Statistical shape analysis of featureless objects

Asger Hobolth

Ph.D. thesis

Department of Mathematical Sciences
University of Aarhus

January 2002

Preface

A hand has key points such as finger tips and indentations between fingers. Objects like biological cells and cell nuclei with no such obvious landmarks are called featureless. In malignancy grading of cancer tissue the shape of cells and cell nuclei is a crucial parameter and more generally characterisation of biological shape variation is of great importance in clinical medicine. By means of the deformable template model Ulf Grenander has provided a tool for describing shape variability. This Ph.D. thesis is concerned with the use of Grenanders model for exploring shape variability of featureless objects.

The thesis consists of a review together with six independently written enclosed papers and is submitted to the Faculty of Science, University of Aarhus. The purpose of the review is to provide an introduction to the statistical analysis of featureless objects and to present the main results of the accompanying papers. The core of the thesis is the enclosed papers, where various different representations and modelling strategies for featureless objects are presented and discussed.

During the four years of my Ph.D. study I have had the fortune of learning from several people. First of all I am deeply grateful to Eva B. Vedel Jensen for excellent supervision. Her committed guidance and constant support throughout the whole period is greatly acknowledged. I am also grateful to Jan Pedersen for joyful collaboration and many stimulating discussions. I want to thank John T. Kent and Ian L. Dryden from Leeds University and Ulf Grenander from Brown University for sharing their knowledge with me and for hospitality during my stays abroad. Many thanks are also due to Niels Væver Hartvig for various kinds of feedback and a gentle introduction to the programming language IDL. The main part of my work is grounded in biomedical problems, and I would like to thank Hans Jørgen Gundersen, Jens R. Nyengaard and Flemming B. Sørensen for providing data and insight into the applied aspects of my research.

Finally I would like to thank Mette Lena Olsen for steadfast emotional support and encouragement.

Århus, January 31, 2002.

Asger Hobolth

Summary

In a wide variety of disciplines it is of great practical importance to measure, describe and compare the shape of objects. Statistical shape analysis involves methods for studying the shape of objects. Here, shape refers to geometrical properties which are invariant under a translation, rotation and scaling of the object. Spatial objects with no obvious landmarks are called featureless, and perhaps the most familiar examples are biological cells and cell nuclei. This thesis is concerned with statistical shape analysis of featureless objects.

The first ingredient in shape analysis is the choice of an appropriate object representation. Many reasonable representations apply but the deformable template representation, advocated by Ulf Grenander, has played a dominant role. Grenanders basic idea is to consider an observed object as a deformation of a template. To be more specific, suppose the outline of a planar object is given in terms of a number of vertices. The vertices could be chosen equally spaced in terms of arc length or from rays emanating from a centre of the object relative to a fixed axis. To remove any translation effect one could also consider the edges of the object. Using the deformable template representation the vertices or edges are measured relative to an underlying template. The resulting vertex or edge transformation vector is a useful representation for analysing the shape of the planar object.

Having decided on the shape description vector the next topic is model specification. It is often desired to capture the most important object features such as boundary smoothness and local and global deviations from the template with a few model parameters. For the planar vertex or edge transformation vectors the classical model is the Gaussian first-order Markov model, where only neighbouring terms contribute to the probability density function. If the objects can be described as a small deformation of a circle then rotational symmetry is also desirable.

The literature on shape representation vectors and corresponding models is growing rapidly and increasingly realistic and complex methods are being formulated. This thesis discusses a variety of representations and models and provide insight to their mutual relationships. A part of the thesis is concerned with the use of planar object representations with underlying circular templates, and a complex linear relationship between the circular vertex and edge representations is derived. Descriptions in terms of ellipses or other non-circular templates are also discussed and corresponding models and probability distributions provided. Finally a continuous generalised p-order model is suggested as a more widely applicable model than the discrete Gaussian first-order Markov model.

With the increasing availability of digital images an important applied area of study is the interpretation of images. In high-level Bayesian image analysis, shape is an important element of prior modelling, and the use of planar deformable templates for object recognition is discussed.

Biological cells and cell nuclei are three-dimensional objects and some can be viewed as small deformations of a sphere. In this case it is useful to represent the object by a global translation and scaling together with a normal deformation of a sphere. A Gaussian model with the desired rotational symmetries can be derived from the physical laws of pressure fields acting on thin membranes. It is of interest to make statistical inference about the shape parameters from a finite number of systematic or random measurements of the surface. The thesis presents a detailed investigation of the spherical deformation model.

Stereology is concerned with methodology for making statistical inference about a population of spatial objects from geometric samples of the objects such as line and plane sections. The objective is not to reconstruct the objects, but instead to make inference about quantitative properties such as volume or surface area. In classical design-based stereology very few assumptions are made about the shape of the objects. Statistical shape analysis provides a promising alternative model-based tool for investigating quantitative properties of populations of objects, as discussed in the thesis.

The thesis consists of a review and six independently written papers published or submitted for publication in international statistical journals. Co-authours of the papers are my supervisor Eva B. Vedel Jensen and Jan Pedersen from Aarhus University, John T. Kent from Leeds University and Ian L. Dryden from Nottingham University.

Contents

Accompanying papers

A Hobolth, A. and Jensen, E.B.V. (2000).
Modelling stochastic changes in curve shape, with an application to cancer diagnostics.
Adv. Appl. Prob. (SGSA), **32**, 344-362.

B Hobolth, A., Kent, J.T. and Dryden, I.L. (2002)
On the relation between edge and vertex modelling in shape analysis.
To appear in *Scand. J. Statist.*

C Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2000)
A continuous parametric shape model.
Research report no. 13, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

D Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2001)
A deformable template model, with special reference to elliptical templates.
Research report no. 14, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

E Hobolth, A. and Jensen, E.B.V. (2001).
A note on design-based versus model-based variance estimation in stereology.
Research report no. 17, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

F Hobolth, A. (2002).
The spherical deformation model.
Research report no. 18, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

1 Introduction

A face has easily recognisable points such as the nose, lips, ears and eyes, and similarly a hand has landmarks such as finger tips and indentations between fingers. Objects with no obvious landmarks like biological cells and cell nuclei are called featureless. The purpose of this review is to summarise and discuss statistical methods for analysing featureless objects.

The literature on statistical analysis of featureless objects is growing rapidly, and increasingly realistic and complex stochastic models are being formulated. Choosing an appropriate shape representation is the first step towards a high standard shape analysis and the deformable template representation, advocated by Ulf Grenander, has played a dominant role. Grenanders basic idea is to represent the observed object as a deformation of an underlying template. In particular the hand and mitochondria studies (Grenander, Chow and Keenan, 1991, and Grenander and Miller, 1994) demonstrate the huge potential of the template representation and have inspired many researchers.

The shape representation generates a shape description vector and the next step in the analysis is to specify a model for this vector. It is desired to capture important object features such as boundary smoothness and global and local deviations from the template with a few model parameters. This review discusses a variety of representations and models and provide insight to their mutual relationships.

The paper is organised as follows. In Section 2 representations of planar objects are presented and compared. The representations are classified into deformations of circular and non-circular templates. Section 3 is concerned with models for the shape vectors from Section 2, and in particular Markov random field models are discussed in detail. Spherical and non-spherical template deformations for objects in three-dimensional space are considered in Section 4 and the corresponding models are described in Section 5. In Section 6 two applications of deformable templates are presented. First the problem of recognising an object in a noisy or blurred image is described and second an application in model-based stereology is discussed. The paper concludes with a description of dynamic space-time shape models.

2 2D deformable template representations

Consider a closed curve $\mathcal{V} = \{v(t) \in \mathbb{C} : 0 \leq t \leq T\}$, which is the boundary of a solid object in the complex plane indexed in an anti-clockwise manner. There are several possible choices of the index 'time' t such as arc length or angle of rays emanating from a centre of the object relative to a fixed axis. Similarly let $\mathcal{V}^0 = \{v^0(t) : 0 \leq t \leq T^0\}$ be the template curve in the complex plane and recall that the fundamental task is to describe $\mathcal{V}$ as a stochastic deformation of $\mathcal{V}^0$. In this section we first discuss the case when $\mathcal{V}^0$ is the unit circle and next we discuss the non-circular case.

2.1 Deformable circle representations

Let $\mathcal{V}^0 = \{e^{2\pi it} : 0 \leq t \leq 1\}$ be the unit circle and suppose the boundary $\mathcal{V}$ is only known at $n \geq 3$ points $v_j = v(t_j)$, $0 \leq t_1 \leq \cdots \leq t_n \leq T$. In this subsection we present several ways of describing the polygon $P = (v_j)$ as a deformation of the regular polygon $P^0 = (v_j^0)$. Here, $v_j^0 = v^0(j/n) = e^{2\pi ij/n}$ are regularly spaced vertices on the unit circle. The most useful representations for exploring shape variability are the standardised vertex and standardised edge transformation vectors. A complex linear relation between the transformation vectors is also provided.

2.1.1 Circular vertex transformation vectors

We now recall the construction of the standardised vertex transformation vector introduced in Hobolth, Kent and Dryden (2002). First we centre the polygon P by translating to the vertices

$$\tilde{v}_j = v_j - v_{av}, \quad j = 0, \ldots, n-1, \tag{2.1}$$

where $v_{av} = \sum_{j=0}^{n-1} v_j/n$. Next define the complex-valued *vertex transformation vector* $\mathbf{d} = (d_j)$ by

$$d_j = \tilde{v}_j / v_j^0, \quad j = 0, \ldots, n-1. \tag{2.2}$$

In order to remove effects due to rotation and scaling we define the *standardised vertex transformation vector* $\mathbf{u} = (u_j)$ with entries

$$u_j = d_j/d_{av}, \quad j = 0, \ldots, n-1,$$

where $d_{av} = \sum_{j=0}^{n-1} d_j/n$. It is easily verified that the original vertices can be written

$$v_j = v_{av} + d_{av}(1 + u_j)v_j^0, \quad j = 0, \ldots, n-1,$$

and thus the positions of the original vertices are separated into location v_{av}, orientation and size d_{av}, and a standardised vertex transformation vector $\mathbf{u} = (u_j)$ providing the deformation of P^0 to the shape of P. Note that $\mathrm{Re}(u_j) = r_j$ determines the component of the jth vertex of P *tangential* to $d_{av}v_j^0$ while $\mathrm{Im}(u_j) = s_j$ determines the *normal* component. The transformation vector $\mathbf{u}$ is invariant under a translation, rotation and isotropic scaling of the object, and therefore it only contains shape information. Due to the standardisation and centring there are two (complex) constraints on the transformation vector, and they can be written

$$\sum_{j=0}^{n-1} u_j = 0 \quad \text{and} \quad \sum_{j=0}^{n-1} u_j e^{2\pi ij/n} = 0. \tag{2.3}$$

An often used method of describing the boundary $\mathcal{V}$ is the *radial representation* as described below. Suppose the object is star-shaped relative to the centre of gravity $c \in$

$\mathbb{C}$, i.e. the radial vectors from c to $\mathcal{V}$ all remain inside the object. Find the vertices at $\mathcal{V}$ by rays starting at c with angles $2\pi j/n$ relative to some fixed axis. Then the sequence of distances $d_j, j = 0, \dots, n-1$, provides a description of the boundary. Consider the polygon given by the vertices

$$\tilde{v}_j = d_j v_j^0, \quad d_j \in \mathbb{R}, \quad j = 0, \dots, n-1. \tag{2.4}$$

The standardised vertex transformation vector of this polygon is given by

$$u_j = d_j/d_{av} - 1 \in \mathbb{R}, \quad j = 0, \dots, n-1, \tag{2.5}$$

and there is variability in the tangent component $\mathbf{r} = \mathrm{Re}(\mathbf{u})$ of the vertex transformation vector only. Note that in this construction of the vertex transformation vector we avoid centring the vertices as in (2.1), but take (2.2) as our starting point. This minor violation of the construction of the standardised transformation vector implies that the average of the vertices is only approximately zero, and therefore the second equation in (2.3) only holds approximately. If the average of the vertices defined from the centre of gravity is not approximately zero, then we suggest iterating the construction until a centre is found with the property that the average of the corresponding vertices is approximately zero.

In the limit when n tends to infinity we have that $d(t), 0 \le t \le 1$, is the distance from c to $\mathcal{V}$ in direction $2\pi t$, where $2\pi t$ is the angle between a fixed axis and a ray starting in c. The standardised radius-vector function $\{r(t) : 0 \le t \le 1\}$ is obtained from

$$r(t) = d(t)/d_{av} - 1, \;\; 0 \le t \le 1,$$

where $d_{av} = \int_0^1 d(t)dt$ is the mean distance length.

There is a large amount of literature on vertex transformation vectors and usually the radial representation is used. The prior model in Rue and Syversveen (1998) is one example. In a procedure for identifying cells in a digital image, the regular n-sided polygon P^0 is deformed and the generated object is determined by (2.2), where the vertex transformation vector $\mathbf{d} = (d_j) \in \mathbb{R}^n$ is real and follows a multivariate normal distribution, which is invariant under cyclic permutation. Note that when the components of $\mathbf{d}$ are real and positive then the generated objects are star-shaped relative to the origin. Some other examples are given by Stoyan and Stoyan (1994, p. 80), Mardia and Qian (1995) and Mardia *et al.* (1996) who use radial representations for modelling sand grains, leaves and mushrooms.

2.1.2 Circular edge transformation vectors

We now change the focus from the vertices to the edges and recall the standardised edge transformation vector introduced in Kent, Dryden and Anderson (2000). Focusing on the edges implies location invariance immediately. The edge vector $\mathbf{e} = (e_j)$ of the polygon P has components

$$e_j \;\; = \;\; v_{j+1} - v_j, \;\; j = 0, \dots, n-1, \tag{2.6}$$

whereas the edge vector of the regular polygon P^0 is given by

$$e_j^0 = v_{j+1}^0 - v_j^0, \quad j = 0, \ldots, n-1.$$

All subscripts throughout the paper are interpreted modulo n. The complex-valued *edge transformation vector* $\mathbf{t} = (t_j)$ is obtained by

$$t_j = e_j / e_j^0, \quad j = 0, \ldots, n-1. \tag{2.7}$$

Invariance under changes in scale and rotation is obtained by dividing with $t_{av} = \sum_{j=0}^{n-1} t_j / n$ leading to the *standardised edge transformation vector* $\mathbf{z} = (z_j)$ defined by

$$z_j = t_j / t_{av} - 1, \qquad j = 0, \ldots, n-1.$$

The standardisation and the closure constraint $\sum_{j=0}^{n-1} e_j = 0$ imply two complex constraints on $\mathbf{z}$ given by

$$\sum_{j=0}^{n-1} z_j = 0 \quad \text{and} \quad \sum_{j=0}^{n-1} z_j e^{2\pi i j/n} = 0, \tag{2.8}$$

and therefore $\mathbf{z}$ has $2n-4$ free parameters. Note that the constraints on $\mathbf{z}$ and $\mathbf{u}$ are the same. As in the previous subsection it follows from

$$e_j = t_{av}(1 + z_j) e_j^0, \qquad j = 0, \ldots, n-1, \tag{2.9}$$

that $\mathrm{Re}(z_j) = x_j$ determines the component of the jth edge of P *tangential* to $t_{av} e_j^0$ while $\mathrm{Im}(z_j) = y_j$ determines the *normal* component.

Another way of defining vertices on $\mathcal{V}$ is in terms of equal arc length, and if the boundary is a small deformation of a circle, then the value of the tangent component $\mathbf{x}$ of the edge transformation vector $\mathbf{z}$ is approximately zero. In practice, one finds the full edge transformation vector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ of the object and consider the normal component $\mathbf{y}$ as the data. This way of describing the boundary is known as the *constant length representation* and is closely related to the tangent angle function, cf. Zahn and Roskies (1972) or Stoyan and Stoyan (1994, p. 69).

If there is no variability in the normal component $\mathbf{y}$ of the standardised edge transformation vector, the angles between successive edges in P are the same as for the regular polygon P^0 and equal to $2\pi/n$. This is called the *constant angle representation* and all realised configurations are convex objects.

The full edge transformation vector is used in Grenander and Miller (1994) for Bayesian object recognition of mitochondria, while Kent *et al.* (2000) use the constant length representation to explore shape variability for a sample of sand grains and ceramic material particle sections. Hansen *et al.* (2000) also report on an object recognition study and essentially use the constant angle representation for describing arteries in a sequence of images.

2.1.3 Relations between circular transformation vectors

We now have two different ways of describing the shape of P relative to the regular polygon P^0, the standardised vertex transformation vector $\mathbf{u}$ and the standardised edge transformation vector $\mathbf{z}$. The edges are determined by the vertices and, together with a specification of the location, the edges determine the vertices. Hobolth, Kent and Dryden (2002) derive a complex linear relation between the two vectors. Letting

$$z_j = x_j + iy_j, \;\; u_j = r_j + is_j, \;\; j = 0, \ldots, n-1,$$

the relation can be written in real coordinates as

$$2x_j \;\; = \;\; r_{j+1} + r_j + \frac{1}{\tan(\pi/n)}(s_{j+1} - s_j), \;\; j = 0, \ldots, n-1, \tag{2.10}$$

and

$$2y_j \;\; = \;\; -\frac{1}{\tan(\pi/n)}(r_{j+1} - r_j) + s_{j+1} + s_j, \;\; j = 0, \ldots, n-1. \tag{2.11}$$

If the radial representation is used, the standardised vertex transformation vector is real. Furthermore if the objects are small deformations of a circle, the (real) vertex transformation vector is approximately zero, and for large n we obtain

$$x_j \approx 0, \;\; y_j \approx -\frac{n}{2\pi}(r_{j+1} - r_j), \;\; j = 0, \ldots, n-1. \tag{2.12}$$

When the radius-vector function is considered we get

$$x(t) \approx 0, \;\; y(t) \approx -r'(t), \;\; 0 \leq t \leq 1, \tag{2.13}$$

and thus the angular change of $\mathcal{V}$ relative to $\mathcal{V}^0$ approximately equals the derivative of the radius-vector function.

In conclusion the circular vertex and edge transformation vectors are useful for representing an object as a deformation of a circle. The full representations are needed when the vertices are arbitrarily spaced, but choosing the radial or constant length representation leads to a more succinct description of the data and reduces the dimension from $2n$ to n. Finally there is a complex linear relation between the two transformation vectors, and for a large number of regularly spaced vertices the derivative of the radius-vector function approximately equals the angular change of the object boundary relative to the circular template.

2.2 Deformable non-circular representations

Recall that the basic idea is to describe the observed boundary $\mathcal{V} = \{v(t) : 0 \leq t \leq T\}$ as a stochastic deformation of a closed template curve $\mathcal{V}^0 = \{v^0(t) : 0 \leq t \leq T^0\}$. Stoyan and Stoyan (1994, p. 90) and Hobolth, Pedersen and Jensen (2001) consider objects given by

$$v(t) = (r(t) + r^0(t))v^0(t), \;\; v^0(t) = e^{2\pi it}, \;\; 0 \leq t \leq 1, \tag{2.14}$$

where $r^0(t)$ represents the template (for example an ellipse), and $r(t)$ is a zero mean stochastic fluctuation around the template. A basic assumption for this representation to hold is that the deviations appear in the radial direction.

Hobolth and Jensen (2000) represent the boundary as a normal deformation

$$v(t) = v^0(t) + r(t)\omega^0(t), \quad 0 \le t \le T, \tag{2.15}$$

where $\omega^0(t)$ is the outer unit normal vector to $\mathcal{V}^0$ at $v^0(t)$, and $r(t)$ is the signed distance between $v(t)$ and $v^0(t)$. In Subsection 2.1.1 the curve $\mathcal{V}^0$ was the unit circle and $\{r(t) : 0 \le t \le 1\}$ the standardised radius-vector function. It can be shown that a similar result as (2.13) holds. To be more precise, Hobolth and Jensen (2000) show that if the deformations are small then

$$r'(t) \approx \Phi_{\mathcal{V}^0}(t) - \Phi_{\mathcal{V}}(t),$$

where $\Phi_{\mathcal{V}^0}(t)$ and $\Phi_{\mathcal{V}}(t)$ are the angles that $\mathcal{V}^{0\prime}(t)$ and $\mathcal{V}'(t)$ make with a fixed axis.

In an application concerning defect detection in potatoes Grenander and Manbeck (1993) describe an edge deformation model. The template $P^0 = \{v_j^0 \in \mathbb{C} : j = 0, \ldots, n-1\}$ is a discretized ellipse with fixed eccentricity, and the allowed deformations of the template edges $e_j^0 = v_{j+1}^0 - v_j^0$ consist of global scaling s, global rotation θ and local rotations θ_j. Since the local rotations θ_j are expected to be small the following approximation holds

$$e_j = se^{i\theta}e^{i\theta_j}e_j^0 \approx se^{i\theta}(1 + i\theta_j)e_j^0, \quad j = 0, \ldots, n-1. \tag{2.16}$$

A comparison with (2.9) show that this representation is closely related to the circular constant length representation, but with a non-circular template. In order to ensure the outline to be closed the constraint $\sum_{j=0}^{n-1} e_j = 0$ needs to be imposed.

Rue and Hurn (1999) use a similar representation as the prior in a Bayesian object recognition problem. They represent an object by

$$e_j = se^{i\theta}(1 + z_j)e_j^0, \quad z_j \in \mathbb{C}, \quad j = 0, \ldots, n-1, \tag{2.17}$$

where the edges e_j^0 are the edges of an ellipse or a circle. Since $z_j \in \mathbb{C}$ this representation is related to the full circular edge representation.

Whereas most of the circular representations generalise to non-circular representations it should be stressed that handling the constraints due to template fitting gets more difficult. Consider for example the representation (2.17). Using a circular template we have

$$e_j^0 = v_{j+1}^0 - v_j^0 = e^{2\pi i(j+1)/n} - e^{2\pi ij/n} = (e^{2\pi i/n} - 1)e^{2\pi ij/n},$$

and the closure constraint $\sum_{j=0}^{n-1} e_j = 0$ is given by $\sum_{j=0}^{n-1} z_j e^{2\pi ij/n} = 0$ in terms of $\mathbf{z}$. When the circulant symmetry no longer applies the closure constraint in terms of $\mathbf{z}$ is more complicated. Similarly the representations (2.14) and (2.15) implicitly assumes that the template and the observed boundary have been aligned, but the alignment procedure usually implies complicated constraints on the stochastic fluctuation.

3 2D deformation models

In this section some commonly used models for the standardised transformation vectors and standardised transformation functions are described. Emphasis is on the pth-order Markov Random Field model and its various extensions.

3.1 Circular deformation models

Suppose the vector $\mathbf{r} \in \mathbb{R}^n$ from the radial representation follows a multivariate normal distribution $N_n(\mu, K)$ with mean $\mu \in \mathbb{R}^n$ and $n \times n$ covariance matrix K. The lack of features on the object implies that any statistical model should be invariant under cyclic permutation of the vertices. Therefore the mean $\mathrm{E}r_i = \mu_i$ should not depend on i, and together with the first constraint in (2.3), with u_j replaced by r_j, this implies that $\mu = \mathbf{0}$. Similarly the covariance $\mathrm{Cov}(r_i, r_{i+j})$ should only depend on j, and therefore the covariance matrix K becomes a circulant matrix.

The circulant symmetry implies that it is useful to analyse the models in the spectral domain. Suppose n is odd and write $\mathbf{r}$ in terms of a Fourier series

$$r_j = a_0 + \sqrt{2} \sum_{k=1}^{(n-1)/2} \big(a_k \cos(2\pi kj/n) + b_k \sin(2\pi kj/n)\big), \tag{3.1}$$

where

$$a_0 = \frac{1}{n} \sum_{j=0}^{n-1} r_j, \tag{3.2}$$

$$a_k = \frac{\sqrt{2}}{n} \sum_{j=0}^{n-1} r_j \cos(2\pi jk/n), \quad k = 1, \ldots, (n-1)/2, \tag{3.3}$$

$$b_k = \frac{\sqrt{2}}{n} \sum_{j=0}^{n-1} r_j \sin(2\pi jk/n), \quad k = 1, \ldots, (n-1)/2. \tag{3.4}$$

Since a_0 and a_k, b_k, $k = 1, \ldots, (n-1)/2$, represent a rotation from the original coordinates $\mathbf{r}$ to the orthogonal eigenvectors of K they are all independent random variables with mean zero and variances λ_k determined by

$$\frac{1}{\lambda_0} = \sum_{i,j=0}^{n-1} k^{ij}, \ \frac{1}{\lambda_k} = 2 \sum_{i,j=0}^{n-1} k^{ij} \cos(2\pi ik/n) \cos(2\pi jk/n), \ k = 1, \ldots, (n-1)/2, \tag{3.5}$$

where $K^{-1} = (k^{ij})$. One reason why it is useful to formulate the models in the spectral domain is that the constraints (2.3), with u_j replaced by r_j, are easy to take into account. They imply that $a_0 = a_1 = b_1 = 0$, which is fulfilled by setting $\lambda_0 = \lambda_1 = 0$. In applications the high frequency Fourier coefficients are often poorly determined due to digitization effects, and therefore only the low frequency Fourier coefficients are used for subsequent analysis. A description of Fourier series analysis of radial vector functions can also be found in Stoyan and Stoyan (1994, p. 80).

3.1.1 Parametric models

Grenander and Miller (1994) use the standardised circular edge transformation vector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ to represent the outline of mitochondria. They assume that $\mathbf{z}$ follows a zero mean complex symmetric normal distribution, which means that $(\mathbf{x}, \mathbf{y}) \sim N_{2n}(\mu, K)$ with $K_{xx} = K_{yy}$ and $K_{xy} = -K_{yx}$, cf. Goodman (1963). Hobolth, Kent and Dryden (2002) use the complex linear relation (2.10) and (2.11) between $\mathbf{z}$ and the standardised vertex transformation vector $\mathbf{u} = \mathbf{r} + i\mathbf{s}$ to show that the complex symmetric edge transformation model can be rewritten as a complex symmetric vertex transformation model.

