THE ESTIMATION OF EDGES IN MAGNETIC RESONANCE IMAGES

ALISON V. WHEELWRIGHT

Presented for the degree of
Doctor of Philosophy

University of Edinburgh

1992

ACKNOWLEDGEMENTS

I would first like to thank the many people who have supported me over the last four years. Most of all to my parents and sister, for their constant interest and the knowledge that they were always there. Secondly, to Betty Heyburn for her encouragement, invaluable practical help and not least her removal skills!

For their friendship and company in the Scottish hills I would like to thank Helen Boden, Anne Burnand, and Ian Cook. Thanks are also due to Ian for convincing me I needed a mountain bike! I am grateful to Alvin Milner for his help, the many cups of tea and for uncomplainingly putting up with sharing a flat over the last year. I would also like to thank Deena Mobbs and Jose Seraphin for their moral support, advice and willingness to listen, and Howard Hughes for proof-reading this thesis.

For providing me with a chance to escape from the thesis I would like to thank my friends from the Scottish Agricultural Statistics Service; Florence, Joan, Joyce, Olwyn and Rosa for the Friday afternoons working in the Womens Royal Voluntary Service hospital canteen; Betty and Hugh Purves and family for the Stonehaven weekends; Audrey and Alan Hutchinson for their interest continuing from when I was in Newcastle ..... and Bryan Adams for his company during the long hours word-processing.

For their suggestions and assistance during the preparation of this thesis I would like to thank my two supervisors, Dr Colin Aitken and Dr Chris Glasbey. I would also like to thank Dr Graham Horgan for his practical help at the start of the project; Dr Geraldine McNeill of the Rowett Research Institute, Aberdeen and Drs Margeret Foster and Paul Fowler of the Deparment of Bio-Medical Physics and Bioengineering, University of Aberdeen for making available the data analysed in this thesis, and thanks again to Dr Margeret Foster for correcting the technical content of Chapter 2.

I am grateful to the Science and Engineering Research Council for funding me throughout this research.

ABSTRACT

This thesis is concerned with the estimation of edges in magnetic resonance images (MRI), which may be seen as a first step in the automatic classification of such data. The estimation is taken as a two-stage process. A set of points lying on a single edge is first identified. Secondly, some form of closed curve is fitted to this set of points to describe the edge.

The data analysed in this study are MRI of cross-sections through human thighs. Although the subject of the images exists in continuous two-dimensional space, in practice data values are only recorded at discrete, sampled points. This is due to quantisation of the underlying continuous function for storage on a computer. A major theme for this study is the recovery of the underlying continuous function from the sampled data: it is expected that this will allow edges to be estimated more accurately.

Bivariate kernel regression is used in the first stage to fit a smooth function to the observed data. Edge points are identified as positions of zero-crossings of the smoothed function. The accuracy with which edge points are located is influenced by the amount of smoothing, and several data-based methods are discussed for estimating an appropriate smoothing parameter.

In the second stage, an edge is modelled as a simple, closed curve by fitting a Fourier series (FS) to the set of edge points. Geometric properties, such as perimeter length, can be determined from the fitted series. The accuracy of the estimation of such properties is used as a criterion to determine the number of terms to be included in the series. The choice of variable with which to label consecutive points prior to fitting the FS is also discussed. Using simulated data, one labelling variable is analysed theoretically and the distribution is derived of the lengths of the intervals (defined as the Euclidean distance) between consecutive edge points, in order to identify the optimum variable.

PUBLICATIONS

1. A summary of the work on kernel regression and the estimation of a smoothing parameter, covered in Chapters 3 and 4 respectively, was presented under the title: 'The application of kernel regression to the detection of tissue boundaries in magnetic resonance images', by Wheelwright, A.V., Aitken, C.G.G. and Glasbey, C.A., at the 15th International Biometric Conference, Hungary, July 1990.

2. Part of the work on the distribution of interval lengths which appears in Chapter 7 is to be published under the title: 'Distances between censored intersections between a square lattice and a random smooth path', by Wheelwright, A.V. and Glasbey, C.A., in Journal of Applied Probability, March 1993.

CONTENTS

3 KERNEL REGRESSION FOR THE IDENTIFICATION OF EDGES

4 METHODS FOR SMOOTHING PARAMETER ESTIMATION

PREFACE

The purpose of the project is the estimation of edges in magnetic resonance images (MRI), taken as a two-stage process. In the first stage, statistical methods are used to identify the locations of a set of edge points. The second stage is the fitting of some form of smooth curve to describe the edge. This can be seen as a step towards 'automatic' analysis of such data. These methods of estimating edges are less subjective and less labour-intensive than the methods currently used: a human operator views the image on a computer screen and, uses a mouse to move a cursor round the screen, so outlining the edge.

Edges of objects in an image have been found to be one of the key factors for the human vision system, as will be discussed in Chapter 1. Therefore it is natural that for computer interpretation of such images, the edges should similarly be one of the first features extracted from the image. The approach taken combines relevant techniques from two complementary fields. The analysis of such images has traditionally been viewed as a problem belonging to the field of computer vision, where more emphasis is placed on how the image was generated and the physical properties causing the change in the observed data. Here statistical methods are used to give some theoretical justification and quantitative assessment of existing methods, and to improve the accuracy of estimation.

The underlying subject of the images typically exists in continuous two-dimensional space, but due to quantization for storage on a computer, the data are only recorded at discrete, sampled points. The image is represented as an array of picture elements or 'pixels'; for the current MRI the arrays are of size 128 x 128. The recovery of the continuous function underlying an image from the sampled data is a major theme for this study. It is expected that this will allow more accurate estimation of the parameters of interest.

The data analysed in this study are MRI of cross-sections through human thighs, obtained as part of a nutrition study at the Rowett

Research Institute, Aberdeen. Each image is of two legs. Tissues are nested and labelling from the inside are marrow, bone, muscle and fat. The identification of a set of edges between individual tissues has several applications. It may aid automatic classification of different tissue regions, or it may be used to screen for abnormalities by comparing a template of a typical healthy organ, such as the heart, with the corresponding (possibly diseased) region identified in a given image. Alternatively, attention may be focussed on the region enclosed by an edge. For example, in the nutrition study, the amount of body fat is of interest, so an estimate of the area of a particular (fat) region is required.

The data and magnetic resonance imaging are discussed in Chapter 2. The data values are recorded at each pixel in the image. The values correspond to an 'intensity': the definition of this term which holds in this thesis is given in Chapter 1. Each tissue has a different characteristic 'intensity' value. Data are available for three variables, though only two are studied here, since the third variable is a non-linear combination of the first two, and so does not provide any additional information. Such images are subject to sampling error or 'noise', that must be taken into account in any analysis.

In computer vision, the aim is to account for the variation in the observed data in terms of the physical characteristics of the underlying image. Separate features are often characterised by 'significant' changes in intensity. This leads to the definition of an edge as a 'significant' change in intensity, which assuming intensity is a continuous function, is mathematically equivalent to a maximum in the first derivative of the intensity function, or alternatively a 'zero-crossing' of the second derivative. In practice, such ideal edges are not observed since the function is sampled only at discrete data points, and further, is subject to random noise. One solution is to 'average' values over neighbouring pixels. This is termed 'smoothing' in the statistical literature: intensities at discrete data points are smoothed or 'spread-out' over neighbouring pixels.

For the first stage of the project, kernel regression is used (Chapter 3) to fit a smooth function to the observed data. The underlying function can be estimated at any point in continuous space. The Gaussian function is selected for the kernel model, and is easily extended for data in two dimensions. The Gaussian has been shown to be 'near-optimal', and has many favourable properties including simplicity and computational efficiency. Edge points are defined as positions of zero-crossings of the second derivative of the smoothed function. The orientation-independent Laplacian operator is used so only one set of derivatives needs to be calculated. Zero-crossings on individual edges are tracked sequentially to the nearest pixel, followed by interpolation to obtain sets of edge points in continuous space.

The amount of smoothing is an important factor for the kernel model and has considerable influence on how well the edges are estimated. Several 'automatic' data-based methods for estimating the appropriate amount of smoothing are considered in Chapter 4. A compromise must be reached between the degree of bias and variance for the set of edge points identified. Bias in the location of points occurs if data are oversmoothed and important local detail is lost, while the variance is due to insufficient smoothing so random fluctuations in the data result in spurious 'edges'.

The second stage of the project involves fitting a smooth, closed curve to points on individual edges. The aim is to describe the edge using a model with fewer parameters, rather than the full set of points identified as zero-crossings of the smoothed image. Since the edge points lie on a closed curve, this allows representation by a periodic function. Fourier descriptors (FD) are chosen as the simplest case, and are defined and fitted to one of the images in Chapter 5. Two aspects of the FD model require more detailed consideration, which motivates the theoretical analysis of Chapters 6 and 7. The results from these two chapters are used in the final section of Chapter 5, when several sets of FD are compared when fitting a Fourier series (FS) to points lying on a single edge in the image.

The first aspect of the FS model to be studied in more detail is the determination of the number of terms to be included in the FS. The data are fully described by the FS, but the complete fitted model contains as many parameters as the set of tracked (zero-crossing) edge points. This poses the question of whether it is possible to obtain a sufficiently accurate estimate of the edge when fitting only the first few terms in the FS. Methods for determining a suitable number of FD at which to truncate the FS are discussed in Chapter 6. Geometric properties, such as the perimeter length and area of a region enclosed by an edge, can be calculated from the set of FD. Defining a suitable truncation point for the FS is equivalent to identifying the subset of FD which best estimates the 'true' value of a given property. Because the true perimeter length and area are not available for the image data, simple geometric shapes with known properties are simulated. The simulated data are used to investigate how the number of terms to be included in the series is affected by changes in signal-to-noise ratio or in the size of the shape, where the size is measured by the area covered.

The choice of variable for labelling consecutive points on an edge is the second aspect of the FD model requiring further analysis. One option considered is labelling points in terms of 'cumulative distance' from an arbitrary start-point. This is extended to include a variant where points are interpolated to equal distances round the polygonal approximation of the edge. A second option is to label points in terms of 'index number', that is, the relative position of the current point in the ordered sequence round the edge. A variant of this second labelling aims to equalise the distance or 'interval' between points, by censoring points where the length of the interval between two consecutive points is less than a prespecified threshold value. Based on simulations of random lines intersecting a square lattice, the distribution of interval lengths with respect to different censoring thresholds is derived in Chapter 7. An optimum threshold value is identified.

A summary of the methods used for the estimation of edges in the MRI is given in Chapter 8.

NOTATION

a_r	Fourier coefficient, x coordinate, cos( ) term [5,6]
$\tilde{a}_\lambda$	constant weight on main diagonal of A_λ [4]
$a_\lambda(p,q)$	(single weight) typical element of matrix A_λ [4]
A	true area of region [6]
$\hat{A}$	general estimator of area, representing one of $\{A(r) : r = 1,2,\ldots,R, A(t)\}$ [6]
A_g	true area for g^{th} shape [6]
$\hat{A}_g$	general estimator of area for g^{th} shape [6]
A_s	general estimator of area for s^{th} simulation [6]
$\hat{A}_{gs}$	general estimator of area for s^{th} simulation of g^{th} shape [6]
$A(t)$	estimator of area calculated from tracked edge points [6]
$A(r)$	estimator of area calculated from Fourier series with $\{r = 1,2,\ldots,R\}$ harmonics [6]
A_λ	M x M matrix of weights $a_\lambda(p,q)$ for model in 'stacked vector' form [4]
b	size of smoothing window, corresponding to bandwidth of non-zero weights in A_λ) [3,4]
b_r	Fourier coefficient, x coordinate, sin( ) term [5,6]
c_r	Fourier coefficient, y coordinate, cos( ) term [5,6]
C_1	scaling probability [7]
C_t	scaling probability as a function of threshold t [7]
d_i	cumulative distance labelling variable [5]
d_i^*	cumulative distance labelling variable after equal distance interpolation [5]
$d_i(S_r)$	value of S_r for fitting to points labelled in terms of d_i [5]
$d_i^*(S_r)$	value of S_r for fitting to points labelled in terms of d_i^* [5]
d_n	'last' point on contour when labelled in terms of cumulative distance [5]
d_r	Fourier coefficient, y coordinate, sin( ) term [5,6]

D	total cumulative distance around single edge [5]
D/n	equal distance between points for cumulative distance interpolation [5]
$E_t(z\|\theta)$	expected value of interval length z, conditional on slope θ [7]
$E_t(z)$	expected value of interval length z [7]
$\hat{f}_\lambda(x-i, y-j)$	contribution from individual observation (i,j) to smoothed estimate at (x,y) [3]
$f_t(z\|\theta)$	conditional probability density function for distribution of intervals z [7]
$f_t(z)$	probability density function for distribution of intervals z [7]
$F(r^*)$	F-ratio for testing significance of r^*th harmonic [5]
g	number of different geometric shapes simulated [6]
G_x	horizontal difference calculated between pixels (over columns) [1]
G_y	vertical difference calculated between pixels (over rows) [1]
i	{i = 1,2,...,m} column coordinate of data site on lattice [1,2,3,4]
i	{i = 1,2,....,n} index/subscript of edge points on a single contour [3,5,6]
i	index number labelling variable [5]
i	subscript for labelling consecutive intersections of random path with lattice [7]
I'	'change of sign' intensity in relation between variables v1 and v2 [2]
I	indicator function for whether given value of z falls within specified range [7]
j	{j = 1,2,...,m} row coordinate of data site on lattice [1,2,3,4]
k	integer intensity after rescaling observed data to integer-incremental scale [2]
k	integer multiple of λ for truncation of smoothing region [4]

K	kernel function [1,3,4]
K_i	univariate kernel function for smoothing over columns [3]
K_j	univariate kernel function for smoothing over rows [3]
l_i	length of individual line segment joining (x_{i-1}, y_{i-1}) and (x_i, y_i) [5]
m	number of columns and of rows in integer lattice, assumed equal [1,2,3,4]
M	total number of pixels on lattice [4]
n	number of edge points on single tracked contour [3,5,6]
n	last point on contour when labelled in terms of index number [5]
n^*	number of points on contour after any adjustment for use of fast Fourier transform [5,6]
$N_{ij}(x,y)$	first order neighbourhood of point (x,y) [4]
O_k	observed intensity on original image scale, associated with integer k on integer-incremental scale [2]
p	$\{p = 1,2,....,M\}$ index giving position of single estimate in 'stacked vector' notation , see $a_\lambda(p,q)$ [4]
p_i	normalised angular form for index label [5]
$p_t(z\|\theta)$	conditional probability of interval z, for given threshold t [7]
$p_t(z)$	probability of interval z, for given threshold t [7]
P	true length of perimeter of simulated region [6]
$\hat{P}$	general estimator of perimeter length representing one of $\{P(r) : r = 1,2,...,R, P(k), P(t)\}$ [6]
P_g	true perimeter length of g^{th} shape [6]
$\hat{P}_g$	general estimator for perimeter length of g^{th} shape [6]
$\hat{P}_s$	general estimator of perimeter length for s^{th} simulation [6]
$\hat{P}_{gs}$	general estimator of perimeter length for s^{th} simulation of g^{th} shape [6]
$P(t)$	estimator of perimeter length calculated from tracked edge points [6]

P(k)	estimator of perimeter length calculated using Kulpa method [6]
P(r)	estimator of perimeter length calculated from Fourier series with {r = 1,2,...,R} harmonics [6]
q	{q = 1,2,....,M} index giving position of a single observation in 'stacked vector' notation, see $a_\lambda(p,q)$ [4]
q_i	normalised angular form for distance label [5]
Q_1	lower quartile of distribution of neighbourhood residuals for estimation of σ^2 [4]
Q_1	lower quartile for distribution of interval lengths [7]
Q_3	upper quartile of distribution of neighbourhood residuals for estimation of σ^2 [4]
Q_3	upper quartile for distribution of interval lengths [7]
r	{r = 1,2,....,R} index/subscript for individual harmonics in Fourier series [5,6]
R	maximum number of harmonics to be fitted [5,6]
r^*	specified number of harmonics to be fitted [5,6]
r_i	radius vector for individual edge point, expressed in polar coordinate form [1]
r1(i,j)	residual calculated within local 2 x 2 neighbourhood of observation (i,j) [4]
r2(i,j)	residual calculated within local 3 x 3 neighbourhood of observation (i,j) [4]
s	{s = 1,2,...,m} column coordinate as for i, but in denominator summation in estimator μ_λ [3,4]
s	{s = 1,2,...,S} subscript for individual simulation with fixed parameters {α,Π,g} [6]
s	x-intercept of short interval of random path [7]
S	total number of simulations in a given set with fixed parameters [6]
S_r	residual standard deviation after fitting r harmonics [5]
S1	estimator of error variance σ^2 based on squared residuals r1(i,j) within a local 2 x 2 neighbourhood [4]
S2	estimator of error variance σ^2 based on squared residuals r2(i,j) within a local 3 x 3 neighbourhood [4]

t	$\{t = 1,2,....,m\}$ row coordinate as for j, but in denominator summation in estimator μ_λ [3,4]
t	censoring threshold for individual intervals between consecutive edge points [5,7]
v	index in Fourier series summation fitting only odd or only even terms [5]
v1	variable used to denote inversion recovery image before standardisation of raw data to common scale for all subjects [2]
v2	variable used to denote T_1 relaxation time image before standardisation of raw data to common scale for all subjects
v3	variable used to denote proton density image before standardisation of raw data to a common scale for all subjects [2]
$w_\lambda(x-i, y-j)$	typical element of matrix W_λ: weight given to observation at position (i,j) contributing to estimate required at position (x,y) [1,3,4]
W_λ	m x m matrix of weights $w_\lambda(x-i,y-j)$ for set of all observations $\{(i,j)\}$ contributing to single estimate required at position (x,y) [4]
x	$\{x = 1,2,...,m\}$ column coordinate of position at which estimate is required (continuous space for Chapters 1,3; restricted to integer values for Chapter 4 [1,3,4]
x_i	horizontal Cartesian coordinate of individual edge point in continuous space [1,3,5,6]
x_i	horizontal coordinate of intersection of random path with integer lattice [7]
$\hat{x}_i(r^*)$	fitted horizontal coordinate when including r^* harmonics in Fourier series [5,6]
X	part of defining equation for test ellipse [5]
y	$\{y = 1,2,...,m\}$ row coordinate of position at which estimate required (continuous space for Chapters 1,3; restricted to integer values for Chapter 4 [1,3,4]
y_i	vertical Cartesian coordinate of individual edge point in continuous space [1,3,5,6]
y_i	vertical coordinate of intersection of random path with integer lattice [7]

$\hat{y}_i(r^*)$ fitted vertical coordinate when including r^* harmonics in Fourier series [5,6]

Y part of defining equation for test ellipse [5]

z interval (Euclidean distance) between two intersections of random path with integer lattice [7]

z(i,j) (or z(x,y), for Chapter 4 only) observed intensity at data site (i,j) [1,3,4]

$\underline{z}$ 'stacked vector' of observations with typical element z(q) [4]

z_i general labelling variable for individual edge points, corresponding to either i or d_i [1,5]

z_n 'last' point on contour labelled in terms of general variable, corresponding to either n or d_n [5]

z1 inversion recovery variable after standardising raw data to a common scale for all subjects [2,3,4]

z2 proton density variable after standardising raw data to a common scale for all subjects [2,3,4]

α	general integer, when defining p as a multiple of m [4]
α	level of area factor in simulations [6]
β	general integer, when defining q as a multiple of m [4]
γ	low threshold on magnitude of first derivative at those points identified as zero-crossings of LoG [3]
γ	number of neighbourhood subsets into which image is divided [4]
$\Gamma(x,y)$	summation over all neighbourhood subsets of each point (x,y) [4]
$\varsigma\{N_{ij}\}$	value of discretised Laplacian calculated within each local neighbourhood [4]
$\nabla^2[z]$	Laplacian (second derivative) operator evaluated with respect to discrete data [1]
$\nabla^2\{\hat{\mu}_\lambda\}$	Laplacian (second derivative) operator [3]
$\nabla_i(S_r)$	magnitude of the difference between values of S_r for unequal (d_i) and equal distances (d_i^*) [5]
ϵ	random error when fitting function μ to data z [1,3,4]
$\underline{\epsilon}$	'stacked vector' form for all random errors [4]
η^2	noise level for simulated shape [6]
θ	angle in defining equation for test ellipse [5]
θ_i	angle of radius vector for individual edge point when expressed in polar coordinate form [1]
θ_i	normalised angular form with respect to general labelling, corresponding to either p_i or q_i [5]
θ	slope of short interval of random path [7]
λ	smoothing parameter [1,3,4]
λ_i	smoothing parameter for univariate kernel K_i, smoothing over columns [3]
λ_j	smoothing parameter for univariate kernel K_j, smoothing over rows [3]
λ(CHI)	estimate of λ from chi-squared method [4]
λ(EDF)	estimate of λ from empirical degrees of freedom method [4]
λ(RICE)	estimate of λ from Rice method [4]

$\lambda(CV)$ estimate of λ from cross-validation method [4]

$\lambda(CVD1)$ estimate of λ from (First Difference) cross-validation method [4]

$\lambda(CVD2)$ estimate of λ from (Second Difference) cross-validation method [4]

$\lambda(CVKAY)$ estimate of λ for cross-validating Laplacian directly using 'Kay' method [4]

μ underlying 'true' continuous intensity function [1,3,4]

μ_λ estimator of smooth function μ [1,3,4]

$\hat{\mu}_\lambda(x,y)$ estimate of intensity at point (x,y), that is, specific value when data supplied [3,4]

$\hat{\mu}_\lambda^c$ 'leave-one-out' cross-validation estimate [4]

$\underline{\hat{\mu}}_\lambda$ 'stacked vector' of estimates, with typical element $\hat{\mu}_\lambda(p)$ [4]

ν general order of derivative of kernel function [3]

ρ for simulated shapes, ratio of major to minor axes of ellipse, or of sides of rectangle [6]

σ^2 error variance [1,3,4]

$\hat{\sigma}^2$ estimate of error variance [4]

τ time-dependent constant in relation between v1 and v2 [2]

τ nearest integer approximation to $k\lambda$ truncation limit of support of Gaussian kernel, when defining smoothing window with respect to discrete lattice [3,4]

ψ number of corner points on an edge (for Kulpa method) [6]

$\Omega(p)$ shorthand notation for summation over restricted subset of $\{p = 1,2,\ldots,M\}$ for cross-validation (first and second difference) methods [4]

$\Omega(r^*)$ residual sum of squares after fitting r^* harmonics in a truncated Fourier series [5]

1 INTRODUCTION: REVIEW OF THE LITERATURE

1.1 Identification of edges in computer vision

1.1.1 Introduction: the importance of edges

The processing of computer images has traditionally been viewed as a problem in image analysis. Such problems occur in a wide variety of fields, for example biology, medicine, geography and in engineering. One aspect is computer vision which aims to automatically extract 'significant' information from an image, in a form suitable for subsequent computer processing. Applications include automatic character recognition, identification or assembly of machine parts, processing of satellite images and the screening of medical images.

An image is defined as a two-dimensional or 'flat scene' representation of a subject area that varies in visual properties from point to point. Such varying visual properties include brightness, colour or reflectance. This variation can be described mathematically as a function of two variables; these two variables are the horizontal and vertical coordinates of a point in the image. The value of the function at a point is often termed the 'grey-level' in the literature, but in this thesis the function will be termed more loosely the 'intensity'. Intensity is strictly defined as the reflected light from the surface of an object: the amount of reflected light will depend on the physical characteristics of the objects in the image. In this thesis the term intensity is not restricted to the reflectance properties but will be used in a rather more broad sense. The intensity at any point is used to mean the value of the function for any physical property which is measured for a particular image type.

A subject underlying the image exists in continuous two-dimensional space, and in theory an intensity value can be recorded at any point within this space. Therefore in this thesis the 'underlying' image is considered to exist in continuous two-dimensional space and intensity is viewed as a continuous

function. When an image is to be processed by computer it is stored in 'digitized' form: the underlying continuous function is sampled at regular intervals to obtain a finite array of 'observed' intensity values.

Computer vision can be related to the study of the biological and physiological aspects of vision, see for example Marr and Hildreth (1980) and references therein. Changes in observed intensity should be accounted for in terms of the physical characteristics of objects in the underlying image. Separate features in an image are often characterised by 'significant' changes in intensity. This is an important concept, since an edge is defined as a 'significant' change in intensity between neighbouring pixels; the size of the change will depend on the relative magnitude of the intensity values and the level of random noise in the image. Edges are usually considered as one of the most important features for vision; see Hildreth (1983).

A change in the underlying intensity function is mathematically equivalent to a maximum of the first derivative or a zero-crossing of the second, so potential edges can be identified by looking at derivatives of the intensity function. This is usually preceded by 'smoothing' or averaging over groups of neighbouring pixels, to reduce the random noise and emphasise features such as edges. It also allows the continuous underlying function to be estimated from the discrete data, so the calculation of derivatives is well-defined.

Whatever the application, the identification of edges is a first crucial step in describing an image. An image represented in terms of its edges has a simpler description and vast amounts of redundant information are removed. Many papers give a general overview of different aspects of edge identification, for example see papers by Canny (1986), Torre and Poggio (1986) and Ulupinar and Medioni (1990).

1.1.2 Identifying edges in discrete data

Locations of edges in discrete images are identified by

approximating derivatives by differences calculated between pixels. The first derivative is approximated by the magnitude of the digital gradient, $[G_x^2 + G_y^2]^{1/2}$, where G_x, G_y correspond to differences in two orthogonal directions, typically taken as the horizontal and vertical directions with respect to a Cartesian coordinate system. Note also that the direction of the edge can be calculated as $\tan^{-1}[G_y/G_x]$. Computational efficiency can be improved by using absolute values to approximate the maximum gradient as $\{|G_x| + |G_y|\}$.

One of the most commonly used gradient operators is due to Roberts (1965). It is based on the observed intensity z(i,j) within each 2 x 2 neighbourhood, where $G_x = \{z(i,j) - z(i+1,j+1)\}$ and $G_y = \{z(i+1,j) - z(i,j+1)\}$. Sobel defined an operator (see Rosenfeld and Kak, 1982) based on a 3 x 3 neighbourhood:

$$G_x = [\{z(i+1,j-1) + 2z(i+1,j) + z(i+1,j+1)\} - \{z(i-1,j-1) + 2z(i-1,j) + z(i-1,j+1)\}]$$

and $$G_y = [\{z(i-1,j+1) + 2z(i,j+1) + z(i+1,j+1)\} - \{z(i-1,j-1) + 2z(i,j-1) + z(i+1,j-1)\}].$$

An operator similar to that used by Sobel but without the '2' scaling factor for the second term in each bracket was proposed by Prewitt (1970). These two operators are less sensitive to noise since they average over a larger region. For second derivatives one approximation is given by the orientation-independent discrete Laplacian:

$$\nabla^2[z(i,j)] = \{z(i+1,j) + z(i-1,j) + z(i,j+1) + z(i,j-1) - 4z(i,j)\}.$$

These and other such edge operators are discussed in more detail in books on digital image processing: see Rosenfeld and Kak (1982) or Pratt (1977). Peli and Malah (1982) considered several measures for assessing how well edges were identified by different operators. They looked at both quantitative measures, for example the percentage of points detected on the true edge, and also qualitative measures, such as whether the edge was broken.

1.1.3 Identifying edges in a continuous image

Edges are to be identified by looking at derivatives of the smoothed image. This must be made more specific, by defining which derivative should be calculated and choosing a function for smoothing the image. Specific details of the general algorithm are discussed by many authors, see for example Hildreth (1983), Torre and Poggio (1986), Canny (1986) and references therein.

Two decisions have to be made when calculating derivatives: whether to use first or second derivatives and further, to use isotropic (rotation-invariant or orientation-independent) operators. Edge positions are characterised by a maximum in the first derivative or equivalently, a 'zero-crossing' of the second derivative. Canny (1986) defined an 'optimal' first derivative operator and discussed ways of improving the accuracy with which edges were located. Second derivative operators have been discussed by many authors. In particular see Haralick (1984) and Ulupinar and Medioni (1990) who gave methods of correcting for additional false 'edges' introduced due to the properties of zero-crossings. This is discussed further in Chapter 3.

Important criteria for the choice between directional and isotropic operators are speed of computation and simplicity of the operator. A further consideration is whether the direction of the edge is important: this is not an issue in this thesis since it must be possible to detect edges at any orientation in the image. Usually more than one directional derivative (Canny, 1986) must be calculated to allow detection of edges which can occur in any orientation. In comparison, only a single isotropic operator is required, for example the second derivative in the direction of the gradient (Haralick, 1984), or the Laplacian. The Laplacian was used by Marr and Hildreth (1980), who proposed conditions on local intensity variation under which it approximated the second directional derivative, and which they assumed would be satisfied in most natural images. A more rigorous justification was given by Torre and Poggio (1986), who derived properties of two operators, the Laplacian and second directional derivative along the gradient. They compared the

operators theoretically and considered when each set of zero-crossings identified by the two different operators coincided. A good general comparison of several derivative operators with the Laplacian was given by Shah et al (1988).

The observed data are discretised, sampled from the underlying continuous intensity function. The data will be subject to random sampling errors (or noise), which will be emphasised by differentiation. The calculation of edges is not well-defined for discrete data. Torre and Poggio (1986) proposed 'regularising' as a prerequisite for differentiation when identifying edges; see also Poggio et al (1988). In this context, regularisation is equivalent to estimating the underlying intensity function, observed only at discrete pixels, which is piecewise regular and so well behaved under differentiation. It is achieved by locally averaging or 'smoothing' over the sampled data, which also reduces the random noise.

The function used for smoothing the data will be termed the smoothing function in this thesis, equivalent to the 'filter' in computer vision terminology. The most commonly used function is the Gaussian, proposed in this context by Marr and Hildreth (1980). Their claim to optimality was based on physical constraints of the underlying image. An alternative approach, defining a strict mathematical model for the edge has shown that calculating derivatives of Gaussian smoothed images was a near-optimal method of identifying edges (Dickey and Shanmugan, 1977, Canny, 1986).

1.1.4 The Laplacian of Gaussian (LoG)

So far, the identification of edges has been viewed as two distinct stages: the smoothing of the image, to be followed by the calculation of derivatives. The two stages can be combined by defining a suitable operator. The Gaussian smoothing function and the Laplacian derivative combine to give the 'mexican-hat' or Laplacian of Gaussian operator (LoG), proposed by Marr and Hildreth (1980). Potential edges are identified as positions of zero-crossings. The LoG has several important properties, those relevant to the current

application will be discussed in Chapter 3.

The LoG is extensively used in computer vision, favoured for its computational efficiency and relative simplicity. It is often used as a standard against which to compare other operators. Lunscher and Beddoes (1986) confirmed the near optimality of the LoG, showing that it compared favourably with a second operator proposed by Dickey and Shanmugan (1977). This second operator was motivated by defining a strict mathematical model for edges, in contrast to the Marr and Hildreth (1980) approach of describing edges in terms of features which give rise to intensity changes in the natural world.

Bias, that is a loss in accuracy in the location of edges, is a problem for many derivative operators. Further smoothing of the image results in an increase in the amount of bias. Using the LoG localisation is significantly worse for curved edges or where two neighbouring edges interact: see Huertas and Medioni (1986) for a qualitative discussion. Berzins (1984) analysed theoretically typical edge features such as corners, curves and 'noisy' edges. He summarised the magnitude and direction of the displacement from the 'true' edge position as a function of the amount of smoothing.

If the value of the intensity function is to be estimated at each pixel in turn, then a large number of computations is required when identifying edges, rapidly increasing with further smoothing. Proposals have been made to improve computational efficiency. One possibility is to use a 'Difference of Gaussian' (DoG) function, that is two Gaussian filters, each with a different smoothing parameter. The DoG would be preferred to the LoG on physiological grounds since it is more consistent with the proposed model for vision (Marr and Hildreth, 1980, Hildreth, 1983). However it is only an approximation to the 'optimal' LoG.

A second option is to decompose the bivariate LoG, the sum of partial derivatives in two orthogonal directions, typically taken as the horizontal and vertical with respect to a Cartesian coordinate system. Each partial derivative is factorised into the product of a univariate Gaussian smoothing function and a univariate Laplacian

operator. Each partial derivative is implemented in two successive one-dimensional calculations, greatly reducing the number of computations required. Much of the previous work on the decomposition of the LoG was summarised by Sotak and Boyer (1989) and references therein. They suggested further improvements and gave a formal design for the choice of parameters (such as the amount of smoothing), aiming to obtain a balance between computational efficiency with accurate location of edges.

A very different approach to reducing the number of computations was taken by Shann and Oakley (1990). Derivatives of the smoothed image were no longer evaluated at all lattice points but only in the neighbourhood of points where an edge was assumed to lie, automatically interpolating the position of edges to sub-pixel accuracy. Results were compared for the Gaussian directional first derivative and the LoG operator.

1.1.5 The smoothing parameter

The amount of smoothing is determined by a parameter which in this thesis will always be termed the smoothing parameter and denoted by λ. In the computer vision literature this parameter is usually termed the 'scale' parameter and denoted by σ. Note that the computer vision use of σ for the scale parameter should not be confused with the term σ^2 used in this thesis for the error variance (when fitting a model to the data), introduced later.

For any image, it will be virtually impossible to define a single value for the parameter λ that will optimise the edges identified. It may be unrealistic to attempt to do so, since physical changes in an image occur over different distances and at different levels of detail. For example in the current MRI, there is local variation within each region, in addition to the more significant changes between different regions. Models for vision assume that changes over different distances are analysed simultaneously so a single smoothed image is inadequate (Marr and Hildreth, 1980).

A complete description of the image can only be obtained by smoothing at several different scales and combining the results. Rosenfeld and Thurston (1971) were among the first to suggest such multiple smoothing of images when identifying edges. Marr and Hildreth (1980) combined information from a single image smoothed with different values of λ. They suggested that edges which coincided when identified with different λ should be taken as the 'true' edges in this image, but did not give any rigorous justification. A multiple scale approach using an arbitrary sequence of discrete λ can be refined by varying the parameter continuously. Bergholm (1987) used logarithmically spaced values of λ in a method termed 'edge focusing'. The image was smoothed with a 'large' value of λ and edges identified. The amount of smoothing was then reduced to 'focus' or improve the accuracy with which edges were located, but only those edges first identified at large λ were retained. 'Scale-space filtering' is the name given to the concept of treating λ as a continuous parameter (Witkin, 1983). Several authors have analysed how zero crossings change with λ, for example Shah et al (1988), Yuille and Poggio (1986), but this will not be pursued in this thesis.

1.2 Fitting a model to the data

1.2.1 Nonparametric kernel regression

A major theme for this study is the fitting of a model to the discrete data sampled from the continuous function underlying the image. By 'smoothing' over neighbouring pixels, the intensity function can be estimated at any point in continuous two-dimensional space. More importantly for the identification of edges, derivatives of the smoothed function can be calculated at any location in the image. A nonparametric approach to model fitting is favoured in biological or medical applications, since there is often little prior information about a suitable parametric model. Only general assumptions are made concerning the smoothness of the function underlying the sampled data.

Consider bivariate data $z(i,j)$, on a lattice indexed by $\{i,j = 1,2,\ldots,m\}$, which are assumed to follow a model of the form:

$$z(i,j) = \mu(i,j) + \epsilon(i,j), \qquad \{i,j = 1,2,\ldots,m\} \qquad (1.1)$$

where μ is the regression function on $\mathbb{R}^2$, (that is continuous two-dimensional space) to be estimated, and $\epsilon(i,j)$ is the random error, assumed to have zero mean and variance σ^2. The estimator μ_λ of μ, indexed by a smoothing parameter λ, is taken to be a linear, weighted sum of the m^2 observations, of the general form:

$$\mu_\lambda(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{m} w_\lambda(x-i,y-j)z(i,j), \qquad (1.2)$$

for $\{i,j = 1,2,\ldots,m\}$. The function $\mu_\lambda(x,y)$ is to be estimated at any point (x,y) in continuous two-dimensional space. The weights are $w_\lambda(x-i,y-j)$, the exact form of which is to be defined.

Further discussion of the form of the estimators will be given in Chapter 3. Estimators are favoured which have a simple structure and are not dependent on the form of the data. Many of the existing smoothing techniques were reviewed by Collomb (1981). Series estimators and smoothing splines were considered in some detail by Eubank (1988). Härdle (1990) discussed k-nearest neighbour, orthogonal series and spline estimators. Methods considered for the current data are the fitting of a global regression surface, either thin-plate or natural neighbour splines (Sibson, 1980, 1988) or local estimation by interpolation (Nalwa, 1987), all of which will be summarised in Chapter 3.

Differences between the methods are due to the form of weighting function for the estimator μ_λ in (1.2). In this thesis a kernel function will be used to define the weights $w_\lambda(x-i,y-j)$. Kernel functions have been used for the estimation of densities, for both probability density functions (Rosenblatt, 1956) and for spectral densities (Parzen, 1957); see also Silverman (1986) for a general overview of density estimation. The use of kernels for regression was covered by Müller (1988) and Eubank (1988, Chapter 4). They discussed

several forms for the estimator μ_λ, different kernel functions K and methods for determining a suitable smoothing parameter λ, all of which had to be made explicit before the general model (1.2) was fitted to the data.

1.2.2 The form of estimator

Kernel regression is to be used to estimate a function μ relating observations z(i,j) to the data points (i,j). The estimator μ_λ is a linear weighted sum of the observations. Several forms of estimator have been proposed for univariate data, which can be generalised for multivariate design points by appropriate choice of kernel function. Some estimators are specifically for equally-spaced data (Priestley and Chao, 1972), but others are more generally applicable, for example to unequally-spaced or random design points (Nadaraya, 1964, Watson, 1964, Benedetti, 1977). Estimators have also been proposed which interpolate between data points (Gasser and Müller, 1979). Eubank (1988, Chapter 4) listed five univariate estimators. In Chapter 3 two of these estimators are generalised to bivariate data and fitted to the current data. Analagous forms of estimator have been defined for the estimation of derivatives of μ, see Gasser and Müller (1984).

A bivariate kernel function is required for the current data. Most of the proposed kernels are univariate functions and so must be modified. The bivariate function used in this thesis will be a product kernel, that is a product of two univariate functions. Such products were originally considered for density estimation (Cacoullos, 1966, Epanechnikov, 1969) and discussed for kernel regression by Müller (1988, Chapter 6).

How well the estimated function fits the underlying data is assessed in terms of some distance measure, such as the mean square error (MSE) between the true and estimated functions. The 'optimal' kernel which minimises the MSE has been studied theoretically (Benedetti, 1977). Asymptotically optimal polynomial kernels have been studied extensively: see Gasser and Müller (1979, 1984), Gasser

et al (1985). Härdle (1990) compared these 'optimal' polynomials with other commonly used kernels, including the Gaussian, a function widely used in computer vision (Section 1.1.3). He concluded that in terms of MSE there was little to choose between them and the choice of a suitable function should be based on other factors, for example computational efficiency.

Kernels for the estimation of derivatives of the regression function are also of interest. A general discussion was given by Eubank (1988, Chapter 4) or Härdle (1990, Chapter 3). Gasser and Müller (1984) defined further polynomial kernels specifically for estimating derivatives.

1.2.3 Estimation of a smoothing parameter

The kernel function and estimator μ_λ are both indexed by a parameter λ. In this thesis, λ is termed the smoothing parameter, though in the literature it may be termed the 'bandwidth' and is equivalent to the 'scale' parameter used in computer vision (Section 1.1.5). Changing the value of λ adjusts both the weights and the size of the local neighbourhood of the point over which observations are to be smoothed when estimating intensity at a single point. The estimation of a suitable parameter is critical in determining how well the fitted, smoothed model estimates the underlying continuous function.

A quick, subjective method for estimating λ is simply to smooth the data using several values of λ and assess the 'best' visually. The definition of 'best' will depend on the estimation problem: in this study a value of λ is selected which results in the most representative set of edges identified between different regions in the image. Usually a more 'automatic' method is preferred and many data-based methods of estimating λ have been proposed. Most have assumed univariate data, and so must be extended for bivariate data; see Thompson et al (1991a) who compared several estimation methods when analysing two-dimensional images.

Many of the data-based methods select a value which minimises some measure of lack-of-fit between the underlying true function μ and its smoothed estimate μ_λ. One such measure is the MSE. Since the true function μ is unknown, in practice it is replaced by its observed value z and the MSE is estimated by the corresponding residual sum of squares (RSS). Parameters estimated using such methods minimise the global lack-of-fit: the value of λ which results in the best fit averaged over all the data points. The alternative is local or variable parameters which adjust the amount of smoothing around each data site, depending on the local variation of the data values.

Four of the more commonly used methods are now outlined. The bivariate extensions of these methods will be covered in detail in Chapter 4, where the results obtained when the methods are used to estimate a smoothing parameter for the current images will be discussed. The first method considered was proposed by Rice (1984), in the context of kernel regression with univariate data points. Using this method the parameter λ is taken as the value which minimises a data-based estimate of the MSE. The method requires an estimate of the error variance σ^2; Rice considered two estimators which will also be discussed in Chapter 4.

The second method, cross-validation (CV), also aims to minimise the MSE lack-of-fit. Unlike the Rice (1984) method, CV does not require an estimate of σ^2 and so is a completely data-based method of estimating λ. Each observation is omitted in turn and the remainder of the data is used to predict the value of the function at that position. The CV 'score' is calculated as the average, over all data points, of the squared differences between the observed data and the predicted value. Cross-validation was proposed by Stone (1974) and in particular for kernel regression by Clark (1977). Thompson et al (1989) discussed the use of CV for image restoration and observed that although the method usually worked well, cases did occur where the value of λ estimated was unsatisfactory, for example resulting in undersmoothing of the data.

Craven and Wahba (1979) discussed CV for estimating a parameter λ

for smoothing splines. They expressed the CV score in a computationally more convenient form and showed how the number of computations could be reduced by using the average of the weights for the estimator μ_λ, rather than calculating individual weights which will generally be different at each point. This is the generalised cross-validation (GCV) method, the form of cross-validation discussed by many authors. For the special case where observations are made at equally-spaced data points and the function μ is estimated only at this same set of points, CV and GCV are equivalent. In the MRI intensity is observed only at integer lattice points, so this property holds for the current data and will be illustrated in terms of the kernel model in Chapter 4. In this thesis the method will always be termed cross-validation, with the understanding that it is equivalent to generalised cross-validation.

Two further modifications of the CV method have been proposed. Asymptotic generalised cross-validation (AGCV) was used by Silverman (1984) to further improve computational efficiency. Rather than calculating differences between the estimate μ_λ and the observed data, Chan and Kay (1991) calculated differences between the estimate μ_λ (based on all the data) and the corresponding estimate omitting each data point in turn. This is termed estimation cross-validation (ECV), but neither this nor AGCV will be considered further in this thesis.

The Rice (1984) and CV methods are examples of estimation methods based on lack-of-fit, as measured by the MSE. An alternative approach is based on goodness-of-fit, as measured by the RSS between the observations and the estimated function. This approach exploits the properties of residuals and the number of degrees of freedom. The third selection method considered in this thesis is the chi-squared (χ^2) method, considered first by Phillips (1962) for the numerical solution of integral equations and then for smoothing splines by Reinsch (1967). The basic assumption is that differences between the observed data and estimated smoothed value, that is the RSS, should be equal to errors in the input data, as measured by the error variance σ^2. The fourth method covered is the empirical degrees of freedom (EDF) method (Wahba, 1983), a modification of the χ^2 method.

It attempts to adjust for the loss in degrees of freedom for error when estimating the unknown function μ. Both the χ^2 and EDF methods require an estimate of the error variance σ^2 and to some extent, the estimation of a 'suitable' parameter λ is affected by the reliability of the value estimated for σ^2.

Many of the methods proposed for smoothing parameter estimation were referenced by Müller (1988, Chapter 7) and Eubank (1988, Chapter 2). In addition to the four methods introduced above, Eubank outlined four more and showed that they were all closely related asymptotically to the more common methods. Titterington (1985) considered several estimation methods and summarised many of the smoothing techniques used in different areas of statistics, such as parametric and nonparametric regression, density estimation and image restoration. In the context of image restoration, a common structure was established between the different techniques and the estimated values of λ compared. Hall and Titterington (1987) considered two general approaches to parameter estimation, which were illustrated for both ridge regression and smoothing splines and the parameters estimated were compared.

For the restoration of two-dimensional images, Thompson et al (1991a) discussed the χ^2, EDF and CV methods when estimating a smoothing parameter. The methods were compared algebraically and simulations used to verify the expected relative magnitude of the corresponding estimated values of λ. Härdle (1990, Chapter 5) compared theoretically several techniques using a kernel regression model. He considered asymptotically which should be used and also summarised the results from two simulation studies (Rice, 1984 and Härdle et al, 1988) which compared the methods for small samples.

Derivatives of the estimated function are similarly indexed by λ, so a related problem is the estimation of a smoothing parameter for derivatives. Derivatives, or functions of them, are of interest in the study of growth curves (Gasser and Müller, 1984), or as in this thesis to characterise the intensity changes which correspond to the position of edges. One important question is whether the same value of λ should be used for the derivative as was estimated for the

function itself. Several methods have been proposed for the estimation of a smoothing parameter specifically for derivatives. Rice (1986) generalised his original method (Rice, 1984) by using difference quotients of the observed data. This modification was one of the methods discussed by Härdle (1990). He outlined an analagous CV function based on difference quotients, also discussed by Müller (1988, Chapter 7).

One disadvantage of the modified Rice (1986) and CV methods for estimating parameters for derivatives is that they rely on difference quotients which are unstable, becoming progressively less stable as higher differences are calculated. An alternative which does not require such differences is the 'factor' method of Müller et al (1987). With this method the theoretically 'optimal' parameter for the derivative was equivalent to the 'optimal' parameter for the function itself, except for a factor depending on the derivative required, the kernel function used and the number of continuous derivatives of the kernel. Müller (1988, Chapter 7) and Härdle (1990, Chapter 5) both summarised the factor method and compared the parameters estimated with those estimated using the modified Rice and CV methods.

1.2.4 Estimation of the error variance

An estimate of the error variance σ^2 is often of interest when fitting the regression model (1.1) or may be required when calculating confidence intervals. In this thesis, three of the methods proposed for smoothing parameter estimation (Section 1.2.3) are functions of σ^2. Several of the estimators of σ^2 considered previously have been discussed in relation to some form of smoothing. More generally, since the exact form of the underlying function will not usually be known, it is desirable that the estimator chosen should be valid for a wide variety of regression functions. Thompson et al (1991b) summarised several univariate estimators which were quadratic functions of the data. The relative performance of the different estimators was compared empirically using four simulated one-dimensional functions.

The general class of estimators which are quadratic in the data can be divided into three subclasses. The first subclass covers difference estimators. Anderson (1971) was the first to propose data-differencing for the estimation of σ^2, although it is a common procedure for removing trend in the analysis of time series. In the context of kernel regression, Rice (1984) proposed a method for smoothing parameter estimation which required an estimate of σ^2. He considered two estimators of σ^2, based on either first or second differences, calculated from either consecutive pairs or triples of the data respectively. For equally-spaced data the second difference estimator is equivalent to one proposed by Gasser et al (1986). They estimated the residual error variance for a general nonlinear regression model as a weighted sum of squared 'pseudo-residuals', calculated from continuous triples of design points. Such estimators were all based on symmetric differences of the data; the definition was extended to estimators of any general order by Thompson et al (1991b). In comparison, Hall et al (1990) considered asymptotically optimal difference estimators which aimed to minimise a MSE criterion between the true and estimated value of σ^2.

Wahba (1983) combined properties from both the CV and EDF methods for smoothing parameter estimation (Section 1.3) to define an estimator for σ^2. The Wahba (1983) estimator is an example of the second subclass of estimators which are quadratic in the data and covers those which are weighted functions of the RSS. The third subclass covers the minimax estimators of Buckley et al (1988): see Thompson et al (1991b) and references therein for further details of these and other related estimators.

The estimators discussed by Thomson et al (1991b) were all defined in terms of univariate data points, but in this thesis a bivariate estimator is required. Kay (1988) proposed a family of bivariate estimators of σ^2, calculating differences in the local neighbourhood of each point. A further consideration is the robustness of an estimator to discontinuities in the data, such as edges in an image, which may result in a biased estimator. Two methods for correcting for edges (Kimber, 1983, Grimson and Pavlidis, 1985) were discussed by Thompson et al (1991b) and compared

empirically with the other estimators they covered in their study.

1.3 Modelling the edge as a closed curve

1.3.1 Methods for fitting closed curves

In the second part of the study the aim is to estimate an edge by fitting some form of closed curve to a set of points which have been identified using the methods of Sections 1.1 and 1.2. Whatever method of curve fitting is used, the estimated edge should be transformation invariant to scaling, rotation and translation. It should not depend on an arbitrary choice of parameters, such as the point defined as the starting point of the edge or the choice of a centre from which to define a polar coordinate system. The method used to obtain the set of edge points should not affect the fitted curve, and reconstruction from the fitted model should always result in a unique, closed curve.

In addition to these basic properties of the closed curve, one extension is the generalisation of the selected method to the case where not all the points identified as 'edge' lie on a given edge, for example 'noise' points within regions enclosed by an edge. Robust estimation could be used to pick out and exclude such outlying points. This extension is not considered here: it is assumed that each set of identified points lie on one of the edges in the image.

There are many curve fitting methods available. The simplest closed curve is a polygon, that is a piecewise linear 'curve' joining points sampled from a tissue edge: see for example, Sriraman et al (1989) who reconstructed planar curves at sub-pixel resolution. Since edges in the MRI are smooth and continuous, this representation would not usually be adequate, although the fitted polygon could subsequently be smoothed. Aiming to obtain a smooth representation for the data, splines were suggested and modified for the fitting of curves in two dimensions, for example Pham (1989). Polynomial B-splines were fit interactively through the set of points and could then be extended to allow for curves such as circles and conic

sections.

Bookstein (1978) considered the fitting of conic sections and piecewise conic sections. Berman and Culpin (1986) considered only the circle which they modelled in polar coordinate form. Analysing the statistical properties of the centre and radius of the circle they defined two least squares estimators, the second of which was equivalent to Bookstein's method. Two error models were considered, depending on whether the error in fit at each point was expressed in Cartesian or polar (radial) form.

Kashyap and Chellappa (1981) adopted a very different approach and used a stochastic model. The edge was approximated by a series of straight line segments joining the identified points which were obtained in an ordered sequence and parameterised in terms of their position in the sequence. Coordinates of the edge points, x_i and y_i say, were modelled separately as circular autoregressive (CAR) processes. For those curves where the radius vector from the centroid to each edge point only intersected the edge once, coordinates (x_i,y_i) were reparameterised in polar coordinate form, (r_i,θ_i), about the centroid and the set of radius vectors $\{r_i\}$ were modelled as a single CAR process. Parameters were estimated by maximum liklihood or by least squares.

1.3.2 Fourier descriptors

If the edges are closed, this suggests representation by a periodic function, one of the simplest examples of which is a Fourier series (FS). The method adopted in this thesis is that of the fitting of Fourier descriptors (FD), first suggested by Cosgriff (1960). This method is appropriate for points which are obtained in an ordered sequence and labelled by an index, z_i say. The closed curve can be represented as a function in complex space, or as in this thesis, coordinates (x_i,y_i) can separately be approximated as functions of z_i. A reduction in the number of model parameters is often a consideration. A key feature of the FD is that fitting only the first 'few' (to be quantified) terms in the series provides a good

description of the basic shape. Higher order terms are less important, describing local detail of the fitted curve.

Much of the previous use of FD has been in the field of pattern recognition, for the discrimination and classification of closed curves. Several parameterisations, that is choice of labelling variable z_i, have been used by different authors. Granlund (1972) proposed FD for hand-print character recognition. Measuring from an arbitrary 'start' point, z_i was parameterised as a function of arc length along the edge. The coordinates of each edge point were modelled as a function of cumulative distance round the edge. The FD for one edge were not unique, but dependent on the 'start' point.

In addition to character recognition, Zahn and Roskies (1972) reported work on the analysis and synthesis of closed curves in the plane. They aimed to extract a finite set of numerical features from closed curves which tended to separate different classes of shapes. The curve was again parameterised as a function of arc length round the edge, but now a cumulative angular function was fitted to each point. Defined as the net amount of angular bend between the starting point and the current point, the function was re-expressed in its polar form so the FD then represented harmonic amplitude and phase angle.

Lin and Chellappa (1987) considered a slightly different classification problem, namely that of classifying shapes from partial edges, that is edges with missing segments. Using the cumulative distance labelling of Granlund (1972), FD were estimated for the unknown complete shape, conditional on the number of missing points being unknown. The accuracy of the estimates was improved by imposing additional constraints, such as the compactness defined as $(\text{perimeter})^2/\text{area}$.

1.3.3 Shape properties from the Fourier descriptors

Once a suitable edge model has been defined, the estimation of geometric properties of the region enclosed by the edge is often of

interest. Typical properties are the length of the perimeter of the region, the area, or the location of the centroid. For the estimation of perimeter length in particular, several methods have been proposed. For a given application the suitability of individual methods depends on the form of data, for example whether the edge is modelled as a continuous curve or represented as a piecewise-linear curve with respect to the digitised image.

Dorst and Smeulders (1987) considered the estimation of the length of the continuous curve from discrete data. Each 'true' curve was represented by a set of edge points assumed to lie at the corner of a pixel. The set of points was denoted by a 'chain code string' giving the relative direction of the next point in the ordered sequence. Several length estimators were compared and also considered for the estimation of the lengths of circular arcs. Estimators based on chain codes were used by Koplowitz and Bruckstein (1989), who discussed methods for estimating the perimeter length of the original two-dimensional continuous region from its discrete binary representation. The edge of such a region is a stepwise 'curve' formed from the horizontal and vertical links corresponding to the pixel outlines. In most cases the length of the perimeter of the digitised region is considerably longer than the perimeter of the original region. Corners of the digitised edge contribute to the overestimation, so methods were discussed which adjust for corners, see also Kulpa (1977).

If the set of edge points lie on a smooth, closed curve and are identified in an ordered sequence, then the edge can be described in terms of its Fourier series (Section 1.3.2). Geometric properties of the enclosed region can be derived as functions of the FD. Formulae for estimators of perimeter length and area were given by Lin and Chellappa (1987). Kiryati and Maydan (1989) calculated the area, coordinates of the centroid and second order moments with respect to the axes passing through the centroid, from which the orientation of the central axes and central moments of inertia were derived. Since the estimation of the perimeter length was more difficult, they only established lower and upper bounds for its true value.

2 THE DATA

2.1 Source of the data and magnetic resonance imaging

2.1.1 Background to the data

The data were obtained as part of a human nutrition study at the Rowett Research Institute (RRI), Aberdeen. The RRI study was set up to assess the usefulness of magnetic resonance imaging in detecting the loss of fat tissue during a period of sustained weight loss in women (Fuller et al, 1987). The images analysed in this thesis are cross-sections through a pair of human thighs. One image is shown in Figure 2.1, where colours have been arbitrarily assigned. The figure is included here simply for illustration: an outline of how such images are produced will be given in Section 2.1.2. Each image is of two legs, with the flat base indicating where the subject was lying on the bed. Within each leg the tissues are nested and roughly concentric; labelling from the inside the tissues are marrow, bone, muscle and fat. The black 'background' to the two legs corresponds to the air surrounding the subject.

The aim of this thesis is the estimation of edges between tissues, as a first step towards automatic classification of images. The estimation is taken in two stages: first, the identification of locations of edge points and second, the fitting of some form of closed curve to describe the edge. The importance of edges in any image analysis project and computational methods of edge identification were discussed in Section 1.1. Currently-employed methods of identifying edges are labour-intensive and rather subjective. A human operator views the image on a computer screen and the tissue edge is outlined by using a mouse to move the cursor round the edge, as viewed on the screen.

For the RRI study the identification of edges has several applications. It may be used to aid automatic classification of different tissues, or to compare a template of a typical healthy tissue with the same tissue identified in a patient's image, in order to screen for abnormalities. Alternatively attention may be focussed

Figure 2.1: MRI cross-section through human thighs - Subject et1
Superimposing data for variables v1, v2, v3

on the region enclosed by an edge. The area of a tissue can be estimated: for example in the RRI study the amount of fat, and the difference in area of a particular fat region between two consecutive images, is of interest.

When dealing with biological data it is unrealistic to expect a totally automated system and a human assessment of an image should be used in parallel with a computer evaluation. A qualitative assessment of an image is best achieved by a human operator: looking at the relative position and shape of different tissues and checking for abnormalities. In contrast, a computer (numerical) analysis is preferable for a quantitative assessment, to evaluate the relative proportions of pixels of each tissue or to obtain the distribution of intensity values for a given tissue. Ideally the automated analysis should be used to complement the biological interpretation of such images.

Data were available for two sets of images from each of six subjects. For each subject one image set was recorded pre-dieting and the second set post-dieting. Since the images are very similar and have the same basic structure, attention is restricted to two subjects, for each considering both the pre- and post-dieting image sets. Throughout this thesis, these data sets will be denoted by {et1, et2, ft1, ft2}, where the letters {e,f} are used to denote the two subjects and the {t} to indicate that the images are through the thigh. The numbers {1,2} indicate whether the images were obtained pre- or post-dieting. Each image set consists of three separate images, corresponding to three of the variables recorded by the imager, as discussed in the next section.

2.1.2 Magnetic resonance imaging

Magnetic resonance imaging is a relatively new technique, developed in the late 1970's, when it was originally termed nuclear magnetic resonance (NMR) imaging. It is favoured in medical practice because it is non-invasive and provides views within living tissues. It is considered biologically safe under normal conditions of use.

The frequency of the radiowaves required to produce magnetic resonance images (MRI) is nine orders of magnitude smaller than the frequency of conventional x-rays, so the radiation frequency lies within the range of conventional radio and television frequencies. It has advantages over other relatively new imaging techniques: the images are not degraded by the presence of bone or gas, which are problems for x-ray computer tomography and ultrasound respectively. Images can be obtained in any plane, but attention is usually restricted to three main planes which are perpendicular to each other. These are the coronal (face-on) slice and the sagital (sideways) slice, in addition to the usual transaxial (cross) slice.

Magnetic resonance imaging exploits the magnetic properties of the nuclei of atoms. The most commonly used nucleus is that of hydrogen, that is a single proton, but other nuclei with the required magnetic properties include those of carbon, sodium and phosphorus. The subject to be imaged is placed in a large, static magnetic field; this field stays on throughout the MRI experiment. The system is then provided with electromagnetic (EM) radiation at the frequency required to excite the nuclei of interest. The required frequency relates to both the type of nucleus to be examined and to the strength of the magnetic field applied. One frequency excites one nucleus at a given static field strength.

The function of the EM radiation is to supply a small, oscillating magnetic field from which energy can be absorbed in order to excite the system. The nuclei have magnetic moments and so align (parallel and anti-parallel) with the static magnetic field. The alignment is disturbed by the radiofrequency (RF) radiation and the nuclei are excited. The nuclei oscillate in this high energy state, then gradually relax back to the ground state, so losing the absorbed energy and returning to their original distribution of alignments.

