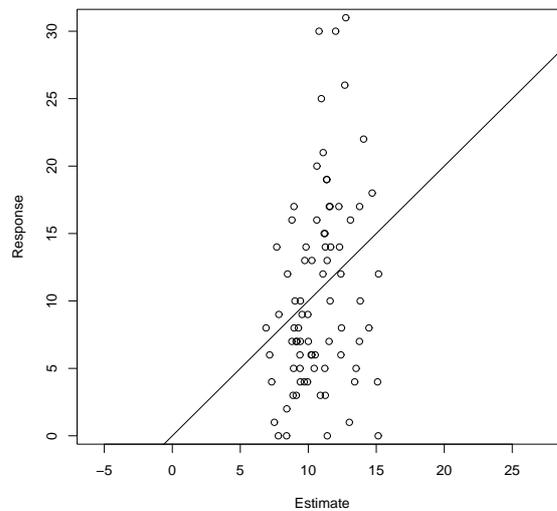
increases the value returned by the function increases. The effect of increasing the first principal component score is changing the curvature at the end of the nasal base from negative to positive therefore it appears that subjects with negative curvature at the end of the nasal base will have a lower psychological score.

When fitting the additive model an arbitrary number of nearest neighbours were used in the nonparametric regression stage. To investigate the effect of changing the number of nearest neighbours an additive model using xy curvature of the nasal rim and xy curvature of the nasal base as predictors is calculated using a range of number of nearest neighbours from 1 to 79. Figure 7.19 shows the significance of the model, tested using a test of no effect, and the  $R^2$  value for all nearest neighbours.

When more than 11 nearest neighbours are used the test of no effect shows that the additive model has a significant effect on psychological score. The optimal  $R^2$  value is 0.070 and this occurs when 33 nearest neighbours are used. Figure 7.20 shows a plot of estimated psychological score using an additive model with xy curvature of nasal rim and xy curvature of nasal base as predictors and 33 nearest neighbours, against true psychological score.



**Figure 7.19:** Plot of the p-value from the test of no effect (left) and  $R^2$  value (right) returned from a generalised additive model using xy curvature of the nasal rim and xy curvature of the nasal base.



**Figure 7.20:** Estimated psychological score, using an additive model, against true psychological score.

It is clear from both Figure 7.20 and the low  $R^2$  value that the additive model is not a strong predictor of psychological score. However, the plots of first principal component score against the value returned by the model for each function do give a useful insight into how each of the functional predictors may affect psychological score. It is also interesting to note that the significant predictors in the functional additive model, xy curvature of nasal rim and xy curvature of nasal base, are also significant predictors in the functional linear model and the

nonparametric regression. This provides even further evidence that these two curvature functions have a significant relationship with psychological score.

### 7.3 Concluding Remarks

This chapter has investigated potential relationships between facial shape, defined by curvature functions of five features, and psychological score. Throughout the chapter it appears that the features which have the strongest relationship with psychological score are the nasal base, nasal rim and to a lesser extent the nasal bridge. It seems entirely reasonable that bending of the nasal base and nasal rim have a relationship with psychological score as these areas are likely to be affected in some way by the subjects cleft. It is interesting that it is the  $xy$  curvature (bending up and down the face) which is significant as presumably bending in this direction is more noticeable when viewing a subject from front on.

The exact interpretation of relationships is difficult to state confidently. Tools such as the `rpanel` and additive model discussed in this chapter have proved useful in gaining some understanding of the relationships, although the discussion here is only a start. The fact that the `rpanel` is both simple to use and to understand may make it possible to enlist the help of a facial surgeon or psychologist in interpreting the relationships.

It seems that a functional linear model with multiple functional predictors produced the best estimates of the known psychological scores. However, it may be that this is simply the case for this data set and it would be useful to try the methods out on other data sets. In the nonparametric regression the area between the functions metric was used to give a measure of closeness. This was due to the fact that for many features it was the magnitude of curvature at certain characteristic points which most clearly defined individual curvature functions. Other simple semi-metrics were used to give a measure of closeness with none improving the model but perhaps a more sophisticated (semi-)metric may have produced better regression models.

# Chapter 8

## Conclusions

This thesis can be broken down into two related but distinct parts. Chapters 3, 4 and 5 used the ideas of curvature to describe and compare facial shape with data coming from a study into 1 year old children. The major aim of this part of the thesis is to make comparisons between control children and children with a cleft lip and/or palate. Chapters 6 and 7 investigate potential relationships between facial shape, defined by curvature functions encountered in previous chapters, and psychological score with the data coming from a study into 10 year old children. The major aim of this section is, for cleft children only, to discover and interpret links between facial shape and psychological score. This chapter will be split into three distinct sections. Section 8.1 will discuss the major issues with, and conclusions drawn from, using curvature to investigate differences between shapes while Section 8.2 will provide an overview of the techniques used and conclusions drawn from an investigation into relationships between curvature functions and psychological score. Section 8.3 will introduce and discuss potential future work.

### 8.1 Curvature

Many standard shape analysis techniques require the alignment of shapes through translation, rotation and in some instances rescaling. While these techniques are effective, for high-dimensional data they can be computationally expensive and further discrepancies that are found between shapes may in part be due to the alignment process. When analysing shape curves the necessity for alignment, and in turn the issues that come with it, can be avoided by analysing curves using curvature. This is due to the fact that the curvature of a curve is invariant with respect to location and rotation. This is not to say that analysing shape curves

using curvature is more or less effective than aligning the curves but it does offer a useful alternative.