A very useful class of parametric models for the radial vector $\mathbf{r}$ is the Gaussian Markov random field (MRF) model, sometimes also referred to as conditional autoregression (CAR) model, cf. Besag (1974). A multivariate normal model with a real circulant $n \times n$ covariance matrix K is a pth order MRF model if

$$k^{ij} = 0 \text{ unless } |(i-j) \bmod n| \leq p,$$

where $(i-j) \bmod n$ is interpreted to lie in the interval $]-n/2, n/2]$ and $p < n/2$. The first-order MRF model can be written as

$$K^{-1} = \text{circ}(\frac{\alpha_0}{n} + 2\alpha_1 n, -\alpha_1 n, 0, \ldots, 0, -\alpha_1 n), \quad \alpha_0, \alpha_1 > 0, \tag{3.6}$$

where we use the notation $\text{circ}(a_0, \ldots, a_{n-1})$ for a circulant matrix with first row $(a_0, \ldots, a_{n-1})$. With the reparametrization

$$\sigma^2 = \frac{n}{\alpha_0 + 2n^2\alpha_1}, \quad \eta = \frac{2n^2\alpha_1}{\alpha_0 + 2n^2\alpha_1},$$

the conditional distribution of the first-order model is given by

$$r_j | \{r_i : i \neq j\} \sim N(\frac{\eta}{2}(r_{j-1} + r_{j+1}), \sigma^2),$$

and therefore the distribution of r_j given all the remaining coordinates of $\mathbf{r}$ depends only on its first-order neighbours. For a pth-order MRF model it holds that the conditional distribution of r_j given all the remaining coordinates depends only on its p nearest neighbours. Note that

$$\mathbf{r}^T K^{-1} \mathbf{r} = \alpha_0 \frac{1}{n} \sum_{j=0}^{n-1} r_j^2 + \alpha_1 n \sum_{j=0}^{n-1} (r_{j+1} - r_j)^2 = \alpha_0 S_0 + \alpha_1 S_1,$$

and thus the sufficient statistic (S_0, S_1) has a nice geometric interpretation. The statistic S_0 is a discrete measure of the distance between $\mathcal{V}$ and $\mathcal{V}^0$ and it follows from (2.12) that S_1 is a discrete measure of the angular changes of $\mathcal{V}$ relative to $\mathcal{V}^0$. It follows from (3.5) that the variances of the first-order MRF model (3.6) decrease as

$$1/\lambda_k = \alpha_0 + 2\alpha_1 n^2 (1 - \cos(2\pi k/n)), \; k = 0, \ldots, (n-1)/2. \tag{3.7}$$

The radial representation with a first-order MRF model was used by Mardia and Qian (1995) to model the shape of leaves. A full edge representation with a complex symmetric version of the first-order MRF model was used by Rue and Hurn (1999) for modelling cells. The same model was used by Hansen, Møller and Tøgersen (2001) for the shape of the artery, but it was found that the model had too much variability and it was refined to what is essentially a constant angle representation with a first-order MRF model. Kent *et al.* (2001) and Hobolth, Kent and Dryden (2002) fitted a truncated constant length first-order MRF model and a truncated radial second-order MRF model to the shape of sand grains.

3.1.2 Continuous models

In the limiting case when n tends to infinity the sums in (3.2)-(3.4) become integrals with the discrete variable j/n replaced by a periodic variable $t \in [0,1]$, and the Fourier series expansion takes the form

$$r(t) = a_0 + \sqrt{2}\sum_{k=1}^{\infty}\big(a_k\cos(2\pi kt) + b_k\sin(2\pi kt)\big), \quad 0 \leq t \leq 1. \tag{3.8}$$

Similarly the multivariate normal model with mean zero and circulant covariance matrix is replaced by a zero mean periodic stationary Gaussian process with covariance function

$$\sigma(t) = \mathrm{Cov}(r(0), r(t)) = \lambda_0 + 2\sum_{k=1}^{\infty}\lambda_k\cos(2\pi kt). \tag{3.9}$$

Rue and Syversveen (1998) describe a procedure for identifying cells in a digital image. The standardised radius-vector function is modelled by a zero mean periodic stationary Gaussian process with covariance function

$$\sigma(t) = \begin{cases} \tau^2 e^{-\beta t}\cos(4\pi t), & 0 \leq t \leq 1/2, \\ \sigma(1-t), & 1/2 \leq t \leq 1, \end{cases} \tag{3.10}$$

where $\tau^2, \beta > 0$. In terms of eigenvalues the model is given by

$$\lambda_k = \frac{2\beta\tau^2}{\beta^2 + (2\pi k)^2} \times \begin{cases} 1 - e^{-\beta/2}, & k = 0, 2, 4, \ldots, \\ 1 + e^{-\beta/2}, & k = 1, 3, 5, \ldots. \end{cases} \tag{3.11}$$

Hobolth, Pedersen and Jensen (2000) consider various extensions of the continuous pth-order MRF model. From a first-order Taylor expansion of (3.7) it follows that the zero mean periodic stationary process obtained in the limit from the discrete first-order MRF model (3.6) is given by

$$1/\lambda_k = \alpha_0 + \alpha_1(2\pi k)^2, \ k \in \mathbb{N}_0, \ \alpha_0 > 0, \ \alpha_1 > 0.$$

In general the limiting process of a pth-order MRF model has eigenvalues given by

$$1/\lambda_k = \alpha_0 + \sum_{l=1}^{p} \alpha_l (2\pi k)^{2l},\ k \in \mathbb{N}_0,$$

where the parameters $\alpha_0, \alpha_p > 0$ and the remaining parameters $\alpha_l, l = 1, \ldots, p-1$, are chosen such that all eigenvalues are positive. The sample paths are continuous for the first-order MRF model, and in general $p-1$ times continuously differentiable for a pth-order MRF model. Note that the eigenvalues (3.11) of Rue and Syversveen (1998) are decreasing according to $p = 1$, and thus the sample paths are continuous, but not differentiable. In most applications the two parameter model with $\alpha_1 = \cdots = \alpha_{p-1} = 0$ and eigenvalues

$$1/\lambda_k = \alpha + \beta(2\pi k)^{2p},\ k \in \mathbb{N}_0,\ \alpha > 0,\ \beta > 0,$$

is sufficiently flexible. If the parameter $p > 1/2$ is allowed to take real values it determines the *fractal* dimension of the sample paths.

Letting

$$a_k = \sqrt{2A_k}\cos(2\pi k\theta_k),\ b_k = \sqrt{2A_k}\sin(2\pi k\theta_k),\ k \in \mathbb{N},$$

we obtain the continuous polar form

$$r(t) = a_0 + 2\sum_{k=1}^{\infty} \sqrt{A_k}\cos\big(2\pi k(t-\theta_k)\big),\ 0 \le t \le 1,$$

of (3.8), where $A_k = (a_k^2 + b_k^2)/2 \ge 0$ and $\theta_k \in [0, 1/k[,\ k \in \mathbb{N}$. Under the normal model A_k follows an exponential distribution with mean λ_k, while θ_k is uniformly distributed on the interval $[0, 1/k[$. In Hobolth, Pedersen and Jensen (2000) the adequacy of the exponential distribution for a particular sample of objects is investigated by considering the more general class of *generalised gamma distributions* $\Gamma(\gamma, \rho, \delta)$. The density of a $\Gamma(\gamma, \rho, \delta)$ distribution is given by

$$f(x) = \frac{\delta x^{\delta\gamma - 1}}{\Gamma(\gamma)\rho^{\delta\gamma}} \exp\big(-(\frac{x}{\rho})^{\delta}\big), \quad x > 0,$$

where $\gamma, \delta > 0$ are shape parameters and $\rho > 0$ is a scale parameter. The ordinary gamma distribution is obtained for $\delta = 1$, the Weibull distribution for $\gamma = 1$, and the exponential distribution corresponds to $\delta = \gamma = 1$. Hobolth, Pedersen and Jensen (2000) conclude that the phase amplitudes has a tendency towards heavier tails than the exponential distribution (corresponding to estimated values of δ less than 1), but the tendency is not significant.

3.2 Non-circular deformation models

Grenander and Manbeck (1993) model the local rotations $\theta_j \in \mathbb{R}$ of the elliptical template edges in (2.16) by a first-order MRF model. Rue and Hurn (1999) also work with an elliptical template. They model the edge transformation vector $\mathbf{z} \in \mathbb{C}^n$ in (2.17) by a complex normal distribution with $K_{xy} = 0$ and a first-order MRF model for the real and imaginary part of $\mathbf{z}$.

Hobolth, Pedersen and Jensen (2001) derive distributional results for the representation (2.14) of Stoyan and Stoyan (1994). Let

$$R(t) = r^0(t) + r(t),\ 0 \le t \le 1,$$

be the radius-vector function of the object, where $\{r^0(t) : 0 \le t \le 1\}$ represents the template and $\{r(t) : 0 \le t \le 1\}$ represents the stochastic fluctuation around the template. Furthermore let

$$r^0(t) = a_0^0 + \sqrt{2}\sum_{k=1}^{\infty}\big(a_k^0 \cos(2\pi kt) + b_k^0 \sin(2\pi kt)\big),$$

and

$$r(t) = a_0 + \sqrt{2}\sum_{k=1}^{\infty}\big(a_k \cos(2\pi kt) + b_k \sin(2\pi kt)\big),$$

be the corresponding Fourier series expansions. Suppose the stochastic fluctuation is a zero mean stationary Gaussian process. Then a_0 and a_k, b_k, $k \in \mathbb{N}$, are all mutually independent, $a_0 \sim N(0, \lambda_0)$ and $a_k \sim b_k \sim N(0, \lambda_k)$, $k \in \mathbb{N}$. It follows that the Fourier expansion of R has the same distributional properties as those of r, except that zero mean values are substituted by the relevant Fourier coefficients from the template.

For a polar Fourier expansion

$$R(t) = \sqrt{A_0} + 2\sum_{k=1}^{\infty}\sqrt{A_k}\cos(2\pi k(t - \theta_k)),\ 0 \le t \le 1,$$

where $A_k \ge 0$ and $\theta_k \in [0, 1/k[$ we have that A_0 and A_k, θ_k, $k \in \mathbb{N}$, are all independent. Furthermore the observed phase amplitudes

$$A_k = \begin{cases} (a_0 + a_0^0)^2, & k = 0, \\ \big((a_k + a_k^0)^2 + (b_k + b_k^0)^2\big)/2, & k \in \mathbb{N}, \end{cases}$$

follow a non-central χ^2-distribution with mean

$$EA_k = A_k^0 + \lambda_k,\ k \in \mathbb{N}_0,$$

where A_k^0 is the kth phase amplitude of the template. Finally, the conditional distribution of θ_k given A_k is given by

$$2\pi k\theta_k | A_k \sim vM(2\pi k\theta_k^0, \frac{2\sqrt{A_k A_k^0}}{\lambda_k}),\ k \in \mathbb{N},$$

where $\theta_k^0 \in [0, 1/k[$ is the kth phase angle of the template. Here, $vM(\mu, \kappa)$ stands for the von Mises distribution with mean $\mu \in [0, 2\pi[$ and concentration parameter $\kappa \geq 0$, cf. Mardia and Jupp (2000, p. 36). For $\kappa = 0$ we get the uniform distribution on $[0, 2\pi[$ while for $\kappa > 0$ large the distribution is concentrated around the mean direction. If the template is a circle, then $a_k^0 = b_k^0 = 0 = A_k^0$, $k \in \mathbb{N}$. In this case A_k and θ_k are independent, A_k follows an exponential distribution with mean λ_k and θ_k is uniformly distributed on $[0, 1/k[$, $k \in \mathbb{N}$.

By equating the first and second moments the distribution of A_k can be approximated by a $(A_k^0 + \lambda_k)\chi^2(f_k)/f_k$-distribution, where

$$f_k = 2\frac{(A_k^0 + \lambda_k)^2}{\lambda_k^2 + 2A_k^0\lambda_k},$$

cf. e.g. Jensen (1991). For a circular template it holds that $A_k^0 = 0$, $f_k = 2$, and the result is exact. If $A_k^0 \gg \lambda_k$ then f_k will be large and the distribution of A_k is concentrated around A_k^0. Note that the result justifies the use of the generalised Gamma distribution used in Hobolth, Pedersen and Jensen (2000) to model the amplitudes. The reason why a sample of phase amplitudes follows a generalised Gamma distribution, but not an exponential distribution, could be a non-circular template!

3.3 Solid deformation models

Solid deformation models are not only defined on the boundary of the object, but the *interior* is also taken into account. The *elastic deformation model* arises from physical properties of elastic materials while *Gibbs pixel-particle model* is defined in terms of motion invariant geometrical functionals.

3.3.1 Elastic deformation model

Continuum mechanics is concerned with the deformation of continuous materials under some external effect. For some materials, the deformation caused by the application of moderate loads disappears with the removal of the load; this property is known as *elasticity*. It is useful to think of the material as a sheet of rubber. Let $\Omega^0 \subset \mathbb{R}^2$ be the material (template) before the load has been applied. The idea is to consider the observed object Ω as an elastic deformation of the material, and the deformation takes the form

$$\Omega = \{(x, y) + (f(x, y), g(x, y)) : (x, y) \in \Omega^0\},$$

where $f(x, y)$ is the displacement in the x-direction and $g(x, y)$ is the displacement in the y-direction. The penalty function (also called the energy) associated with an elastic deformation is

$$E(\Omega) = \int\int_{\Omega^0} \{(\lambda + 2\mu)(f_x^2 + g_y^2) + 2\lambda f_x g_y + \mu(f_y + g_x)^2\} dxdy, \quad \mu > 0, \ \lambda \geq -\mu,$$

where μ and λ are the Lamè elastic material constants, cf. e.g. Lai *et al.* (1993). Here we use the notation $f_x = \partial f(x,y)/\partial x$ and similar for f_y, g_x and g_y. The ratio λ/μ determines the appearance of the deformations. Suppose the boundary of the material has been displaced at a localised place. If λ/μ is small the resulting deformation is very localised, remaining in the region of the original displacement, whereas when the ratio increases, the deformation spreads further throughout the whole region.

Another example of a penalty function is the simple sum of squared derivatives

$$E(\Omega) = \int\int_{\Omega^0} \{f_x^2 + f_y^2 + g_x^2 + g_y^2\} dxdy,$$

cf. e.g. Glasbey and Mardia (1998). Glasbey and Mardia (2001) suggest using these penalties as priors to ensure smooth transformations when warping images. They also suggest various modified forms of the penalty functions in order to ensure desired properties such as invariance under bilinear or affine transformations.

Godwin (2000) demonstrate how the elastic deformation model can be used to recover the deformation of the interior of an object when only the boundary displacements are known. Miller *et al.* (1993) seek a transformation of one anatomy (e.g. a template brain) into another (e.g. the brain of a patient), and use an elastic deformation model as a prior for the transformation. When the displacements are not too large a penalty based on the energy of an elastic body ensures a resulting smooth deformation, but the elastic deformation model may result in a 'folding' of the material for severe or large deformations. Christensen *et al.* (1996) use fluid dynamics to give a method for handling severe or large deformations.

3.3.2 Gibbs pixel-particle model

In real data examples a planar object Ω is often represented as a connected set of pixels without holes on a lattice $\mathbb{Z}^2$. The Gibbsean probability measure of a pixel-particle is given by

$$P(\Omega) = \frac{1}{Z} e^{-E(\Omega)}, \quad Z = \sum_{\Omega} e^{-E(\Omega)},$$

where E is the penalty function (or energy) and Z is a normalizing constant often referred to as the partition function. We assume that each admissible configuration Ω includes all spatial shifts of the same particle, and the sum is taken over all these admissible configurations.

Many interesting choices of penalty functions can be defined. Anastassiou and Sakrison (1981) consider the choice

$$E(\Omega) = bU(\Omega) + h\sqrt{A(\Omega)}, \tag{3.12}$$

where $U(\Omega)$ is the length of the perimeter, $A(\Omega)$ is the area, and b and h are parameters of the model. Bearing in mind that area and perimeter length are perhaps the two most important quantities when describing a planar object and that the area-perimeter ratio $A(\Omega)/U(\Omega)^2$ is an efficient shape-ratio for determining the deviation from circular shape this choice seems sensible. In applications $A(\Omega)$ is simply the number of pixels in Ω, while more care has to be taken when calculating the perimeter length. In particular a line with a 45^0 tilt with respect to a horizontal line will appear in a staircase manner with a length equal to $\sqrt{2}$ times its real length, independent of the resolution of the lattice. It is interesting to note that

$$\text{mean}\big(U(\Omega)\big) = \sum_{\Omega} P(\Omega)U(\Omega) = \frac{1}{Z}\sum_{\Omega} e^{-bU(\Omega)-h\sqrt{A(\Omega)}}U(\Omega) = -\frac{1}{Z}\frac{\partial Z}{\partial b},$$

and thus a statistical quantity such as the mean boundary length can be expressed in terms of the partition function $Z(b,h)$. Similar formulas are available for the second moment of the boundary length and first and second moments of the square root of the area, but unfortunately no closed form expression for the partition function exists.

Stoyan *et al.* (2001) consider a penalty function given by

$$E(\Omega) = \frac{(A(\Omega)-\mu_A)^2}{2\sigma_A^2} + \frac{(U(\Omega)-\mu_U)^2}{2\sigma_U^2} + \frac{(T(\Omega)-\mu_T)^2}{2\sigma_T^2}, \tag{3.13}$$

where $T(\Omega)$ is the number of towers and hollows, μ_A, μ_U, μ_T are preferred area, perimeter and tower parameters, while $\sigma_A, \sigma_U, \sigma_T$ are the remaining model parameters. Note that the energy only depends on motion invariant geometrical functionals. Anastassiou and Sakrison (1981) and Stoyan *et al.* (2001) give examples of random samples from the probability distributions determined by (3.12) and (3.13) for various choices of model parameters. Since the probability distribution is only known up to a constant they apply Metropolis-Hastings algorithms (Gilks *et al.*, 1996) to simulate from the model. It still remains to clarify how statistical inference should be performed in the class of Gibbs pixel-particle models.

4 3D deformable template representations

Consider a regular surface $\mathcal{M} \subset \mathbb{R}^3$, which is the boundary of a solid object in three-dimensional space, and let $\mathcal{M}^0 \subset \mathbb{R}^3$ be a regular template surface. Loosely speaking the boundary of a solid object is a regular surface if it can be obtained by taking pieces of a plane, deforming them in a smooth way and arranging them so that the resulting surface has no sharp points, edges or self-intersections, cf. Carmo (1976, p. 52). As in the planar case the basic idea is to describe $\mathcal{M}$ as a stochastic deformation of $\mathcal{M}^0$. It is not immediate obvious how to generalise the planar situation to three-dimensional space, the main obstacle being the loss of index-time. Whereas the vertices in the plane can be arranged in an anti-clockwise order, there is no simple way of arranging the vertices on a regular surface. In this section we first discuss the situation where the observed objects are star-shaped and the template surface is the unit sphere. Next we discuss the non-spherical case.

4.1 Deformable sphere representations

Suppose the surface $\mathcal{M} \subset \mathbb{R}^3$ is star-shaped relative to $c \in \mathbb{R}^3$, i.e. the radial vectors from c to the boundary all remain inside the object. Using spherical coordinates the surface is determined by

$$\{c + d(\theta,\phi)\omega(\theta,\phi) : \ 0 \leq \theta < 2\pi, \ 0 \leq \phi \leq \pi\},$$

where $\omega(\theta,\phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)$ is the vector on the unit sphere with polar longitude θ and polar latitude ϕ, and $d(\theta,\phi)$ is the distance from c to $\mathcal{M}$ in direction $\omega(\theta,\phi)$. It is useful to express the radius-vector function $d(\theta,\phi)$ in terms of the spherical harmonics

$$\{\varphi_n^m(\theta,\phi) : \ n \in \mathbb{N}_0, \ m = -n, \ldots, n\},$$

which constitute an orthonormal basis on the sphere. The spherical harmonics are given by

$$\varphi_n^m(\theta,\phi) = \begin{cases} k_n^{|m|} P_n^{|m|}(\cos\phi)\cos m\theta, & m = -n, \ldots, -1 \\ k_n^0 P_n^0(\cos\phi), & m = 0 \\ k_n^m P_n^m(\cos\phi)\sin m\theta, & m = 1, \ldots, n, \end{cases}$$

where

$$k_0^0 = \frac{1}{\sqrt{2\pi}}, \ k_n^0 = \sqrt{\frac{2n+1}{4\pi}}, \ k_n^m = \sqrt{\frac{2n+1}{2\pi}\frac{(n-m)!}{(n+m)!}}, \ n \in \mathbb{N}, \ m = 1, \ldots, n, \quad (4.1)$$

are normalizing constants and P_n^m are the associated Legendre functions of the first kind. Now consider the Fourier-Legendre series expansion

$$d(\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(\theta,\phi)$$

of the radius-vector function, where the Fourier-Legendre coefficients are given by

$$a_n^m = \int_0^{2\pi} \int_0^{\pi} d(\theta,\phi)\varphi_n^m(\theta,\phi)\sin\phi d\phi d\theta. \quad (4.2)$$

The mean radius-vector length is determined by

$$d_{av} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} d(\theta,\phi)\sin\phi d\phi d\theta = \frac{1}{\sqrt{2\pi}} a_0^0,$$

and hence a_0^0 can be used as a measure of the size of the object. To remove size one may consider the standardised radius-vector function $d(\theta,\phi)/d_{av} - 1$.

As in the planar case the choice of centre implies constraints on the radius-vector function. If c is the centre of mass of the object then the constraints can be written explicitly, as shown in the appendix of Hobolth (2002). In this case the constraints involve the fourth power of the radius-vector function, but assuming that the object is a small deformation of a sphere, a first-order Taylor expansion leads to the approximate constraints $a_1^m \approx 0, \ m = -1, 0, 1$.

4.2 Deformable non-spherical representations

For a small deformation of a regular template surface it may suffice to extend the planar representation (2.15) of Hobolth and Jensen (2000) to three-dimensional space. Then the object $\mathcal{M}$ can be represented by

$$\mathcal{M} = \{x + U(x)\omega^0(x) : x \in \mathcal{M}^0\}, \tag{4.3}$$

where $\omega^0(x)$ is the outer unit normal vector to $\mathcal{M}^0$ at $x \in \mathcal{M}^0$. The function $\{U(x) \in \mathbb{R} : x \in \mathcal{M}^0\}$ determines the transformation from $\mathcal{M}^0$ to $\mathcal{M}$.

Joshi *et al.* (1997) suggest another representation of non-spherical objects. In order to describe the shape of the cortical and hippocampal surfaces of macaque and human brains they first establish a cortical or hippocampal template $\mathcal{M}^0$. Second a complete orthonormal basis φ_n, $n \in \mathbb{N}$, of the template is determined. Joshi *et al.* (1997) suggest choosing the basis functions to correspond to eigenfunctions associated with a differential operator L, derived from thin elastic shell theory. Thus $L\varphi_n = \eta_n\varphi_n$, $n \in \mathbb{N}$, where η_n is the eigenvalue associated with the eigenfunction φ_n. A cortical or hippocampal surface is then represented by

$$\mathcal{M} = \{x + U(x) : x \in \mathcal{M}^0\}, \tag{4.4}$$

where the function $\{U(x) \in \mathbb{R}^3 : x \in \mathcal{M}^0\}$ is given by

$$U(x) = \sum_{n=1}^{\infty} a_n\varphi_n(x),\ x \in \mathcal{M}^0, \tag{4.5}$$

in terms of the orthonormal basis φ_n, $n \in \mathbb{N}$. Note that while the value of the transformation function $U(x)$ is real in (4.3) it belongs to $\mathbb{R}^3$ in (4.4).

5 3D deformation models

In this section we introduce non-parametric and parametric rotational symmetric models for the standardised radius-vector function. Next the non-spherical representations are considered and models for the transformation functions $\{U(x) : x \in \mathcal{M}^0\}$ are discussed.

5.1 Spherical deformation models

Consider the Fourier-Legendre series expansion of the standardised radius-vector function

$$r(\theta, \phi) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(\theta, \phi).$$

As argued in Subsection 4.1 the three Fourier-Legendre coefficients corresponding to $n = 1$ are approximately zero if c is the centre of mass and the object is a small deformation of a sphere. Alternatively one could choose the centre as the point satisfying $a_1^m = 0,\ m = -1, 0, 1$ (assuming the object is star-shaped relative to this point). In any case these coefficients are treated as non-random nuisance parameters, and the remaining coefficients

$$a_n^m \sim N(0, \lambda_n^m),\ n \geq 2,\ m = -n, \ldots, n,$$

are modelled as independent Gaussian random variables with mean zero (the average shape is a sphere) and variance λ_n^m. We also suppose that we have *stationarity on the sphere*, in the sense that the covariance between two points on the sphere depends only on the angle between the points. Stationarity is obtained by assuming

$$\lambda_n^m = \lambda_n \geq 0,\ n \geq 2,\ m = -n, \ldots, n, \tag{5.1}$$

and the covariance becomes

$$\begin{aligned} \mathrm{Cov}\big(r(\theta_1,\phi_1), r(\theta_2,\phi_2)\big) &= \sum_{n=2}^{\infty} \lambda_n \sum_{m=-n}^{n} \varphi_n^m(\theta_1,\phi_1)\varphi_n^m(\theta_2,\phi_2) \\ &\overset{(*)}{=} \sum_{n=2}^{\infty} \lambda_n (k_n^0)^2 P_n(\cos\psi_{12}), \end{aligned} \tag{5.2}$$

where $\cos\psi_{12} = \omega(\theta_1,\phi_1) \cdot \omega(\theta_2,\phi_2)$, and we at $(*)$ have used the *addition theorem*, cf. Müller (1966, Theorem 2).

The covariance is thus determined by the variances λ_n, and to proceed further we seek a parametric model for the variances. Miller *et al.* (1994) use a model induced from Poisson's equation for pressure fields acting on thin membranes. The potential associated with the model is given by

$$E(r) = \frac{1}{2} \int_0^{2\pi} \int_0^{\pi} |Lr(\theta,\phi)|^2 \sin\phi d\phi d\theta,$$

where

$$L = \frac{\partial^2}{\partial \theta^2} + \frac{\cos\theta}{\sin\theta}\frac{\partial}{\partial \theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial \phi^2}$$

is the Laplacian operator on the surface of the sphere. Since $\varphi_n^m(\theta,\phi)$ is an eigenfunction of the Laplacian operator with eigenvalue $\eta_n = -n(n+1)$ the potential corresponds to a model where the variances decrease as $1/\lambda_n = \eta_n^2 = (n(n+1))^2$.

Grenander and Miller (1998, Section 5.3) suggest obtaining more general models by introducing polynomials $p(L) = a_0 I + a_1 L + \cdots + a_d L^d$ of the basic operator L. The bi-harmonic operator is for example obtained by choosing $p(L) = L^2$. For such models the variances decrease as $1/\lambda_n = p(\eta_n)^2$.

Hobolth (2002) suggest letting the variances decrease according to

$$1/\lambda_n = \alpha + \beta n^p,\ n \geq 2,\ p > 2,\ \beta > 0,\ \alpha > -\beta 2^p. \tag{5.3}$$

There are several reasons why this is expected to be a good model. The parameter p makes the model very flexible with regard to the smoothness of the radius-vector function. From Stein (1999, Chapter 2) it follows that the degree of mean square differentiability is determined by the behaviour of the covariance function and its derivatives near the origin. By making repeated use of the relation

$$P'_{n+1}(x) = (2n+1)P_n(x) + P'_{n-1}(x), \quad -1 \leq x \leq 1,$$

and using $P_0(x) = 1$, $P_1(x) = x$ and $P_n(1) = 1$ it can be shown that the radius-vector function is k times mean square differentiable when $2(k+1) < p \leq 2(k+2)$. When $2 < p \leq 4$ the radius-vector function is mean square continuous. Note that in the model used by Miller *et al.* (1994) the variances decrease according to $p = 4$, while for the bi-harmonic operator the variances decrease according to $p = 8$.