The decay in the oscillations over time is measured using a simple receiver coil. The spin population, that is the nuclei, generates a fluctuating signal in the coil. The magnitude of this signal relates to the size of the spin population and to the relaxation characteristics. The relaxation characteristics are

defined by the mobility and molecular neighbourhood of the spins being interrogated. the different molecular composition of the tissues will yield different relaxation times.

To obtain a display of the image a 'field gradient' is applied, so that each set of protons requires a different RF to excite it. Then a broad band RF is applied. A Fourier transform is used to convert the frequency output (that is, amplitude and phase angle of the signal) into the space domain. The formation of the image is described further in Section 2.2.1. For further information on the technical details of magnetic resonance imaging which are beyond the scope of this thesis, see Foster and Hutchison (1988).

The subject underlying an image exists in continuous two-dimensional space, so theoretically intensity (using the definition given in Section 1.1.1) is a continuous function which can be recorded at any point (x,y) in two-dimensional space. In practice the image intensity function is sampled at a discrete number of points, usually on a square lattice or grid. This spatial quantisation results in an integer array of intensity values which is an approximation to the original continuous function. Each element of the lattice is referred to as a 'pixel', short for 'picture element'. The scale of the discrete images is determined by the number of points at which the original image is sampled, and hence the area represented by a pixel. For the images analysed in this thesis, each pixel is a square with sides of length 3½ millimetres.

Magnetic resonance imaging techniques can be used to provide a variety of different displays of the basic information on PD and relaxation time. Relaxation occurs by two separate processes: either energy is lost into the surrounding environment (longitudinal relaxation) or spin orientation is exchanged between nuclei. As well as a display of observable protons (which is little affected by relaxation processes) it is possible to obtain a large variety of images containing PD information that is weighted to different extents by the two relaxation processes. Five different image types are described by Foster et al (1984). For the RRI study, data are available for three of these image types: proton density,

longitudinal relaxation T_1 and inversion recovery (IR), which is a T_1 weighted measure of PD.

Although the IR values can be either positive or negative, it is normal practice to display only the magnitude of the signal, that is without the sign. Most MRI systems, including the one used to produce the images studied in this thesis, are designed to measure the amplitude of the signal induced by the net magnetization in the receiver coil. The phase of the signal can also be measured, but this is not usually done. The IR image without the sign, that is magnitude only, is easy to collect and process, and is known as the 'unrectified' image. The sign can be derived if the signal phase is known and a 'rectified' image produced, though this is rather laborious to carry out. There is no improvement in the displayed image if the rectified, rather than the unrectified IR image is used. The unrectified image is preferred because of ease of interpretation. Further, using the unrectified image, the timings of the RF pulse can be varied to highlight particular tissue types, so in clinical practice most people would have no use for the rectified image at all.

Whatever the variable the data value (or intensity) at each pixel is of 'byte' form, that is an integer value in the range 0 to 255. When displaying the images, the intensities are represented on a grey-scale, with values of 0 represented as white, values of 255 as black, and intermediate values as shades of grey. The colour image (Figure 2.1) presented in Section 2.1.1 is a combined display of the IR, T_1 and PD variables. The data for each variable are arbitrarily assigned to one of the red, blue or green colour bands and then superimposed. The colour image was produced by displaying the three superimposed data sets on a computer screen and taking a photograph using a standard SLR camera. Note that the vertical distortion (which has the effect of 'compressing' the legs) is a consequence of displaying the data on a rectangular screen, so individual pixels appear rectangular, rather than square.

In this thesis only the PD and IR variables will be considered in detail, since T_1 is a non-linear combination of PD and IR and

provides no additional information: see Section 2.2.2 for further discussion. Grey-scale plots of data for the PD and IR variables, for subject et1, are given in Figures 2.2a and 2.2b respectively. The variation in intensity recorded for the IR variable (Figure 2.2a) is dependent on the type of bonding between the atoms. For this image type, there is a clear distinction between the muscle and fat tissues. It is also possible to pick out small areas (covering two or three pixels) of blood within the muscle regions, which represent the main artery and vein in each leg. Pixels of blood are not always observed since their presence is dependent on the relation between the pumped blood flow and the time of imaging. The bone and muscle tissues have very similar intensity values and it is difficult to identify the edge between the bone and muscle. On close inspection however, it is possible to pick out a few slightly lighter pixels around the marrow tissue: these are pixels of bone.

For this image type the edges between the marrow and bone, the bone and muscle, and also the fat and background are quite distinct. The PD variable is effectively a count of the number of hydrogen nuclei. Although many of the hydrogen nuclei are present in the form of water molecules, the form in which hydrogen is held does also depend on the tissue type. For example, the big darkish area at the outside of the legs is fat tissue. The high intensity value is due to the protons of the mobile lipid chains in these regions. The grey muscle tissue contains about 75% water and has an intermediate intensity value. In contrast, the bone region has a very low intensity value and appears white. Although bone tissue contains about 40% water this is too little to be recorded, partly due to the amount and partly due to the structure of the water.

Magnetic resonance imaging was introduced as a diagnostic technique in the 1980's. The range of differences in MRI characteristics between different normal tissues and associated with different disease states means the images are an effective method of demonstrating normal and pathological anatomy. The RRI nutrition study (Fuller et al, 1987) is one example of its value for human studies: many others were outlined by Foster and Hutchison (1985).

Figure 2.2a: Display of inversion recovery variable z1
Subject et1

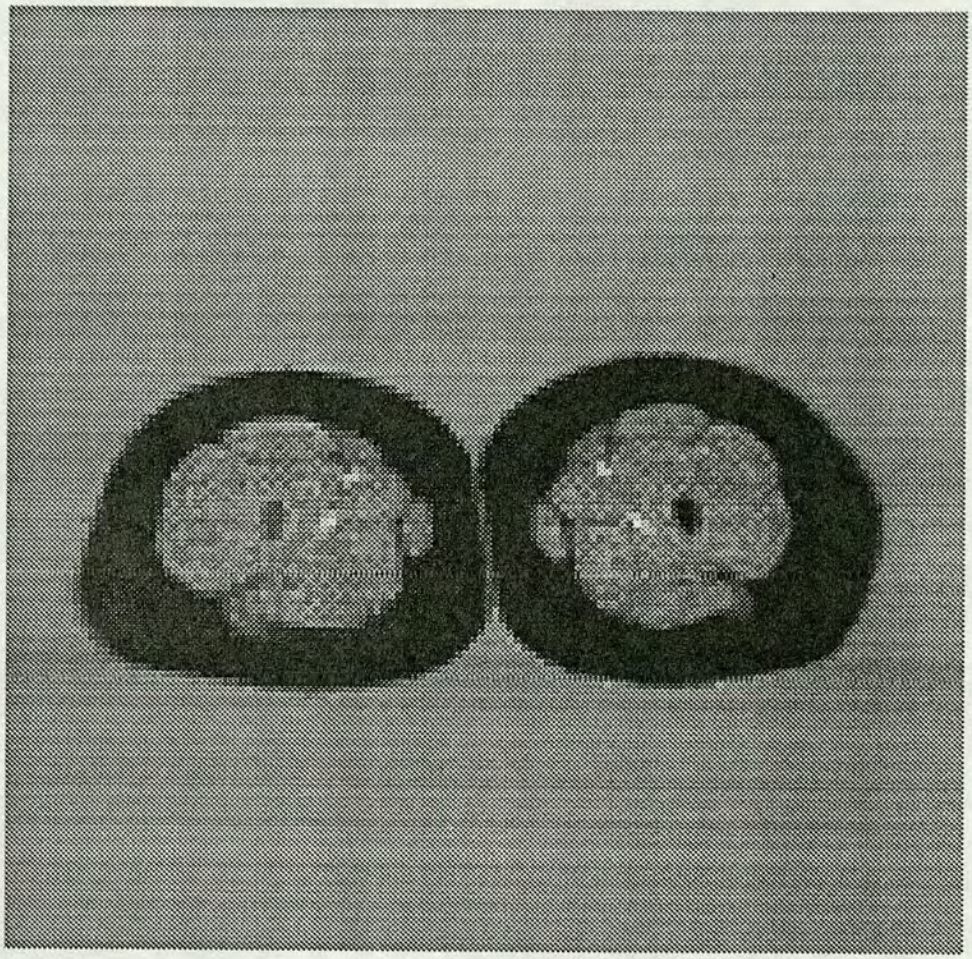
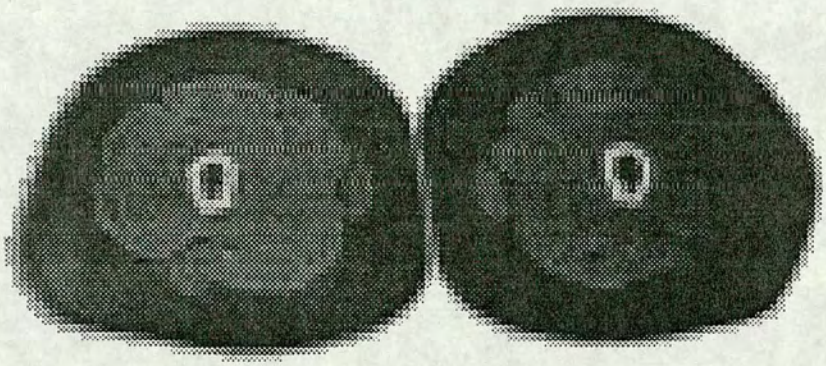


Figure 2.2b: Display of proton density variable z2
Subject et1



This form of imaging has also been used in animal studies. One example is the estimation of the body composition of meat animals prior to slaughter (Foster et al, 1985). By adjusting the imaging (pulse) sequence it is possible to discriminate between different tissues and so assess the proportion of fat to lean tissue on an animal. The technique has also been used for non-biological applications, for example in archaelogical studies of the timbers of the Mary-Rose and in the investigation of the Cheshire peat bog man.

2.2 Form of the data

2.2.1 Defining the limits of the image

For the MRI analysed in this thesis, intensities are recorded at each of 128 x 128 pixels on an integer lattice. That is, the data in each image are arranged in {i = 1,2,...,m=128} columns and {j = 1,2,...,m=128} rows, indexed with respect to an origin at the top left-hand corner of the image. The 128 x 128 matrix defines the limit of the imaged slice. This is inevitable because of the method of image collection (outlined in Section 2.1.2). An image collection consists of 128 repetitions of a pulse sequence. Each pulse sequence is 'chopped' into 128 segments. This gives a 128 set in the horizontal signal direction, and a 128 set in the vertical set of signals direction. The horizontal direction is time during collection of the signal. The vertical direction is known as 'pseudo-time' and is the same segment in each signal of the 128 set. A two-dimensional Fourier transform is then used to convert the time-domain signals to space domain. The output of this is a 128 x 128 matrix of amplitudes (or 'intensities'), in the terminlology of this thesis.

Everything within the receiver coil is imaged. However because the legs are small compared to the diameter of the receiver coil, they only occupy a smallish area of the image. There is a lot of air space or 'background' around the legs, so intensities of zero are recorded at pixels in the outer columns and rows of the 128 x 128 matrix.

As outlined in Section 1.2.1, the intensity at each lattice point is estimated as the weighted sum of m^2 observations on the lattice. Particular attention must be given to the lattice points 'close' to the boundary limits of the image, that is 1 or m (= 128). Here the set of observations over which the intensity is averaged will extend beyond the boundaries of the 128 x 128 image. Thus to allow estimation at such points it is assumed that the intensities observed on a two-dimensional lattice are 'wrapped-round' as on a torus.

For the thigh images analysed in this thesis where the study object occupies only the central area of the 128 x 128 lattice, an intensity of zero will always be assigned when positions near the image boundary are wrapped-round. For this special case wrapping-round on a torus is equivalent to automatically assigning an intensity of zero at positions outside the m x m image lattice. This can be thought of as an infinite background of zero intensities extending out in all directions around the observed image. However for a more general solution and to make the structure of the data explicit, the algebraic notation of wrapping-round on a torus will be used in the following when discussing the estimation of intensity at a given point.

2.2.2 Rescaling the data

For this study, data were available for three variables (Section 2.1.2): inversion recovery (IR), T_1 relaxation time and proton density (PD), which for convenience will be denoted by v1, v2 and v3 respectively. Since the emphasis in most of the studies involving MRI is the display of the images this obviously influences the form in which the data are recorded. The modification of the raw data required prior to the analysis of thesis is covered in this section.

The IR variable (v1) is a PD (v3) image, weighted with T_1 (v2) information. In this thesis only two of the variables {v1,v3} are studied, since the third (v2) does not provide any additional information. The form of the data available for v1 is the unrectified image, that is the magnitude of the signal (intensity) without the

sign. It was considered, however, that the sign of the intensity should be included in any statistical analysis using the numerical values of the data. Further, since variables are individually scaled for each image results cannot be compared directly between subjects. Therefore after recovering the negative sign for v1, both variables v1 and v3 are rescaled to a common integer-incremental scale. The data analysed in Section 2.3 and in subsequent chapters consist of two variables on an integer-incremental scale, which are directly comparable between different subjects.

The phase of the signal, required to derive the sign of the signal and so provide a rectified IR image, was not available for the MRI. The negative sign for v1 is therefore recovered based on the relationship between {v1,v3} and v2. The T_1 relaxation time, that is variable v2, is derived from v1 and v3 intensities using the relation:

$$v2 = \frac{\tau}{\ln\left[\frac{2v3}{v1 - v3}\right]}, \quad \text{which rearranges to} \quad \frac{v1}{v3} = \left\{\frac{2}{\exp\left[\frac{\tau}{v2}\right]}\right\} - 1,$$

where τ is a time-dependent constant. In practice, this theoretical expression is not used explicitly: only the indicated relation between (v1/v3) and v2 is used, in order to recover the negative sign for v1 empirically.

Only positive v2 intensities are observed but for some values of v2, v1 is positive and for others it is negative. The first step in the rescaling is to find the v2 intensity, I' say, at which v1 (or equivalently (v1/v3) since v3 is always positive) changes sign from negative to positive. The 'change-of-sign' (or 'null-point') intensity I' is identified by plotting (v1/v3) against v2 and reading off the value of v2, that is I', at which (v1/v3), takes the value zero. The value of I' is a function of an operator variable for the imager, the value of which will depend on the sequence of RF pulses used to produce the images. For the MRI considered in this study, I' corresponds to a v2 intensity of 74.

Using the typical intensities of the fat and muscle tissues the

negative value is assigned to the correct range of v1 intensities such that if v2 > I' then v1 is negative, if v2 = I' then v1 = 0 and if v2 < I' then v1 is positive. In all subsequent discussion it will be assumed that the negative sign has been replaced (where appropriate) for variable v1 and no further use will be made of v2.

As noted in Section 2.1.2, the data is recorded in byte form, that is integer intensities in the range 0 to 255 (inclusive). In practice, not all intensities within this theoretical range are observed for either v1 or v3. The intensities which are observed do not occur on an integer incremental scale, and even for a single subject the increments between successive observed values are not equal. To enable direct comparisons of images for different subjects it is considered necessary to rescale the raw data to be of an integer-incremental form. The following outline of the rescaling is given in terms of variable v3 for a single subject, but the rescaling follows similarly for v1, with some minor adjustments to allow for the negative sign. For both variables, an observed intensity of zero is treated as a special case. Such values are identified and automatically assigned a corresponding integer value of zero, then excluded from any further rescaling adjustments.

Let the observed intensities be denoted by $\{O_k : k = 1,2,....\}$. Associating the observed values with an integer-incremental scale, pairs $\{k, O_k\}$, that is integer and observed intensities are matched: for example for variable v3 for subject et1:

Integer-incremental scale	k	1	2	3	4	5	6	.....
Observed Intensity	O_k	1	3	5	7	8	10	.....

This is then used as a form of 'look-up table' (LUT) for rescaling the raw data: an observed intensity O_k will be reassigned as integer intensity k. So for example, an observed intensity of $O_4 = 7$ would be assigned as integer intensity 4, $O_6 = 10$ as integer value 6, and so on. The automatic assignment of an integer value of zero corresponding to an observed value of zero can be inserted as the first entry in the LUT once the rescaling of the remaining observed

intensities has been determined.

The rescaling is slightly more complex for variable v1, since the possible negative sign of the observed intensity must be taken into account. An observed intensity of zero is assigned an integer intensity of zero. Then all observed intensities less than zero are associated with negative integer intensities (incrementing backwards from zero) and observed intensities greater that zero are associated with positive integers (incrementing upwards from zero). For both variables v1 and v3 some minor adjustments may be necessary at the ends of the ranges of the assigned integer intensities. Because the observed data is subject to recording errors, some of the intensities close to the minimum or maximum observed values may be 'missing' and so the corresponding integer scale is slightly distorted. Missing values can usually be identified and the integer scale subsequently corrected by a simple analysis regressing the integer intensities against the observed values.

Look-up tables are formed in this way for all subjects for both variables v1 and v3. The conversion from unequally-incremented observed intensities to integer-incremental scale need only be carried out once. All further discussion assumes the data has been rescaled and for a given variable images are directly comparable between subjects. After rescaling, the variable previously written as v1 will be denoted by z1, and the variable previously written as v3 will be denoted by z2.

Although z1 and z2 are usually considered individually the methods developed in subsequent chapters are equally applicable to either and so the variable z will generally be used to represent either z1 or z2. Extending the notation further, then the intensity at a single pixel can be denoted by $\{z(i,j) : i,j = 1,2,\ldots,128\}$, where i and j correspond to column and row positions with respect to the image lattice.

2.3 Summary Statistics

2.3.1 Introduction

Before fitting any form of regression model to describe the images, summary statistics are evaluated for the raw data. This enables a rough quantitative assessment of qualitative features observed simply by looking at the images (Figures 2.1, 2.2a and 2.2b), for example the characteristic intensity levels for each tissue. The validity of assumptions such as a Normal distribution of intensities for each tissue and homogeneity of variance which will be made in later chapters can also be verified.

It must be stressed that the results are very subjective. It is difficult to be objective in discriminating between tissues and assigning pixels which fall in the 'intermediate' area between two tissues. Neither do the methods make any use of the spatial structure of the data. Although the methods covered later in this thesis are more theoretically involved and computationally more expensive much of the subjectiveness of the ad-hoc methods described here can be eliminated. Improved accuracy of estimation of features such as tissue area or shape should be sufficient justification for the extra time and cost involved in using the more technical methods.

Three ways of partitioning the data are considered. First, the variation in intensity across the image is assessed by looking at one-dimensional transects, fixing either the column or row position. Second, data are extracted for small homogeneous regions within each tissue. Third, the edges between tissues are drawn by eye and all pixels subjectively assigned to the appropriate tissue type. Attention is restricted to the fat and muscle tissues, since the number of pixels covered by bone and marrow tissues is too small to give a representative sample of intensities.

2.3.2 Summary results

The variation in intensity across the image can be assessed by

considering one-dimensional transects of the data. For example, the row position j is fixed and intensity values extracted for columns {i = 1,2,...,m}. Alternatively, the column position i could be fixed and intensities extracted for rows {j = 1,2,...,m}. Since the patterns are very similar only results for fixed rows will be presented. For subject et1, data for three fixed rows (the first two of which pass through the bone region) are plotted in Figures 2.3a and 2.3b. The plots illustrate the different mean levels which are characteristic of each tissue for the two variables. The 'significant' change in intensity at the edges of each tissue is obvious, though the edge between the fat region for the two legs is not always so clear.

It is possible to pick out more specific features for each variable. For variable z1 (IR) in Figure 2.3a, the difference in intensity for fat (high values) and muscle (lower values) is very clear. Particularly clear in row 58 are the negative intensities (around columns 44 and 85) which correspond to pixels of blood. The higher intensities around column 36 in row 52 correspond to pixels of marrow tissue; marrow pixels are present at a similar position in row 58 but row 64 is below the (bone) marrow and so only fat and muscle tissues are present. In Figure 2.3b the distinction between fat and muscle tissues is not so clear. However it is possible to distinguish pixels of both bone and marrow. This is well illustrated in row 58: moving across rows 34 - 40 (and again, rows 90 - 96) the lower values correspond to bone, with the higher intensities in between representing the marrow. For both variables the random variation about the mean intensity for each tissue and between different parallel rows is reasonably constant. This supports the assumption of constant error variance over the image, irrespective of tissue type.

The second set of results are obtained by looking at small, homogeneous regions within each tissue. The 128 x 128 array of intensity values are first printed out. Using this data matrix in combination with the colour photographs and grey-level images of Figures 2.1, 2.2a and 2.2b, regions of size 8 x 8 pixels are identified within the fat and muscle tissues, for both left and right legs. The four regions are chosen randomly in the sense that the

Figure 2.3a: Change in intensity across columns
Subject et1, Variable z1

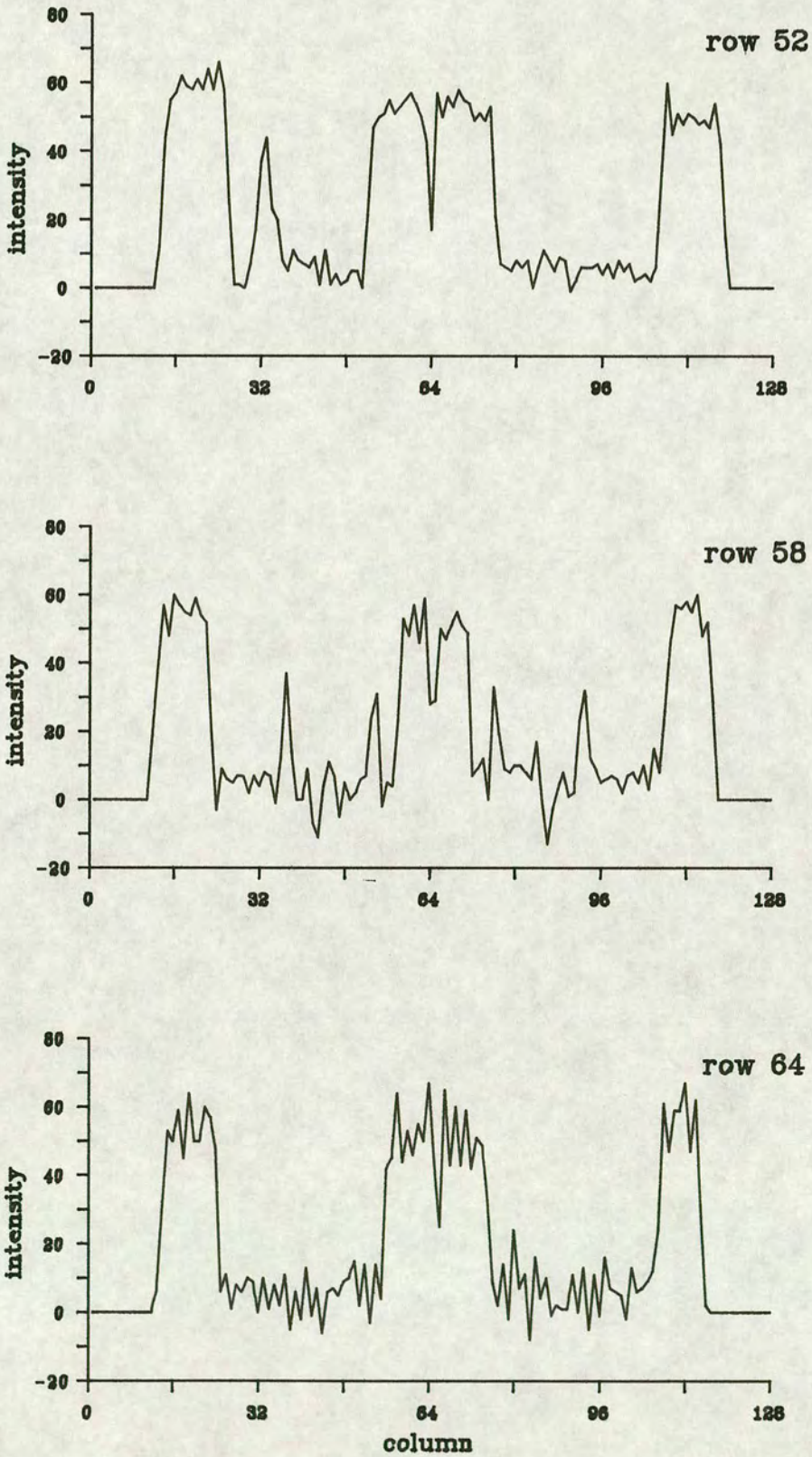
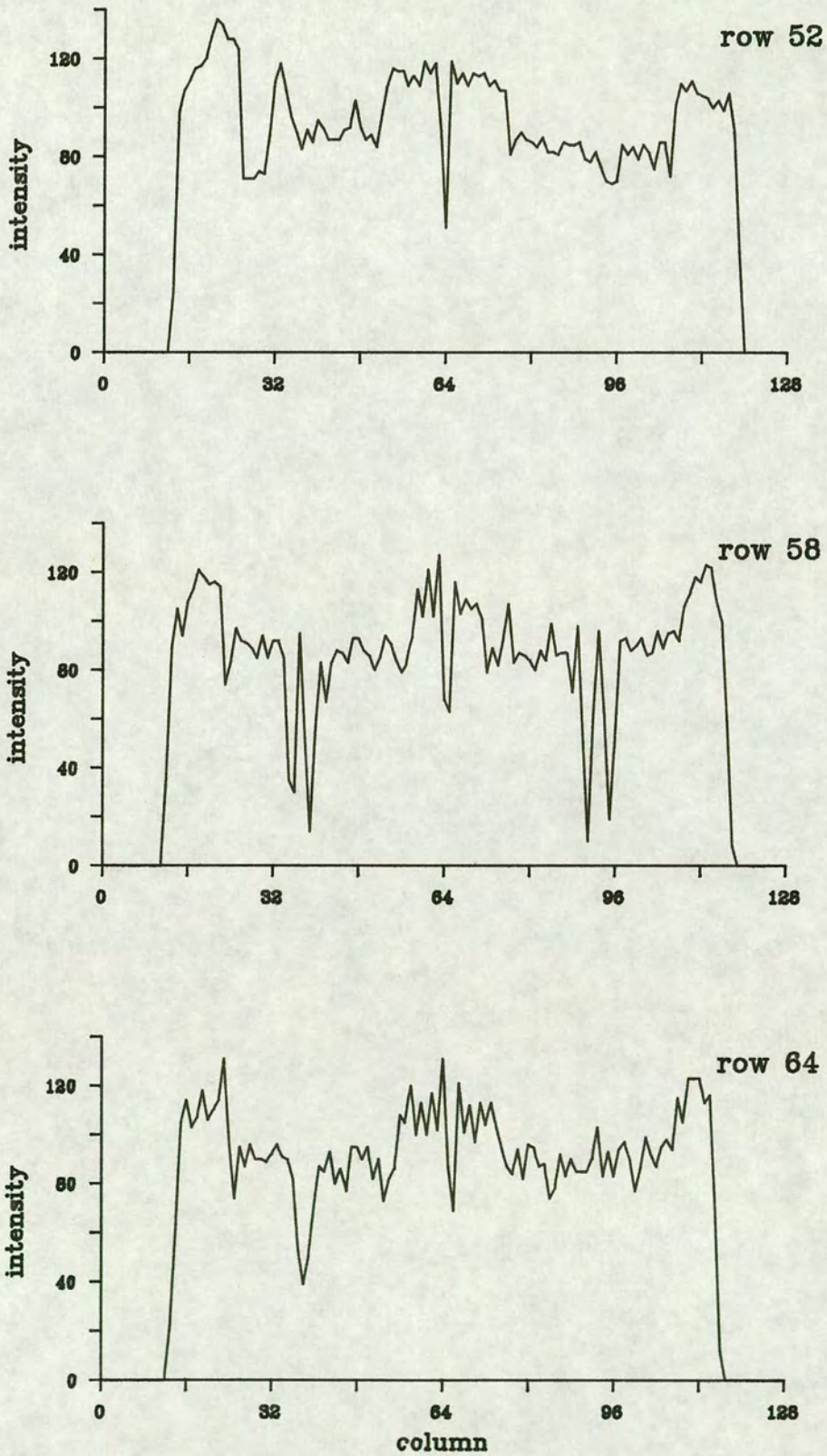


Figure 2.3b: Change in intensity across columns
Subject et1, Variable z2



regions are not specifically sampled at locations where the intensities are 'similar'. However the regions are sampled from approximately the same positions for each subject, to avoid confounding any real differences between regions with possible systematic changes in data values caused by the imager.

Ignoring the spatial structure (that is the relative position of each pixel within the 8 x 8 region) the mean and standard deviation of the 64 intensity values are calculated. The calculations are repeated for the fat and muscle regions in the left and right legs for both variables z1 and z2 for each of the four subjects et1, et2, ft1 and ft2. The results are presented in Table 2.1. In most cases there is a systematic decrease in mean level for the right leg in comparison with the left. Such systematic differences could be due to inhomegeneities in the magnetic field or RF radiation. It is accepted by anyone working with MRI that the PD signal (and anything depending on it) does vary from day to day. For example, the signal can be changed by the introduction of equipment into the room, or the temperature of the room. As a consequence, it is inevitable that the mean level for each tissue will be image dependent.

For a given subject and variable, the values of the standard deviation for the four regions are fairly consistent, again indicating that an assumption of constant error variance is reasonable. Histograms of the distribution of intensity values are plotted for all regions. For brevity the plots are not presented here, but do confirm the validity of the assumption of a Normal distribution of values for muscle and fat tissues.

The third set of results are based on the image area covered by the two legs, after excluding the 'background' air pixels. For the four subjects et1, et2, ft1 and ft2 the distribution of intensity values are plotted in Figure 2.4a, 2.4b for variables z1 and z2 respectively. The two distributions of values for muscle (low) and fat (high) are quite distinct for z1, though less well defined for z2. These plots indicate that for each tissue, an assumption of Normal distribution of intensity values is reasonable. For z1 there are also a small number of pixels with negative intensities

Table 2.1: Summary statistics for 8 x 8 homogeneous regions

Table 2.1a: Variable z1

Fat n = 64	Mean Left	Mean Right	St. Dev. Left	St. Dev. Right
et1	61.7	48.7	3.99	3.68
et2	62.1	52.1	4.34	4.67
ft1	63.9	51.3	4.75	4.25
ft2	67.3	61.7	6.00	4.53

Muscle n = 64	Mean Left	Mean Right	St. Dev. Left	St. Dev. Right
et1	7.0	5.5	3.40	3.42
et2	12.5	8.2	4.79	4.39
ft1	5.7	5.0	4.71	3.09
ft2	11.6	12.4	5.78	3.73

Table 2.1b: Variable z2

Fat n = 64	Mean Left	Mean Right	St. Dev. Left	St. Dev. Right
et1	125.3	104.5	5.51	4.46
et2	132.1	106.0	5.70	7.43
ft1	134.1	117.0	6.50	7.21
ft2	130.1	121.6	7.52	8.54

Muscle n = 64	Mean Left	Mean Right	St. Dev. Left	St. Dev. Right
et1	87.2	89.7	6.49	3.57
et2	96.4	90.6	8.15	5.68
ft1	87.2	89.9	5.67	5.15
ft2	91.7	88.8	4.85	4.78

Figure 2.4a: Distribution of intensities – Variable z1

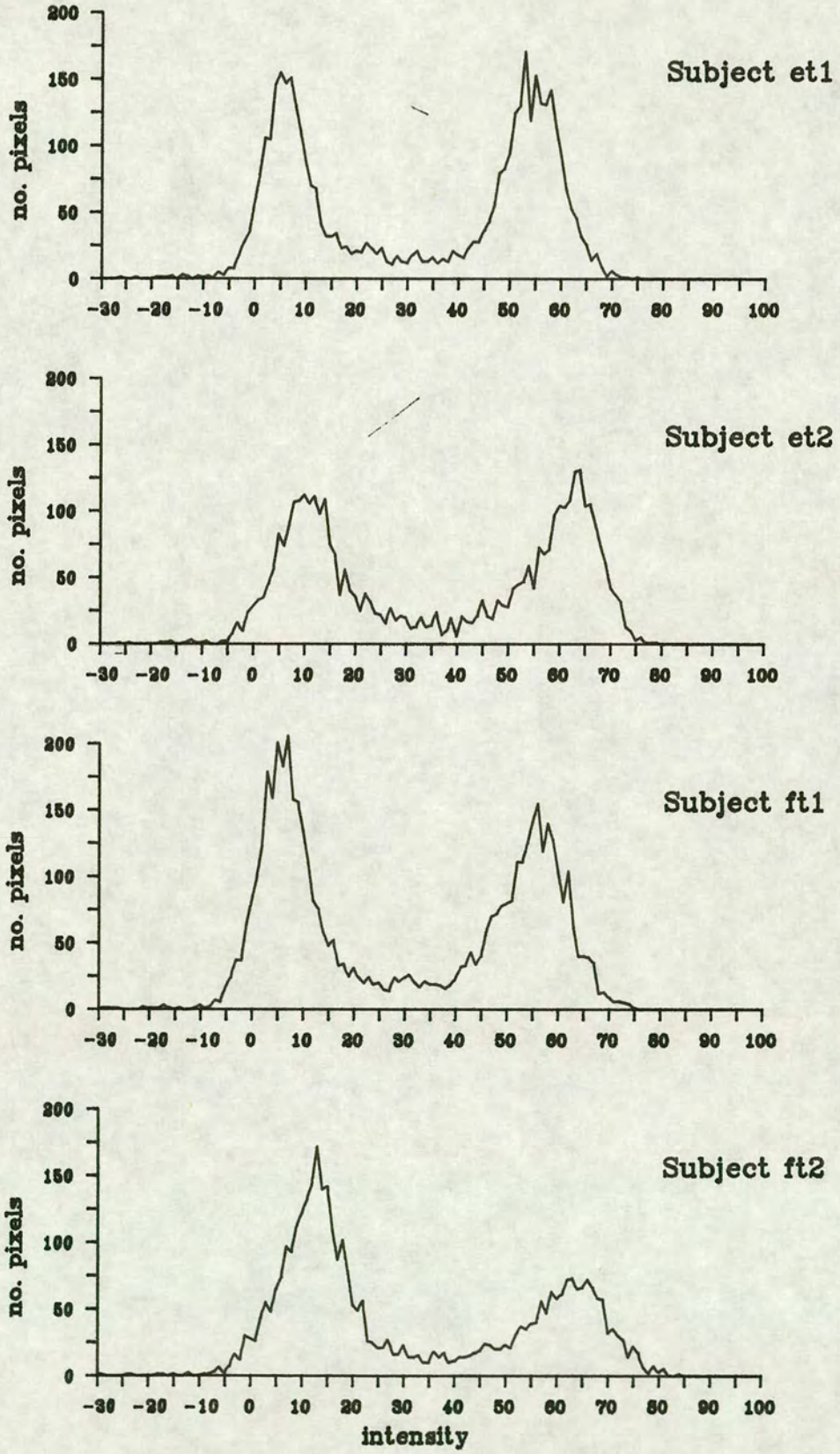
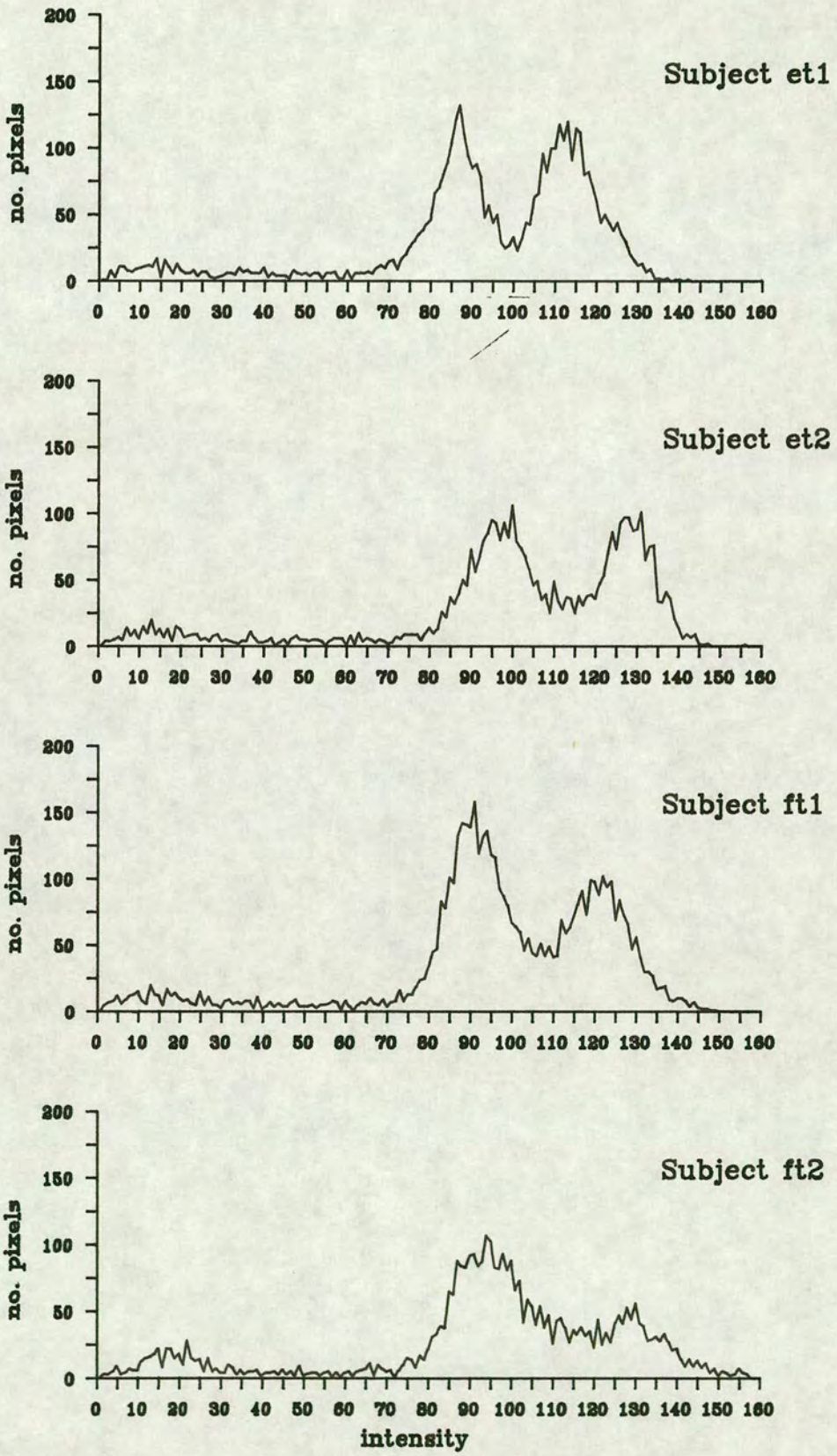


Figure 2.4b: Distribution of intensities – Variable z2



corresponding to blood (from the artery and vein passing through the muscle). The long left tail for z2 corresponds to pixels of bone.

The 128 x 128 data arrays (printed out for the second set of results in order to identify the 8 x 8 homogeneous regions) are now used to assign all pixels in the image to one of the groups fat, muscle or other. The two legs are considered separately. Note that this part of the analysis is very subjective: to obtain results without using time-consuming methods, simplifying assumptions are made concerning the form of data which must inevitably bias the results. For each image, the plots of Figure 2.4 are used to estimate a threshold intensity to discriminate between the distribution of values for fat and muscle tissues. 'Edges' are hand-drawn between tissue regions, and ignoring the spatial structure of the data (that is, the relative position of each pixel) all pixels falling within the outlined region are assigned as fat or muscle, for both left and right legs. Some allowance is made for the 'indentations' which occur on tissue edges, but because of the ad-hoc nature of this method of assigning pixels it is inevitable that some will be misclassified. This is particularly true of those falling on the edge between the fat and muscle tissues. Note that bone and marrow pixels are excluded from this assignment, but no allowance is made for any pixels of blood which may be present within the muscle region.

The mean and standard deviation of the intensities assigned to (left and right) fat and muscle tissues are calculated. To illustrate typical results, summary values for the subjects et1 and et2 for both variables z1 and z2 are given in Table 2.2. The decrease in mean value for the right leg compared to the left is apparent, though smaller in magnitude than for the 8 x 8 regions. In comparison to Table 2.1 for the 8 x 8 regions, the mean levels for muscle are quite similar. The fat mean level is noticeably smaller, probably due to pixels of muscle tissue which occur in 'indentations' of the fat/muscle edge being incorrectly assigned as fat tissue. The standard deviations for the four tissues for each subject and variable are again fairly constant. Values are higher than the equivalent statistics in Table 2.1: this suggests that although tissues are locally quite homogeneous there is greater variation in

Table 2.2: Summary statistics for assigning all tissue intensities

Table 2.2a: Variable z1

Fat	Mean		St. Dev.		Sample size n
	Left	Right	Left	Right	
et1	53.9	52.2	7.81	7.60	2231
et2	59.7	56.5	10.74	9.83	2076

Muscle	Mean		St. Dev.		Sample size n
	Left	Right	Left	Right	
et1	7.5	7.6	7.34	6.78	1560
et2	12.6	11.0	7.36	7.63	1427

Table 2.2b: Variable z2

Fat	Mean		St. Dev.		Sample size n
	Left	Right	Left	Right	
et1	115.6	111.5	8.24	7.13	2031
et2	124.5	119.5	10.93	12.18	2049

Muscle	Mean		St. Dev.		Sample size n
	Left	Right	Left	Right	
et1	86.7	85.7	6.82	7.13	1556
et2	97.1	91.5	7.61	7.52	1283

intensity across the image and the whole tissue is considered.

2.3.3 Conclusions

The two main conclusions from the analysis covered in this section are now summarised. Looking at the level of random variation, as measured by the standard deviation σ, the assumption of constant variance σ^2 appears justified by all three sets of results. Although the mean level for each tissue varies, the amount of variation about the mean level is virtually constant and independent of the tissue type and position in the image. Secondly, the assumption of a Normal distribution of intensities for each tissue holds, as is illustrated in Figure 2.4, with the separation of intensities into two fairly distinct distributions.

The methods used are subjective, particularly the assignment of all pixels discussed at the end of Section 2.2.2. Although not covered here, a more objective approach would have been to use a log regression model to fit a mixture of two Normal distributions to the data used to produce the plots in Figure 2.4 and hence obtain estimates of the mean variance as parameters of the Normal distribution. However although individual results may be slightly biased, the general trends in the data do, on the whole, justify the assumptions made. In practice if the assumptions are not met this is not too critical for the current estimation of edges. The two legs are compared with each other directly and because the data is smoothed prior to fitting the regression model the assumptions of constant variance σ^2 and homeogeneity of intensity for a given tissue are not so critical. The regression is used only as a means by which to estimate edge points rather than aiming to fit a regression curve exactly.

3 KERNEL REGRESSION FOR THE IDENTIFICATION OF EDGES

3.1 Locations of edge points

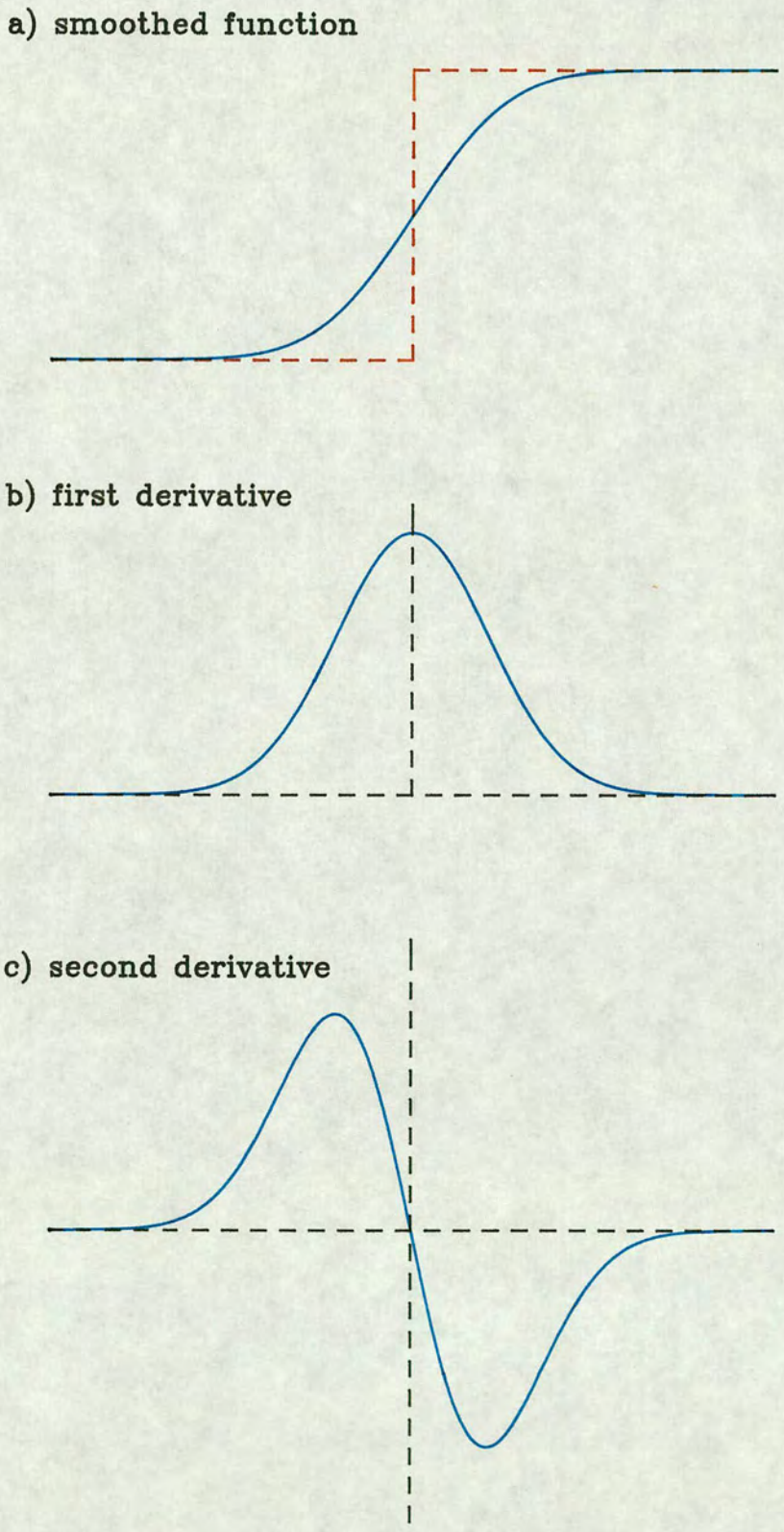
3.1.1 Definition of edges

The first stage in the estimation of edges is the identification of the locations of sets of points on the edges between different tissue regions in the image. Using the methods described in this chapter, points on the edges between all the different tissues can be identified and plotted so the edges can be assessed visually. Alternatively, the set of points lying on a single edge of interest can be identified and used as the input data for the curve fitting in the second stage of the edge estimation (Chapter 5).

An edge is said to be located at a point in the image where there is a 'significant' (to be quantified) change in intensity value within a local neighbourhood. The size of this neighbourhood depends on the type of intensity changes in the image: there may be a gradual change in intensity over several pixels, or alternatively the intensity may change significantly between two adjacent pixels. Whatever the size of neighbourhood considered, edges can equivalently be identified as locations where the gradient (first derivative) has a maximum or there is a 'zero-crossing' of the second derivative. A 'zero-crossing' is defined more precisely as the position at which the second derivative of the intensity function takes the value zero, and the sign of this second derivative changes.

A simple example is given in Figure 3.1, for ease of illustration shown here in one dimension but the principle extends similarly to two dimensions. A 'step-edge' is plotted in Figure 3.1a, typically occuring between two homogeneous regions where intensity is observed only at discrete data points. Superimposed on this step-edge is the underlying continuous function from which the data was sampled. The corresponding first and second derivatives of the underlying function are plotted in Figures 3.1b and 3.1c respectively, with the position of the edge marked for each.

Figure 3.1: Illustration of edge position (1–d)



The definition of an edge as a significant change in intensity between neighbouring pixels is not directly suited to discrete data. Anything other than constant intensity could result in an 'edge', depending on the magnitude of the change to be identified. Also, changes between homogeneous regions often occur more gradually over two or three pixels, rather than the 'ideal' step-edge shown in Figure 3.1a. Differentiation is not well-defined for discrete data so it is not immediately possible to use the properties of derivatives to identify edges. However because the observed data is sampled from an underlying continuous function, it is valid to estimate image intensity at any point in continuous space, achieved by locally averaging the discrete data using some form of 'smoothing'. The aim is to eliminate (or 'smooth out') 'small' changes in intensity, corresponding to noise or local fluctuations and which should not be identified as edge positions. Greater differences between discrete intensity values would still result in a significant change in smoothed intensity, so 'true' edge points are still identified.

Existing methods can be grouped into quite broad classes according to whether they use first or second derivative operators (Section 3.1.2.), and whether directional or isotropic derivatives are calculated (Section 3.1.3). Criteria for the two alternatives are now summarised in more detail.

3.1.2 First or second derivative?

Canny (1986) identified three goals when identifying edges: edges should be located accurately, only 'true' edges should be detected and each edge should only be detected once. Expressing the three goals mathematically an optimal operator was derived, which in one dimension was well approximated by the first derivative of a Gaussian. The disadvantage of such first derivative operators is that there is often a broad peak around the maximum. Edge location accuracy can be improved by a maximum detection step, to identify the local maximum in the neighbourhood of this peak.

Edge locations can equivalently be identified as zero-crossings

of the second derivative. Zero-crossings correspond to points of inflexion and so need not necessarily correspond to maxima in the first derivative (so indicating the position of an edge). Zero-crossings also correspond to a change in the rate of intensity variation and so spurious edges may be identified. However the first derivative at such locations is small and can be used as a condition to exclude such spurious points: see for example Haralick (1984) and Ulupinar and Medioni (1990).

3.1.3 Directional or isotropic operators?

Computational efficiency and simplicity of the edge identification algorithm are two criteria for the choice between directional or isotropic (equivalently orientation-independent or rotation-invariant) operators. Using derivatives calculated with respect to a particular direction edges can be located more accurately, but more than one derivative is usually calculated to identify all possible edges which may occur at any orientation in the image. Canny (1986) discussed briefly the computation of several directional derivatives which were then combined into a single operator. Torre and Poggio (1986) concluded that in a noise-free image, only two directional derivatives were necessary. In comparison, Haralick (1984) calculated a single directional (second) derivative taken in the direction of the gradient, using local interpolation to estimate the underlying intensity function.

Isotropic operators respond equally well to edges at any orientation and so only a single operator is required. Typical examples are the Laplacian or the directional derivative along the gradient, both second derivative operators discussed by Torre and Poggio (1986). Directional derivatives are only approximated by isotropic operators, with the bias for isotropic operators greatest for curved edges. Marr and Hildreth (1980) proposed conditions on local intensity variation for the approximation to hold and concluded that they should be satisfied for most natural images. A further important property of isotropic operators is that they ensure closed zero-crossing contours for the identified edge points, which is not

generally the case for directional derivatives (Torre and Poggio, 1986).

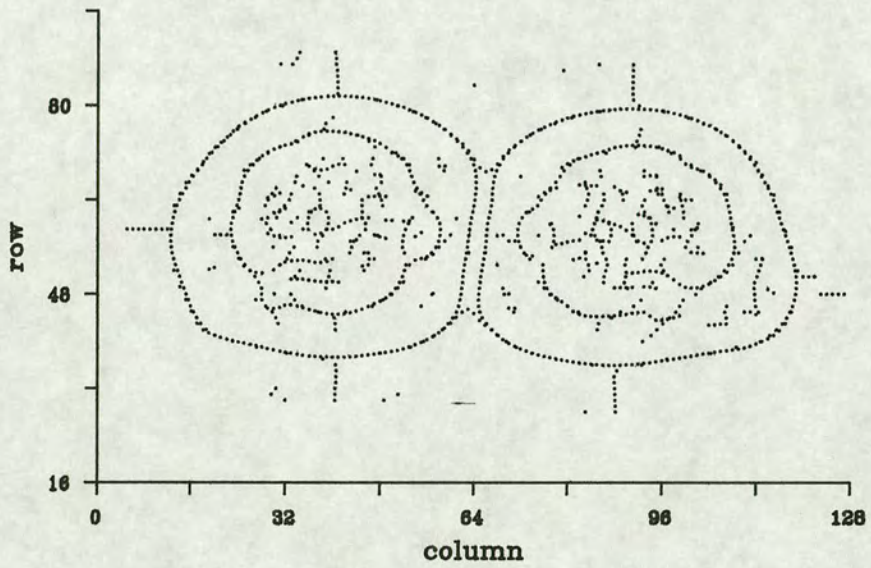
To a certain extent, the choice of operator and derivative to be calculated depends on the application and any further analysis of the set of edge points. Using the two derivative operators, edge points are identified for the grey-level image (Figure 2.2a) for subject et1, variable z1. The points are plotted in Figure 3.2. Prior to the calculation of derivatives the image is first smoothed using a Gaussian function; this smoothing and the algorithm used to identify edge points will be discussed later in this chapter.

The edge points plotted in Figure 3.2a are identified as local maxima in the first directional derivative. At each pixel, partial derivatives are calculated in the horizontal and vertical directions, where the horizontal direction is taken to be parallel to the rows of the image lattice, and the vertical direction parallel to the columns. The two partial derivatives are then combined to obtain the maximum gradient at each pixel position. The grid of maximum gradient values is searched sequentially but as can be seen from the figure this results in some false 'edge' points identified as local maxima in a single direction, rather than a global first derivative maximum. Points identified as zero-crossings of the Laplacian operator are plotted in Figure 3.2b.

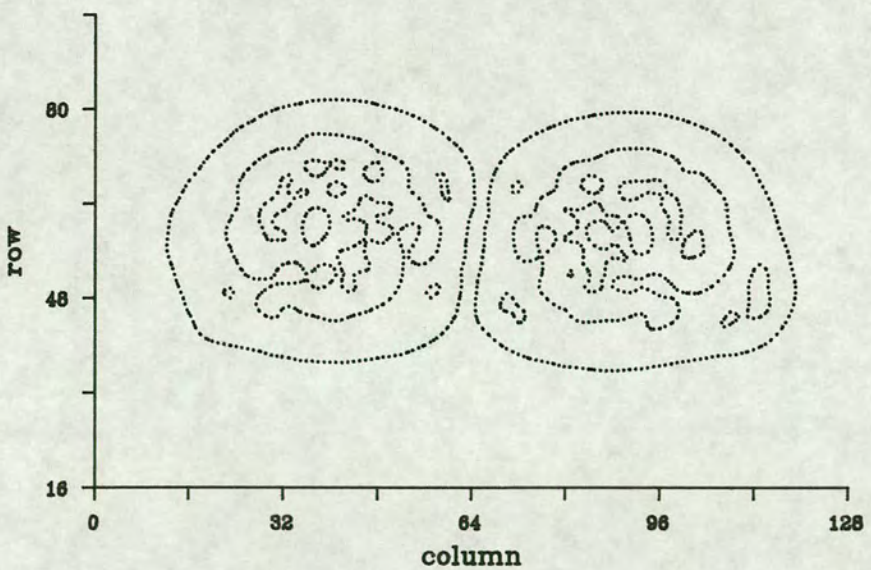
Comparing the two plots, the set of points in Figure 3.2b gives the 'best' representation of edges. This visual evidence combined with the previously-discussed properties suggests that the Laplacian is a suitable operator for this application. Thus (unless otherwise stated) it is assumed throughout the remainder of this thesis that edge points are identified as zero-crossings of the Laplacian of the smoothed image.

Figure 3.2: Edge points: Gaussian kernel, $\lambda = 1.5$

a) Maximum of first derivative



b) Zero–crossings of second derivative (Laplacian)



3.2 The data and methods of smoothing

3.2.1 The data

The data consist of a set of m^2 observations $\{z(i,j) : i,j = 1,2,\ldots,m\}$, where (i,j) are the two explanatory variables, the column and row locations on the m x m lattice; for this image type m = 128. The value of the intensity response variable z, observed at position (i,j) is given by $z(i,j)$. A model is required relating observed intensity $z(i,j)$ to (i,j), the position on the image lattice. Here only a regular lattice of design points is considered but the methods can be generalised to unequal spacing. The data is assumed to follow a model of the form:

$$z(i,j) = \mu(i,j) + \epsilon(i,j), \qquad i,j = 1,2,\ldots,m, \qquad (3.1)$$

where $\mu(i,j)$ is the intensity function to be estimated and $\epsilon(i,j)$ is the random error, assumed to be a zero mean uncorrelated random variable, with variance σ^2. The estimation of the error variance σ^2 will be discussed in Chapter 4.

Although the image is observed only at discrete lattice points, μ, the underlying intensity function, exists in continuous space. As already noted, a major theme for this study is the estimation of the intensity function at any point (x,y) in continuous space. Thus a method is required by which a smooth function μ_λ can be fitted in order to model the underlying continuous function μ. A nonparametric approach, often used in biological or medical applications when there is no prior information about a suitable model, would seem the most appropriate. Unlike a parametric approach, no assumptions are made about the form of model to be fitted. Nonparametric methods rely heavily on the data and any assumptions made are concerned only with qualitative properties, such as smoothness or differentiability. A regression curve, or surface in two dimensions is fitted to the data but there is great flexibility in the exact form of this curve.

A general nonparametric model is given by (1.2), where the smoothed estimate is written as a linear weighted average of the

observations. Often differences between nonparametric methods can be reduced to different ways of defining these weights. Several nonparametric methods for fitting a smooth function are summarised in Section 3.2.2, before the method used in this study is introduced in Section 3.2.3.

3.2.2 Nonparametric regression methods

Many methods of smoothing have been proposed, but the most commonly used estimators are those which have a simple structure and are not designed for specific problems. A general review of smoothing techniques was given by Collomb (1981). Eubank (1988) included chapters on series estimators and spline smoothers, while Härdle (1990, Chapter 3) considered k-nearest neighbour, orthogonal series and spline estimators in some detail. Several methods which were considered for the current images are now summarised. For some of these methods, reasons are given as to why they were not implemented in this thesis.

(*i*) Global regression surface. This is fit to the data as some polynomial function of the two explanatory variables (i,j). However this results in a representation which is too smooth.

(*ii*) Thin-plate or Laplacian splines are a generalisation to two dimensions of the one-dimensional cubic spline, which is often used for fitting regression curves to univariate data (Silverman, 1986). The fitting of splines for multivariate data was considered by Eubank (1988, Chapter 6), though Sibson (1988) observed that only in two dimensions was the resulting spline smooth enough to be of any use for data analysis. The thin-plate spline was estimated by fitting piecewise quadratics over the data sites. Because this method can only cope with a few hundred data sites and is more normally applied to scattered data, it is not particularly suited to the present large, regularly spaced data sets.

(*iii*) Natural neighbour splines (Sibson, 1980). These are a further generalisation of the spline method, which in two dimensions

give results very similar to the thin-plate splines. Fitted values and gradients are evaluated at the data sites and a quadratic function is bridged across them. Fitting is a localised procedure within a neighbourhood (to be defined) of each data site, so that only intensities which are 'reasonably' close influence the value at a given data site. This localization means that it is now possible to apply the method for 100,000 or more data sites. However, as for the thin-plate splines, this method is more commonly used for scattered data, rather than for observations on a regular lattice.