There is a large amount of literature on defining and calculating curvature in geometry books. Many of the ideas are summarised in this thesis. For plane curves the curvature value is dependent on the positions of the tangent and normal vectors and also the arc length of the curve. Using some simple geometry, an equation can be found which calculates curvature at each point on the curve using the first and second derivatives of the  $x$  and  $y$  positions of the curve with respect to arc length. Since this allows curvature to be calculated at all arc lengths it is possible to produce a function which shows the curvature for all arc lengths. Curvature functions for different curves can then be compared with relative ease. However, to allow direct comparability, it is necessary to rescale the arc lengths of each curve to a common value; arc length 1 was chosen throughout this thesis, without changing the shape.

Using the derivative formula to calculate curvature is technically equivalent but in practice some issues occur. Numerical calculation of the derivatives can be produced in **R** using the `predict` function. For most relatively simple curves the first derivatives are calculated with reasonable accuracy; however, for second and particularly higher level derivatives, the calculated values are subject to some error, especially at the ends of the curve. In practice it is shown that for second derivatives the error is not sufficiently large to invalidate the curvature values calculated from the derivative formula although the errors in the tails have small undesirable effects on the calculation of curvature. Practical interpretation of the curvature functions provides another issue. A major aim of this methodology is to produce curvature functions which are both simple to interpret and compare. Even when analysing relatively smooth shape curves the curvature functions produced have large amounts of roughness; therefore to aid interpretation and comparison the curvature functions need to be smoothed. The method for doing this here was to smooth the functions of  $x$  and  $y$  position against arc length before calculating the derivatives. The advantage of this method over smoothing the final function of curvature is that the smoother the position functions the lower the error in the calculation of derivatives.

A useful method for investigating the accuracy of the calculated curvature function is to attempt to reproduce the original shape curve from the curvature function. There are formulae to reproduce the shape curve up to the similarity transformations of location and rotation and Procrustes methods can be used

to align the original and reconstructed shape curves so that differences between the curves are caused by errors in the curvature function. Using this method to assess the accuracy of the derivative formula curvature functions calculated for an example midline profile produced an accurate reconstruction of the original curve as long as the degrees of freedom for the smoothing of the position functions were not too low.

Comparing curvature functions is a useful way to compare the shapes of features and in particular compare different groups. An average curvature function for a group can be calculated by taking the piecewise average of the set of curvature functions and from this average curvature function it is possible to produce an average reconstructed shape curve. Comparing average reconstructed shape curves is a useful way to gain understanding of the anatomical differences present between groups. An interesting extension to simply taking a pointwise average of the curvature functions is to firstly align curvature functions according to points of anatomical/geometrical interest and then take the average. Averaging of this type is called structural averaging. The anatomical/geometrical points of interest are often shown as turning points on the curvature functions. For a collection of curvature functions the turning points which correspond to interesting points on the shape curve can be aligned by warping the arc length axis. Structural averaging is useful in that it gives a better indication of the average magnitude of curvature at points of interest than averaging the original curves.

Calculating curvature of space curves introduces extra complexity. In standard methodology the bending of a space curve is defined by two scalar values at each point on the curve called curvature and torsion. Curvature and torsion are dependent on a moving trihedron which contains the tangent, normal and binormal vectors. Using geometry it is possible to produce equations to calculate curvature, using the first and second derivatives of the  $x$ ,  $y$  and  $z$  positions of the curve with respect to arc length, and torsion, using the first second and third derivatives of the  $x$ ,  $y$  and  $z$  positions of the curve with respect to arc length. Curvature and torsion functions can then be produced to define the shape curve. However, due to the problems in calculating high order derivatives in  $\mathbf{R}$  the torsion functions are subject to a large amount of error making this method unreliable. Further methods were introduced to find more accurate curvature and torsion functions. Two methods which were investigated, one which sets the moving trihedron at various points on the curve and used formulae based on the tangent, normal and binormal vectors to calculate curvature and torsion, and

another which finds the curvature and torsion functions which optimise the accuracy of the reconstructed shape curve, produced curvature and torsion functions which allowed accurate reconstructions of the shape curve.

An issue with defining shape curves using curvature and torsion functions is that it is difficult to attach anatomical meaning to the functions. An alternative method is suggested where the space curves are projected onto two perpendicular planes and the curvature of these projected curves can be found using planar methods. The perpendicular planes are chosen so that bending in each plane corresponds to a simply definable change in anatomical direction. This method produces two curvature functions which are easily interpretable and define the shape of the curve up to similarity transformations of translation and rotation. As with plane curves reconstructing the original shape curve from the curvature functions gives a useful indicator of the accuracy of the functions. The reconstructed perpendicular plane curves can be combined by rotating the starting angle of the reconstructions to find a common axis. This technique is used to show that the perpendicular plane method produces curvature functions which give an accurate reconstruction of the original shape curve.

As mentioned above comparing curvature functions, and in particular average curvature functions, is a useful way to compare the shape of a feature for different groups. Piecewise averaging of both curvature functions is again trivial and the average curvature functions are used to produce average shape curves for simple comparability between groups. However, aligning curvature functions to produce structural averages is somewhat less straightforward. The major issue lies in the fact that it is important to ensure that the aligned arc length axis on both curvature functions for one subject correspond to the same point on the space curve. Therefore, any warping carried out on one curvature function must be carried out on the other curvature function. A choice must be made on whether one, both or a combination of the curvature functions are used to define important anatomical/geometrical points.