For fixed p the value of β determines the 'local' shape of the object since the variances with large indices are determined by β. The third parameter in the model is most easily interpreted when making the reparametrization $\tilde{\alpha} = \alpha + \beta 2^p$, in which case $\tilde{\alpha}$ controls the first few variances, and thereby the 'global' shape of the object.

5.2 Non-spherical deformation models

For the normal representation of Hobolth and Jensen (2000) the transformation $\{U(x) \in \mathbb{R} : x \in \mathcal{M}^0\}$ is considered a random field on a regular surface. It seems appropriate to let the covariance $\mathrm{Cov}(U(x_1), U(x_2))$, $x_1, x_2 \in \mathcal{M}^0$, depend on geometrical quantities such as the geodesic distance on $\mathcal{M}^0$ between x_1 and x_2 and the principal curvatures and directions at x_1 and x_2, but an explicit model still needs to be formulated.

Joshi *et al.* (1997) consider two models for the Gaussian random variables a_n, $n \in \mathbb{N}$ in (4.5). In both models the random variables are independent with zero means, but in the first model the variances λ_n are chosen to be the inverse of the squared eigenvalues $1/\lambda_n = \eta_n^2$ and in the second model the variances are estimated empirically from a sample of surfaces. The relation to the spherical deformation model is obtained by letting $\mathcal{M}^0 = S^2$ and letting the operator L be the Laplacian operator with eigenfunctions φ_n, $n \in \mathbb{N}$, equal to the spherical harmonics. In the spherical deformation model the independence assumption is reasonable on ground of rotational symmetry, but in general the assumption seems rather arbitrary and should be justified.

6 Applications

In this section two applications of deformable template models are considered. First the problem of recognising an object in a noisy or blurred image is treated. We take a Bayesian approach and use deformable templates as prior distributions. Second the problem of variance estimation in stereology is discussed, and it is demonstrated how tools from model-based stereology may improve the methods used at present in design-based stereology.

6.1 Bayesian object recognition

In this subsection we discuss the Bayesian object recognition problem following Kent, Mardia and Walder (1996). Suppose we do not observe the outline itself, but a digital image

$$y = \{y_s : s = (s_1, s_2), 1 \leq s_1 \leq S_1, 1 \leq s_2 \leq S_2\}.$$

A simple observation model is

$$y_s \sim N(\mu_1, \sigma^2),\ s \in \Omega,\ \ y_s \sim N(\mu_2, \sigma^2),\ s \notin \Omega,$$

where Ω is the interior of the polygon $P = (v_j)$ in the plane defined by the vertices (v_j). This model is called the blur-free independent noise model, and the log-likelihood is given by

$$L = -\frac{1}{2\sigma^2}\Big(\sum_{s\in\Omega}(y_s - \mu_1)^2 + \sum_{s\notin\Omega}(y_s - \mu_2)^2\Big). \tag{6.1}$$

Suppose the object is given in terms of the standardised radial representation (2.3), with u_j replaced by r_j, together with a specification of scale and orientation d_{av} and position v_{av}. Furthermore assume the standardised radial vector $\mathbf{r}$ follows the first-order MRF model (3.6). Then the log-prior becomes

$$Q = -\frac{1}{2}\Big(\alpha_0\frac{1}{n}\sum_{j=0}^{n-1} r_j^2 + \alpha_1 n\sum_{j=0}^{n-1}(r_{j+1} - r_j)^2\Big),\ \ \alpha_0 > 0,\ \alpha_1 > 0, \tag{6.2}$$

and the log-posterior density of P given the image y equals

$$Q + L, \tag{6.3}$$

plus an irrelevant constant. The goal in Bayesian object recognition is often to determine the maximum a posteriori (MAP) estimate which maximizes (6.3) over P. A locally-based iterative method for finding the MAP estimate can be developed by updating the values of r_j, $j = 0, \ldots, n-1$, several times. This algorithm is an example of the iterative conditional modes (ICM) algorithm of Besag (1986), and a good starting point is essential for the algorithm to achieve useful results.

During the last decade this simple Bayesian object recognition problem has been generalised in various ways. In order to handle digital images with several objects Baddeley and van Lieshout (1993) embedded the deformable template model in a marked point process framework. For a marked point process each object is specified as a marked point, where the point gives the location and the mark determines the shape and size of the object.

The point process is responsible for the number of objects and the spatial relationship between objects. The reference process is the Poisson point process, where the number of objects are Poisson distributed and, conditional on this number, the locations of the objects are independent and uniformly distributed. To avoid clustering of objects around each true object (called 'multiple response') the area-interaction point process, the hard-core point process or the Strauss point process are often used. Baddeley and van Lieshout (1993) in an application concerning pellets in an image use the Strauss point process. The pellets are modelled by a circle of fixed radius and the observation model is a blurred independent noise model.

The distribution of the marks is often given by a deformable template model. Examples include Mardia *et al.* (1997) where mushrooms in a growing bed are analysed. The mushrooms are modelled by the radial representation with a Gaussian first-order MRF model. In Rue and Syversveen (1998) cells in a confocal microscopy image are located and the radial representation with the zero mean stationary Gaussian radius-vector process given by (3.10) is used. A third example is provided by Rue and Hurn (1999), where two types of cells are located. The templates are either a circle or an ellipse with fixed eccentricity and the edge tranformations follow a Gaussian first-order MRF model.

Perhaps the prior model has attracted too much attention in the literature compared to the treatment of the observation model. A first extension is to relax the independence assumption of the observation model, and for this purpose a local MRF model where neighbouring pixels are dependent is often used. For an application of a combined low-level MRF model and a high-level deformable template model see Qian, Titterington and Chapman (1996), where an irregular boundary of a magnetic domain is identified. Husby (2001) also use carefully defined realistic observation models together with a deformable templates in applications concerning ultrasound images.

There are close relations between the theory of snakes (Kass *et al.*, 1988) and Bayesian object recognition with Markov Random Field priors. A snake is a curve in an image which minimizes a certain energy. Often the energy takes the form (6.3), where (6.1) is called the image energy and connects the snake to the image, the first term in (6.2) forces the snake to have a shape similar to a template snake, and the second term in (6.2) ensures the snake to be smooth. Several other energies can be added in order to ensure certain desired properties of the snake.

6.2 Model-based stereology

Stereology is a collection of methods for making inference about a population of spatial objects from geometric samples of the objects such as line and plane sections. The interest is in quantitative properties such as volume or surface area in three-dimensional space and area or boundary length in two-dimensional space. If a typical object from the population can be regarded as a realization of a stochastic process R, then the quantity of interest can be expressed as a function f of R. Using a geometrical design ϕ, independent of R, a design-unbiased predictor $\hat{f}(R, \phi)$ of $f(R)$ can often be constructed such that $\hat{f}(R, \phi)$ has the property

$$\mathrm{E}(\hat{f}(R, \phi)|R) = f(R).$$

To fix ideas let $R = \{R(2\pi t) : 0 \leq t \leq 1\}$ be some power of the radius-vector function and suppose the parameter of interest is $f(R) = \int_0^1 R(2\pi t)dt$. For instance the area of the boundary can be represented in this manner. Furthermore suppose R is only measured in the points $\phi + j/n$, $j = 0, \ldots, n-1$, with ϕ uniformly distributed in $[0, 1/n]$. Then

$$\hat{f}(R, \phi) = \frac{1}{n} \sum_{j=0}^{n-1} R(2\pi(\phi + j/n)). \tag{6.4}$$

is design-unbiased for $f(R)$.

In most cases it is of interest to make statements about the population of objects and not only about the sampled objects. A relevant quantity is here the prediction error

$$\mathrm{E}(\hat{f}(R, \phi) - f(R))^2.$$

Using that $\hat{f}(R, \phi)$ is design-unbiased the prediction error can be rewritten as

$$\begin{aligned} &\mathrm{E}(\hat{f}(R, \phi) - f(R))^2 \\ &= \mathrm{Var}\big(\hat{f}(R, \phi) - f(R)\big) \\ &= \mathrm{Var}\big(\mathrm{E}(\hat{f}(R, \phi) - f(R)|R)\big) + \mathrm{E}\big(\mathrm{Var}(\hat{f}(R, \phi) - f(R)|R)\big) \\ &= \mathrm{E}\big(\mathrm{Var}(\hat{f}(R, \phi)|R)\big) \\ &= \mathrm{E}\sigma_R^2. \end{aligned} \tag{6.5}$$

It is also part of the methodology of design-based stereology to construct a design-unbiased estimator $\hat{\sigma}_R^2(\phi)$ of the conditional variance

$$\sigma_R^2 = \mathrm{Var}(\hat{f}(R, \phi)|R). \tag{6.6}$$

From (6.5) and (6.6) it is evident that $\hat{\sigma}_R^2(\phi)$ or an average of such estimators for a sample of objects can be regarded as an unbiased estimator of the prediction error. In design-based stereology the predictor $\hat{\sigma}_R^2(\phi)$ is often based on the empirical covariogram, cf. Gual-Arnau and Cruz-Orive (2000), but Hobolth and Jensen (2001) demonstrate that a model-based statistical setting may lead to more efficient estimators.

Gual-Arnau and Cruz-Orive (2000) use the design-based setting. Translated to a model-based setting they suggest to model R by a stationary, random periodic process with mean μ and covariance function (3.9) with variances given by

$$\lambda_0 = \beta_0 - 2\sum_{k=1}^{\infty} \lambda_k, \ \lambda_k = \frac{(2p)!}{k^{2p}}\beta, \ k \in \mathbb{N},$$

where λ_0, $\beta > 0$ and $p \in \mathbb{N}$. In terms of model parameters Hobolth and Jensen (2001) show that the prediction error is given by

$$\mathrm{E}(\hat{f}(R,\phi) - f(R))^2 = (-1)^{p-1}(2\pi)^{2p} B_{2p} \frac{1}{n^{2p}} \beta, \tag{6.7}$$

where B_{2p} is the Bernoulli number of order $2p$. Hobolth and Jensen (2001) suggest using the maximum likelihood estimate of $\hat{\beta}$ of β. Taking aliasing into account

$$\hat{\beta} = \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{a_k^2 + b_k^2}{\kappa_k} \sim \beta \chi^2(n-1)/(n-1), \tag{6.8}$$

where

$$\kappa_k = \sum_{j \in \mathbb{Z}} \frac{(2p)!}{(k+nj)^{2p}}, \quad k = 1, \ldots, n-1,$$

and a_k, b_k are given by (3.3) and (3.4). Since $\hat{\beta}$ is a function of the sufficient statistics it follows from the Rao-Blackwell theorem that it is also the unique unbiased minimum variance estimator of β. By plugging the estimator (6.8) of β into (6.7) we obtain an estimator of the prediction error, and since the distribution of $\hat{\beta}$ is known we can also supply the point estimator with a confidence interval.

7 Spatio-temporal shape models

In recent years the evolution of shapes over time has been studied, and there is an extensive engineering literature dealing with space-time noisy image sequences. In these studies the Bayesian approach is often used. Kervrann and Heitz (1999) present a model for the evolution in *shape and motion* of the heart ventricle where the prior consists of a deformable template representation of the heart ventricle and the motion is modelled by an affine velocity field. Furthermore Kervrann and Heitz (1999) derive a procedure for finding maximum likelihood estimates for shape and motion parameters.

Another challenging problem is to model the *growth of shapes*. Cressie and Hulting (1992) define a random-set growth model to describe the growth of human breast cancer tumors. A tumor X_t at time t is a Boolean model

$$X_t = \cup_{i=1}^{n} D \oplus s_i,$$

with circular grains D of fixed radius and germs $\{s_i \in \mathbb{R}^2 : i = 1, \ldots, n\}$. Here we use the notation $D \oplus s_i = \{d + s_i : d \in D\}$. The model is Markovian in time and the step from X_{t-1} to X_t is given by letting the number of grains n be Poisson distributed with parameter $\lambda \cdot \mathrm{area}(X_{t-1})$ and letting the germs be independent and uniformly distributed in X_{t-1}. A radial growth model somewhat similar in spirit to Cressie and

Hulting (1992) is defined in Stoyan *et al.* (2002), where small sub-particles of varying shapes and sizes are added to the radius-vector function.

The growth of star-shaped objects in three-dimensional space can be described by the radius-vector function $d(\theta, \phi)$ with polar longitude θ and polar latitude ϕ considered in Subsection 4.1. The dynamic radius-vector function is given by

$$d(\theta, \phi; t) = d_{av}(t)\big(1 + \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_n^m(t)\varphi_n^m(\theta, \phi)\big), \;\; t \in \mathbb{N}_0, \; 0 \leq \theta \leq 2\pi, \; 0 \leq \phi \leq \pi,$$

where $d_{av}(t)$ determines the size and $\{a_n^m(t) : n \in \mathbb{N}, m = -n, \ldots, n\}$ the shape of the object at time t. Several models for the shape of an object at a given time were discussed in Subsection 5.1, and interesting questions for a dynamic model include how size and shape changes over time, and whether shape depends on size. If shape does not change over time, then a reasonable model for $\{a_n^m(t) : t \in \mathbb{N}\}$ could be a general pth-order autoregressive $AR(p)$ process.

References

Anastassiou, D. and Sakrison, D.J. (1981). A probability model for simple closed random curves. *IEEE Transactions on information theory*, **27**(3), 376-381.

Baddeley, A.J. and van Lieshout, M.N.M. (1993). Stochastic geometry models in high-level vision. In *Statistics and Images, Advances in Applied Statistics, a supplement to the Journal of Applied Statistics*, K.V. Mardia and G.K. Kanji (eds), vol. 20, Chapter 11, 231-261. Carfax publishing.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J.R. Statist. Soc.* B, **36**, 192-236.

Besag, J.E. (1986). On the statistical analysis of dirty pictures. *J.R. Statist. Soc.* B, **48**, 259-302.

Carmo, M.P.d. (1976). *Differential geometry of curves and surfaces.* Prentice-Hall, New Jersey.

Christensen, G.E., Rabbitt, R.D. and Miller, M.I. (1996). Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing, **5**(10), 1435-1447.

Cressie, N. and Hulting, F.L. (1992). A spatial statistical analysis of tumor growth. *American Statistical Assoc.*, **87**, 272-283.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.G. (1996). *Introducing Markov Chain Monte Carlo.* Chapman and Hall, London.

Glasbey, C.A. and Mardia, K.V. (1998). A review of image warping methods. *Journal of Applied Statistics*, **25**, 155-171.

Glasbey, C.A. and Mardia, K.V. (2001). A penalized likelihood approach to image warping. *J.R. Statist. Soc.* B, **63**, 465-514.

Godwin, D.J. (2000). Deformations in shape analysis. Ph.D. thesis, University of Leeds.

Goodman, N.R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.*, **34**, 152-177.

Grenander, U., Chow, Y. and Keenan, D.M. (1991). *Hands: A Pattern Theoretic Study of Biological Shapes.* Research notes on Neural Computing. Springer: Berlin.

Grenander, U. and Manbeck, K.M. (1993). A stochastic shape and color model for defect detection in potatoes. *J. Comp. Graph. Statist.*, **2.** 131-151.

Grenander, U. and Miller, M. (1994). Representations of knowledge in complex systems (with discussion). *J.R. Statist. Soc.* B, **56**, 549-603.

Grenander, U. and Miller, M.I. (1998). Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, **5**, 617-694.

Gual-Arnau, X. and Cruz-Orive, L.M. (2000). Systematic sampling on the circle and on the sphere. *Adv. Appl. Prob. (SGSA)*, **32**, 628-647.

Hansen, M.B., Møller, J. & Tøgersen, F.Aa. (2000). Bayesian contour detection in a time series of ultra-sound images through dynamic deformable template models. *Research Report* **17**, Centre for Mathematical Physics and Stochastics, University of Aarhus. To appear in *Biostatistics.*

Hobolth, A. (2002). The spherical deformation model. Research report no. 18, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

Hobolth, A. and Jensen, E.B.V. (2000). Modelling stochastic changes in curve shape, with an application to cancer diagnostics. *Adv. Appl. Prob. (SGSA)*, **32**, 344-362.

Hobolth, A. and Jensen, E.B.V. (2001). A note on design-based versus model-based variance estimation in stereology. Research report no. 17, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

Hobolth, A., Kent, J.T. and Dryden, I.L. (2002). On the relation between edge and vertex modelling in shape analysis. To appear in *Scand. J. Statist.*.

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2000). A continuous parametric shape model. Research report no. 13, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2001). A deformable template model, with special reference to elliptical templates. Research report no. 14, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

Husby, O.K. (2001). High-level image models with applications to medical ultrasonics. Ph.D. thesis, Norwegian University of Science and Technology.

Jensen, J.L. (1991). A large deviation-type approximation for the "Box class" of likelihood ratio criteria. *Journal of American Statistical Association*, **86**, 437-440.

Joshi, S.C., Miller, M.I. and Grenander, U. (1997). On the Geometry and Shape of Brain Submanifolds. *International Journal of Pattern Recognition and Artificial Intelligence.* **11**, 1317-1343.

Kass, M., Witkin, A. and Terzopoulos, D. (1988). Snakes: Active contour models. *Int. J. Computer Vision,* **1**, 321-331.

Kent, J.T., Dryden, I.L. and Anderson, C.R. (2000). Using circulant symmetry to model featureless objects. *Biometrika*, **87**, 527-544.

Kent, J.T., Mardia, K.V. and Walder, A.N. (1996). Conditional cyclic Markov random fields. *Adv. Appl. Probab. (SGSA)*, **28**, 1-12.

Kervrann, C. and Heitz, F. (1999). Statistical deformable model-based segmentation of image motion. *IEEE Transactions on Image Processing,* **8**(4), 583-588.

Lai, W.M., Rubin, D. and Krempl, E. (1993). *Introduction to continuum mechanics,* 3rd edn, Pergamon Press, New York.

Mardia, K.V. and Jupp, P.E. (2000). *Directional statistics.* Chichester: Wiley.

Mardia, K.V. and Qian, W. (1995). Bayesian method for compact object recognition from noisy images. In *Complex stochastic systems in science and engineering*, D.M. Titterington (ed), 155-165, Clarendon Press, Oxford.

Mardia, K.V., Qian, W., Shah, D. and de Souza, K.M.A. (1997). Deformable template recognition of multiple occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 1035-1042.

Miller, M.I., Christensen, G.E., Amit, Y. and Grenander, U. (1993). Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Science,* USA, **90**, 11944-11948.

Miller, M.I., Joshi, S., Maffitt, D.R., McNally, J.G. and Grenander, U. (1994). Membranes, mitochondria and amoeba: shape models. In *Advances in Applied Statistics*, vol. II, K. Mardia (ed), 141-163, Carfax Publishing, Abingdon, Oxfordshire, England.

Müller, C. (1966). *Spherical harmonics.* Springer-Verlag, Berlin.

Qian, W., Titterington, D.M. and Chapman, J.N. (1996). An image analysis problem in electron microscopy. *Journal of the American Statistical Association*, **91**, 944-952.

Rue, H. and Hurn, M.A. (1999). Bayesian object recognition. *Biometrika*, **86**, 649-660.

Rue, H. and Syversveen, A.R. (1998). Bayesian object recognition with Baddeley's delta loss. *Advances in Applied Probability (SGSA)*, **30**, 64-84.

Stein, M.L. (1999). *Interpolation of Spatial Data.* Springer, New York.

Stoyan, D., Davtyan, A. and Turetayev, D. (2002). Statistics for random domains and particles. In *Morphology of Condensed Matter. Physics and Geometry of Spatially Complex Systems. Lecture Notes in Physics.* K. Mecke and D. Stoyan (eds).

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes and point fields.* Chichester, Wiley.

Zahn, C.T. and Roskies, R.Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers.* **21**, 269-281.

A

Hoboth, A. and Jensen, E.B.V. (2000).
Modelling stochastic changes in curve shape,
with an application to cancer diagnostics.
Adv. Appl. Prob. (SGSA), **32**, 344-362.

Modelling stochastic changes in curve shape, with an application to cancer diagnostics

Asger Hobolth and Eva B. Vedel Jensen
Laboratory for Computational Stochastics
University of Aarhus

Abstract

Often, the statistical analysis of the shape of a random planar curve is based on a model for a polygonal approximation to the curve. In the present paper, we describe instead the curve as a continuous stochastic deformation of a template curve. The advantage of this continuous approach is that the parameters in the model do not relate to a particular polygonal approximation. A somewhat similar approach has been used in Kent *et al*. (1996), who describe the limiting behaviour of a model with a first-order Markov property as the landmarks on the curve become closely spaced, see also Grenander (1993). The model studied in the present paper is an extension of this model. Our model possesses a second-order Markov property. Its geometrical characteristics are studied in some detail and an explicit expression for the covariance function is derived. The model is applied to the boundaries of profiles of cell nuclei from a benign tumour and a malignant tumour. It turns out that the model with the second-order Markov property is the most appropriate, and that it is indeed possible to distinguish between the two samples.

Keywords: Cancer diagnostics, cyclic stationarity, deformable templates, featureless objects, Markov random field, shape, stochastic geometry.

1 Introduction

In the grading of malignancy of cancer tissue, many morphological parameters may be used. At low magnification the architecture of the cellular tissue may be considered while on a medium scale of magnification the variation in cell nuclear size is an important feature. At high magnification the shape, size and colour of each single nuclei are studied. Stereological techniques may be used to determine the size of nuclei, cf. e.g. Srensen (1991) and Jensen (1998). The remaining parameters are, however, usually subjectively estimated by the pathologists without using any quantitative methods. The pathologists' opinions may differ, and particularly in the cases where the malignancy is in an intermediate stage this leads to different grading. In order to make precise diagnoses there is therefore a need for supplementary methods, which may objectively quantify important features at each scale of magnification. This paper presents a method of describing stochastic changes of the shape of solid objects in the plane with no obvious landmarks. The method is applied to nuclear profiles, obtained by sectioning normal tissue and cancer tissue from the human skin.

The basic idea is to model an observed planar curve $\mathcal{F} = \{F(t) : 0 \leq t \leq T\}$, which is the boundary of a solid object, as a stochastic deformation of a non-random closed template curve $\mathcal{C} = \{c(t) : 0 \leq t \leq T\}$. One of the stochastic geometry models considered takes the form

$$F(t) = c(t) + X(t)\omega(t), \quad 0 \leq t \leq T,$$

where $\omega(t)$ is the inner unit normal vector to $\mathcal{C}$ at $c(t)$, cf. Figure 1. Note that $X(t)$ is the signed distance between $F(t)$ and $c(t)$, and may therefore be regarded as a residual. The challenge is to model the residual process $\{X(t) : 0 \leq t \leq T\}$.

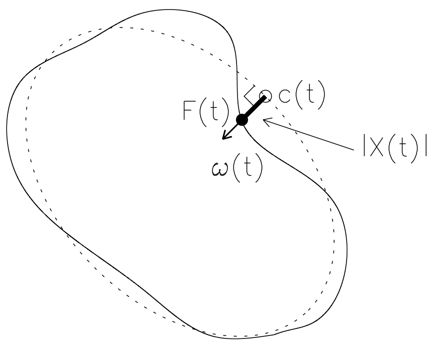

Figure 1: The observed curve is a realisation of a stochastic deformation of a template.

We will assume that the residual process is distributed as

$$\{TX_1(t/T) : 0 \leq t \leq T\},$$

where the distribution of the 'normalized' process $\{X_1(t) : 0 \leq t \leq 1\}$ belongs to a parametrized class $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of distributions of cyclic and stationary stochastic processes with zero mean. Since the curves considered are closed, the residual process should be cyclic. In the application, the residual process is assumed to be Gaussian with zero mean and with a second-order Markov property. Recall that a Gaussian process $\{X(t)\}$ with zero mean is stationary if and only if the covariance between $X(t)$ and $X(s)$ depends on (t, s) through $s - t$ only. Stationarity is a natural requirement when the object has no obvious landmarks and hence no reference point.

Note that under translations, rotations and rescaling in the plane, the normalized process $\{X_1(t) : 0 \leq t \leq 1\}$ remains unchanged. Therefore, this process describes the stochastic changes in curve shape of the observed curve $\mathcal{F}$ relative to the template curve $\mathcal{C}$, and the changes can be quantified by estimating the parameter of the distribution P_θ of the normalized process. One of the major advantages of this continuous type model is its independence from the need to specify the number of landmarks. The idea of relating to a continuous process has also been used in e.g. Kent *et al.* (1996), Rue and Hurn (1997) and Rue and Husby (1998).

The idea of describing objects such as potatoes, cells, hands or leaves as deformations of a template has been advocated by Ulf Grenander. His work on pattern theory has been collected in Grenander (1993), see also Grenander and Miller (1994).

In the above mentioned examples the template is a closed polygon, representing the outline of a typical object. Grenander and Manbeck (1993) use a discretized ellipse with fixed eccentricity as template in an application concerning defect detection in potatoes. In order to determine whether an object has the shape of a potato, the angles of the edges of the discretized object are compared with the angles of the edges of the template. In our model a similar comparison naturally occurs when we consider the derivative of the residual process.

In Rue and Syversveen (1998) an example of a stochastic geometry model of the type described in the present paper is considered. The template curve $\mathcal{C}$ is a circle with a radius r, and the residual process

$$\{X(t) : 0 \leq t \leq 2\pi r = T\}$$

is a cyclic and stationary Gaussian process with zero mean, variance $r^2\sigma^2$ and correlation function

$$\rho(h) = \begin{cases} e^{-\alpha h/T}\cos(4\pi h/T), & 0 \leq h \leq T/2, \\ \rho(T-h), & T/2 \leq h \leq T, \end{cases}$$

where $\alpha > 0$.

Kent *et al.* (1996) consider multivariate normal models for edges and vertices of a closed polygonal outline in the plane. The inverse covariance matrix of the edges is a circulant matrix with a first-order Markov property. They describe the limiting behaviour when the vertices become closely spaced. We extend this approach to the case where the inverse covariance matrix is a circulant matrix with a second-order Markov property. This turns out to be a better choice for our purpose. The general second-order model, described in the Appendix of the present paper, contains as a special case a second-order model suggested by Grenander (1993).