(*iv*) Local estimation. This approach belongs more to the field of computer vision, rather than being a method of nonparametric regression. The basic aim is to improve resolution by reconstructing the underlying image intensity surface from the discrete, sampled observations. Haralick (1986) estimated the underlying intensity from observations at pixels in a given neighbourhood. Discrete Chebychev polynomials in the row and column coordinates were used for the interpolation.

3.2.3 The kernel estimate

In this study, the underlying function μ is estimated by nonparametric kernel regression. Kernel estimates were used first for density estimation (Rosenblatt, 1956) and introduced for regression by Priestley and Chao (1972). For a general introduction to kernel regression, see Müller (1988) or Eubank (1988). This approach is favoured for its simplicity and computational advantages; the estimators have good statistical properties and result in a model which is asymptotically as efficient as fitting splines. One disadvantage of the kernel method is that it is not resistant to the affect of outliers. Each observation z(i,j) can contribute to the smoothed estimate at any point within a local neighbourhood of (i,j), so a single outlying observation can adversely affect the intensity estimated at any point within this neighbourhood.

As outlined in Section 1.2, smoothed estimates $\hat{\mu}_\lambda$ are calculated as weighted averages of the observations; here the weights are

defined using kernel functions K(x-i,y-j). Substituting this form of weights into (1.2), a more explicit form of linear estimator is given by:

$$\mu_\lambda(x,y) = \frac{1}{m^2\lambda^2} \sum_{i=1}^{m} \sum_{j=1}^{m} z(i,j)\, K\left[\frac{x-i}{\lambda}, \frac{y-j}{\lambda}\right]. \qquad (3.2)$$

The amount of smoothing is controlled by the parameter λ. Using (3.2), intensity can be estimated at any point in continuous space; of specific interest here are the individual sets of (x,y) coordinates lying on the edges between different tissue regions. Eubank (1988) listed five univariate estimators for μ, two of which will be discussed in more detail in Section 3.3.1. With two explanatory variables (column and row lattice coordinates), a bivariate kernel function is required (Section 3.3.2), but the form of estimator μ_λ is unchanged.

An expression of the form (3.2) assumes that each estimate is a weighted average of all the observations on the m x m lattice. Specific boundary kernels have been defined which adjust the form of the function fitted near the boundary of the observed data: see for example Müller (1988). Such boundary kernels are not considered here: the 'wrapping-round' of the image on a torus (Section 2.2.2) means that in effect the observed data extends infinitely, without boundaries. The estimator (3.2) will be made more general in Chapter 4, and the summation limits modified to take into account the 'wrapping-round' of the data. However in this chapter, when introducing the kernel regression method, to be consistent with the literature the summation limits used will be {i,j = 1,2,....,m}.

The kernel K must satisfy certain moment conditions related to qualitative properties of the chosen function. Only the minimum detail is given here so as to indicate the shape of a suitable function, see Eubank (1988, Chapter 4) for mathematical expressions. Theoretically, functions with finite support would be favoured, though in practice those with infinite support 'suitably truncated' for discrete data are also used. The function K should be symmetric

about a maximum at the origin so that z(i,j) contributes most to the intensity estimated at position (i,j). It is not essential that individual weights K(x-i,y-j) are greater than or equal to zero, but if this is satisfied for all (i,j), then together with the condition of weights integrating to one, this implies that the kernel K is a probability density function (PDF). The order of a kernel is a further consideration, determined by 'how smooth' the regression function can be assumed to be, or equivalently, how many continuous derivatives it will have. The order of a kernel is mentioned again briefly in Section 3.3.3, but for a more complete discussion see Eubank (1988, Chapter 4). The choice of a specific function for K (Section 3.3.3) may also be limited by practical considerations, for example it may be preferable to restrict attention to functions that are zero outside some fixed interval to avoid numerical underflow on a computer.

Whichever kernel is selected, it is indexed by a parameter λ. The role of this smoothing parameter and criteria for its choice are discussed in Section 3.3.4, while different methods for 'automatic' selection are compared in Chapter 4. It is worth noting here that although the selection of the linear estimator μ_λ, the kernel function K and the parameter λ are all discussed in separate sections, the choice of one of these is not independent of the choice made for the other two. This is illustrated by a simple example in Section 3.3.3, where the exact form of the weights $w_\lambda(x-i,y-j)$ used for smoothing the data is defined. A description of how the estimates are calculated is given in Section 3.4, while Section 3.5 covers the identification of sets of points lying on the edges between different tissue regions.

3.3 Defining the kernel weights

3.3.1 Choice of form of linear estimator

Recall from Section 3.2.1 that a linear estimator $\mu_\lambda(x,y)$ is of the general form (3.2). This estimator is approximating a convolution

integral of the form:

$$\int\int \mu(x',y')\ K(x - \lambda x', y - \lambda y')\ dx'dy',$$

where (x',y') is the position at which the estimate is required and integration is over the whole continuous image. The summation form of the estimator (3.1) is an approximation for this integral when the true intensity μ is unknown and replaced by its observed value z, which is only sampled at discrete data points.

Forms of estimators differ only in the definition of the weights. Priestley and Chao (1972) proposed the most basic form of estimator for regular design points, later generalised for unequally-spaced data, and then modified by Benedetti (1977) to ensure that weights summed exactly to unity. The Benedetti estimator has also been considered for random design points (Nadaraya, 1964, Watson, 1964).

The convolution integral can also be approximated by an 'integral' form of estimator which interpolates between data points and so may be expected to give a better approximation to the underlying continuous function: see Gasser and Müller (1979, 1984). Such estimators are not considered here, although Eubank (1988) and Müller (1988) both used an integral form of estimator when defining the kernel model and deriving theoretical properties. However the general methodology they cover is common to all forms of kernel estimators of μ, since each can always be written as a weighted sum (3.2). Eubank compared the different forms, concluding that they all had similar properties for equally-spaced data, although the 'integral' form of estimator was superior if data was irregularly spaced.

Eubank (1988, Chapter 4) listed five variants for univariate data, but only the two most applicable for equally-spaced data are considered here. The two forms are bivariate extensions of univariate estimators. They differ only in the normalisation (discussed further in Section 3.3.3): the summation to unity of individual weights is only approximate for the first form but exact for the second. They

are defined in terms of a general bivariate kernel, written as a product of two univariate functions K, as will be described in Section 3.3.2.

The first form of estimator was proposed for univariate data by Priestley and Chao (1972); for bivariate data it is given by:

$$\mu_\lambda(x,y) = \frac{1}{m^2\lambda^2} \sum_{i=1}^{m} \sum_{j=1}^{m} z(i,j)\, K\left[\frac{x-i}{\lambda}\right] K\left[\frac{y-j}{\lambda}\right]. \qquad (3.3)$$

Benedetti (1977) used the second form for univariate data; it is extended to the bivariate case as:

$$\mu_\lambda(x,y) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} z(i,j)\, K\left[\frac{x-i}{\lambda}\right] K\left[\frac{y-j}{\lambda}\right]}{\sum_{s=1}^{m} \sum_{t=1}^{m} K\left[\frac{x-s}{\lambda}\right] K\left[\frac{y-t}{\lambda}\right]}. \qquad (3.4)$$

Taking into account the function chosen for the kernel K, the relative merits of the two estimators are assessed in Section 3.3.3 and reasons given justifying the use of (3.4) in this study.

3.3.2 Bivariate kernel functions

A bivariate kernel function is required for this application, with column and row grid coordinates as the two explanatory variables. Most of the kernels previously suggested have been for univariate data and must be modified for bivariate data. Fukunaga and Hostetler (1975) and Singh (1976,1981) estimated partial derivatives. Product kernels, proposed for multivariate density estimation (Cacoullos, 1966, Epanechnikov, 1969) were used for multivariate kernel regression by Müller (1988, Chapter 6). Here product bivariate kernels are used, defined as the product of two univariate functions:

$$\frac{1}{\lambda^2} K\left[\frac{x-i}{\lambda}, \frac{y-j}{\lambda}\right] = \frac{1}{\lambda_i} K_i\left[\frac{x-i}{\lambda_i}\right] \frac{1}{\lambda_j} K_j\left[\frac{y-j}{\lambda_j}\right], \qquad (3.5)$$

where each univariate kernel K_i, K_j must satisfy the necessary moment conditions. The general case (3.5) allows for different kernels K_i,K_j and different levels of smoothing (determined by λ_i,λ_j) for the two univariate functions; a common value is more usual, denoted here by K and λ respectively.

By modifying the kernel function, any of the proposed estimators μ_λ (Section 3.3.1) can be extended for bivariate (or multivariate) design points. Provided the bivariate function is separable, that is it can be factorised, then in practice the number of computations can be reduced, by decomposing the kernel into two univariate functions. Intensities are estimated by smoothing successively with two (orthogonal) univariate functions. Data is first smoothed over columns and intermediate estimates stored; these intermediate 'column-smoothed' estimates are then further smoothed over rows. The estimate after the second stage of smoothing is equal to that which would have been estimated if the bivariate function had been used directly on the data. Although there is an increase in the number of estimates to be stored, for an m x m data array the number of computations can be reduced from order m^2 to order 2m, a considerable saving for large m. In the field of computer vision, considerable attention has been given to methods of improving computational efficiency, and in particular for the calculation of derivatives for the identification of edges. Such technicalities are not a prime concern for this study, but see for example Sotak and Boyer (1989) for a summary.

3.3.3 Choice of kernel function

The choice of function for the kernel K will, to some extent, depend on the form of the data and any subsequent analysis of the smoothed estimate. Many of the existing kernels were proposed for univariate data, so the ease of extension to bivariate data must also be taken into account. Thus some criterion is required, on the basis of which different functions can be compared in order to determine the 'best' kernel.

A statistical approach often involves a quantitative comparison of possible functions, defining some measure of closeness between the true and estimated functions. Several such distance measures were outlined by Härdle (1990, Chapter 4), including the MSE criterion to be defined in Chapter 4. Considering the asymptotic properties of the MSE led Gasser and Müller (1979, 1984) to define a wide class of polynomial kernels; see also Müller (1988). Kernels previously proposed for density estimation (Epanechnikov, 1969, Benedetti, 1977) are special cases of these polynomial kernels. However, it is generally concluded that although the shape of the functions differ, there is little to choose between them in terms of MSE.

In computer vision, an attempt is made to relate the choice of kernel function (termed 'filter' in the literature) to a proposed model for early visual processing of the image. The emphasis is on the physical properties of the underlying image which result in the observed changes of intensity. Marr and Hildreth (1980) identified two conflicting physical requirements which should be satisfied when the image is smoothed. First, the range of scales over which intensity changes take place should be reduced. Second, the estimate at each point in the image should be calculated as an average of neighbouring points, rather than smoothing over all widely scattered points. So, although the terminology and motivation behind these two criteria for computer vision differ from the more familiar statistical concepts, the objective is the same. A kernel function should be chosen to obtain the globally 'best' fitting model, by reducing local variation between neighbouring values but without greatly increasing bias, so edges can be located accurately.

The kernel used in this thesis is the Gaussian function, often denoted by G. It is easily extended for bivariate data, when the function is of the form:

$$\frac{1}{\lambda^2} \; K\left[\frac{x-i}{\lambda}, \frac{y-j}{\lambda}\right] = \frac{1}{2\pi\lambda^2} \exp\left[\frac{-1}{2\lambda^2}\{(x-i)^2+(y-j)^2\}\right], \qquad (3.6)$$

illustrated in Figure 3.3a.

Figure 3.3a: Bivariate Gaussian function G

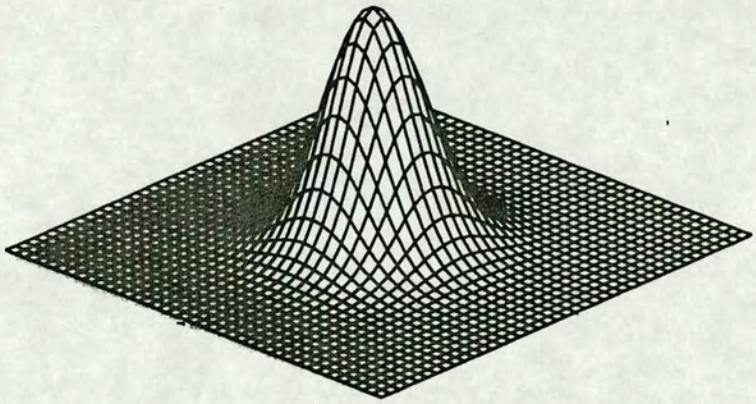
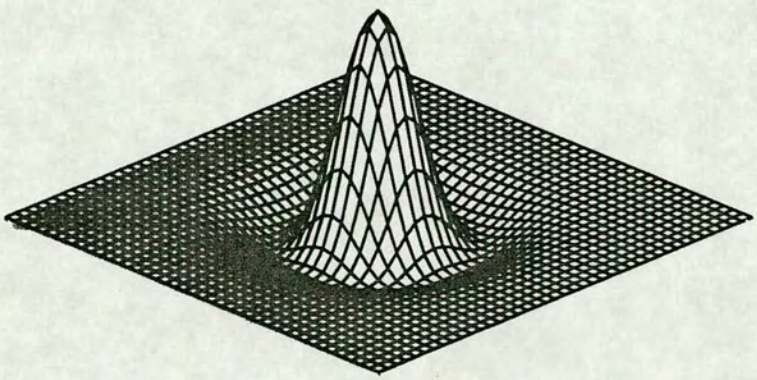


Figure 3.3b: Bivariate Laplacian of Gaussian function LoG



The Gaussian was shown (Leipnik, 1960) to be the only function which optimises the trade-off between the two conflicting requirements discussed by Marr and Hildreth (1980). Defining three goals for edge identification Canny (1986) derived mathematically an optimal filter and concluded that it was well approximated by the Gaussian for most real images. It is one of the kernels considered by Härdle (1990) who showed that there was little difference (in terms of MSE) between the Gaussian and the polynomial kernels of Gasser and Müller (1979).

The Gaussian is extensively used in computer vision, favoured for its simplicity and computational efficiency. The bivariate kernel can be factorised into the product of two univariate functions, so the number of computations can be reduced, as discussed in Section 3.3.2. The Gaussian is also rotation-invariant, an important property for the current data. Multiple images for a single subject will typically be collected over a period of time, but as a result of the rotation-invariance, several images can be compared directly without the subject being repositioned identically.

It is easy to adjust the amount of smoothing by varying λ. When identifying edges as zero-crossings of linear derivative operators, in particular the Laplacian used here, then the Gaussian is the only kernel which has the 'nice scaling behavior' of Yuille and Poggio (1986): no additional zero-crossings are introduced as the amount of smoothing is decreased, but existing ones never disappear. This is important if comparing edges identified in a single image smoothed with several different values of λ.

The Gaussian kernel has a theoretically infinite support which must be suitably truncated in practice for data observed only at discrete points. The level of truncation should be such that individual weights are effectively zero at the limits of the support of the kernel, to avoid possible spurious features as a result of too much truncation. This is to be balanced against the additional computation when defining an unneccessarily wide support for the kernel, so multiplying observations by weights that are effectively zero. It is a property of the Gaussian distribution that 99.7% of the

observations lie within ±3 standard deviations (of the mean), equivalent here to a distance of ±3λ from the position (x,y) at which the intensity is to estimated.

Preliminary analysis suggested that ±3λ is not always sufficient to completely smooth out each observation. The support was increased to ±10λ and individual weights renormalised and compared with those from the ±3λ support. For the ±10λ support, those weights at points within ±3λ of (x,y) were almost identical to those calculated if only a ±3λ support is assumed. However, with this increased support it could be seen that non-zero weights did extend to a distance of approximately ±5λ from (x,y). Although the weights were very small, it did indicate that observations at a distance greater than ±3λ should be included in the estimate $\hat{\mu}_\lambda(x,y)$. Therefore in this study the support of the Gaussian kernel is truncated at ±5λ. For the current data this limit provides the best compromise between incomplete smoothing at ±3λ and unnecessary computation (using virtually zero weights) at ±10λ.

As mentioned at the end of Section 3.2.3, the choice of a function for the kernel K is not necessarily independent of the form of estimator chosen for μ_λ. To some extent, the choice is also dependent on the value of the smoothing parameter λ. This lack of independence is illustrated by the following example comparing estimators (3.3) and (3.4), substituting the Gaussian function (3.6) for the general kernel K. Bearing in mind the desirable properties of the kernel estimate, the relative merits of (3.3) and (3.4) are assessed before choosing the estimator to be used for this study. For simplicity estimates are calculated in one dimension; in two dimensions a second univariate function smoothes orthogonal to the first, so the only difference for the bivariate estimate is in the scaling.

The two forms of estimator (3.3) and (3.4) differ only in the definition of the weights, and in particular, the normalisation of the weights. Using a Gaussian kernel, individual weights for estimator (3.3) are given by:

$$w_\lambda(x-i,y-j) = \frac{1}{2\lambda^2} \exp\left[\frac{-1}{2\lambda^2} \{(x-i)^2+(y-j)^2\}\right], \qquad (3.7a)$$

and for estimator (3.4):

$$w_\lambda(x-i,y-j) = \frac{\exp\left[\frac{-1}{2\lambda^2} \{(x-i)^2+(y-j)^2\}\right]}{\sum_{s=1}^{m}\sum_{t=1}^{m} \exp\left[\frac{-1}{2\lambda^2} \{(x-s)^2+(y-t)^2\}\right]}. \qquad (3.7b)$$

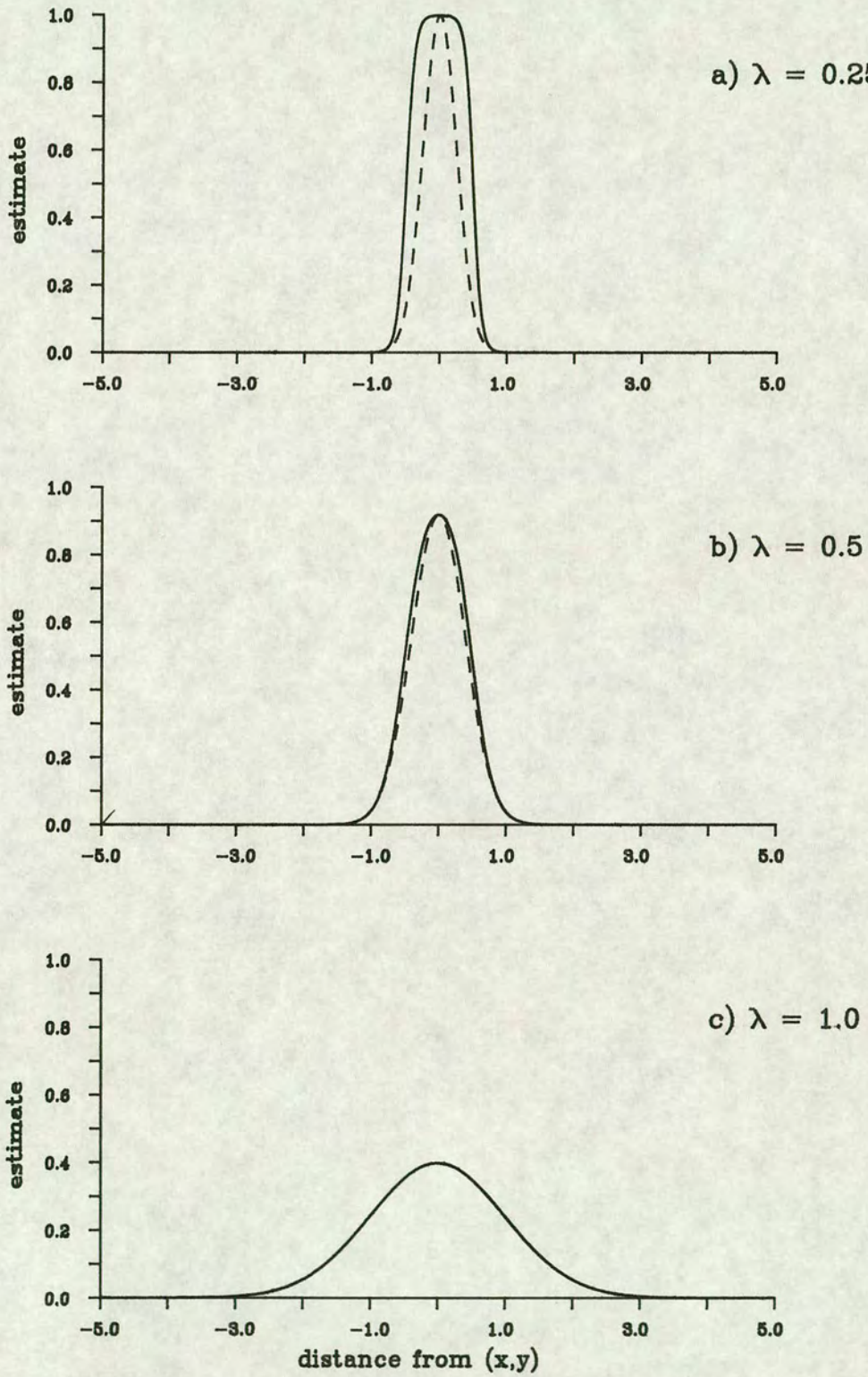
The importance of this here is that normalising using the sum of weights in the denominator of (3.7b) means that the kernel weights are no longer sampled from an exact Gaussian function, though the magnitude of the difference between weight (3.7a) and (3.7b) is very small.

A second desirable property for the estimator is that weights $w_\lambda(x-i,y-j)$ should sum to unity. This is not satisfied for weights (3.7a): although individual weights integrate exactly to unity, for discrete points the summation is only approximate. Thus each of the estimators satisfies only one, but not both, of these two conflicting properties: weights (3.7a) are exact Gaussian weights, but only with weights (3.7b) do individual weights sum exactly to unity.

The two forms of estimator with increased amounts of smoothing are compared in Figure 3.4. Estimates calculated using (3.3), that is, weights (3.7a) are rescaled to allow more direct comparison with those from (3.4), that is weights (3.7b). Although there is an observable difference for $\lambda = 0.25$ (Figure 3.4a), estimates are virtually identical for $\lambda = 0.5$ (Figure 3.4b) and cannot be distinguished visually for $\lambda = 1.0$ (Figure 3.4c). With values of $\lambda < 0.5$, data are no longer smoothed over neighbouring sites, so such λ would not be used in practice. This approximation (to a true Gaussian function) of the weights in estimator (3.4) is sufficiently accurate for all λ of practical interest.

Thus form (3.4) is the estimator used in this study. The kernel is still termed a Gaussian function, but with the understanding that this is only a (very close) approximation when using this estimator.

Figure 3.4: Comparison of estimators: Gaussian kernel (1−d)
– – – – estimator (3.3) ——— estimator (3.4)



The Gaussian kernel (3.6) can be substituted for general kernel K in (3.4), so a more explicit form of the linear estimator used here is given by:

$$\mu_\lambda(x,y) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} z(i,j)\,\exp\left[\frac{-1}{2\lambda^2}\{(x-i)^2+(y-j)^2\}\right]}{\sum_{s=1}^{m}\sum_{t=1}^{m}\exp\left[\frac{-1}{2\lambda^2}\{(x-s)^2+(y-t)^2\}\right]}. \qquad (3.8)$$

3.3.4 Choice of smoothing parameter λ

The smoothing parameter λ is perhaps the most important single factor in determining how well the fitted model estimates the underlying intensity function. Although a full discussion of the different methods for selecting λ will be given in Chapter 4, a few comments are included here by way of introduction. The value of λ determines the size of the local neighbourhood over which the observed data is smoothed, that is how far observations z(i,j) are allowed to be from the point (x,y) and still contribute to the estimate $\hat{\mu}_\lambda(x,y)$. With smaller λ, estimates are much rougher, relying heavily on information near (x,y) but as λ increases, smoother estimates result.

With regard to edge detection, the value of λ will also determine the accuracy of the set of edges identified, balancing bias and variance of the fitted model. For 'small' λ , there is negligible bias in position so edges can be accurately located, but local variability of the smoothed function may result in the detection of 'false' edges and failure to identify all the 'true' edges. As smoothing increases with λ, the detection accuracy improves, correctly identifying 'true' edges and no longer including 'false' edges due only to local fluctuations. However this reduced variability is at the expense of increased bias, so individual edges are located less accurately.

A theoretical expression for the optimal λ can be derived for

polynomial kernels, mentioned in Section 3.3.3 as possible functions for K. More generally, there are two alternatives which are considered here for estimation of a smoothing parameter for the Gaussian kernel. As a first step in any nonparametric regression analysis, it is often recommended to smooth the data at different levels of λ, which may highlight any strong patterns in the data. This represents a subjective approach for estimating λ, since the value which gives the 'best' results is assessed visually. Here data is smoothed at different levels of λ and the edge points identified and plotted using methods which will be described in Section 3.5.

Secondly, data-based methods are considered, which represent a more automatic approach to smoothing parameter estimation. The methods discussed are cross-validation (for the data, and for first and second differences of the data); the 'Rice' method, the chi-squared method and the empirical degrees of freedom method. Definitions in terms of the kernel model will be given in Chapter 4, together with a discussion of the motivation for each method. Results are reported for their application to several image data sets, for both variables z1 and z2.

The above discussion has assumed a global value for λ, but it is possible to define locally optimal values. Parameters λ which are allowed to vary tend to smooth more in regions of high variation and less in regions of slowly varying intensity. Such local smoothing parameters will not be considered further here, but see Eubank (1988, Chapter 4).

3.4 Kernel smoothing of the data

Recall that the image is to be smoothed using (3.4) for the linear estimator $\mu_\lambda(x,y)$, where individual weights are normalised to give an exact weighted average of the observations z(i,j). The kernel chosen is a two-dimensional Gaussian function, in which case the estimator is of the form (3.8). It is assumed that a suitable value for the parameter λ has been determined using one of the methods to be described in Chapter 4.

As is observed in Section 3.3.3, the theoretically infinite support of the Gaussian distribution must be truncated in practice, in this case at $\pm 5\lambda$. In terms of the summation over observations on the image lattice, this means that for each position (x,y) at which an estimate is required the only points of interest in the calculation and given a non-zero kernel weight are those (i,j) satisfying:

$$|x - i| \leqslant 5\lambda \qquad \text{and} \qquad |y - j| \leqslant 5\lambda. \tag{3.9}$$

All other observations on the lattice will be weighted zero.

In practice, for data observed only at discrete grid points the theoretical limit 5λ is approximated by τ, where τ is the maximum integer less than 5λ. For the special case (holding in Chapter 4) where estimates are required only at integer lattice points, then the set of points contributing to a single estimate will fall within a region centred on (x,y) and of size b x b, where $b = 2\tau+1$. This b x b region of non-zero weights is termed the 'smoothing window', discussed further in Chapter 4.

For each (x,y) at which intensity is to be estimated, the data points $\{(i,j) : i,j = 1,2,\ldots,m\}$ are considered in turn. The set of (i,j) satisfying (3.9) for the given (x,y) are determined and for each (i,j) the individual contribution to the estimate at that point is evaluated as:

$$\hat{f}_\lambda(x-i,y-j) = \frac{z(i,j)\ \exp\left[\frac{-1}{2\lambda^2}\ \{(x-i)^2+(y-j)^2\}\right]}{\sum_{s=1}^{m}\sum_{t=1}^{m} \exp\left[\frac{-1}{2\lambda^2}\ \{(x-s)^2+(y-t)^2\}\right]}.$$

This value is added to the cumulative total of contributions from previous data points (i,j), so the intensity estimate at (x,y) can be represented as:

$$\hat{\mu}_\lambda(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{f}_\lambda(x-i,y-j).$$

Similar calculations are repeated at all other positions (x,y) at which a smoothed estimate is required. For this application, typical sets of {(x,y)} are points on the edges between different tissues as will be described in Section 3.5.

3.5 The identification of edge points

Edge points are identified as positions of zero-crossings of the second derivative of the smoothed image. As outlined in Section 3.1.3 the isotropic Laplacian is the second derivative operator used in this study. The derivatives of the estimated function μ, required for the Laplacian can also be estimated by the kernel method. Kernel derivative estimators are defined by differentiating the weight function K with respect to the 'location' explanatory variable, hence the necessary condition of smoothness of the kernel function to ensure continuous second derivatives. The kernel estimate of the (general) νth derivative of μ is a local average of the observations z of the form (3.2), where the νth derivative of the kernel is used as a weighting function.

With multivariate design points, partial derivatives must be calculated with respect to each of the explanatory variables. For the bivariate case, the Laplacian estimate is the sum of the two second partial derivatives of the smoothed estimate, that is differentiating $\hat{\mu}_\lambda(x,y)$ with respect to x and y:

$$\nabla^2\{\hat{\mu}_\lambda(x,y)\} = \left[\frac{\delta^2}{\delta x^2}\{\hat{\mu}_\lambda(x,y)\} + \frac{\delta^2}{\delta y^2}\{\hat{\mu}_\lambda(x,y)\}\right]. \qquad (3.10)$$

The second partial derivative (with respect to x) of the Gaussian kernel has the algebraic form:

$$\frac{\delta^2}{\delta x^2}\{K(x-i,y-j)\} = \frac{1}{2\pi\lambda^2}\left[\frac{-1}{\lambda^2}\right] \exp\left[\frac{-1}{2\lambda^2}\{(x-i)^2+(y-j)^2\}\right] \left\{1-\frac{(x-i)^2}{\lambda^2}\right\}$$

and similarly for the y partial derivative.

So far, smoothing and differentiation have been viewed as two distinct stages: the whole image must be smoothed prior to calculation of derivatives. In practice, the two stages are combined by using the Laplacian of Gaussian, denoted by LoG or $\nabla^2 G$. The LoG combines the two directional derivatives as in (3.10), calculating partial derivatives of the Gaussian smoothed estimate (3.8). A plot of the two-dimensional LoG function is given in Figure 3.3b.

The use of the LoG is supported by current understanding of vision (Hildreth, 1983), but it also has several other desirable properties. It is isotropic and so responds equally well to edge features in any orientation. The LoG combines the Laplacian operator and Gaussian kernel and so retains properties attributable to these two components. It is a linear operator involving the Gaussian and so has the 'nice scaling behaviour' of Yuille and Poggio (1986): see Section 3.3.3. Further, due to the Laplacian, sets of edge points identified as edges always result in closed curves.

As outlined in Section 3.3.2, when calculating the Gaussian smoothed estimate, the number of computations can be reduced and a similar reduction is possible for the LoG (Wiejak et al, 1985). The bivariate function can be separated (factorised) and edges identified using two orthogonal univariate kernels. The partial derivative with respect to x can be written as the product of a (univariate) second derivative $\hat{\mu}_\lambda''(x)$ and a (univariate) smoothed estimate $\hat{\mu}_\lambda(y)$. In its most general form the Laplacian is written as the sum of two univariate products:

$$\nabla^2\{\hat{\mu}_\lambda(x,y)\} = \hat{\mu}_\lambda''(x)\hat{\mu}_\lambda(y) + \hat{\mu}_\lambda(x)\hat{\mu}_\lambda''(y).$$

For the particular case of the Laplacian of Gaussian estimates are calculated using a univariate Gaussian kernel. The order of smoothing, that is first over columns x and then rows y, or vice versa is unimportant. Similarly, the order of smoothing or differentiating is unimportant. In this study, the LoG is calculated as the sum of two univariate products, so reducing the number of

computations required for such large data sets. Further improvements in computational efficiency have been discussed in the computer vision literature (see Sotak and Boyer (1989), for example) but since efficiency is not a prime concern here, these are not implemented .

Edge points are defined as locations of zero-crossings of the LoG smoothed image. However, preliminary searches for zero-crossings identified some spurious 'edge' points within regions. Closer inspection of the values of the LoG function in the neighbourhood of such points indicates that this is due to a slight discontinuity in the second derivative function, corresponding to a small 'step-edge' in smoothed intensity between two adjacent locations near the limit of support of a smoothing kernel. The discontinuity is the result of truncating a kernel with theoretically infinite support; extending the truncated support (for example to $\pm 10\lambda$) does not eliminate the problem.

At those locations in the image where there is a spurious 'step-edge' in the smoothed intensity, matched by an additional zero-crossing of the LoG function, there is also an equivalent maximum of the first derivative function. The magnitude of the first derivative is very small at such spurious edge points and this can be used as an extra condition to exclude these additional points. Thus the set of edge points can be summarised as those locations where the LoG function has a zero-crossing, and further that the magnitude of the first derivative at these locations is less than γ, for suitably small γ; $\gamma = 0.01$ is used here.

The set of edge locations can be identified in one of two ways. A sequential grid search (or raster-scan) is used to identify positions of zero-crossings to the nearest lattice point (i,j). The LoG function in the region of (i,j) is then optimised to obtain (x,y), the coordinates of the zero-crossing in continuous space. In practice, the grid is searched twice, separately in both horizontal and vertical directions. Since it can easily be implemented on a computer, this is the method which was used initially to identify the sets of edge points, plotted in Figure 3.5. It is still useful when a general impression of all the edge points is sufficient, for example

when choosing the smoothing parameter subjectively and comparing sets of edge points identified from images smoothed with several different values of λ.

For this first method, the order in which individual points are identified and stored is determined solely by the sequence in which the points were accessed, that is the direction of the raster scan. The form of data required when fitting a closed curve to describe the edge (Chapter 5) is an ordered sequence of points lying on a single edge. A major disadvantage of the first method is that it is not immediately possible to extract a set of points in this form. Edge points required in an ordered sequence must therefore be identified using an alternative method, now described.

A sequential grid search is used to identify a single zero-crossing as a suitable 'start-point' for one edge. Subsequent zero-crossings on this single edge are tracked, individually optimising positions to obtain an ordered set of boundary coordinates (x,y) in continuous space. Once a single edge has been closed by tracking back to the 'start-point', the sequential grid search resumes to identify the 'start-point' of the next edge. All other edges in the image are tracked similarly, to obtain separate sets of ordered coordinates for individual edges. The edges so identified are virtually identical to those identified by the first method.

Sets of edge points identified by the first method are plotted in Figure 3.5, for both variables z1 and z2. Sets of edge points are compared for subjectively chosen values of λ = 1.0, 2.0 and 3.0; see also the earlier Figure 3.2b for a further plot of points for z1 when λ = 1.5. Note that even with the same amount of smoothing edges do not coincide for the two variables, although they originate from the same physical edge. It is easy to see that as λ increases there is a reduction in the number of edges and amount of noise, but this improvement must be balanced against increased bias in location of the edge points.

Multiple smoothing of the image is one solution to the problem of choosing a single value of λ to identify the 'best' set of edges but without greatly increasing bias in edge position. The image is

Figure 3.5a: Edge points from zero-crossings of LoG
(Subject et1, Variable z1)

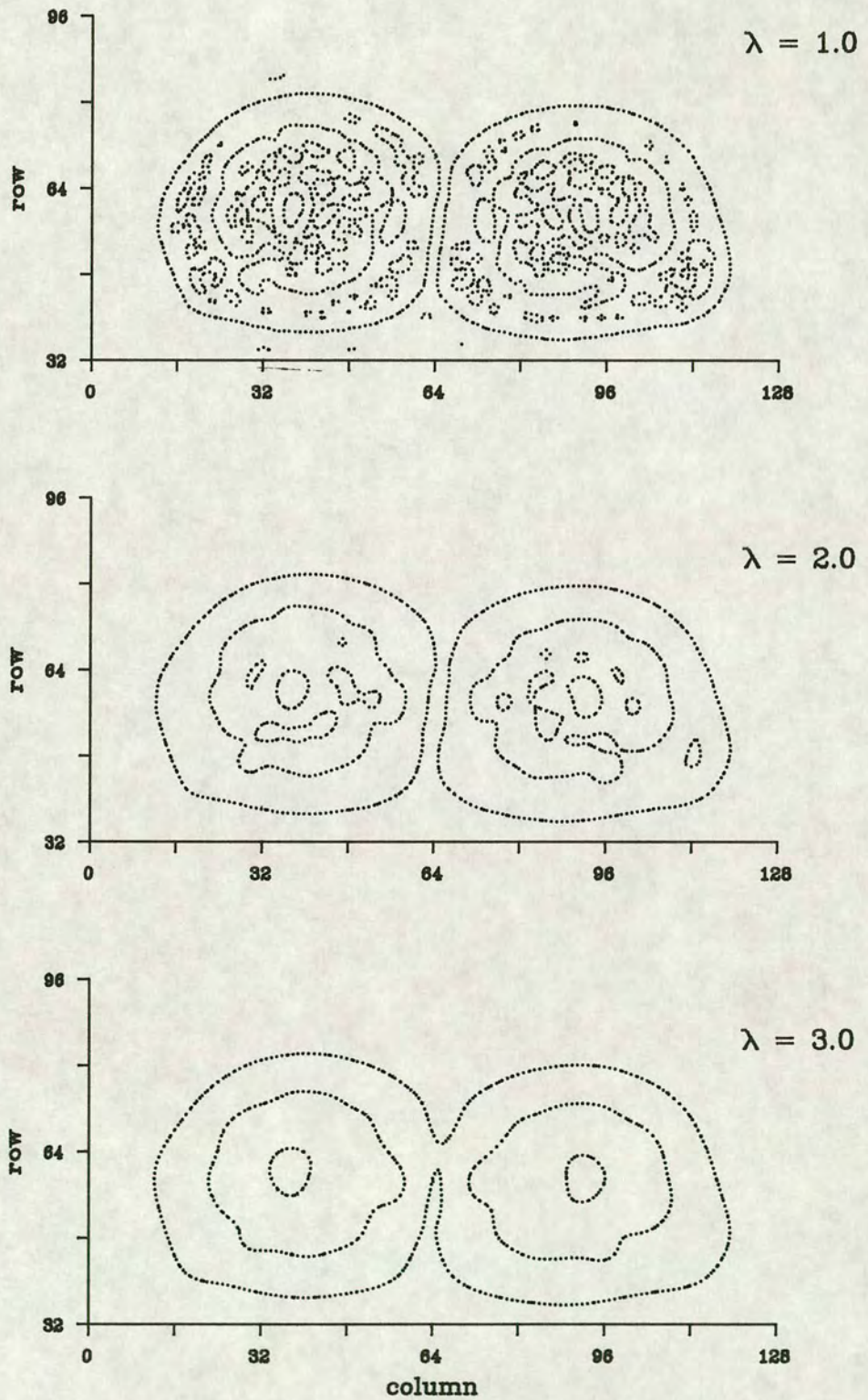
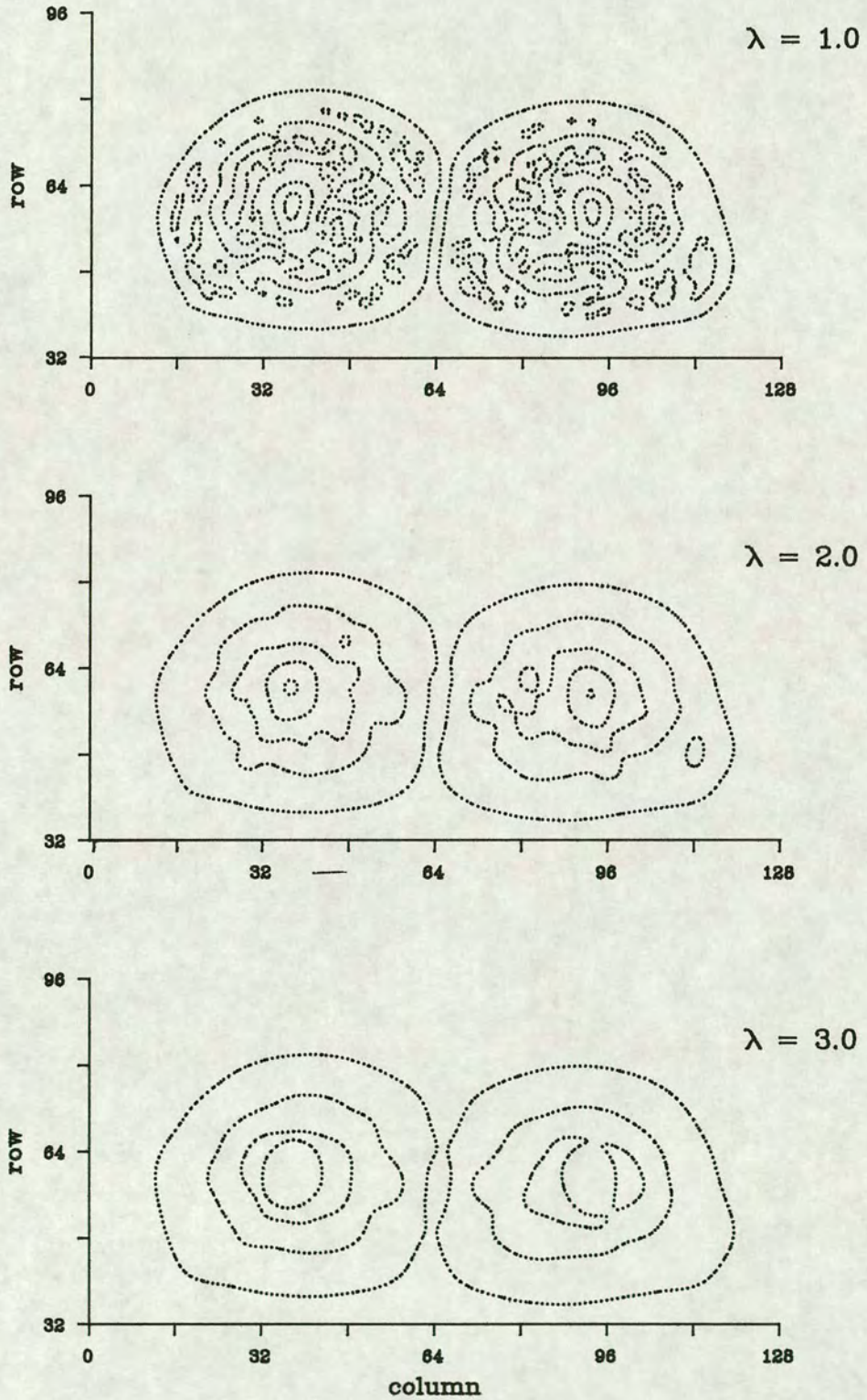


Figure 3.5b: Edge points from zero-crossings of LoG
(Subject et1, Variable z2)



smoothed initially with a 'large' value of λ to identify the main edges of interest: here these are the edges between the bone, muscle, fat and air regions. The image is then smoothed with a sequence of progressively smaller values of λ. Edge points are only retained if they have already been identified when smoothing with a larger value of λ, but are now located more accurately due to reduced smoothing of the image. Such an approach is not covered here, but it has received considerable attention in the computer vision literature: see for example Witkin (1983), Bergholm (1987).

Using the second method to identify edge points, it is possible to extract separate sets of coordinates for individual edges. Such an ordered sequence of points is the form of data required for the second stage of this project, that is the fitting of some form of closed curve as a model for a single edge. Methods of curve fitting are introduced in Chapter 5. Each edge is modelled separately. The ordered set of coordinates for a single edge is denoted by $\{(x_i,y_i) : i = 1,2,....,n\}$, where (x_i,y_i) denotes the location of the point in continuous two-dimensional space, and the subscript i denotes the relative position in the ordered sequence of points for a single edge. Note the slightly different use of notation (for variables x, y and i) to that used in Chapters 3 and 4. Since the two stages are considered separately (a set of points must be identified from the first stage before the second can begin), this should not cause confusion.

4 METHODS FOR SMOOTHING PARAMETER ESTIMATION

4.1 Introduction

Chapter 3 covered in detail the use of nonparametric kernel regression for fitting a continuous function μ_λ as a model for the underlying image. The observed discrete data is smoothed to obtain an estimate of intensity in continuous space, prior to differentiation and the identification of edge points. The amount of smoothing is determined by the parameter λ, a suitable value for which may be estimated by one of several methods. The simplest and quickest alternative is to choose λ subjectively; the data is smoothed using different values of λ and the 'best' is assessed visually.

Any definition of 'best' must be related to how the smoothed estimate is to be interpreted. A slightly oversmoothed image may be desirable if the purpose of smoothing is to suggest a possible parametric model, or to improve the visual appearence by increasing the signal-to-noise ratio. By comparison, if it is more important to accurately estimate local detail of the regression curve, then a slightly undersmoothed curve may be more appropriate. When identifying edges in MRI it is usually more important to optimise the set of edges obtained, that is, those between the fat, muscle and bone tissues, rather than imposing very strict location accuracy. In this context a slightly oversmoothed image may be preferred. Thus when identifying edges as in Section 3.5, visual comparisons could be made between several sets of edge points, after smoothing with increasing values of λ. In this case the subjective choice of λ would be the value which results in the most 'representative' set of edges for the tissue regions of interest.

Usually however, a more 'automatic' method of estimating λ is preferred. Eubank (1988, Chapter 4) and Müller (1988, Chapter 7) mentioned briefly and gave references for many of the alternative data-based methods which have been proposed for estimating λ. Such methods are usually defined in terms of univariate data; the extension to bivariate data is considered in this chapter. The background and motivation for four methods are described in more

detail in Section 4.2. These are the chi-squared method (Section 4.2.2); the empirical degrees of freedom method (Section 4.2.3); the Rice method (Section 4.2.4) and the cross-validation method (Section 4.2.5). In Sections 4.3.2 - 4.3.5 corresponding definitions are given for fitting the nonparametric kernel regression model (3.4) to bivariate data.

Smoothing parameters must also be estimated for derivatives of a kernel function. The basic problem is whether the same value of λ should be used for the derivative as was appropriate for the function itself, and if not, how a value for the derivative kernel should be estimated. The extension of existing methods to include derivative functions is covered in Section 4.4, and two variants of the cross-validation method are proposed.

Section 4.5 covers estimation of the error variance σ^2, a value for which is required for the X^2, EDF and Rice methods. Numerical results for the different methods applied to several images for both variables z1 and z2 are given in Section 4.6. Cases where any of the methods perform poorly will be highlighted. In Section 4.7 results are interpreted taking account of the nature of the MRI data and recommendations made as to the most appropriate method for smoothing parameter estimation given this form of data.

A final point for smoothing parameter estimation is whether a local or global value is most applicable. All the methods discussed are based on global measures of lack-of-fit, aiming for the best fit, on average, of the smoothed function to the underlying continuous image. The alternative is local or variable smoothing parameters, which may increase the estimation accuracy at one particular point since they smooth more in regions where the intensity is varying rapidly and less where values are more constant. Local parameters will not be covered here, see Eubank (1988, Chapter 4) or Härdle (1990, Chapter 5) who outlined two methods for locally adapting the amount of smoothing.

4.2 Methods considered for smoothing parameter estimation

4.2.1 Introduction

The intensity data and the fitted model are as defined in Sections 3.2.1 and 3.2.3 respectively, but the notation is modified slightly. A point at which an estimate is required is denoted by (x,y), but for the special case of methods for smoothing parameter estimation covered in this chapter, estimates are only evaluated at lattice points, so x and y will be restricted to be integer valued, $\{(x,y) : x,y = 1,2,\ldots,m\}$. As previously, (i,j) is used to denote the position of an observation contributing to the estimate of intensity at (x,y).

For the theoretical criteria and methods to be introduced in the following sections, summation will be over all positions at which the intensity is estimated where x and y are necessarily integer-valued. Therefore for this chapter only, the observed data z will be indexed interchangeably by (i,j) or (x,y) depending on the context: by (i,j) if summation is over points contributing to an estimate at (x,y), or over (x,y) if summing over quantities involving this smoothed estimate. Thus the observed data $\{z(i,j) : i,j = 1,2,\ldots,m\}$ can equivalently be written as $\{z(x,y) : x,y = 1,2,\ldots,m\}$, where m is the number of columns and rows in the image, assumed equal.

Let M ($= m^2$) be the total number of observations and hence total number of degrees of freedom (DF). Thus nonparametric kernel regression is used to fit a model of the form (3.1), which using the modified indexing appropriate for this chapter is given by:

$$z(x,y) = \mu(x,y) + \epsilon(x,y), \qquad x,y = 1,2,\ldots,m, \tag{4.1}$$

where $\mu(x,y)$ is the underlying 'true' function at the point (x,y), estimated by $\hat{\mu}_\lambda(x,y)$. The random error $\epsilon(x,y)$ is assumed to have zero mean and variance σ^2.

Most of the methods proposed for smoothing parameter estimation are based on some measure of lack-of-fit between the underlying true

function μ and the smoothed estimate $\hat{\mu}_\lambda$. One such measure is the mean square error (MSE), which for bivariate data is given by:

$$MSE_\lambda = \frac{1}{M} \sum_{x=1}^{m} \sum_{y=1}^{m} \{\mu(x,y) - \hat{\mu}_\lambda(x,y)\}^2. \qquad (4.2)$$

In practice the 'true' function μ is unknown and so must be estimated, usually by the observed data $z(x,y)$. In this case the 'true' MSE lack-of-fit is replaced by its estimated value, equivalent to the residual sum of squares, RSS:

$$RSS_\lambda = \frac{1}{M} \sum_{x=1}^{m} \sum_{y=1}^{m} \{z(x,y) - \hat{\mu}_\lambda(x,y)\}^2. \qquad (4.3)$$

As before, the subscript λ is used to indicate the dependence on λ of the RSS. All four methods studied here can be written as functions of the RSS. Differences between methods arise depending on how the theoretical lack-of-fit is modified when estimating the unknown μ.

The different methods for parameter selection are generally defined in the context of a specific smoothing application and for univariate data. In the remainder of this section the four methods of interest are introduced in terms of the original estimation problem. Modified definitions for kernel estimates and bivariate data are given in Section 4.3.

4.2.2 The chi-squared method

The value of λ estimated from the chi-squared (χ^2) method is denoted by $\lambda(CHI)$. First considered by Phillips (1962), it is based on the assumption that differences between the observed data $z(x,y)$ and the smoothed estimate $\hat{\mu}_\lambda(x,y)$, as measured by the RSS, should be equal to errors in the input data, as given by σ^2. Assuming errors are Normally distributed with mean zero and variance σ^2, then by standard properties of χ^2 random variables the 'true' residual sum of squares, RSS_T say, when scaled by the error variance σ^2, has a χ^2 distribution on M degrees of freedom, written $\chi^2(M)$, that is,

$RSS_T / \sigma^2 \sim \chi^2(M)$. Replacing the true residual sum of squares by RSS_λ, its observed value for a given value of λ, the χ^2 choice $\lambda(CHI)$ is the solution for λ of $RSS_\lambda / M = \sigma^2$, where χ^2 is used to denote the value of the chi-squared score function ($= RSS_\lambda/M$) evaluated at a given value of λ. An explicit solution is not possible so $\lambda(CHI)$ must be evaluated numerically.

The solution for $\lambda(CHI)$ is dependent on the error variance σ^2 which is generally unknown; possible data-based estimators are discussed in Section 4.5. To some extent, the reliability of the estimate $\hat{\sigma}^2$ will affect the estimated smoothing parameter. Thompson et al (1991a) concluded from a simulation study that for the χ^2 method the estimate $\lambda(CHI)$ was not too sensitive to the value estimated for σ^2. However, they did suggest that values of $\lambda(CHI)$ tend to be higher than other estimates for λ, and so consequently oversmooth the data relative to parameters estimated by different methods. A similar conclusion was reached by Craven and Wahba (1979).

4.2.3 The empirical degrees of freedom method

The parameter λ estimated according to the empirical degrees of freedom (EDF) method is denoted by $\lambda(EDF)$. This method is an extension of the χ^2 method described in Section 4.2.2. As the name suggests it allows for the loss in degrees of freedom of the true RSS when the true function μ is replaced by its estimate $\hat{\mu}_\lambda$. The equivalent degrees of freedom remaining for error is denoted by EDF_λ and depends on the value of λ used in the estimation of μ_λ. A definition of EDF_λ in terms of the kernel model will be given in Section 4.3.1. The EDF estimate of λ is the solution $\lambda = \lambda(EDF)$ of $RSS_\lambda / EDF_\lambda = \sigma^2$. As for the chi-squared method, a solution for $\lambda = \lambda(EDF)$ must be evaluated numerically.

Thompson et al (1991a) gave a theoretical comparison of $\lambda(CHI)$ and $\lambda(EDF)$. Since the EDF method was motivated by allowing for the loss in degrees of freedom of the estimated RSS, they expected that $\lambda(EDF)$ would be smaller than $\lambda(CHI)$; that this held in practice was confirmed in a simulation study. The error variance σ^2 must also be

estimated; from their simulations Thompson et al (1991a) concluded that λ(EDF) is quite sensitive to the value used for σ^2, though data-based variance estimators work well.

4.2.4 The Rice method

The parameter λ estimated by this method will be denoted by λ(RICE). Rice (1984) proposed the method in the context of nonparametric kernel regression, though in practice he simplified the analysis by replacing the kernel estimate by a tapered Fourier series and his results were derived based on this modification. Note that compared to the currently defined kernel estimate, Rice defined his smoothing parameter in such a way that it is the reciprocal of the parameter in (3.5).

The mean square error (4.2) is stochastic and therefore not readily mathematically tractable. Rice (1984) considered the associated expected mean square error (EMSE), the bivariate form of which is given by:

$$\mathrm{EMSE}_\lambda = \frac{1}{M} E\left[\sum_{x=1}^{m}\sum_{y=1}^{m} \{\mu(x,y) - \hat{\mu}_\lambda(x,y)\}^2\right]. \qquad (4.4)$$

The EMSE, also termed the 'risk', is asymptotically equivalent to the IMSE, the theoretical value which would be obtained by integrating the MSE over all data points.

It is important, however, to note a slight difference in interpretaion of the MSE and EMSE criteria. Minimisation of the MSE criterion means that the estimated μ_λ is made as close as possible to the true function μ, for a particular data set. In comparison, the optimal parameter for the EMSE criterion aims to minimise the distance averaged over all data sets. Note also the difference between the terminology of this chapter and that used by Eubank (1988, Chapter 2): the MSE and EMSE here are equivalent to the Eubank 'loss' and 'risk' functions respectively.

In theory, the minimiser of the EMSE is taken as an estimate of λ, that is the value of λ so that, on average, the mean of the observed data will be most closely predicted. However the EMSE involves the unknown function μ. In practice, the EMSE is estimated from the data, which Rice achieved by replacing $\mu(x,y)$ with the observed value $z(x,y)$. An elementary calculation showed that an unbiased estimator of the EMSE can be written as a weighted function of the residual sum of squares (4.3) and the error variance, σ^2. This is termed the 'Rice' score; $RICE_\lambda$ will be used to denote the value of this score with a given value of λ. Further details and an expression in terms of the kernel model are given in Section 4.3.4. The value of λ minimising $RICE_\lambda$ is taken as the estimate of the parameter λ. Further, since the $RICE_\lambda$ is a data-based estimate of the EMSE, then λ(RICE) should be a good estimate of the value of λ which would minimise the theoretical EMSE criterion.

The Rice score is a function of the unknown σ^2, which must be estimated. Rice himself suggested two alternatives, discussed in Section 4.5.1. The parameter λ(RICE) estimated by the Rice method has been shown to be quite sensitive to the value estimated for σ^2. Underestimation of σ^2 results in a parameter which is too small and hence there is insufficient smoothing of the data. On the basis of a simulation study, Müller et al (1987) illustrated this tendency to undersmooth when the Rice estimate for λ was calculated from an inappropriate value of σ^2. Consequently there is less chance of oversmoothing the data if the Rice method is used to estimate λ.

Although the value for λ(RICE) selected by this method is asymptotically equivalent to several other data-based estimates of λ, estimates may vary for finite samples as Rice showed with a small simulation. The results of this simulation together with that of Härdle, Hall and Marron (1988) were summarised by Härdle (1990). He suggested that the apparent superiority of the Rice method was due to the design of the first simulation; the results of the second simulation with a different design indicated the Rice method was no longer the best.

4.2.5 The cross-validation method

The parameter λ estimated by cross-validation (CV) is denoted by $\lambda(CV)$. The CV method was proposed for kernel regression by Clark (1977), while Wahba and Wold (1975) considered a similar technique for smoothing splines. It is motivated by prediction: select a value for λ so that, on average, the mean of the data would have been most closely predicted. As for the Rice method, the aim is to estimate that λ which minimises the MSE lack-of-fit. The two methods are asymptotically equivalent, except for a constant σ^2 and the relationship between the two methods will be shown in Section 4.3.5. However cross-validation does not require an estimate of the error variance and so represents a totally data-based approach to smoothing parameter estimation.

Independence of σ^2 is achieved by replacing $\hat{\mu}_\lambda$ by its 'leave-one-out' cross-validation estimate $\hat{\mu}_\lambda^c$. To calculate $\hat{\mu}_\lambda^c$, at each data point in turn the observation is omitted and the remainder of the data used to predict the value of the function at that position. Thus the cross validation estimate for λ is that value $\lambda = \lambda(CV)$ which minimises the difference between the observed and the predicted value, as given by $\hat{\mu}_\lambda^c$, averaged over the data points. This average difference is the cross-validation score; when calculated with respect to a specific value of λ the score is denoted by CV_λ and given by:

$$CV_\lambda = \frac{1}{M} \sum_{x=1}^{m} \sum_{y=1}^{m} \{z(x,y) - \hat{\mu}_\lambda^c(x,y)\}^2. \qquad (4.5)$$

This score was used by Craven and Wahba (1979) for estimating a smoothing parameter for polynomial splines but their cross-validation score was defined in a slightly different form. By writing the 'leave-one-out' cross-validation estimate $\hat{\mu}_\lambda^c$ as a weighted sum of $\hat{\mu}_\lambda$, the estimate based on all the data, and $z(x,y)$, the omitted observation, the cross-validation score can be re-expressed as a weighted version of the RSS (4.3). This alternative form is computationally more convenient, since it does not require the calculation of explicit 'leave-one-out' cross-validation estimates.

The algebraic simplification will be given for the kernel estimator (3.4) in Section 4.3.5.

The weights will generally be different at each contributing point (x,y) but Craven and Wahba (1979) showed that a further computational reduction was possible if individual weights were replaced by their average value. Justification for the resulting generalised cross-validation (GCV) score as a method for smoothing parameter selection was given formally by Craven and Wahba in the GCV theorem. The name generalised cross-validation was considered by Eubank (1988, Chapter 2) to be slightly misleading, since CV is not a special case of GCV but rather the two methods correspond to alternative weighting of the RSS. The two scores are often closely related: if the data is recorded only at integer lattice points, then for the special case where intensity is to be estimated at this same set of points, CV and GCV are equivalent. This property will be illustrated for the kernel model in Section 4.3.5, so for the current data, only the cross-validation method need be considered.

Silverman (1984) proposed asymptotic generalised cross-validation (AGCV), with the aim of improving computational efficiency. However, for equally-spaced data, the parameters estimated from AGCV were shown to be virtually identical to those estimated by GCV. Therefore since it is the accuracy of the parameters estimated which is of interest here, rather than the efficiency of computation, this method is not considered.

Wahba (1983) proposed using the CV method as the basis of an estimator for σ^2. The estimator incorporated the definition of EDF (Section 4.2.3) into the CV score. If this estimator of σ^2 is used in the EDF method, values for λ(EDF) and λ(CV) were expected to be similar (Thompson et al, 1991a).

Cross-validation has been used in many applications as a completely data-based method of estimating an 'optimum' smoothing parameter. Härdle and Marron (1985a, 1985b) proved that choosing λ to minimise the CV function was an asymptotically optimal method. This property was unaffected by the smoothness of the underlying function.