As a practical application, curvature functions, calculated using the perpendicular planes method, offer the basis of a comparison of the shape of the upper lip between the control children and the children with a cleft lip or cleft lip and palate in the 1 year study. A variety of mainly exploratory techniques are used for the comparison. The shape of average curvature functions and the average reconstructed shape curves give a visualisation of differences between the groups while analysis of the position and magnitude curvature at points of interest give

a more formal analysis of differences between the groups. Principal components analysis is shown to be a useful tool to investigate variation in the curvature functions where component score plots can indicate if variation described by a certain component is dominated by a certain group. In producing structural averages all curves were aligned to the average position of each point of interest. Considering that the ultimate aim of surgery is to remove all traces of the cleft, the structural average is produced for just the control group to give a baseline (or *gold standard*) curvature function and the position warping required to align each individual function to the average of the control group is quantified. Once all curvature functions are aligned the major difference between them is in the magnitude of curvature thus it is interesting to quantify the amplitude adjustment which would be required to produce the structural average control curvature function from each aligned function. Comparing the magnitude of both position warping and amplitude adjustment gives an interesting insight into which group has curvature functions most like the *gold standard*.

The practical application produced some interesting results. Investigating the upper lip, the major interest occurs in the middle of the upper lip where control children experience a turn down the face, sharply followed by a turn up the face, sharply followed by a turn down the face. This area is termed the Cupid's bow. There is evidence that the start of the Cupid's bow occurs nearer the start of the upper lip and the end of the Cupid's bow occurs further along the upper lip for children born with a cleft lip and children born with a cleft lip and palate than for control children. Perhaps of more interest there is evidence that neither children born with a cleft lip or those born with a cleft lip and palate have the upturn in the middle of the Cupid's bow and those children have lower curvature down the face on average at the ends of the Cupid's bow than control children. Therefore it seems that cleft children have a flat area in the middle of the upper lip where control children have their Cupid's bow. The cleft children also require more position and amplitude warping on average to produce the *gold standard* curvature functions. It is clear that there is a large amount of evidence that on average cleft children (both those born with a cleft lip and those born with a cleft lip and palate) have very different upper lip shapes than control children.

In summary, curvature functions have been shown to be useful in describing the shape of feature curves. A perpendicular planes method is an appropriate way to describe space curves as the curvature functions produced offer clear anatomical interpretation. The curvature functions are useful to compare the

shape of feature curves across different groups. An application comparing the upper lips of control children with children born with a cleft lip and children born with a cleft lip and palate at 1 year of age showed clear and anatomically interesting differences between the groups.

## 8.2 Functional Regression

There is a huge amount of literature on the problem of investigating relationships between a response and a predictor. The relationship between a scalar response and scalar predictor can be examined using standard linear regression techniques. It is further possible to examine the relationship between a scalar response and a functional predictor; however, this introduces some complexities. These complexities do not just affect the formal regression analysis but also graphical data description techniques.

The major issue with graphical displays is that, unlike the scalar case, there is no natural ordering of the predictors. One method for imposing ordering on the functional predictors is to carry out a functional principal components analysis and order according to the first principal component scores. Informally investigating the relationship between first principal component score and response may give an indication of the relationships in the functional regression. Visualising the effect on the mean functional predictor of changing the first principal score further increases understanding of potential relationships. A useful application of the `rpanel` tool allows a panel to be produced which investigates the effect on the response of changing the first principal component score and in turn visualises the curvature function and the estimated reconstruction of the feature which corresponds to that component score.

Two methods are suggested for formally applying the functional regression. A functional linear model is a simple extension of the standard linear model where a functional parameter is calculated for the functional predictor. Using least squares methods to estimate the functional parameter is not trivial as the continuous nature of the parameter allows a sum of squared error of zero. To allow least squares methods to be used the functional predictors and the functional parameters must be regularised. One method for doing this is to interpolate the predictors and consider a multivariate linear regression using the value of the function at the interpolation points as the predictors. However, this method loses information from the functional predictors, particularly when the sample size is

small. A preferable alternative is to regularise the functions using basis functions (in this thesis B-spline bases are used). Some simple algebra produces a least squares solution of the functional parameter in terms of coefficients of the basis functions.

The major interest of the functional linear model comes from the interpretation of the functional predictor. However, using least squares often results in a functional predictor which is too ‘rough’ to offer simple interpretation so some form of smoothing should be applied. This can be done, somewhat artificially, by reducing the number of bases functions used to define the functional parameter, however this has the undesirable effect of introducing inaccuracies in the basis representations of the functional predictors. An alternative solution is to penalise the residual sum of squares using a term based on the second derivative of the functional predictor with a smoothing parameter controlling the trade-off between smoothness of the functional parameter and accuracy of the functional linear model. It is sufficient to manually select the value of the smoothing parameter; however, a cross-validation procedure can be used to find the value which gives the ‘best’ trade off between smoothness and accuracy.

Nonparametric regression techniques can also be extended to deal with functional predictors. The method estimates a response given the functional predictor as a weighted average of the known responses. The responses are weighted according to the ‘closeness’ between their functional predictor and the functional predictor of the unknown response. Closeness can be measured using a semi-metric or a metric. A semi-metric is chosen if functions which differ in magnitude but are exactly the same in shape are to be given high weights. Throughout this thesis the magnitude of the functions was considered of major importance so the  $L_2$  metric was applied to quantify closeness. Asymmetric kernels are used to give weights to the responses depending on the measure of closeness between the functional predictors. As usual with kernel functions the choice of bandwidth is crucial. In this thesis the bandwidth is selected using a technique called nearest neighbours where a predetermined number of responses, with the closest functional predictors, are given non-zero weight.