Kass *et al.* (1988) provide through the theory of snakes a way of performing boundary detection in an image. A snake is a curve in an image which minimizes a certain energy functional. In our set-up $\mathcal{F}$ is the snake while $\mathcal{C}$ can be regarded as a template snake. A very first choice of the energy functional could take the form

$$\mathrm{E}^{\star}_{snake} = \int_0^T (\alpha X(t)^2 + \beta X'(t)^2 + \mathrm{E}_{image}(t))dt, \quad \alpha, \beta > 0. \tag{1}$$

The first term represents an external constraint energy, which forces the snake to have a shape similar to the template snake. The next term represents an internal energy and makes the snake smooth. The last term connects the snake to the image. A simple image energy functional could be

$$\mathrm{E}_{image}(t) = \delta I(F(t)), \quad 0 \leq t \leq T,$$

where $I(x)$ is the intensity of the image at the position x. Depending on the sign of δ the snake is attracted to either black or white pixels. This approach is very similar to a Bayesian algorithm for object detection with the first-order Markov model described in Section 4.1 below as the prior model.

In Section 2 the data is described, and an ellipse is fitted to each of the objects. In Section 3 the stochastic geometry models are presented, and some geometrical characteristics of the residual process are explored under these models. In Section 4 we consider first- and second-order Markov models for the residual process, and they are fitted to the data in Section 5. Finally in Section 6 we discuss further properties of the models considered in Section 4 and consider topics for future work.

2 The data

The data set consists of 27 nuclear profiles from a malignant tumour and 27 nuclear profiles from a benign tumour of the human skin. The profiles were observed using light microscopy. The silhouettes were traced manually because light microscopic images are very complex with low contrast, containing many other components than the profiles of interest. The data have previously been analysed with respect to size and variability of size in Jensen and Srensen (1991) and Srensen (1991).

By visual inspection the nuclear profiles were smoothed and rescaled such that an area of approximately 75,000 pixels was obtained for each profile, cf. Figure 2. As may be seen from Figure 2, nuclear profiles from the malignant tumour appear to be less smooth than those from the benign tumour.

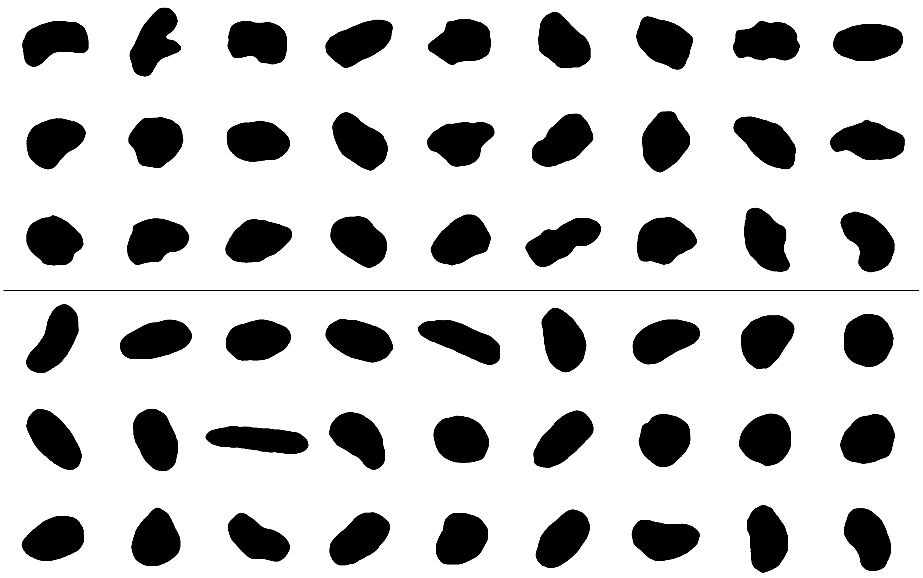

Figure 2: The nuclear profiles after scaling and smoothing. The upper panel is from the malignant tumour while the lower is from the benign tumour.

As a first analysis we fitted an ellipse to each of the profiles, using $n = 50$ (approximately) equidistant points $(x_1, \ldots, x_n) \in \mathbb{R}^{2n}$ on the boundary of the profile. Let $\mathcal{C} = \mathcal{C}(a, b, \theta, x_0, y_0)$ be an ellipse with semi-axes a, b, orientation θ and center (x_0, y_0), and denote by $c_i(a, b, \theta, x_0, y_0)$ the point on $\mathcal{C}$ closest to x_i, $i = 1, \ldots, n$, cf. Figure 3.

The fitted ellipse was then determined as that having parameter values (a, b, θ, x_0, y_0) that minimizes

$$\sum_{i=1}^{n} |x_i - c_i(a, b, \theta, x_0, y_0)|^2,$$

where $|\cdot|$ is Euclidean norm.

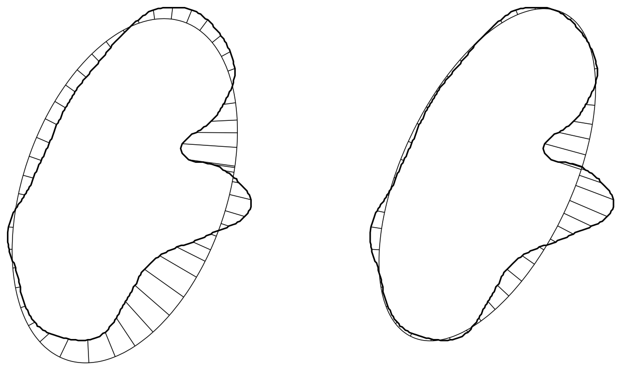

Figure 3: The fitted ellipse minimizes the sum of squares of the indicated distances. Left: An initial fit. Right: The fitted ellipse.

The ratio between the minor axis and the major axis of the fitted ellipses of the profiles does not show significant difference between the two samples, cf. Figure 4. Therefore, a study of the residual process is needed in order to describe the difference in shape of the two samples. Note also that the simpler description of the profiles by means of circles does not seem to be adequate.

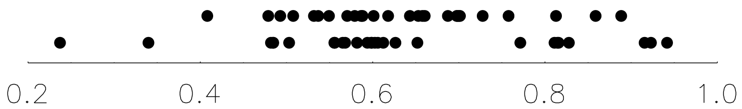

Figure 4: The ratio between the minor axis and the major axis of the fitted ellipses of the profiles. The upper points are from the profiles from the malignant tumour while the lower points are from the profiles from the benign tumour.

3 Stochastic geometry models for random curves

When modelling featureless objects using templates some of the frequently posed questions concern the choice, matching and number of landmarks. We propose a model where the observed curve $\mathcal{F}$ is matched to the template curve $\mathcal{C}$ by a perpendicular projection. A continuous type description makes it possible to study how the parameters of the finite-dimensional distributions depend on the number of landmarks. There are several ways of choosing the landmarks. In Model I below we parametrize $\mathcal{C}$ by arc length, which in the application corresponds to choosing the landmarks equidistantly on $\mathcal{C}$. In Model II we take the reverse approach and parametrize $\mathcal{F}$ by arc length.

3.1 Model I

The observed curve $\mathcal{F} = \{F(t) \in \mathbb{R}^2 : 0 \le t \le T\}$ is assumed to be a realization of a stochastic process

$$F(t) = c(t) + X(t)\omega(t), \quad 0 \le t \le T. \tag{2}$$

Here, $\mathcal{C} = \{c(t) \in \mathbb{R}^2 : 0 \le t \le T\}$ is a non-random closed ($c(0) = c(T)$) smooth curve in the plane parametrized by arc length. Furthermore, $\omega(t)$ is the inner unit normal vector to $\mathcal{C}$ at $c(t), 0 \le t \le T$, and

$$\{X(t) \in \mathbb{R} : 0 \le t \le T\}$$

is a real-valued cyclic and stationary stochastic process with zero mean. The process $\{X(t)\}$ models the deviations between the observed curve $\mathcal{F}$ and the expected curve $\mathcal{C}$. We will call $\{X(t)\}$ the residual process.

Note that the construction (2) puts some restrictions on how 'wild' the random curve $\mathcal{F}$ may look. Thus, each point $c(t) \in \mathcal{C}$ generates exactly one point on the random curve, positioned on the line $c(t) + \text{span}\{\omega(t)\}$. In particular, if $\mathcal{C}$ is a circle with center e and radius r and $X(t) \le r$, then $\mathcal{F}$ will be the boundary of a random set which is star-shaped relative to e.

The model is closed under translations and rotations in $\mathbb{R}^2$. Under such transformations, the curve $\mathcal{C}$ will be translated and rotated correspondingly, while $X(t)$ is unaffected. The model is also closed under scale transformations in $\mathbb{R}^2$

$$(x, y) \to \alpha(x, y) = (\alpha x, \alpha y), \quad \alpha > 0.$$

Thus, parametrizing the rescaled curve $\alpha\mathcal{C}$ by arc length and letting $c_\alpha(t) = \alpha c(t/\alpha)$, we have

$$\alpha\mathcal{C} = \{c_\alpha(t) : 0 \le t \le \alpha T\}.$$

Furthermore, if we likewise define F_α and X_α, but let $\omega_\alpha(t) = \omega(t/\alpha)$, the equation for the scale-transformed process becomes

$$F_\alpha(t) = c_\alpha(t) + X_\alpha(t) \cdot \omega_\alpha(t), \quad 0 \le t \le T_\alpha,$$

where $T_\alpha = \alpha T$.

The features of the model (2) which are invariant under changes in location, orientation and scale are thus the shape of $\mathcal{C}$ and the distribution of the process

$$X_1(t) = X(Tt)/T, \quad 0 \le t \le 1.$$

The latter process will be called the normalized residual process.

In order to analyze the model (2) we need to find, for selected values of $t \in [0, T]$, the point $F(t)$ on the line $c(t) + \mathrm{span}\{\omega(t)\}$. If the fluctuations of $F(t)$ around $c(t)$ are not too large, then $F(t)$ is expected to be the point on $c(t) + \mathrm{span}\{\omega(t)\}$ which is nearest to $c(t)$.

The process $\{X(t)\}$ and its derivatives (if they exist) contain interesting geometric information about the difference between the random curve $F(t)$ and its expectation $c(t)$.

Proposition 1 *Suppose that $X(t)$ is differentiable. Let $\mathcal{C}$ be orientated anti-clockwise. Furthermore, let $\Phi_F(t)$ and $\Phi_{\mathcal{C}}(t)$ be the angles that $F'(t)$ and $c'(t)$ make with a fixed axis, respectively. Then,*

$$X'(t) = \tan \Psi(t)(1 - \kappa_{\mathcal{C}}(t)X(t)), \tag{3}$$

where $\Psi(t) = \Phi_F(t) - \Phi_{\mathcal{C}}(t)$ and $\kappa_{\mathcal{C}}(t)$ is the curvature of $\mathcal{C}$ at $c(t)$.

Proof. Using that

$$\begin{aligned} F(\tilde{t}) &= F(t) + F'(t)(\tilde{t} - t) + o_F(\tilde{t} - t) \\ c(\tilde{t}) &= c(t) + c'(t)(\tilde{t} - t) + o_{\mathcal{C}}(\tilde{t} - t) \end{aligned}$$

we find that

$$\begin{aligned} 2X(t)X'(t) &= (X(t)^2)' \\ &= \lim_{\tilde{t} \to t} \frac{X(\tilde{t})^2 - X(t)^2}{\tilde{t} - t} \\ &= 2\langle F(t) - c(t), F'(t) - c'(t) \rangle, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is Euclidean inner procuct. Therefore,

$$\begin{aligned} X'(t) &= \langle \frac{F(t) - c(t)}{X(t)}, F'(t) - c'(t) \rangle \\ &\overset{(\star)}{=} \langle \frac{F(t) - c(t)}{X(t)}, F'(t) \rangle \\ &= \langle \frac{F(t) - c(t)}{X(t)}, \frac{F'(t)}{|F'(t)|} \rangle |F'(t)|, \end{aligned} \tag{4}$$

where at $(\star)$ we have used that $c'(t) \perp F(t) - c(t)$.

Since $\mathcal{C}$ is parametrized by arc length, $|c'(t)| = 1$ and

$$c'(t) = (\cos \Phi_{\mathcal{C}}(t), \sin \Phi_{\mathcal{C}}(t)).$$

It follows that

$$c''(t) = \Phi'_{\mathcal{C}}(t)(-\sin \Phi_{\mathcal{C}}(t), \cos \Phi_{\mathcal{C}}(t)) = \kappa_{\mathcal{C}}(t)\omega(t).$$

Using that

$$\frac{F'(t)}{|F'(t)|} = (\cos \Phi_F(t), \sin \Phi_F(t))$$

we therefore find

$$\langle \frac{F(t)-c(t)}{X(t)}, \frac{F'(t)}{|F'(t)|} \rangle = \sin(\Phi_F(t) - \Phi_{\mathcal{C}}(t)).$$

On the other hand,

$$\langle c'(t), \frac{F'(t)}{|F'(t)|} \rangle = \cos(\Phi_F(t) - \Phi_{\mathcal{C}}(t)) \tag{5}$$

and rewriting the left-hand side of (5)

$$\begin{aligned} \langle c'(t), \frac{F'(t)}{|F'(t)|} \rangle &= \frac{1}{|F'(t)|}(\langle c'(t), c'(t) \rangle + \langle c'(t), F'(t) - c'(t) \rangle) \\ &= \frac{1}{|F'(t)|}(1 - \langle c''(t), F(t) - c(t) \rangle) \\ &= \frac{1}{|F'(t)|}(1 - \kappa_{\mathcal{C}}(t) X(t)), \end{aligned}$$

(3) follows immediately. □

Note that if the variance of $X(t)$ is small then $\Phi_F(t) - \Phi_{\mathcal{C}}(t)$ is also expected to be small and

$$X'(t) \approx \Phi_F(t) - \Phi_{\mathcal{C}}(t)$$

can be approximated by the process of angular differences. These differences have been considered in a discrete set-up by Grenander and Manbeck (1993). If $X(t)$ is twice differentiable, then under the same assumption

$$X''(t) \approx \Phi'_F(t) - \Phi'_{\mathcal{C}}(t) = |F'(t)|\kappa_F(t) - \kappa_{\mathcal{C}}(t) \approx \kappa_F(t) - \kappa_{\mathcal{C}}(t).$$

3.2 Model II

As an altervative, we may take a reverse approach and parametrize $\mathcal{F}$ by arc length. The idea is then to construct $\mathcal{F}$ from a residual process $\{X(t) : 0 \leq t \leq T\}$ such that $X(t)$ is the signed distance from $F(t)$ to $\mathcal{C}$ and the parameter t represents arc length on $\mathcal{F}$. It is evidently also necessary in order to start the construction to specify a point $c \in \mathcal{C}$ such that $F(0)$ has signed distance $X(0)$ to c.

As in Model I, $\{X(t) : 0 \leq t \leq T\}$ is a real-valued cyclic and stationary stochastic process with zero mean. Under Model II it is therefore natural to sample points on $\mathcal{F}$ which are equally spaced in terms of arc length on $\mathcal{F}$. Such data points are, for instance, also used for the shape model described in Grenander and Manbeck (1993) and for the constant length or articulated model described in Kent *et al.* (1998).

Notice that it is not always possible to construct $\mathcal{F}$ in this way. As a simple example, suppose that $\mathcal{C}$ is a circle of radius r. Then, to avoid pathological cases we must assume $X(t) \leq r$. Furthermore, for small $\epsilon > 0$, we also need

$$|X(t+\epsilon) - X(t)| \leq \epsilon,$$

cf. Figure 5. In particular, if $X(t)$ is differentiable, then $|X'(t)| \leq 1$.

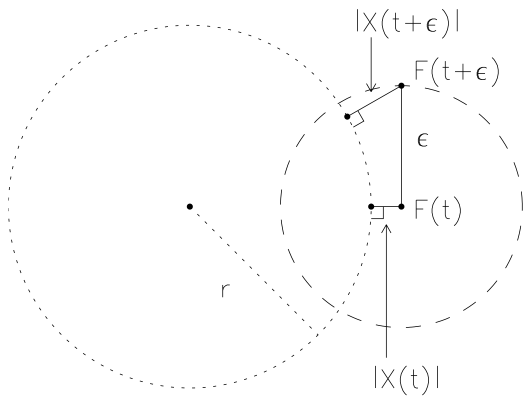

Figure 5: Illustration of the condition on $X(t)$.

Note that as under Model I, the 'shape features' of Model II are the shape of $\mathcal{C}$ and the distribution of the normalized process $\{X(Tt)/T : 0 \leq t \leq 1\}$.

Similarly to the case where $\mathcal{C}$ is parametrized by arc length, the process $X'(t)$ can be approximated by the process of angular differences.

Proposition 2 *Suppose that $X(t)$ is differentiable. Let $\mathcal{C}$ be orientated anti-clockwise. Furthermore, let $\Phi_F(t)$ and $\Phi_{\mathcal{C}}(t)$ be the angles that $F'(t)$ and $c'(t)$ make with a fixed axis, respectively. Then,*

$$X'(t) = \sin \Psi(t),$$

where $\Psi(t) = \Phi_F(t) - \Phi_{\mathcal{C}}(t)$.

Proof. Since $\mathcal{F}$ is parametrized by arc length, $|F'(t)| = 1$ and

$$F'(t) = (\cos \Phi_F(t), \sin \Phi_F(t)).$$

Using that

$$\frac{c'(t)}{|c'(t)|} = (\cos \Phi_{\mathcal{C}}(t), \sin \Phi_{\mathcal{C}}(t)) \perp \omega(t)$$

we see that

$$\omega(t) = (-\sin \Phi_{\mathcal{C}}(t), \cos \Phi_{\mathcal{C}}(t)).$$

The proposition now follows from (4). $\square$

4 Statistical inference

We will assume that the normalized residual process $\{X_1(t) : 0 \leq t \leq 1\}$ has a distribution belonging to a parametrized class of cyclic and stationary Gaussian processes with zero mean. In what follows we will omit the index 1 which should not cause any confusion.

There are as many choices of classes of Gaussian processes as there are parametrized classes of covariance functions. We will here concentrate on classes, having the property that the finite-dimensional (multivariate normal) distributions have a first- or second-order Markov property.

4.1 First-order Markov model

One such class which has been suggested, among others, by Grenander (1993, p. 476) and Kent *et al.* (1996) has the property that the finite-dimensional distributions have a first-order Markov property, approximately.

The class is characterized by the fact that $X(t)$ has zero mean, variance τ^2 and correlation function

$$\rho(h) = \rho(X(t), X(t+h)) = \frac{e^{(h-1/2)\phi} + e^{-(h-1/2)\phi}}{e^{\phi/2} + e^{-\phi/2}}, \quad 0 \leq h \leq 1, \quad \phi > 0. \qquad (6)$$

Note that the correlation $\rho(h)$ is always positive and it is a decreasing function on the interval $[0, 1/2]$. Note also that $\rho(h) = \rho(1-h)$ which is a general property of cyclic and stationary processes on $[0, 1]$.

In the proposition below, it is shown that the finite-dimensional distributions of $\{X(t)\}$ have a first-order Markov property, approximately. Recall that a multivariate normally distributed random vector

$$(X_0, \ldots, X_{n-1}) \sim N_n(0, \Sigma)$$

has a first-order Markov property if $(\Sigma^{-1})_{ij} = 0$ unless $j = (i+k) \bmod n$ with $|k| \leq 1$, cf. e.g. Lauritzen (1996).

For a $n \times n$ circulant matrix we use the notation $\operatorname{circ}(a_0, a_1, \ldots, a_{n-1})$ if its first row is $(a_0, a_1, \ldots, a_{n-1})$.

Proposition 3 *Let $\{X(t)\}$ be a cyclic and stationary Gaussian process with zero mean, variance τ^2 and correlation function (6). Furthermore, let $\{X^{(n)}(t)\}$ be the cyclic Gaussian process defined by*

$$X^{(n)} = (X^{(n)}(t_0), \ldots, X^{(n)}(t_{n-1})) \sim N_n(0, \Sigma_n),$$

$t_i = i/n, i = 0, \ldots, n-1$, with linear interpolation between t_i and t_{i+1}, and with $X^{(n)}(0) = X^{(n)}(1)$. Here, Σ_n is the regular $n \times n$ matrix with inverse

$$\Sigma_n^{-1} = \operatorname{circ}(\alpha/n + 2\beta n, -\beta n, 0, \ldots, 0, -\beta n), \quad \alpha, \beta > 0. \qquad (7)$$

Then, $X^{(n)}$ is first-order Markov and $\{X^{(n)}(t)\}$ converges weakly to $\{X(t)\}$. The 1-1 correspondence between (τ^2, ϕ) and (α, β) is

$$\phi^2 = \alpha/\beta \quad \tau^2 = \frac{e^{\phi/2} + e^{-\phi/2}}{2\phi\beta(e^{\phi/2} - e^{-\phi/2})}.$$

A proof of Proposition 3 may be found in Grenander (1993, p. 476-480). It follows immediately from the form of Σ_n^{-1} that $X^{(n)}$ is first-order Markov.

The parameters τ^2 and ϕ contain, under the model given in Proposition 3, the information about the deviation of the random curve shape from $\mathcal{C}$, cf. Section 3. The parameter τ^2 is a measure of the overall difference while ϕ regulates the smoothness of the random curve. The smaller the value of ϕ, the smoother the curve, cf. Figure 6.

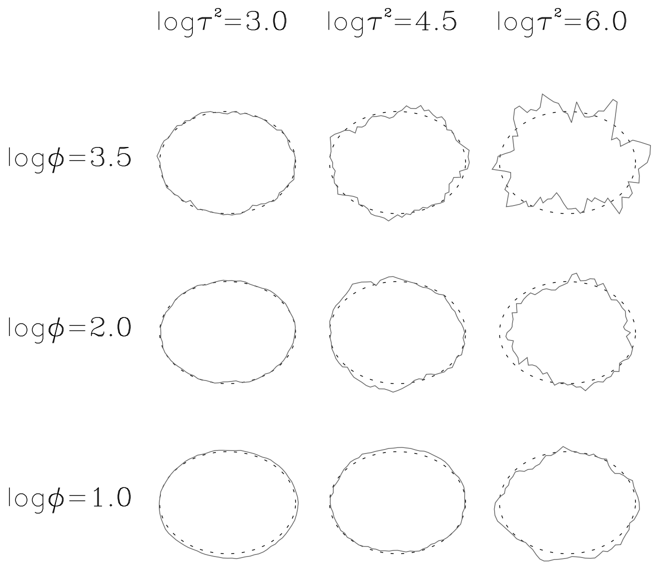

Figure 6: Simulations from the first-order Markov model for different values of ϕ and τ^2. We have simulated from Model I, cf. Section 3.1, with $\mathcal{C}$ taken to be an ellipse.

Estimation of the parameters (τ^2, ϕ) or equivalently (α, β) can be done by an approximate likelihood analysis. Let

$$X_n = (X(t_0), \ldots, X(t_{n-1})).$$

Then, because of Proposition 3, we can approximate the distribution of X_n with that of $X^{(n)}$ and use the likelihood based on $X^{(n)}$

$$L_n(\alpha, \beta) = (2\pi)^{-n/2}\det(\Sigma_n^{-1})^{1/2} \exp(-\frac{1}{2}X_n^*\Sigma_n^{-1}X_n),$$

where Σ_n^{-1} is given by (7). Denoting the eigenvalues of Σ_n^{-1} by $\lambda_i, i = 0, \ldots, n-1$, we get the log likelihood

$$l_n(\alpha, \beta) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=0}^{n-1}\log(\lambda_i) - \frac{1}{2}X_n^*\Sigma_n^{-1}X_n.$$

The actual form of the eigenvalues is given in the Appendix. The maximum of $l_n(\alpha, \beta)$ can be found by standard numerical methods. Usually, it will be a good idea to estimate α and β for a collection of values of n and as a model check investigate whether the estimation is stable. Given that the model is suitable it appears to be a good idea to base the estimation on an n as large as possible.

Note that

$$\begin{aligned} X_n^* \Sigma_n^{-1} X_n &= \alpha \frac{1}{n} \sum_{i=0}^{n-1} X(t_i)^2 + \beta \frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{X(t_i) - X(t_{i-1})}{1/n}\right)^2 \\ &= \alpha Y_n + \beta Z_n, \end{aligned}$$

say. The sufficient statistics (Y_n, Z_n) have a nice geometrical interpretation. Thus, Y_n is a discrete measure of the distance between the observed and the expected curve while Z_n compares local orientation,

$$Z_n = \frac{1}{n} \sum_{i=0}^{n-1} f(\Psi_i)^2,$$

where $\Psi_i = \Phi_{F_i} - \Phi_{c_i}$, and Φ_{F_i} and Φ_{c_i} are the angles that the line segments $F(t_i) - F(t_{i-1})$ and $c(t_i) - c(t_{i-1})$ make with a fixed axis, respectively. Furthermore, f=tan or sin, depending on whether $X(t)$ has been constructed using Model I or II of Section 3. Note also the connection between $X_n^* \Sigma_n^{-1} X_n$ and the sum of the external and internal energy from the theory of snakes, cf. (1).

4.2 Second-order Markov model

In Grenander (1993, p. 484) a model class with an approximate second-order Markov property is suggested. This class is characterized by the fact that $\{X(t)\}$ has mean zero, variance τ^2 and correlation function, cf. Figure 7,

$$\rho(h) = \frac{(\psi_2\psi_3 + \psi_1\psi_4)\psi_1(h)\psi_3(h) + (\psi_2\psi_3 - \psi_1\psi_4)\psi_2(h)\psi_4(h)}{\psi_1\psi_2 + \psi_3\psi_4}, \qquad \psi > 0, \quad 0 \le h \le 1, \tag{8}$$

where we use the notation

$$\psi_1(h) = \cos(\psi(h - 1/2)),\ \psi_2(h) = \sin(\psi(h - 1/2)),\ \psi_3(h) = \cosh(\psi(h - 1/2)),$$
$$\psi_4(h) = \sinh(\psi(h - 1/2)),\ \psi_i = \psi_i(1),\ i = 1, \ldots, 4.$$

The explicit form of the correlation function is not given in Grenander (1993), but can be derived from the spectral density of $X(t)$, as demonstrated in the Appendix. Note that the correlation function only depends on ψ.

In Grenander (1993, p. 484), it is mentioned that the finite-dimensional distributions of $X(t)$ have a second-order Markov property, approximately. Recall that a multivariate normally distributed random vector

$$(X_0, \ldots, X_{n-1}) \sim N_n(0, \Sigma)$$

has a second-order Markov property if $\Sigma_{ij}^{-1} = 0$ unless $j = (i + k) \bmod n$ with $|k| \le 2$. The result by Grenander is formulated in the proposition below.