However although it does work well in most situations, cases do occur where the cross-validation estimate of λ is unsatisfactory. This was the conclusion of Thompson et al (1989), who from a simulation study for image restoration illustrated typical cases where the λ(CV) estimate was unsatisfactory. For example, for values of $\lambda > 0$ there was no minimum of the CV score function (CV_λ), or the value of λ minimising CV_λ resulted in undersmoothing of the data, or multiple minima were observed for the CV_λ function. So, although on the whole cross-validation provides a 'reasonable' estimate of the 'optimal' λ, care should be taken before automatically adopting the minimising λ.

4.3 Estimation of the smoothing parameter

4.3.1 Introduction and notation

Many of the methods are motivated in the context of a specific smoothing problem. For example, Craven and Wahba (1979) used cross-validation for smoothing splines, while Thompson et al (1991a) compared the χ^2, EDF and CV methods for the restoration of one and two-dimensional images. It is worth comparing the role of the smoothing parameter for image restoration and the current estimation. Image restoration is an 'inverse' problem, since the observed data is often a blurred image, and the aim is to 'unsmooth' the data to recover the underlying 'true' image. In this thesis, kernel regression is used to fit a smooth function to the observed data, to smooth out random variation, with the primary aim of identifying edges. There is a corresponding difference in the interpretation of the smoothing parameter selected. For image restoration, the problem is to measure the level of smoothing that has already been applied to the observed image. In comparison, for kernel regression the aim is to determine how much smoothing should be applied to allow the best representation of the edges of interest.

Titterington (1985) summarised and established a common structure between a number of smoothing techniques and methods for parameter estimation; see also Hall and Titterington (1987). Several of the more common methods for estimating λ have been shown to be

asymptotically equivalent to minimising the MSE criterion (Rice, 1984). Using a kernel regression model Härdle (1990, Chapter 5) compared different methods and considered theoretically which should be used. For any given data set one method may be superior, since an estimate of λ calculated from observed data need not necessarily achieve its data-based optimum. Härdle (1990) summarised the results from two simulation studies (Rice, 1984 and Härdle and Marron, 1988) which indicated that there were real differences between estimates when compared for small samples and concluded that overall, the CV method gave the best results.

Individual methods are now defined in terms of the kernel regression model (3.4) of the preceeding chapter. As outlined in Section 4.2.1, the notation is modified slightly for this chapter, in that x and y are restricted to integer lattice points, that is, $\{(x,y) : x,y = 1,2,\ldots,m\}$ and hence $\{x,y\}$ can be used interchangeably with $\{i,j\}$ to index grid points, that is positions at which intensity is observed.

The estimator μ_λ (3.2) is a linear function of the weighted observations. The calculation of estimates at positions (x,y) 'close' to the boundary of the m x m image is not made explicit in (3.2). Here a more general expression is given for the estimator μ_λ, reflecting how estimates at such positions are calculated in practice. The infinite range of summation allows smoothing windows to extend beyond the m x m lattice, assuming there has been 'wrap-round' of observed intensities at the boundary, as described in Section 2.2.1. If $w_\lambda(x-i,y-j)$ is the weight given to the observation at position (i,j), contributing to the estimate required at position (x,y), then the estimate at any point (x,y) on the m x m lattice can be written as:

$$\hat{\mu}_\lambda(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} w_\lambda(x-i,y-j)\, z(i,j), \qquad (4.6)$$

where $|x-i| \leqslant k\lambda$ and $|y-j| \leqslant k\lambda$ for any $k \in \mathbf{Z}$, and any $x,y \in \{1,2,\ldots,m\}$.

The additional conditions on i and j are a consequence of

truncating the support of the kernel function, so that only a subset of the m^2 lattice points (i,j) contribute to the estimate at (x,y). For data observed at discrete grid points, the theoretical $k\lambda$ limit is approximated by τ, where τ is the maximum integer less than $k\lambda$. Thus when intensity is to be estimated only at integer lattice points, the smoothing window (and matrix W_λ) of non-zero weights will be of size b x b, where $b = 2\tau+1$.

The estimated intensity at any point (x,y) on the m x m lattice is given by (4.6). However, most of the methods for smoothing parameter selection are defined for univariate data, so in order to derive algebraically the corresponding expression for bivariate data it is more convenient to rewrite the set of smoothed estimates in terms of a linear model. For the restoration of two-dimensional images, Thompson et al (1991a) used a model of this form when comparing methods of estimating λ. Their expressions were used as the basis of the scores to be defined in subsequent sections, with appropriate modifications to the notation.

Observations $z(i,j)$ are recorded on an m x m grid but can be stacked as in a raster scan (see Gonzales and Wintz, 1987 or Kay, 1988) to give an M x 1 vector $\underline{z}$ as:

$$\underline{z} = [z(1,1), z(2,1), \ldots, z(m,1), z(1,2), \ldots, z(1,m), \ldots, z(m,m)]'.$$

A typical element of the stacked vector $\underline{z}$ is denoted by $\{z(q) : q = 1,2,\ldots,M\}$, where $M = m^2$. The index q denoting position is calculated from $\{(i,j) : i,j = 1,2,\ldots,m\}$, the row and column indices of this element with respect to the m x m lattice, as $q = m(j-1) + i$.

Estimates $\hat{\mu}_\lambda(x,y)$ and random errors $\epsilon(x,y)$ are similarly stacked for M x 1 vectors $\underline{\hat{\mu}}_\lambda$ and $\underline{\epsilon}$ respectively. Typical elements of $\underline{\hat{\mu}}_\lambda$ are denoted by $\{\hat{\mu}_\lambda(p) : p = 1,2,\ldots,M\}$, where the index p is calculated from lattice indices $\{(x,y) : x,y = 1,2,\ldots,m\}$ as $p = m(y-1) + x$. Estimates $\underline{\hat{\mu}}_\lambda$ are calculated as weighted, linear combinations of the observations $\underline{z}$, so (4.6) may be rewritten as $\underline{\hat{\mu}}_\lambda = A_\lambda \underline{z}$, where A_λ is an M x M, that is m^2 x m^2, matrix of weights. The matrix A_λ is of

block-Toeplitz form, with m blocks in each column and in each row. Further, each individual block of size m x m is also of Toeplitz form. Entries $\{a_\lambda(p,q) : p,q = 1,2,\ldots,M\}$ in the weight matrix A_λ are given by $a_\lambda(p,q) = w_\lambda(x-i,y-j)$, where x,y,i,j are defined such that:

$$x = \begin{cases} m & \text{if } \mathrm{rem}(p,m) = 0, \\ \mathrm{rem}(p,m) & \text{otherwise;} \end{cases}$$

$$y = \begin{cases} (p/m) & \text{if } p=\alpha\, m \text{ for any } \alpha \in \mathbb{Z}, \\ \mathrm{int}(p/m) & \text{otherwise;} \end{cases}$$

$$i = \begin{cases} m & \text{if } \mathrm{rem}(q,m) = 0, \\ \mathrm{rem}(q,m) & \text{otherwise;} \end{cases}$$

$$j = \begin{cases} (q/m) & \text{if } q=\beta\, m \text{ for any } \beta \in \mathbb{Z}, \\ \mathrm{int}(q/m) & \text{otherwise;} \end{cases}$$

where rem(a,b) denotes the remainder of (a/b) and int(a/b) denotes the integer part of (a/b). With this notation an individual smoothed estimate is given by:

$$\hat{\mu}_\lambda(p) = \sum_{q=1}^{M} a_\lambda(p,q)z(q), \quad \text{for any } p = 1,2,\ldots,M. \tag{4.7}$$

Usually the matrix W_λ, representing the smoothing window will be much smaller than the size of the image, that is $b \ll m$. The matrix A_λ will have a sparse, diagonally banded structure and individual blocks within A_λ will have the same banded form. The block bandwidth of A_λ will be b, the number of non-zero blocks, that is m x m matrices, in each block row (and column) of A_λ. Similarly, the bandwidth within each block matrix will be b, the number of non-zero diagonals.

In order to establish the general structure of A_λ, it is useful to consider a detailed example of one of the most basic cases of a 3 x 3 smoothing window and 5 x 5 image, resulting in a 25 x 25 matrix of weights A_λ, so that m=5, b=3, M=25. Even for this example the derivation of the matrix A_λ is a non-trivial problem but it is included here since it is helpful for establishing patterns in the weights. The 5 x 5 image is illustrated in lattice form in Figure 4.1, with the assumed 'wrap-round' of intensities shown explicitly

Figure 4.1: 5 x 5 image of intensities z(i,j), showing 'wrap-round'

z(5,5)	z(1,5)	z(2,5)	z(3,5)	z(4,5)	z(5,5)	z(1,5)
z(5,1)	z(1,1)	z(2,1)	z(3,1)	z(4,1)	z(5,1)	z(1,1)
z(5,2)	z(1,2)	z(2,2)	z(3,2)	z(4,2)	z(5,2)	z(1,2)
z(5,3)	z(1,3)	z(2,3)	z(3,3)	z(4,3)	z(5,3)	z(1,3)
z(5,4)	z(1,4)	z(2,4)	z(3,4)	z(4,4)	z(5,4)	z(1,4)
z(5,5)	z(1,5)	z(2,5)	z(3,5)	z(4,5)	z(5,5)	z(1,5)
z(5,1)	z(1,1)	z(2,1)	z(3,1)	z(4,1)	z(5,1)	z(1,1)

for this small example. Equation (4.8) gives the smoothing window of weights W_λ, centred on the position (x,y) at which the estimate is required.

$$\begin{bmatrix} w_\lambda(1,1) & w_\lambda(0,1) & w_\lambda(-1,1) \\ w_\lambda(1,0) & w_\lambda(0,0) & w_\lambda(-1,0) \\ w_\lambda(1,-1) & w_\lambda(0,-1) & w_\lambda(-1,-1) \end{bmatrix} \qquad (4.8)$$

Assuming the observations and estimates recorded on an m x m lattice have been stacked into M x 1 vectors $\underline{z}$ and $\underline{\hat{\mu}}_\lambda$, respectively, then the matrix A_λ is given in Figure 4.2. Individual weights $w_\lambda(x-i,y-j)$ for entries $a_\lambda(p,q)$ are written in the shortened form (x-i,y-j). As discussed in Section 3.4, intensity observed at positions (i,j) contributes to estimates of intensity at positions (x,y) and this notation makes explicit the dependence of the non-zero weights on the distance between (i,j) and (x,y).

The structure of the matrix A_λ remains the same, whatever function is used to define the weights $w_\lambda(x-i,y-j)$. Expressions for the weights as functions of the Gaussian kernel will be given in (4.9). The following numerical example illustrates how individual entries of the matrix A_λ (Figure 4.2) are evaluated. A single entry of A_λ is given by $a_\lambda(p,q) = w_\lambda(x-i,y-j)$. Suppose p=14 and q=10. For the index p, rem(p,m) = 4, so x=4; $p \neq \alpha\, m$, so int(p/m) = 3, giving y=3. Analagous calculations for the index q give rem(q,m) = 0, so i=5 and $q=\beta\, m$, so (q/m) = 2 and j=2. Thus (x-i)=-1 and (y-j)=1, so the weight for the individual entry $a_\lambda(14,10)$ is $w_\lambda(-1,1)$. Other entries of the matrix are calculated similarly for {p,q = 1,2,...,M}.

The structure of the matrix A_λ is clearly visible from Figure 4.2. The most important property is that entries $a_\lambda(p,p)$ on the main diagonal are constant, equal to the weight $w_\lambda(0,0)$. This allows simplification of the expressions for estimating λ derived for the different methods, as will be shown in the following sections. Since individual weights are calculated using a kernel function, it follows that $w_\lambda(0,0)$ is the maximum weight, so each observation contributes the most to its own smoothed estimate.

Figure 4.2: M x M matrix A_λ

(0, 0)(-1, 0)(0, 0)(0, 0)(1, 0) (1, 0)(0, 0)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 0)(0, 0)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 0)(0, 0)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 0)(0, 0)	(0,-1)(-1,-1)(0, 0)(0, 0)(1,-1) (1,-1)(0,-1)(-1,-1)(0, 0)(0, 0) (0, 0)(1,-1)(0,-1)(-1,-1)(0, 0) (0, 0)(0, 0)(1,-1)(0,-1)(-1,-1) (-1,-1)(0, 0)(0, 0)(1,-1)(0,-1)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 1)(-1, 1)(0, 0)(0, 0)(1, 1) (1, 1)(0, 1)(-1, 1)(0, 0)(0, 0) (0, 0)(1, 1)(0, 1)(-1, 1)(0, 0) (0, 0)(0, 0)(1, 1)(0, 1)(-1, 1) (-1, 1)(0, 0)(0, 0)(1, 1)(0, 1)
(0, 1)(-1, 1)(0, 0)(0, 0)(1, 1) (1, 1)(0, 1)(-1, 1)(0, 0)(0, 0) (0, 0)(1, 1)(0, 1)(-1, 1)(0, 0) (0, 0)(0, 0)(1, 1)(0, 1)(-1, 1) (-1, 1)(0, 0)(0, 0)(1, 1)(0, 1)	(0, 0)(-1, 0)(0, 0)(0, 0)(1, 0) (1, 0)(0, 0)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 0)(0, 0)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 0)(0, 0)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 0)(0, 0)	(0,-1)(-1,-1)(0, 0)(0, 0)(1,-1) (1,-1)(0,-1)(-1,-1)(0, 0)(0, 0) (0, 0)(1,-1)(0,-1)(-1,-1)(0, 0) (0, 0)(0, 0)(1,-1)(0,-1)(-1,-1) (-1,-1)(0, 0)(0, 0)(1,-1)(0,-1)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)
(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 1)(-1, 1)(0, 0)(0, 0)(1, 1) (1, 1)(0, 1)(-1, 1)(0, 0)(0, 0) (0, 0)(1, 1)(0, 1)(-1, 1)(0, 0) (0, 0)(0, 0)(1, 1)(0, 1)(-1, 1) (-1, 1)(0, 0)(0, 0)(1, 1)(0, 1)	(0, 0)(-1, 0)(0, 0)(0, 0)(1, 0) (1, 0)(0, 0)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 0)(0, 0)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 0)(0, 0)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 0)(0, 0)	(0, 1)(-1,-1)(0, 0)(0, 0)(1,-1) (1,-1)(0, 1)(-1,-1)(0, 0)(0, 0) (0, 0)(1,-1)(0, 1)(-1,-1)(0, 0) (0, 0)(0, 0)(1,-1)(0, 1)(-1,-1) (-1,-1)(0, 0)(0, 0)(1,-1)(0, 1)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)
(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 1)(-1, 0)(0, 0)(0, 0)(1, 1) (1, 1)(0, 1)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 1)(0, 1)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 1)(0, 1)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 1)(0, 1)	(0, 0)(-1, 0)(0, 0)(0, 0)(1, 0) (1, 0)(0, 0)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 0)(0, 0)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 0)(0, 0)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 0)(0, 0)	(0,-1)(-1,-1)(0, 0)(0, 0)(1,-1) (1,-1)(0,-1)(-1,-1)(0, 0)(0, 0) (0, 0)(1,-1)(0,-1)(-1,-1)(0, 0) (0, 0)(0, 0)(1,-1)(0,-1)(-1,-1) (-1,-1)(0, 0)(0, 0)(1,-1)(0,-1)
(0,-1)(-1,-1)(0, 0)(0, 0)(1,-1) (1,-1)(0,-1)(-1,-1)(0, 0)(0, 0) (0, 0)(1,-1)(0,-1)(-1,-1)(0, 0) (0, 0)(0, 0)(1,-1)(0,-1)(-1,-1) (-1,-1)(0, 0)(0, 0)(1,-1)(0,-1)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0) (0, 0)(0, 0)(0, 0)(0, 0)(0, 0)	(0, 1)(-1, 1)(0, 0)(0, 0)(1, 1) (1, 1)(0, 1)(-1, 1)(0, 0)(0, 0) (0, 0)(1, 1)(0, 1)(-1, 1)(0, 0) (0, 0)(0, 0)(1, 1)(0, 1)(-1, 1) (-1, 1)(0, 0)(0, 0)(1, 1)(0, 1)	(0, 0)(-1, 0)(0, 0)(0, 0)(1, 0) (1, 0)(0, 0)(-1, 0)(0, 0)(0, 0) (0, 0)(1, 0)(0, 0)(-1, 0)(0, 0) (0, 0)(0, 0)(1, 0)(0, 0)(-1, 0) (-1, 0)(0, 0)(0, 0)(1, 0)(0, 0)

For the current image m=128, with b, the size of the smoothing window, determined by the $\pm 5\lambda$ truncation limit of the support of the Gaussian kernel, as discussed in Section 3.4. Although the block bandwidth, and bandwidth within each block, will generally be wider than for the example shown in Figure 4.2 and there will be many more zero entries, the matrix A_λ will have the same structure. In particular, there will be a constant value $a_\lambda(p,p) = w_\lambda(0,0)$ on the main diagonal. To simplify subsequent expressions a constant weight is defined so that $a_\lambda(p,p) = \tilde{a}_\lambda$, for all $p = 1,2,\ldots,M$. Since from Figure 4.2 it can be seen that this constant value is equal to the weight $w_\lambda(0,0)$ then $\tilde{a}_\lambda = w_\lambda(0,0)$, and these two expressions may be used interchangeably.

The exact form of weights depends on the method used to estimate μ. For a general bivariate kernel K (where no assumptions have been made on the separability of the function) and estimator (3.4), individual weights are given by:

$$w_\lambda(x-i,y-j) = \frac{K\left[\frac{x-i}{\lambda}, \frac{y-j}{\lambda}\right]}{\sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} K\left[\frac{x-s}{\lambda}, \frac{y-t}{\lambda}\right]},$$

where $|x-i| \leqslant k\lambda$, $|y-j| \leqslant k\lambda$, $|x-s| \leqslant k\lambda$ and $|y-t| \leqslant k\lambda$ for any $k \in \mathbb{Z}$, and any $x,y \in \{1,2,\ldots,m\}$. Note that the summation of weights over an infinite range means that the estimator μ_λ is made more general than that given in Section 3.3.1, now taking into account the 'wrapping-round' of the image at the boundaries (Section 2.2.2). As before individual weights are normalised so that:

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} w_\lambda(x-i,y-j) = 1,$$

where $|x-i| \leqslant k\lambda$ and $|y-j| \leqslant k\lambda$ for $k \in \mathbb{Z}$ and for any $x,y \in \{1,2,\ldots,m\}$.

The extension to general multivariate kernel functions is straightforward. A bivariate kernel involving a separable function is given by (3.5), but here common values of K and λ are used for the

two univariate kernels. For the example of a bivariate Gaussian kernel (Section 3.3.3), with support truncated at $\pm 5\lambda$, then weights are:

$$w_\lambda(x-i,y-j) = \frac{\exp\left[-\frac{1}{2\lambda^2}\{(x-i)^2 + (y-j)^2\}\right]}{\sum_{s=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\exp\left[-\frac{1}{2\lambda^2}\{(x-s)^2 + (y-t)^2\}\right]}, \tag{4.9}$$

where $|x-i| \leqslant 5\lambda$, $|y-j| \leqslant 5\lambda$, $|x-s| \leqslant 5\lambda$ and $|y-t| \leqslant 5\lambda$, for any $x,y \in \{1,2,\ldots,m\}$.

The weight $w_\lambda(0,0)$, on the main diagonal of the matrix A_λ, is of particular importance for a given (x,y). It is the contribution of each observation to its own smoothed estimate, that is, when $i=x$ and $j=y$. For the Gaussian kernel:

$$w_\lambda(0,0) = \left[\sum_{s=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\exp\left[-\frac{1}{2\lambda^2}\{(x-s)^2 + (y-t)^2\}\right]\right]^{-1},$$

for $|x-s| \leqslant 5\lambda$ and $|y-t| \leqslant 5\lambda$, for any $x,y \in \{1,2,\ldots,m\}$.

The dependence on the $\{a_\lambda(p,q)\}$ of the different smoothing parameter scores can often be written as a function of the trace of the matrix A_λ, $Tr\{A_\lambda\}$. Using the previously defined notation and properties of the M x M matrix A_λ, the following expressions may be derived:

$$Tr\{A_\lambda\} = M\,\tilde{a}_\lambda, \tag{4.10a}$$

$$Tr\{A_\lambda^2\} = M\,\tilde{a}_\lambda^2, \tag{4.10b}$$

$$Tr\{(I - A_\lambda)^2\} = \sum_{p=1}^{M}\{1 - a_\lambda(p,p)\}^2 = M(1 - 2\tilde{a}_\lambda + \tilde{a}_\lambda^2). \tag{4.10c}$$

In the following sections, expressions for the smoothing parameter scores will be derived in terms of the algebraically more convenient linear model (4.7), with weights A_λ. However, to emphasise the lattice structure of the data and that smoothing occurs in two

dimensions, the final form of individual scores in Sections 4.3.2 - 4.3.5 will be expressed in terms of the bivariate kernel model (4.6), that is, with coordinates (x,y) and weights W_λ.

Due to the nature of the thigh MRI, when calculating the smoothing parameter scores in practice, the summation is not over all m^2 points (x,y). Recall from Section 2.2.1 that for the data used here, intensities of zero are recorded at many of the points in the outer columns and rows of the 128 x 128 lattice, corresponding to the 'background air'. Since the objective is to estimate the amount of smoothing required over the image thigh area, such background points are excluded from the summation. Instead, new approximate column and row limits are defined which enclose the thigh area, but exclude most of the background zeros. Scores are calculated with respect to this smaller lattice area. However for generality and because the new column and row limits are different for each subject, the scores defined in the following sections will be defined in terms of the limits $\{x,y : 1,2,\ldots,m\}$.

4.3.2 The chi-squared estimate, λ(CHI)

As given in Section 4.2.2, the chi-squared estimate for λ is obtained as the solution to:

$$RSS_\lambda \; / \; M \;=\; \sigma^2. \qquad (4.11)$$

Since an explicit solution is not possible, the left-hand side of (4.11) is evaluated for a sequence of λ values and compared against the estimated error variance $\hat{\sigma}^2$. Once an interval (with respect to λ) within which λ(CHI) lies has been identified, a more precise estimate is optimised using a golden section search (see Fletcher, 1987).

4.3.3 The empirical degrees of freedom estimate, λ(EDF)

Recall from Section 4.2.3 that λ(EDF) is evaluated numerically as the solution to $RSS_\lambda \; / \; EDF_\lambda = \sigma^2$. The total number of degrees of

freedom is M, while the number of degrees of freedom lost when estimating μ_λ is given by $Tr\{A_\lambda\}$. Using expression (4.10a), the equivalent degrees of freedom for error can be written as:

$$EDF_\lambda = M - Tr\{A_\lambda\} = M - M\tilde{a}_\lambda = M[1 - w_\lambda(0,0)]. \qquad (4.12)$$

Thus λ(EDF) is the solution of:

$$RSS_\lambda / M[1 - w_\lambda(0,0)] = \sigma^2. \qquad (4.13)$$

A solution can be obtained numerically as for the χ^2 method: the left-hand side of (4.13) is evaluated at a sequence of λ values, and the estimate λ(EDF) optimised by a golden section search.

4.3.4 The Rice estimate, λ(RICE)

Recall from Section 4.2.4 that the motivation for the Rice method was the minimisation of the EMSE criterion. The bivariate form of the EMSE (4.4) can be rewritten in terms of the linear model of Section 4.3.1 as:

$$EMSE = \frac{1}{M} E\left[\sum_{p=1}^{M} \{\mu(p) - \hat{\mu}_\lambda(p)\}^2\right] = \frac{1}{M} E\left[\, \| \underline{\mu} - A_\lambda \underline{z} \|^2 \,\right],$$

where $\| \quad \|^2$ is the Euclidean norm. Evaluating the expectation operator, then using the mean and covariance properties of $\underline{\epsilon}$ and expression (4.10a) for $Tr\{A_\lambda\}$, $EMSE_\lambda$ may be written as:

$$EMSE_\lambda = \frac{1}{M} \| (I - A_\lambda)\underline{\mu} \|^2 + \frac{\sigma^2}{M} Tr\{A_\lambda\} = \frac{1}{M} \| (I - A_\lambda)\underline{\mu} \|^2 + \sigma^2 \tilde{a}_\lambda^2.$$

In theory, the minimiser of EMSE is taken as the estimate of the optimal smoothing parameter λ.

The EMSE must be estimated from the data since the 'true' function $\underline{\mu}$ is not usually available. This estimated criterion is then minimised with respect to λ. So, replacing $\underline{\mu}$ by its observed value $\underline{z}$

in the original MSE lack-of-fit (4.2) gives the RSS as:

$$RSS_\lambda = \frac{1}{M} \sum_{p=1}^{M} \{z(p) - \hat{\mu}_\lambda(p)\}^2 = \frac{1}{M} \| (I - A_\lambda)\underline{z} \|^2,$$

where $\underline{\hat{\mu}}_\lambda = A_\lambda \underline{z}$. The estimated RSS should be unbiased for the EMSE, that is, it is required that $E[RSS_\lambda] = EMSE$. Taking the expected value of the RSS it can be shown that:

$$E[RSS_\lambda] = \frac{1}{M} \| (I - A_\lambda)\underline{\mu} \|^2 + \frac{\sigma^2}{M} Tr\{(I - A_\lambda)^2\},$$

or simplifying using expression (4.10c):

$$E[RSS_\lambda] = \frac{1}{M} \| (I - A_\lambda)\underline{\mu} \|^2 + \sigma^2(1 - 2\tilde{a}_\lambda + \tilde{a}_\lambda^2) = EMSE_\lambda + \sigma^2(1 - 2\tilde{a}_\lambda).$$

Thus RSS_λ is a biased estimator for $EMSE_\lambda$, with the bias equal to $\sigma^2(1 - 2\tilde{a}_\lambda)$. However, an unbiased estimator of EMSE is $RICE_\lambda$, the so-called Rice score:

$$RICE_\lambda = RSS_\lambda + \sigma^2(2\tilde{a}_\lambda - 1), \qquad (4.14)$$

where RSS_λ is as given in (4.3). Minimisation with respect to λ of the data-based score should result in a good estimator of EMSE, the theoretical lack-of-fit criterion. Hence the minimiser of $RICE_\lambda$ should be a good estimate of the optimal smoothing parameter λ.

4.3.5 The cross-validation estimate, $\lambda(CV)$

As outlined in Section 4.2.5, the cross-validation method is motivated by minimisation of the MSE, which provides an estimate for λ which is totally data-based, since unlike the Rice method it does not require an estimate of σ^2. The estimate $\hat{\mu}_\lambda$ is replaced by its 'leave-one-out' cross-validation estimate $\hat{\mu}_\lambda^C$ and λ is estimated such that, on average, the mean of the data is most closely predicted. Thus the estimate $\lambda(CV)$ is the minimiser of the cross-validation score, CV_λ (4.5). In terms of the linear model (4.7) of Section 4.3.1

CV_λ is given by:

$$CV_\lambda = \frac{1}{M} \sum_{p=1}^{M} \{z(p) - \hat{\mu}_\lambda^c(p)\}^2. \qquad (4.15)$$

The CV estimate $\hat{\mu}_\lambda^c(p)$ can be expressed (Craven and Wahba, 1979) as a weighted function of the estimate $\hat{\mu}_\lambda(p)$ based on all the data and the omitted observation $z(p)$, that is:

$$\hat{\mu}_\lambda^c(p) = \frac{\hat{\mu}_\lambda(p) - a_\lambda(p,p)z(p)}{1 - a_\lambda(p,p)}. \qquad (4.16)$$

Substituting (4.16) for $\hat{\mu}_\lambda^c(p)$ into (4.15) and simplifying gives a computationally more efficient form of the score as:

$$CV_\lambda = \frac{1}{M} \sum_{p=1}^{M} \frac{\{z(p) - \hat{\mu}_\lambda(p)\}^2}{\{1 - a_\lambda(p,p)\}^2}. \qquad (4.17)$$

To reduce the number of separate weights $a_\lambda(p,p)$ to be computed, Craven and Wahba (1979) proposed replacing individual weights by the average value as given by:

$$\frac{1}{M} \sum_{p=1}^{M} a_\lambda(p,p) = \frac{1}{M} \mathrm{Tr}\{A_\lambda\}.$$

This modification gives the generalised cross-validation (GCV) score:

$$GCV_\lambda = \frac{1}{M} \sum_{p=1}^{M} \frac{\{z(p) - \hat{\mu}_\lambda(p)\}^2}{\{1 - (\mathrm{Tr}\{A_\lambda\})/M\}}. \qquad (4.18)$$

For the special case holding here of estimates restricted to integer lattice points, then the CV and GCV methods are equivalent. This is a consequence of the structure of the matrix A_λ, that is, with constant weight $\tilde{a}_\lambda$ on the main diagonal, so substituting (4.10a) into (4.17) for GCV_λ:

$$GCV_\lambda = \frac{1}{M} \sum_{p=1}^{M} \frac{\{z(p) - \hat{\mu}_\lambda(p)\}^2}{\{1 - (M\tilde{a}_\lambda)/M\}}.$$

Here the method will be termed cross-validation, with the understanding that it is equivalent to generalised cross-validation. As noted earlier, to emphasise the lattice structure of the data, it is helpful to rewrite the scores for estimating λ in terms of grid coordinates $\{(x,y) : x,y = 1,2,\ldots,m\}$ and kernel weights W_λ, whereby CV_λ is given by:

$$CV_\lambda = \frac{1}{M} \sum_{x=1}^{m} \sum_{y=1}^{m} \frac{\{z(x,y) - \hat{\mu}_\lambda(x,y)\}^2}{\{1 - w_\lambda(0,0)\}^2}.$$

The Rice and CV methods are both motivated by minimisation of the theoretical MSE criterion. The respective scores have been shown (Müller, 1988) to be asymptotically equivalent to the IMSE, that is $E[RICE_\lambda] \approx IMSE_\lambda$ and $E[CV_\lambda] \approx IMSE_\lambda + \sigma^2$. Thus the two scores are asymptotically equivalent except for a constant σ^2, a conclusion which was also reached by Craven and Wahba (1979). By rearranging (4.14) and (4.17) and substituting for the residual sum of squares (4.3), the relationship between the Rice and the CV scores can be written algebraically as:

$$CV_\lambda = \frac{RICE_\lambda - \sigma^2 \{2\tilde{a}_\lambda - 1\}}{\{1 - \tilde{a}_\lambda\}^2},$$

$$RICE_\lambda = \{1 - \tilde{a}_\lambda\}^2 CV_\lambda + \sigma^2 \{2\tilde{a}_\lambda - 1\}.$$

The estimates $\lambda(RICE)$ and $\lambda(CV)$ minimising $RICE_\lambda$ and CV_λ respectively, should be asymptotically equivalent, and equal to the theoretical minimiser of the MSE lack-of-fit.

Asymptotic expressions for the limiting values of the scores can be derived and expressed in terms of the observed data. As $\lambda \to 0$:

$$CV_\lambda \to \frac{1}{4M} \sum_{x=1}^{m} \sum_{y=1}^{m} \{4z(x,y) - [z(x-1,y)+z(x+1,y)+z(x,y-1)+z(x,y+1)]\}^2.$$

This limiting score implies that the intensity at any point is estimated by the average of its four nearest neighbours. Although it is much easier to calculate this simple average, rather than using a kernel function to estimate intensity, such an approach is not suitable for the current application. Using this average intensity, values are only estimated at discrete data points. The underlying continuous function cannot be estimated and so derivatives, in particular the Laplacian, cannot be evaluated.

As $\lambda \to \infty$, the limiting score is given by:

$$CV_\lambda = \frac{1}{M} \sum_{x=1}^{m} \sum_{y=1}^{m} \{z(x,y)\}^2.$$

The limiting value implies that a common intensity at all positions is estimated as the average intensity calculated over all data points.

4.4 Estimation of the smoothing parameter for derivatives

4.4.1 Existing methods

Derivatives of a regression function are often also of interest, for example in the study of growth curves to estimate changes in the rate of growth (Gasser and Müller, 1984), or as in this study, the characterisation of intensity changes corresponding to edge positions. As for the function itself the derivative kernel is indexed by a smoothing parameter λ, for which a 'suitable' value must be selected.

It is not immediately obvious how a smoothing parameter for derivatives should be estimated. Any noise in the original data will be emphasised by calculating derivatives of a continuous function, or equivalently, differences for discrete data. Müller (1988, Chapter 7) concluded that the optimal smoothing parameter was different for each order of derivative and that asymptotically it would be wrong to use the same λ for derivatives as for the function itself. This is in

contrast to the observations of Craven and Wahba (1979), who considered cross-validation for smoothing splines. They used simulations to show that if an estimated parameter (for example, the minimiser of the CV score function) was optimal for the smoothed function μ_λ, then the same parameter was a good estimate of the optimal parameter for the derivative of μ_λ. The motivation for this approach is the assumption that if the underlying data is well estimated by the fitted function, then a good estimate of the derivative is possible by differentiating this smoothed function. Polynomial functions were mentioned briefly (Section 3.3.3) as possible kernels. For such kernels an asymptotically optimal smoothing parameter can be defined depending on the kernel used and derivative required (Gasser and Müller, 1984).

Methods have been proposed for the estimation of parameters specifically for derivatives. Using difference quotients, Rice (1986) generalised his original method, while the CV method was similarly modified by Müller (1988, Chapter 7). Difference quotients of the observed data were compared with derivatives of the kernel estimates, the latter constructed omitting the observations which were used for the difference quotients; see also Härdle (1990).

One disadvantage of both the modified Rice and CV methods is that they rely on difference quotients which are unstable, becoming progressively worse as higher differences are calculated. One option which does not require such differences is the 'factor' method of Müller et al (1987), summarised by Müller (1988, Chapter 5). According to the factor method, the theoretically optimum parameter for the derivative is equivalent to the optimum parameter for the function itself, except for a factor depending on the kernel used, the derivative required and the number of continuous derivatives of the kernel. Asymptotic results were derived without assuming a particular functional form for the kernel, although polynomial kernels (Section 3.3.3) were used in simulations to compare methods for small samples. Smoothing parameters for derivatives of the function estimated by the factor method were compared with values of λ from the cross-validation and Rice quotient scores, and also from the CV score (Section 4.3.5) for the function itself. The values of λ

estimated from the factor method were superior in almost all cases and were also closer to the theoretically optimal value.

Derivatives of the kernel function are required for the estimation of edges, but because the intensity derivative function is not available none of the methods previously discussed are directly applicable. Data are observed only at discrete points on an integer lattice, so first and higher order derivatives are approximated by the corresponding differences. A new method based on differences is defined here for the estimation of a smoothing parameter appropriate for the approximated derivatives. First difference and second difference scores are defined as an extension of the original cross-validation score (4.16), but replacing the function or observation by the corresponding difference. Higher difference scores could be similarly defined but are not considered here since only first and second derivatives are required for edge identification. The cross-validation (first difference) method will be described in detail (Section 4.4.2); the extension to an analagous cross-validation (second differences) method (Section 4.4.3) is straightforward. A further method estimating the parameter for second derivatives which cross-validates the Laplacian of the kernel estimate directly was suggested by Kay (1991, personal communication). This method is outlined in Section 4.4.3.

4.4.2 The cross-validation (first difference) method

In two dimensions, the cross-validation (first difference) score, $CVD1_\lambda$, is the sum of two components calculated separately in the horizontal direction (over columns, x) and in the vertical direction (over rows, y). Since both components are calculated in the same way, formulae will be given only for the horizontal score but analagous results hold for the vertical score. In the horizontal direction, first differences between observations and cross-validation estimates are calculated as $\{z(x,y)-z(x-1,y)\}$ and $\{\hat{\mu}_\lambda^c(x,y)-\hat{\mu}_\lambda^c(x-1,y)\}$ respectively. Alternatively, with respect to the matrix notation of Section 4.3.1, comparisons are between positions p and (p-1), so corresponding first differences are $\{z(p)-z(p-1)\}$ and

$\{\hat{\mu}_\lambda^c(p)-\hat{\mu}_\lambda^c(p-1)\}$. Then, rather than comparing the observed value with the CV estimate as for the CV score (4.14), a comparison is now made between the differences in observed values and the differences in estimates μ_λ.

A corresponding cross-validation first difference sum of squares in the horizontal direction, $CVD1H_\lambda$ say, is calculated as:

$$CVD1H_\lambda = \sum_{\Omega(p)} \left[\{z(p)-z(p-1)\} - \{\hat{\mu}_\lambda^c(p)-\hat{\mu}_\lambda^c(p-1)\} \right]^2, \qquad (4.19)$$

where $\Omega(p)$ denotes summation over $\{p = 2,3,\ldots,M,\ p \neq \alpha\, m \text{ for } \alpha \in \mathbb{Z}\}$. First differences are similarly calculated in the vertical direction between lattice points (x,y) and (x-1,y), or equivalently, with respect to the matrix model, positions p and (p-m) and an analagous vertical cross-validation sum of squares, $CVD1V_\lambda$, can be defined. The cross-validation (first difference) score is defined as the average of the sum of squares in the horizontal and vertical directions:

$$CVD1_\lambda = \frac{1}{(m-1)m} (CVD1H_\lambda + CVD1V_\lambda).$$

The minimiser $\lambda(CVD1)$ of $CVD1_\lambda$ is the estimate of the smoothing parameter for the first derivative function. The difference score is calculated at a sequence of values of λ, followed by a golden section optimisation to locate the minimum.

As for the CV score (4.15), calculation of the explicit CV estimates can be avoided by using a computational simplification. As given by (4.16), the CV estimate $\hat{\mu}_\lambda^c(p)$ can be re-expressed as a weighted sum of $\hat{\mu}_\lambda(p)$, the estimate based on all the data and z(p), the omitted observation. Analagous simplifications can be made for $\hat{\mu}_\lambda^c(p-1)$ and $\hat{\mu}_\lambda^c(p-m)$, for first differences in the horizontal and vertical directions respectively. Considering $\hat{\mu}_\lambda^c(p-1)$ for example, since $a_\lambda(p-1,p-1)$ is an entry on the diagonal of A_λ it is equal to the constant $\tilde{a}_\lambda$ and hence:

$$\hat{\mu}_\lambda^c(p-1) = \frac{\hat{\mu}_\lambda(p-1) - a_\lambda(p-1,p-1)z(p-1)}{1 - a_\lambda(p-1,p-1)} = \frac{\hat{\mu}_\lambda(p-1) - \tilde{a}_\lambda z(p-1)}{1 - \tilde{a}_\lambda}.$$

Substituting the expressions for $\hat{\mu}_\lambda^c(p)$ and $\hat{\mu}_\lambda^c(p-1)$ into (4.19) for $CVD1H_\lambda$ and simplyfying gives:

$$CVD1H_\lambda = \sum_{\Omega(p)} \frac{[\,\{z(p)-z(p-1)\} - \{\hat{\mu}_\lambda(p)-\hat{\mu}_\lambda(p-1)\}\,]^2}{[1 - \tilde{a}_\lambda]^2},$$

where $\Omega(p)$ denotes summation over $\{p = 2,3,\ldots,M,\ p \neq \alpha\, m,$ for $\alpha \in \mathbb{Z}\}$. Finally, rewriting in terms of the lattice coordinates (x,y):

$$CVD1H_\lambda = \sum_{x=2}^{m} \sum_{y=1}^{m} \frac{[\,\{z(x,y)-z(x-1,y)\} - \{\hat{\mu}_\lambda(x,y)-\hat{\mu}_\lambda(x-1,y)\}\,]^2}{[1 - w_\lambda(0,0)]^2},$$

and similarly for $CVD1V_\lambda$.

The asymptotic value of $CVD1_\lambda$ as $\lambda \to 0$ is evaluated, but for brevity the complex algebraic expression is omitted. The asymptotic values of $CVD1_\lambda$ for $\lambda \to \infty$ are also considered. Expressed here in terms of the lattice coordinates (x,y), the limiting values of the sums of squares in the horizontal and vertical directions respectively are given by:

$$CVD1H_\lambda \to \sum_{x=2}^{m} \sum_{y=1}^{m} \{z(x,y)-z(x-1,y)\}^2, \quad CVD1V_\lambda \to \sum_{x=1}^{m} \sum_{y=2}^{m} \{z(x,y)-z(x,y-1).$$

As before the score $CVD1_\lambda$ is calculated as the average of the sum of $CVD1H_\lambda$ and $CVD1V_\lambda$. Thus for 'large' λ, the limiting value of $CVD1_\lambda$ is equivalent to the sum of the average of first differences of the observations calculated separately in the horizontal and vertical directions.

4.4.3 The cross-validation (second difference) method

The cross-validation (second difference) score is denoted by $CVD2_\lambda$. As for first differences, sums of squares in the horizontal and vertical directions, $CVD2H_\lambda$ and $CVD2V_\lambda$ respectively, are calculated separately. Only $CVD2H_\lambda$ will be defined explicitly but

analagous formulae apply for $CVD2V_\lambda$. In each direction the second difference is calculated as the first difference of each first difference. For the observed data the second difference is calculated as $\{z(x,y)-z(x-1,y)\} - \{z(x-1,y)-z(x-2,y)\}$, which simplifies to $\{z(x,y)-2z(x-1,y)+z(x-2,y)\}$. Similarly, the (second difference) of the CV estimates simplifies to $\{\hat{\mu}_\lambda^c(x,y)-2\hat{\mu}_\lambda^c(x-1,y)+\hat{\mu}_\lambda^c(x-2,y)\}$. In terms of the matrix notation of Section 4.3.1, comparisons are made between positions p, (p-1) and (p-2) in the horizontal direction and positions p, (p-m) and (p-2m) in the vertical direction.

Second differences of observed intensities and CV estimates are compared to obtain $CVD2H_\lambda$ as:

$$CVD2H_\lambda = \sum_{x=3}^{m} \sum_{y=1}^{m} \Big[\{z(x,y)-2z(x-1,y)+z(x-2,y)\} - \{\hat{\mu}_\lambda^c(x,y)-2\hat{\mu}_\lambda(x-1,y)+\hat{\mu}_\lambda^c(x-2,y)\} \Big]^2 \qquad (4.20)$$

and similarly for $CVD2V_\lambda$, the corresponding sum of squares in the vertical direction. A cross-product term is not included since the second differences are viewed as an approximation to the Laplacian second derivative operator (3.10), which is calculated as the sum of the two partial derivatives with respect to x and y. Thus the cross-validation (second difference) score, $CVD2_\lambda$, is defined as the average of the sum of the two second difference sums of squares in the horizontal and vertical directions:

$$CVD2_\lambda = \frac{1}{(m-2)m} (CVD2H_\lambda + CVD2V_\lambda).$$

The value of the parameter $\lambda = \lambda(CVD2)$ minimising this score is the cross-validation (second difference) estimate, optimised by a golden section search after $CVD2_\lambda$ has been evaluated at a sequence of values of λ. As was shown in detail for $CVD1H_\lambda$, the sum of squares (4.20) can be simplified algebraically using the constant diagonal weight property of the matrix A_λ. Thus with respect to the lattice coordinates an alternative computational form for $CVD2H_\lambda$ is given by:

$$\sum_{x=3}^{m}\sum_{y=1}^{m}\frac{\left[\{z(x,y)-2z(x-1,y)+z(x-2)\}-\{\hat{\mu}_\lambda(x,y)-2\hat{\mu}_\lambda(x-1,y)+\hat{\mu}_\lambda(x-2,y)\}\right]^2}{[1-w_\lambda(0,0)]^2}.$$

and similarly for the vertical sum of squares, $CVD2V_\lambda$.

For brevity the complex expression is omitted for the asymptotic value of $CVD2_\lambda$ as $\lambda \to 0$. Asymptotic values of $CVD2_\lambda$ as $\lambda \to \infty$ are also evaluated. Expressed here in terms of the lattice coordinates, the limiting values of the sum of squares in the horizontal and vertical directions are given by:

$$CVD2H_\lambda \to \sum_{x=3}^{m}\sum_{y=1}^{m}\{z(x,y)-2z(x-1,y)-z(x-2,y)\}^2,$$

$$CVD2V_\lambda \to \sum_{x=1}^{m}\sum_{y=3}^{m}\{z(x,y)-2z(x,y-1)-z(x,y-2)\}^2.$$

The score $CVD2_\lambda$ is calculated as the average of the sum of $CVD2H_\lambda$ and $CVD2V_\lambda$. Thus for 'large' λ the limiting value of $CVD2_\lambda$ is equivalent to the sum of the average of the second differences of the observations calculated separately in the horizontal and vertical directions.

Since the Laplacian of the smoothed image is estimated when identifying edges, Kay (1991, personal communication) suggested that the Laplacian function (or its estimates) should be cross-validated directly. The Laplacian at position (x,y), that is $\nabla^2\{\hat{\mu}_\lambda(x,y)\}$, is estimated by (3.10), or more generally by (4.6), where the range of summation for the subset of points contributing to the estimate at (x,y) is made explicit. The first order neighbourhood of pixel (x,y), that is the points $\{(i-1,j),(i,j),(i+1,j),(i,j-1),(i,j+1)\}$ is denoted by $N_{ij}(x,y)$. The image can be divided into γ subsets of such neighbourhoods: here $N_{ij}(x,y)$ is considered for each point (x,y) in turn, so the neighbourhoods overlap. The image could alternately be partitioned in several ways into non-overlapping neighbourhoods, but this is not considered here. The discretised Laplacian can be

calculated in each neighbourhood as:

$$\mathcal{S}\{N_{ij}(x,y)\} = z(i-1,j) + z(i+1,j) + z(i,j-1) + z(i,j+1) - 4z(i,j).$$

The 'leave-one-out' CV estimate of the Laplacian is denoted by $\nabla^{2c}\{\hat{\mu}_\lambda(x,y)\}$, where the range of summation over (i,j) as given by (4.6) is modified to further exclude the subset of points $N_{ij}(x,y)$.

Kay (1991, personal communication) defined his CV score function as:

$$CVKAY_\lambda = \frac{1}{\gamma} \sum_{\Gamma(x,y)} \sum [\mathcal{S}\{N_{ij}(x,y)\} - \nabla^{2c}\{\hat{\mu}_\lambda(x,y)\}]^2,$$

where $\Gamma(x,y)$ denotes the summation over all subsets (x,y). The $CVKAY_\lambda$ score function is evaluated at a sequence of values of λ and the value $\lambda = \lambda(CVKAY)$ minimising the function is the selected smoothing parameter.

4.5 Estimation of the error variance σ^2

4.5.1 Univariate estimators

Three of the methods for smoothing parameter estimation, that is the χ^2, EDF and Rice methods, are functions of the error variance σ^2. Several data-based estimators of σ^2 are now considered. The estimation of an appropriate smoothing parameter is dependent on the availability of a 'good' estimator. Underestimation of the error variance can mean that parameters estimated using the Rice method are too small, and consequently the data is undersmoothed (Müller et al, 1987). Parameters selected by the χ^2 method appear to be unaffected by the estimator of σ^2 used, while those selected by the EDF method are more sensitive, though data-based estimators work well (Thompson et al, 1991a).

For the current data, the error variance is estimated for bivariate design points: the bivariate estimator used in this thesis will be discussed in the next section. Several estimators of σ^2 have

been proposed, but most are defined for one-dimensional data. Univariate estimators previously discussed in the context of nonparametric regression incorporating some form of smoothing are now summarised.

In the context of nonlinear regression, Gasser et al (1986) estimated the residual error variance as a weighted sum of squared pseudo-residuals, calculated from continuous triples of design points. In addition to proposing the Rice method (Section 4.2.4) for smoothing parameter estimation, Rice (1984) defined two estimators for σ^2. For univariate data, he fitted constants to successive pairs or triples of points. Both of these methods of estimating σ^2 are based on differences of the data. In contrast, Wahba (1983) defined an estimator combining properties from both the (generalised) cross-validation and the EDF methods.

Thompson et al (1991b) summarised several estimators which were quadratic functions of the data. In addition to the three outlined here, they discussed an extension of the difference-based estimators to consider asymptotically optimal differences (Hall et al, 1990) and also estimators based on the minimax method of Buckley et al (1988). The performance of the estimators was compared empirically on the basis of four simulated one-dimensional functions. One consideration when choosing an estimator is how robust it is to discontinuities in the data, for example the presence of edges in an image, which can result in a biased estimator. Thompson et al (1991b) also discussed two estimators which aim to account for this lack of continuity, either by 'trimming-out' the observations which bias the estimator (Kimber, 1983) or by identifying edge points and 'editing out' the corresponding observations (Grimson and Pavlidis, 1985). However such methods of bias correction are not considered in this thesis.

4.5.2 Bivariate estimators

The bivariate estimators of σ^2 used in this thesis belong to a family of estimators proposed by Kay (1988). Although Kay discussed the estimators for the restoration of images, rather than for

smoothing (see Section 4.3.1 for a comparison of the two contexts), the presence of edges in the image can affect the estimator in the same way and so this is viewed as the most appropriate method of estimating σ^2. The estimators are a generalisation to two dimensions of the difference-based estimators of Rice (1984). Kay (1988) defined the estimators for any general order of difference, but only those (S1 and S2, say) based on first and second differences are considered here.

Residuals, r1(i,j) and r2(i,j) say, are calculated within a local 2 x 2 or 3 x 3 neighbourhood of an observation z(i,j) as:

$$r1(i,j) = [\{z(i,j)+z(i+1,j+1)\} - \{z(i+1,j)+z(i,j+1)\}] / 2,$$

$$r2(i,j) = [\{z(i,j)+z(i,j+2)+4z(i+1,j+1)+z(i+2,j)+z(i+2,j+2)\} + \{2z(i,j+1)+2z(i+1,j)+2z(i+1,j+2)+2z(i+2,j+1)\}] / 6.$$

The lattice coordinates i and j are used here to define the residuals in a more general form, rather than indices x and y which are restricted to be integer valued for this chapter only. Kay (1988) used the mean of the squared residuals to define empirical estimators S1 and S2 as:

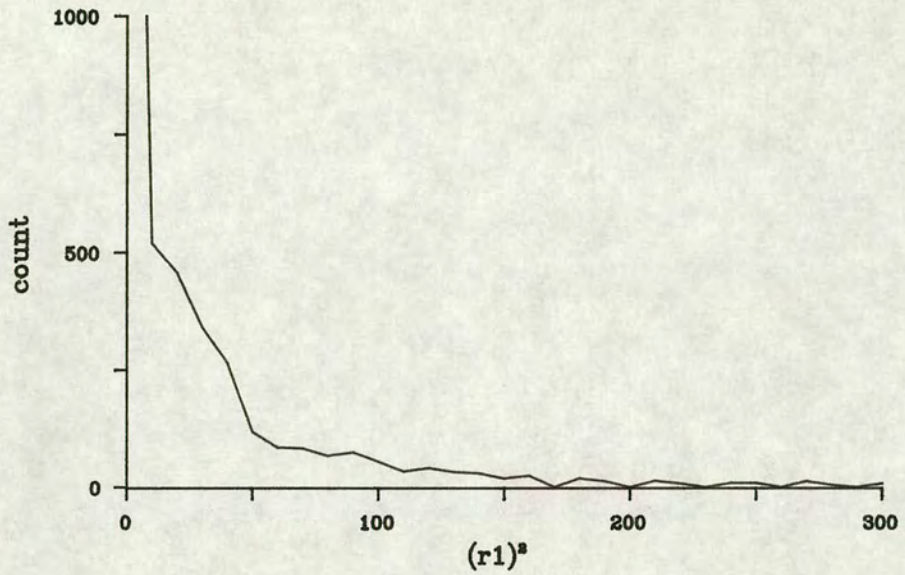
$$S1 = \frac{1}{(m-1)^2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \{r1(i,j)\}^2,$$

$$S2 = \frac{1}{(m-2)^2} \sum_{i=1}^{m-2} \sum_{j=1}^{m-2} \{r2(i,j)\}^2.$$

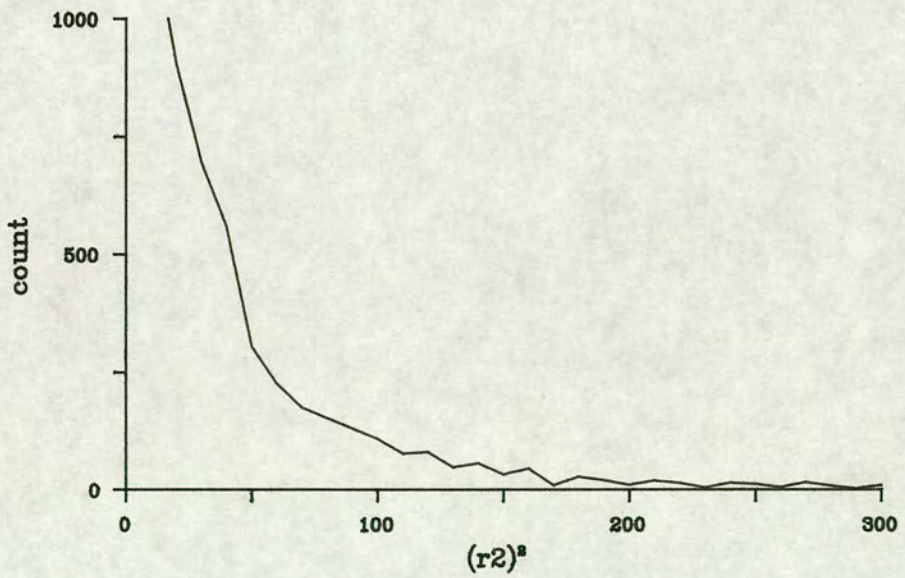
So defined, S1 and S2 are consistent estimators for σ^2, but in practice it may be necessary to correct for bias. The typical distributions of squared residuals for variables z1 and z2 for one image (for subject et1, variable z1) are illustrated in Figure 4.3. In both cases the distribution is strongly skewed to the right, due to the large number of edge points in the image. Kay (1988) suggested either trimming out the bias or identifing and excluding from the estimation any neighbourhoods containing an edge, methods for which were discussed by Thompson et al (1991b).

Figure 4.3: Distribution of neighbourhood residuals
– Subject et1, Variable z1

a) 2 x 2 neighbourhood



b) 3 x 3 neighbourhood



Note: For a) and b), both x– and y– axes truncated to show more detail of main body of distribution

A different approach is taken here: the median of the squared residuals is proposed as a more robust estimator to allow for such bias. The two estimators for σ^2 used in this thesis are defined as:

$$S1 = \underset{i,j}{\text{median}} \left[\{r1(i,j)\}^2 \right],$$

$$S2 = \underset{i,j}{\text{median}} \left[\{r2(i,j)\}^2 \right].$$

Estimates of S1 and S2 for the data sets for four subjects are given in Table 4.1, for both variables z1 and z2. Values of the lower and upper quartiles (Q_1, Q_3) are also tabulated for the respective distribution for which S1 and S2 are the median values. The high value of the upper quartiles confirms the right-skewed distribution of the data.

4.6 Results: Estimated parameters λ

4.6.1 Introduction

The respective scores for the methods described in Sections 4.3 and 4.4 are evaluated at a sequence of values of λ incremented in a geometric progression. The minimum value of λ considered is $\lambda = 0.5$: for values of $\lambda < 0.5$ there is no longer smoothing over neighbouring points and intensity is estimated as the observation at each point. Cases where the parameter selected would be greater than the maximum value of λ for which the score would be evaluated are denoted by $\lambda > 20.0$. For values of λ of this magnitude, the calculated scores change very slowly as λ increases and the smoothed estimate at each point is calculated as the weighted average of almost all the other observations in the image. Such values of λ would not be used in practice: if such high levels of smoothing were appropriate, the preferred method of estimating intensity at each point would be as in the limiting case for $\lambda \to \infty$, that is as the average of all the observations in the image. This avoids the extensive computation involved in smoothing over such large windows.

Table 4.1: Distribution of neighbourhood residuals for estimation of error variance σ^2

Table 4.1a: Variable z1

	et1		et2		ft1		ft2	
	S1	S2	S1	S2	S1	S2	S1	S2
Median	9.0	11.1	9.0	12.3	9.0	10.0	6.3	8.0
Q_1	1.0	1.4	1.0	1.8	1.0	1.0	0.3	0.4
Q_3	30.3	36.0	36.0	36.0	30.3	34.0	30.3	32.1

Table 4.1b: Variable z2

	et1		et2		ft1		ft2	
	S1	S2	S1	S2	S1	S2	S1	S2
Median	6.3	5.4	6.3	6.3	6.3	4.7	6.3	5.4
Q_1	0.3	0.7	1.0	0.7	0.3	0.4	0.3	0.3
Q_3	30.3	21.8	30.3	23.4	25.0	20.3	30.3	23.4

The parameters estimated are presented in Tables 4.2a and 4.2b, for variables z1 and z2 respectively, and for four images: et1, et2, ft1 and ft2. Much of the following discussion will refer to the variable z1, since most of the results for z2 are inconclusive and the selected values of λ are unsuitable for smoothing the images.

4.6.2 Parameters estimated for the data

When smoothing the observed data, the parameter λ may be estimated by one of the χ^2, the EDF, the Rice or the CV methods. For the first three methods, two values of λ are given depending on whether the error variance σ^2 was estimated by S1 or S2. In practice, there was only a slight difference between the parameters estimated for a given data set using the two estimators of σ^2.

For a given method the values of λ are reasonably consistent across data sets. All the parameters are 'small', indicating only low levels of smoothing. In cases where the minimum of the CV score function occurs below the minimum value of λ (= 0.5) for which the scores were evaluated, this is recorded as $\lambda < 0.5$. For both variables z1 and z2 the parameter estimated by the Rice, χ^2 or EDF methods is very small; these 'automatically' estimated parameters are not suitable for edge identification as will be discussed in Section 4.7.1.

Even though the parameters estimated are not particularly useful for this application, it is interesting to verify the expected theoretical relations between the methods. In all cases the parameters λ(CHI) are greater than the corresponding parameter λ(EDF), as discussed by Thompson et al (1991a). The parameters λ(RICE) and λ(CV) are of a similar magnitude (Müller, 1988), but because in most cases there is no minimum of the CV score function it is not possible to compare individual parameters exactly.

For variable z1 and data set ft1, there is a minimum of the CV score function at $\lambda = 0.39$, the value of λ which would therefore be taken as the smoothing parameter. In all other cases for z1, and all

Table 4.2: Estimates of the smoothing parameter λ

Table 4.2a: Variable z1

Method	Parameter		et1	et2	ft1	ft2
χ^2	λ(CHI)	S1	0.55	0.53	0.56	0.50
		S2	0.58	0.58	0.57	0.53
EDF	λ(EDF)	S1	0.44	0.42	0.44	0.40
		S2	0.46	0.46	0.45	0.42
RICE	λ(RICE)	S1	0.44	0.42	0.44	0.40
		S2	0.46	0.45	0.45	0.42
CV	λ(CV)		<0.50	<0.50	0.39	<0.50
CVD1	λ(CVD1)		0.84	0.80	0.87	0.82
CVD2	λ(CVD2)		>20.00	>20.00	>20.00	>20.00

Table 4.2b: Variable z2

Method	Parameter		et1	et2	ft1	ft2
χ^2	λ(CHI)	S1	0.46	0.45	0.47	0.46
		S2	0.45	0.45	0.45	0.45
EDF	λ(EDF)	S1	0.37	0.36	0.38	0.36
		S2	0.36	0.36	0.36	0.36
RICE	λ(RICE)	S1	0.37	0.36	0.38	0.36
		S2	0.36	0.36	0.36	0.36
CV	λ(CV)		<0.50	<0.50	<0.50	<0.50
CVD1	λ(CVD1)		<0.50	<0.50	<0.50	0.40
CVD2	λ(CVD2)		<0.50	<0.50	0.72	>20.00

data sets for z2, the CV score function is monotonically increasing. The score calculated for $\lambda = 0.5$ is virtually identical to the limiting score in the limit as $\lambda \to 0$ (Section 4.3.5). The score calculated at the highest value of λ considered is tending towards the limiting score in the limit as $\lambda \to \infty$, thus indicating that there is unlikely to be a minimum of the score function outside the range of λ considered. This is consistent with the results of Thompson et al (1989) who concluded that the parameters estimated by CV were not always satisfactory.