Clearly it is important to have a method for formally testing the relationships shown by the functional linear model and nonparametric regression. A simple extension of the test of no effect proposed by Bowman and Azzalini (1997) can be applied to test the relationships. The test is based on a pseudo-likelihood ratio statistic which is proportional to the usual  $F$  statistic but is not dependent on

the degrees of freedom of the two models.

The techniques used to investigate the relationship between a scalar response and functional predictor can be extended to allow for multiple predictors with relative ease. Ordering the functional predictors with respect to their first principal component score enables informative graphical displays of potential relationships between multiple functional predictors and the response. The extension of the functional linear model to incorporate multiple predictors is trivial. The smoothing parameters for each of the functional predictors need not be the same. However, unless there is reason to use different smoothing parameters, it seems sensible to keep the smoothing parameter common across the predictors.

Extending nonparametric regression to incorporate multiple predictors is less straightforward. Each known response is weighted according to a combination of the closeness of all functional predictors. This weight can be found as the product of individual kernel weights for each functional predictor. The bandwidth is again chosen so that a pre-determined number of responses with the closest set of functional predictors are given non-zero weight. Clearly for the product of the individual kernel weights to be non-zero all individual weights have to be non-zero. To avoid scale issues dominating the selection of nearest neighbours the closeness measure for each predictor were standardised.

With multiple functional predictors it is possible to define a functional version of an additive model. The analogue of the smoothing of the partial residuals in a standard additive model is using the weights given by closeness between the functions to estimate the partial residuals. A natural extension of the backfitting algorithm can be used to find the value returned by each function in the final model. Again the functional nature of the predictors makes the process difficult to visualise and the effect of each predictor difficult to display. However, a plot of first principal component score against the value returned by the function in the additive model can be useful to investigate the effect of the function on the response in addition to the other functional predictors.

It is important to be able to test the significance of each functional predictor in addition to the other functional predictors in all three models; functional linear model, nonparametric regression and additive model. A simple and natural extension of the test of no effect for one variable gives a method for these tests. This method can be incorporated into a stepwise selection procedure to find a model which contains only significant functional predictors.

Functional regression techniques were applied to the data from the 10 year

study to investigate the relationship between psychological score and facial shape, defined by curvature functions, for cleft children. The single functional predictor which appeared to have the strongest relationship with psychological score related to bending up and down the face of the nasal base. The suggestion is that cleft children who have a more rounded nasal base at landmark *sn* and have this landmark occurring in the middle of the nasal base have lower psychological score than those with a flat area and a skewed position of the *sn* landmark. However, the relationship between this single functional predictor and psychological score is not strong.

When models with multiple functional predictors were considered the functions corresponding to bending up and down the face of the nasal base, bending up and down the face of the nasal rim and bending in and out the face of the nasal bridge were returned by the stepwise selection method for both the functional linear model and nonparametric regression whilst the functional additive model returned a model containing the functions of bending up and down the face of the nasal base and bending up and down the face of the nasal rim. The functional linear model also contained a function of bending up and down the face of the nasal bridge whilst the nonparametric regression also contained a function of bending across the face of the midline profile. The multiple functional linear model appeared to be a better predictor of psychological score than the nonparametric regression or additive model. A further advantage of the multiple functional linear model is that analysing the functional parameters gave a very interesting insight into the effect of the shape of the curvature functions, and in turn the shape of the original features, on the psychological score.

In summary, functional regression techniques are useful to investigate the relationship between a scalar response and functional predictor. An interesting challenge is in how best to display the data to both obtain an initial impression and then to further analyse significant results. The technique used in this thesis involved ordering the functional predictors according to their principal component scores. The `rpanel` displays provided a very helpful means for investigating a change of component score both on the response and the functional predictor. Using curvature functions in the functional regression models produced some interesting results. From the 10 year study it appears that there is evidence that the psychological state of cleft children may be mostly related to the bending up and down the face of the nasal base and nasal rim. The effect of these predictors can be investigated using both `rpanel` and the functional parameters from the

functional linear model. However, while suggestions of the psychological effect of certain anatomical phenomenon were made in the thesis where possible, this is by no means the expert opinion which is necessary.

### 8.3 Potential for Further Work

The work in this thesis touches the surface of two major issues. Firstly, finding an appropriate method for defining the shape of an object and secondly using this measure of shape to investigate relationships with a response. There is room to build on methods used throughout this thesis and also a number of interesting areas in which the work could move.

Barry (2008) carried out longitudinal analysis on landmarks and curves from the faces of children. Procrustes methods were used for alignment before the analysis. As an interesting extension it may be possible to carry out longitudinal analysis of curvature functions as opposed to the shape curves. This could return information on how the curvature of facial features change throughout time and if these changes are similar for control and cleft children.

Standard practice for positioning landmarks, and therefore producing shape curves, is for manual selection by trained operators. However, it appears that it may be possible to extract curves from a three-dimensional image using changes in surface curvature. Discussion of this curve extraction method is outwith the scope of this thesis; however, if curves can be extracted automatically curvature functions may provide a useful tool in landmarks placement. When aligning curvature functions in the thesis it was said that turning points in the curvature functions often correspond to landmarks of interest. It would be interesting to select the position of landmarks using curvature functions and compare these to the positions which are manually selected by an operator. Combined with a curve extraction method based on surface curvature this could potentially lead to a fully automated system for extracting curves and landmarks.

The investigation into the relationship between facial shape and psychological score produced some interesting results. However, it may be possible to improve and extend the analysis. It appeared that the nonparametric regression was less effective in finding relationships than the functional linear model. This may have been due to a rather simple metric being chosen to measure closeness between functions. As mentioned, any suitable metric can be chosen so time may be well spent considering and testing improved metric functions. Furthermore, it

would be both interesting and helpful to have some method for visualising the nonparametric regression process in the way that the functional linear model returns informative functional parameters.