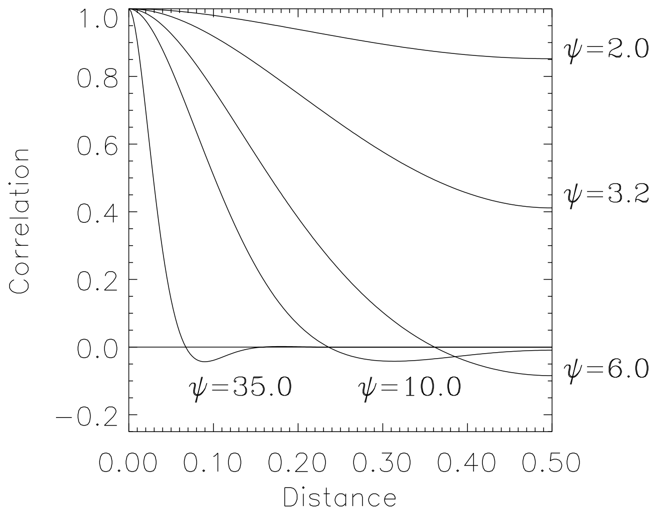

Figure 7: Correlation functions for the second-order Markov model for different values of ψ.

Proposition 4 *Let $\{X(t)\}$ be a cyclic and stationary Gaussian process with zero mean, variance τ^2 and correlation function (8). Furthermore, let $X^{(n)}(t)$ be constructed as in Proposition 3, but with*

$$\Sigma_n^{-1} = \mathrm{circ}\,(\alpha/n + 6\gamma n^3, -4\gamma n^3, \gamma n^3, 0, \ldots, 0, \gamma n^3, -4\gamma n^3), \ \alpha, \gamma > 0. \qquad (9)$$

Then, $X^{(n)} = (X^{(n)}(t_0), \ldots, X^{(n)}(t_{n-1})$ is second-order Markov and $\{X^{(n)}(t)\}$ converges weakly to $\{X(t)\}$. Here, the 1-1 correspondence between (τ^2, ψ) and (α, γ) is

$$4\psi^4 = \alpha/\gamma \quad \tau^2 = \frac{\psi}{2\alpha}\frac{\psi_1\psi_2 + \psi_3\psi_4}{(\psi_2\psi_3)^2 + (\psi_1\psi_4)^2}.$$

As in the first-order case the parameter τ^2 is a measure of the overall difference, while ψ regulates the smoothness of the curve, cf. Figure 8.

Similarly to the first-order case we can estimate the parameters by an approximate likelihood analysis. In this case

$$\begin{aligned} X_n^*\Sigma_n^{-1}X_n &= \alpha\frac{1}{n}\sum_{i=0}^{n-1} X(t_i)^2 + \gamma\frac{1}{n}\sum_{i=0}^{n-1}(\frac{X(t_i) - 2X(t_{i-1}) + X(t_{i-2})}{(1/n)^2})^2 \\ &= \alpha Y_n + \gamma V_n. \end{aligned}$$

V_n can be interpreted geometrically as a discrete measure of the change in local orientation

$$V_n = \frac{1}{n}\sum_{i=0}^{n-1}(\frac{f(\Psi_i) - f(\Psi_{i-1})}{(1/n)^2})^2,$$

where f and Ψ_i are defined as in the first-order case.

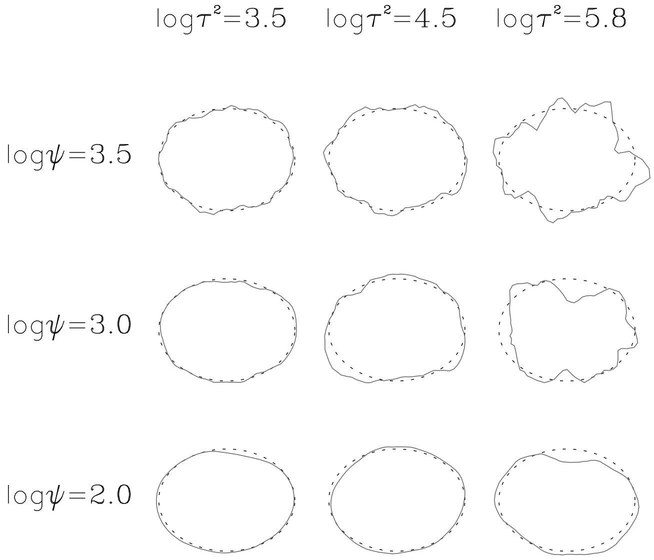

Figure 8: Simulations from the second-order Markov model for different values of ψ and τ^2. We have simulated from Model I, cf. Section 3.1, with $\mathcal{C}$ equal to an ellipse.

5 The application

For each nuclear profile we calculated the boundary length T and considered the $n = 50$ points

$$(F(Tt_0), \ldots, F(Tt_{n-1})), \quad t_i = i/n, \quad i = 0, \ldots, n-1,$$

collected (approximately) equidistantly on the boundary of the nuclear profile. For $\mathcal{C}$ we chose the ellipse from Section 2. This is the set-up in Model II, cf. Section 3.2. The reason for preferring Model II to Model I was that Model II was expected to be a more sensitive tool for distinguishing between the profiles from the malignant and the benign tumour since under Model I curve segments from $\mathcal{F}$ which are close to being perpendicular to $\mathcal{C}$ are undersampled.

Now consider the corresponding n observations from the normalized residual process

$$X_n = (X(Tt_0), \ldots, X(Tt_{n-1}))/T, \quad t_i = i/n, \quad i = 0, \ldots, n-1.$$

Our initial model was a general second-order Markov model

$$M_0 : X_n \sim N_n(0, \Sigma_n),$$

where

$$\Sigma_n^{-1} = \operatorname{circ}(\alpha/n + 2\beta n + 6\gamma n^3, -\beta n - 4\gamma n^3, \gamma n^3, 0, \ldots, 0, \gamma n^3, -\beta n - 4\gamma n^3), \qquad \alpha > 0, \quad \beta, \gamma \geq 0,$$

see also the Appendix. Note that this model contains as special cases both the model described by (7) and that described by (9). For all the profiles, we estimated the covariance matrix, using the ellipses fitted in Section 2. For some of the profiles, the ellipse and the covariance were also estimated simultaneously. The ellipses fitted in this way did only differ slightly from those determined in Section 2.

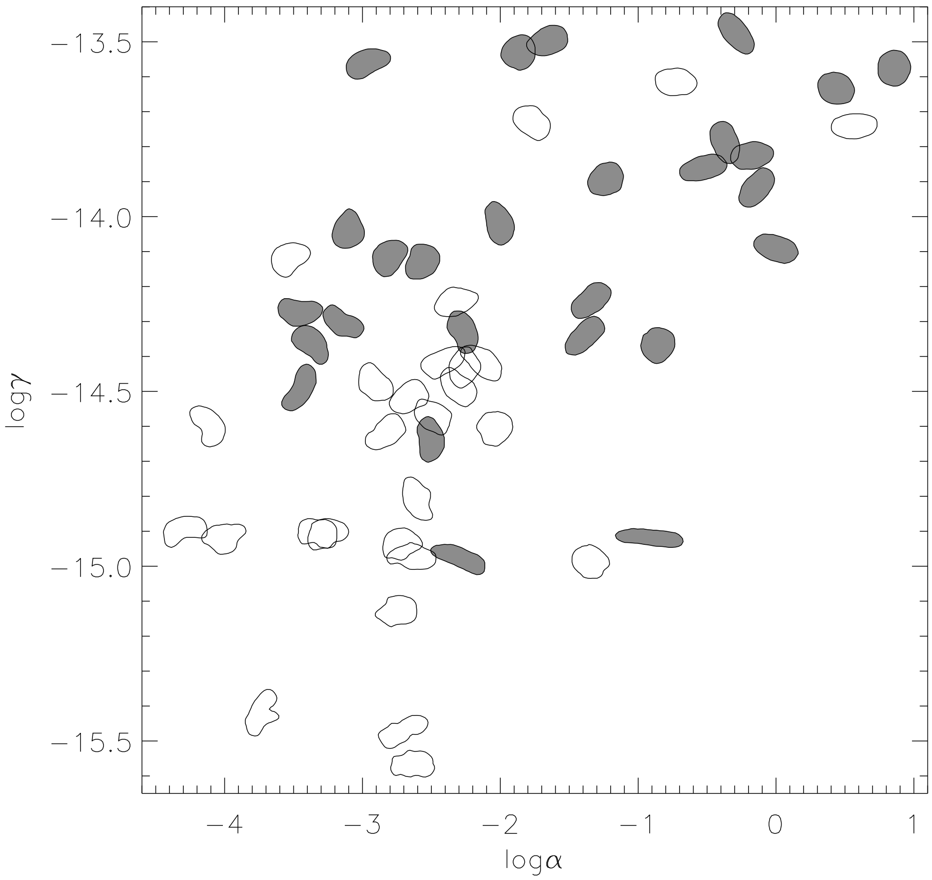

Figure 9: The estimates of (α, γ) under the two parameter second-order Markov model. The white nuclei are from the malignant tumour while the black nuclei are from the benign tumour.

To see whether the first- or second-order Markov model from Section 4 could be used we tested the two hypotheses

$$H_1 : \gamma = 0 \qquad H_2 : \beta = 0.$$

We found that the second-order Markov model with $\beta = 0$ described in Section 4.2 was the most appropriate. Since the nuclear profiles look like the ones simulated in Figure 8 this was of course not surprising. Figure 9 shows the estimates of (α, γ) under the two parameter second-order Markov model. Note that it is indeed possible to distinguish between the two samples.

To verify that the estimates do not depend on the number n of landmarks, we fitted the model for landmark numbers between 30 and 90. From Figure 10 it is clear, that the estimation is stable.

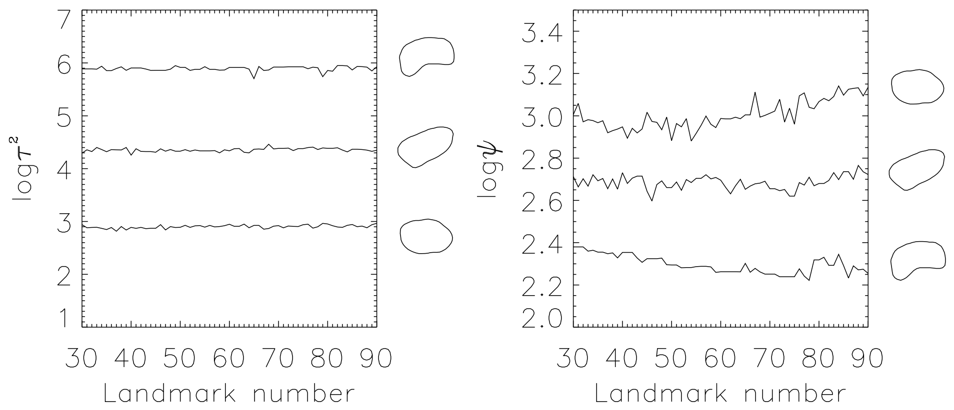

Figure 10: The estimates of $\log \tau^2$ and $\log \psi$ for different numbers of landmarks.

6 Discussion

The smoothness of the sample paths of the residual process is an important topic when constructing realistic models. It follows from (15) in the Appendix below that the first- and second-order processes considered in Sections 4.1 and 4.2 have a spectral density of the form

$$(\alpha + \beta(2\pi s)^2)^{-1}, \qquad s = 0, \pm 1, \ldots \tag{10}$$
$$(\alpha + \gamma(2\pi s)^4)^{-1}, \qquad s = 0, \pm 1, \ldots, \tag{11}$$

respectively, cf. Cramer and Leadbetter (1967, p. 126-128). Since (10) decreases as $1/s^2$ and (11) decreases as $1/s^4$ the sample paths of the first type of process are continuous, while the sample paths of the second type are continuously differentiable, cf. Cramer and Leadbetter (1967, p. 181). For each nuclear profile it seems natural to assume that the observed curve $\mathcal{F}$ is differentiable, and since the expected curve $\mathcal{C}$ is infinitely often differentiable the model from Section 4.2 appears to be the most appropriate in our application, also from the point of view of smoothness.

The simulations presented in Figures 6 and 8 involve the generation of an n-dimensional normally distributed random vector with mean zero, constant variance τ^2 of its coordinates and correlation matrix determined by (7) and (9). It is however also posssible to simulate from the continuous process, cf. Hobolth and Pedersen (1999). Furthermore, it is part of our future research plans to develop estimation procedures based on observations of the continuous process.

The model suggested in the present paper seems to provide the right framework for solving our original problem, viz. quantifying in a simple manner the shape differences between the two samples of nuclear profiles. It should however be noticed that our model appears to be difficult to use as a prior model in a Bayesian object recognition problem with an unknown number of objects, cf. Rue and Hurn (1997, p. 10). See also the related papers on Bayesian object recognition, Rue and Husby (1998) and Rue and Syversveen (1998).

Kent *et al.* (1998) describe a deformable template model with a circle as template. An object is described by edge deformations of the template. The model, which is used for exploring shape variability, is only valid for samples of objects which can be regarded as deformations of a circle. The sample of objects from the benign tumour as well as from the malignant tumour cannot be regarded as deformations from a circle.

The objects in Kent *et al.* (1998) and in the present paper are assumed to be observed without measurement error. In a Bayesian approach, Hurn *et al.* (1999) consider parameter estimation in deformable template models for objects observed with measurement error. The object prior is described by edge deformations of the template. The distribution of the edges are assumed to have a first-order Markov property similar to the one described in the present paper.

Our model for the normalized residual process $\{X_1(t) : 0 \leq t \leq 1\}$ may be generalized in various ways. Instead of having zero mean a mean value of the residual process given by either

$$\mu_1(t) = \alpha_1 \cos(6\pi t + \omega_1), \quad 0 \leq t \leq 1,$$

or

$$\mu_2(t) = \alpha_2 \cos(8\pi t + \omega_2), \quad 0 \leq t \leq 1,$$

or a sum of $\mu_1(t)$ and $\mu_2(t)$ would allow for a systematic variation from the shape of an ellipse. Still modelling the correlation structure by the second-order Markov property it might be even more easy to distinguish between the two samples. Another generalization would be to let the residual variance depend on the curvature of the template curve.

Acknowledgement

The authors thank Hans Jrgen Gundersen, Flemming Srensen and colleagues from Department of Mathematical Sciences, University of Aarhus, for fruitful discussions. We also want to thank the referee for helpful comments and suggestions. This work was supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation.

Appendix

In this appendix we will derive the explicit form (8) of the correlation function for the second-order model. We will use an approach such that we at the same time derive $\rho(h)$ for the first-order model, cf. (6).

Thus, consider the matrix

$$\Sigma_n^{-1} = \mathrm{circ}\,(\alpha/n + 2\beta n + 6\gamma n^3, -\beta n - 4\gamma n^3, \gamma n^3, 0, \ldots, 0, \gamma n^3, -\beta n - 4\gamma n^3),$$
$$\alpha > 0,\ \beta, \gamma \geq 0.$$

For $v = (v_0, \ldots, v_{n-1}) \in \mathbb{R}^n$ we have that

$$v\Sigma_n^{-1}v^* = \alpha/n \sum_{i=0}^{n-1} v_i^2 + \beta n \sum_{i=0}^{n-1}(v_i - v_{i-1})^2 + \gamma n^3 \sum_{i=0}^{n-1}(v_i - 2v_{i-1} + v_{i-2})^2,$$

where $v_{-1} = v_{n-1}$ and $v_{-2} = v_{n-2}$. Therefore, Σ_n^{-1} is positive definite. Note that $\gamma = 0$ is the case treated in Proposition 3 and $\beta = 0$ is the case treated in Proposition 4.

Since Σ_n^{-1} is circular we have that Σ_n is also circular. The eigenvalues of Σ_n^{-1} are given by, cf. Anderson (1958, p. 280, 282),

$$\lambda_s = (\alpha/n + 2\beta n + 6\gamma n^3) - (\beta n + 4\gamma n^3)2\cos(2\pi s/n) + \gamma n^3 2\cos(4\pi s/n),$$

with corresponding orthonormal (complex) eigenvectors u_s, $s = 0, \ldots, n-1$, where

$$(u_s)_l = e^{2\pi lsi/n}/\sqrt{n}, \qquad l = 0, \ldots, n-1.$$

Note that $u_0 = (1, \ldots, 1)/\sqrt{n}$. Let

$$\Lambda = \operatorname{diag}(\lambda_0, \ldots, \lambda_{n-1}) \text{ and } U = \begin{pmatrix} u_0 \\ \vdots \\ u_{n-1} \end{pmatrix}.$$

Then, we have for $l = 0, 1, \ldots, n-1$ that

$$(\Sigma_n)_{l,0} = (U\Lambda^{-1}\overline{U}^*)_{l,0} = \sum_{s=0}^{n-1} \frac{1}{\sqrt{n}}(U\Lambda^{-1})_{l,s} = \sum_{s=0}^{n-1} \frac{e^{2\pi lsi/n}}{n\lambda_s} = \sum_{s=0}^{n-1} \frac{\cos(2\pi ls/n)}{n\lambda_s}. \tag{12}$$

Now, we will study a sequence $X^{(n)}(t)$ of cyclic Gaussian processes, constructed as in Propositions 3 and 4. Thus, let

$$(X_{n0}, \ldots, X_{n(n-1)}) \sim N_n(0, \Sigma_n)$$

and define a continuous process $\{X^{(n)}(t) : 0 \leq t \leq 1\}$ by

$$X^{(n)}(j/n) = X_{nj} \quad j = 0, \ldots, n-1,$$

with linear interpolation in between, and with $X^{(n)}(0) = X^{(n)}(1)$. Using (12), we find the following expression for the covariance between $X^{(n)}(0)$ and $X^{(n)}([nh]/n), 0 \leq h < 1$,

$$\operatorname{Cov}(X^{(n)}(0), X^{(n)}([nh]/n)) = \operatorname{Cov}(X_{n0}, X_{n([nh])}) = \sum_{s=0}^{n-1} \frac{\cos(2\pi[nh]s/n)}{n\lambda_s}.$$

Assuming that at least one of the parameters β and γ is positive, we obtain, using dominated convergence,

$$\lim_{n\to\infty} \operatorname{Cov}(X^{(n)}(0), X^{(n)}([nh]/n)) = \frac{1}{\alpha} + 2\sum_{s=1}^{\infty} \frac{\cos(2\pi hs)}{\alpha + \beta(2\pi s)^2 + \gamma(2\pi s)^4}. \tag{13}$$

By construction $X^{(n)}(t), 0 \leq t < 1$, is given by

$$X^{(n)}(t) = X_{n([nt])} + n(t - [nt]/n)(X_{n([nt]+1)} - X_{n([nt])}).$$

Hence we have

$$\begin{aligned} |X^{(n)}(h) - X_{n([nh])}| &= n(h - [nh]/n)|X_{n([nh]+1)} - X_{n([nh])}| \\ &\leq |X_{n([nh]+1)} - X_{n([nh])}| \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}((X_{n([nh]+1)} - X_{n([nh])})^2) &= \mathrm{E}((X_{n1} - X_{n0})^2) \\ &= 2(\mathrm{Var} X_{n0} - \mathrm{Cov}(X_{n1}, X_{n0})) \\ &= 2(\sum_{s=0}^{n-1} \frac{1}{n\lambda_s} - \sum_{s=0}^{n-1} \frac{\cos(2\pi s/n)}{n\lambda_s}) \\ &\to 0 \text{ for } n \to \infty. \end{aligned} \tag{14}$$

From (13) and (14) it is clear that the limiting covariance function is given by the Fourier series

$$\sigma(h) = \frac{1}{\alpha} + 2\sum_{s=1}^{\infty} \frac{\cos(2\pi hs)}{\alpha + \beta(2\pi s)^2 + \gamma(2\pi s)^4}, \quad 0 \leq h \leq 1. \tag{15}$$

In order to find an explicit expression of $\sigma(h)$ note that in the sense of generalized functions and the theory of distributions

$$\alpha\sigma(h) - \beta\sigma''(h) + \gamma\sigma''''(h) = 1 + 2\sum_{s=1}^{\infty} \cos(2\pi hs) = \sum_{s=-\infty}^{\infty} \cos(2\pi hs) = \sum_{s=-\infty}^{\infty} \delta_s,$$

where δ_s denotes the δ-function at s. Therefore, $\sigma(h)$ is the solution of a homogeneous linear differential equation with constant coefficients, cf. Hirsch and Smale (1974, p. 138). In general, $\sigma(h)$ is thus a linear combination of exponential and trigonometric functions depending on the roots of the characteristic polynomial

$$\alpha - \beta y^2 + \gamma y^4 = 0.$$

We will for simplicity only consider the two cases previously mentioned, $\beta > 0, \gamma = 0$ and $\beta = 0, \gamma > 0$.

Before treating these two cases, it is convenient to change the interval from $[0, 1]$ to $[-1/2, 1/2]$. Thus, let

$$\sigma_*(h) = \sigma(h + 1/2), \qquad -1/2 \leq h \leq 1/2.$$

The function $\sigma_*(h)$ is the solution to the same homogeneous linear differential equation,

$$\alpha\sigma_*(h) - \beta\sigma_*''(h) + \gamma\sigma_*''''(h) = 0. \tag{16}$$

First consider the case $\beta > 0, \gamma = 0$. With $\phi^2 = \alpha/\beta$ the solution of (16) is given by, cf. Hirsch and Smale (1974, p. 139),

$$\sigma_*(h) = c_1 e^{\phi h} + c_2 e^{-\phi h}, \qquad -1/2 \leq h \leq 1/2.$$

From $\sigma(h) = \sigma(1-h), 0 \leq h \leq 1$, it follows that

$$\sigma_*(h) = \sigma_*(-h), \qquad -1/2 \leq h \leq 1/2,$$

and hence we have $c_1 = c_2 = c$. Using that the first Fourier coefficient is given by $1/\alpha$ we can find c from the equation

$$1/\alpha = 2\int_0^{1/2} \sigma_*(h)dh, \tag{17}$$

and we obtain the form (6) of $\rho(h)$, stated in the main text, and the correspondence between (τ^2, ϕ) and (α, γ) given in Proposition 3.

Now consider the case $\beta = 0, \gamma > 0$. Again, using $\sigma_*(h) = \sigma_*(-h)$ and letting $4\psi^4 = \alpha/\gamma$, the solution of (16) is given by, cf. Hirsch and Smale (1974, p. 139)

$$\sigma_*(h) = c_1(e^{\psi h} + e^{-\psi h})\cos(\psi h) + c_2(e^{\psi h} - e^{-\psi h})\sin(\psi h), \; -1/2 \leq h \leq 1/2.$$

The constants c_1 and c_2 can be found from (17) and $\sigma_*'(1/2) = 0$, and we obtain the explicit form (11) of $\rho(h)$ and the correspondence between (τ^2, ψ) and (α, γ) given in Proposition 4.

References

Anderson, T.W. (1958). *The Statistical Analysis of Time Series.* John Wiley and Sons, New York.

Cramer, H and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes.* Wiley, New York.

Grenander, U. (1993). *General Pattern Theory.* Oxford University Press, Oxford.

Grenander, U. and Manbeck, K.M. (1993). A stochastic model for defect detection in potatoes. *J. Comp. Graph. Statist.* **2**, 131-151.

Grenander, U. and Miller, M.I. (1994). Representations of knowledge in complex systems (with discussion). *J. R. Statist. Soc.* B **56**, 549-603.

Hirsch, M.W. and Smale, S. (1974). *Differential Equations, Dynamical Systems and Linear Algebra.* Academic Press, San Diego.

Hobolth, A. and Pedersen, J. (1999). Estimation in a class of circular Gaussian processes. *Research Report* **8**, Laboratory for Computational Stochastics, Institute of Mathematical Sciences, University of Aarhus, Denmark. In preparation.

Hurn, M.A., Steinsland, I. and Rue, H. (1999). Parameter estimation for a deformable template model. *Statistics Research Report* **99.02**, University of Bath, UK.

Jensen, E.B.V. (1998). *Local Stereology.* World Scientific, Singapore.

Jensen, E.B. and Srensen, F.B. (1991). A note on stereological estimation of the volume-weighted second moment of particle volume. *J.Microsc.* **164**, 21-27.

Kass, M., Witkin, A., Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision* **1**, 321-331.

Kent, J.T., Dryden, I.L. and Anderson, C.R. (1998). Using circulant symmetry to model featureless objects. *Technical Report* STAT**98/14**, Department of Statistics, University of Leeds, UK.

Kent, J.T., Mardia, K.V. and Walder, A.N. (1996). Conditional cyclic Markov random fields. *Adv. Appl. Probab. (SGSA)* **28**, 1-12.

Lauritzen, S.L. (1996). *Graphical Models.* Oxford University Press, Oxford.

Rue, H. and Hurn, M.A. (1997). Bayesian object identification. *Statistics* **6**, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Rue, H. and Husby, O.K. (1998). Identification of partly destroyed objects using deformable templates. *Statistics and Computing* **8**(3), 221-228.

Rue, H. and Syversveen, A.R. (1998). Bayesian object recognition with Baddeley's Delta Loss. *Adv. Appl. Prob. (SGSA)* **30**, 64-84.

Srensen, F.B. (1991). Stereological estimation of mean and variance of nuclear volume from vertical sections. *J. Microsc.* **162**, 203-229.

B

Hobolth, A., Kent, J.T. and Dryden, I.L. (2002).
On the relation between edge and vertex modelling in shape analysis.
To appear in *Scand. J. Statist.*

B

On the relation between edge and vertex modelling in shape analysis

ASGER HOBOLTH
University of Aarhus
JOHN T. KENT
University of Leeds
IAN L. DRYDEN
University of Nottingham

ABSTRACT. Objects in the plane with no obvious landmarks can be described by either vertex transformation vectors or edge transformation vectors. In this paper we provide the relation between the two transformation vectors. Grenander and Miller (1994) use a multivariate normal distribution with a block circulant covariance matrix to model the edge transformation vector. This type of model is also feasible for the vertex transformation vector and in certain cases the free parameters of the two models match up in a simple way. A vertex model and an edge model are applied to a data set of sand particles to explore shape variability.

Key words: Circulant symmetry, complex symmetry, deformation, edge transformation, featureless objects, outline, shape, vertex transformation.

1 Introduction

Consider a solid object in the plane with no obvious landmarks. Suppose the object can be regarded as a stochastic deformation of a circle. It is often fruitful to describe the object by deforming an n-sided regular polygon using either vertex transformations or edge transformations. In order to describe the variability in shape, it is useful to standardize the edge transformation vector, cf. Kent *et al*. (2000). A standardized edge transformation vector is invariant under a translation, rotation and isotropic scaling of the object. In this paper we standardize the vertex transformation vectors similarly, and we establish a linear relation between the transformation vectors. An immediate advantage of this unification is that it allows for a comparison of strengths and weaknesses of models independently of whether they are defined as vertex or edge transformation models.