4.6.3 Parameters for derivatives

For variable z1, the cross-validation (first difference) score $CVD1_\lambda$ does have a minimum; the parameter λ(CVD1) is taken as the value of λ where this minimum occurs. The parameters estimated are reasonably consistent across data sets, around $\lambda = 0.8$. Although values of λ(CVD1) are higher than those for smoothing the data, that is the parameters λ(CHI), λ(EDF), λ(RICE), λ(CV), such λ(CVD1) are still indicating only low levels of smoothing. Evaluated at $\lambda = 0.5$, the $CVD1_\lambda$ score is tending towards the limiting score in the limit $\lambda \to 0$. Similarly, the $CVD1_\lambda$ score calculated for the highest value of λ considered is tending towards the limiting score in the limit as $\lambda \to \infty$. This suggests that the value λ(CVD1) in Table 4.2a is the only minimum of the $CVD1_\lambda$ score and hence the 'automatically' estimated parameter for first differences of z1. For variable z2, there is a minimum of the $CVD1_\lambda$ function only for data set ft2, estimating a parameter of λ(CVD1) = 0.4. For the other data sets the score at the lowest value of λ considered tends towards the score in the limit $\lambda \to 0$. Similarly for the highest value of λ, the $CVD1_\lambda$ score tends to the score in the limit $\lambda \to \infty$, suggesting that there is unlikely to be a minimum outside the range of λ considered.

There is a marked difference in the magnitude of the parameters λ(CVD2) estimated by the cross-validation (second difference) method. This is particularly true for variable z1; for all data sets the selected parameter implies a high level of smoothing. The value of the $CVD2_\lambda$ score evaluated at $\lambda = 0.5$ is almost equal to the limiting

score as $\lambda \to 0$. Similarly the score $CVD2_\lambda$ evaluated at the highest λ considered is almost equal to the limiting score for $\lambda \to \infty$, implying the minimum of the $CVD2_\lambda$ function is in the limit as $\lambda \to \infty$.

For variable z2 the parameters estimated are less consistent across images. For both et1 and et2, $CVD2_\lambda$ is a monotonically increasing function, so a value of $\lambda < 0.5$ is indicated. The $CVD2_\lambda$ function for ft1 does have a minimum, indicating a parameter of $\lambda = 0.72$. For these three data sets, $CVD2_\lambda$ evaluated at the lowest and highest values of λ considered are tending towards the limiting scores, in the limit as $\lambda \to 0$ and $\lambda \to \infty$ respectively. In contrast, the $CVD2_\lambda$ function for ft2 follows the same pattern as for z1, that is a monotonically decreasing function with the minimum in the limit as $\lambda \to \infty$.

The CVKAY(λ) score function was outlined at the end of Section 4.4.3. The function was evaluated for both variables z1 and z2, but for subject et1 only. In each case a parameter $\lambda > 20.0$ was indicated. The shape of the score function was very similar to that for $CVD2_\lambda$, with the score still decreasing at the highest value of λ at which it was evaluated. Since this method did not appear to select parameters which were satisfactory for the MRI, the $CVKAY_\lambda$ score was not evaluated for any further data sets.

4.7 Interpretation and discussion

4.7.1 Interpretation in terms of the MRI

For smoothing the data, the parameters are very similar and for all data sets parameters $\lambda < 0.5$ would be estimated. However the kernel method of estimating the smoothed intensity (Section 3.4) breaks down for such values of λ which are less than half a pixel width. There is no longer any smoothing over adjacent data sites and rather than calculating a weighted average, intensity is estimated simply as the observed value at each data point. If such small amounts of smoothing were appropriate then the kernel method need not be used and intensity can be estimated by the limiting case of the CV

score (Section 4.3.5). In the limit as $\lambda \to 0$, intensity at each data point is estimated as the average of its four nearest neighbours. This requires much less computation than the kernel method. However it is not possible to estimate intensity at positions between data sites or to evaluate the Laplacian of the smoothed intensity and so is not suitable for identifying edge points located in continuous space.

Although the parameters selected are very small, this can be explained by considering the characteristics of the MRI. All the methods aim to estimate a value of λ to locally smooth out 'small' changes between neighbouring pixels. Since the current images are already reasonably smooth, in the sense that the observed intensities are very similar within a localised neighbourhood, it is not necessary to further smooth by calculating weighted averages over large neighbourhoods. The smallest neighbourhood over which data can be smoothed is the four horizontal and vertical neighbours of a point, equivalent to the limiting case ($\lambda \to 0$) of the CV score as discussed above.

The parameters selected by CV for first differences are consistent across data sets for both z1 and z2. Although there is a minimum of the $CVD1_\lambda$ function for z1, the parameters selected do still only indicate small amounts of smoothing. The values selected for z2 would not be used in practice, as discussed previously. Such parameter values can be interpreted by looking at the distribution of first differences of the data. For each data set, the lower and upper quartiles of the distribution of the first differences are of the order of -4 and 4 respectively, indicating that even pixels at a distance one apart are quite similiar and little smoothing is required.

The results for CV for second differences for all data sets for z1 and data set ft2 for z2 are in complete contrast. One explanation is that because neighbouring pixel intensities are so similar, the second differences calculated between such values are very small. The distribution of the second differences is therefore dominated by the random noise, so the CV is trying to smooth out the error, rather

than smoothing the second differences of the data. The 'best' solution in this case is to average over the whole image, hence the high values of λ indicated. This hypothesis can be given some justification by looking at the distribution of the second differences. The lower and upper quartiles are typically of the order of -7 and 7 respectively and of a similar magnitude to the error variance estimates in Table 4.1. There are still a number of 'large' second differences, which could be attributed to cases where the second difference was calculated over an edge in the image.

4.7.2 Smoothing parameter selected for the MRI

The parameters selected by the methods discussed in this chapter can be given a reasonable interpretation in terms of the MRI, although in practice none of the values in Table 4.2 are used in this thesis. For any smoothing problem it is important to relate the choice of smoothing parameter λ to the aims of the project, rather than immediately adopting a parameter automatically estimated using one of the data-based methods. In Chapter 3 the aim is to locate the edges between the main (air, fat, muscle, bone) tissue regions. In this context, the appropriate value of λ optimises the subset of edges identified: ideally only these main edges should be identified. High positional accuracy of individual edges is of less importance, since further approximations to the 'true' edge position will occur when edges are modelled as closed curves in Chapter 5.

Determining the correct amount of smoothing conditional on the subset of edges to be identified has been discussed in the computer vision literature. There are different types of intensity change in an image: for example local detail between individual muscles contrasted with edges between tissue regions where the changes are more gradual and spread over a wider neighbourhood. To identify these two distinct types of edges two different values of λ would be required.

In this thesis the parameter λ is choson subjectively. Intensity is estimated by kernel regression (Section 3.4) and edge points

identified as zero-crossings of the Laplacian of the smoothed image. The data are smoothed using several values of λ and the value which results in the 'best' subset of edges is assessed visually. Here 'best' is interpreted as the value such that only the main tissue edges are identified, but without greatly increasing the bias when locating edges. For subject et1, for both variables z1 and z2, sets of edge points identified for different values of λ are plotted in Figures 3.2 and 3.5. These figures illustrate why the small values of λ selected by the data-based methods are not suitable. Although all the main edges are identified, even with $\lambda = 1.0$ unnecessary local detail is also apparent and the identification of sets of points lying on individual edges would be very difficult. As λ is increased the local detail is smoothed out so only the main edges are identified, though this improvement in the number of edges identified is at the expense of an increase in bias in location of individual edges. The bias is particularly noticeable around the bone region since the number of bone pixels is small relative to the number of pixels over which the data are smoothed.

The accuracy of locating the main edges could be improved by using a multi-scale approach. The data are smoothed with a 'large' value of λ (= 3.0, say) and the main edges identified and marked in some way. The data are then smoothed with progressively smaller values of λ but only the set of edge points identified at higher values of λ are retained. In this way the bias in locating the main edges is reduced. Such an approach is not used in this thesis. In addition to the method being computationally very expensive, any bias in edge location will be confounded with the approximations incurred when fitting a closed curve to the data in Chapter 5 and which makes precise locating of edges unnecessary.

For subject et1, a value of $\lambda = 2.0$ is chosen. It must be stressed that although this value is appropriate for this data set, the value will vary between subjects and a similar procedure of smoothing the data with several values of λ will be required for each data set.

4.7.3 Summary

This concludes the first stage of this project so it is appropriate to review the main points covered in the first four chapters and highlight any progress made. The aim of this thesis is to identify the location of edges between tissues. The method used in clinical analysis of the MRI requires an operator to view the digitised image on a computer screen and uses a hand-held mouse to outline a tissue edge as viewed on the screen.

Using the kernel method described in Chapter 3, intensity can be estimated at any point in continuous space. By estimating the underlying continuous intensity function in this way edge points are located to sub-pixel accuracy as zero-crossings of the Laplacian of the smoothed image. The accuracy with which edges are located is dependent on the degree of smoothing. In view of the second stage of this project, provided the edges of interest are identified, the precise value of the smoothing parameter is not so critical. Approximations are made when the edge points are fitted as a closed curve, so it is unnecessary to locate edges with great accuracy at the identification stage.

Since the 'automatic' data-based methods discussed in this chapter do not appear to estimate a suitable smoothing parameter for the MRI, a 'semi-automatic' method is proposed. A computer program would be run to locate the edges in the smoothed image and the points identified as lying on an edge plotted out. This would be repeated several times, smoothing the image with different values of λ. The operator would then assess visually the value of λ giving the best representation of edges of interest; this would be the value of λ chosen as the smoothing parameter for that particular image. The image is smoothed using the chosen value of λ and edge points identified. Sets of points lying on individual edges are extracted in an ordered sequence and used as the input data for the fitting of a closed curve to describe the edge, discussed in Chapter 5.

5 FITTING FOURIER DESCRIPTORS TO CLOSED CURVES

5.1 Introduction

The data to be discussed in the second stage of this thesis takes the form of an ordered set of n points lying on a closed curve in continuous two-dimensional space. In Chapter 5 methods will be introduced in terms of this general definition of the data before investigating their usefulness for analysing the images. For the images, the data consists of an ordered set of coordinates $\{(x_i,y_i) : i = 1,2,...,n\}$ lying on the edge of a tissue region. The points are identified as zero-crossings of the Laplacian of Gaussian smoothed image (Section 3.5). Each set of points is identified by sequentially tracking the zero-crossings sampled from a single edge and which form an n-polygon approximating the underlying zero-crossing contour. The method used to fit the curve must also be valid if the ordered set of points is obtained by an alternative method, as now outlined but not considered further until Chapter 6. Rather than smoothing the image the data are thresholded to produce a discrete binary image. Points lying on each pixellated edge are tracked sequentially to obtain separate sets of coordinates for each edge.

The aim is to model, as a closed curve, the data sampled from individual tissue edges. This is a realistic representation for edges in the image which exist in continuous two-dimensional space. It may also reduce the number of parameters required to describe the edge, rather than specifying the 2n coordinates $\{(x_i,y_i) : i = 1,2,...,n\}$. A Cartesian model allows errors in the x and y directions to be modelled separately. Independently minimising the two errors implies that the overall error in position is minimised. For this reason Cartesian coordinates are preferred to a polar coordinate system, used for example by Berman and Culpin (1986), and which required choice of an arbitrary centre from which to define the polar system.

There are several methods available for fitting a closed curve to a set of points. Sriraman et al (1989) fitted a polygon, that is a piecewise linear curve joining points. Aiming to obtain a smooth

curve, Pham (1989) fitted splines through the set of points. A stochastic model was suggested by Kashyap and Chellappa (1981). They approximated the edge by a series of straight line segments between points obtained in an ordered sequence. Coordinates x_i and y_i were modelled independently as circular autoregressive processes.

Since the data lie on a closed curve, this suggests representation by a periodic function. The fitting of a Fourier series (FS) to a set of edge points was considered by Granlund (1972) and Zahn and Roskies (1972), and used even if the edge was incomplete (Lin and Chellappa, 1987). In this thesis edges will be described in terms of their Fourier descriptors (FD), that is the coefficients in the FS, a model for which is defined in Section 5.2. The edges can be described to arbitrary accuracy if sufficient descriptors are included in the model. Often a reasonable fit is possible if only the first few descriptors are included. One method of determining a suitable truncation point of the FS is discussed in Section 5.2. Geometric properties (such as area, perimeter) of the region enclosed by an edge can be computed from the FD. The relative accuracy of the area or perimeter values calculated with increasing numbers of coefficients can alternatively be used as a criterion for truncation. This motivates the work of Chapter 6. Methods of estimating the area and perimeter length are discussed with the aim of identifying the 'optimum' number of FD to be included in the series.

Points in the ordered sequence must be labelled consecutively. Several labelling variables are discussed in Section 5.3. An extension of one of the labelling variables (the 'censored index number' of Section 5.3.3) is analysed theoretically in Chapter 7. By omitting some of the edge points there may be a reduction in the variability in the lengths (defined as the Euclidean distance) of the 'intervals' calculated between each pair of points which remain. Hence fewer descriptors will be needed to produce a good fit.

In addition to the area and perimeter of a region enclosed by an edge, other geometric properties such as rotational symmetry can be derived from the FD. In Section 5.4, FS are fit to an ellipse so such properties can be assessed for a shape with known geometry. The

labelling variables introduced in Section 5.3 are compared in Section 5.5, for fitting a curve to a set of image edge points.

5.2 Fourier descriptors

5.2.1 Introduction

Cosgriff (1960) suggested the fitting of FD to describe closed curves. Several definitions of the FD have been given, depending on how the author parameterised the points to obtain a labelling variable, z_i say, for consecutive points. Granlund (1972) labelled points in terms of cumulative distance round the edge, where distance was measured from an arbitrary start point. Since the FD were not unique but dependent on the start point, functions of the FD were defined which could distinguish between true shape properties and those which were dependent on translation, rotation and scaling. Granlund used the FD to indicate the degree of symmetry of a curve and gave conditions for rotational symmetry.

The cumulative distance labelling of Granlund (1972) was also used by Lin and Chellappa (1987). They considered a more difficult problem, that of classifying shapes from edges with missing segments. The FD were estimated for the unknown complete shape, conditional on the number of missing points being unknown. The FD for the partial edge were calculated numerically based on an arbitrarily truncated FS. The accuracy of the estimated FD was improved by imposing additional constraints, such as the compactness, defined as $(\text{perimeter})^2/\text{area}$.

Zahn and Roskies (1972) fitted FD in terms of the cumulative angular function, defined as the net amount of angular bend between an (arbitrary) start point and the current point. The angular function was expressed in polar form, so the FD represented harmonic amplitude and phase angle. The harmonic amplitudes and functions of some pairs of phase angles for this form of FD were invariant to translations, rotations and scaling. Conditions on the FD for rotational and axial symmetry were given. Measures were derived based

on a truncated FS to assess how close the curves were to being symmetric and how similar they were. Synthetic curves were also generated from the FD of a truncated FS and conditions were given for a curve to be closed.

The labelling of (consecutive) edge points used in this thesis is now defined and the notation required to define the FS model introduced.

5.2.2 Definitions and notation

The data consist of an ordered set of n points $\{(x_i,y_i) : i = 1,2,\ldots,n\}$, where (x_i,y_i) denotes the location of the point in continuous space and the subscript i gives the relative position in the ordered sequence of points on the curve. Since the curve is closed, $(x_{n+1},y_{n+1}) = (x_1,y_1)$. Cartesian coordinates x_i and y_i are fitted separately. Only the fitting of a FS to the x_i coordinate will be described in detail, but results for the y_i coordinate follow similarly.

Let z_i denote a general labelling variable for consecutive locations in the ordered sequence of points on the curve. The general variable may be defined as $z_i = i$, representing the 'index number', that is the relative position in the ordered sequence. A variant of this, allowing deletion of points which are 'close' together is the 'censored index number'. The general z_i may alternatively be defined as $z_i = d_i$, representing the 'cumulative (Euclidean) distance' from an arbitrary 'start-point' to the ith point, measured with respect to the n-polygon approximating the curve. A variant, where new points are estimated at equal distances measured round the n-polygon, is the 'interpolated distance' labelling. The 'last' nth point on the curve is labelled as z_n, where $z_n = n$ for index number and $z_n = d_n$ for distance. The four labelling variables are considered in more detail in Section 5.3.

Labelling points in terms of z_i, coordinates x_i and y_i are represented in terms of their respective FS as:

$$x_i = a_o + \sum_{r=1}^{R} \left\{ a_r \cos\left[\frac{r2\pi z_i}{z_n}\right] + b_r \sin\left[\frac{r2\pi z_i}{z_n}\right] \right\},$$

$$y_i = c_o + \sum_{r=1}^{R} \left\{ c_r \cos\left[\frac{r2\pi z_i}{z_n}\right] + d_r \sin\left[\frac{r2\pi z_i}{z_n}\right] \right\},$$

for $\{i = 1,2,\ldots,n\}$. The above expressions can be simplified by normalising the range to $(0,2\pi]$. Define $\{\theta_i = 2\pi z_i/z_n,\ \theta_i \in (0,2\pi],\ i = 1,2,\ldots,n\}$. Then coordinates (x_i,y_i) for $\{i = 1,2,\ldots,n\}$ are given by:

$$x_i = a_o + \sum_{r=1}^{R} \{a_r \cos(r\theta_i) + b_r \sin(r\theta_i)\},$$

$$y_i = c_o + \sum_{r=1}^{R} \{c_r \cos(r\theta_i) + d_r \sin(r\theta_i)\}, \qquad (5.1)$$

The Fourier descriptors are the vector coefficients $\{a_0,a_r,b_r : r = 1,2,\ldots,R\}$ and $\{c_0,c_r,d_r : r = 1,2,\ldots,R\}$, for x_i and y_i respectively. With n points on the curve, then R, the maximum number of terms which can be fitted, is such that $R = (n-1)/2$, if n is odd, and if n is even then $R = n/2$, with coefficients $b_{n/2} = 0$ and $d_{n/2} = 0$.

The data are fully described by their FS (5.1), but the full series, fitting all harmonics up to the R^{th}, contains as many terms as the original data set. With a Fourier representation of the curve a reduction in the number of model parameters is possible, by using a truncated series of the form:

$$\hat{x}_i(r^*) = \hat{a}_o + \sum_{r=1}^{r^*} \{\hat{a}_r \cos(r\theta_i) + \hat{b}_r \sin(r\theta_i)\},$$

for $\{i = 1,2,\ldots,n\}$ and $\theta_i \in (0,2\pi]$. The number of harmonics to be

fitted is $r^* \leqslant R$. One method of determining the appropriate number of terms, r^*, is discussed at the end of Section 5.3.2. Including fewer terms in the series has the effect of providing a smoother curve. The curve fitted using a truncated series is only an approximation to the true edge. The approximation is usually sufficiently accurate to allow estimation of geometric properties which will be discussed in Chapter 6. The estimation of the area and perimeter length of a region are discussed in Chapter 6 and an assessment made of the values estimated from a truncated series.

5.2.3 Fitting the Fourier descriptors

Least squares (LS) regression can be used to calculate the x coordinate coefficients $\{\hat{a}_0, \hat{a}_r, \hat{b}_r : r = 1,2\ldots,r^*\}$ which minimise the residual sum of squares for fitting r^* harmonics:

$$\Omega(r^*) = \sum_{i=1}^{n} \{x_i - \hat{x}_i(r^*)\}^2,$$

for $\{r^* = 1,2,\ldots,R\}$. Least squares regression can still be used to estimate coefficients even if points are unequally spaced on the labelling variable which means that successive harmonics are not orthogonal. Often coefficients are re-estimated after progressively increasing the number of harmonics r^* ($\leqslant R$) and in theory the series would be truncated at the highest significant term.

An alternative and computationally more efficient method of estimating the FD is to use the discrete Fourier transform (DFT). Unless otherwise stated, it is assumed that this is the method of estimation used in this thesis. If the data are equally spaced on the labelling variable, then using the DFT the x_i coordinate coefficients for $\{r = 0,1,2,\ldots,R\}$ are estimated as:

$$\hat{a}_r = \frac{2}{n} \sum_{i=1}^{n} x_i \cos(r\theta_i), \qquad \hat{b}_r = \frac{2}{n} \sum_{i=1}^{n} x_i \sin(r\theta_i),$$

and similarly for the coefficients for the y coordinate:

$$\hat{c}_r = \frac{2}{n}\sum_{i=1}^{n} y_i\cos(r\theta_i), \qquad \hat{d}_r = \frac{2}{n}\sum_{i=1}^{n} y_i\sin(r\theta_i), \qquad (5.2)$$

Fourier coefficients (5.2) are computed using the NAG Pascal procedure C06EAC, initialised by procedure A00ABC. This procedure uses the Fast Fourier Transform (FFT). Although computationally more efficient, the FFT does impose constraints on the value of n allowed, since no prime factor of n may exceed 19. In practice, if n is unsuitable for use of the FFT, then the number of points is adjusted to a value, n^* say, close to n but for which the FFT can be calculated. If a FS is truncated at $r = r^*$ harmonics, with FD $\{\hat{a}_o,\hat{c}_o,\hat{a}_r,\hat{b}_r,\hat{c}_r,\hat{d}_r : r = 1,2,\ldots,r^*\}$, then using the inverse FFT fitted coordinates $\{\hat{x}_i(r^*),\hat{y}_i(r^*) : i = 1,2,\ldots,n\}$ can be calculated as:

$$\hat{x}_i(r^*) = \hat{a}_o + \sum_{r=1}^{r^*} \{\hat{a}_r\cos(r\theta_i) + \hat{b}_r\sin(r\theta_i)\},$$

$$\hat{y}_i(r^*) = \hat{c}_o + \sum_{r=1}^{r^*} \{\hat{c}_r\cos(r\theta_i) + \hat{d}_r\sin(r\theta_i)\}.$$

To determine a suitable number of harmonics r^* at which the series should be truncated, the significance of individual terms in the FS can be tested by stepwise regression. Expressing the sums of squares in terms of the FD, the total sum of squares (TSS) is given as:

$$TSS = \sum_{i=1}^{n}\left[\hat{a}_o + \sum_{r=1}^{R}\{\hat{a}_r\cos(r\theta_i) + \hat{b}_r\sin(r\theta_i)\}\right]^2 = n\hat{a}_o^2 + \frac{n}{2}\sum_{r=1}^{R}\{\hat{a}_r^2 + \hat{b}_r^2\}.$$

For fitting r^* harmonics $\{r^* = 1,2,\ldots,R\}$, the TSS can be partitioned into the regression sum of squares (REG SS(r^*)) and the residual sum of squares (RESID SS(r^*)) as:

$$TSS = \left[n\hat{a}_o^2 + \frac{n}{2}\sum_{r=1}^{r^*}\{\hat{a}_r^2 + \hat{b}_r^2\}\right] + \left[\frac{n}{2}\sum_{r=r^*+1}^{n}\{\hat{a}_r^2 + \hat{b}_r^2\}\right]. \qquad (5.3)$$

Since successive terms are orthogonal, as r^* is increased and higher harmonics included in the fitted model, the significance of r^*th harmonic $\{\cos(r^*\theta), \sin(r^*\theta)\}$ can be tested. The REG SS(r^*) can be calculated as in (5.3), based on $2r^*$ degrees of freedom (DF). Calculating the difference between the REG SS(r^*) and the analagous REG SS(r^*-1) gives the difference due to fitting the r^*th harmonic, for which the significance is to be tested. The difference, on two DF, is compared against RESID SS(r^*), based on $(n-1)-2r^*$ DF. The significance of the r^*th harmonic is tested by calculating the F-ratio, F(r^*) as:

$$F(r^*) = \frac{\{\text{REG SS}(r^*)-\text{REG SS}(r^*-1)\} \;/\; 2.}{\text{RESID SS}(r^*) \;/\; \{(n-1)-2r^*\}}$$

The calculated F-ratio F(r^*) is compared against tabulated values of the F-distribution on $\{2, (n-1)-2r^*\}$ DF. In this thesis, the significance probability is obtained using the NAG Pascal function G01BBC. F-ratios can be calculated for any value of $\{r = 1,2,\ldots,R-1\}$; the R^{th} $(=n/2)$ term cannot be tested since there are no residual DF remaining.

As further harmonics are included in the fitted model each additional harmonic can be tested in this way using a predetermined level of significance. In theory, the series would be truncated at the highest significant term. In practice, when a FS is fit to a set of edge points, some but not all, harmonics even up to the $(R-1)^{st}$ are significant. Thus, although this is perhaps the most obvious criterion for determining the number of terms r^* to be included in the series, it is not conclusive. A suitable value of r^* could be determined by comparing the relative accuracy of perimeter lengths calculated including an increasing number of terms in the truncated FS fitted to an edge, as will be discussed in Chapter 6.

The relative fit of a FS truncated at a prespecified number of harmonics could also be used to aid the choice of variable for labelling successive points. One criterion is S_r, the residual standard deviation (SD) after fitting r harmonics:

$$S_r = [\text{RESID SS}(r) \;/\; \{(n-1)-2r\}]^{1/2}, \qquad (5.4)$$

for {r = 1,2,...,R-1}. The RESID SS(r) is calculated as in (5.3). The optimum labelling could be defined as that with minimum residual standard deviation after fitting the prespecified number of harmonics. Fourier series will be fit to points labelled in terms of the four variables as introduced in Section 5.2.2 and discussed further in Section 5.3. By plotting S_r against the number of harmonics {r = 1,2,...,R-1}, the different labelling options can be compared, as is discussed in Section 5.5 for fitting a FS to a set of edge points identified from one of the images.

5.3 Choice of labelling variable

5.3.1 Introduction

The set of points $\{(x_i,y_i) : i = 1,2,...,n\}$ are the coordinates of an n-polygon sampled from a closed curve in continuous two-dimensional space. In Section 5.2.1 points were labelled in terms of the general variable z_i, and four possible definitions of z_i were introduced. In this section these four labelling variables are discussed in more detail. As will be discussed in Section 5.5, the residual standard deviation S_r (5.4) is used as a criterion with which to compare the four labelling variables. The S_r criterion provides a quantitative assessment of the fit of the model with respect to a specific labelling variable. Three further factors should be considered before a recommendation is made for the 'best' labelling variable. First, if points are equally spaced on the labelling variable, then the FD can be calculated efficiently using the FFT. Second, some additional computation is required to derive variables other than the index number of Section 5.3.2. Third, as further approximations are made to the sampled n-polygon, the bias in the fitted curve is increased.

The choice of labelling variable also has implications for geometric properties of the fitted curve, since such properties are functions of the FD. In Section 5.4, the simple test case of an ellipse is discussed, in order to study the relation between labelling variable and properties such as rotational symmetry.

5.3.2 The index number labelling variable

The index number corresponds to the relative position of the current point in the ordered sequence of points round the curve. The general variable z_i is defined such that $z_i = i$ and $z_n = n$. Normalising the range to $(0,2\pi]$ gives angular increments $\{p_i = 2\pi i/n, p_i \in (0,2\pi], \text{ for } i = 1,2,\ldots,n\}$. The angular increments p_i calculated between consecutive points are equal. One advantage of this variable is the constant integer increments between labels $\{i = 1,2,\ldots,n\}$. This results in equal spacing between consecutively labelled points and so the Fourier coefficients can be computed using the FFT. This holds provided that the value of n is suitable for the FFT algorithm to be used, though in practice the value of n (and hence the coordinates of points on the curve) can be modified slightly. Any modification of the value of n has only negligible effect on the curve fitted.

Using the index number labelling, the points will not be equally spaced round the curve, in the sense that the 'intervals', calculated as the Euclidean distance between consecutively labelled points, will not be equal. A variant of the index number aims to reduce the variation in the size of these intervals. This is the censored index number, discussed in the next section.

5.3.3 The censored index number labelling variable

For points labelled in terms of the index number variable, the length of the interval between each pair of consecutive points is calculated as the Euclidean distance between the current and preceeding point. Since the curve is closed, the interval length to the first labelled point is calculated as the distance between (x_n,y_n) and (x_1,y_1). If the length of the interval between two points is less than a prespecified censoring threshold distance, then this interval is censored and subsequent points on the curve are reindexed. By censoring points which are 'close' together local variation is reduced and it may be possible to reduce the number of FD required to produce a good fit, as measured by S_r (5.4).

The censoring threshold, t say, is an important parameter. In this chapter values in the range [0,1] are considered, where '1' is the width of a unit pixel. Intuitively, the 'best' threshold is expected to be that which minimises the range of interval lengths, so individual interval lengths are as similar as possible. The theoretical analysis of Chapter 7 is motivated by the problem of how to define the 'best' threshold. By simulating a random smooth path of infinite length and looking at the intersections of the path with the lattice, the distribution is derived of the interval lengths between consecutive intersections. Similarly, a distribution of (censored) interval lengths can be derived after censoring points at different threshold values. Since the distribution is a function of the threshold t, it is possible to derive analytically the value of t which minimises the range of interval lengths. This value is then used as the 'best' threshold when censoring points on a closed curve, prior to the fitting of FD.

The number of points remaining after censoring will depend on the threshold used, and may need to be modified before the FFT algorithm can be used. By perturbing the threshold by a small increment around the nominal value t derived as the 'best' censoring threshold (Chapter 7), it is possible to either increase or decrease the value of n by one or two points, where here n is the number of points remaining after censoring. This rather ad-hoc method would seem sufficiently accurate in view of the approximation which will be made anyway when fitting a truncated series. However it may initially require considerable recalculation of edge point coordinates and interval lengths before a suitable value of n is found.

5.3.4 The cumulative distance labelling variable

Points may alternatively be labelled in terms of cumulative distance, defined as the total distance travelled round the curve from the zeroth point (x_0,y_0). Since the curve is closed $(x_0,y_0) = (x_n,y_n)$. For distance the general variable z_i is defined such that $z_i = d_i$ and $z_n = d_n$. A point is chosen as the zeroth point such that if points are labelled by distance d_i, then the subscript i

is the same as the index i if points are labelled in terms of index number. Normalising the range to $(0,2\pi]$ gives angular increments $\{q_i = 2\pi d_i/d_n,\ q_i \in (0,2\pi],\ \text{for } i = 1,2,\ldots,n\}$. Labelling in terms of distance, the angular increments between consecutive points are unequal. In Section 5.4 the affect of such unequal angular increments, when fitting a truncated series to a test ellipse is compared to the equal angular increments of points labelled in terms of index number.

The length l_i of the line segment joining points (x_{i-1},y_{i-1}) and (x_i,y_i), is defined as the Euclidean distance between these two points. Therefore $l_1 = \{(x_1 - x_n)^2 + (y_1 - y_n)^2\}^{1/2}$ and $l_i = \{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2\}^{1/2}$ for $\{i = 2,3,\ldots,n\}$. The cumulative distance d_i is defined as the total distance to the i^{th} point, so that if $d_0 = 0$ then:

$$d_i = \sum_{k=1}^{i} l_k \quad \text{for} \quad i = 1,2,\ldots,n \quad \text{and} \quad D = \sum_{i=1}^{n} l_i = d_n.$$

Thus $D = d_n$ is the total cumulative distance round the n-polygon which approximates the true curve.

The cumulative distance labelling allows for unequal change in x and y coordinates between consecutive points. This is in contrast to the index number labelling where all such changes between points are equal. However a major disadvantage of the cumulative distance labelling is that points are no longer equally spaced in terms of the cumulative distance variable. This means successive terms in the FS are not orthogonal and so must be fit using LS regression (Section 5.2.3). Since it is computationally more convenient to use the FFT to estimate the FD, a variant of this labelling variable, where distances between points are equalised, is proposed in the next section.

5.3.5 The interpolated equal distance labelling variable

For points labelled in terms of the cumulative distance of

Section 5.3.4, a new set of points at equal distances, d_i^* say, are calculated by interpolating along the n-polygon approximating the original curve. The new points lie on the n-polygon rather than on the curve. The total distance round the polygon is D, and linear interpolation is used to estimate the coordinates of n points at equal distances, $d_i^* = D/n$. The 'start' point of the interpolated polygonal curve is again the zeroth point, $(x_o, y_o) = (x_n, y_n)$.

By design, the distances between consecutive points on the interpolated polygon are equal, so FD can be estimated using the FFT. If the number of points n is unsuitable for use of the FFT, then with this labelling the number and position of points on the interpolated curve is easily modified. The number of points, n^* say, closest to n, is found for which the FFT can be calculated. Rather than at distances D/n, new points are interpolated at equal distances D/n^* round the n-polygon approximating the sampled curve.

By interpolating with respect to the n-polygon, a form of smoothing is imposed on the data, which reduces local variation and possibly improves the global fit. To be balanced against the reduction in variation is an increase in bias in the location of points as further approximations are made to the 'true' curve. Points no longer lie on the 'true' curve but on its polygonal approximation. Thus it is possible that labelling in terms of the interpolated equal distance variable, rather than the unequal cumulative distance of Section 5.3.4 may result in a poorer fitting model. This is illustrated for the simple case of an ellipse in Section 5.4.

5.4 General properties for fitting Fourier descriptors

5.4.1 Introduction

In early references to the fitting of FD (Granlund, 1972, Zahn and Roskies, 1972) geometric properties such as symmetry were attributed to the fitted curves. Such properties held for a given subset of significant coefficients and were dependent on the variable used to label consecutive points. In this section FD are fit to an

ellipse, a test shape with known geometry. Geometric properties are related to the different labelling variables in Sections 5.4.2 and 5.4.4 and any differences between them highlighted. The unequal distance (Section 5.3.4) and equal distance (Section 5.3.5) variables are compared in Section 5.4.3.

5.4.2 Fitting a Fourier series to an ellipse

A test ellipse, centred at the origin (0,0) is defined by the equation $\{X = 5\cos\Theta,\ Y = 10\sin\Theta\}$. Coordinates of points on the ellipse are evaluated at equal angular increments, the $p_i = 2\pi i/n$ of Section 5.3.2. For this example $n = 64$ points are considered sufficient. The points are labelled in terms of unequal cumulative distance d_i and a FS is fit using LS regression (Section 5.2.1). Two aspects of the estimated FD are now discussed.

5.4.2.1 Fitting the first harmonic only

The fitted ellipse is positioned accurately, with centre coefficients estimated as $\hat{a}_0 \approx 0.0$ and $\hat{c}_0 \approx 0.0$. For the first FD, coefficients (5.2) are estimated as $\{\hat{a}_1 = 5.4,\ \hat{b}_1 = -0.7,\ \hat{c}_1 = 1.2,\ \hat{d}_1 = 9.2\}$. The FD are obviously not equal to the true parameters $\{a_1 = 5.0,\ b_1 = 0.0,\ c_1 = 10.0,\ d_1 = 0.0\}$. This shows that even when the shape is an ellipse, the true curve cannot be recovered exactly if points are labelled in terms of d_i. With this unequal distance labelling, the angular increments between consecutive locations (the q_i of Section 5.3.4) are unequal and so do not reproduce an exact sinusoid. Defining points to be at equal angular increments p_i is equivalent to labelling in terms of index number. Hence if a set of points on a curve are such that angular increments between them are equal, the index number labelling is superior to the cumulative distance variable. It is unlikely that the set of points sampled from an edge in the image would be located at equal angular increments, so in practice any benefit from labelling points in terms of the index number will not be attainable.

5.4.2.2 Fitting a truncated Fourier series

A FS with 16 harmonics is fitted to the same ellipse. As previously the centre coefficients are estimated as $\{\hat{a}_o \approx 0.0, \hat{c}_o \approx 0.0\}$. Estimated coefficients for the highter terms decrease with increasing harmonic number r, but for both coordinates x_i and y_i there is a pattern in the FD: the FD for all odd harmonics are non-zero, while for all even harmonics the FD are identically zero.

5.4.3 Comparison of the two distance labelling variables

A test ellipse centred at the origin (0,0) is defined by $\{X = 5\cos\Theta, Y = 10\sin\Theta\}$, as in Section 5.4.2. Coordinates of points on the ellipse are evaluated at 60 equal angular increments, $\{p_i = \pi i/30, : i = 1,2,...,60\}$. Using LS regression, a FS is fit to points labelled in terms of unequal cumulative distance d_i (Section 5.3.4). Recall that for the unequal distance variable points lie on the original curve. The results (from fitting in terms of d_i) are compared with those from a FS fit (using the FFT) to points labelled in terms of interpolated equal distance d_i^* (Section 5.3.5). For the equal distance variable points do not lie on the original curve but only on the approximating n-polygon.

The fit of the FS is assessed in terms of S_r (5.4), the residual standard deviation after fitting r harmonics. For fitting a FS to x_i and y_i coordinates individually, Table 5.1 summarises values of S_r for both equal and unequal distances, d_i and d_i^* respectively. The magnitude of the difference, $\nabla(S_r) = |d_i(S_r) - d_i^*(S_r)|$, is also tabulated and the difference expressed as a percentage of $d_i(S_r)$, the residual lack-of-fit for unequal distance. In general values of S_r decrease with increasing harmonic number r, but for x_i and y_i coordinates for both d_i and d_i^* there is a slight increase in the value of S_r for r = 2, compared to r = 1. This is reasonable, since for r = 1 an ellipse is fitted to the points, so the underlying shape is correct although the precise dimensions and location are not accurately estimated. Fitting a second harmonic, that is r = 2, destroys the equivalence of the basic shapes. If at least two

Table 5.1 Comparison of values of residual standard deviation S_r for fitting a Fourier series in terms of distance

Table 5.1a Fourier series fitted to x_i

H. No. r	$d_i(S_r)$	$d_i^*(S_r)$	$\nabla_i(S_r)$	$^{+}\frac{\nabla_i(S_r)}{d_i(S_r)}$ %
1	0.3419	0.3331	0.0088	2.6
2	0.3481	0.3391	0.0090	2.6
4	0.0907	0.0861	0.0046	5.1
8	0.0126	0.0119	0.0007	5.4
12	0.0025	0.0028	-0.0002	9.8
16	0.0006	0.0016	-0.0010	160.1

Table 5.1b Fourier series fitted to y_i

H. No. r	$d_i(S_r)$	$d_i^*(S_r)$	$\nabla_i(S_r)$	$^{+}\frac{\nabla_i(S_r)}{d_i(S_r)}$ %
1	0.4966	0.4793	0.0173	3.5
2	0.5055	0.4879	0.0018	0.3
4	0.1204	0.1147	0.0057	4.7
8	0.0154	0.0146	0.0008	5.1
12	0.0030	0.0037	-0.0007	23.7
16	0.0007	0.0026	-0.0019	277.3

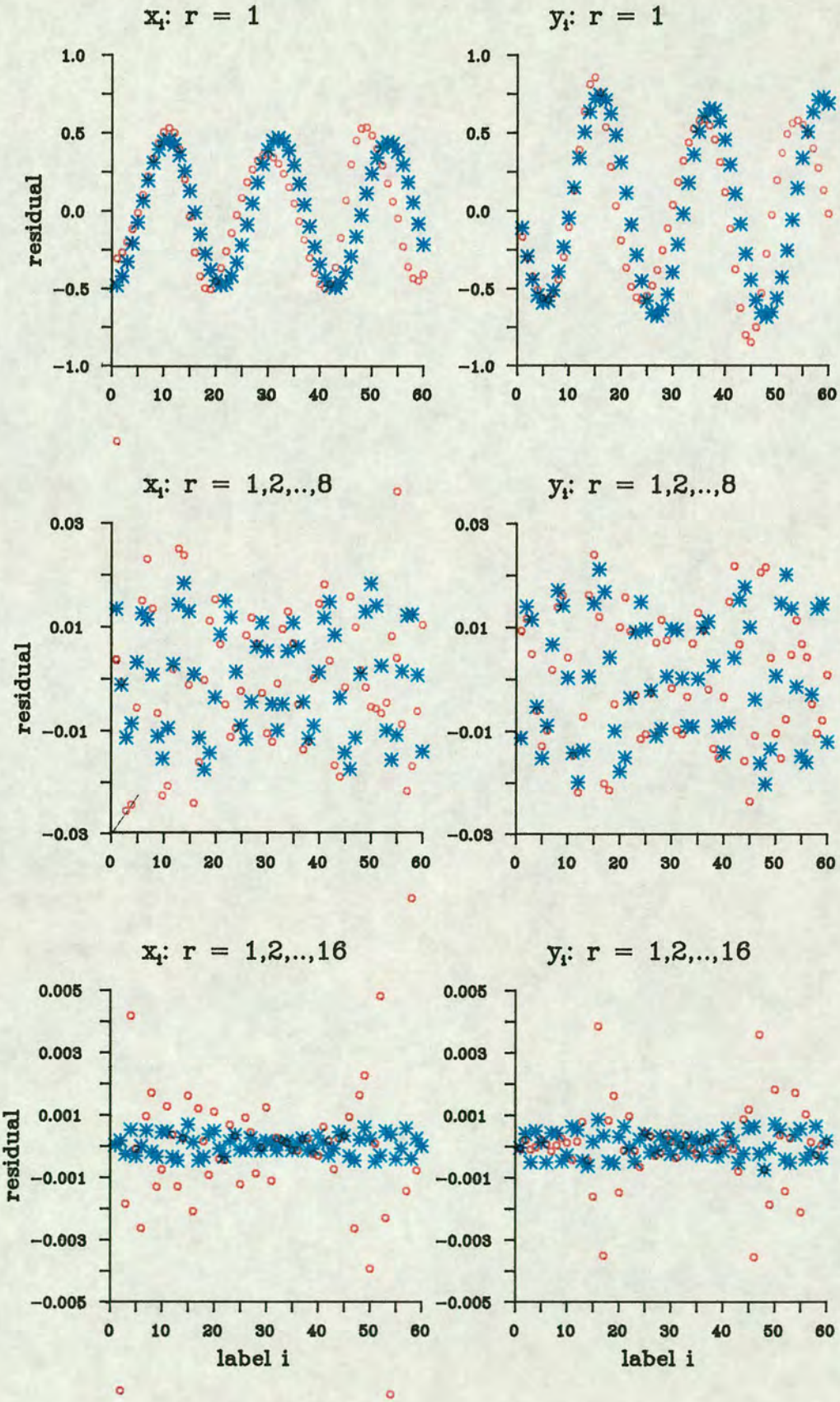
+ Note: define $\nabla_i(S_r) = | d_i(S_r) - d_i^*(S_r) |$

further harmonics are fitted ($r \geqslant 4$) then local detail is modelled, for higher r, and an increasingly more accurate estimate of the ellipse is obtained. Consequently values of S_r begin to decrease with r for $r \geqslant 4$.

Comparing the results for the two distance variables, values of S_r for fitting up to eight harmonics are very similar, with d_i^* giving a slightly better fitting model. If higher harmonics are included in the model, the fit is considerably better for points labelled in terms of d_i. Although for $r > 8$, $\nabla_i(S_r)$ is much smaller compared to that for $\{r = 1,2,\ldots,8\}$, as a percentage of $d_i(S_r)$ it is much greater. This suggests that after about eight harmonics the local smoothing of the n-polygon which occurs as a result of interpolation for equal distance becomes more apparent. The interpolated points do not lie on the ellipse and further approximation (through smoothing) of the true curve means that bias in position begins to dominate the S_r value.

For fitting a FS to the x_i coordinate, residuals are defined as $\{x_i - x_i(r)\}$, and similarly for y_i. The residual at each point $\{i = 1,2,\ldots,n\}$ after fitting $\{r = 1,8,16\}$ harmonics are plotted in Figure 5.1. Values are compared for fitting a FS to points labelled in terms of both d_i and d_i^*, plotted as '*' and ''' respectively. Although residuals decrease in magnitude as more harmonics (beyond the second) are included in the series, the difference between residuals for d_i and d_i^* does increase. Labelling points in terms of equal distance d_i^*, with 16 harmonics in the series there are a number of residuals which are 'large' compared to the majority of residuals for the model fitted in terms of the d_i^* variable. An examination of the residuals in relation to the position of the corresponding points on the ellipse indicates that 'large' residuals occur in regions of greatest curvature of the ellipse. Although points on the ellipse are defined to be at equal angular increments, it is in the greatest curvature regions where points are closest together in terms of Euclidean distance. When interpolating for equal distance, some of these points may be omitted. This effectively smooths the fitted curve and increases bias in position of the fitted point, hence the large residual at such points.

Figure 5.1: Residuals after fitting in terms of unequal (*) and equal (∘) distance



This example illustrates one case where the unequal cumulative distance labelling d_i results in a better fitting model than the interpolated equal distance d_i^*. The crucial difference between the two variables is whether points lie on the underlying curve or only on its polygonal approximation. In the latter case, the curve is smoothed as a consequence of the interpolation and the inclusion of higher terms in the series results in an overfitted model. In practice, when fitting a FS to a set of edge points sampled from an image edge, the affect of the interpolation on the n-polygon will not be so apparent. Unlike the ellipse, the underlying edge is not smooth and encloses an irregularly shaped region. Thus the overfitting observed when labelling points on an ellipse in terms of d_i should not be a problem when fitting to an edge in the image.

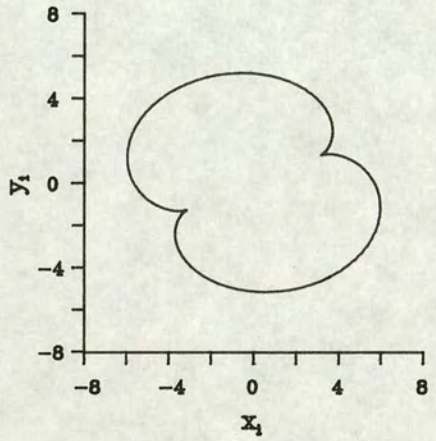
5.4.4 Fitting specific subsets of harmonics

In this section, the investigation of the properties which can be attributed to FD is approached from a different angle. A set of FD $\{a_r,b_r,c_r,d_r\}$ is prespecified and substituted into (5.1). The corresponding Cartesian coordinates (x_i,y_i) are evaluated and the curve so defined is plotted and any symmetry properties considered. For this example, 128 increments of θ are used, angular increments are equal so $\{\theta_i = \pi i/64 : i = 1,2,....128\}$. Note that specifying equal angular increments is equivalent to fitting a FS to points labelled in terms of the index number i. The FD are arbitrary, subject to $\{a_r,b_r,c_r,d_r\}$ decreasing faster than (1/r), where r is the harmonic number. This empirical constraint would appear desirable for a simple, closed curve; it may be possible to justify this mathematically though this is not considered here.

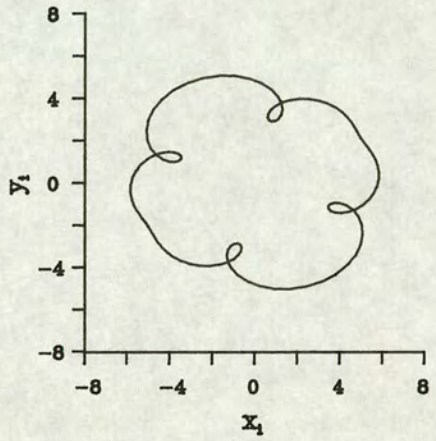
Two cases, for fitting only odd (Section 5.4.4.1) and only even harmonics (Section 5.4.4.2) are considered separately. The resultant shapes for fitting (only odd) harmonics are illustrated in Figure 5.2a (*i - iv*) and for (even only) in Figure 5.2b (*i - iv*).

Figure 5.2a: Fitting odd harmonics only

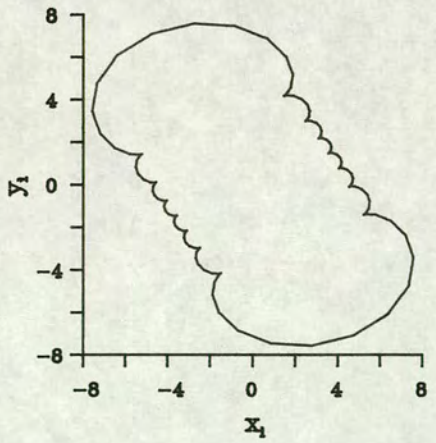
(i) r = {1,3}



(iii) r = {1,5,7}



(ii) r = {1,3,5,7,9,11,13,15}



(iv) r = {1,3,9,11}

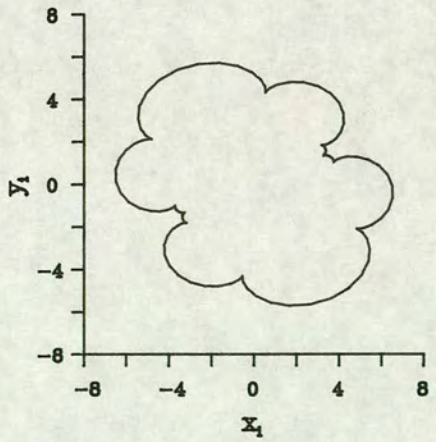
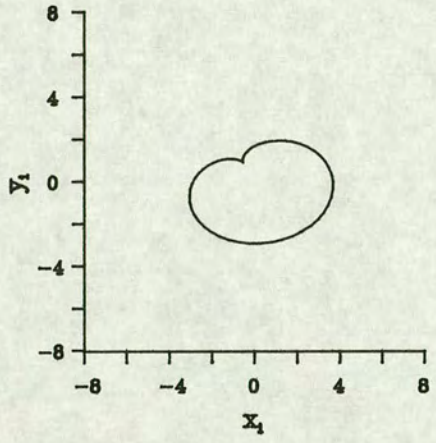
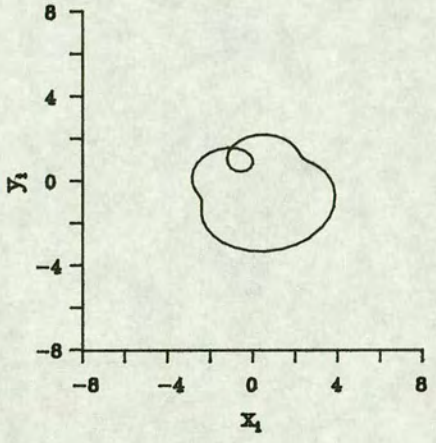


Figure 5.2b: Fitting even harmonics only

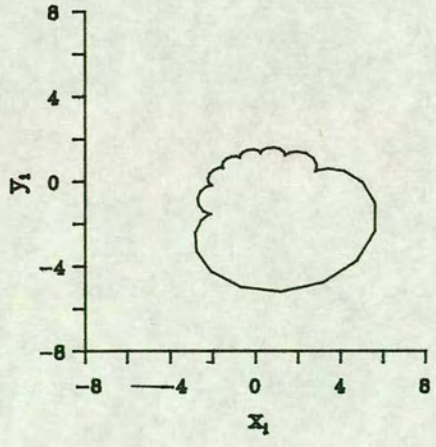
(i) r = {2,4}



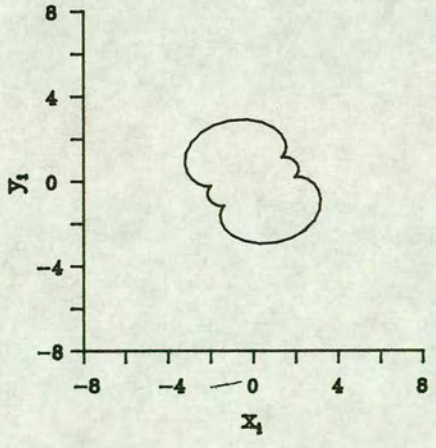
(iii) r = {2,4,8}



(ii) r = {2,4,6,8,10,12,14,16}



(iv) r = {2,6,10}



5.4.4.1 Fitting odd harmonics only

For a FS truncated at the r^*th harmonic (where r^* is assumed to be even), then if only odd terms are included in the FS, fitted coordinates $\{(x_i,y_i) : i = 1,2,\ldots,n\}$ are evaluated as:

$$\hat{x}_i(r^*) = \sum_{v=1}^{r^*/2} \{a_{2v-1}\cos([2v-1]\theta_i) + b_{2v-1}\sin([2v-1]\theta_i)\},$$

$$\hat{y}_i(r^*) = \sum_{v=1}^{r^*/2} \{c_{2v-1}\cos([2v-1]\theta_i) + d_{2v-1}\sin([2v-1]\theta_i)\}, \qquad (5.5)$$

Four subsets of FD are considered: subset *(i)* fits harmonics $\{r = 1,3\}$, *(ii)* fits $\{r = 1,3,5,7,9,11,13,15\}$, *(iii)* fits $\{r = 1,5,7\}$ and *(iv)* fits $\{r = 1,3,9,11\}$. The resultant curves are plotted in Figure 5.2a *(i - iv)* respectively, from which it is concluded that the sinusoid has rotational symmetry of degree 2 about its origin. This property still holds even if some of the first few odd harmonics are omitted from the FS, as is the case for subsets *(iii)* and *(iv)*.

This conclusion is in contrast to the results of Section 5.4.2.2, in which all odd harmonics are non-zero and illustrates how such properties are conditional on the variable with which individual points are labelled. Recall from Section 5.4.2.2 that a FS was fit to an ellipse, with points labelled in terms of unequal cumulative distance d_i. A standard property of the ellipse is that the shape has rotational symmetry of degree 2, in common with the sinusoids fitted in this section. However when fitting to the ellipse (in terms of d_i), all odd FD are required. This is in contrast to the observations of this section, that is when labelling points in terms of index number, only a subset of the odd harmonics is required.

The contrasting results can be explained in terms of the angular increments between points labelled by the two variables, in relation to the smoothly changing ellipse. For the index number variable, increments p_i are equal so only a subset of the odd harmonics is required to capture the regularity of the changes between successive points on the ellipse. In contrast, the unequal increments q_i of the

distance variable mean the regularity cannot be exploited in the same way and so all odd terms are required to describe the curve.

5.4.4.2 Fitting even harmonics only

For a FS truncated at the r*th harmonic, then if only even terms are included in the series, fitted coordinates (x_i, y_i) are evaluated as given by (5.5), now defining centre coefficients $\{\hat{a}_0 = 0.0, \hat{c}_0 = 0.0\}$ and writing the subscript for harmonic number as $r = 2v$. Four subsets of FD are considered: subset *(i)* fits harmonics $\{r = 2,4\}$, *(ii)* fits $\{r = 2,4,6,8,10,12,14,16\}$, *(iii)* fits $\{r = 2,4,8\}$ and *(iv)* fits $\{r = 2,6,10\}$. The resultant curves are plotted in Figure 5.2b *(i - iv)* respectively. In each case the sinusoid would appear to have reflection symmetry about the line Y = -X, followed by an affine transformation; the property still holds if some of the first few even harmonics are omitted from the FS, as in subsets *(iii)* and *(iv)*.

5.5 Comparison of labelling variables

5.5.1 Fitting a Fourier series to an image edge

General properties were established for the different labelling variables in Section 5.4, for fitting FS to a simple simulated shape. In this section the set $\{(x_i, y_i) : i = 1,2,\ldots,n\}$ is defined more precisely so the labelling variables can be compared for fitting a FS to real data. The set of points used in this example lie on the edge between the muscle and fat regions of the right leg (as viewed in Figure 2.1a) for subject et1, variable z1. Using a value of the smoothing parameter of $\lambda = 2.0$, edge points (as plotted in Figure 3.5) were identified as zero-crossings of the LoG smoothed image (Section 3.5). Points on the muscle/fat edge were tracked sequentially to obtain a set of n points $\{(x_i, y_i)\}$, where for this example n = 168.

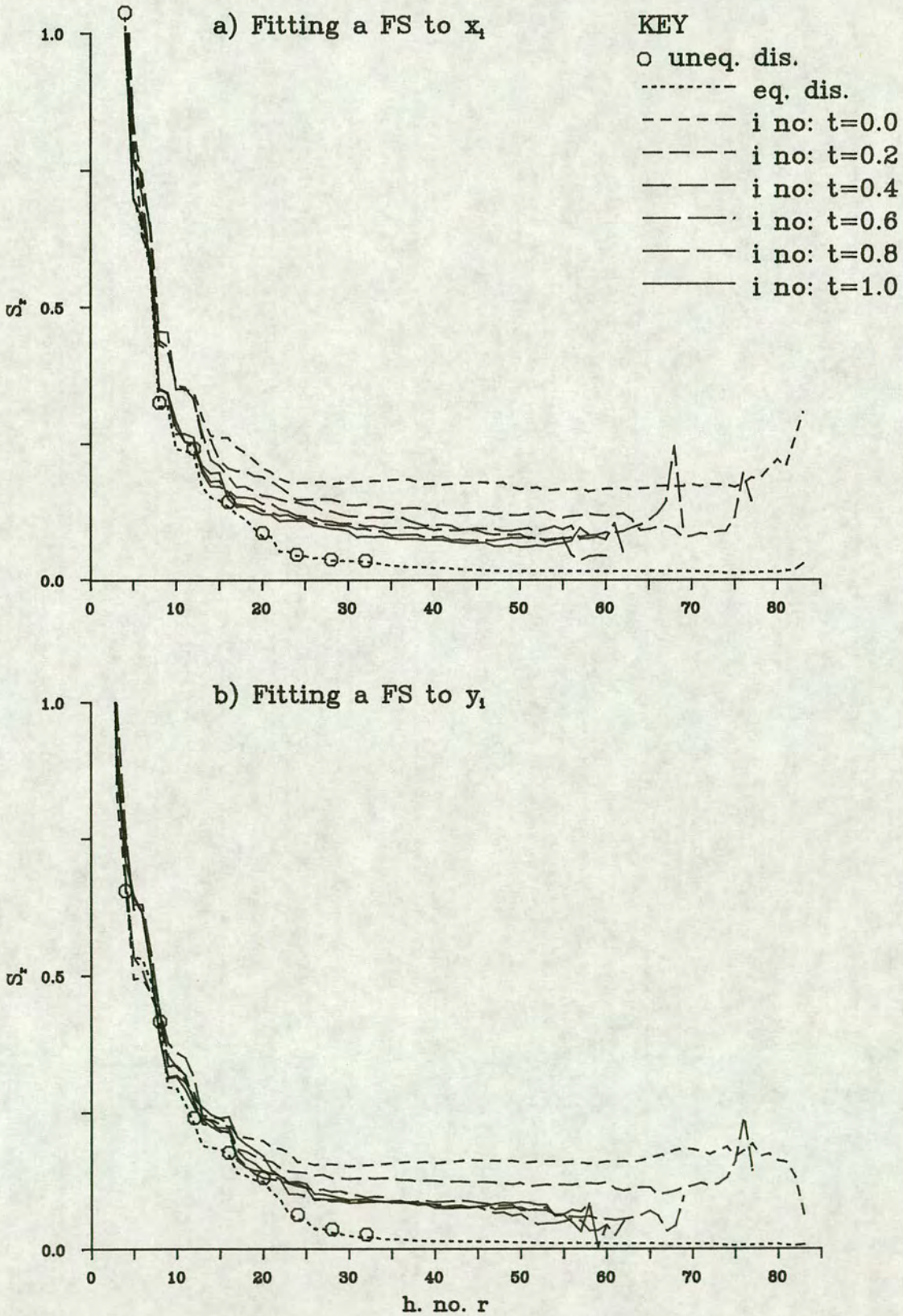
The points are labelled in terms of each of the four variables of

Section 5.3 in turn. For the index number labelling, a FS is fit to points without censoring, equivalent to setting a threshold of $t = 0.0$, and to points after censoring at thresholds of $\{t = 0.1,0.2,\ldots,1.0\}$. Details of the algorithm for censoring points and the relabelling of subsequent points will be given in Chapter 7. The FS is fit, by LS regression, to points labelled in terms of unequal cumulative distance, d_i. For the other three variables, the FS is fit using the FFT, subject to possible slight modification of the number (and hence position) of edge points for a suitable value of n. As outlined at the end of Section 5.2.3, the residual standard deviation S_r (5.4) is used as a criterion with which to compare the four labelling variables in terms of the fit of their respective FS, at $\{r = 1,2,\ldots,R\}$ harmonics.