In the 10 year study both children and parents were asked to respond to questionnaires; however, this was only requested for cleft children. It would have been interesting to have psychological information for control children and to analyse whether the facial shape of control children is related to psychological score. In fact if enough data were available, comparisons in the results of a functional regression analysis between the control group and both cleft groups (those born with a cleft lip and those born with a cleft lip and palate) may provide interesting insights. In the analysis in this thesis, psychological score was measured according to the results for a questionnaire completed by the children's parents. It would clearly be useful and interesting to repeat the analysis using results from questionnaires completed by the children themselves.

A major issue in the functional regression was how best to display both the initial data and the results. The `rpanel` displays can be used to visualise the effect of changing principal component scores, not only in terms of the response but, on the effect on the mean curvature function and in turn the reconstructed shape curve. With some work it may be possible to produce an `rpanel` which investigates the effect of the functional parameters from the functional linear model. Throughout this thesis there was some difficulty in identifying clear effects of the functional parameter but by changing the shape of the mean curvature function, and in turn the corresponding shape curve, in directions deemed to affect the response by the functional parameter, interpretation may become clearer. How this would work in an `rpanel` requires much thought but the resulting display would be very informative.

In the functional regression, facial shape is defined by curvature functions. It has already been mentioned that there are a variety of methods to define facial shape and any of these methods can be used in a regression setting. One interesting analysis could be to use the work of Bowman and Bock (2006) to produce asymmetry measures for all the landmarks of each subject. Some combination of these asymmetry measures could form part of standard linear regression of the psychological score. This is just one example of a number of similar regression analyses that could be used to relate facial shape to psychological score and a systematic comparison of the methods to define facial shape in the regression analysis may be useful.

In summary, there are a number of natural extensions to the work in this thesis. Considering the relationship between facial shape and psychological score of cleft children is an extremely worthwhile pursuit and may in future prove useful to surgeons and psychologists. How best to define facial shape and psychological score is certainly open to debate and whilst this thesis suggests certain methods a systematic analysis of various techniques may in fact provide improved methods. An improvement and extension of the methods for displaying the results of a functional regression may be helpful in allowing access of the important results to surgeons and psychologists who offer help to the children.

# Bibliography

- Alkhodre, A., Belarbi, M., Langs, G., Tosovic, S. and Sicard, N. (2001), Shape description via fourier analysis, *in* ‘ERASMUS Intensive Program’.
- Alshabani, A., Dryden, I. and Litton, C. (2007), ‘Bayesian analysis of human movement curves’, *Journal of Applied Statistics* **56**, 415–428.
- Apanasovich, T. and Goldstein, E. (2008), ‘On prediction error in functional linear regression’, *Statistics & Probability Letters* **78**, 1807–1810.
- Aykroyd, R. and Mardia, K. (2003), ‘A wavelet approach to shape analysis for spinal curves’, *Journal of Applied Statistics* **30**, 605–623.
- Ayoub, A., Garrahy, A., Hood, C., White, J., Bock, M., Siebert, J., Spencer, R. and Ray, A. (2003), ‘Validation of a vision-based, three-dimensional facial imaging system.’, *The Cleft Palate-Craniofacial Journal* **40**, 523–529.
- Baíllo, A. and Grané, A. (2007), ‘Local linear regression for functional predictor and scalar response’, *Statistics and Econometric Series* **15**, 1–15.
- Barry, S. (2008), Longitudinal Analysis of Three-dimensional Facial Shape Data, PhD thesis, University of Glasgow.
- Besse, P., Cardot, H., Faivre, R. and Goulard, M. (2005), ‘Statistical modelling of functional data’, *Applied Stochastic Models in Business and Industry* **21**, 165–173.
- Bock, M. and Bowman, A. (2006), ‘On the measurement and analysis of asymmetry with applications to facial modelling’, *Applied Statistics* **55**, 77–91.
- Bookstein, F. (1989), ‘Principal warps: Thin-plate splines and the decomposition of deformations’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 567–585.

- Bosq, D. (1991), Modelization, non-parametric estimation and prediction for continuous time processes, in G. Roussas, ed., 'Nonparametric Functional Estimation and Related Topics', Kluwer Academic Publishers, pp. 509–529.
- Bowman, A. and Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations.*, Oxford University Press.
- Bowman, A. and Bock, M. (2006), 'Exploring variation in three-dimensional shape data', *Journal of Computational and Graphical Statistics* **15**, 524–541.
- Boyce, W. E. and DiPrima, R. C. (1992), *Elementary Differential Equations*, John Wiley & Sons.
- Brent, R. (1973), *Algorithms for Minimization Without Derivatives*, Prentice-Hall.
- Cai, T. and Hall, P. (2006), 'Prediction in functional linear regression', *The Annals of Statistics* **34**, 2159–2179.
- Cardot, H., Crambes, C., Kneip, A. and Sarda, P. (2007), 'Smoothing splines estimators in functional linear regression with errors-in-variables', *Computational Statistics & Data Analysis* **51**, 4832–4848.
- Cardot, H., Ferraty, F. and Sarda, P. (1999), 'Functional linear model', *Statistics and Probability Letters* **45**, 11–22.
- Cardot, H., Goia, A. and Sarda, P. (2004), 'Testing for no effect in functional linear regression models, some computational approaches', *Communications in Statistics - Simulation and Computation* **33**, 179–199.
- Cardot, H. and Sarda, P. (2006), Linear regression models for functional data, in 'The Art of Semiparametrics', Physica-Verlag HD, pp. 49–66.
- Cardot, H. and Sarda, P. (2008), 'Varying-coefficient functional linear regression models', *Communications in Statistics - Theory and Methods* **37**, 3186–3203.
- Chong, M., Gay, R., Tan, H. and Liu, J. (1992), 'Automatic representation of fingerprints for data compression by b-spline functions', *Pattern Recognition* **25**, 1199–1210.