Grenander and Miller (1994) use a multivariate normal distribution to model the edge transformation vector, which is also a feasible model for the vertex transformation vector. The lack of features on the object implies that any statistical model should be invariant under a cyclic permutation of the vertices and this invariance implies that the mean should be zero and that the covariance matrix should be block circulant. To limit the number of free parameters one can propose symmetries or restrictions to the model. We demonstrate how easy it is to go back and forth between the free parameters in the edge and vertex models in some of these cases.

By spacing the vertices in certain ways the data analyst can reduce the dimension of the data from $2n$ to n. This choice of spacing also enables the data analyst to focus on real shape differences rather than differences of the object due to vertex spacing. A vertex and an edge representation of such spacings are presented and in order to compare the two representations we apply them to a data set of sand particles.

Many examples of shape modelling of objects without landmarks can be found in the literature. The objects under study in Grenander and Manbeck (1993), Grenander and Miller (1994), Stoyan and Stoyan (1994, pp167), Mardia *et al.* (1996), Rue and Syversveen (1998), Hansen *et al.* (2000) and Hobolth and Jensen (2000) are potatoes, mitochondria, sand grains, mushrooms, cells, arteries and cell nuclei, respectively, and they all fall in this category. This list also suggests the wide range of applications of the type of models considered in this paper.

In Section 2 the vertex and edge transformation vectors are defined and two simple representations are described. In Section 3 the relationship between the transformation vectors is established. In Section 4 we describe the model proposed by Grenander and Miller, and in Section 5 we consider some special cases of the model where it is easy to go back and forth between the free parameters of a vertex and an edge representation. In Section 6 we apply and compare the two simple representations on a data set of sea sand particles collected from a beach at the Baltic Sea.

2 Transformation vectors

In Figure 1 we show a collection of objects with no obvious landmarks. Furthermore, the objects can be regarded as deformations of a circle. The sample consists of 24 sand grains collected from a beach at the Baltic Sea, as given by Stoyan (1997). In this section we demonstrate how standardized transformation vectors can be used to describe the shape variability of this type of object.

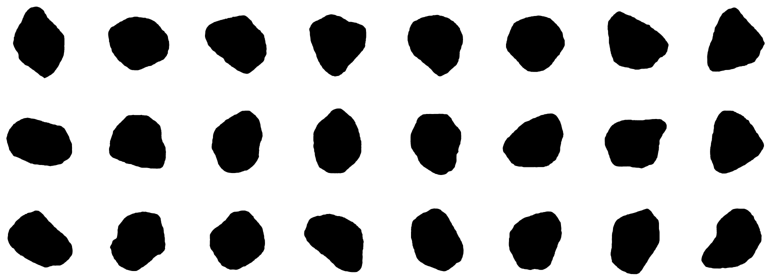

Figure 1: A sample of 24 sand grains from the Baltic Sea. The sand grains are scaled so that they all have approximately the same area.

The transformation vectors are defined from a number of vertices collected on the boundary of the object. For arbitrarily chosen vertices, the dimension of the vectors is $2n$, but under certain regular arrangements of the vertices the data analyst can reduce the dimension to n. Examples of such spacings include the radial representation and the constant length representation, which are described in Section 2.1 and Section 2.2.

$\boxed{\mathbf{B}}$

2.1 Vertex transformation vectors

Consider an object P in the complex plane with $n \geq 3$ vertices $\tilde{v}_j, j = 0, \ldots, n-1$, located round the outline of the object in an anti-clockwise order. First, make the vertices independent of the location by translating to the new centred vertices

$$v_j = \tilde{v}_j - \frac{1}{n}\sum_{k=0}^{n-1} \tilde{v}_k, \qquad j = 0, \ldots, n-1. \tag{2.1}$$

The sample mean of the vertices is now located at the origin. Let P^0 denote the regular n-sided polygon with vertices

$$v_j^0 = e^{2\pi i j/n}, \qquad j = 0, \ldots, n-1. \tag{2.2}$$

The basic idea is to describe the object P as a stochastic deformation of the polygon P^0. One way of doing this is through a complex-valued "vertex transformation vector" $\mathbf{d} = (d_j)$ defined by

$$d_j = v_j / v_j^0, \qquad j = 0, \ldots, n-1. \tag{2.3}$$

We want to explore the variability in shape and thus we have to investigate the effect of a scaling and rotation of the vertices. Scaling and rotating the vertices by $\alpha \in \mathbb{C}$ leads to

$$\alpha v_j = \alpha d_j v_j^0 = (1 + u_j) v_j^0, \quad j = 0, \ldots, n-1,$$

where

$$u_j = \alpha d_j - 1, \quad j = 0, \ldots, n-1.$$

One possible standardization for α is to rescale and rotate so that

$$u_{av} = \frac{1}{n}\sum_{j=0}^{n-1} u_j = 0, \tag{2.4}$$

which implies

$$1/\alpha = \frac{1}{n}\sum_{j=0}^{n-1} d_j = d_{av}.$$

Thus, replacing $\mathbf{d}$ by the "standardized vertex transformation vector" $\mathbf{u} = (u_j)$ with components

$$u_j = d_j / d_{av} - 1, \qquad j = 0, \ldots, n-1, \tag{2.5}$$

gives invariance under changes in scale and rotation. The representation in terms of $\mathbf{u}$ breaks down if $d_{av} = 0$, but since $d_{av} = 0$ only for extreme deformations from P^0 this does not pose a practical problem.

The translation (2.1) implies that

$$\sum_{j=0}^{n-1} v_j = 0,$$

which can be recast in terms of $\mathbf{u}$ as

$$0 = \sum_{j=0}^{n-1} v_j/d_{av} = \sum_{j=0}^{n-1}(1+u_j)v_j^0 = \sum_{j=0}^{n-1} u_j v_j^0 = \sum_{j=0}^{n-1} u_j e^{2\pi ij/n}, \qquad (2.6)$$

where in the third equality we have used

$$\sum_{j=0}^{n-1} v_j^0 = \sum_{j=0}^{n-1} e^{2\pi ij/n} = 0.$$

The complex standardization constraint (2.4) and the complex translation constraint (2.6) imply that $\mathbf{u}$ has $2n-4$ free (real) parameters. Note that from

$$v_j = d_{av}(1+u_j)v_j^0, \qquad j = 0,\ldots,n-1,$$

it follows that $\mathrm{Re}(u_j) = r_j$ determines the component of the jth vertex of P *tangential* to $d_{av}v_j^0$ while $\mathrm{Im}(u_j) = s_j$ determines the *normal* component.

One often used method of describing a boundary is the *radial representation* as described below. Let $c \in \mathbb{C}$ be the centre of gravity of the solid two-dimensional object. Find the vertices of the object by rays starting at c with angles $2\pi j/n, j = 0,\ldots,n-1$, relative to some fixed axis. Suppose the object is star shaped relative to the centre of gravity, i.e. the radial vectors from a single central point to the boundary all remain inside the object. Then the sequence of distances $d_j, j = 0,\ldots,n-1$, provides a description of the boundary. Consider the polygon given by the vertices

$$v_j = d_j v_j^0, \quad d_j \in \mathbb{R}, \quad j = 0,\ldots,n-1.$$

The standardized vertex transformation vector of this polygon is given by

$$u_j = d_j/d_{av} - 1 \in \mathbb{R}, \quad j = 0,\ldots,n-1. \qquad (2.7)$$

Hence, there is variability only in the tangent component $\mathbf{r} = (r_j)$ of the vertex transformation vector, cf. Figure 2. Note that in this construction of the vertex transformation vector we avoid centering the vertices as in (2.1), but take (2.3) as our starting point. We recommend living with this minor violation of the assumptions since the average of the vertices defined from the centre of gravity is usually close to zero. If the average of the vertices defined from the centre of gravity is not close to zero, then we suggest iterating the construction until a centre is found with the property that the average of the corresponding vertices is close to zero.

There is a large literature on vertex transformation models and usually the radial representation is used. The prior model in Rue and Syversveen (1998) is one example.

B

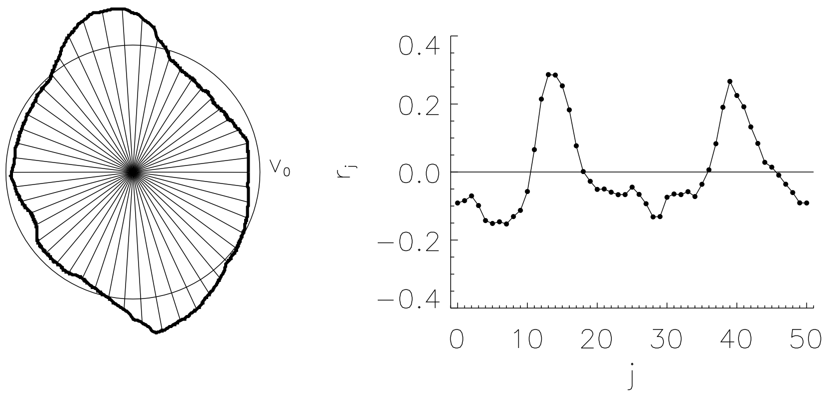

Figure 2: Left: The rays from the center of gravity determine the vertices. Right: The value of the tangent component $\mathbf{r}$ of the standardized vertex transformation vector.

In a procedure for identifying cells in a digital image, the regular n-sided polygon P^0 is deformed and the generated object is determined by (2.3), where the vertex transformation vector $\mathbf{d} = (d_j) \in \mathbb{R}^n$ is real and follows a multivariate normal distribution, which is invariant under cyclic permutation. Note that when the components of $\mathbf{d}$ are real and positive then the generated objects are star shaped relative to the origin. Some other examples are given by Rohlf and Archie (1984), Johnson *et al.* (1985), Mardia and Qian (1995) and Mardia *et al.* (1996) who use such representations in modelling mosquito wings, mouse vertebrae, leaves and mushrooms. A description of Fourier series analysis of radial vector functions is given by Stoyan and Stoyan (1994, pp80), applied to sand grains. Fourier series analysis has close connections to the analysis of radial vectors, as seen in Section 5.4.

2.2 Edge transformation vectors

We now change the focus from the vertices to the edges and recall the standardized edge transformation vector as introduced in Kent *et al.* (2000). Focusing on the edges removes any translation effect immediately. The vector of edges $\mathbf{e} = (e_j)$ for the object P has components

$$e_j \quad = \quad v_{j+1} - v_j, \qquad j = 0, \ldots, n-1, \tag{2.8}$$

whereas the vector of edges for the regular polygon P^0 are given by

$$e_j^0 = v_{j+1}^0 - v_j^0 = (e^{2\pi i/n} - 1)e^{2\pi ij/n}, \qquad j = 0, \ldots, n-1.$$

All subscripts throughout the paper are interpreted modulo n. The complex-valued "edge transformation vector" $\mathbf{t} = (t_j)$ can be defined by

$$t_j = e_j/e_j^0, \qquad j = 0, \ldots, n-1. \tag{2.9}$$

Dividing $\mathbf{t}$ by $t_{av} = \frac{1}{n}\sum_{j=0}^{n-1} t_j$ gives invariance under changes in scale and rotation and leads to the "standardized edge transformation vector" $\mathbf{z} = (z_j)$ defined by

$$z_j = t_j/t_{av} - 1, \qquad j = 0, \ldots, n-1.$$

There are two complex constraints on $\mathbf{z}$. As in (2.4) the standardization implies that

$$\frac{1}{n}\sum_{j=0}^{n-1} z_j = z_{av} = 0. \tag{2.10}$$

Secondly, to ensure that the outline P is closed we also require $\mathbf{z}$ to satisfy

$$0 = \sum_{j=0}^{n-1} e_j/t_{av} = \sum_{j=0}^{n-1}(1+z_j)e_j^0 = \sum_{j=0}^{n-1} z_j e_j^0 = \sum_{j=0}^{n-1} z_j e^{2\pi i j/n}, \tag{2.11}$$

and therefore $\mathbf{z}$ has $2n-4$ free parameters. Note that the constraints on $\mathbf{z}$ and $\mathbf{u}$ are the same. As in the previous subsection it follows from

$$e_j = t_{av}(1+z_j)e_j^0, \qquad j = 0, \ldots, n-1,$$

that $\mathrm{Re}(z_j) = x_j$ determines the component of the jth edge of P *tangential* to $t_{av}e_j^0$ while $\mathrm{Im}(z_j) = y_j$ determines the *normal* component.

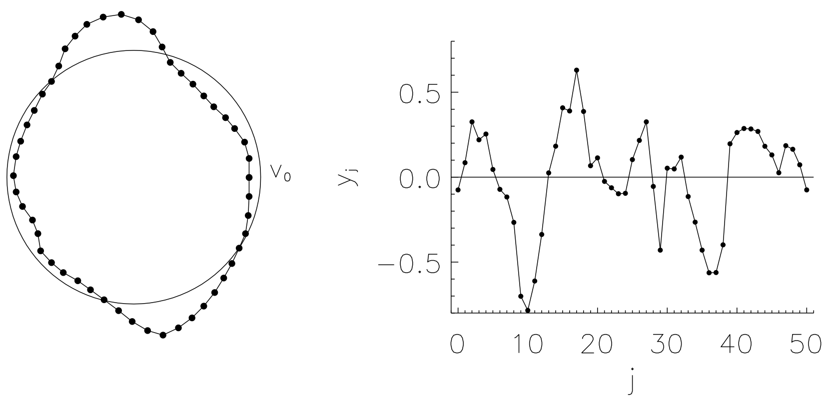

Figure 3: Left: The vertices are equally spaced in terms of arc length. Right: The value of the normal component $\mathbf{y}$ of the standardized edge transformation vector.

Another way of collecting vertices on a boundary is that of equal spacing in terms of arc length. If the deformations from a circle are small, the length of each edge remains constant, to first order. This implies that the value of the tangent component $\mathbf{x}$ of the standardized edge transformation vector is zero to first order, and it suffices to consider the normal component $\mathbf{y}$. In practice, one finds the edge transformation vector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ of the object and consider $\mathbf{y} = \mathrm{Im}(\mathbf{z})$ as the data, cf. Figure 3.

This way of describing the boundary is called the *constant length representation*, and is closely related to the tangent angle function, cf. Zahn and Roskies (1972).

Applications of edge transformation models may be found in Kent *et al*. (2000), where data from sand grains and ceramic material particle sections are analyzed, in Grenander and Miller (1994) who describe mitochondria, or in Hansen *et al*. (2000) where images of arteries are considered.

3 The relation between the transformation vectors

We now have two different ways of describing the shape of the object P relative to the regular polygon P^0, the standardized vertex transformation vector $\mathbf{u}$ and the standardized edge transformation vector $\mathbf{z}$. The edges are determined by the vertices and, together with a specification of the location, the edges determine the vertices. In this section we show a 1-1 correspondence between the two standardized transformation vectors.

To derive the relationship between the transformation vectors we first note that

$$e_j^0 = (e^{2\pi i/n} - 1)v_j^0 = (e^{2\pi i/n} - 1)e^{-2\pi i/n}v_{j+1}^0.$$

From (2.3), (2.8) and (2.9) it now follows that

$$(e^{2\pi i/n} - 1)t_j = e^{2\pi i/n}d_{j+1} - d_j, \qquad j = 0, \ldots, n-1,$$

which gives a complex linear relation between $\mathbf{t}$ and $\mathbf{d}$. The relation implies that $t_{av} = d_{av}$, which shows that the edges and vertices are scaled and rotated in the same way to obtain invariance. Division by $t_{av} = d_{av}$ yields the complex linear relation

$$(e^{2\pi i/n} - 1)z_j = e^{2\pi i/n}u_{j+1} - u_j, \qquad j = 0, \ldots, n-1, \tag{3.1}$$

between the standardized vertex transformation vector $\mathbf{z}$ and the standardized edge transformation vector $\mathbf{u}$. Let B be the circulant $n \times n$ matrix

$$B = \mathrm{circ}(0, 1, 0, \ldots, 0).$$

Then equation (3.1) can be written as

$$(e^{2\pi i/n} - 1)\mathbf{z} = (e^{2\pi i/n}B - I_n)\mathbf{u}. \tag{3.2}$$

The eigenvalues of B are given by, cf. Anderson (1971, p. 280-282),

$$\mu_k = e^{2\pi ik/n}, \qquad k = 0, \ldots, n-1,$$

with corresponding unit eigenvectors $\mathbf{w}_k$, $k = 0, \ldots, n-1$, with components

$$(\mathbf{w}_k)_j = \frac{1}{\sqrt{n}}e^{2\pi ijk/n}, \qquad j = 0, \ldots, n-1. \tag{3.3}$$

Note that $\mathbf{w}_k$ is also an eigenvector for B^T with corresponding eigenvalue $\overline{\mu}_k$, $k = 0, \ldots, n-1$, where $\overline{\mu}_k$ denotes the complex conjugate of μ_k. The eigenvalues of $(e^{2\pi i/n}B - I_n)$ are therefore

$$\lambda_k = \mathbf{w}_k^*(e^{2\pi i/n}B - I_n)\mathbf{w}_k = e^{2\pi i(k+1)/n} - 1, \quad k = 0, \ldots, n-1,$$

where $\mathbf{w}_k^* = \overline{\mathbf{w}}_k^T$ is the transposed complex conjugate of $\mathbf{w}_k$. Since $\lambda_k = 0$ for $k = n-1$ and $\lambda_k \neq 0$ otherwise the matrix has a singularity corresponding to the eigenvector $\mathbf{w}_{n-1}$. The constraints (2.6) and (2.11) can be written as

$$\mathbf{w}_{n-1}^*\mathbf{z} = \mathbf{w}_{n-1}^*\mathbf{u} = 0,$$

and thus the mapping (3.2) is 1-1 on the domain of $\mathbf{z}$ and $\mathbf{u}$. The two other constraints (2.4) and (2.10) can be written in terms of the eigenvector $\mathbf{w}_0$ and thus the 1-1 correspondence between the two standardized transformation vectors follows.

We now recast the relationship between the transformation vectors in terms of real coordinates. Using the equality

$$\tan(\pi/n) = \frac{1 - \cos(2\pi/n)}{\sin(2\pi/n)}$$

it is straightforward to show that

$$\frac{e^{2\pi i/n}}{e^{2\pi i/n} - 1} = \frac{1}{2}(1 - \frac{i}{\tan(\pi/n)}) \text{ and } \frac{1}{e^{2\pi i/n} - 1} = -\frac{1}{2}(1 + \frac{i}{\tan(\pi/n)}).$$

From (3.1) it now follows that the relation between the vertex transformation vector $\mathbf{u}$ and the edge transformation vector $\mathbf{z}$ is given by

$$2z_j = (1 - \frac{i}{\tan(\pi/n)})u_{j+1} + (1 + \frac{i}{\tan(\pi/n)})u_j, \qquad j = 0, \ldots, n-1. \tag{3.4}$$

Letting

$$z_j = x_j + iy_j, \quad u_j = r_j + is_j, \quad j = 0, \ldots, n-1,$$

we can write (3.4) in real coordinates as

$$2x_j \;=\; r_{j+1} + r_j + \frac{1}{\tan(\pi/n)}(s_{j+1} - s_j), \qquad j = 0, \ldots, n-1, \tag{3.5}$$

and

$$2y_j \;=\; -\frac{1}{\tan(\pi/n)}(r_{j+1} - r_j) + s_{j+1} + s_j, \qquad j = 0, \ldots, n-1. \tag{3.6}$$

If we let

$$A = \begin{pmatrix} 1 & 1/\tan(\pi/n) \\ -1/\tan(\pi/n) & 1 \end{pmatrix},$$

then the $2n$ equations (3.5) and (3.6) can be written as

$$2\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = (A \otimes B + A^T \otimes I_n)\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}, \tag{3.7}$$

where $\otimes$ denotes the Kronecker product. The basic equations (3.5) and (3.6) can also be justified using Figure 4.

B

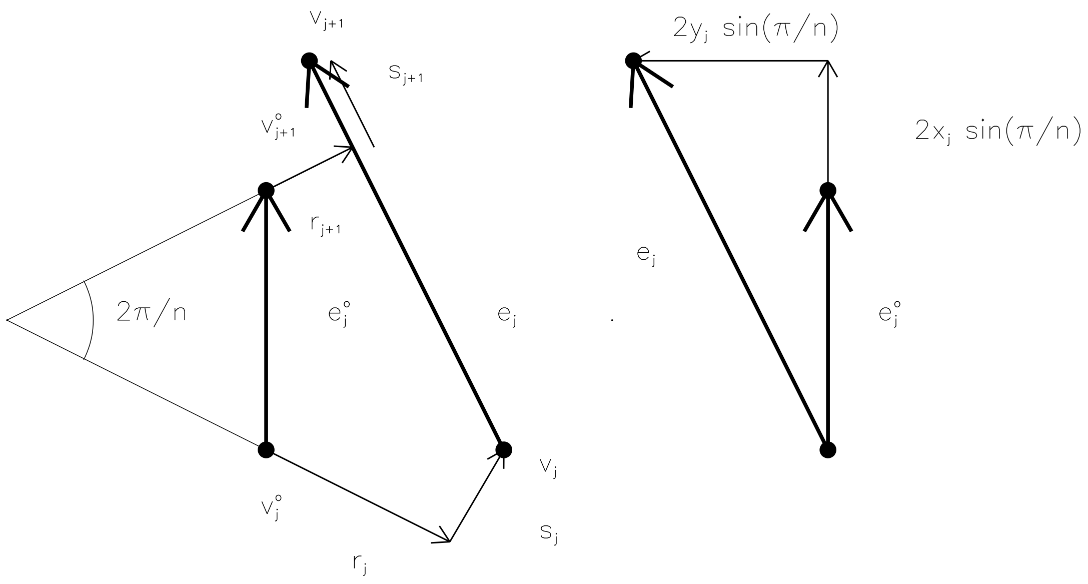

Figure 4: Illustration of the interpretation of r_j, s_j and x_j, y_j in the case $d_{av} = 1$. The factor $2\sin(\pi/n)$ comes from the fact that the length of e_j^0 is $2\sin(\pi/n)$. The basic equations (3.5) and (3.6) can be found from this figure.

4 A model for the transformation vectors

In this section we consider a multivariate normal model for the standardized vertex transformation vector $(\mathbf{r}, \mathbf{s})$. The considerations for a model for the standardized edge transformation vector are similar. To a large extent we follow Kent *et al.* (2000, p. 529-531), but formulate the results in terms of vertices.

It is convenient to partition the $(2n) \times (2n)$ covariance matrix of $(\mathbf{r}, \mathbf{s})$ by

$$\mathrm{Var}\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} = K = \begin{pmatrix} K^{(r,r)} & K^{(r,s)} \\ K^{(s,r)} & K^{(s,s)} \end{pmatrix}.$$

The lack of features on the object implies that any statistical model should be invariant under cyclic permutation of the vertices. Cyclic invariance implies that $\mathrm{E}(r_j, s_j)^T = 0$ and

$$\mathrm{E}\begin{pmatrix} r_i \\ s_i \end{pmatrix}(r_{i+j}, s_{i+j}) = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}$$

cannot depend on i. Hence there exist numbers $a_j, b_j, c_j, d_j, j = 0, \ldots, n-1$ such that

$$K^{(r,r)} = \mathrm{circ}(a_j),\ K^{(r,s)} = \mathrm{circ}(b_j),\ K^{(s,r)} = \mathrm{circ}(c_j),\ K^{(s,s)} = \mathrm{circ}(d_j),$$

where $\mathrm{circ}(a_j)$ denotes the circulant $n \times n$ matrix with first row $(a_0, \ldots, a_{n-1})$. The scale and rotation constraint (2.4) leads to a rank 2 deficiency in K. The translation constraint for the vertices (2.6) leads to another rank 2 deficiency in K.

Since the four partitioned matrices are all circulant we can make a partial diagonalization of K by the unitary matrix

$$W = (\mathbf{w}_0, \dots, \mathbf{w}_{n-1}), \tag{4.1}$$

where $\mathbf{w}_k$ is the coloumn vector defined in (3.3), cf. Anderson (1971, p. 280,281). Let $W^* = \overline{W}^T$ be the transpose of the complex conjugate of W. Then there exist numbers $\alpha_k, \beta_k, \rho_k, \delta_k$ such that,

$$(I_2 \otimes W^*)K(I_2 \otimes W) = \begin{pmatrix} \operatorname{diag}(\alpha_k) & \operatorname{diag}(\beta_k) \\ \operatorname{diag}(\rho_k) & \operatorname{diag}(\delta_k) \end{pmatrix} = D, \tag{4.2}$$

where $\operatorname{diag}(\alpha_k)$ denotes the diagonal $n \times n$ matrix with $\alpha_0, \dots, \alpha_{n-1}$ on the main diagonal.

By diagonalising the blocks

$$D(k) = \begin{pmatrix} \alpha_k & \beta_k \\ \rho_k & \delta_k \end{pmatrix}$$

one can obtain a complete diagonalization of K. Denote the eigenvalues of $D(k)$ by $\kappa_{1,k}$ and $\kappa_{2,k}$ and write the eigenvectors in the form

$$\boldsymbol{\gamma}_{1,k} = \begin{pmatrix} \cos\theta_k \\ e^{i(\phi_k - \pi/2)} \sin\theta_k \end{pmatrix}, \quad \boldsymbol{\gamma}_{2,k} = \begin{pmatrix} -\sin\theta_k \\ e^{i(\phi_k - \pi/2)} \cos\theta_k \end{pmatrix}, \quad 0 \le \theta_k, \phi_k < \pi,$$

for suitable choices of θ_k and ϕ_k. It then follows that $\boldsymbol{\gamma}_{m,k} \otimes \mathbf{w}_k$ is an eigenvector of K for $\kappa_{m,k}, m = 1, 2$. The conditions on $\{a_j, b_j, c_j, d_j\}$ imply conditions on $\{\alpha_k, \beta_k, \rho_k, \delta_k\}$, which can also be expressed as conditions on $\{\kappa_{m,k}, \theta_k, \phi_k\}$. It turns out that for $1 \le k < n/2$, $D(k)$ and $D(n-k)$ have the same eigenvalues $\kappa_{m,k}, m = 1, 2$ with eigenvectors $\boldsymbol{\gamma}_{m,k}, m = 1, 2$ and $\overline{\boldsymbol{\gamma}}_{m,k}, m = 1, 2$, respectively. Together with $\mathbf{w}_k = \overline{\mathbf{w}}_{n-k}$ this implies that $\kappa_{m,k}, m = 1, 2$ has multiplicity 2 for $1 \le k < n/2$. Furthermore, the standardization constraint (2.4) implies that the eigenvalues for $k = 0$ must vanish, $\kappa_{1,0} = \kappa_{2,0} = 0$. Similarly, the translation constraint (2.6) implies that for $k = 1$ the first eigenvalue must vanish, $\kappa_{1,1} = 0$.