Values of S_r are plotted against r in Figures 5.3a and 5.3b, for fitting a FS to x_i and y_i respectively. For clarity, only a subset of the results are presented for the censored index number, for thresholds $\{t = 0.0,0.2,0.4,0.6,0.8,1.0\}$. The increasing threshold is depicted by the increasing lengths of the dashes in the figure. For equal distance, values of S_r are denoted by a dotted line. Values of S_r for points labelled in terms of d_i are denoted by a circle; since for this variable the estimation of FD by LS is computationally more involved, values of S_r are not evaluated for the whole range of $\{r = 1,2,\ldots,R\}$.

When plotting the graphs, the high values of S_r are truncated for the first few harmonics so that the scale then used for the y-axis allows better discrimination between the plotted lines at higher values of r. Censoring points for the index number variable ($t > 0.0$) means there are fewer points on the curve and consequently the number of harmonics which can be fitted is less than the maximum number R for $t = 0.0$. Note that in some cases, as the cumulative number of harmonics fitted tends towards the maximum number which can be fitted, then values of S_r increase. This is reasonable, since both the RESID SS and the scaling degrees of freedom in the expression (5.4) for S_r are tending to zero, but at different rates.

Figure 5.3: Comparison of labelling variables
Subject et1, Variable z1



5.5.2 Discussion

For points labelled in terms of index number, the fit of the model at t = 0.0 is always improved upon by censoring points by raising the threshold t. This empirical observation is confirmed by the theoretical results of Chapter 7. In Chapter 7 it will be shown that the range of interval lengths between consecutive points is minimised by setting t = 1.0. Therefore it is not surprising that the best fit to the image edge for points labelled in terms of censored index number is obtained with a threshold of t = 1.0

An intuitive explanation can be given by considering which intervals are excluded when points are censored at different thresholds, and comparing the respective values of S_r for a fixed number of harmonics r. In Figure 5.3a, for example with r = 30, there is a greater difference between S_r values at 'lower' thresholds (< 0.6 say), than between S_r values at 'higher' thresholds, where there is negligible improvement if the threshold is increased further. At a threshold of t = 0.0, points are approximately equally spaced with interval lengths of approximately one unit, but there are also a number of small intervals. When raising the threshold the points are censored, between which the length of such intervals are small, and hence the lengths of intervals between the remaining points become more equal. Consequently fewer harmonics are required to describe the edge and a better fitting model, in terms of smaller S_r, is possible. Although in general S_r values decrease with increasing r, there is only negligible difference in S_r values for $t > 0.6$. This is because once the threshold has been increased to a critical level where all such 'small' interval lengths have been excluded, raising it still further makes only negligible difference to the set of points remaining and so values of S_r are more stable.

For points labelled in terms of distance, values of S_r are very similar, with the interpolated equal distance d_i^* giving a marginally better fit. Edges of the image regions are not regular so the effect of further approximation to the true edge when interpolating to equal distance should be negligible. Thus for the image data the bias, at higher values of r, from further approximation to the original curve

when interpolating to equal distance d_i^* does not dominate (and increase) S_r, as was the case for the test case of the ellipse (Section 5.4.3).

5.5.3 Recommendations for the labelling variable

The distance labelling variable is preferred, over the index number variable, even after attempting to improve the fit for the index number labelling by censoring points. Although there is little to be gained in terms of improvement of fit using the equal distance variable d_i^*, computationally variable d_i^* is preferred to the unequal distance d_i because of the ease in which the full model can be fit using the FFT. Hence the recommendation would be to fit a FS to points labelled in terms of interpolated equal distance d_i^* of Section 5.3.5.

6 USE OF FOURIER DESCRIPTORS FOR THE ESTIMATION OF PERIMETER LENGTH AND AREA

6.1 Introduction

The data considered in this chapter is an ordered set of n points $\{(x_i,y_i) : i = 1,2,...,n\}$. The points lie on a closed curve so $(x_{n+1},y_{n+1}) = (x_1,y_1)$. The fitting of a FS (5.1) to coordinates (x_i,y_i) was discussed in Chapter 5. The curve is fully described by fitting all the FD, but this requires as many parameters as describing the curve by the set of n coordinates $\{(x_i,y_i)\}$. When describing a closed curve in terms of its FS the aim is to reduce the number of model parameters by fitting only the first few FD. The problem lies in defining a suitable number of harmonics, r^*, at which to truncate the FS. One criterion, truncating the series at the highest significant term, was discussed briefly in Section 5.2.3. but did not provide a satisfactory solution.

A different approach is to consider the perimeter length and area of the region enclosed by an edge. Both properties can be calculated from the FD. After truncating the series at an increasing number of harmonics r^*, the estimated perimeter length and area can be compared against the corresponding 'true' value. The relative accuracy of the estimated value can be used as a criterion to define the optimum truncation point as the value of $r = r^*$ which results in the best estimate of the true value. In this chapter methods are considered for estimating the perimeter length and area.

The estimation of geometric properties of a region from a set of edge points has been considered by several authors. Dorst and Smeulders (1987) and Koplowitz and Bruckstein (1989) compared estimators of perimeter length which were designed for a piecewise linear edge in the digitised image. Typically edge points were assumed to be at unit distance apart, and an estimate was obtained by counting the number of points on the edge. For some of the estimators adjustments were made for corners, the presence of which resulted in over-estimation of the length. Since in this thesis the assessment of the usefulness of FD is of greater interest, (rather than precise

estimation of perimeter length), it is more appropriate to use estimators based on FD, although one estimator based on the digitised edge will be calculated for comparison. Lin and Chellappa (1987) gave formulae for the perimeter length and area in terms of the FD. Other shape properties, such as coordinates of the centroid of the region and central moments of inertia have been derived from the FD (Kiryati and Maydan, 1989) but will not be covered here.

Throughout this thesis it has been assumed that two separate stages are required for the estimation of edges. The image is first smoothed with the Laplacian of Gaussian (LoG) function and edge points identified as positions of zero-crossings. As a second stage a truncated FS is fit to this set of points and the edge is modelled as a closed curve. Fitting a subset of the first few harmonics in the FS has the effect of smoothing the edge but as may be observed from Figure 3.5, the set of points identified as zero-crossings already lie on a smooth curve. If only the edges of regions are of interest, then rather than smoothing the whole image an alternative method would be to track an edge (to the nearest pixel) in the discrete image. By including only the first few terms in the FS fit to the points (x_i, y_i) it is possible to obtain a smooth representation of the pixellated edge which is faster than smoothing the whole image using the methods as described in Chapter 3. It is worth noting that this second method would only be feasible if the images were already 'reasonably' smooth, as is the case for the MRI.

The results and conclusions discussed in this chapter are derived for fitting a FS to points on a pixellated edge which has been tracked with respect to the discrete image. The 'true' values of the perimeter length and area are not known for the human MRI, so simple geometric shapes (such as an ellipse or a rectangle) are simulated (Section 6.2.1). The known properties of these shapes are used as a standard against which to compare the estimated values. The tracking of edge points with respect to the discrete image is discussed in Section 6.2.2, to obtain a set of points to which a FS is fit. Perimeter length and area are calculated from the FD (Section 6.3) and the statistic used to assess the accuracy of the estimated properties is defined in Section 6.4. Results for estimating

perimeter length and area are discussed in Section 6.5 and a recommendation made for the 'optimum' number of harmonics to be fitted.

6.2 Simulated data and identification of edges

6.2.1 Simulation of geometric shapes

The perimeter length and area of a region in the available MRI are unknown, so to allow a quantitative assessment of how well a calculated property estimates the true value, images are simulated for which the true values are known. A total of $g = 7$ geometric shapes are simulated: ellipses with axis ratios, ρ say, of 2 and 4, circles, rectangles with side ratios ρ of 2 and 4, squares and equilateral triangles. Note that circles and squares are special cases of ellipses and rectangles respectively, with axis (or side) ratios of unity. Each geometric shape is simulated at four different sizes, where size is defined by α, the area covered. Random noise, denoted by η, is superimposed at three levels. The simulation algorithm is common to all shapes and so is discussed in general for any shape, with comments for modifications required for individual shapes where appropriate.

Each shape is simulated at a random position and orientation with respect to a 40 x 40 grid of intensity values. Pseudo-random numbers from a U(0,1) distribution are simulated using the NAG Pascal routine G05CAC. For each shape the random number generator is initialised to the same value by the routine G05CBC. The orientation of each shape is generated as a realisation from a $U(0,2\pi)$ distribution. Using the standard result that if a random variable X has a U(0,1) distribution, then a derived random variable $Y = a + (b-a)X$, with constants a,b, has a U(a,b) distribution. Similarly, the row and column coordinates of the random centre of each shape are generated as independent realisations of a U(19,20) distribution. Such a distribution restricts the centre to lie within a single pixel and is chosen in order to locate the shape in the 'middle' of the 40 x 40 array and ensure the simulated shape lies entirely within the grid.

Except for the triangle, in addition to the centre and orientation, two further parameters are required to define the shape to be simulated. These are the area to be covered by the shape and the ratio of the major to minor axes for an ellipse, or for a rectangle the ratio of the length to the width. Shapes are simulated with four areas $\{a = 50,100,200,400\}$ and with ratios $\{\rho = 2,4\}$. For a triangle, only the area α is required. Each pixel in the 40 x 40 array is considered in turn. Using standard trigonometric properties of the particular shape, the orientation, area and ratio parameters are used to evaluate the position of each pixel relative to the centre. The pixel can then be assigned either as 'object', that is lying within the area covered by the shape, or 'background' if it lies outside. It is assumed that an intensity of 1 is assigned to object pixels, that is pixels which are completely covered by the shape, and 0 to those lying outside the object. To evaluate intensity on a continuous scale an intermediate grey-level is assigned to pixels which lie on the edge between object and background pixels. The grey-level intensity represents the proportion of each edge pixel covered by the shape.

In order to assess the robustness to noise of the methods for estimation of geometric properties, random noise is superimposed on the simulated shapes. The NAG Pascal routine G05DDC is used to generate pseudo-random numbers from a $N(0,\eta^2)$ distribution. To each shape is added Gaussian noise with a standard deviation η of $\frac{1}{8}$, $\frac{1}{4}$ or $\frac{1}{2}$ corresponding to signal-to-noise ratios (SNR) of 8, 4 or 2 respectively.

Although the next sections will consider any single general shape it is worth remembering the range of different shapes that can be simulated. There are three 'factors' which can be varied: the area $\{\alpha = 50,100,200,400\}$, the random noise $\{\eta = \frac{1}{8},\frac{1}{4},\frac{1}{2}\}$ and the specific geometric shape $\{g = 1,2,\ldots,7\}$. In the following sections any of these 84 different shapes simulated with respect to fixed levels of $\{\alpha,\eta,g\}$ factors will be referred to as a (simulated) 'shape'. This should not be confused with the term 'geometric shape' which will be used to refer to one of the g = 7 levels of this factor, such as the

circle, triangle. These factors and levels are chosen to allow comparisons between 'regular' shapes (circle, square, triangle) against 'elongated' shapes (ellipse, rectangle), or between shapes with straight edges and corners (rectangle, square, triangle) against those with curved boundaries (ellipse, circle). It should also be possible to assess any interaction between robustness to noise and increasing area. Typical shapes are illustrated in Figure 6.1, showing different combinations of the levels of the shape, area (size) and noise factors.

6.2.2 Tracking an edge in discrete space

The grey-level image as illustrated in Figure 6.1 is thresholded to obtain the corresponding binary images in Figure 6.2, where white represents an intensity of zero and black an intensity of one. For the simulated data, recorded on a continuous [0,1] grey-scale, a threshold of 0.5 is used. All pixels where the intensity exceeded 0.5 were identified as object and assigned a value of one, the remaining pixels are background and assigned a value of zero.

Edges of objects in the binary image are identified by the tracking algorithms {BDFL1, BDFL2} as implemented in the SPIDER package. The tracking algorithm is based on topological properties of digitized binary images, as derived by Yokoi et al (1975). Note that for the quantity defined as an 'edge' in this thesis, the term used by Yokoi et al (1975) is 'border' and in the SPIDER algorithm, 'boundary'. For consistency with the rest of this thesis, the term 'edge' will be used in the following discussion of the implemented algorithm. Two changes are made to the algorithm as supplied in the SPIDER package. First, the code is converted from Fortran to Pascal. Second, it proved necessary to augment the binary image to half-pixel resolution to ensure tracking of all fine details of the edge, for example on an extended straight section round an indentation of only a single pixel.

As can be seen from the thresholded image of Figure 6.2, there are a number of 'small' (of size two or three pixels in area) black

Figure 6.1a: Simulated ellipse: $\rho = 2$, $\alpha = 200$, $\eta = \frac{1}{4}$

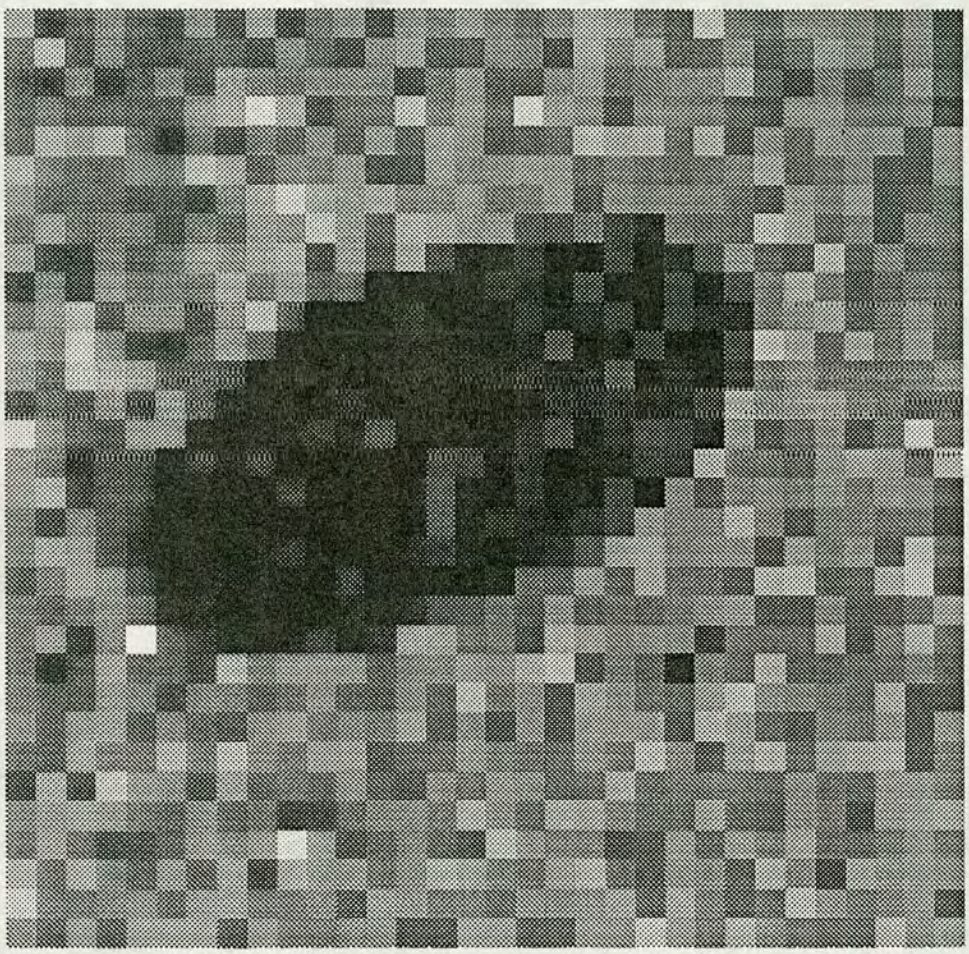


Figure 6.1b: Simulated square: $\rho = 1$, $\alpha = 400$, $\eta = \frac{1}{8}$

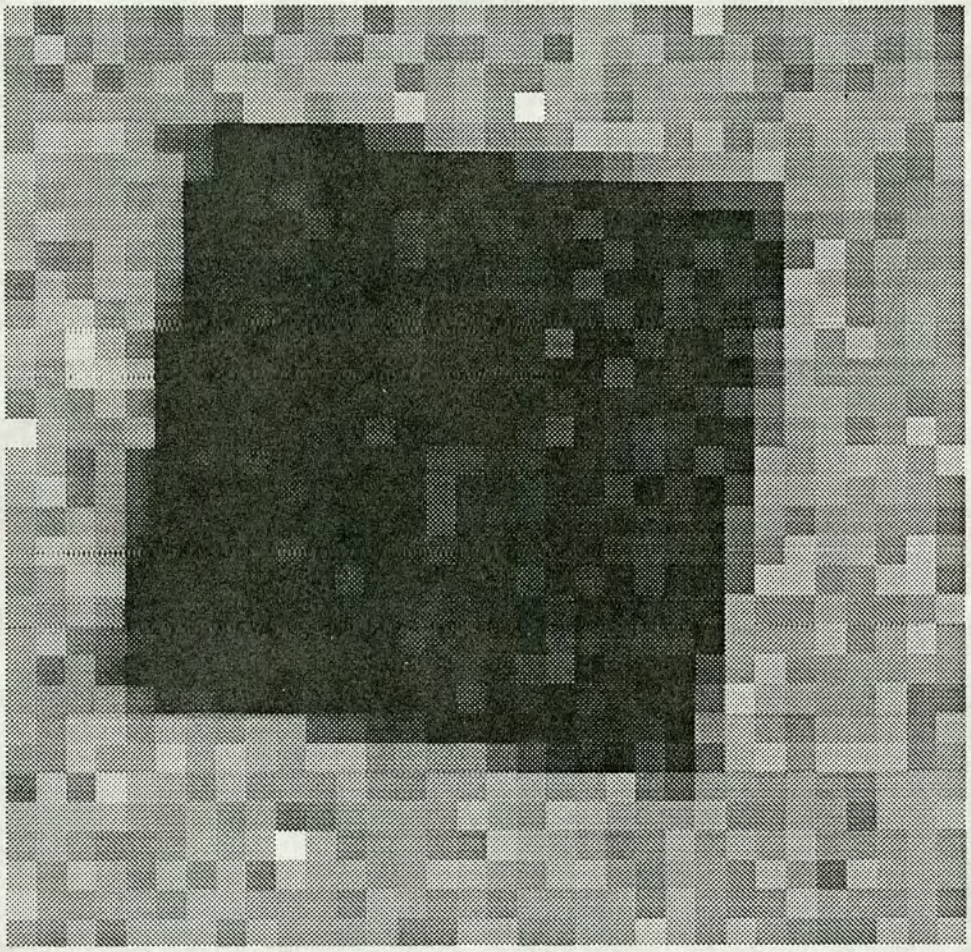


Figure 6.1c: Simulated triangle: $\alpha = 50$, $\eta = \frac{1}{2}$

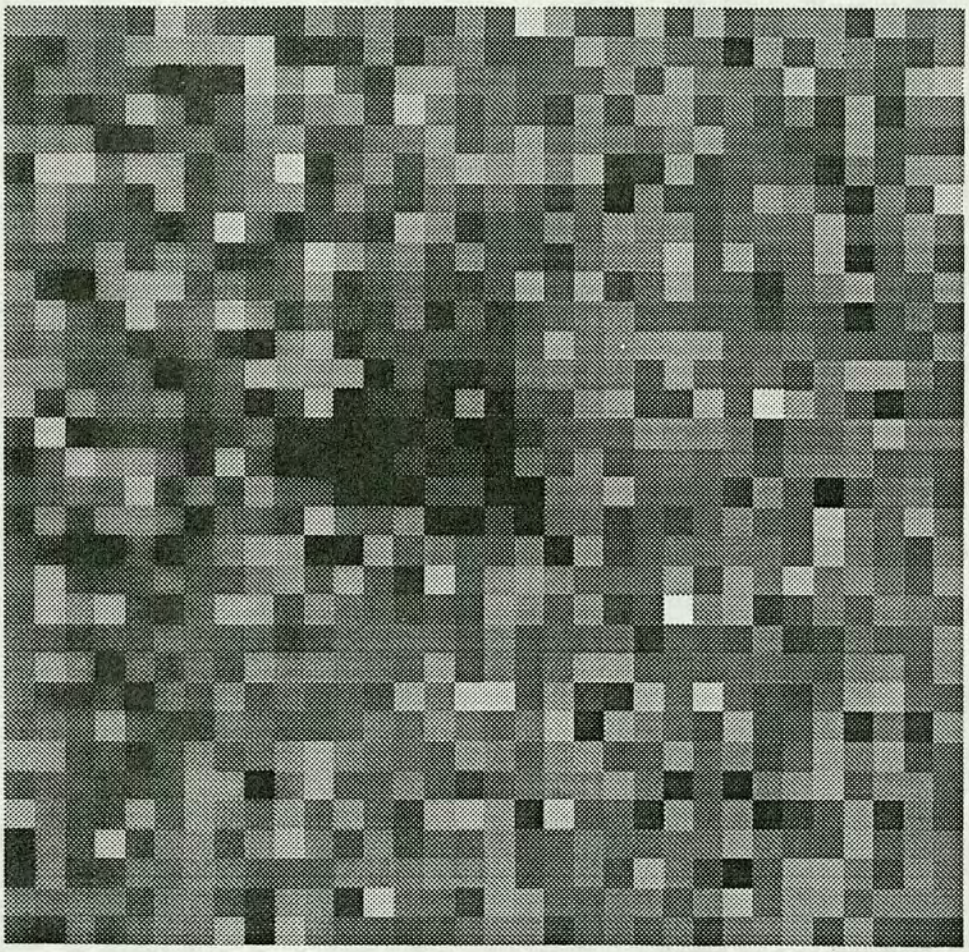


Figure 6.2a: Thresholded (binary) ellipse: $\rho = 2$, $\alpha = 200$, $\eta = \frac{1}{4}$

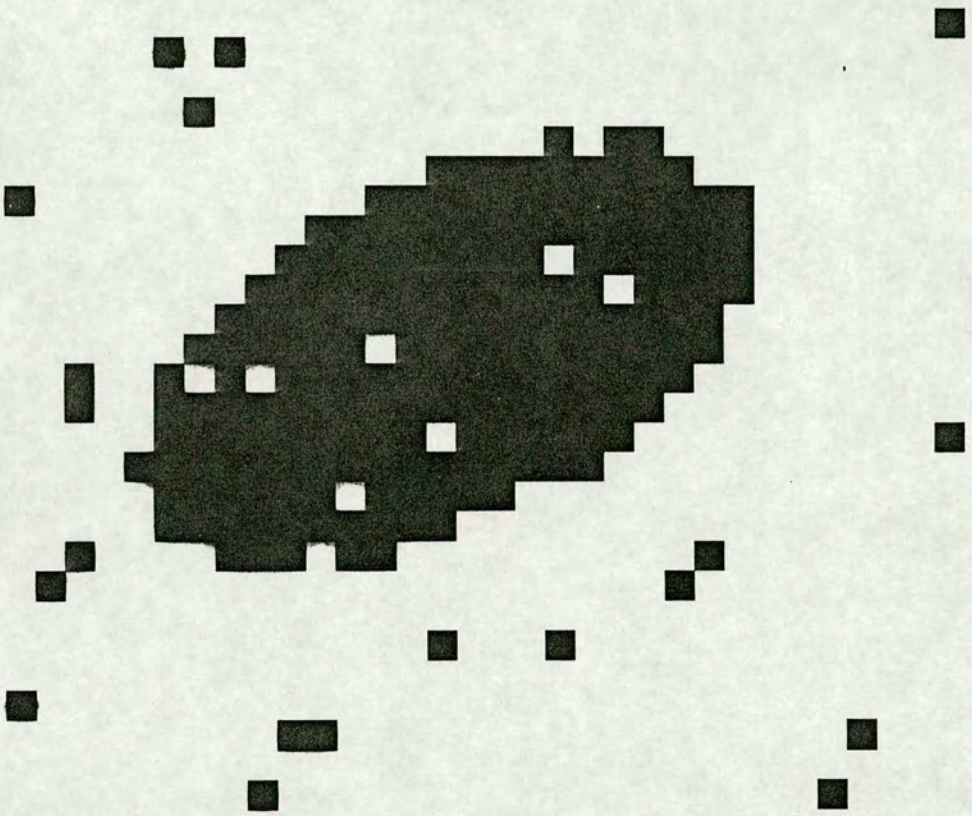


Figure 6.2b: Thresholded (binary) square: $\rho = 1$, $\alpha = 400$, $\eta = \frac{1}{8}$

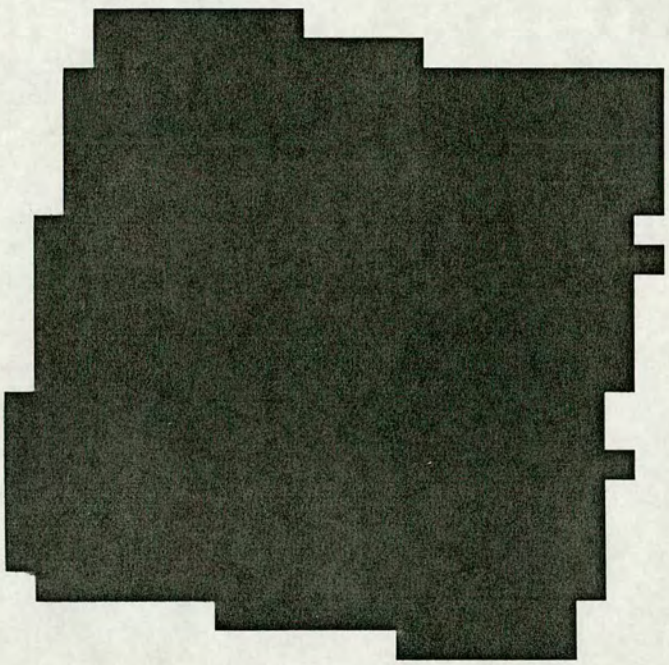
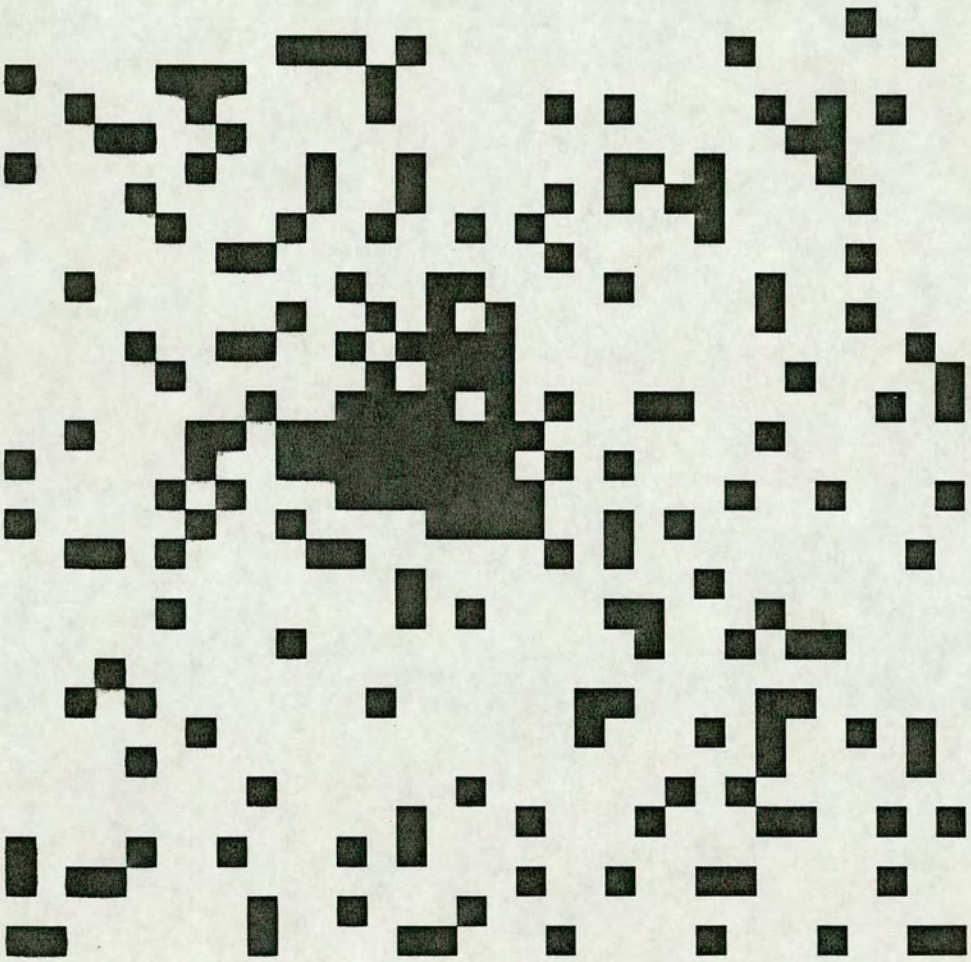


Figure 6.2c: Thresholded (binary) triangle: $\alpha = 50$, $\eta = \frac{1}{2}$



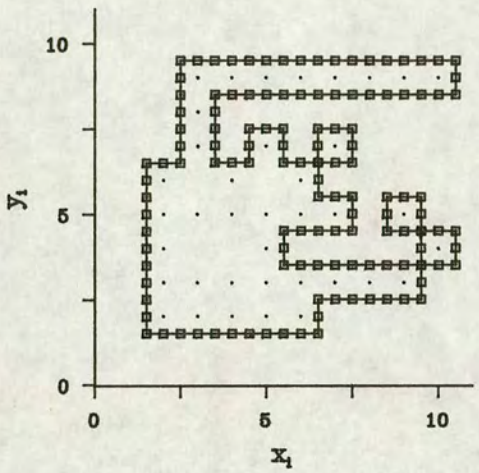
regions (of intensity 1) in the background to the main object and similarly a few isolated white pixels (of zero intensity) within the main object. This is particularly true for the simulations at higher noise levels. The algorithm BDFL2 tracks the edges of all individual connected regions in the image, calling a second algorithm BDFL1 to track round each connected region of intensity 1 say, where the neighbouring pixels are of intensity zero. The required edge of the main shape object is defined as that enclosing the largest area, where the area is calculated using a formula to be given in Section 6.3.3. In the remainder of this chapter all discussion of shape (or object) will be with respect to this single 'main' edge.

After tracking, the edge is stored as a set of coordinates (x_i, y_i) at half-pixel resolution, that is the Euclidean distance between each consecutive pair of points is 0.5, where each pixel is of unit length. The edge tracked at half-pixel resolution is illustrated for a simple test example in Figure 6.3. The object pixels are recorded as a dot for the top left vertex of each pixel and backgound pixels are left blank. In Figure 6.3a a square represents the coordinates of edge points at half-pixel resolution. Since the simulated image is observed only at pixel resolution, the edge tracked at half-pixel resolution contains redundant information and is sampled to give a subset of pixels at pixel resolution. The two possible subsets of sampled coordinates are shown in Figure 6.3b, represented by a circle and asterisk.

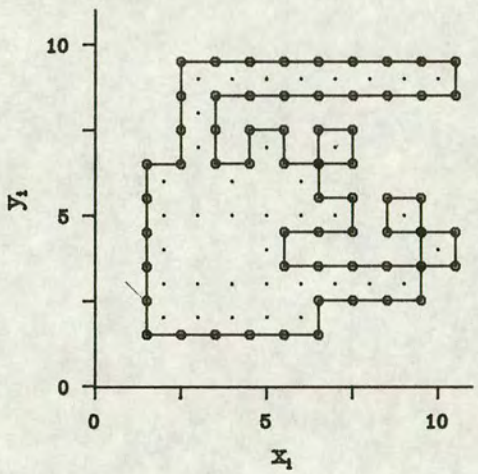
As can be seen from Figure 6.3c, the subset of points represented by a circle are equally-spaced at a distance of one pixel. The Fourier coefficients, the estimate of the area based on the tracked coordinates and the Kulpa (1977) estimate of perimeter length (to be defined in Section 6.3.2) will be calculated from this set of edge points. In contrast, the subset of points represented by an asterisk (Figure 6.3d) are not equally-spaced. The tracked edge has its 'corners cut-off' in an attempt to reduce the overestimation which results when tracking on a discrete grid the edge of an object which exists in continuous space. This subset of points is used only to calculate the estimate of perimeter length based on the tracked coordinates, the P(t) to be defined in Section 6.3.2. The two

Figure 6.3: Tracking an edge in discrete space
using SPIDER algorithm BDFL1

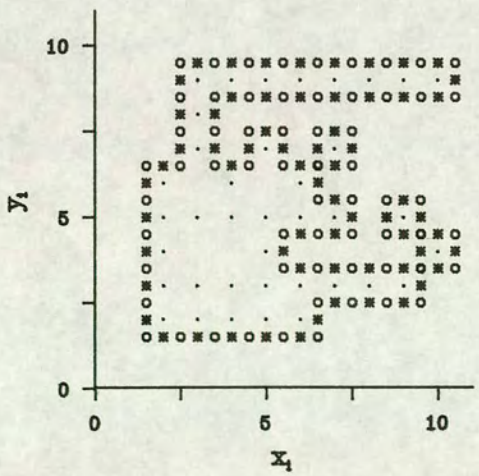
a) half–pixel edge



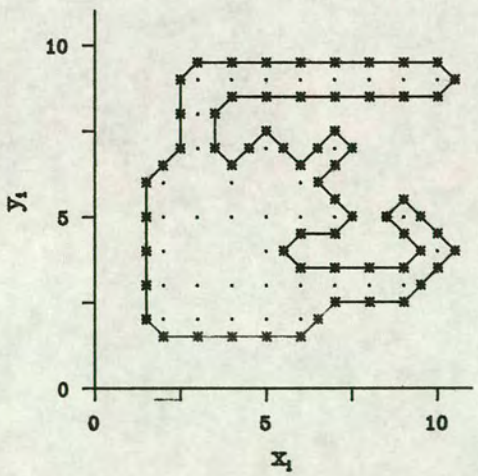
c) Subset 1 (calculation of FD)



b) pixel–level sampling



d) Subset 2 (perimeter length)



subsets of points will not be referred to separately in the following and except for the estimation of the 'tracked' perimeter length, the first subset of points (at equal spacing, and marked by a circle) will be assumed.

The SPIDER algorithm is used to track round the (pixellated) edge of an object to obtain an ordered set of points or vertices $\{(x_i,y_i) : i = 1,2,...,n\}$. The distances between adjacent points are equal, and may without loss of generality be taken to be unity. As a consequence of the equal spacing of points, the interpolated equal distance and index number labelling (as discussed in Section 5.3) are equivalent. The choice of 'best' labelling variable is therefore not an issue in this chapter, and it is assumed that points are labelled in terms of index number.

The set of vertices $\{(x_i,y_i) : i = 1,2,...,n\}$ is of the form discussed in Chapter 5: an ordered set of points lying on a closed curve. Using the FFT, a FS is fit to the coordinates (x_i,y_i) as described in Section 5.2.3 and the FD estimated as in (5.1), as the coefficients $\{a_o,c_o,a_r,b_r,c_r,d_r : r = 1,2,..,R\}$. Coordinates $(x_i(r^*),y_i(r^*))$ are estimated by using the inverse FT to fit a truncated series with $r^* \leqslant R$ harmonics. In this chapter, R sets of fitted coordinates $\{(x_i(r),y_i(r))\ r = 1,2,...,R\}$ are estimated using the inverse FT, truncating the series at $\{r = 1,2,...,R\}$ harmonics. Corresponding values of the perimeter length and area can be calculated from the estimated coordinates as will be described in Section 6.3.

Recall that the value of n, and hence the coordinates of the edge points may require some modification before the FFT can be used to fit the FD (Section 5.2.3). A method is now outlined which could be used to adjust the number of points on a pixellated edge. The longest 'straight' section (that is without 'corners') of the edge is identified. One point (that is, a pixel vertex) is added or deleted and the remainder of the vertices interpolated to equal distances with respect to that single section. This is repeated for a sequence of the next longest straight sections until a suitable value of $n = n^*$ is found. This modification means that although points on

individual straight sections are equally spaced, there is no longer overall equal spacing for the revised set of n^* points on the closed edge.

In practice if n is unsuitable for the FFT, then rather than modifying the number of points the data is discarded and a further random shape simulated. Obviously this is not an option for the MRI but since edges in the MRI are more irregular the modification of n should have negligible affect on the estimation of the FD.

6.3 Estimation of perimeter length and area

6.3.1 Introduction

For all shapes except the ellipse, the true perimeter length P can be calculated using standard geometric properties. Since no formula exists for the ellipse the perimeter length must be evaluated numerically. The true area A of each shape is already known since it is one of the parameters (α) used to define the shape. The estimates of perimeter length will be defined in terms of unit pixel length, and the estimates of area in terms of unit pixel area.

The calculation of the perimeter length and area, P(t) and A(t) say, from the tracked edge points $\{(x_i,y_i) : i = 1,2,\ldots,n\}$ is described in the following two sections. Estimated coordinates $\{(x_i(r),y_i(r) : r = 1,2,\ldots,R\}$ can be calculated, including an increasing number of harmonics r in the fitted series. By replacing coordinates (x_i,y_i) by the estimated $(x_i(r),y_i(r))$ in formulae for perimeter length and area, corresponding properties P(r) and A(r) can be estimated for fitting $\{r = 1,2,\ldots,R\}$ harmonics. Recall that the motivation for this chapter is the determination of the number of harmonics r to be included in the series. This may now be obtained by estimating $\{P(r),A(r) : r = 1,2,\ldots,R\}$ and assessing which best estimates the corresponding true values P and A.

The properties P(t) and A(t) are also calculated for the tracked points $\{(x_i,y_i) : i = 1,2,\ldots,n\}$. These values are compared with P

and A to assess whether accurate estimates of the properties are possible wthout fitting the FD, so the additional computation required to calculate the FD can be avoided. This would be a consideration if the estimation of such properties is the priority. Here the estimation of values of P and A is of secondary interest: the estimation is used in order to develop a criterion for truncation of the FS.

6.3.2 Estimation of perimeter length

Recall from Section 6.2.2 that the set of tracked coordinates, to be used only for the estimation of 'tracked' perimeter length, is defined as the second subset of sampled points (marked by an asterisk in Figure 6.3). The perimeter length is calculated as the cumulative sum of the (Euclidean) distance between each pair of consecutive points. With respect to the tracked coordinates the estimator P(t) is defined as:

$$P(t) = \sum_{i=1}^{n} \{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2\}^{1/2}. \qquad (6.1)$$

Since points lie on a closed curve, $(x_{n+1}, y_{n+1}) = (x_1, y_1)$.

The reason for using this second subset of sampled coordinates is to 'cut-off the corners' of the edge tracked at half-pixel resolution. This is an attempt to reduce overestimation of the perimeter length which occurs when the edge of an object in continuous space is located only to the nearest pixel when tracking with respect to the discrete grid. A different approach to adjust for this overestimation is to use a simple, unbiased estimator for smooth edges. One example is given by Kulpa (1977), which will be denoted by P(k) and defined as:

$$P(k) = \frac{\pi}{8(\sqrt{2}-1)} \left\{ n - \left[1 - \frac{1}{\sqrt{2}} \right] \Psi \right\}, \qquad (6.2)$$

where Ψ is the number of corner points on the edge. The edge is that given by the first subset of points sampled from the half-pixel edge,

that is those used for the calculation of the FD from the tracked coordinates and marked by a circle in Figure 6.3.

Fitted coordinates $\{(x_i(r),y_i(r)) : r = 1,2,...,R\}$ are calculated from the FD by using an inverse FT (Section 5.2.3). Corresponding estimators of perimeter length $\{P(r) : r = 1,2,...,R\}$ are defined by replacing the tracked coordinates $\{(x_i,y_i) : i = 1,2,...,n\}$ in (6.1) by the fitted coordinates for the corresponding value of r $\{(x_i(r),y_i(r)) : i = 1,2,...,n\}$. Since the estimator P(r) is obtained by fitting r FD to the edge points, it will be termed the FD 'fitted' estimator (of perimeter length) in the following.

The results for the R+2 estimators of perimeter length $\{P(r) : r = 1,2,...,R,$ P(t) and P(k)$\}$ as calculated for the simulated images will be discussed fully in Section 6.5, but it is worth making a few preliminary comments here. Even using the second subset of sampled points, the estimator P(t) (6.1) calculated from the tracked coordinates, is not expected to be very robust to increases in noise level η. As η increases the edges become longer, more winding and fragmented. By tracking round all the 'indentations' of the edge the perimeter length estimated from the tracked coordinates will increase as η increases. This is illustrated by considering the simple example of a circle with true area A = 100 and perimeter P = 35.4. For noise levels of η = 1/8, 1/4 and 1/2 the estimates of P(t) are 38.6, 41.5 and 56.8 respectively. As η becomes very large the perimeter length will tend towards a limit of ∞, though once η exceeds some critical value, then by eye there will be no visible change in the plotted edge.

The global properties of the fitted edge will be captured by coordinates $\{x_i(r),y_i(r) : i = 1,2,...,n\}$ calculated from a FS including only the first few harmonics. Adding higher terms in the FS will describe more local detail of the noisy edge. The estimated perimeter length P(r) is expected to increase as r, the number of fitted harmonics, increases. The simulations are used to test a hypothesis that the 'best' estimator of perimeter length is one calculated from the fitted coordinates from a FS including only the 'first few' harmonics.

6.3.3 Estimation of area

In theory, for the simulated images of Figure 6.1 which exist in continuous two-dimensional space, the area of a region enclosed by a curve is calculated by integration. In practice, for the thresholded data observed with respect to the discrete grid, the calculated area is that of the polygonal approximation (as in Figure 6.2) of the continuous region. The area A(t) calculated with respect to the tracked coordinates is defined as:

$$A(t) = \left| \sum_{i=1}^{n} (x_i\, y_{i+1} - x_{i+1}\, y_i) \right|. \qquad (6.3)$$

The points lie on a closed curve, so $(x_{n+1},y_{n+1}) = (x_1,y_1)$. The sign associated with the calculated value will depend on the direction of tracking of the edge, but since this does not affect the magnitude of the estimate the absolute value is used and is denoted by $|\ |$. Estimates of area can also be calculated from the fitted coordinates $\{(x_i(r),y_i(r)) : r = 1,2,\ldots,R\}$. Corresponding area estimates $\{A(r) : r = 1,2,\ldots,R\}$ for fitting an increasing number of harmonics in the FS are defined by replacing the tracked coordinates $\{(x_i,y_i) : i = 1,2,\ldots,n\}$ in (6.2) by the appropriate fitted coordinates $\{(x_i(r),y_i(r)) : i = 1,2,\ldots,n\}$. In the following the estimator A(r) wiil be termed the FD 'fitted' estimator (for area).

The results for the R+1 estimators of area $\{A(r) : r = 1,2,\ldots,R$ and $A(t)\}$ calculated for the simulated images will be discussed in Section 6.5. The area estimator A(t) (6.3) based on the tracked coordinates is expected to be reasonably robust to increases in the noise level σ. As σ increases the edge becomes more winding and indented. The perimeter length increases but the edge does not enclose much additional area. Any extra area included in the estimate from small adjacent connected regions is, on average, cancelled out by area from similar regions omitted. There is little change in the estimate of the area at the three σ levels, as is illustrated by again considering the simple example of a circle with true area A = 100. At σ levels of ⅛, ¼ and ½ the estimates of the area A(t) are 100, 104 and 115 respectively. Thus particularly at low noise levels

it may be possible to obtain a reasonably accurate estimate of the area based on the tracked coordinates. Alternatively, if area A(r) is calculated from the fitted coordinates, it is expected that only two or three harmonics will be needed in the FS to accurately estimate the area.

6.4 Assessing the estimated properties

6.4.1 The assessment criterion

Recall from Section 6.2.1 that g = 7 geometric shapes are simulated, each at four different sizes (α) and at three levels (η) of superimposed random noise. Perimeter length and area are estimated for 84 shapes with different combinations of the levels of these factors. For any given simulated shape, let $\hat{P}$ and $\hat{A}$ denote the estimators for perimeter length and area respectively, where properties may be estimated by any of the methods in Sections 6.3.2 and 6.3.3. Some statistic is sought to quantify 'how well' $\hat{P}$ or $\hat{A}$ estimates the corresponding true value, P or A. One quantity, defined for perimeter length, is the bias, $BIAS_P = (P - \hat{P})$, and similarly for the area bias, $BIAS_A$. Since the magnitude of deviation is more important than the direction, the preferred criterion is the square of the bias, $(BIAS_P)^2 = (P - \hat{P})^2$.

A more robust measure of the quality of an estimator is obtained by calculating $\hat{P}_s$ and $\hat{A}_s$ for $\{s = 1,2,\ldots,S\}$ simulations for a given shape. The same fixed area, noise and geometric shape parameters $\{\alpha,\eta,g\}$ are retained but the S shapes are simulated with different random centres and orientations. The $(BIAS_P)^2$ is calculated for all S simulations and results averaged to define the root mean square error (RMSE) criterion for assessing the quality as:

$$RMSE_P = \left[\sum_{s=1}^{S} \{P - \hat{P}_s\}^2 / S \right]^{1/2} \qquad (6.4)$$

and similarly for area, $RMSE_A$.

The subscript s in $\hat{P}_s$ or $\hat{A}_s$ denotes the estimates of $\hat{P}$ or $\hat{A}$ calculated for the s^{th} simulation. Estimates $\hat{P}_s$, $\hat{A}_s$ are not considered individually, since only the results for averaging over S such estimates are required. Therefore for convenience the subscript s is usually omitted, except in formulae involving summation, in which case it will be made explicit. Most of the following will refer to the general estimators $\hat{P}$ and $\hat{A}$, with the understanding that S such estimates $\{\hat{P}_s, \hat{A}_s : s = 1,2,\ldots,S\}$ are calculated for each shape. For perimeter length, the estimator $\hat{P}$ is one of P(t), the estimator (6.1) calculated from the tracked coordinates or P(k), the Kulpa estimator (6.2) or one of the R FD fitted estimators $\{P(r) : r = 1,2,\ldots,R\}$, from fitting a FS with r harmonics. Individual estimates corresponding to $\hat{P}_s$ for the s^{th} simulation would be $\{P_s(t), P_s(k)$ or $P_s(r) : r = 1,2,\ldots,R\}$ respectively. Similarly for area the estimator $\hat{A}$ represents one of A(t) (6.3) or $\{A(r) : r = 1,2,\ldots,R\}$, with estimator $\hat{A}_s$ for individual simulations representing $\{A_s(t)$ or $A_s(r) : r = 1,2,\ldots,R\}$.

The RMSE is a combined measure of the average bias for each simulation, and the variation between simulations with the same parameters. The optimal estimator for perimeter length or area is defined as that which minimises the $RMSE_P$ or $RMSE_A$ respectively. The $RMSE_A$ for area is used as a criterion to determine the number of simulations S over which to average results for a given shape with fixed parameters $\{\alpha,\eta,g\}$. For computational efficiency the aim would be to average over as few simulations as possible while still correctly identifying the optimum estimator. Obviously a larger number of simulations result in more accurate estimates of the magnitude of the $RMSE_A$ but here the magnitude is not the main interest. The result required from each set of S simulations is the optimal estimator $\hat{A}$ which minimises the $RMSE_A$. The value of S required is the smallest number of simulations with which the minimum is correctly identified. The method used to determine S is now summarised.

For a given shape with fixed (α,η,g) parameters, 1000 simulations were carried out to identify the 'ideal' estimator $\hat{A}$. Those simulations where the value of n was unsuitable for use of the FFT

were discarded and another shape generated. Averaging over the 1000 simulations, the $RMSE_A$ was calculated as in (6.4) and the estimator $\hat{A}$ minimising the $RMSE_A$ identified as the 'ideal' estimator. For the shape simulated this ideal estimator was $A(r^*)$, that is the $RMSE_A$ was minimised by fitting a truncated FS with r^* harmonics.

The series of 1000 simulations was repeated for the same shape with fixed $\{\alpha,\eta,g\}$, but now recalculating the $RMSE_A$ after each additional simulation s. For each cumulative number of simulations $\{s = 1,2,\ldots,1000\}$ the FD fitted estimator minimising the corresponding $RMSE_A$ was identified and compared with the ideal estimator $A(r^*)$, in order to assess with how few simulations this estimator can be correctly identified.

The calculation of the $RMSE_A$ for a cumulative number of simulations is repeated for several shapes with different fixed parameters $\{\alpha,\eta,g\}$. In each case the conclusion is the same: a reasonably consistent estimate of the estimator $\hat{A}$ minimising the $RMSE_A$ can be determined from S = 100 simulations. Simulations where n, the number of points on the tracked edge, is unsuitable for use of the FFT are discarded, and a further shape generated. Therefore in the following analysis it is assumed that each of the 84 possible shapes is simulated 100 times and the results given are summaries from averaging over S = 100 simulated random shapes.

Even with fixed parameters (α,η,g) the number of observations n and hence R = n/2, varies for each of the 100 random simulations. Thus for each of the 84 shapes, the smallest value of R over the 100 simulations is identified. The RMSE values are only considered for values of r up to this minimum, since only these RMSE results will be based on exactly 100 simulations.

6.4.2 Comparison of results between shapes

For each of the 84 possible shapes from all combinations of the levels of the $\{\alpha,\eta,g\}$ factors, the $RMSE_P$ or $RMSE_A$ is calculated for each estimator $\hat{P}$ or $\hat{A}$. The $RMSE_P$ and $RMSE_A$ require some modification

before direct comparisons can be made between the wide range of simulated shapes. Some allowance must be made for the higher $RMSE_P$ and $RMSE_A$ for those shapes which, by definition, have a larger area ($\alpha = 200$ or $\alpha = 400$) and hence longer perimeters. Although individual biases $BIAS_P = (P - \hat{P})$ may be of several orders of magnitude greater than the bias for a shape with smaller area, $\alpha = 50$ say, as a proportion of the true perimeter length P the corresponding scaled bias may be smaller. Thus it is a measure of the lack-of-fit, as measured by the $RMSE_P$ as a proportion of the true value P, which should be compared between sets of S simulations. Therefore when calculating the $RMSE_P$ for perimeter length, the $BIAS_P$ is first scaled by the true perimeter length P and converted to a percentage, so the definition (6.4) is modified to:

$$\%RMSE_P = 100 \left[\frac{1}{100} \sum_{s=1}^{100} \left\{ \frac{P - \hat{P}_s}{P} \right\}^2 \right]^{1/2}. \qquad (6.5)$$

Similarly for $\%RMSE_A$, where $BIAS_A$ is scaled by the true area A.

Any increase in the area α and level of noise η of the shape is matched by an increase in the length of the tracked edge enclosing the shape. This increase in perimeter length means there is an increase in n, the number of points (x_i, y_i) on the edge. Consequently there is an increase the maximum number of harmonics which can be fitted in a FS, that is $R = n/2$, where n is assumed even. For the 84 different shapes there is a wide range in the values of R, from $R = 14$ for a circle or square of area $\alpha = 50$ and $\eta = 1/6$, to $R = 75$ for the ellipse of axis ratio $\rho = 4$, area $\alpha = 400$ and $\eta = 1/2$. In some situations it may be desirable to compare RMSE values over all shapes for numbers of harmonics greater than 14. Therefore for those shapes where the value of R is less than 75 (the maximum value of R over all 84 sets of simulations for the different shapes), results are extended by repeating the RMSE value for fitting R harmonics calculated with respect to that shape, up to a maximum of 75 terms. For perimeter length the $\%RMSE_P$ for fitting r harmonics $\{r = 1,2,\ldots,R=75\}$ is defined as:

$$\%RMSE_P(r) = 100 \left[\frac{1}{100} \sum_{s=1}^{100} \left\{ \frac{P - P_s(r)}{P} \right\}^2 \right]^{1/2}. \qquad (6.6)$$

The $\%RMSE_A(r)$ is defined similarly for area. Direct comparisons can now be made between any of the 84 shapes, at any fixed number of harmonics $\{r = 1,2,\ldots,75\}$.

Analagous $\%RMSE_P$ can be defined for the other estimators of perimeter length: for Kulpa's method $P(r)$ is replaced by $P(k)$ (6.2) to give $\%RMSE_P(k)$ and for the estimator from the tracked points $P(t)$ (6.1) replaces $P(r)$, with corresponding $\%RMSE_P(t)$. For area, $\%RMSE_A(t)$ is calculated from $A(t)$ (6.3), that is the area estimator based on the tracked coordinates.

Since image regions usually have an irregular shape, it is important that the number of harmonics (as identified as the minimum of the $\%RMSE_P$ or $\%RMSE_A$ criterion) is optimal for any general shape and not just one of the seven specific geometric shapes simulated here. Extending the notation introduced earlier let P_g, A_g denote the true perimeter and area for the g^{th} geometric shape, $\{g = 1,2,\ldots,7\}$. Corresponding estimators for the g^{th} geometric shapes are $\hat{P}_g$ and $\hat{A}_g$, where $\hat{P}_g$ is one of $\{P_g(r) : r = 1,2,\ldots,75, P_g(t) \text{ or } P_g(k)\}$ and $\hat{A}_g$ is one of $\{A_g(r) : r = 1,2,\ldots,75 \text{ or } A_g(t)\}$. Further, estimates for individual simulations $\{s = 1,2,\ldots,100\}$ are $\{P_{gs}(r) : r = 1,2,\ldots,75, P_{gs}(t) \text{ or } P_{gs}(k)\}$ for perimeter estimator $\hat{P}_{gs}$, and $\{A_{gs}(r) : r = 1,2,\ldots,75 \text{ or } A_{gs}(t)\}$ for area estimator $\hat{A}_{gs}$.

Averaging over all $g = 7$ geometric shapes with the same fixed $\{\alpha,\eta,g\}$ parameters, then with respect to the general estimator of perimeter length $\hat{P}_g$, a summary %RMSE for perimeter length and area are defined as:

$$\%RMSE7_P = 100 \left[\frac{1}{7} \frac{1}{100} \sum_{g=1}^{7} \sum_{s=1}^{100} \left\{ \frac{P_g - \hat{P}_{gs}}{P_g} \right\}^2 \right]^{1/2}, \qquad (6.7a)$$

and

$$\%RMSE7_A = 100 \left[\frac{1}{7} \frac{1}{100} \sum_{g=1}^{7} \sum_{s=1}^{100} \left\{ \frac{A_g - A_{gs}}{A_g} \right\}^2 \right]^{1/2} \qquad (6.7b)$$

Note that the '7' is included in the abbreviated name to indicate that the %RMSE is averaged over all seven geometric shapes.

Analagous $\%RMSE_P$ values could be defined for specific estimators, for example for $\{P_g(r) : r = 1,2,\ldots,75\}$ by extending the definition given in (6.6) to average over the g = 7 geometric shapes as in (6.7) to calculate $\{\%RMSE_P(r) : r = 1,2,\ldots,75\}$; similarly for $\%RMSE_P(t)$, $\%RMSE_P(k)$, $\{\%RMSE_A(r) : r = 1,2,\ldots,75\}$ and $\%RMSE_A(t)$.

6.5 Discussion of results

6.5.1 Perimeter

The results are summarised in Table 6.1 for the 12 combinations of the levels of the fixed $\{\alpha,\eta\}$ factors, averaging over the g = 7 geometric shapes. For the FD fitted estimators the values tabulated are r^*, the optimal number of coefficients and $\%RMSE7_P(r^*)$, the RMSE for that number of harmonics. The corresponding $\%RMSE_P(k)$ and $\%RMSE_P(t)$ are also given. In each case the $\%RMSE7_P(r^*)$ is less than $\%RMSE_P(k)$ or $\%RMSE_P(t)$. All the %RMSE values increase with increasing η but this increase is much greater for $\%RMSE_P(k)$ and $\%RMSE_P(t)$ and the superiority of the FD fitted estimator $P(r^*)$ is more apparent. The rapid increase of $\%RMSE_P(k)$ and $\%RMSE_P(t)$ is due to the overestimation of the length since the edge becomes increasingly more 'winding' as η increases. The Kulpa method is not sufficiently robust to account for such length increases and even 'cutting-off' the corners of the tracked edge (as described in Section 6.2.2) does not overcome the estimation.

For the FD fitted estimator $P(r^*)$ the optimal r^* does vary depending on the levels of the $\{\alpha,\eta\}$ factors, but trends can be identified for increasing α and η. The value of r^* increases as the area α increases. As the area of a region increases there will be a

Table 6.1: Perimeter
- Optimal number of Fourier descriptors with resulting %RMSE7 for Fourier descriptor estimator and corresponding %RMSE for Kulpa and tracked estimators

Noise η	Area α	Optimal No. of FD r^*	Minimum $\%RMSE7_P(r^*)$ $P(r^*)$	Kulpa $\%RMSE_P(k)$ $P(k)$	Tracked $\%RMSE_P(t)$ $P(t)$
1/8	50	7	3.9	4.8	4.7
	100	10	2.9	3.3	5.2
	200	13	2.2	2.5	5.2
	400	21	3.6	4.1	5.0
1/4	50	5	5.6	10.5	15.0
	100	7	3.8	9.5	14.5
	200	9	3.0	8.8	14.1
	400	13	4.3	8.2	13.1
1/2	50	1	12.9	80.8	90.1
	100	2	10.4	76.6	85.8
	200	3	7.1	72.4	81.6
	400	5	6.6	61.8	65.1

corresponding increase in the number of points on its tracked edge and consequently in R, the maximum possible number of harmonics which can be fitted. Thus it is reasonable that there is a proportional increase in the minimum, that is optimal, number of coefficients required to describe the global structure of the edge enclosing the region.

The optimal r^* increases with increasing signal-to-noise ratio (SNR), or equivalently decreasing η. The smaller the value of r^*, the greater the amount of smoothing of the edge: the first few coefficients describe only the global structure, while higher terms model more local detail. Thus it is reasonable that at high noise levels, the value of r^* is fairly small as more smoothing is required for the longer noisy and 'winding' edge. For small η the contrast between global shape structure and local noise fluctuations is not so pronounced. In this case the so-called 'optimal' number of coefficients may in fact be taking account of some of the random behaviour.