- Cleveland, W. (1979), 'Robust locally-weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**, 829–836.
- Cleveland, W. (1981), 'Lowess: A program for smoothing scatterplots by robust locally weighted regression', *The American Statistician* **35**, 54.
- Cootes, T., Taylor, C., Cooper, D. and Graham, J. (1995), 'Active shape models - their training and application', *Computer Vision and Image Understanding* **61**, 38–59.
- Costa, L. F., dos Reis, S., Arantes, R., Alves, A. and Mutinari, G. (2004), 'Biological shape analysis by digital curvature', *Pattern Recognition* **37**, 515–524.
- Craizer, M., Pesco, S. and Teixeira, R. (2005), 'A numerical scheme for the curvature equation near the singularities', *Journal of Mathematical Imaging and Vision* **22**, 89–95.
- Cuevas, A., Febrero, M. and Fraiman, R. (2002), 'Linear functional regression: The case of fixed design and functional response', *The Canadian Journal of Statistics* **30**, 285–300.
- Decker, L., Berge, C., Renous, S. and Penin, X. (2007), 'An alternative approach to normalization and evaluation for gait patterns: Procrustes analysis applied to the cyclograms of sprinters and middle-distance runners', *Journal of Biomechanics* **40**, 2078–2087.
- Dryden, I., Kume, A., Le, H. and Wood, A. (2008), 'A multi-dimensional scaling approach to shape analysis', *Biometrika* **95**, 779–798.
- Dryden, I. and Mardia, K. (1998), *Statistical shape analysis*, Wiley.
- Eilers, P. and Marx, B. (1996), 'Flexible smoothing using b-splines and penalized likelihood', *Statistical Science* **11**(2), 89–121.
- Elander, J. and Rutter, M. (1996), 'Use and development of the rutter parents' and teachers' scales', *International Journal of Methods in Psychiatric Research* **6**, 63–78.
- Faraway, J. (1997), 'Regression analysis for a functional response', *Technometrics* **39**, 254–261.

- Faraway, J., Reed, M. and Wang, J. (2007), ‘Modelling three-dimensional trajectories by using bézier curves with application to hand motion’, *Journal of Applied Statistics* **56**, 571–585.
- Ferraty, F., Mas, A. and Vieu, P. (2007), ‘Nonparametric regression on functional data: Inference and practical aspects’, *Australian & New Zealand Journal of Statistics* **49**, 267–286.
- Ferraty, F. and Vieu, P. (2006), *Nonparametric Functional Data Analysis*, Springer.
- Fredman, M. and Willard, D. (1994), ‘Tree-dichotomous algorithms for minimum spanning trees and shortest paths’, *Journal of Computer and System Sciences* **48**, 533–551.
- Frost, F. and Wolstenholme, J. (2001), *A Treatise on Solid Geometry*, Adamant Media Corporation.
- Gasser, T., Kneip, A. and Ziegler, P. (1990), ‘A method for determining the dynamics and intensity of average growth’, *Annals of Human Biology* **17**, 459–474.
- Gervini, D. and Gasser, T. (2004), ‘Self-modelling warping functions’, *Journal of the Royal Statistical Society. Series B* **66**, 959–971.
- Gervini, D. and Gasser, T. (2005), ‘Nonparametric maximum likelihood estimation of the structural mean of a sample of curves’, *Biometrika* **92**, 801–820.
- Gray, A. (1998), *Modern Differential Geometry of Curves and Surfaces with Mathematica*, CRC Press.
- Green, P. and Silverman, B. (1994), *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall.
- Guliatto, D., Rangayyan, R., Carvalho, J. and Santiago, S. (2008), ‘Polygonal modelling of contours of breast tumors with the preservations of spicules’, *IEEE Transactions on Biomedical Engineering* **55**, 14–20.
- Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall.

- Hausrath, A. and Goriely, A. (2007), ‘Continuous representation of proteins: Construction of coordinate models from curvature profiles’, *Journal of Structural Biology* **158**, 267–281.
- James, G. (2002), ‘Generalized linear models with functional predictors’, *Journal of the Royal Statistical Society. Series B* **64**, 411–432.
- James, G. and Silverman, B. (2005), ‘Functional adaptive model estimation’, *Journal of the American Statistical Association* **100**, 565–576.
- Johnson, N. and Kotz, S. (1972), *Distributions in Statistics: Continuous Univariate Distributions, Vol II.*, Wiley.
- Jolliffe, I. (1986), *Principal Component Analysis*, Springer.
- Karousos, E., Ginnis, A. and Kaklis, P. (2008), Controlling torsion sign, in ‘Advances in Geometric Modeling and Processing’, Springer, pp. 92–106.
- Khambay, B., Nairn, N., Bell, A., Miller, J., Bowman, A. and Ayoub, A. (2008), ‘Validation and reproducibility of a high-resolution three-dimensional facial imaging system’, *British Journal of Oral and Maxillofacial Surgery* **46**, 27–32.
- Kneip, A. and Gasser, T. (1992), ‘Statistical tools to analyze data representing a sample of curves’, *The Annals of Statistics* **20**, 1266–1305.
- Korn, G. and Korn, T. (1968), *Mathematical handbook for scientists and engineers :definitions, theorems, and formulas for reference and review*, McGraw-Hill.
- Kume, A., Dryden, I. and Le, H. (2007), ‘Shape-space smoothing splines for planar landmark data’, *Biometrika* **94**, 513–528.
- Lewiner, T., Gomes Jr, J., Lopes, H. and Craizer, M. (2005), ‘Curvature and torsion estimators based on parametric curve fitting’, *Computer & Graphics* **29**, 641–655.
- Li, Y. and Hsing, T. (2007), ‘On rates of convergence in functional linear regression’, *Journal of Multivariate Analysis* **98**, 1782–1804.
- Mokhtarian, F. and Mackworth, A. (1992), ‘A theory of multiscale, curvature-based shape representation for planar curves’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**, 789–805.