In total, the eigenvalues and eigenvectors of $D(k)$ satisfy the following conditions:

$k = 0:$	$\kappa_{1,0} = \kappa_{2,0} = 0;$
	θ_0, ϕ_0 irrelevant.
$k = 1:$	$\kappa_{1,1} = 0, \kappa_{2,1} \ge 0$ a free parameter;
	$\theta_1 = 3\pi/4, \phi_1 = 0.$
$1 < k < n/2:$	$\kappa_{m,k} \ge 0, m = 1, 2,$ free parameters;
	$\theta_k \in [0, \pi[, \phi_k \in [0, \pi[$ free parameters.
$k = n/2,$ (n even) :	$\kappa_{m,n/2} \ge 0, m = 1, 2,$ free parameters;
	$\theta_{n/2} \in [0, \pi[$ a free parameter; $\phi_{n/2} = \pi/2.$

Let

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} \sim \mathrm{N}_{2n}(0, K).$$

With $R = A \otimes B + A^T \otimes I_n$ it follows from (3.7) that

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}_{2n}(0, \frac{1}{4}RKR^T).$$

In general, the free parameters $\{\kappa_{m,k}, \theta_k, \phi_k\}$ for K do not match up in a simple way with the free parameters in a similar diagonalization of $RKR^T/4$.

5 Special cases

To limit the number of free parameters one can propose symmetries or restrictions to the model. In this section we consider such cases. We demonstrate how easy it is to go back and forth between the free parameters of the edge and the vertex model in the complex symmetric case and in the radial representation. The eigenvalues are often restricted through a Markov random field (MRF) model. We show that a first-order MRF model for the edges implies a second-order MRF model for the vertices in the complex symmetric and in the radial representations.

5.1 Complex symmetry

A complex Gaussian random vector $\mathbf{u} = \mathbf{r} + i\mathbf{s}$ with mean $\mathbf{0}$ is said to possess complex symmetry if

$$\mathrm{E}(r_{j_1} r_{j_2}) = \mathrm{E}(s_{j_1} s_{j_2}), \quad \mathrm{E}(r_{j_1} s_{j_2}) = -\mathrm{E}(s_{j_1} r_{j_2}), \quad 0 \le j_1, j_2 \le n-1.$$

This property implies that $\mathbf{u}$ and $e^{i\theta}\mathbf{u}$ have the same distribution for any θ. If $\mathbf{u}$ possesses complex symmetry then from (3.5) and (3.6) it is straightforward to show that $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ also possesses complex symmetry. In the complex symmetric case it can be shown, see Kent *et al.* (2000, p.532), that the blocks of $D(k)$ take the form

$$D(k) = \begin{pmatrix} \alpha_k & i\beta_k \\ -i\beta_k & \alpha_k \end{pmatrix}, \quad 1 \le k < [n/2],$$

where $\alpha_k \ge 0, \beta_k \in \mathbb{R}$, $-\alpha_k \le \beta_k \le \alpha_k$ and $\beta_{n/2} = 0$ (n even). Thus, the eigenvalues of $D(k)$ are $\alpha_k \pm \beta_k$ with corresponding unit eigenvectors

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad \mathbf{q}_2 = \overline{\mathbf{q}}_1.$$

In this case the blocks of D can be expressed in the following parameters:

$k = 0:$	$\kappa_{1,0} = \kappa_{2,0} = 0;$
	θ_0, ϕ_0 irrelevant.
$k = 1:$	$\kappa_{1,1} = 0, \kappa_{2,1} \ge 0$ a free parameter;
	$\theta_1 = 3\pi/4, \phi_1 = 0.$
$1 < k < n/2:$	$\kappa_{m,k} \ge 0, m = 1, 2$ free parameters;
	$\theta_k = \pi/4, \phi_k = 0.$
$k = n/2,$ (n even) $:$	$\kappa_{m,n/2} \ge 0, m = 1, 2$ free parameters;
	$\theta_{n/2} = 0, \phi_{n/2} = \pi/2.$

Hence, the complex symmetric model is determined by the sequence of eigenvalues $\kappa_{1,k}$ and $\kappa_{2,k}$. The covariance matrix of $\mathbf{u}$ and the covariance matrix of $\mathbf{z}$ have the same eigenvectors and the eigenvalues $\kappa^u_{m,k}$ and $\kappa^z_{m,k}$ are related by the following proposition.

Proposition 1 *In the complex symmetric model the free eigenvalues for the vertex transformation vector* $\kappa^u_{m,k}$ *and the free eigenvalues for the edge transformation vector* $\kappa^z_{m,k}$ *are related by*

$$\kappa^z_{1,k} = g_1(k)\kappa^u_{1,k}, \qquad 2 \le k \le [n/2],$$

and

$$\kappa^z_{2,k} = g_2(k)\kappa^u_{2,k}, \qquad 1 \le k \le [n/2],$$

where

$$g_1(k) = \frac{1}{2}(1 + \frac{1}{\tan(\pi/n)^2}) + \frac{1}{2}(1 - \frac{1}{\tan(\pi/n)^2})\cos(2\pi k/n) - \frac{\sin(2\pi k/n)}{\tan(\pi/n)},$$

and

$$g_2(k) = \frac{1}{2}(1 + \frac{1}{\tan(\pi/n)^2}) + \frac{1}{2}(1 - \frac{1}{\tan(\pi/n)^2})\cos(2\pi k/n) + \frac{\sin(2\pi k/n)}{\tan(\pi/n)}.$$

Proof. First note that $\mathbf{q}_1$ and $\mathbf{q}_2$ are also eigenvectors of A (defined in (3.7)) with corresponding eigenvalues

$$\omega_1 = 1 + i/\tan(\pi/n) \;\text{ and }\; \omega_2 = \overline{\omega}_1.$$

The Proposition now follows from a direct calculation since

$$\begin{aligned}
\kappa^z_{1,k} &= (\mathbf{q}_1 \otimes \mathbf{w}_k)^* \frac{1}{4} R K R^T (\mathbf{q}_1 \otimes \mathbf{w}_k) \\
&= \frac{1}{4}(\mathbf{q}_1 \otimes \mathbf{w}_k)^*(A \otimes B + A^T \otimes I_n) K (A^T \otimes B^T + A \otimes I_n)(\mathbf{q}_1 \otimes \mathbf{w}_k) \\
&= \frac{1}{4}(\omega_1 \mu_k + \overline{\omega}_1)(\overline{\omega}_1 \overline{\mu}_k + \omega_1)(\mathbf{q}_1 \otimes \mathbf{w}_k)^* K (\mathbf{q}_1 \otimes \mathbf{w}_k) \\
&= g_1(k)\kappa^u_{1,k}, \quad 2 \le k \le [n/2].
\end{aligned}$$

The relation between $\kappa^z_{2,k}$ and $\kappa^u_{2,k}$ is shown in a similar way. An alternative proof can be based on the complex representation (4.3). The quantities $g_1(k)$ and $g_2(k)$ appear as the squared absolute eigenvalues of $(e^{2\pi i/n}B - I_n)/(e^{2\pi i/n} - 1)$ for the eigenvectors $\mathbf{w}_{n-k}$ and $\mathbf{w}_k$, respectively. □

Note that g_1 and g_2 are non-linear and largely monotonic (except near $k = n/2$). For large n and small k the relationship is approximately quadratic, as $g_1(k) \approx (k-1)^2$ and $g_2(k) \approx (k+1)^2$; see Figure 5.

The complex symmetric model is a rather restrictive model since it forces the tangential and normal components r_j, s_j (x_j, y_j) of the standardized vertex (edge) transformation vector to have the same variance and to be uncorrelated.

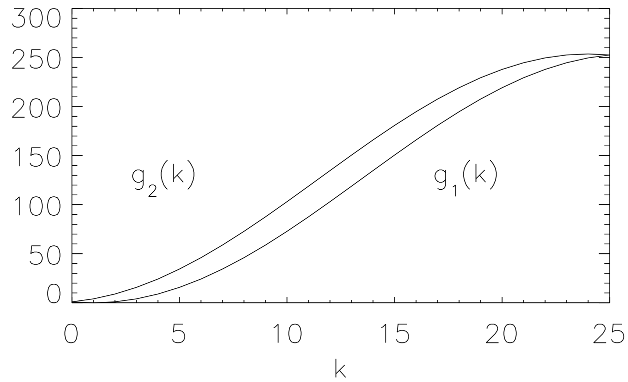

Figure 5: Plot of $g_1(k)$ and $g_2(k)$ for $n = 50$.

5.2 Complex symmetry with a MRF structure

One way to impose further structure on the eigenvalues is through a pth-order Markov random field (MRF) model. The complex form of this model for an n-vector $\mathbf{u}$ takes the form

$$\mathrm{E}\mathbf{u} = \mathbf{0}, \qquad \mathrm{E}\mathbf{u}\mathbf{u}^* = 2\Sigma,$$

where $\Sigma^{-1} = (\sigma^{jk})$ is banded with nonzero entries

$$\sigma^{j,j+l} = \overline{\sigma}^{j+l,j} = \alpha_l, \qquad l = 0, \ldots, p.$$

Here $\alpha_0 > 0$ and $\alpha_1, \ldots, \alpha_p$ are allowed to be complex. The eigenvalues η_k^u of this circulant covariance matrix are given by

$$1/\eta_k^u = \alpha_0 + 2\sum_{l=1}^{p} \mathrm{Re}(\alpha_l e^{2\pi ikl/n}), \quad k = 0, \ldots, n-1.$$

In the terminology of Section 4, $\kappa_{1,k} = \eta_{n-k}^u, \kappa_{2,k} = \eta_k^u$. The α_l parameters must be chosen to ensure the eigenvalues are positive in order to guarantee a positive definite covariance matrix. A sufficient condition is $2\sum_{l=1}^{p} |\alpha_l| < \alpha_0$.

We now describe how to include the constraints (2.4) and (2.6), or equivalently (2.10) and (2.11). Let

$$\Sigma = W^* \mathrm{diag}\{\eta_k^u\} W$$

be the spectral decomposition of a complex covariance matrix satisfying the MRF assumption. Then the covariance matrix

$$W^* \mathrm{diag}\{0, \eta_0^u, \ldots, \eta_{n-2}^u, 0\} W$$

with the eigenvalues for $k = 0, n-1$ replaced by 0 satisfies the required constraints. Note that the parameter space for the α_l parameters is larger in the constrained model since fewer eigenvalues are required to be positive.

Consider for simplicity the complex symmetric first-order MRF model for the edge transformation vector $\mathbf{z}$ given by

$$\alpha_0 = 1/\sigma^2,\ \alpha_1 = -\beta/(2\sigma^2),$$

and subject to the constraints (2.10) and (2.11). A sufficient condition for positive definiteness for the covariance matrix is

$$\sigma^2 > 0,\ \beta \in \mathbb{C} \text{ and } |\beta| < 1,$$

but note that this parameter space can be enlarged as just described.

Proposition 2 *The constrained complex symmetric first-order MRF model for the edge transformation vector* $\mathbf{z}$ *corresponds to a constrained complex symmetric second-order MRF model for the vertex transformation vector* $\mathbf{u}$ *given by*

$$\begin{aligned}
\alpha_0 &= \{2(1+\frac{1}{\tan(\pi/n)^2}) - (1-\frac{1}{\tan(\pi/n)^2})\operatorname{Re}(\beta) + 2\frac{\operatorname{Im}(\beta)}{\tan(\pi/n)}\}/(4\sigma^2) \\
\alpha_1 &= \{(1-\frac{i}{\tan(\pi/n)})^2 - (1+\frac{1}{\tan(\pi/n)^2})\beta\}/(4\sigma^2) \\
\alpha_2 &= \{-\frac{1}{2}(1-\frac{i}{\tan(\pi/n)})^2\beta\}/(4\sigma^2).
\end{aligned}$$

Proof. In terms of $\mathbf{z}$ and $\mathbf{u}$ we can write (3.4) as

$$2\mathbf{z} = (1-\frac{i}{\tan(\pi/n)})B\mathbf{u} + (1+\frac{i}{\tan(\pi/n)})\mathbf{u},$$

and the proposition follows from the change of variables formula. Another way to prove Proposition 2 is to use the real version of the model and apply the relationship between the eigenvalues established in Proposition 1. □

Thus, a conditioned first-order MRF model for the edge transformation vector is equivalent to a conditioned second-order MRF model for the vertex transformation vector. Note that in general a pth-order MRF model for the edge transformation vector corresponds to a $(p+1)$th-order MRF model for the vertex transformation vector. A similar result can be found in Kent *et al.* (1996).

An example application of the complex symmetric edge transformation model is given by Hurn *et al.* (1999) who use the model as the prior for one type of cell. The complex symmetric edge transformation model has a MRF structure with $\beta \in \mathbb{R}$, and their model incorporates a constraint for closure, but not for size and rotation.

5.3 The radial representation

Suppose there is variability only in the radial component $\mathbf{r}$ of the vertex transformation vector $\mathbf{u} = \mathbf{r} + i\mathbf{s}$ and no variability in the normal component $\mathbf{s}$. This restriction leads to zero entries in K apart from $K^{(r,r)}$. Denote the eigenvalues of the covariance matrix $K^{(r,r)}$ by κ_k^r with corresponding eigenvectors $\mathbf{w}_k$, $k = 0, \ldots, n-1$. Similarly, let κ_k^x and κ_k^y denote the eigenvalues of the covariance matrix of the tangent component $\mathbf{x}$ and the normal component $\mathbf{y}$ of the edge transformation vector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$.

B

Proposition 3 *In the radial representation the free eigenvalues of the tangent component of the vertex transformation vector, κ_k^r, and the free eigenvalues of the tangent and normal component of the edge transformation vector, κ_k^x and κ_k^y, respectively, are related by*

$$\kappa_k^x = \frac{1}{2}(1+\cos(2\pi k/n))\kappa_k^r, \quad k = 0,\dots,n-1,$$

and

$$\kappa_k^y = \frac{1-\cos(2\pi k/n)}{2\tan(\pi/n)^2}\kappa_k^r, \qquad k = 0,\dots,n-1. \tag{5.1}$$

Proof. Since there is no variability in the normal component of the vertex transformation vector $\mathbf{s} = \mathbf{0}$, and we get from (3.5) that

$$2\mathbf{x} = (B+I)\mathbf{r}. \tag{5.2}$$

A direct calculation gives

$$\begin{aligned}
\kappa_k^x &= \mathbf{w}_k^* K^{(x,x)} \mathbf{w}_k \\
&= \frac{1}{4}\mathbf{w}_k^*(B+I)K^{(r,r)}(B^T+I)\mathbf{w}_k \\
&= \frac{1}{4}(\mu_k+1)(\overline{\mu}_k+1)\mathbf{w}_k^* K^{(r,r)}\mathbf{w}_k \\
&= \frac{1}{2}(1+\cos(2\pi k/n))\kappa_k^r, \quad k = 0,\dots,n-1.
\end{aligned}$$

From (3.6) we have that

$$2\mathbf{y} = -\frac{1}{\tan(\pi/n)}(B-I)\mathbf{r}, \tag{5.3}$$

and (5.1) follows from a similar calculation. □

Note that for large n and small k we have $\kappa_k^x \approx \kappa_k^r$ and $\kappa_k^y \approx k^2\kappa_k^r$. The major part of the variability from a circle is described by the eigenvectors with small and large k. Thus, in the radial representation with many landmarks the large part of the variability from a circle is in the normal direction of the edge transformation vector.

5.4 The radial representation with a MRF structure

The real form of a pth-order model for an n-vector $\mathbf{r}$ takes the form

$$\mathrm{E}\mathbf{r} = \mathbf{0}, \qquad \mathrm{E}\mathbf{r}\mathbf{r}^T = \Sigma,$$

where $\Sigma^{-1} = \sigma^{jk}$ is banded with non-zero entries

$$\sigma^{j,j+l} = \sigma^{j+l,j} = \alpha_l, \qquad l = 0,\dots,p.$$

Here $\alpha_0 > 0$ and $\alpha_1, \ldots, \alpha_p$ are real and chosen to ensure the eigenvalues

$$\kappa_k = 1/(\alpha_0 + 2\sum_{l=1}^{p} \alpha_l \cos(2\pi kl/n))$$

of the covariance matrix are positive.

A periodic stationary stochastic process can be derived by letting $n \to \infty$ and letting the α_l vary with n. The limiting process is most simply described as a random Fourier series

$$r(t) = \sum_{k\in\mathbb{Z}} \{\zeta_k^c (\kappa_k^{(\infty)})^{1/2} \cos(2\pi kt) + \zeta_k^s (\kappa_k^{(\infty)})^{1/2} \sin(2\pi kt)\},$$

where the ζ_k^c, ζ_k^s are iid N(0,1) random variables and the eigenvalues take the form

$$1/\kappa_k^{(\infty)} = a_0 + \sum_{l=1}^{p} a_l (2\pi k)^{2l},$$

and the a_l are real. If this process is restricted to the equally-spaced points $t = 2\pi j/n, j = 0, \ldots, n-1$, for some value of n, then a circulant random vector results, with eigenvalues $\kappa_k^{(C)}$, say, given by

$$1/\kappa_k^{(C)} = \sum_{m\in\mathbb{Z}} 1/\kappa_{k+nm}^{(\infty)},$$

after taking aliasing into account.

Consider the radial model with no variability in the normal component of the vertex transformation vector. The relation between $\mathbf{y}$ and $\mathbf{r}$ is then given by (5.3). The constraints on $\mathbf{r}$ are

$$\sum_{j=0}^{n-1} r_j = \sum_{j=0}^{n-1} r_j \cos(2\pi j/n) = \sum_{j=0}^{n-1} r_j \sin(2\pi j/n) = 0,$$

and these constraints can be recast in terms of $\mathbf{y}$ as

$$\sum_{j=0}^{n-1} y_j = \sum_{j=0}^{n-1} y_j \cos(2\pi j/n) = \sum_{j=0}^{n-1} y_j \sin(2\pi j/n) = 0.$$

As in the complex symmetric case the constraints are most simply implemented in the spectral domain where they correspond to setting $\kappa_0^r = \kappa_1^r = \kappa_{n-1}^r = 0$.

For simplicity let the model for the normal component of the edge transformation vector be a first-order MRF model given by

$$\alpha_0 = 1/\sigma^2,\ \alpha_1 = -\beta/(2\sigma^2),$$

with the sufficient condition

$$\sigma^2 > 0,\ \beta \in \mathbb{R} \text{ and } |\beta| < 1,$$

for positive definiteness for the covariance matrix.

Proposition 4 *The constrained radial first-order MRF model for the normal component of the edge transformation vector corresponds to a constrained second-order MRF model for the tangent component of the vertex transformation vector given by*

$$\alpha_0 = c(2+\beta),\ \alpha_1 = -c(1+\beta),\ \alpha_2 = c\beta/2,$$

where $1/c = 4\sigma^2 \tan(\pi/n)^2$.

Proof. From Proposition 3 it follows that

$$\begin{aligned} 1/\kappa_k^r &= \frac{1-\cos(2\pi k/n)}{2\tan(\pi/n)^2}\frac{1}{\kappa_k^y} \\ &= \frac{2+\beta-(1+\beta)2\cos(2\pi k/n)+\beta\cos(4\pi k/n)}{4\sigma^2\tan(\pi/n)^2} \\ &= \frac{1}{4\sigma^2\tan(\pi/n)^2}\mathbf{w}_k^*\Big((2+\beta)I_n-(1+\beta)(B+B^T)+\frac{\beta}{2}(B^2+(B^T)^2)\Big)\mathbf{w}_k, \\ & \quad k = 2,\ldots,n-2. \end{aligned}$$

The proposition can also be shown directly from (5.3) and the change of variables formula. □

6 Data analysis

6.1 Data representation and model specification

For arbitrarily spaced vertices around the outline of an object, the full normal models for the edge and vertex representations are interchangeable. However, in practice the data analyst can often choose the spacing of the vertices. Choosing the vertices to be regularly spaced, or the edges to have constant length, leads to a more succinct description of the data. In Section 2.1 and 2.2 we discussed the radial and the constant length representations. They both have the advantage of reducing the dimension of the data from $2n$ to n without losing any important information.

Consider the radial representation. Since there is no variability in the normal component $\mathbf{s}$ of the standardized vertex transformation vector we have zero entries in K apart from $K^{(r,r)}$. As in the previous section denote the eigenvalues of $K^{(r,r)}$ by κ_k^r with corresponding eigenvectors $\mathbf{w}_k, k = 0,\ldots,n-1$. Since $K^{(r,r)}$ is symmetric and circulant and $\mathbf{w}_k = \overline{\mathbf{w}}_{n-k}$ we have $\kappa_k^r = \kappa_{n-k}^r, k = 1,\ldots,[n/2]$. Therefore we only have to consider $\kappa_0^r,\ldots,\kappa_{[n/2]}^r$. The constraints (2.4) and (2.6) imply $\kappa_0^r = \kappa_1^r = \kappa_{n-1}^r$. The considerations for the constant length representation are similar. For the constant length representation there is no variability in the radial component $\mathbf{x}$ of the standardized edge transformation vector, and thus we only have to consider the normal component $\mathbf{y}$, which is modelled by the symmetric and circulant covariance matrix $K^{(y,y)}$ with the same constraints.

To impose further structure on the eigenvalues we consider a real version of the pth-order continuous MRF model discretized to n points as described in Section 5.4.

The general pth-order model has $p+1$ parameters, but from exploratory fits we found that in our application a two parameter pth-order model with $a_1 = \cdots = a_{p-1} = 0$ is sufficiently flexible. Thus, the free eigenvalues are given by

$$\kappa_k^r = \sum_{m \in \mathbf{Z}} \frac{1}{\alpha + \beta\{2\pi(k+nm)\}^{2p}}, \qquad 2 \leq k \leq [n/2]. \tag{6.1}$$

The continuous two parameter first-order model with $p = 1$ was considered in Kent *et al.* (2000) and the continuous two parameter second-order model with $p = 2$ was considered in Hobolth and Jensen (1999), although without taking aliasing into account (i.e. only the term with $m = 0$ is present). The parameter space is given by

$$\beta > 0, \qquad \alpha > -(4\pi)^{2p}\beta,$$

to ensure that the eigenvalues are positive. The parameters α, β determine the overall shape variability. The parameter β in (6.1) determines how fast the eigenvalues are decreasing and can be viewed as a smoothness parameter. The parameter α is mainly important for the first few eigenvalues and therefore reflects global shape variability.

It is often useful to consider the marginal likelihoods for the first few eigenvalues only. Following Kent *et al.* (2000) we shall label the model $Full(k^*)$ when we ignore the eigenvalues with indices greater than k^*, where k^* satisfies $2 \leq k^* \leq [n/2]$. Similarly the two parameter model with eigenvalues given by (6.1) is labelled two parameter $CMRF(p, k^*)$ when we ignore the eigenvalues with indices greater than k^*.

6.2 Statistical inference

Next we consider statistical inference for some of the radial models. Inference in the full normal model of Section 4 may be found in Kent *et al.* (2000). Consider a random sample of N standardized vertex transformation vectors from the radial representation

$$\mathbf{r}_l = (r_{l,0}, r_{l,1}, \ldots, r_{l,n-1})^T, \qquad l = 1, \ldots, N,$$

drawn independently from the multivariate normal model

$$\mathbf{r}_l \sim \mathrm{N}(\mathbf{0}, K^{(r,r)}), \qquad l = 1, \ldots, N,$$

where $K^{(r,r)}$ is a symmetric, circulant $n \times n$ matrix with the constraints $\kappa_0^r = \kappa_1^r = \kappa_{n-1}^r = 0$. The maximum likelihood estimator (m.l.e.) of $K^{(r,r)}$ is given by

$$\hat{K}^{(r,r)} = \frac{1}{Nn} \sum_{l=1}^{N} \sum_{p=0}^{n-1} \mathbf{r}_l^{(p)} (\mathbf{r}_l^{(p)})^T,$$

where $\mathbf{r}_l^{(p)}$ denotes $\mathbf{r}_l$ cyclically permuted by p sites, ie.

$$\mathbf{r}_l^{(p)} = (r_{l,p}, r_{l,1+p}, \ldots, r_{l,n-1+p})^T,$$

and the sample eigenvalues are

$$\hat{\kappa}_k^r = \mathbf{w}_k^* \hat{K}^{(r,r)} \mathbf{w}_k, \qquad k = 0, \ldots, n-1.$$

Ignoring the components κ_0^r and $\kappa_1^r = \kappa_{n-1}^r$ and using $\kappa_k^r = \kappa_{n-k}^r, k = 2, \ldots, [n/2]$, the marginal likelihood becomes

$$(\prod_{k=2}^{[n/2]} 2\pi(\kappa_k^r)^{\tau_k})^{-N/2} \exp(-\frac{N}{2} \sum_{k=2}^{[n/2]} \tau_k \frac{\hat{\kappa}_k^r}{\kappa_k^r}), \tag{6.2}$$

where $\tau_k = 2, 1 \leq k < [n/2]$, and $\tau_{n/2} = 1$, the last term being present only for n even. Thus, under the radial representation the m.l.e.s of the parameters $\kappa_k^r, k = 2, \ldots, [n/2]$ are the sample eigenvalues $\hat{\kappa}_k^r, k = 2, \ldots, [n/2]$.

The likelihood for the more restricted two parameter $CMRF(p, k^*)$ is given by (6.2) with $[n/2]$ replaced by k^* and κ_k^r given by (6.1). To calculate the m.l.e.s of (α, β) we use numerical maximization of the log-likelihood.

Testing if the two parameter $CMRF(p, k^*)$ is a reasonable simplification of the $Full(k^*)$ is carried out using a classical likelihood ratio test. Let L_1 be the maximized likelihood under the $Full(k^*)$ with $\nu_1 = k^* - 1$ free parameters and L_2 the maximized likelihood under the two parameter $CMRF(p, k^*)$ with $\nu_2 = 2$ free parameters. For a large sample size N we have the chi-squared approximation

$$-2 \log (L_2/L_1) \simeq \chi^2_{(\nu_1 - \nu_2)}.$$

6.3 Application

We consider inference for the parameters in the models using the sand particle data for illustration. For each of the N sand grains we first determine the vector

$$\mathbf{r}_l = (r_{l,0}, r_{l,1}, \ldots, r_{l,n-1})^T, \qquad l = 1, \ldots, N,$$

using the radial representation described in Section 2.1. Each sand grain is represented by a binary image (1 for an object grey level and 0 for a background grey level). The centre of gravity of the image is obtained, and then equally spaced radii are traced out to the boundary. A boundary point is defined as the pixel location before the first zero is obtained. The number of vertices n on each outline should be chosen high enough to capture the shape of the object, but since the objects are only represented as binary images not much extra information is gained by choosing a very high n. We have chosen the rather high $n = 50$.

Now consider the constraints. From $\sum_{j=0}^{n-1} r_j = 0$ it follows that $\hat{\kappa}_0^r = 0$. If the sample mean of the vertices is zero then we get from (2.6) that $\hat{\kappa}_1^r = \hat{\kappa}_{n-1}^r = 0$. In practice the sample mean of the vertices is only approximately zero and hence $\hat{\kappa}_1^r = \hat{\kappa}_{n-1}^r$ is only approximately zero.