Similar patterns can be identified if results are considered for the seven geometric shapes individually, rather than looking at their average. The typical shape of the $\%RMSE_P$ function is illustrated in Figure 6.4a. Two combinations of factor levels are shown: in *(i)* $\{\eta = \frac{1}{8}, \alpha = 100\}$, and in *(ii)* $\{\eta = \frac{1}{2}, \alpha = 200\}$. Note that since the key is superimposed on the top graph, for clarity the plotted lines are truncated at r = 40 harmonics since they remain constant after this point. For each geometric shape the $\%RMSE_P(r^*)$ at the optimal r^* is always less than $\%RMSE_P(k)$ or $\%RMSE_P(t)$. As expected (see Section 6.3.2) the improvement is more pronounced for the higher noise in the second graph.

For the seven geometric shapes, the optimal r^* for the fitted estimator P(r) are tabulated in Table 6.2. The circle, and to a slightly lesser extent the ellipses, are well estimated by fitting only the first few harmonics. These shapes have smooth, continuously 'curving' edges and the equation of the 'true' shape is just the weighted sum of the first $(\cos\theta, \sin\theta)$ term in the FS. More harmonics are required to describe the triangle, rectangles and the square.

Figure 6.4a: Perimeter %RMSE$_p$ for increasing number of harmonics

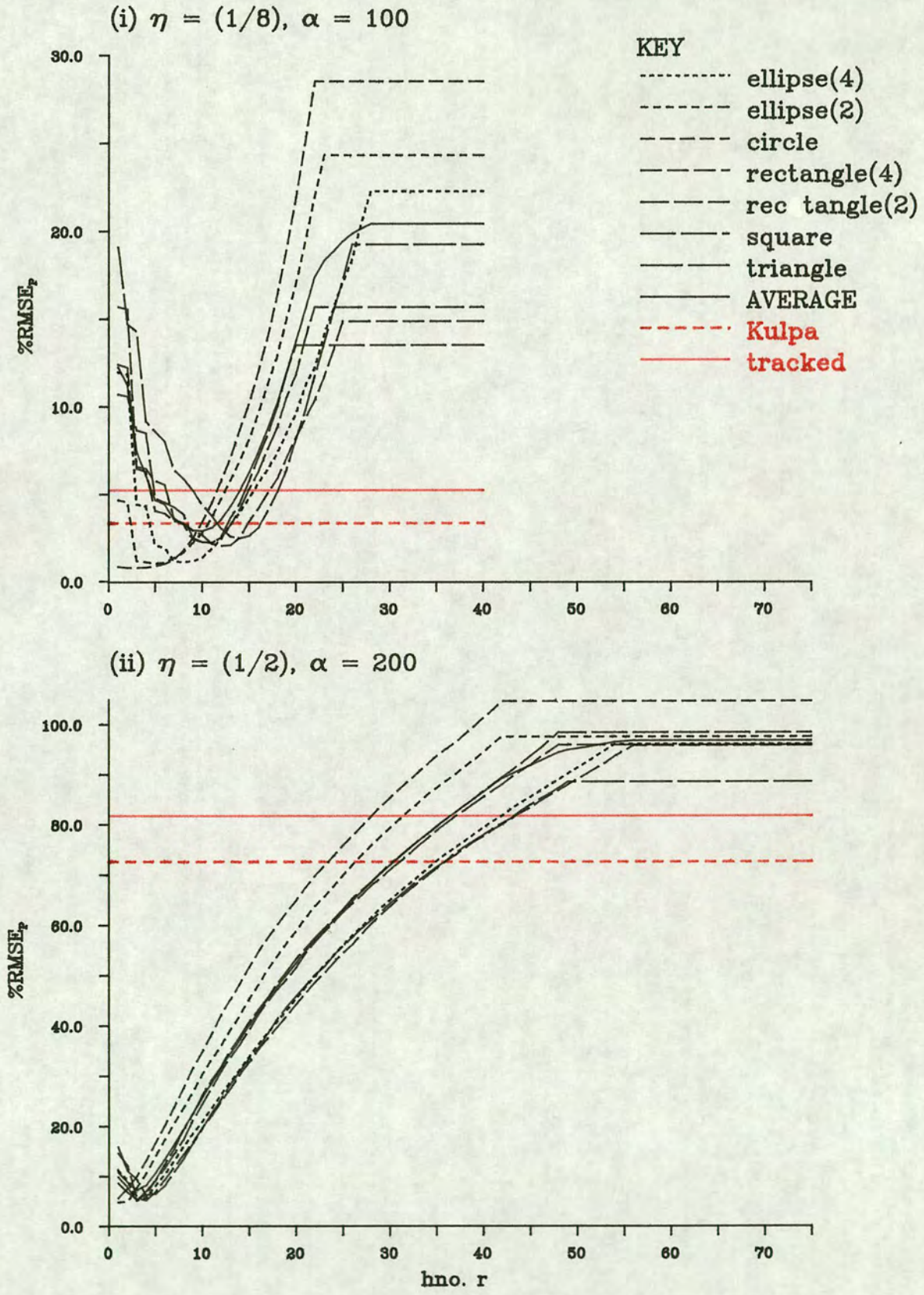


Table 6.2: Perimeter
- Optimal number of coefficients for each geometric shape

Noise η	Area α	Ellipse ρ=4	Ellipse ρ=2	Circle	Rectangle ρ=4	Rectangle ρ=2	Square	Triangle
1/8	50	7	4	3	9	9	8	11
	100	8	5	3	12	11	11	14
	200	11	6	5	15	13	13	19
	400	34	6	4	34	18	17	24
1/4	50	5	3	1	6	6	5	8
	100	6	3	1	8	7	7	10
	200	7	3	2	11	9	9	12
	400	24	5	1	24	13	11	16
1/2	50	2	1	1	3	1	1	2
	100	3	1	1	3	2	1	4
	200	3	1	1	4	3	3	4
	400	8	2	1	9	5	5	7

The straight edges and corners of these shapes are inconsistent with the underlying Fourier model of continuously curving edges, so it is to be expected that some higher harmonics will be required to estimate the 'sharp' changes of direction.

Such trends in the relative number of coefficients for the seven geometric shapes are repeated in Table 6.3, which shows the range of the number of coefficents which are 'near optimal'. The near optimal range is defined as the range of the number of coefficents for which the $\%RMSE_P(r)$ for each geometric shape is no more than 25% greater than the minimum $\%RMSE7_P(r^*)$ at the optimal r^* when averaging over the seven geometric shapes. As before, circles require least coefficients and rectangles the most. In all cases the range covers the optimal r^* for each geometric shape (Table 6.2). More importantly, the range also covers the value of r^* when averaging over the seven shapes (Table 6.1) indicating that the average value is reasonably robust to changes in the type of shape over which results are averaged. This is advantageous for the MRI data, since any image region of interest is not a precise geometric shape which is known a priori.

6.5.2 Area

In Table 6.4, results are summarised for the 12 combinations of the levels of the fixed $\{\alpha,n\}$ factors, averaging over the $g = 7$ geometric shapes. The optimal r^* and the minimum $\%RMSE7_A(r^*)$ are tabulated for the FD fitted estimator $P(r)$, together with the $\%RMSE_A(t)$ for the area estimate calculated from the tracked points. Unlike the results for the perimeter estimators (Table 6.1) where the estimator from the fitted FS is clearly superior, for area there is little difference between $RMSE_A$ values for $A(r^*)$ or $A(t)$, though the difference does increase slightly with n. Since there is such a negligible improvement in area estimation from fitting a FS it does pose the question of whether the additional computation required to calculate the FS is worthwhile, or whether the estimate of area $A(t)$ from the tracked edge points is sufficiently accurate.

Table 6.3: Perimeter
- Range of number of Fourier descriptors for which %RMSE is no more than 25% greater than the minimum

Noise η	Area α	Ellipse ρ=4	Ellipse ρ=2	Circle	Rectangle ρ=4	Rectangle ρ=2	Square	Triangle
1/8	50	1-3	1	1	1-3	1-2	1-2	1-3
	100	1-4	1-2	1	2-5	1-3	1-3	2-5
	200	3-5	1-3	1-2	3-5	2-4	1-4	4-6
	400	6-10	1-5	1-5	5-11	3-8	2-6	4-9
1/4	50	3-7	1-5	1-5	3-9	3-7	4-7	7-9
	100	3-9	2-7	1-6	5-11	5-9	5-9	8-11
	200	5-12	3-9	1-9	6-14	7-12	7-11	10-13
	400	19-30	1-16	1-14	15-31	5-20	5-18	8-23
1/2	50	4-10	1-8	1-7	5-12	7-10	5-10	9-13
	100	5-13	3-10	1-10	7-15	8-13	9-12	12-16
	200	5-18	3-14	1-13	11-20	9-17	10-16	16-22
	400	26-40	1-24	1-22	24-41	7-28	5-26	11-33

Table 6.4: Area
- Optimal number of Fourier descriptors with resulting %RMSE7 for Fourier descriptor estimator and corresponding %RMSE for Kulpa and tracked estimators

Noise η	Area α	Optimal No. of FD r^*	Minimum $\%RMSE7_{A(r^*)}(r^*)$	Tracked $\%RMSE_{A(t)}(t)$
1/8	50	16	3.2	3.2
	100	25	1.7	1.7
	200	39	1.1	1.1
	400	5	3.8	3.8
1/4	50	7	4.9	5.0
	100	10	2.7	2.8
	200	17	1.7	1.7
	400	7	3.8	3.8
1/2	50	10	20.9	21.6
	100	18	13.0	13.6
	200	23	7.6	8.0
	400	28	4.8	5.0

To a certain extent there seems to be some interaction between the $\{\alpha,\eta\}$ factors as the levels of each are increased. Patterns in the optimal r^* are not as strong or as consistent as for the perimeter results. The value of r^* generally increases as the area α increases, though r^* does decrease again for the two combinations $\{\alpha = 400,\ \eta = 1/8\}$ and $\{\alpha = 400,\ \eta = 1/4\}$. The general increase in r^* with area is reasonable, since more points lie on the edges of larger regions. There is an increase in the maximum R and a proportional increase in the minimum, that is the optimal r^*, the number required to describe the global shape. One explanation for the two exceptions is that for such large areas the small irregularities in the edge at low noise levels have negligible effect and the global shape is captured by the first few harmonics. However for $\alpha = 400$ at the highest noise level $\eta = 1/2$, the change in area due to any 'irregularities' in the edge must be accounted for by fitting higher harmonics.

As the noise increases, the optimal r^* decreases from $\eta = 1/8$ to $\eta = 1/4$, so there is more smoothing of the noisier edge. For $\eta = 1/2$, r^* increases again, suggesting that any area which is omitted from or added to the main region because of 'indentations' of the noisy edge is now accounted for by fitting some higher harmonics to describe more local detail.

Considering the seven geometric shapes individually, the $\%RMSE_A$ function is plotted for two examples in Figure 6.4b. Particularly for example *(i)* at low noise $\eta = 1/8$, it is easy to see that in terms of $\%RMSE_A$ there is little improvement when fitting a FS as opposed to using an estimate based on the tracked points. The optimal r^* for the fitted estimator $A(r^*)$ is given in Table 6.5 for each of the seven geometric shapes. The trends in r^* for the different shapes are similar to those for the perimeter (Table 6.2) though the values of r^* are generally higher and in some cases (indicated by +) the optimal r^* is equivalent to R, the maximum number which can be fitted.

The range of the number of coefficents which are 'near optimal', that is the $\%RMSE_A$ is no more than 25% greater than the minimum

Figure 6.4b: Area %RMSE$_A$ for increasing number of harmonics

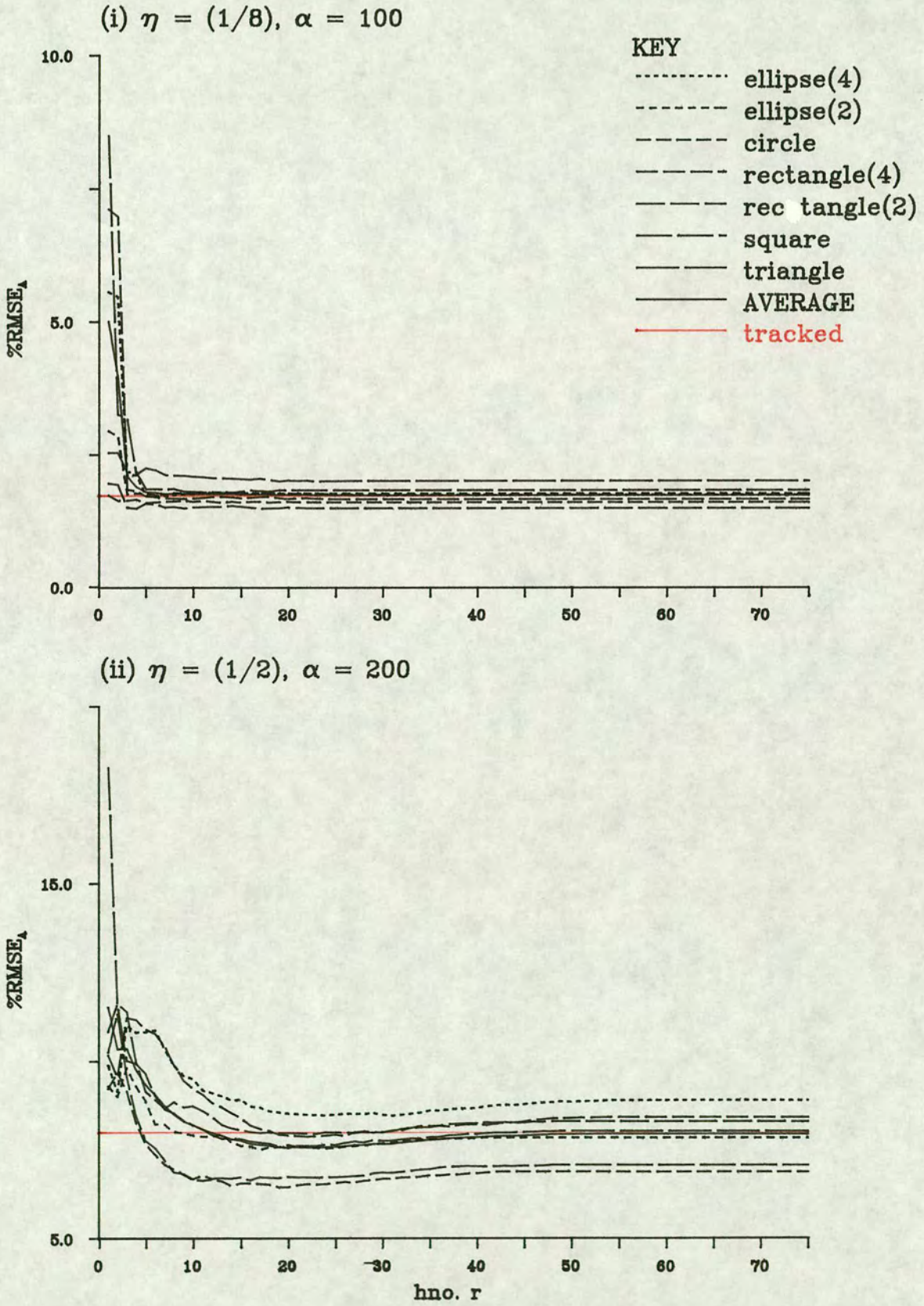


Table 6.5: Area
- Optimal number of coefficients for each geometric shape

Noise η	Area α	Ellipse ρ=4	Ellipse ρ=2	Circle	Rectangle ρ=4	Rectangle ρ=2	Square	Triangle
1/8	50	19+	3	2	14	12	7	16
	100	7	6	5	25+	17	12	19
	200	39+	9	3	37+	10	26	31
	400	12	7	6	5	45+	29	4
1/4	50	14	7	5	11	8	8	10
	100	12	10	11	10	12	9	9
	200	17	14	16	17	11	11	17
	400	7	25	18	6	19	16	22
1/2	50	10	10	8	8	10	10	12
	100	19	14	12	18	17	15	15
	200	32	17	20	22	21	16	20
	400	57	30	28	52	27	26	31

+ Note: optimal number r^* is equal to R, the maximum number which can be fitted

$\%RMSE7_A(r^*)$, are given in Table 6.6. Except for a few cases (marked /) where none of the $\%RMSE_A$ fall within the acceptable range, for most of the other combinations of factor levels almost any number of harmonics r can be classed as near optimal. Thus for area the precise number of terms to be included in the FS is not critical, and using a non-optimal value will only result in a small increase in $\%RMSE_A$.

6.5.3 Conclusions

The use of a reduced number of FD works well when estimating the perimeter length of a pixellated edge enclosing a region in a binary image. The results in Table 6.3 show that the method is reasonably robust when estimating the perimeter for different geometric shapes, so it would be reasonable to extend the results to estimate the perimeter length of an irregular region in the image. Provided some estimate of typical area and noise level is available then the optimal number of coefficients to be used when estimating the fitted edge can then be taken from Table 6.1. In Figure 6.5 the tracked edge is compared with the fitted edge including r^* harmonics in the FS. For this particular example of a shape of area $\alpha = 200$ and $\eta = \frac{1}{4}$, a value of $r^* = 9$ can be read off from Table 6.1.

Only a slight improvement in the accuracy of the area estimate is achieved when using a reduced number of FD rather than the tracked points. If only an estimate of area is of interest, and not the fitting of a model to describe the edge, then the additional computation required to fit a FS appears unnecessary: the area can be estimated with sufficient accuracy using the tracked coordinates. However since it will often be the case that a FS will be fit to the edge points in order to calculate other geometric properties, then for consistency it may be preferable to calculate all properties of interest with respect to the same model of the edge. For example the optimal r^* with respect to perimeter length may be identified from Table 6.1 and the area estimated from the corresponding FS with r^* harmonics. In all cases the ranges of near optimal r^* for area as given in Table 6.6 include the optimal r^* for perimeter (Table 6.1). There will be negligible difference in the estimate of the area

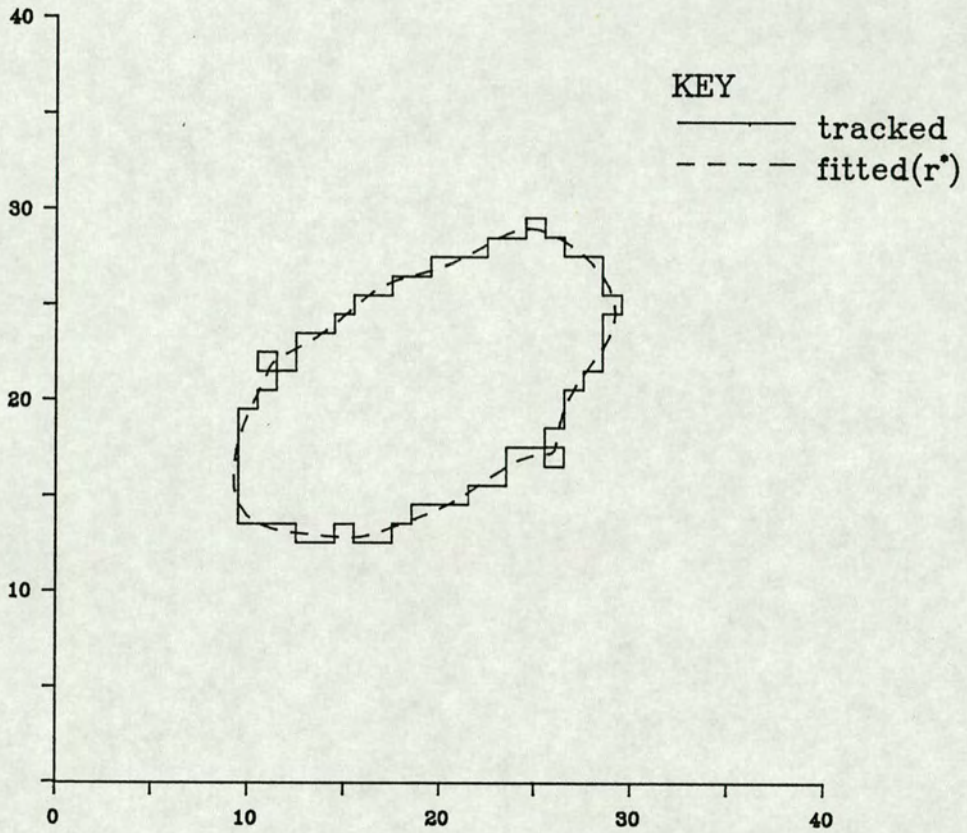
Table 6.6: Area
- Range of number of Fourier descriptors for which %RMSE is no more than 25% greater than the minimum

Noise η	Area α	Ellipse ρ=4	Ellipse ρ=2	Circle	Rectangle ρ=4	Rectangle ρ=2	Square	Triangle
1/8	50	$^{+}$3-	1-	1-	3-	2-	1-	4-
	100	3-	3-	1-	3-	1-	5-	7-
	200	8-	3-	1-	3-	3-	5-	7-
	400	$^{\oplus}$/	1-	1-	/	1-	1-	2-
1/4	50	3-	1-	1-	3-	2-	1-	2-
	100	3-	2-	1-	3-	2-	3-	4-
	200	5-	3-	1-	3-	4-	3-	4-
	400	/	1-	1-	/	1-	1-	2-
1/2	50	2-	1-	1-	1-	1-	1-	2-
	100	2-	2-	1-	7-	1-	3-	3-
	200	11-	4-	3-	9-	5-	4-	6-
	400	6-	4-	4-	5-	7-	5-	11-

$^{+}$ Note: n- ($n \in \mathbb{Z}$) indicates that all numbers of Fourier descriptors $\geqslant$n fall within required range

$^{\oplus}$ Note: / indicates that none of the numbers of Fourier descriptors fall within required range

Figure 6.5: Comparison of tracked edge with edge from fitting first 9 Fourier descriptors
Ellipse: $\rho = 2$, $\alpha = 200$, $\eta = \frac{1}{4}$



which is virtually constant over all r, as can be seen in Figure 6.4b.

As can be seen from Figure 6.5, using a reduced number of coefficients in a FS fitted to points lying on a pixellated edge in a thresholded image does produce a smooth representation of the edge. This method is much quicker than smoothing the whole image and works well when noise levels are moderate. At higher noise levels the thresholded edge will be too poor a representation and more computer-intensive methods will be necessary, for example smoothing the whole image (with the LoG function) and identifying edge points as positions of zero-crossings (Section 3.5). Neither fitting a reduced number of FD or smoothing the whole image are applicable methods if the original region is not smooth, since roughness is confounded with noise in a pixellated image. If objects are of a known shape, for example circular, then methods such as template matching or deriving perimeter length as a function of area may be expected to perform better.

7 DEFINITION OF AN OPTIMUM CENSORING THRESHOLD USING THE DISTRIBUTION OF INTERVAL LENGTHS

7.1 Introduction

The ordered set of points $\{(x_i,y_i) : i = 1,2,...,n\}$ considered in the previous two chapters were identified as points of intersection of a tissue edge with the discrete grid or lattice (Section 3.5). The edge can be modelled as a closed curve by fitting Fourier descriptors (FD) (5.1) to this set of points, as discussed in Chapter 5. The analysis of this chapter is motivated by the need to define a variable with which to label consecutive points. One option is the index number i (Section 5.3.2), corresponding to the relative position of each point in the ordered sequence round the edge. With this labelling points are not equally spaced round the curve, in the sense that the 'interval' lengths, that is the Euclidean distance between each pair of consecutively labelled points, are not equal. By omitting some points, the variation in the lengths of the intervals calculated between the remaining points may be reduced and consequently there may be a reduction in the number of FD required to produce a good fit. Points are censored if the distance between them is less than a prespecified censoring threshold length, and subsequent points are reindexed.

The choice of an optimum censoring threshold is considered in this chapter, see also Wheelwright and Glasbey (1993). The analysis examines the relationship between computer representations of an image, that is data on a lattice, and the underlying true images which exist in continuous space. Several properties of random lines in the plane are derived by Kendall and Moran (1963), but none of the results solve the censoring problem considered in this chapter. There are also connections with stereology: see Coleman (1969).

Rather than looking at intervals between edge points on a curve, an assumption of local linearity is made in order to simplify the analysis. This assumption seems reasonable for most of the image edges, but is not adequate in regions of high curvature of an edge, as will be discussed in Section 7.5. Attention is restricted to the integer lattice in the x-y plane. A random smooth path of infinite

length is simulated and intersects the square lattice. The simulated data approximates an image edge considered within a local neighbourhood of two or three pixels. Defining it to be smooth implies that at this scale the path can be considered to be straight. An assumption of randomness implies that the path exhibits no long term dependencies and that short sections of the path have isotropic uniform randomness (IUR). By IUR it is meant that the short sections of the path have no preferred direction and that the slope of each short section is a random variable uniformly distributed over the range of angles $(0,2\pi]$.

The path is tracked in a single direction, and the length of an interval calculated as the Euclidean distance between each pair of consecutive intersections of the path with the lattice. An intersection point is censored if it lies within a threshold distance of the preceeding uncensored intersection. In this chapter the distribution is derived of the length of the intervals between consecutive intersections after censoring, for several censoring threshold values, t say. From this, the theoretically optimal threshold which minimises the variability in the distribution of interval lengths can be defined. Using the analysis of this chapter as a starting point Coleman (1992) considers a modified censoring algorithm. By adopting a different rule for selecting which intersections are censored, he derives an exact expression for the distribution of interval lengths. Since the distribution derived is conditional on the censoring algorithm, his results are not directly comparable with those derived in this chapter.

The notation used in the derivation of the distribution will be introduced in Section 7.2.1, and the random path defined. The simulation of a path and the censoring of intervals will be discussed in Section 7.2.2. The distribution of interval lengths is derived in two stages. Section 7.3 covers the distribution conditioning on the slope of the interval. In Section 7.4 the conditioning is removed and the required distribution is derived by numerical integration. From this distribution summary statistics can be computed for a given censoring threshold t in the range [0.0,1.0], where 1.0 is the width of a unit pixel on the (integer) lattice. The results are discussed

in Section 7.5. This includes a description of the distribution for thresholds $t > 1.0$. It also considers the validity of extending the results to intervals between points on an image edge, rather than between points on the 'ideal' straight line. Finally, an optimal censoring threshold is proposed.

7.2 The random path

7.2.1 Introduction and notation

To allow logical use of notation, some of the variables used in Chapter 7 have different definitions to those given elsewhere in this thesis. Wherever a variable has been defined previously and reused for a different purpose in this chapter, a new definition will be given. Since this chapter is to some extent independent of the main problem of describing edges, there should be no confusion between the two different definitions.

As already stated, the aim is to derive the distribution of the lengths of intervals between consecutive intersections, after censoring. Without loss of generality only a short section of the path given by:

$$y = (x-s)\tan\theta \tag{7.1}$$

need be considered. The section (7.1) starts in the interval [0,1) on either the x- or y- axis, and the slope θ is a random variable uniformly distributed in the range $(\pi/4,\pi/2]$. The x-intercept s is also a random variable, whose distribution depends on θ and the censoring threshold, t. Values of the threshold in the range [0,1] will be considered in detail; the distribution for thresholds of $t > 1.0$ will be outlined in Section 7.5.

The distribution is derived in two stages, covered in the next two sections. In the first stage the distribution is derived conditional on θ, the slope of the interval. Assuming θ is fixed then for IUR and in the absence of censoring, the intersection random

variable s is uniformly distributed between $-\cot\theta$ and 1.0. The effect on the distribution of s when censoring small intervals of length z say, less than t, is not immediately obvious. Therefore several random paths were simulated to gain information on the typical shape of the distribution of the censored intervals. The simulations are described in detail in the next section, and observations made about qualitative aspects of the distribution and the effect of censoring. Using these observations as a starting point, an analytic expression for the distribution of interval lengths, based on these simulations, is derived in Sections 7.3 and 7.4.

7.2.2 The simulated distribution

To look at the effect of censoring on the distribution of intervals, a total of 25 straight lines were simulated to model the theoretical random path. Lines were simulated for all 25 combinations of five values of the censoring threshold $\{t = 0.2, 0.4, 0.6, 0.8, 1.0\}$ and five values of the slope random variable $\{\theta = 46.1^{\circ}, 56.1^{\circ}, 66.1^{\circ}, 76.1^{\circ}, 86.1^{\circ}\}$. The choice of these particular angles θ is discussed later.

The simulation of a single line for any $\{\theta, t\}$ pair is now described. The 'start-point' of each line in the unit square is fixed to lie on the $y = 0$ axis, but with random x-intercept s. Since it is not possible to simulate a line of infinite length, the length of each path is fixed at 1500 units, where the units are measured with respect to the integer lattice. This finite length is deemed sufficient to simulate the long-term behaviour of a random path. Intersections of the line with the lattice are identified and the first 50 intersections systematically omitted. This is to avoid any 'start-up' effect which may occur as a result of defining the line from the $y = 0$ axis, and is equivalent to defining a line starting from $y = -\infty$, but still with random x-intercept s.

The interval length z is calculated as the Euclidean distance between each pair of intersections of the line with the lattice. The relative position (along the line) of an interval is unimportant, so

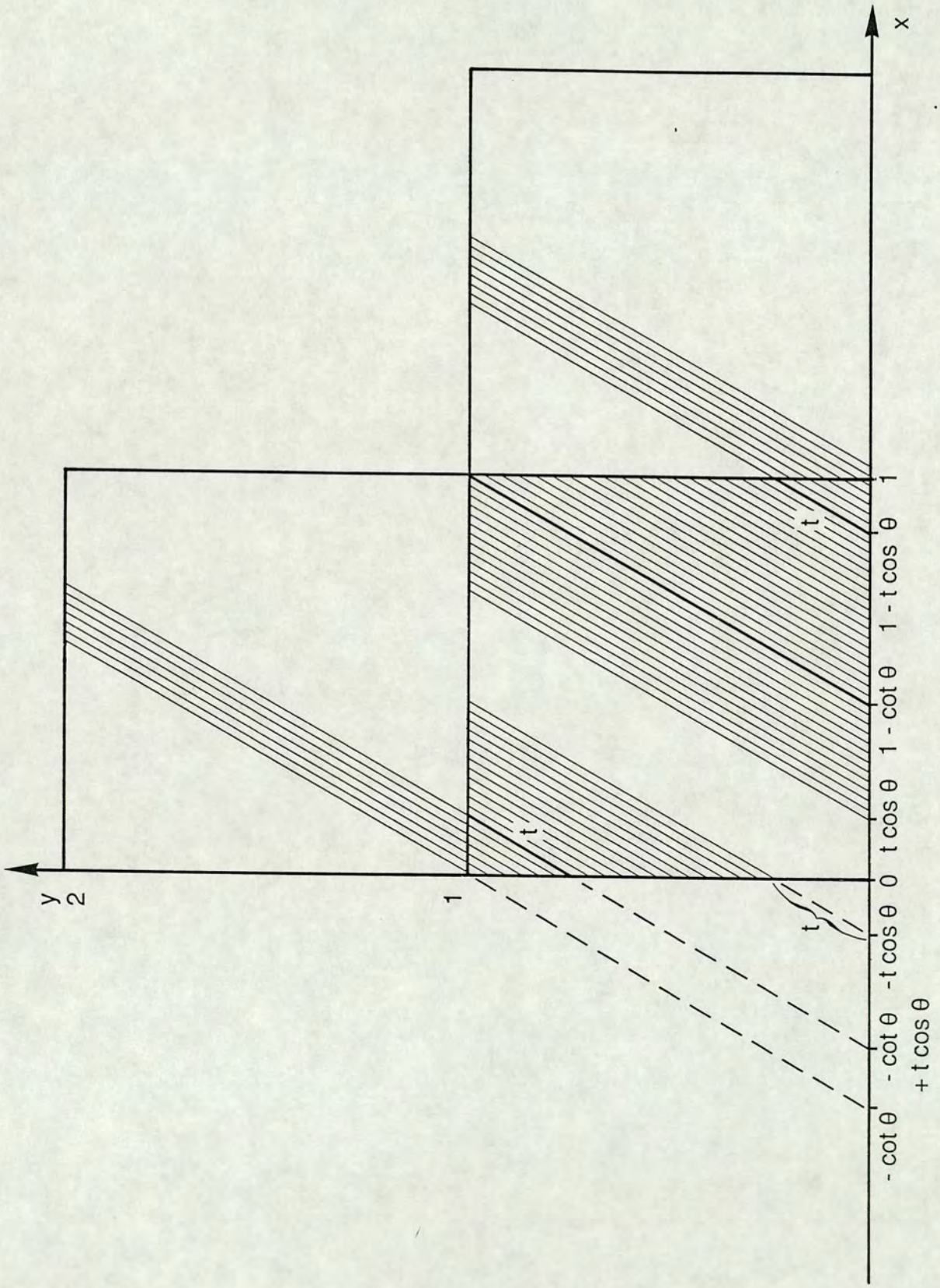
without loss of generality each interval can be repositioned with respect to the unit square. Thus each interval is now of the form (7.1), starting in the interval [0,1) on either the x- or y-axis. In the absence of censoring, the intersection random variable s is uniformly distributed over the range (-cotθ,1).

The values of θ {46.1°,56.1°,...,86.1°} were chosen to cover the range of interest. An attempt was also made to obtain the best illustration that intersections s were in fact uniformly distributed over this range. In preliminary simulations with an angle of θ of say 46°, a 'cyclic' pattern was observed in the lengths of consecutive intervals. This meant that the full range of theoretically possible interval lengths could not be generated. However this cyclic pattern is not a true feature of the distribution, but is a consequence of the method used to generate the intervals. After allowing for the cyclic pattern, checks were made and it was concluded that an assumption of a uniform distribution of intersections s was valid. In subsequent simulations the angles of θ were perturbed slightly, using θ = 46.1° say. This has the desired effect of increasing the 'cycle length' before the pattern of intervals was repeated, and so minimising the effect, on the distribution to be derived, of repeated interval lengths.

The effect of censoring is now considered. Suppose intersections are denoted by (x_i,y_i), labelled consecutively by the subscript i, for {i = 1,2,...}. For those intersections (x_i,y_i) where z is less than the specified censoring threshold t, then the current intersection (x_i,y_i) is censored, and a new cumulative distance calculated between the previous intersection (x_{i-1},y_{i-1}) and (x_{i+1},y_{i+1}), the next intersection along the path. For thresholds in the range $t \in [0.0,1.0]$ censoring a single intersection will always result in a new (cumulative) interval, z' say, which is greater than t.

The typical distribution of censored intervals (after repositioning), as shown by the simulations, is illustrated in Figure 7.1. Note that to avoid obscuring any fine detail only $(1/5)^{th}$ of all such intervals are plotted, since lines begin to merge if all

Figure 7.1: Censored intervals for case (1a)



intervals are shown. As can be seen from Figure 7.1, censoring intersections where the distance z from the previous intersection is less than t means that such intervals are extended into the next square above or to the right. The algorithm for censoring points which is implemented here ensures that 'double counting' of intervals is avoided. Intervals from the previous square (below or to the left) which are shorter than t and so should be extended, are now omitted. This produces the gap shown in Figure 7.1. The results of the simulations led to the development of a theoretical model for interval lengths, as is discussed in the next two sections.

7.3 Distribution conditional on θ

7.3.1 Conditional probability density function

Before deriving the distribution it is useful to establish some notation. The indicator function I is defined as:

$$I(z|a,b) = \begin{cases} 1 & \text{if } z \in (a,b) \\ 0 & \text{otherwise.} \end{cases}$$

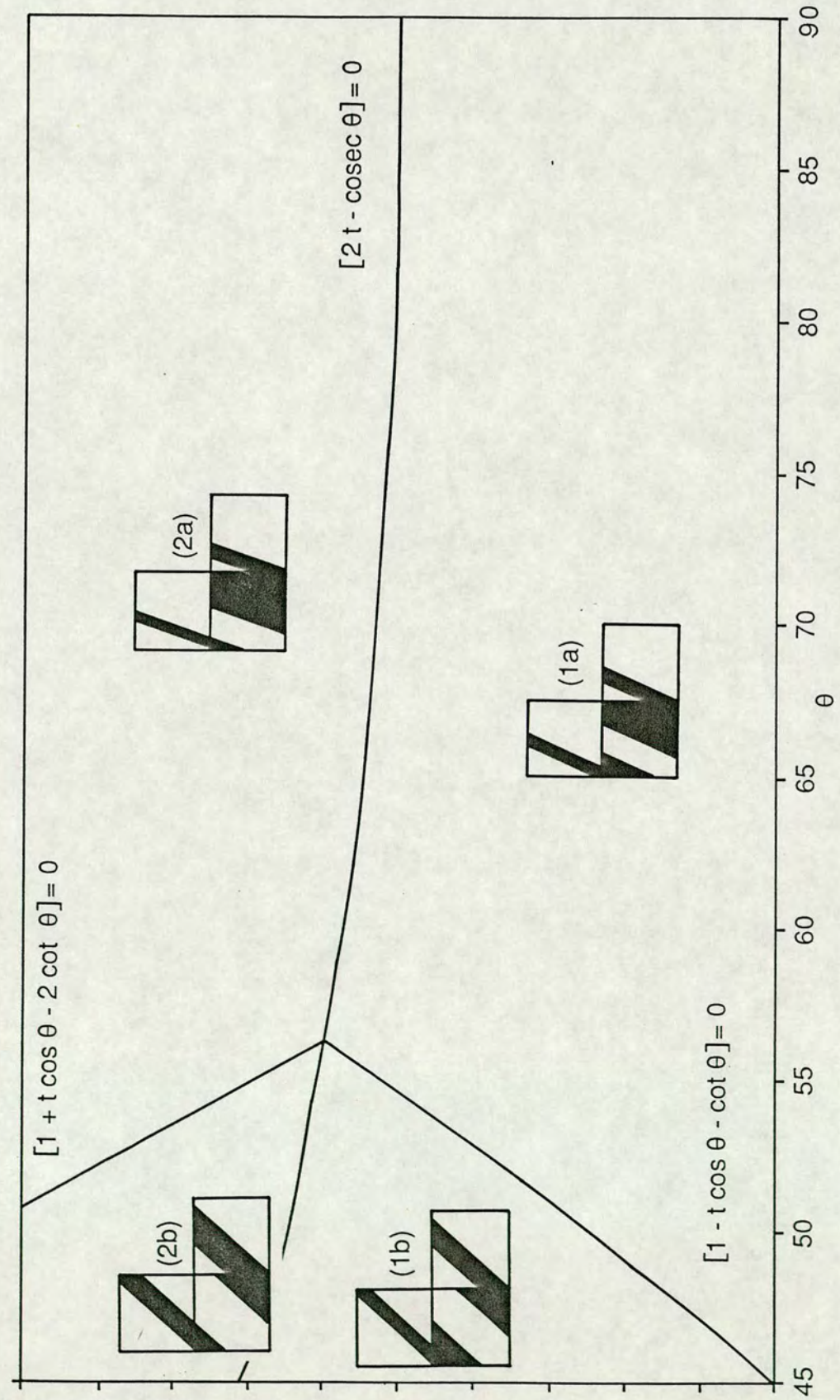
The Dirac delta function δ is defined such that:

$$\delta(z) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \; I(z|-\epsilon,\epsilon).$$

With respect to a given value of t the probability density function (PDF) of z, conditional on θ, is $f_t(z|\theta)$.

Based on the simulations of Section 7.2.2, four distinct cases can be identified depending on the combination of $\{\theta,t\}$ values used. In Figure 7.2 the $\{\theta,t\}$ sample space is partitioned into four regions and examples are given of the typical shape of the distribution of intervals in each region. From Figure 7.2 it can be seen that there are two 'main' cases: case (1) and case (2), each of which is further subdivided into parts (a) and (b). The division between the two main cases is determined by the value of t (conditional on θ). If $2t \leqslant \operatorname{cosec}\theta$ then case (1) holds and intersections s are uniformly

Figure 7.2: Partition of {θ,t} sample space showing four possible cases of distribution of censored intervals



distributed in the ranges (-cotθ, -tcosθ) and (tcosθ, 1.0). Otherwise if $2t \geqslant \text{cosec}\theta$, case (2) is valid and the ranges of intersections s are (-cotθ, -tcosθ) and (cotθ-tcosθ, 1.0). Note that for equality, that is if $2t = \text{cosec}\theta$, the intervals for cases (1) and (2) are identical.

Given $2t \leqslant \text{cosec}\theta$, a subdivision of case(1) occurs if $t\cos\theta \leqslant 1-\cot\theta$. This results in the distribution of intervals shown in Figure 7.1 and is denoted case (1a). Otherwise if $t\cos\theta \geqslant 1-\cot\theta$, case (1b) holds. Alternatively if $2t \geqslant \text{cosec}\theta$ there is a similar subdivision of case (2): if $\cot\theta-t\cos\theta \leqslant 1-\cot\theta$ then case (2a) holds, otherwise case (2b) is valid. Again if there is equality, that is $t\cos\theta = 1-\cot\theta$ for case (1), or $\cot\theta-t\cos\theta = 1-\cot\theta$ for case (2), then the intervals are identical for cases (1a), (1b) and cases (2a), (2b) respectively. The derivation of the PDF will be described in some detail for the first case. The other 3 cases follow similarly and only the summary form of the PDF will be given.

Recall that for case (1), s is uniformly distributed in the ranges (-cotθ, -tcosθ) and (tcosθ, 1.0). A scaling probability, C_1 say, is required in order that the total probability over the two ranges of s is equal to unity. That is:

$$C_1\{-t\cos\theta-(-\cot\theta)\} + C_1(1-t\cos\theta) = 1,$$

which is simplified and rearranged to:

$$C_1 = \frac{1}{1+\cot\theta-2t\cos\theta}.$$

Dividing through by cosθ gives:

$$C_1 = \frac{\sec\theta}{\sec\theta+\text{cosec}\theta-2t}.$$

A subdivision of case (1) occurs if $t\cos\theta \leqslant 1-\cot\theta$, when case (1a) of Figure 7.1 is valid. For each sub-range of intersections s it is possible to associate corresponding ranges of z, which are observed when s takes a given value. Thus the probability of observing any given range of interval lengths z is equal to the probability of observing the corresponding range of intersections s.

For example, intersections in the ranges $s \in (-\cot\theta+\tcos\theta, -\tcos\theta)$ and $s \in (1-\cot\theta+\tcos\theta, 1-\tcos\theta)$ both yield values of $z \in (t, \operatorname{cosec}\theta-t)$. Let $\tilde{p}$ be the probability of observing values of s in this sub-range. Then:

$$\tilde{p} = C_1\{(-\tcos\theta+\cot\theta-\tcos\theta)+(1-\tcos\theta-1+\cot\theta-\tcos\theta)\} = C_1\{2\cot\theta-4\tcos\theta\}$$

Simplifying and substituting for C_1 gives $\tilde{p} = \dfrac{2(\operatorname{cosec}\theta-2t)}{\sec\theta+\operatorname{cosec}\theta-2t}$.

This is the total probability for observing a value of z in the range $(t, \operatorname{cosec}\theta-t)$. Applying the mean value theorem (see Swokowski, 1979), to derive the PDF it is necessary to divide the probability $\tilde{p}$ by the 'width' of the range of z values, that is $\operatorname{cosec}\theta - 2t$. Thus using the indicator function I, the PDF (denoted by $\tilde{f}$) for the sub-range of $z \in (t, \operatorname{cosec}\theta-t)$ can be expressed as:

$$\tilde{f}_t(z|\theta) = 2I(z|t,\operatorname{cosec}\theta-t)\left\{\frac{1}{\sec\theta+\operatorname{cosec}\theta-2t}\right\}.$$

The PDF is derived similarly for $z \in (\operatorname{cosec}\theta-t, \operatorname{cosec}\theta+t)$ which results from intersections in the ranges $s \in (1-\cot\theta, 1-\cot\theta+\tcos\theta)$ and $s \in (-\cot\theta, -\cot\theta+\tcos\theta)$.

Intersections s in the ranges $s \in (\tcos\theta, 1-\cot\theta)$ and $s \in (1-\tcos\theta, 1)$ result in intervals of length $z = \operatorname{cosec}\theta$, with probability:

$$\tilde{p} = \frac{\sec\theta-\operatorname{cosec}\theta}{\sec\theta+\operatorname{cosec}\theta-2t}.$$

The PDF is expressed in terms of the Dirac delta function as:

$$\tilde{p}_t(z|\theta) = \delta(z-\operatorname{cosec}\theta)\left\{\frac{\sec\theta-\operatorname{cosec}\theta}{\sec\theta+\operatorname{cosec}\theta-2t}\right\}.$$

Bringing together the expressions for the PDF for the sub-ranges of intervals z, then for case (1a) the PDF of z, conditional on θ is:

$$f_t(z|\theta) = \frac{1}{\sec\theta+\operatorname{cosec}\theta-2t}\left\{2I(z|t,\operatorname{cosec}\theta-t) + I(z|\operatorname{cosec}\theta-t,\operatorname{cosec}\theta+t)\right.$$

$$\left. + (\sec\theta-\operatorname{cosec}\theta)\ \delta(z-\operatorname{cosec}\theta)\right\}. \qquad (7.2a)$$

The probability density functions of z, conditional on θ for the other 3 cases are derived similarly, considering the ranges of possible interval lengths z and the intersections s from which they result. If $t\cos\theta \geqslant 1-\cot\theta$ then case (1b) holds and:

$$f_t(z|\theta) = \frac{1}{\sec\theta+\operatorname{cosec}\theta-2t}\left\{2I(z|t,\operatorname{cosec}\theta) + I(z|\operatorname{cosec}\theta-t,\sec\theta-t)\right.$$

$$\left. + I(z|\operatorname{cosec}\theta,\sec\theta) + t\ \delta(z-\operatorname{cosec}\theta) + (\operatorname{cosec}\theta-\sec\theta+t)\ \delta(z-\sec\theta)\right\}.$$

(7.2b)

If $2t \geqslant \operatorname{cosec}\theta$, case (2) holds. A subdivision is made if $\cot\theta-t\cos\theta \leqslant 1-\cot\theta$ when case (2a) holds and the PDF of z conditional on θ is given by:

$$f_t(z|\theta) = \cos\theta\{I(z|t,2\operatorname{cosec}\theta-t) + (\sec\theta-2\operatorname{cosec}\theta+2t)\ \delta(z-\cos\theta)\} \qquad (7.2c)$$

Otherwise when $\cot\theta-t\cos\theta \geqslant 1-\cot\theta$, case (2b) holds and:

$$f_t(z|\theta) = \cos\theta\{I(z|t,t+\sec\theta-\operatorname{cosec}\theta) + t\ \delta(z-\operatorname{cosec}\theta)$$

$$+ I(z|\operatorname{cosec}\theta,\sec\theta) + (2\operatorname{cosec}\theta-\sec\theta-t)\ \delta(z-\sec\theta)\}. \qquad (7.2d)$$

7.3.2 Conditional expectation

The conditional expectation is required as a weighting factor when removing the conditioning on θ, as will be shown in Section 7.4.1. For the four cases analytic expressions can be derived for $E_t(z|\theta)$, the expected value of interval length z, conditional on θ. The derivation is shown in detail for case (1a); analogous expressions for the other three cases follow. Looking at each sub-range of z, as defined in the PDF (7.2a), the expected value of z with respect to that range is evaluated and multiplied by $p_t(z|\theta)$, the probability of observing a value of z within that range. The

conditional expectation is evaluated by summing over all sub-ranges.

If $E_t(z \in a,b|\theta)$ denotes the conditional expected value of z given $z \in (a,b)$ then for case (1a):

$$E_t(z|\theta) = \Big\{ E_t(z \in t,\operatorname{cosec}\theta - t|\theta) \times p_t(z|\theta) I(z|t,\operatorname{cosec}\theta - t) +$$

$$E_t(z \in \operatorname{cosec}\theta - t,\operatorname{cosec}\theta|\theta) \times p_t(z|\theta) I(z|\operatorname{cosec}\theta - t,\operatorname{cosec}\theta + t) +$$

$$E_t(\delta(z - \operatorname{cosec}\theta)|\theta) \times (\sec\theta - \operatorname{cosec}\theta) \Big\}.$$

Substituting the expectations and probabilities for each sub-range, first taking out a common multiplying factor:

$$E_t(z|\theta) = \frac{1}{\sec\theta + \operatorname{cosec}\theta - 2t} \Big\{ \operatorname{cosec}\theta \tfrac{1}{2} \quad (\operatorname{cosec}\theta - 2t) + 2\operatorname{cosec}\theta \times \frac{t}{2}$$

$$+ \operatorname{cosec}\theta(\sec\theta - \operatorname{cosec}\theta) \Big\}.$$

Rearranging and simplifying gives $E_t(z|\theta) = \dfrac{\operatorname{cosec}\theta\sec\theta}{\sec\theta + \operatorname{cosec}\theta - 2t}$, and hence

$$E_t(z|\theta) = \frac{1}{\sin\theta + \cos\theta - 2t\sin\theta\cos\theta}.$$

Expressions for $E_t(z|\theta)$ for the other three cases are evaluated similarly. For case (1b), $E_t(z|\theta)$ is the same as case (1a), while for cases (2a) and (2b), $E_t(z|\theta)$ is equal to $1/\sin\theta$. Thus the expected value of interval length z, conditional on θ can be summarised as:

$$E_t(z|\theta) = \begin{cases} \dfrac{1}{\sin\theta + \cos\theta - 2t\sin\theta\cos\theta} & 2t \leqslant \operatorname{cosec}\theta \\ \dfrac{1}{\sin\theta} & \operatorname{cosec}\theta \leqslant 2t \leqslant 2. \end{cases} \qquad (7.3)$$

7.4 Unconditional distribution of interval length

7.4.1 Introduction

To evaluate the PDF $f_t(z)$ for the distribution of interval

lengths assuming random θ, the conditional distribution $f_t(z|\theta)$ (7.2) is integrated over the range of values of θ considered. A weighting factor is required for the integration, as can be seen by considering the uncensored case, $t = 0$. Suppose lines of equal length L say, are simulated at slopes of θ in the range $(\pi/4, \pi/2)$. For each value of θ a count is made of the number of intersections of the line with the integer lattice. If the number of intersections is plotted against the corresponding value of θ, then the greatest number $L\sqrt{2}$ is observed when $\theta = \pi/4$, the number decreasing monotonically to a minimum of L at $\theta = \pi/2$. This illustrates that depending on the value of θ, some interval lengths z will occur more frequently and so it is necessary to weight the conditional distribution (7.2) by the number of intervals of a given length z.

For a given slope of θ the number of intervals is proportional to $E_t(z|\theta)$, the mean interval length (7.3). For the uncensored case, $E_0(z|\theta)$ is given by $1/(\sin\theta + \cos\theta)$ and the number of intervals is proportional to $1/E_0(z|\theta)$. This argument holds similarly for $E_t(z|\theta)$ after censoring. Thus for any general threshold t the unconditional PDF $f_t(z)$ is given by:

$$f_t(z) = C_t \int_{\pi/4}^{\pi/2} \frac{f_t(z|\theta)}{E_t(z|\theta)}\, d\theta, \tag{7.4}$$

where C_t is a scaling constant such that $\int_0^\infty f_t(z)\, dz = 1$.

Substituting the expression for $f_t(z)$ as given by (7.4) and reversing the order of integration, this is equivalent to requiring that:

$$C_t \int_{\pi/4}^{\pi/2} \frac{1}{E_t(z|\theta)}\, d\theta = 1, \tag{7.5}$$

which after integration and standard algebraic manipulation gives $C_t = 1$, for the special case $t = 0$.

In principle it is possible to combine equations (7.2), (7.3) and (7.4) to derive analytic expressions for $f_t(z)$ for any value of t.

However this rapidly becomes very complicated and so the derivation is given only for the simplest case of uncensored intervals. To emphasise the point that the derived distribution holds only for the special case t = 0, in the following the general subscript t will be replaced by the subscript 0. For t = 0, $E_0(z|\theta)$ is given by $1/(\sin\theta + \cos\theta)$, case (1a) holds and the PDF (7.2a) simplifies to:

$$f_0(z|\theta) = \frac{1}{\sec\theta+\operatorname{cosec}\theta}\left\{2I(z|0,\operatorname{cosec}\theta) + (\sec\theta-\operatorname{cosec}\theta)\delta(z-\operatorname{cosec}\theta)\right\}.$$

In the integration (7.4), it is incorrect to evaluate $f_0(z)$ over the whole range of θ for all values of z since the range of z for which the PDF is valid is dependent on θ. All of the distinct ranges of z, as θ varies in the range $(\pi/4,\pi/2)$, need be considered. The range $z \in (0,\operatorname{cosec}\theta)$ is subdivided and $f_0(z)$ integrated over values of z in the range $0 \leqslant z \leqslant 1$, and over the range $1 \leqslant z \leqslant \operatorname{cosec}\theta$. For the second sub-range there is a change in the integration limits when $z = \operatorname{cosec}\theta$, that is when $\theta = \arcsin(1/z)$. Substituting the appropriate forms of $f_0(z|\theta)$ and $E_0(z|\theta)$ into (7.4), then for $0 \leqslant z \leqslant 1$:

$$f_0(z) = \int_{\pi/4}^{\pi/2} \frac{2(\sin\theta+\cos\theta)}{\sec\theta+\operatorname{cosec}\theta}\, d\theta,$$

which after simplyfying and integrating reduces to:

$$f_0(z) = 2\left[\frac{\sin^2\theta}{2}\right]_{\pi/4}^{\pi/2} = \frac{1}{2}.$$

Similarly for $1 \leqslant z \leqslant \operatorname{cosec}\theta$:

$$f_0(z) = \int_{\pi/4}^{\arcsin(1/z)} \frac{2(\sin\theta+\cos\theta)}{\sec\theta+\operatorname{cosec}\theta}\, d\theta = \frac{1}{z^2} - \frac{1}{2}.$$

Considering $\delta(z-\operatorname{cosec}\theta)$ in the range of $\theta \in (\pi/4,\pi/2)$, z may take values $1 \leqslant z \leqslant \sqrt{2}$. Integrating the delta function gives:

$$f_0(z) = \frac{1}{z^2\sqrt{(z^2-1)}} - \frac{1}{z^2}.$$

Finally summing over $f_o(z)$ in the range $1 \leqslant z \leqslant \sqrt{2}$, for the uncensored case the PDF is summarised as:

$$f_o(z) = \begin{cases} 1/2 & 0 \leqslant z \leqslant 1 \\ \dfrac{1}{z^2\sqrt{(z^2-1)}} - \dfrac{1}{2} & 1 \leqslant z \leqslant \sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

This is one of the results derived by Coleman (1969), the conditions imposed here being equivalent to what he termed μ-randomness, through a rectangle with unit sides. However because of the complexity of deriving analagous expressions for $t > 0$, a numerical approach is favoured and described in the next section.

7.4.2 Derivation of distribution by numerical integration

For general t, the PDF $f_t(z)$ of the distribution of interval lengths is derived by numerical integration. A table of {z,t} values of the cumulative distribution function (CDF) is evaluated. For a given value of t and of the random variable Z, the CDF_t is defined such that $CDF_t(z)$ is equal to the probability that $(Z \leqslant z)$. Approximate probabilities can be derived for each z by calculating first differences of the CDF. Summary statistics can then be extracted from the CDF table and the value of t minimsing the range of intervals identified.

The threshold t is assumed fixed in the following description, though in theory the CDF can be derived for any value of t. The CDF is derived in two stages. In the first stage, the slope θ is assumed to be fixed as when deriving the PDF in Section 7.2, and the conditional function $CDF_t(z|\theta)$ is evaluated. In the second stage the conditioning is removed, summing over values of θ to derive $CDF_t(z)$. Assuming t is fixed, $CDF_t(z|\theta)$ is evaluated on a discrete mesh indexed by z and θ, initially specifying values of z such that $\{0.0 \leqslant z \leqslant 2.0$: increments of $0.01\}$ and values of θ such that $\{45.0^{\circ} \leqslant \theta \leqslant 90.0^{\circ}$: increments of $1^{\circ}\}$. Subsequently the $CDF_t(z|t)$ was evaluated at smaller increments of θ to improve the accuracy of the

approximation to the integrated function.

The conditional function $CDF_t(z|\theta)$ is evaluated for each combination of $\{z,\theta\}$ as follows. The case appropriate for the given $\{\theta,t\}$ combination is identified from the partition diagram (Figure 7.2). The correct sub-range of the PDF for this case is identified for the given z and the cumulative probability tabulated. Once the conditional CDF has been evaluated for all values of $\{\theta,t\}$, θ is allowed to vary. This is equivalent to summing the conditional $CDF_t(z|\theta)$ over θ, for each value of z. The summation is weighted by the conditional expected number of intervals $E_t(z|\theta)$; the form of $E_t(z|\theta)$ appropriate will depend on the given combination of $\{\theta,t\}$, as discussed in Section 7.4.1. Thus the unconditional CDF is given by:

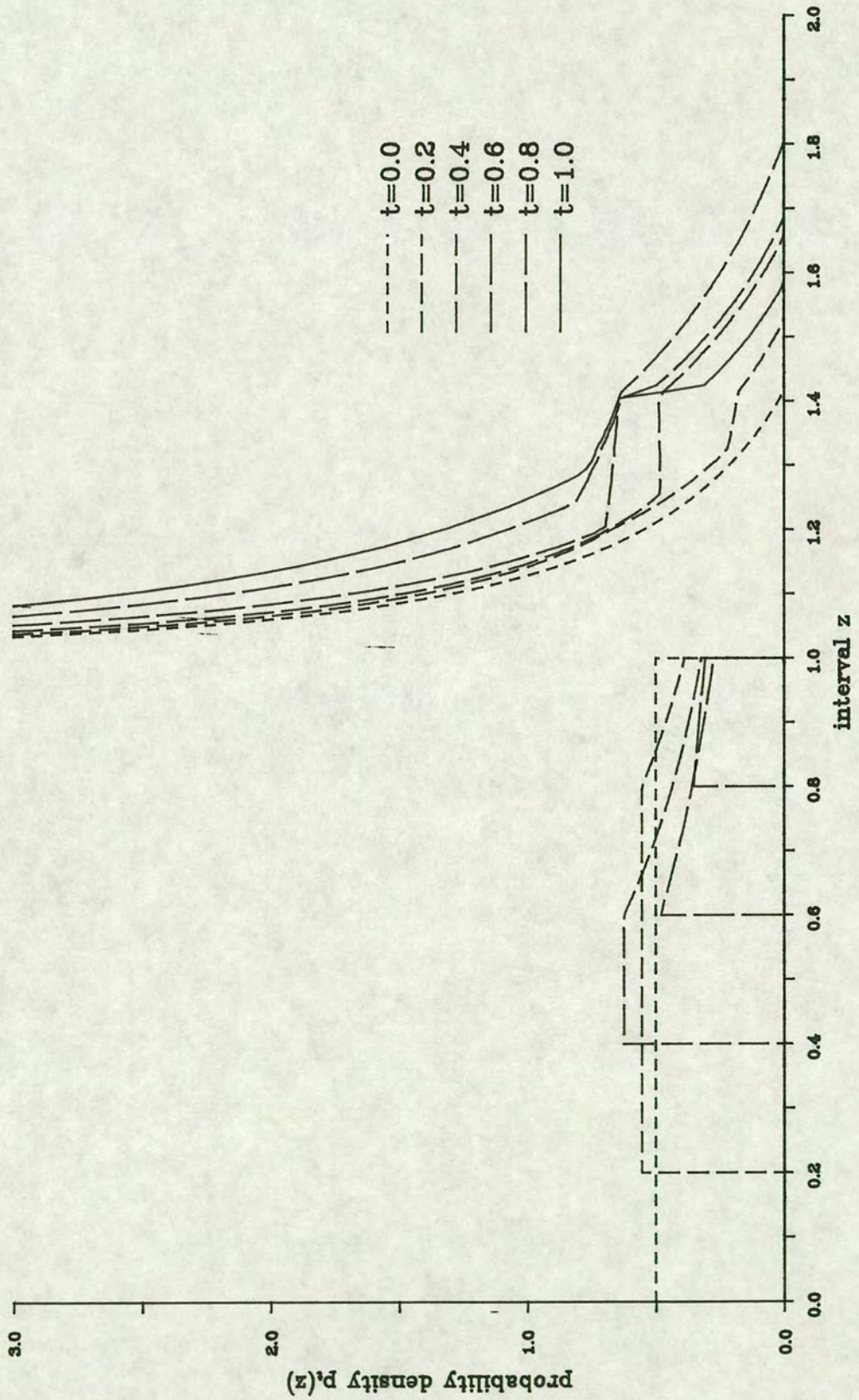
$$CDF_t(z) = C_t \sum_{\theta=45}^{90} \frac{CDF_t(z|\theta)}{E_t(z|\theta)}. \qquad (7.6)$$

The constant C_t is a scaling factor introduced to ensure that all probabilities sum to unity, such that $CDF_t(z=2.0) = 1$, and hence

$$C_t = \left[\frac{CDF_t(z=2.0|\theta)}{E_t(z|\theta)}\right]^{-1}.$$

The $CDF_t(z)$ function (7.6) is evaluated at values of t $\{0.0,0.1,0.2,\ldots,1.0\}$, to obtain a $\{z,t\}$ table of cumulative probabilities. A corresponding table of approximate probabilities $p_t(z)$ for each z is derived by calculating first differences of the cumulative probabilities. As described in Section 7.3.1, the density function $f_t(z)$ is evaluated by dividing $p_t(z)$ by the 'width' of the range of z values for which the probability is valid. The resulting densities $f_t(z)$ are plotted in Figure 7.3, for a subset of the values of t. For the densities of Figure 7.3, when evaluating the conditional function $CDF_t(z|\theta)$, the number of increments of θ in the range 45° to 90° was increased to 5000. This ensures smoothness of the plotted lines when the continuous distribution is only evaluated at discrete increments. The visual smoothness of the densities plotted in Figure 7.3 was used as a criterion to determine the

Figure 7.3: Density function of intervals
at thresholds t ε [0.0,1.0]



number of θ increments and in the following it is assumed that all results quoted are with respect to $CDF_t(z|\theta)$ evaluated at 5000 increments of θ.