- Müller, H.-G. and Stadtmüller, U. (2005), ‘Generalized functional linear models’, *The Annals of Statistics* **33**, 774–805.
- Munsell, B., Dalal, P. and Wang, S. (2008), ‘Evaluating shape correspondence for statistical shape analysis: A benchmark study’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**, 2023–2039.
- Niblack, W. and Yin, J. (1995), A pseudo-distance measure for 2D shape based on turning angle, in ‘ICIP ’95: Proceedings of the 1995 International Conference on Image Processing’, IEEE Computer Society, p. 3352.
- Noriega, A., Tocino, A. and Cervantes, E. (2008), Hydrogen peroxide treatment results in reduced curvature values in the arabidopsis root apex. Article in press with the Journal of Plant Physiology.
- Pajdla, T. and Van Gool, L. (1995), Efficient matching of space curves, in ‘CAIP ’95: Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns’, Springer-Verlag, pp. 25–32.
- Pavlidis, T. and Horowitz, S. (1974), ‘Segmentation of plane curves’, *IEEE Transactions on Computers* **23**, 860–870.
- Pazos, V., Cheriet, F., Danserau, J., Ronsky, J., Zernicke, R. and Labelle, H. (2007), ‘Reliability of trunk shape measurements based on 3-d surface reconstructions’, *European Spline Journal* **16**, 18821891.
- Pressley, A. (2001), *Elementary Differential Geometry*, Springer.
- Ramsay, J. and Li, X. (1998), ‘Curve registration’, *Journal of the Royal Statistical Society. Series B* **60**, 351–363.
- Ramsay, J. and Silverman, B. (1997), *Functional Data Analysis*, Springer.
- Ramsay, J. and Silverman, B. (2006), *Functional Data Analysis - Second Edition*, Springer.
- Reinsch, C. (1967), ‘Smoothing by spline functions’, *Numerische Mathematik* **10**, 177–183.
- Rice, J. and Silverman, B. (1991), ‘Estimating the mean and covariance structure nonparametrically when the data are curves’, *Journal of the Royal Statistical Society Series B* **55**, 233–243.

- Rieger, B. and van Vliet, L. (2002), ‘Curvature of n-dimensional space curves in grey-value images’, *IEEE Transactions on Image Processing* **11**, 738–745.
- Rønn, B. (2001), ‘Nonparametric maximum likelihood estimation for shifted curves’, *Journal of the Royal Statistical Society. Series B* **63**, 243–259.
- Rosin, P. and West, G. (1995), ‘Curve segmentation and representation of super-ellipses’, *IEE Proc.-Vis. Image Signal Process.* **142**, 280–288.
- Rutter, M. (1967), ‘A children’s behaviour questionnaire for completion by teachers: Preliminary findings’, *Journal of Child Psychology and Psychiatry* **8**, 1–11.
- Samir, C., Srivastava, A. and Daoudi, M. (2006), ‘Three-dimensional face recognition using shapes of facial curves’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 1858–1863.
- Shen, L., Rangayyan, R. and Leo Desautels, J. (1994), ‘Application of shape analysis to mammographic calcifications’, *IEEE Transactions on Medical Imaging* **13**, 263–274.
- Shi, J. and Wang, B. (2008), ‘Curve prediction and clustering with mixtures of gaussian process functional regression models’, *Statistical Computing* **18**, 267–283.
- Shi, J., Wang, B., Murray-Smith, R. and Titterton, D. (2007), ‘Gaussian process functional regression modelling for batch data’, *Biometrics* **63**, 714–723.
- Small, C. and Le, H. (2002), ‘The statistical analysis of dynamic curves and sections’, *Pattern Recognition* **35**, 1597–1609.
- Srivastava, D. and Lisle, R. (2004), ‘Rapid analysis of fold shape using bézier curves’, *Journal of Structural Geology* **26**, 1553–1559.
- Steele, J., Shepp, L. and Eddy, W. (1987), ‘On the number of leaves of a euclidean minimal spanning tree’, *Applied Probability* **24**, 809–826.
- Tang, R. and Müller, H.-G. (2008), ‘Pairwise curve synchronization for functional data’, *Biometrika* **95**, 875–889.
- Tieng, Q. and Boles, W. (1994), ‘Recognition of space curves based on the dyadic wavelet transform’, *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference* pp. 137–140.