7.4.3 Summary statistics

Summary statistics for each threshold t can be extracted from the $CDF_t(z)$ tables evaluated using (7.6), and are presented in Table 7.1. The lower limit is obviously equal to the threshold t. Further an analytic solution is possible for the expectation, as is now shown.

The expectation $E_t(z)$ is evaluated by integrating the conditional expectation $E_t(z|\theta)$ over the θ-range $(\pi/4,\pi/2)$, that is:

$$E_t(z) = C_t \int_{\pi/4}^{\pi/2} E_t(z|\theta)\, d\theta,$$

where C_t is a scaling factor. However, as in Section 7.4.1, it is necessary to weight the integral by the mean interval length $1/E_t(z|\theta)$ and hence:

$$E_t(z) = C_t \int_{\pi/4}^{\pi/2} \frac{E_t(z|\theta)}{E_t(z|\theta)}\, d\theta = \frac{\pi}{4} C_t.$$

The scaling factor C_t satisfies (7.5), and for general t is given by:

$$C_t = \left\{ \int_{\pi/4}^{\pi/2} \frac{1}{E_t(z|\theta)}\, d\theta \right\}^{-1}. \qquad (7.7)$$

The expectation is a function of t, and the two forms of $E_t(z|\theta)$ for the two main cases must be taken into account when integrating. Case (1) holds provided $2t \leqslant \operatorname{cosec}\theta$. From the partition diagram (Figure 7.2) it is obvious that the value of t at which the change between cases occurs depends on θ. Solving the equation $(2t-\operatorname{cosec}\theta = 0)$ for $\theta = \pi/4$ gives $t = \sqrt{0.5}$, and for $\theta = \pi/2$, $t = 0.5$.

Table 7.1: Summary statistics for theoretical distributions of distances with censoring thresholds in the range $t \in [0.0,1.0]$

$+_t$ and LLIM	Q_1	M	Q_3	ULIM	$E_t(z)$	SD	IQR	Range
0.0	0.50	1.00	1.04	1.41	0.79	0.36	0.54	1.41
0.1	0.58	1.00	1.05	1.47	0.83	0.32	0.47	1.37
0.2	0.65	1.01	1.06	1.53	0.87	0.30	0.41	1.33
0.3	0.73	1.01	1.08	1.59	0.92	0.28	0.35	1.29
0.4	0.85	1.02	1.11	1.66	0.98	0.26	0.26	1.26
0.5	1.00	1.03	1.15	1.73	1.05	0.24	0.15	1.23
0.6	1.01	1.04	1.19	1.80	1.10	0.22	0.19	1.20
0.7	1.01	1.05	1.19	1.74	1.11	0.18	0.18	1.04
0.8	1.01	1.06	1.18	1.69	1.11	0.15	0.17	0.89
0.9	1.01	1.06	1.17	1.63	1.11	0.13	0.16	0.73
1.0	1.02	1.06	1.17	1.58	1.11	0.12	0.15	0.58
Sample size: intervals on single line of length 1500 units								

+ Note: t denotes censoring threshold
LLIM denotes lower limit
Q_1 denotes lower quartile
M denotes median
Q_3 denotes upper quartile
ULIM denotes upper limit
$E_t(z)$ denotes expected value
SD denotes standard deviation
IQR denotes inter-quartile range

Thus for case (1), the range of t values is split into two sub-ranges: $0.0 \leqslant t \leqslant 0.5$ and $0.5 \leqslant t \leqslant \sqrt{0.5}$. There is a change in the limits of integration for each sub-range, occurring when $2t = \operatorname{cosec}\theta$, that is $\theta = \arcsin(1/2t)$. Case (2) holds for values of t $\sqrt{0.5} \leqslant t \leqslant 1.0$ and integration is over the θ-range $(\pi/4, \pi/2)$.

Considering each sub-range of t in turn, the expectation is now evaluated. For $0.0 \leqslant t \leqslant 0.5$, $2t \leqslant \operatorname{cosec}\theta$ so substituting the appropriate form of $E_t(z|\theta)$ from (7.3) into (7.7) gives:

$$C_t = \left\{ \int_{\pi/4}^{\pi/2} \sin\theta+\cos\theta-2t\sin\theta\cos\theta \, d\theta \right\}^{-1}.$$

On integration:

$$C_t = \left\{ [-\cos\theta+\sin\theta-t\sin^2\theta]_{\pi/4}^{\pi/2} \right\}^{-1},$$

which when substituting integration limits and simplifying gives $C_t = 2/(2-t)$. Hence:

$$E_t(z) = \frac{\pi}{4} \left[\frac{2}{2-t} \right], \qquad 0.0 \leqslant t \leqslant 0.5.$$

For $0.5 \leqslant t \leqslant \sqrt{0.5}$, then $2t \leqslant \operatorname{cosec}\theta$ and the integration follows as for $0.0 \leqslant t \leqslant 0.5$, but substituting θ-limits of integration of $(\pi/4, \arcsin(1/2t))$. Standard algebraic manipulation gives $C_t = 4t/(1+2t^2)$ and hence:

$$E_t(z) = \frac{\pi}{4} \left[\frac{4t}{1+2t^2} \right], \qquad 0.5 \leqslant t \leqslant \sqrt{0.5}.$$

Finally for $\sqrt{0.5} \leqslant t \leqslant 1.0$, case (2) holds and from (7.3) $E_t(z|\theta) = 1/(\sin\theta)$. Substituting into (7.7) for C_t and integrating $(\sin\theta)$ between $\pi/4$ and $\pi/2$ gives $C_t = \sqrt{2}$. Hence $E_t(z) = (\pi/4)\sqrt{2}$ for $\sqrt{0.5} \leqslant z \leqslant 1.0$. The expectation $E_t(z)$ is summarised as:

$$E_t(z) = \frac{\pi}{4} \begin{cases} \dfrac{2}{2-t} & 0.0 \leqslant t \leqslant 0.5 \\ \dfrac{4t}{1+2t^2} & 0.5 \leqslant t \leqslant \sqrt{0.5} \\ \sqrt{2} & \sqrt{0.5} \leqslant t \leqslant 1.0 \end{cases} \tag{7.8}$$

The expectation $E_t(z)$ is evaluated numerically using (7.8) for values of $\{t = 0.0,0.1,\ldots,1.0\}$. When compared with these analytic values, all the simulated results given in Table 7.1 agree to at least three decimal places. Analytic expressions are also given for the upper limit (ULIM) by Wheelwright and Glasbey (1993).

7.5 Discussion and conclusions

7.5.1 Simulated data

The simulations of Section 7.2.2 were used to gain information on the shape of the typical distribution of interval lengths. Using these simulations it was possible to establish limiting equations, in terms of θ and t, which partitioned the sample space into four cases, as shown in Figure 7.2. Based on the simulations and Figure 7.2, conditional PDFs $f_t(z|\theta)$ were defined and numerical integration used to evaluate $f_t(z)$, the PDF for the unconditional distribution. Hence the theoretical results of Table 7.1 are all based on the original simulations of Section 7.2.2.

In this section independent simulations of straight lines are carried out against which to check the theoretical results of Table 7.1. A new line defined by random slope θ and x-intercept s, and of length 20 units (after correcting for any 'start-up' effect) are simulated as described in Section 7.2.2. After censoring at thresholds of $\{t = 0.0,0.1,0.2,\ldots,1.0\}$ the interval lengths z between consecutive intersections are determined. One thousand such lines are simulated and all values of z after censoring stored. For each value of t summary statistics for the distribution of intervals over all 1000 simulations are given in Table 7.2. It can be seen that these results agree reasonably well with the values given in Table 7.1. This indicates that the theoretical results are not only valid for the simulations from which they were derived, but hold for any set of lines simulated under the same conditions.

Table 7.2: Summary statistics for distributions of distances for simulated data

$+_t$ and LLIM	Q_1	M	Q_3	ULIM	$E_t(z)$	SD	IQR	Range
0.0	0.49	0.99	1.03	1.40	0.78	0.35	0.54	1.40
0.1	0.56	1.00	1.04	1.46	0.82	0.32	0.48	1.36
0.2	0.64	1.00	1.05	1.52	0.87	0.30	0.41	1.32
0.3	0.72	1.01	1.07	1.58	0.92	0.28	0.35	1.28
0.4	0.83	1.01	1.10	1.64	0.98	0.26	0.27	1.24
0.5	1.00	1.02	1.15	1.72	1.04	0.24	0.15	1.22
0.6	1.00	1.03	1.19	1.80	1.09	0.22	0.19	1.20
0.7	1.01	1.04	1.19	1.73	1.11	0.18	0.19	1.03
0.8	1.01	1.05	1.18	1.67	1.11	0.15	0.17	0.87
0.9	1.01	1.05	1.17	1.62	1.11	0.13	0.16	0.72
1.0	1.01	1.05	1.18	1.58	1.11	0.12	0.16	0.58
Sample size: 1000 lines, each of length 20 units								

+ Note: t denotes censoring threshold
LLIM denotes lower limit
Q_1 denotes lower quartile
M denoted median
Q_3 denotes upper quartile
ULIM denotes upper limit
$E_t(z)$ denoted expected value
SD denotes standard deviation
IQR denotes inter-quartile range

7.5.2 Distribution at higher thresholds

The distribution of intervals for thresholds between 1.0 and 5.0 is now considered briefly. The censoring algorithm described in Section 7.2.2. must be modified for thresholds $t > 1.0$. For thresholds in the range [0.0,1.0] censoring a single intersection always results in a new cumulative interval z' which is greater than the specified t. For thresholds $t > 1.0$, cases may occur where the cumulative interval z' is still less than t so the intersection (x_{i+1}, y_{i+1}) is censored and the interval z'' calculated between (x_{i-1}, y_{i-1}) and (x_{i+2}, y_{i+2}). Subsequent intersections are similarly censored as necessary until the corresponding cumulative interval exceeds t.

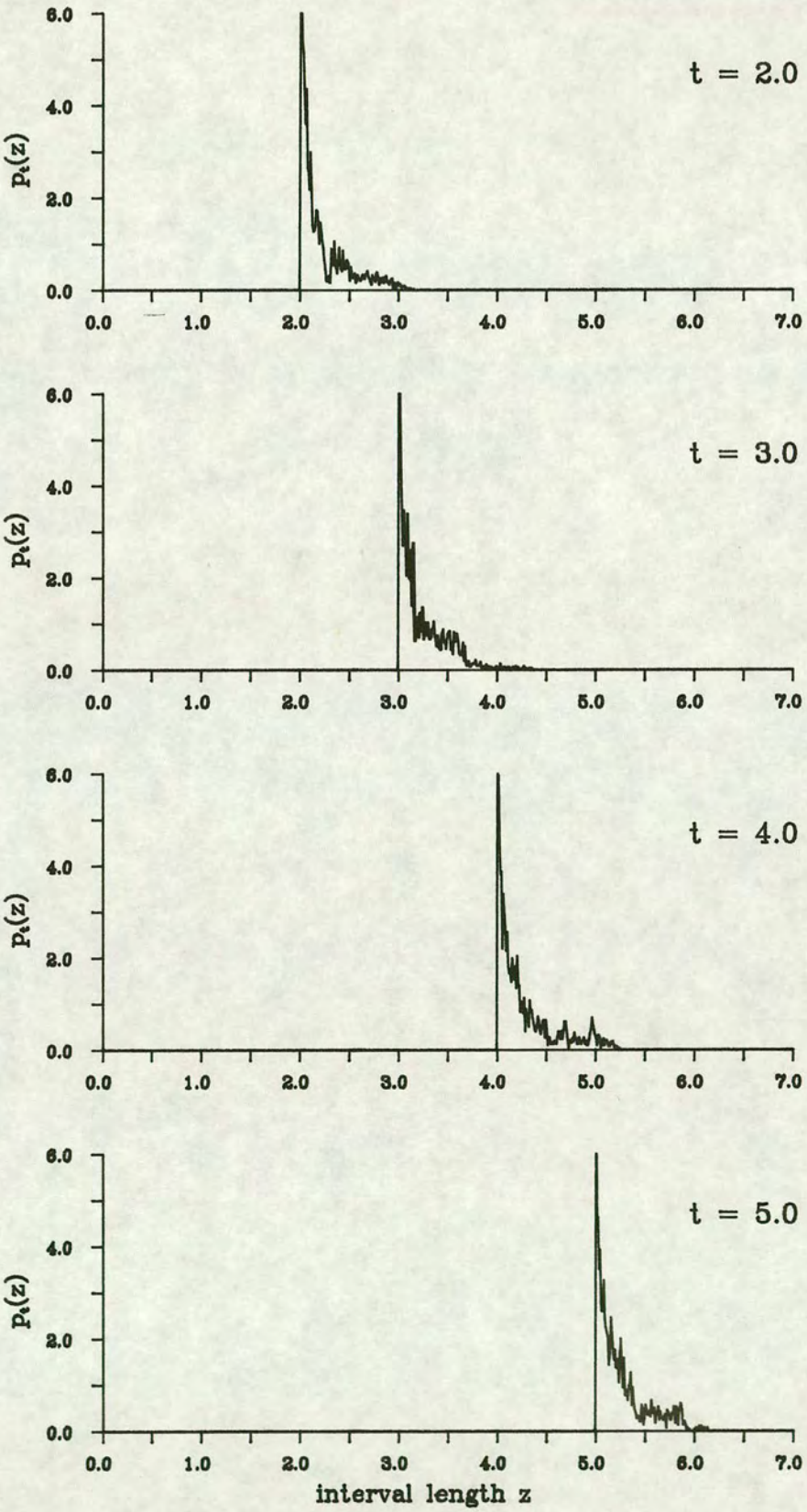
Further simulations with values of t in the range [1.0,5.0] are carried out. Defining a random slope θ and x-intercept s, 5000 lines are simulated to be of length 20 units (after correcting for any 'start-up' effect). Based on these 5000 simulations, the (approximate) probability distribution and density function are evaluated numerically, based on a count of the number of intervals of each length z. The density function $f_t(z)$ is illustrated in Figure 7.4 for values of $\{t = 2.0, 3.0, 4.0, 5.0\}$. To show more detail in the right tail of the density function, the high densities at interval lengths close to the specified threshold are truncated.

The densities plotted in Figure 7.4 indicate that the distributions for higher t have the same general shape as those in Figure 7.3, with a singularity at the integer immediately above t. However there are slight differences, such as the range appearing to oscillate with a frequency of $\sqrt{2}$ as t increases. No other value of t achieves as small a range as a value of t = 1.0.

7.5.3 Image edges

The work of this chapter arose from analysis of fitting a closed curve to points lying on an edge in an image. Therefore it is of interest to compare the summary statistics of the theoretical

Figure 7.4: Density function of interval lengths
when setting higher thresholds



distribution derived in Section 7.4 with similar statistics derived for the distribution of interval lengths between points on an edge in the MRI. The edge considered is that used in Section 5.5: the edge between the muscle and fat regions of the right leg, for subject et1 and variable z1. The set of n = 168 points are identified as positions of zero-crossings of a Laplacian of Gaussian smoothed image (Section 3.5), using a value for the smoothing parameter of $\lambda = 2.0$. This set of points corresponds to intersections of the edge with the lattice. Provided care is taken to define an appropriate start-point (x_1,y_1), and each intersection is checked to see if it matches the (last) point (x_n,y_n) for closure of the curve, intersections can be censored at thresholds of $\{t = 0.0,0.1,0.2,\ldots,1.0\}$. For each value of t, summary statistics are given in Table 7.3 for the distribution of interval lengths on the edge.

Most of the values in the two tables agree reasonably well. From Table 7.3 it can be seen however, that for values of $t \geqslant 0.8$ the upper limit ULIM observed for the image edge is greater than the theoretical upper limit. Although for the theoretical distribution ULIM decreases with increasing t, ULIM for the observed data remains the same. When checks are made of the longer observed interval lengths, relative to their position on the image edge, it is noted that such intervals occur in regions of high curvature of the edge. This suggests that in such regions the assumption of local linearity, made in order to approximate the edge by a random path, is not sufficiently accurate. Improvements could be made by modelling the interval between two censored intersections in terms of arc length, but this is not pursued here. Thus it can be concluded that the summary statistics derived for the theoretical distribution are lower bounds for the values which will be observed for the image edges.

7.5.4 Conclusions

The motivation for deriving the distribution discussed in this chapter was a problem encountered in Chapter 5: choosing a variable with which to label consecutive points round the edge. Interest was focussed on the censored index number labelling variable when

Table 7.3: Summary statistics for distributions of distances for image {Muscle/Fat} edge (Subject et1, Variable z1)

$+_t$ and $\oplus_{LLIM}$	Q_1	M	Q_3	ULIM	$E_t(z)$	SD	IQR	Range
0.0 (0.01)	0.53	1.00	1.03	1.29	0.80	0.34	0.50	1.28
0.1 (0.10)	0.61	1.00	1.03	1.35	0.83	0.31	0.42	1.25
0.2 (0.22)	0.68	1.00	1.04	1.40	0.87	0.28	0.36	1.18
0.3 (0.32)	0.72	1.00	1.05	1.42	0.91	0.26	0.32	1.09
0.4 (0.42)	0.82	1.01	1.07	1.61	0.96	0.26	0.25	1.19
0.5 (0.50)	1.00	1.02	1.10	1.61	1.03	0.25	0.10	1.11
0.6 (0.61)	1.00	1.03	1.14	1.61	1.08	0.22	0.14	1.00
0.7 (0.70)	1.00	1.03	1.14	1.64	1.10	0.20	0.13	0.94
0.8 (0.81)	1.00	1.03	1.14	1.85	1.11	0.20	0.13	1.04
0.9 (0.95)	1.01	1.04	1.17	1.85	1.14	0.20	0.16	0.90
1.0 (1.00)	1.01	1.04	1.17	1.85	1.14	0.20	0.16	0.85
Sample size: intervals calculated between 168 edge points								

+ Note: t denotes censoring threshold
LLIM denotes lower limit
Q_1 denotes lower quartile
M denotes median
Q_3 denotes upper quartile
ULIM denotes upper limit
$E_t(z)$ denotes expected value
SD denotes standard deviation
IQR denotes inter-quartile range

⊕ Note: observed lower limit for image data given in brackets

fitting a Fourier series to model the edge as a closed curve, and in particular the value of the censoring threshold t which minimised the variability in the distribution of interval lengths.

It can be seen from Table 7.1 that when censoring with thresholds in the range [0.0,1.0], the inter-quartile range is minimised when t = 0.5, when the lower quartile is equal to 1.0. However both the range and the standard deviation are minimised when t = 1.0. A minimising threshold of t = 1.0 is compatible with the empirical results of Section 5.5. As illustrated in Figure 5.3, for the FS fit in terms of the censored index number for different t, then the best fit in terms of residual standard deviation S_r (5.4) is obtained with a threshold of t = 1.0.

8 CONCLUSIONS

The purpose of this thesis was to consider methods for estimating edges in magnetic resonance images (MRI), as a first step in the automatic analysis of such data. The data consisted of images of cross-sections through human thighs, with particular attention given to the edges between the different tissue regions. The estimation of such edges was taken as a two-stage process. Firstly, the identification of points lying on the edge between two regions (Chapters 3,4), and secondly, the fitting of some form of closed curve to describe this set of points (Chapters 5,6,7).

An important part of this analysis was the recovery of the underlying continuous image from data sampled only at discrete points. The approach taken combined techniques from two fields. Results from the field of computer vision provided background information on the structure of the images and the physical characteristics of the imaged object which gave rise to specific observed features. Statistical methods were used to give a more formal justification for these intuitive ideas, allowing quantitative estimation and evaluation.

In Chapter 1 the literature was reviewed. Defining an edge as a significant change in intensity, the importance of edges for image analysis was discussed. A summary was given of methods existing for both the identification of edge points and the fitting of closed curves.

The data and magnetic resonance imaging were discussed in Chapter 2. The limits and structure of the image 'pixel' lattice were defined and the two intensity variables analysed in this thesis introduced. Considering the two variables separately, the range of observed intensity values was modified so that the data for all subjects was on the same scale, with integer increments between the intensity values. Summary statistics were calculated based on the numerical value at each pixel. This simplified approach ignored any spatial information on the relative position of each pixel within the lattice, information which was crucial in subsequent chapters.

The identification of edge points was covered in Chapter 3. The data observed at discrete lattice points were smoothed using kernel regression, so intensity could be estimated at any point in continuous space. It was concluded that for the MRI, the best representation of a set of edge points was obtained by smoothing data with a Gaussian kernel and identifying edge points as positions of zero-crossings of the Laplacian of Gaussian smoothed image.

In Chapter 4, several data-based methods were discussed for the estimation of the optimal amount of smoothing both for the data, and for derivatives of the data. The conclusion reached was that although an automatic method for estimating the smoothing parameter λ was preferred, in practice a subjective choice may be more appropriate, as was the case for the MRI. Using the methods of Chapter 3, the data were smoothed several times using different values of λ, and edges identified. The value of λ chosen as the smoothing parameter was that which, on visual assessment, resulted in the 'best' representation of the sets of edge points of interest. The analysis of this chapter did show how important it was to combine human interpretation of the qualitative properties of an image with any computational analysis. A further point to be taken into account was whether there was to be further processing of the data, since this influenced to what degree of positional accuracy individual edge points needed to be located.

As the second stage of the edge estimation, the set of points identified as zero-crossings of the Laplacian of Gaussian smoothed image was described in terms of a simple, closed curve. In Chapter 5, a Fourier series (FS) was fitted to this set of points, and the edge was described in terms of a set of Fourier descriptors (FD). Fitting a truncated FS of only a subset of the first few FD, a reasonably accurate representation of the edge was obtained. This showed that it was possible to define a model with fewer parameters than the full set of 2n edge point coordinates and still obtain a satisfactory description of the edge. The relation between the FD and geometric properties of the fitted shape, such as rotational symmetry, was discussed. Using simple simulated data, it was shown that conditional on the particular terms included in the series, these properties

still held even if only a subset of FD was used.

Prior to fitting a FS to the set of points, it is necessary to define a variable for labelling consecutive points round the edge. Two basic options were discussed in Chapter 5. Points could either be labelled in terms of cumulative distance round the edge (the distance measured from an arbitrary start-point), or alternatively in terms of index number, the relative position of the current point in the ordered sequence round the edge. The two labelling variables were compared for fitting a FS to a set of points sampled from an edge in one of the images. The conclusion reached was that the best fit was obtained using the cumulative distance labelling. If points were interpolated to equal distances round the edge the improvement in fit over using the original (unequal) distances was negligible, but the equal distance variable was preferred because of the ease with which the FS model could be fitted using a fast Fourier transform.

The analysis of both Chapters 6 and 7 was motivated by questions which arose when defining the basic FD model in Chapter 5. The first question was how many terms should be included in a truncated series; methods for determining a suitable number were discussed in Chapter 6. The second question was which of the possible variables (as defined in Chapter 5) should be used to label consecutive edge points. This led to the theoretical analysis of one of these variables in Chapter 7.

The analysis of Chapter 6 aimed to answer the first question arising from Chapter 5, namely the determination of a suitable number of terms at which to truncate the FS. Geometric properties of a region, such as perimeter length and area, can be determined from the FD fitted to an edge enclosing that region. The accuracy with which such properties were estimated when fitting only a subset of FD was used as a criterion to determine the optimum number of FD to be fitted. Using simple simulated shapes with known properties, comparisons were made between three estimators of perimeter length and two estimators of area. It was concluded that a more accurate estimate of perimeter length was obtained when fitting a truncated FS, as opposed to calculating the length based on the set of

identified points. In contrast, a reasonably accurate estimate of area was obtained based on the 2n identified edge points, and there was negligible improvement in accuracy irrespective of the number of terms included in the FS.

A variant of the index number labelling variable of Chapter 5 was analysed theoretically in Chapter 7. In Chapter 5 it was suggested that when labelling consecutive points in terms of index number, an improvement in fit of the model may be possible by censoring points which were less than a prespecified censoring threshold distance apart. Subsequent points were re-indexed and a FS fitted to the reduced subset of points remaining. The problem of how to define an optimum censoring threshold was considered in Chapter 7. By simulating random lines and identifying points of intersection of the lines with a square lattice, the distribution was derived of interval lengths between consecutive points, conditional on the censoring threshold used. The optimum threshold was defined as that which minimised the range of interval lengths.

From the analysis of this thesis it can be concluded that the estimation of edges in MRI using statistical methods is a realistic proposition. It is proposed that these methods should be used as an aid to complement human interpretation of such images, rather than totally replace the assessment by a trained operator.

The usefulness of the methods developed should be assessed in terms of the underlying objectives of the clinical analysis of the MRI. To realistically limit the amount of data considered, most of the analysis has been confined to the edge between the muscle and fat tissues in a single image for one subject. The methods could now be applied to estimate edges between different tissues, such as the muscle and bone, and to analyse images from several subjects. Some indication of the normal biological variation could be obtained by comparing one particular edge (say between muscle and fat tissues) for several subjects. This could then be used as a basis from which to define a model for the edge enclosing a typical healthy region, to be used as a template against which to compare similar potentially diseased regions. The amount of smoothing required for an image will

differ for each subject. It would be necessary to determine some criterion to define the 'correct' amount of when estimating the smoothing parameter subjectively (as proposed in Chapter 4), before making comparisons between images for several subjects.

The data analysed in this thesis were obtained as part of a nutrition study, where the amount of body fat was of interest. The methods of Chapter 6 for estimating the perimeter length and area of a region were shown to work well for simulated data. These methods could now be applied to estimate the area of different tissues in the thigh images. If possible these theoretical estimates should be compared with the values obtained subjectively by an operator outlining the same edge as the image is viewed on a computer screen. This would then give some idea in how successful the developed methods have been in providing a solution to some of the medical problems which motivated this analysis.

REFERENCES

Anderson, T.W. (1971) *The statistical analysis of Time Series.* Wiley, New York.

Babaud, J., Witkin, A.P., Baudin, M. and Duda, R.O. (1986) Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 26-33.

Benedetti, J. (1977) On the nonparametric estimation of regression functions. *Journal of the Royal Statistical Society, Series B* 39, 248-253.

Bergholm, F. (1987) Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 726-741.

Berman, M. and Culpin, D. (1986) The statistical behaviour of some least squares estimators of the centre and radius of the circle. *Journal of the Royal Statistical Society, Series B* 48, 183-196.

Berzins, V. (1984) Accuracy of Laplacian edge detectors. *Computer Vision, Graphics and Image Processing* 27, 195-210.

Bookstein, F.L. (1978) *The measurement of biological shape and shape change.* Springer-Verlag, Berlin and New York.

Buckley, M.J., Eagleson,G.K. and Silverman,B.W. (1988) The estimation of residual variance in nonparametric regression. *Biometrika* 75, 189-99.

Cacoullos, R. (1966) Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics* 18, 179-189.

Canny, J. (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 679-698.

Chan, K.P.-S. and Kay, J.W. (1991) Smoothing parameter selection in image restoration. In *Nonparametric Functional Estimation and Related Topics.* G.Roussas, Ed. Kluwer. pp. 201-211.

Clark, R.M. (1977) Nonparametric estimation of a smooth regression function. *Journal of the Royal Statistical Society, Series B* 39, 107-113.

Coleman, R. (1969) Random paths through convex bodies. *Journal of Applied Probability* 6, 430-441.

Coleman, R. (1992) Extended paths across tiles. Submitted to *Journal of Applied Probability.*

Collomb, G. (1981) Estimation non-parametrique de la regression: revue bibliographique. *International Statistical Review* 49, 75-93.

Cosgriff, R.L. (1960) Identification of shape. Report 820-11, ASTIA-AD-254-792, Ohio State University Research Foundation. Columbus.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degeree of smoothing by the method of generalised cross-validation. *Numerische Mathematik* 31, 377-403.

Dickey, F.M. and Shanmugan, K.S. (1977) Optimum edge detector filter. *Applied Optics* 16, 145-148.

Dorst, L. and Smeulders, A.W.M. (1987) Length estimation for digitized contours. *Computer Vision, Graphics and Image Processing* 40, 311-333.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

Epanechnikov, V.A. (1969) Nonparametric estimation of a multivariate probability density. *Theoretical Probabability and its Applications* 14, 153-158.

Fletcher, R. (1987) *Practical Methods of Optimisation.* 2nd edition. Wiley.

Foster, M.A. and Hutchison, J.M.S. (1985) NMR imaging - Method and applications. *Journal of Biomedical Imaging* 7, 171-182.

Foster, M.A. and Hutchison, J.M.S. (1988) *Practical NMR Imaging.* IRL Press, Oxford.

Foster, M.A., Hutchison, J.M.S., Mallard, J.R. and Fuller,M. (1984) Nuclear magnetic resonance pulse sequence and discrimination of high- and low-fat tissues. *Magnetic Resonance Imaging* 2, 187-192.

Fukunaga, K. and Hostetler, L.D. (1975) Estimation of the gradient of a density function. *IEEE Transactions on Information Theory* IT-21, 32-40.

Fuller, M.F., Stratton, S.M., Geddes, D., Fowler, P.A. and Foster, M.A. (1987) A study of the sites of adipose tissue loss by NMR imaging. In *In vivo Body Composition Studies.* K.J. Ellis, S. Yasanura and W.D. Morgan, Eds. Institute of Physical Sciences in Medicine, London. pp. 55-59.

Gasser, Th. and Müller, H.G. (1979) Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation.* Th. Gasser & M. Rosenblatt, Eds. Springer, Heidelberg. pp. 23-68.

Gasser, Th. and Müller, H.G. (1984) Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics,* 11, 171-185.

Gasser, Th., Müller, H.G. and Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, Series B* 47, 238-252.

Gasser, Th., Sroka, L. and Jennen-Steinmetz, C. (1986) Residual variance and residual pattern in nonlinear regression. *Biometrika* 73, 625-633.

Gonzales, R.C. and Wintz, P.A. (1987) *Digital Image Processing.* Wokinham Addison-Wesley, Reading, Mass.

Granlund, G.H. (1972) Fourier preprocessing for hand-print character recognition. *IEEE Transactions on Computers* C-21, 195-201.

Grimson, W.E.L. and Pavlidis, T. (1985) Discontinuity detection for visual surface reconstruction. *Computer Vision, Graphics and Image Processing* 30, 316-30.

Hall, P., Kay, J.W. and Titterington, D.M. (1990) Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* 77, 521-528.

Hall, P. and Titterington, D.M. (1987) Common structure of techniques for choosing smoothing parameters in regression problems. *Journal of the Royal Statistical Society, Series B* 49, 184-198.

Haralick, R.M. (1984) Digital step edges from zero-crossings of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6, 58-68.

Härdle, W. (1990) *Applied Nonparametric Regression.* Cambridge University Press.

Härdle, W. and Marron, J.S. (1985a) Asymptotic nonequivalence of some bandwidth selectors in nonparametric regression. *Biometrika* 72, 481-484.

Härdle, W. and Marron, J.S. (1985b) Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics* 13, 1465-1482.

Härdle, W., Hall, P. and Marron, J.S. (1988). How far are automatically chosen regression parameters from their optimum? *Journal of the American Statistical Association* 83, 86-95.

Hildreth, E.C. (1983) The detection of edges by computer and biological systems. *Computer Vision, Graphics and Image Processing* 22, 1-27.

Huertas, A. and Medioni,G. (1986) Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 651-664.

Kashyap, R.L. and Chellappa, R. (1981) Stochastic models for boundary analysis: Representation and reconstruction. *IEEE Transactions on Information Theory* IT-27, 627-637.

Kay, J.W. (1988) On the choice of regularisation parameter in image restoration. In *Springer Lecture Notes in Computer Science* 301, pp. 587-596.

Kendall, M.G. and Moran, P.A.P. (1963) *Geometrical Probability.* Griffin, London.

Kimber, A.C. (1983) Trimming in gamma samples. *Applied Statistics* 32, 7-14.

Kiryati, N. and Maydan, D. (1989) Calculating geometric properties from Fourier representation. *Pattern Recognition* 22, 469-475.

Koplowitz, J. and Bruckstein, A.M. (1989) Design of perimeter estimators for digitized planar shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-11, 611-622.

Kulpa, Z. (1977) Area and perimeter measurements of blobs in discrete binary pictures. *Computer Vision, Graphics and Image Processing* 6, 434-451.

Leipnik, R. (1960) The extended entropy uncertainty principle. *Information and Control* 3, 18-25.

Lin, C.C. and Chellappa, A. (1987) Classification of partial 2-d shapes using Fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 686-690.

Lunscher, W.H.H.J. and Beddoes, M.P. (1986) Optimal edge detector design I: Parameter selection and noise effects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 164-177.

Marr, D. and Hildreth, E. (1980) Theory of edge detection. *Proceedings of the Royal Society Of London, Series B* 207, 187-217.

Müller, H-G. (1988) *Nonparametric Regression Analysis of Longitudinal Data.* Springer-Verlag, Berlin.

Müller, H-G., Stadtmüller, U. and Schmitt, T. (1987) Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika* 74, 743-750.

Nadaraya, E.A. (1964) On estimating regression. *Theoretical Probability and its Applications* 9, 141-142.

NAG Limited Wilkinson House, Jordan Hill Road, Oxford, OX2 8DR.

Nalwa, V.S. (1987) Edge detector resolution improvement by image interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 446-451.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329-348.

Peli, T. and Malah, D. (1982) A study of edge detection algorithms. *Computer Vision, Graphics and Image Processing* 20, 1-21.

Pham, B. (1989) Conic B-splines for curve fitting - A unifying approach. *Computer Vision, Graphics and Image Processing* 45, 117-125.

Phillips, D.L. (1962) A technique for the numerical solution of certain integral equations of the first kind. *Journal Assoc. Comput. Mach.* 9, 84-97.

Poggio, T., Voorhees, H. and Yuille, A. (1988) A regularized solution to edge detection. *Journal of Complexity* 4, 106-123.

Pratt, W.K. (1977) *Digital Image Processing.* Academic Press, New York.

Prewitt, J.M.S. (1970) Object enhancement and extraction. In *Picture Processing and Psychopictorics.* B.S. Lipkin and A. Rosenfeld, Eds. Academic Press, New York.

Priestley, M.B. and Chao, M.T. (1972) Nonparametric function fitting. *Journal of the Royal Statistical Society, Series B* 34, 385-392.

Reinsch, C.H. (1967) Smoothing by spline functions. *Numerische Mathematik* 10, 177-183.

Rice, J. (1984) Bandwidth choice for nonparametric regression. *Annals of Statistics* 12, 1215-1230.

Rice, J. (1986) Bandwidth choice for differentiation. *Journal of Multivariate Analysis* 19, 251-264.

Roberts, L.G. (1965) Machine perception of three-dimensional solids. In *Optical and Electro-Optical Information Processing.* J.T.Tippett et al , Eds. M.I.T. Press, Cambridge, Mass. pp. 159-197.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832-835.

Rosenfeld, A. and Kak, A.C. (1982) *Digital Picture Processing.* 2nd edition. Academic Press, New York.

Rosenfeld, A. and Thurston, M. (1971) Edge and curve detection for visual scene. *IEEE Transactions on Computers* C-20, 562-569.

Shah, M., Voorhees, H. and Yuille, A. (1988) A regularized solution to edge detection. *Journal of Complexity* 4, 106-123.

Shann, R.T. and Oakley, J.P. (1990) Novel approach to boundary finding. *Image and Vision computing* 8, 32-36.

Sibson, R. (1980) A brief introduction to natural neighbour interpolation. In *Interpreting Multivariate Data.* V.Barnett, Ed. Wiley, London.

Sibson, R. (1988) Interpolation and Smoothing. In *Lecture Notes for SERC Short Course on Statistics, Computing and Expert Systems,* University of Nottingham.

Silverman, B.W. (1984) A fast and efficient cross validation method for smoothing parameter choice in spline regression. *Journal of the American Statistical Association* 79, 584-589.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Singh, R.S. (1976) Nonparametric estimation of mixed partial derivatives of a multivariate density. *Journal of Multivariate Analysis* 6, 111-122.

Singh, R.S. (1981) Speed of convergence in nonparametric estimation of a multivariate μ-density and it's mixed partial derivatives. *Journal of Statistical Planning Inference* 5, 287-298.

Sotak, G.E. and Boyer, K.L. (1989) The Laplacian-of-Gaussian kernel: A formal analysis and design procedure for fast, accurate convolution and full frame output. *Computer Vision, Graphics and Image Processing* 48, 147-189.

SPIDER (1978-1980) *Subroutine package for image data enhancement and recognition.* Electrotechnical Laboratory, Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Japan.

Sriraman, R., Koplowitz, J. and Mohan, S. (1989) Tree searched chain coding for sub-pixel reconstruction of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-11 95-104.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society, Series B* 36, 111-147.

Swokowski, E.W. (1979) *Calculus with analytical geometry.* Prindle, Weber and Schmidt, Boston, Mass.

Thompson, A.M., Kay, J.W. and Titterington, D.M. (1989) A cautionary note about cross-validatory choice. *Journal of Statistical Computation and Simulation* 33, 199-216.

Thompson, A.M., Brown, J.C., Kay,J.W. and Titterington, D.M. (1991a) A study of methods of choosing the smoothing parameter in image restoration by regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-13, 326-339.

Thompson, A.M., Kay, J.W. and Titterington, D.M. (1991b) Noise estimation in signal restoration using regularization. *Biometrika* 78, 475-488.

Titterington, D.M. (1985) Common structure of smoothing techniques in statistics. *International Statistical Review 53,* 141-170.

Torre, V. and Poggio, T.A. (1986) On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 147-163.

Ulupinar, F. and Medioni, G. (1990) Refining edges detected by a LoG operator. Technical Report, Institute for Robotics and Intelligent Systems, Department of Electrical Engineering and Computer Science, University of Southern California, Los Angeles.

Wahba, G. (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B* 45, 133-150.

Wahba, G. and Wold, S. (1975) A completely automatic French curve: fitting spline functions by cross-validation. *Communications in Statistics* 4, 1-17.

Watson, G.S. (1964) Smooth regression analysis. *Sankhyā, Series A* 26, 359-372.

Wheelwright, A.V. and Glasbey (1993) Distances between censored intersections between a square lattice and a random smooth path. To appear in *Journal of Applied Probability*, March 1993.

Wiejak, J.S., Buxton, H. and Buxton, B.F. (1985) Convolution with separable masks for early image processing. *Computer Vision, Graphics and Image Processing* 32, 279-290.

Witkin, A. (1983) Scale space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Karsruhe, pp 1019-1021.

Yokoi, S., Toriwaki, J. and Fukumura, T. (1975) An analysis of topological properties of digitized binary pictures using local features. *Computer Vision, Graphics and Image Processing* 4, 63-73.

Yuille, A.L. and Poggio, T.A. (1986) Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 15-25.

Zahn, C.T. and Roskies, R.Z. (1972) Fourier descriptors for plane closed curves. *IEEE Transactions on Computers* C-21, 269-281.

APPENDIX

The following paper, based on the work of Chapter 7, is to appear in the *Journal of Applied Probability*, March 1993.

Permission to include this preprint here has been granted by the second author, Dr. C. A. Glasbey.

DISTANCES BETWEEN CENSORED INTERSECTIONS BETWEEN A SQUARE LATTICE AND A RANDOM SMOOTH PATH

A.V. Wheelwright

Department of Mathematics and Statistics, University of Edinburgh

JCMB, The King's Buildings, Edinburgh EH9 3JZ

&

C.A. Glasbey

Scottish Agricultural Statistics Service,

JCMB, The King's Buildings, Edinburgh EH9 3JZ

Abstract

A random smooth path of infinite length crossed a square lattice. Intersections with the lattice were censored if they lay within a threshold distance of a preceding uncensored intersection, defined by tracking along the path in one direction. The distribution of distances between consecutive uncensored intersections is derived.

Keywords

Image analysis, stereology

1. Introduction

A random smooth path of infinite length in the plane intersected a square lattice. By smooth we mean that over the scale of a few lattice intervals the path can be considered to be straight, and by random we mean that short sections of the path have isotopic uniform randomness (IUR), and the path exhibits no long-term dependencies. In tracking along the path in a single direction, an intersection with the lattice was censored if it lay within a threshold distance of the preceding uncensored intersection. In this paper, the distribution is derived of distances between consecutive intersections, after censoring.

The problem arose in image analysis. Fourier descriptors (Grandlund, 1972) were used to approximate a closed curve specified by locations of intersections with the image lattice. That is, given points (x_i, y_i) for $i = 1, \ldots, n$, the fitted curve is

$$\hat{x}_i = \sum_{j=0}^{m} \left(a_j \cos \frac{2\pi ij}{n} + b_j \sin \frac{2\pi ij}{n} \right),$$

for some value of $m < n/2$, and similarly for y, where the a's and b's are the least squares estimates obtained by applying a Fourier transformation. By omitting some points, it was hoped to reduce the variability in distances between the remainder and thereby reduce m, the number of Fourier coefficients needed to produce a good fit.

There are further possibilities for applying the results in image analysis. The relationship between computer representations of images,

that is as data on a lattice, and true images in continuous space remains much in need of study. There are also connections with stereology, in which context the uncensored case was solved by Coleman (1969).

2. Distribution

Without loss of generality, we shall consider the integer lattice in the x-y plane, and a short interval of the path given by

$$y = (x - s) \tan \theta,$$

starting in the interval [0,1) on either the x or y axis, where the slope, θ, is a random variable, distributed in the range $\pi/4$ to $\pi/2$. The x-intercept, s, is also a random variable, whose distribution depends on θ and on the censoring threshold, t. We shall consider t in the range [0, 1], but will return to the general case in the discussion of t taking any positive value.

2.1 Conditioning on θ

In the absence of censoring and conditional upon θ, for IUR, s is uniformly distributed between $-\cot\theta$ and 1. The effect of censoring is that intervals with distances, denoted z, which are less than t are extended into the next square above, or to the right, as shown in Fig 1. But, to avoid double counting, intervals from the previous square (below or to the left) which are shorter than t, and thus have to be extended, have to be omitted. This produces the gap shown in Fig 1, and s is

uniformly distributed between $-\cot\theta$ and $-t\cos\theta$, and between $t\cos\theta$ and 1.

This situation, denoted case (1), pertains provided $2t \le \operatorname{cosec}\theta$. If t exceeds this limit then the upper interval for s changes to being between $\cot\theta - t\cos\theta$ and 1, denoted case (2). Fig 2 illustrates the two cases.

Returning to case (1), the intervals are as shown in Fig 1 provided that $t\cos\theta \le 1 - \cot\theta$, denoted case (1a). The probability density function of z, conditional on θ is

$$p_t(z|\theta) = \frac{1}{\sec\theta + \operatorname{cosec}\theta - 2t}\{2I(z|t, \operatorname{cosec}\theta - t) + I(z|\operatorname{cosec}\theta - t, \operatorname{cosec}\theta + t) + (\sec\theta - \operatorname{cosec}\theta)\,\delta(z - \operatorname{cosec}\theta)\},$$

where I is the indicator function defined as

$$I(z|a,b) = \begin{cases} 1 & \text{if } z \in (a,b) \\ 0 & \text{otherwise,} \end{cases}$$

and δ is the Dirac delta function

$$\delta(z) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon}\, I(z| -\varepsilon, \varepsilon).$$

If $t\cos\theta \ge 1 - \cot\theta$ then case (1b) pertains, and

$$p_t(z|\theta) = \frac{1}{\sec\theta + \operatorname{cosec}\theta - 2t}\{2I(z|t, \operatorname{cosec}\theta - t) + I(z|\operatorname{cosec}\theta - t, \sec\theta - t) + t\,\delta(z - \operatorname{cosec}\theta) + I(z|\operatorname{cosec}\theta, \sec\theta) + (\operatorname{cosec}\theta - \sec\theta + t)\,\delta(z - \sec\theta)\}.$$

Case (2) similarly divides into two. In (2a), $\cot\theta - t\cos\theta \le 1 - \cot\theta$, and $p_t(z|\theta) = \cos\theta \{ I(z|t, 2\operatorname{cosec}\theta - t)$

$+ (\sec\theta - 2\operatorname{cosec}\theta + 2t)\,\delta(z - \cos\theta)\}$.

Otherwise, in case (2b)

$p_t(z|\theta) = \cos\theta\{I(z|t, t + \sec\theta - \operatorname{cosec}\theta) + t\,\delta(z - \operatorname{cosec}\theta)$

$+ I(z|\operatorname{cosec}\theta, \sec\theta) + (2\operatorname{cosec}\theta - \sec\theta - t)\,\delta(z - \sec\theta)\}$.

2.2 Random θ

In removing the conditioning on θ, the probability density function becomes

$$p_t(z) = \int_{\pi/4}^{\pi/2} p_t(z|\theta)\, p_t(\theta)\, d\theta.$$

In the absence of censoring, for IUR, θ and the perpendicular distance from the origin to the line, that is $s \sin\theta$, are uniformly distributed over the region defined by $\pi/4 \le \theta \le \pi/2$, $-\cos\theta \le s\sin\theta \le \sin\theta$. Therefore $p(\theta)$ is proportioned to $\sin\theta + \cos\theta$. As we have seen, censoring restricts the range of s. Therefore,

$$p_t(\theta) = c_t \begin{cases} \sin\theta + \cos\theta - 2t\sin\theta\cos\theta & \text{if } 2t \le \operatorname{cosec}\theta \\ \sin\theta & \text{if } \operatorname{cosec}\theta \le 2t \le 2 \end{cases}$$

where c_t is a normalising constant.

In principle, it is possible to combine these equations with those in section 2.1 to obtain analytical expressions for $p_t(z)$ for different

values of t. However, this rapidly becomes very complicated, as can be judged from the result below for small t, that is the "resolution constraint" sampling case:

$$p_t(z) = \frac{2}{2-t} \begin{cases} \frac{1}{2} & \text{if } t \le z \le 1-t \\ \frac{1}{2(z+t)^2} & \text{if } 1-t \le z \le 1 \\ \frac{1}{2(z+t)^2} + \frac{1}{z^2\sqrt{z^2-1}} - \frac{1}{z^2} & \text{if } 1 < z \le 1+t \\ \frac{1}{2(z+t)^2} + \frac{1}{2(z-t)^2} - \frac{1}{2} + \frac{1}{z^2\sqrt{z^2-1}} - \frac{1}{z^2} & \text{if } 1+t \le z \le \operatorname{cosec}\alpha \\ \frac{1}{2(z-t)^2} - \frac{1}{2z^2} + \frac{t}{z^3} & \text{if } \operatorname{cosec}\alpha \le z \le \sqrt{2} \\ \frac{1}{2(z-t)^2} - \frac{1}{2z^2} - \frac{1}{2} + \frac{1}{z^2\sqrt{z^2-1}} + \frac{t}{z^3} & \text{if } \sqrt{2} \le z \le \sec\alpha \end{cases}$$

where α is the solution in the interval $\pi/4$ to $\pi/2$ of $t = \sec\alpha - \operatorname{cosec}\alpha$, that is

$$\alpha = \frac{\pi}{4} + \arcsin\left[\frac{1}{\sqrt{2}\,t}\left(\sqrt{1+t^2} - 1\right)\right].$$

This result applies provided that $1 + t \le \operatorname{cosec}\alpha$, that is $t < 0.29077$. For $t = 0$, this was one of the results derived by Coleman (1969); the conditions being what he termed μ-randomness, through a rectangle with unit sides.

We resorted to numerical integration, which provided the results in Fig 3 and Table 1. Analytic results are available for the expectation and

the upper limit:

$$E_t(Z) = \frac{\pi}{4} \begin{cases} \dfrac{2}{2 - t} & \text{if} \quad 0 \le t \le 0.5 \\ \dfrac{4t}{1 + 2t^2} & \text{if} \quad 0.5 \le t \le \sqrt{0.5} \\ \sqrt{2} & \text{if} \quad \sqrt{0.5} \le t \le 1. \end{cases}$$

The upper limit is sec α. Here α is as defined above for the range of value of t in the right boundary of case (1b) shown in Fig 2, that is $t \le \sqrt{13} / 6$, and otherwise for $t \le 1$ is the solution, in the range $\pi/4$ to

$\pi/2$, of

$$1 + t \cos \alpha - 2 \cot \alpha = 0.$$

3. Discussion

This project was set up to find what degree of censoring minimised the variability in the distribution. For censoring between 0 to 1, the range and standard deviation are minimised when $t = 1$. However, the interquartile range is minimised by t near 0.5, when the lower quartile is equal to 1.

Simulation results for t between 1 and 10 show the probability densities to have the same general shapes as those in Fig 3, with the singularity at the integer immediately above t. However, there are also subtle differences, such as the range appearing to oscillate with a frequency of $\sqrt{2}$ as t increases. No other values of t achieve as small a range

or standard deviation in the distribution as that reached when t = 1.

Acknowledgement

We thank Dr C G G Aitken for helping to supervise this work, Dr R Coleman for improving the rigour and comprehensibility of the paper, and the Science and Engineering Research Council and the Scottish Office Agriculture and Fisheries Department for financial support.

References

Coleman, R., (1969). Random paths through convex bodies. *Journal of Applied Probability*, **6**, 430-441.

Granlund, G.H., (1972). Fourier preprocessing for hand print character recognition. *IEEE Transactions in Computers*, **C21**, 195-201.

FIG 1

Censored intervals in case (1a)

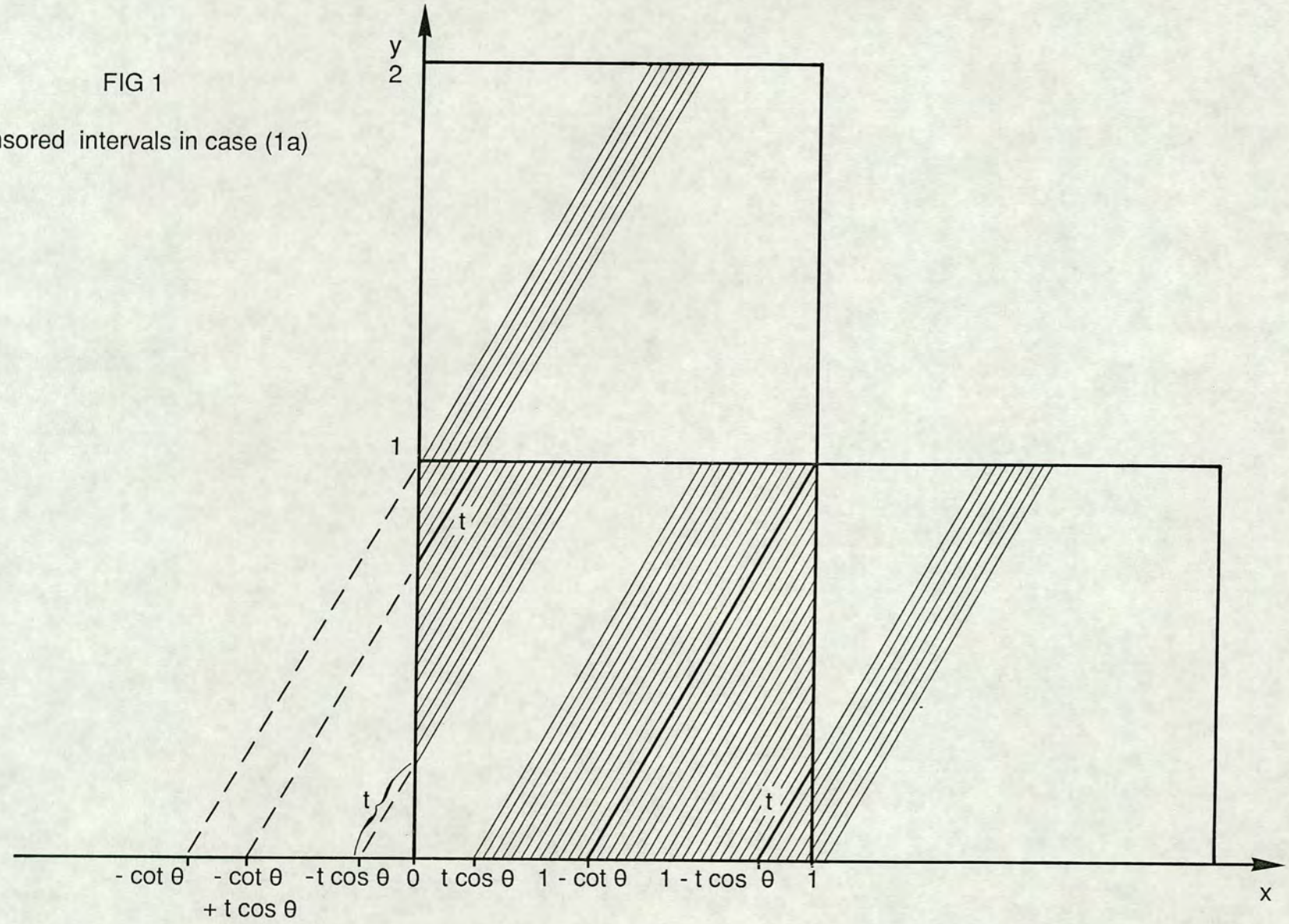


FIG 2

Four possible cases of envelopes of censored intervals

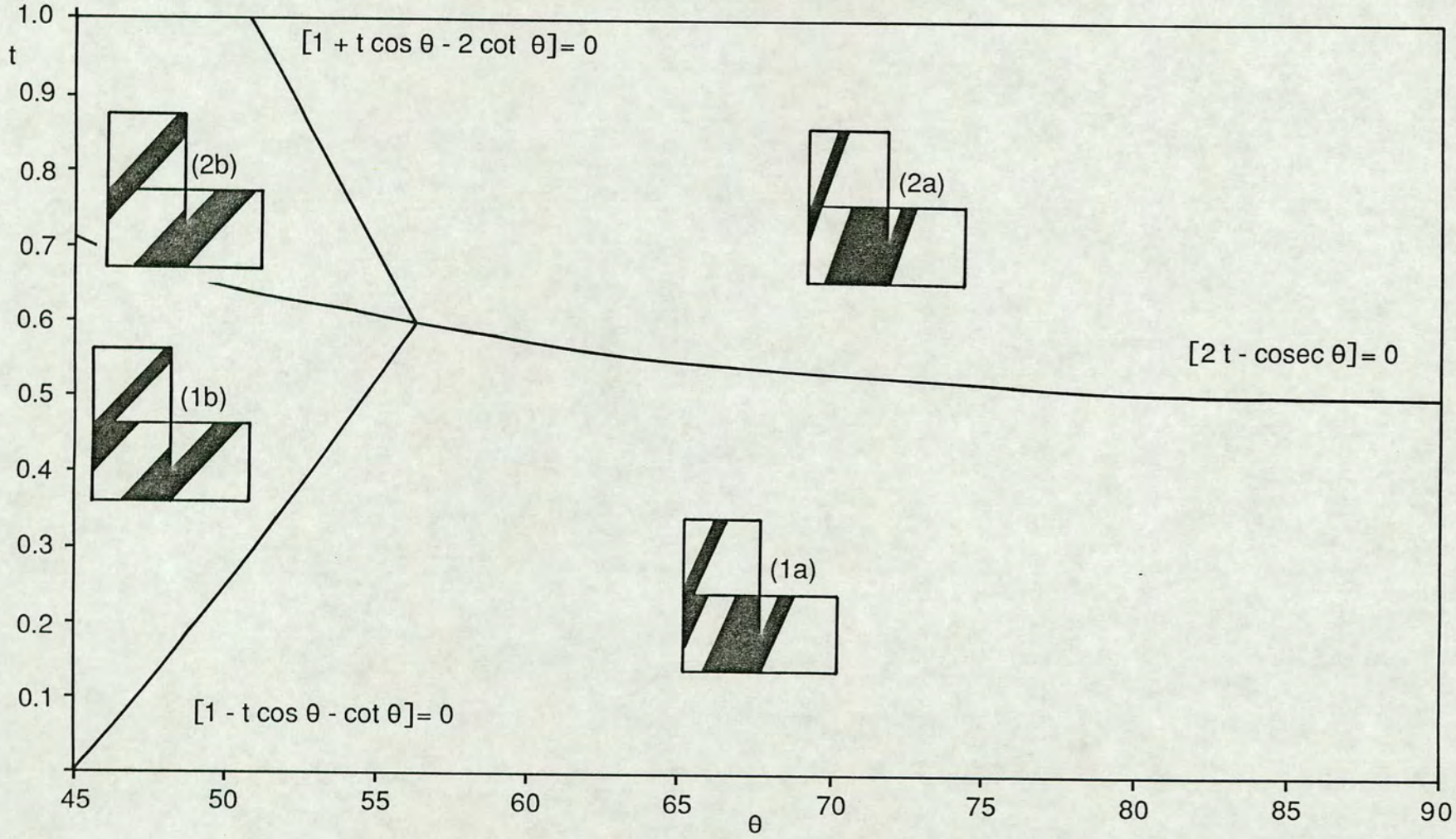


Fig 3 Comparison of distribution at different thresholds