- Trotman, C.-A., Faraway, J. and Phillips, C. (2005), ‘Visual and statistical modeling of facial movement in patients with cleft lip and palate’, *The Cleft Palate-Cranofacial Journal* **42**, 245–254.
- Wang, K. and Gasser, T. (1997), ‘Alignment of curves by dynamic time warping’, *The Annals of Statistics* **25**, 1251–1276.
- Wang, K. and Gasser, T. (1999), ‘Synchronizing sample curves nonparametrically’, *The Annals of Statistics* **27**, 439–460.
- Worring, M. and Smeulders, A. (1992), The accuracy and precision of curvature estimation methods, *in* ‘Pattern Recognition, 1992. Vol.III. Conference C: Image, Speech and Signal Analysis, Proceedings., 11th IAPR International Conference’, pp. 139–142.
- Yang, X., Shen, Q., Xu, H. and Shoptaw, S. (2007), ‘Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures’, *Statistics in Medicine* **26**, 1552–1566.
- Yao, F., Muller, H.-G. and Wang, J.-L. (2005), ‘Functional linear regression analysis for longitudinal data’, *The Annals of Statistics* **33**, 2873–2903.

# Appendix A

## Revised Rutter Questionnaire

## REVISED RUTTER PARENT SCALE FOR SCHOOL-AGE CHILDREN

Child's name: \_\_\_\_\_

Age: \_\_\_\_\_

Below are a series of descriptions of behaviour often shown by children. After each statement are three columns: *Does not apply*, *Applies somewhat* and *Certainly applies*. If your child definitely shows the behaviour described by the statement, place a cross in the box under column 3 *Certainly applies*. If your child shows the behaviour described by the statement but to a lesser degree or less often, place a cross in the box under column 2 *Applies somewhat*. If, as far as you are aware, your child does not show the behaviour, place a cross in the box under column 1 *Does not apply*.

Please complete on the basis of your child's behaviour *during the past three months*.

Put *one* cross against *each* statement. Thank you.

This statement . . .

1. Tries to be fair in games
2. Very restless, has difficulty staying seated for long
3. Considerate of other people's feelings
4. Squirmy, fidgety child
5. Often destroys or damages own or others' property
6. Has had tears on arrival at school or has refused to go into the building in the past 12 months
7. Will try to help someone who has been hurt
8. Frequently fights or is extremely quarrelsome with other children
9. Gives up easily
10. Not much liked by other children
11. Volunteers to help around the house or garden
12. Often worried, worries about many things
13. Tends not to finish things started, short attention span
14. Spontaneously affectionate to family members
15. Tends to be on own, rather solitary

|  | Does not apply           | Applies somewhat         | Certainly applies        |
|--|--------------------------|--------------------------|--------------------------|
| 1. Tries to be fair in games   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Very restless, has difficulty staying seated for long   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. Considerate of other people's feelings  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. Squirmy, fidgety child  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. Often destroys or damages own or others' property   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. Has had tears on arrival at school or has refused to go into the building in the past 12 months | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 7. Will try to help someone who has been hurt  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 8. Frequently fights or is extremely quarrelsome with other children                               | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 9. Gives up easily   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 10. Not much liked by other children   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 11. Volunteers to help around the house or garden  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12. Often worried, worries about many things   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 13. Tends not to finish things started, short attention span                                       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 14. Spontaneously affectionate to family members   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 15. Tends to be on own, rather solitary  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |



2

This statement . . .

|   | <i>Does not apply</i>    | <i>Applies somewhat</i>  | <i>Certainly applies</i> |
|---|--------------------------|--------------------------|--------------------------|
| 16. Irritable, touchy, is quick to 'fly off the handle'               | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 17. Kind to younger children  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 18. Often appears miserable, unhappy, tearful or distressed           | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 19. Resentful or aggressive when corrected                            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 20. Blames others for things  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 21. Comforts a child who is crying or upset                           | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 22. Has a stutter or stammer  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 23. Has other speech difficulty                                       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 24. Truants from school   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 25. Has twitches, mannerisms, or tics of the face and body            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 26. Frequently sucks thumb or finger                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 27. Gets on well with other children                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 28. Has stolen things on more than one occasion in the past 12 months | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 29. Cries easily  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 30. Frequently bites nails or fingers                                 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 31. Is often disobedient  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 32. Tries to stop quarrels or fights                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 33. Has wet or soiled self this year                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 34. Cannot settle to anything for more than a few moments             | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 35. Forceful, determined child  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |



3

Does not apply  
Applies somewhat  
Certainly applies

This statement . . .

- 36. Shares out treats with friends
- 37. Tends to be fearful or afraid of new things or new situations
- 38. Kicks or bites other children
- 39. Stares into space, stares blankly
- 40. Plays imaginatively, enjoys 'pretend' games
- 41. Fussy, or over-particular child
- 42. Inattentive, easily distracted
- 43. Independent, confident child
- 44. Doesn't share toys
- 45. Helps other children who are feeling ill
- 46. Often tells lies
- 47. Bullies other children
- 48. Kind to animals
- 49. Often complains of aches or pains
- 50. Inconsiderate of others

Completed by: \_\_\_\_\_ Date of completion: \_\_\_\_\_

Signed: \_\_\_\_\_ Thank you for your help in this study.



© Michael Rutter, 1993. The *Revised Rutter Scales* by Michael Rutter. Reproduced by kind permission of the author. The scales come in parent and teacher versions for two age groups – preschool and school-age. The measures derive from the questionnaires first developed by Michael Rutter and William Yule; these versions contain certain items developed by and reproduced with permission of Kirk Weir and Robert Goodman, and further items developed in the USA by Lenore Behar and Samuel Stringfield.

This measure is part of *The Child Psychology Portfolio* edited by Irene Sclare. Once the invoice has been paid, it may be photocopied for use **within the purchasing institution only**. Published by nferNelson Publishing Company Ltd, The Chiswick Centre, 414 Chiswick High Road, London W4 5TF, UK. Code 0090006336