In order to avoid the effect of digitization it is useful to consider only the first few well-determined eigenvalues. The choice of the truncation value k^* is based on the eigenvalues, which should be truncated when they are essentially determined by noise

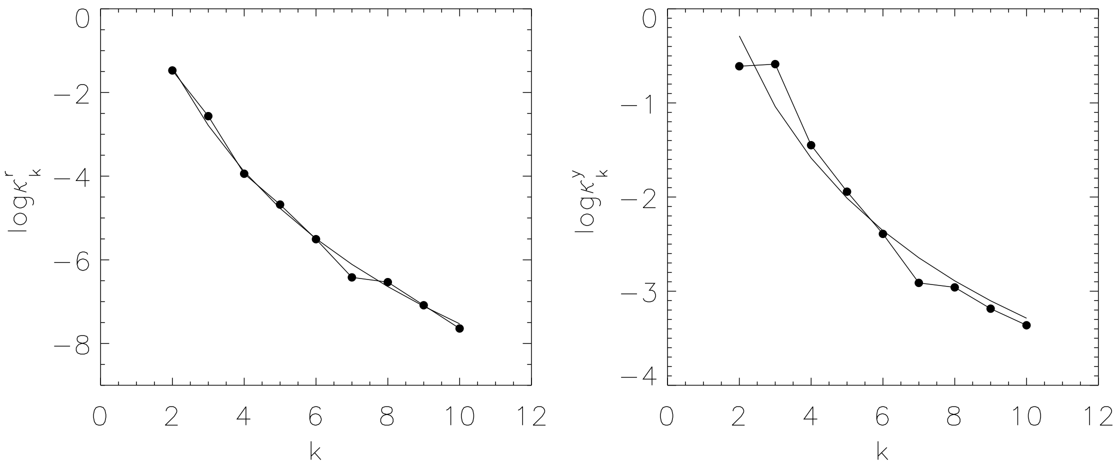

Figure 6: Plot of the logarithm of the eigenvalues under the $Full(10)$ (line with filled points) and under the two parameter $CMRF(p, 10)$ (line without points) for the radial ($p = 2$, left) and constant length ($p = 1$, right) representations.

due to digitization effects. We choose $k^* = 10$. Note that when k^* is small compared to $[n/2]$ then the specific choice of n is less important. The sample log-eigenvalues $\log\hat{\kappa}^r_k$ are shown in Figure 6 for $k = 2, \ldots, k^*$.

A two parameter $CMRF(2, 10)$ is fitted to the data and the m.l.e.s are $(\hat{\alpha}, \hat{\beta}) = (1.108, 0.000120)$. To see whether the two parameter $CMRF(2, 10)$ is reasonable the likelihood ratio test is carried out. We obtain $-2\log(L_2/L_1) = 4.34$ and accept the MRF model on $\nu_1 - \nu_2 = 7$ degrees of freedom, the p-value being 0.74. In Figure 6 we have plotted the log-eigenvalues under the fitted two parameter $CMRF(2, 10)$ model. Similarly one can fit a two parameter $CMRF(1, 10)$ or a two parameter $CMRF(3, 10)$ to the data and it turns out that neither of these models is appropriate.

A similar analysis as for the radial vertex representation was carried out for the constant length edge representation. For each of the N sand grains we first determine the vector

$$\mathbf{y}_l = (y_{l,0}, \ldots, y_{l,n-1})^T, \qquad l = 0, \ldots, N,$$

as described in Section 2.2. As before we denote the eigenvalues of $K^{(y,y)}$ by κ^y_k. The constraints are as in the radial representation. From (2.10) it follows that $\hat{\kappa}^y_0 = 0$. Furthermore, since (2.11) holds approximately with z_j replaced by y_j we have $\hat{\kappa}^y_1 = \hat{\kappa}^y_{n-1} \approx 0$. The two parameter $CMRF(p, 10)$ is then fitted to the data. In this case it turns out that the first-order model with $p = 1$ is appropriate and the m.l.e.s are $(\hat{\alpha}, \hat{\beta}) = (0.125, 0.0077)$. When we test the two parameter $CMRF(1, 10)$ against the $Full(10)$ we obtain the likelihood ratio $-2\log(L_2/L_1) = 10.56$ and accept the two parameter $CMRF(1, 10)$ on 7 degrees of freedom, the p-value being 0.16. In Figure 6 we have plotted the fitted log-eigenvalues under the full model and under the first-order MRF model. Bearing the results from Section 5 in mind it should be no surprise that a first-order MRF model for an edge model 'corresponds' to a second-order MRF model for a vertex model.

An effective way of examining a model is to inspect random samples from the model. In Figure 7 we have simulated outlines from the fitted MRF models. The outlines from the radial and constant length representations look like realistic sea sand grains. Furthermore the simulated samples from the two representations look very much alike. The simulated outlines are a little smoother than the sand grains, but this was expected because of the truncation. From the simulations it seems that both models are adequate for the sea sand grains.

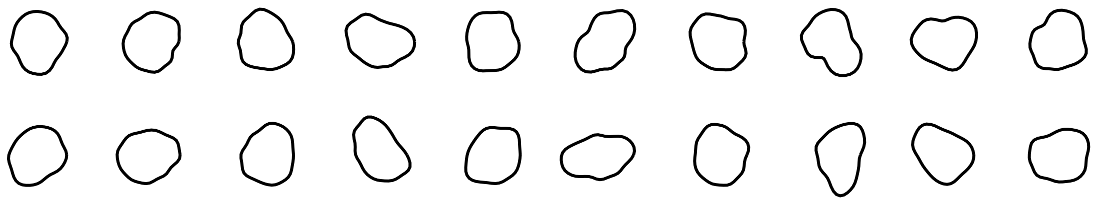

Figure 7: Simulated objects under the fitted MRF models. The first row shows simulations from the radial second-order MRF model and the second row is from the constant length first-order MRF model.

In comparing the two methods it does seem that in this particular application the radial vertex representation follows the second-order MRF model more closely than the constant length edge representation follows the first-order MRF model (although both are deemed reasonable fits from the tests). The radial vertex representation also has the advantage that in general it is easier to construct the discrete data from a continuous outline, although the representation is restricted to star shaped objects.

6.4 Recommendation

The shape of objects in the plane with no obvious landmarks can be described by a variety of different methods. In this paper we have concentrated on an approach based on the standardized edge or vertex transformation vector. When deciding which model to use for the transformation vector one has to take into account how the vertices have been collected. If there is variability in both the real and imaginary part of the transformation vector, then the full normal model described in Section 4 is needed. This model has the required circulant symmetries, but is also rather complicated with many $(2n-4)$ free parameters. One way of imposing constraints on the free parameters is through the more restrictive complex symmetric or even complex symmetric Markov random field models. The complex linear 1-1 relation between the standardized transformation vectors makes it possible to recast any edge model in terms of a vertex model and vice versa. In the general model the free parameters do not match up in a simple way, but in the complex symmetric models there is a simple correspondence as described in Proposition 1 and Proposition 2 of Section 5. See Table 1 for a summary of the possible representations and models when the vertices are arbitrarily spaced.

Choosing the vertices to be regularly spaced, or the edges to have constant length, leads to a more succinct description of the data than when the vertices are arbitrarily

Representation	Data	Models		
vertex	$\mathbf{u} = \mathbf{r} + i\mathbf{s}$	General	Complex symm.	Complex symm. MRF
edge	$\mathbf{z} = \mathbf{x} + i\mathbf{y}$	(Section 4)	(Section 5.1)	(Section 5.2)

Table 1: Table summarizing the possible representations and models when the vertices are arbitrarily spaced.

spaced. The radial and the constant length representations have the advantage of reducing the dimension of the data from $2n$ to n without losing any important information, and furthermore they allow the data analyst to focus on honest differences rather than differences due to the vertex spacing. Unfortunately the vertex spacing is different in the two representations so a direct comparison of the full radial model and the full constant length model is not possible. Instead of the full models we might again impose restrictions on the eigenvalues of the covariance matrix through Markov random field models. For the sand grains the radial representation follows a MRF model more closely than the constant length representation, but this outcome would have been hard to predict. If the vertex spacing can be decided by the data analyst then we suggest analysing both the radial and the constant length model before deciding on a final model. See Table 2 for a summary of the possible representations and models when the vertices are regularly spaced or the edges have constant length.

Representation	Data	Models		
Radial vertex	$\mathbf{r}$	$Full(k^*)$	$CMRF(p, k^*)$	two parameter $CMRF(p, k^*)$
Constant length edge	$\mathbf{y}$	(Section 6.1)	(Section 5.4 and 6.1)	(Section 6.1)

Table 2: Table summarizing the possible representations and models when the vertices are regularly spaced or the edges have constant length.

Acknowledgements

The first author thanks Eva B. Vedel Jensen for fruitful discussions. The authors wish to thank Dietrich Stoyan for kindly providing the sand particle data and an associate editor and two referees for constructive suggestions leading to a better presentation of the contents of the paper.

References

Anderson, T.W. (1971). *The statistical snalysis of time series.* Wiley, New York.

Grenander, U. and Manbeck, K.M. (1993). A stochastic model for defect detection in potatoes. *J. Comput. Graph. Statist.* **2**, 131-151.

Grenander, U. and Miller, M.I. (1994). Representations of knowledge in complex systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **56**, 549-603.

Hansen, M.B., Møller, J. & Tøgersen, F.Aa. (2000). Bayesian contour detection in a time series of ultra-sound images through dynamic deformable template models. *Research Report* **17**, Centre for Mathematical Physics and Stochastics, University of Aarhus. To appear in *Biostatistics.*

Hobolth, A. and Jensen, E.V.B. (1999). Modelling stochastic changes in curve shape, with an application to cancer diagnostics. *Adv. in Appl. Probab. (SGSA)* **32**, 344-362.

Hurn, M.A., Steinsland, I and Rue, H. (1999). Parameter estimation for a deformable template model. Statistics Research Report 99.02, University of Bath. To appear in *Statistics and Computing.*

Johnson, D.R., O'Higgins, P., McAndrew, T.J., Adams, L.M. and Flinn, R.M. (1985). Measurement of biological shape: a general method applied to mouse vertebrae. *Journal of Embryology and Experimental Morphology* **90**, 363-377.

Kent, J.T., Dryden, I.L. and Anderson, C.R. (2000). Using circulant symmetry to model featureless objects. *Biometrika* **87**, 527-544.

Kent, J.T., Mardia, K.V. and Walder, A.N. (1996). Conditional cyclic Markov random fields. *Adv. in Appl. Probab. (SGSA)* **28**, 1-12.

Mardia, K.V. and Qian, W. (1995). Bayesian method for compact object recognition from noisy images. In: Titterington, D.M., editor, *Complex stochastic systems in science and engineering*, pages 155-165, Clarendon Press, Oxford.

Mardia, K.V., Qian, W., Shah, D., and De Souza, K. (1996). Deformable template recognition of multiple occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, 1036-1042.

Rohlf, F.J. and Archie, J. (1984). A comparison of Fourier methods for the description of wing shape in mosquitoes. *Systematic Zoology* **33**, 302-317.

Rue, H. and Syversveen, A.R. (1998). Bayesian object recognition with Baddeley's Delta Loss. *Adv. in Appl. Probab. (SGSA)*, **30**, 64-84.

Stoyan, D. (1997). Geometrical means, medians and variances for samples of particles. *Particle and Particle Systems Characterization* **14**, 30-34.

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes and point fields: methods of geometric statistics.* Wiley, Chichester.

Zahn, C.T. and Roskies, R.Z. (1972). Fourier descriptors for plane closed curves. *IEEE Trans. Comput.* $\mathbf{C-21}$, 269-281.

Asger Hobolth, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark. E-mail address: asho@imf.au.dk.

C

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2000).
A continuous parametric shape model.
Research report no. 13, Laboratory for Computational Stochastics, University of Aarhus.

C

A continuous parametric shape model

Asger Hobolth, Jan Pedersen and Eva B. Vedel Jensen
Laboratory for Computational Stochastics
University of Aarhus

Abstract

In the present paper we propose a flexible continuous parametric shape model for a star-shaped planar object. The model is based on a polar Fourier expansion of the normalized radius-vector function of the object. The expected phase amplitudes are modelled by a simple regression with parameters having simple geometric interpretations. The model is a generalization of first- and second-order Gaussian shape models and is called the generalized $p-$order model. In particular, non-Gaussian errors are allowed. The statistical analysis is straightforward, as demonstrated on a data set concerning shape discrimination of two cell populations.

Keywords: Cancer diagnostics; Featureless objects; Fourier descriptors; Radius-vector function; Shape; Star-shaped objects.

1 Introduction

Recently, shape modelling of featureless objects has attracted a lot of attention in the statistical literature. The Gaussian model with cyclic invariance properties, described by Grenander & Miller (1994), has played a predominant role.

One line of research has been concerned with the application of the Gaussian model as a prior model in Bayesian object recognition. Such an application has been discussed in Grenander & Miller (1994). The group around Håvard Rue has also contributed significantly to this research, cf. e.g. Rue & Syversveen (1998) and Rue & Hurn (1999). In Hansen et al. (2000) a similar Bayesian analysis is performed where also the time aspect has been taken into account.

Another line of research has dealt with likelihood analysis of the Gaussian model. This approach is useful for describing rather than finding the objects. A very important contribution is the paper by Kent et al. (2000) where the model is used for modelling the standardized edge transformation vector, see also Kent et al. (1996). (The standardized edge transformation vector only contains shape information.) In particular, the eigendecomposition of the circulant covariance matrix is described. In the follow-up paper Hobolth et al. (1999) the corresponding theory is developed for the standardized vertex transformation vector. Likelihood analysis has also been considered in Hurn et al. (1999).

In Hobolth & Jensen (2000) a continuous approach is used, which may have a general appeal because the model and its parameters do not relate to a particular choice of the number of landmarks. Apart from that it appears natural to represent the boundary of an object continuously. The continuous counterpart of the standardized vertex transformation vector is the so-called normalized residual process, as introduced in Hobolth & Jensen (2000). Continuous models have also been mentioned in Hobolth et al. (1999) and Kent et al. (2000).

In the present paper we represent the shape of a random planar star-shaped object in terms of the normalized radius-vector function $R = (R(t))_{t\in[0,1]}$. A flexible continuous statistical model is proposed for R. Generally R need not be Gaussian. The main reason for choosing a specific representation of the object is that this allows us to analyse in detail the relationship between the model assumed for the normalized radius-vector function and the random geometry of the object.

Our approach rely on a polar Fourier expansion of the normalized radius-vector function

$$R(t) = 1 + 2\sqrt{c_1}\cos(2\pi(t-d_1)) + 2\sum_{s=2}^{\infty}\sqrt{C_s}\cos(2\pi s(t-D_s)), \quad t \in [0,1].$$

We show that the first phase amplitude c_1 and the first phase angle d_1 play a special role as parameters of asymmetry and discuss in detail how the remaining random phase amplitudes C_s and phase angles D_s influence the shape of the random object. The model proposed is called the generalized $p-$order model. Under this model the expected phase amplitudes $\lambda_s = E(C_s)$ satisfy the simple regression equation

$$\lambda_s^{-1} = \alpha + \beta(s^{2p} - 2^{2p}), \quad s \geq 2,$$

where $\alpha > 0$, $\beta > 0$ and $p > 1/2$. It will be shown that p determines the smoothness of the boundary of the object while the parameters α and β determine the 'global' and the 'local' shape, respectively.

The phase angles D_s are assumed to be uniformly distributed. There are, however, no restrictions on the distributions of the phase amplitudes C_s. Exponentially distributed amplitudes correspond to a Gaussian normalized radius-vector function. Generalized gamma distributed amplitudes offer a simple extension which allows for both heavier and lighter tails than the exponential ones.

In Section 2, the geometry of the radius-vector function is analysed in detail. This analysis is the basis for the construction of the generalized p-order model in Section 3. Its statistical inference is discussed in Section 4. The model is applied in Section 5 to a data set of normal mantle cell nuclei and cell nuclei from a mantle lymphoma. It turns out that the cell nuclei from the mantle lymphoma are more 'irregular' than the normals cells (significantly different β-values in the two groups). Shape discrimination of these two cell nuclei types was our original motivation for studying the continuous shape model. Section 6 contains some ideas for future research.

2 The geometry of the radius-vector function

Let K be a compact subset of $\mathbf{R}^2$. Let us suppose that K is star-shaped with respect to $z \in K$, i.e. the intersection between every line through z and K is a line segment.

We will describe K in terms of its radius-vector function $(r_K(t;z))_{t\in[0,1]}$ with respect to z, where

$$r_K(t;z) = \max\{r : z + r(\cos 2\pi t, \sin 2\pi t) \in K\}, \quad t \in [0,1].$$

The value $r_K(t;z)$ is the distance from z to the boundary of K along the ray, starting at z and with angle $2\pi t$ relative to a fixed axis. Because K is star-shaped we can reconstruct K from $r_K(\cdot;z)$.

The radius-vector function is well-known in the shape literature, cf. e.g. Stoyan & Stoyan (1994, p. 63), Lestrel (1997), Loncaric (1998) and references therein. It is also an important quantity in local stereology and geometric tomography, cf. Jensen (1998, Chapters 4 and 5) and Gardner (1995, Section 0.7). The radius-vector function is in geometric tomography called the radial function.

Using Hobolth & Jensen (2000, Proposition 1) it can be seen that the derivative (if it exists) of the radius-vector function contains interesting geometric information

$$r'_K(t;z) = 2\pi \cot(\varphi_K(t;z) - 2\pi t) r_K(t;z),$$

where $\varphi_K(t;z)$ is the angle that the tangent of the boundary point of K at position t makes with a fixed axis. The tangent-angle function $\varphi_K(\cdot;z)$ is therefore obtainable from the radius-vector function. The reverse statement is also true, but then the radius-vector function is only determined up to a multiplicative constant. The second derivative of $r_K(\cdot;z)$ involves the local curvature of K.

The area and boundary length of K can be expressed in terms of the radius-vector function

$$\begin{aligned} A(K) &= \pi \int_0^1 r_K(t;z)^2 dt \\ B(K) &= 2\pi \int_0^1 \{r_K(t;z)^2 + (2\pi)^{-2} r'_K(t;z)^2\}^{1/2} dt \\ &= 2\pi \int_0^1 r_K(t;z)/|\sin(\varphi_K(t;z) - 2\pi t)| dt. \end{aligned}$$

The formula for area holds without further assumptions, while the formula for boundary length holds under mild assumptions about K, including that the boundary of K is smooth, cf. Stoyan & Stoyan (1994, p. 64) and Jensen (1998, Proposition 5.4). It is also possible to express the total curvature of K in terms of $r_K(\cdot;z)$.

The definition of the radius-vector function can be extended to not-necessarily star-shaped sets K, cf. Gardner et al. (1995). The extended function is the radius-vector function of a star-shaped set associated with K, called the directed chordal symmetrical.

The radius-vector function is invariant under translation and rotation. To be more specific, let $z_0 \in \mathbf{R}^2$ and

$$A = \begin{pmatrix} \cos 2\pi t_0 & -\sin 2\pi t_0 \\ \sin 2\pi t_0 & \cos 2\pi t_0 \end{pmatrix}, \quad t_0 \in [0,1].$$

Then,

$$r_{AK+z_0}(t; Az + z_0) = r_K((t - t_0) \bmod 1; z), \quad t \in [0, 1].$$

A scaling transformation yields

$$r_{\alpha K}(t; \alpha z) = \alpha r_K(t; z), \quad t \in [0, 1], \quad \alpha > 0.$$

The shape of K is thus, up to shifts in t, represented by the normalized radius-vector function

$$\widetilde{r}_K(t; z) = \frac{r_K(t; z)}{\int_0^1 r_K(u; z) du}, \quad t \in [0, 1].$$

Note that the normalized radius-vector function is a continuous analogue of a standardized vertex transformation vector, cf. Hobolth et al. (1999). Below we will simply write $r(\cdot)$ for $\widetilde{r}_K(\cdot; z)$ in cases where it will cause no confusion.

A more detailed description of the shape of K can be obtained from a Fourier series expansion of the normalized radius-vector function r, cf. e.g. Stoyan & Stoyan (1994, p. 80-88) and Lestrel (1997),

$$r(t) = 1 + \sqrt{2} \sum_{s=1}^{\infty} a_s \cos(2\pi st) + \sqrt{2} \sum_{s=1}^{\infty} b_s \sin(2\pi st), \quad t \in [0, 1],$$

where the Fourier coefficients are

$$a_s = \sqrt{2} \int_0^1 r(t) \cos(2\pi st) dt, \quad b_s = \sqrt{2} \int_0^1 r(t) \sin(2\pi st) dt, \quad s \geq 1. \tag{2.1}$$

The Fourier coefficient at phase 0 is 1 because of the normalization of the radius-vector function. Letting

$$a_s = \sqrt{2c_s} \cos(2\pi s d_s), \quad b_s = \sqrt{2c_s} \sin(2\pi s d_s), \quad s \geq 1,$$

we obtain the polar form

$$r(t) = 1 + 2 \sum_{s=1}^{\infty} \sqrt{c_s} \cos(2\pi s(t - d_s)), \quad t \in [0, 1], \tag{2.2}$$

where $c_s = (a_s^2 + b_s^2)/2 \geq 0$ and $d_s \in [0, \frac{1}{s}[$, $s \geq 1$. (If $a_s = b_s = 0$, let $d_s = 0$.) The coefficient c_s is called the sth phase amplitude and d_s the sth phase angle. It is immediate from (2.2) that the c_ss are invariant under shifts in t.

Writing $z = (z_1, z_2)$, the boundary of K can be represented as

$$(f_1(t), f_2(t)) = (z_1, z_2) + qr(t) (\cos(2\pi t), \sin(2\pi t)), \quad t \in [0, 1],$$

where q is the integral of the radius-vector function. Combining this with (2.1) it follows that

$$(a_1, b_1) = \frac{\sqrt{2}}{q} \left(\int_0^1 [f_1(t) - z_1] dt, \int_0^1 [f_2(t) - z_2] dt \right).$$

C

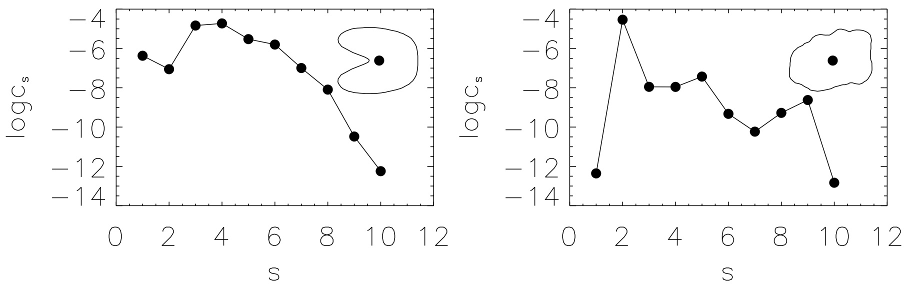

Figure 1: The values of the phase amplitudes c_s are shown as a function of s for an asymmetric object (left) and a fairly symmetric object (right).

Thus, if K is symmetric (with respect to z) then $a_1 = b_1 = c_1 = 0$. Conversely, a high value of c_1 indicates a high degree of asymmetry relative to z, cf. Figure 1. The left object in Figure 1 is an example of an object which is symmetric around the $x-$axis, but rather asymmetric around the $y-$axis. If the angle $2\pi t$ is measured relative to the $x-$axis this means that a_1 is rather large while $b_1 = 0$. In the Appendix we show that the Fourier coefficients a_1 and b_1 can also be expressed as integrals on the interior of K.

To analyse the geometry of the higher order phase amplitudes let us consider an object for which all but the sth phase amplitude are zero such that

$$r(t) = 1 + 2\sqrt{c_s}\cos(2\pi s(t - d_s)). \tag{2.3}$$

For such an object we have that z is the centre of gravity, cf. the Appendix. Moreover, $r(t)$ possesses an s-fold symmetry,

$$r(t) = r(t + \frac{1}{s}) = \cdots = r(t + \frac{s-1}{s}), \quad t \in [0, \frac{1}{s}[.$$

The 2-fold symmetry is the usual type of symmetry. In Figure 2 we have plotted objects with radius-vector function of the form (2.3), corresponding to different values of s, $d_s = 0$ and varying values of c_s. In Figure 3 we have illustrated how the $s-$fold symmetric objects contribute for small s to the 'global' shape of a given object K and for large s to the 'local' shape.

To sum up, we can interpret $c_s, s \geq 1$, as shape parameters. For $s = 1$, c_s is an asymmetry parameter. For $s \geq 2$ small, c_s determines the 'global' shape of K while for s large c_s affects the 'roughness' of the boundary of K. Up to a shift in t, d_s, $s \geq 1$, are also shape parameters. For $s \geq 2$, they determine the relative orientation of the $s-$fold symmetric objects associated with K.

In Zahn & Roskies (1972) the geometric interpretation of a Fourier expansion of the tangent-angle function is studied in a similar way.

Let us conclude this section by discussing how z can be chosen. In some applications z is 'given by nature'. An important example comes from local stereology where K is actually a planar section through a biological cell, passing through the nucleus or nucleolus of the cell, cf. Jensen (1998, Chapter 7). In other cases z is defined from

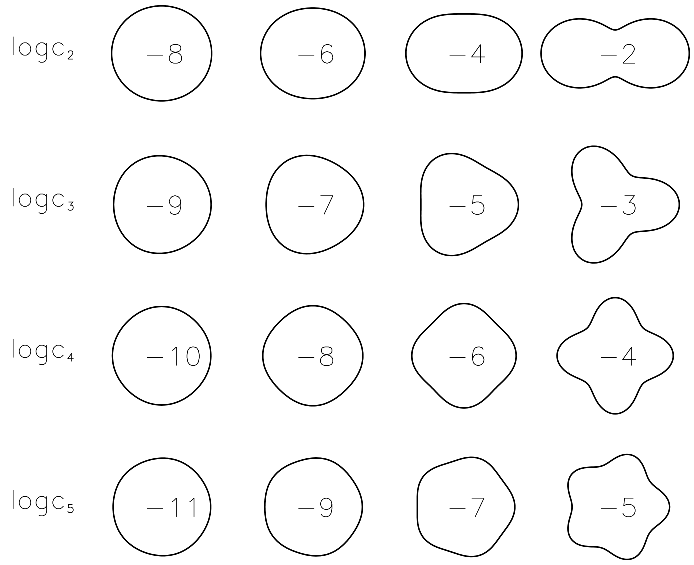

Figure 2: Objects with radius-vector function of the form (2.3) with $d_s = 0$. In each row the value of s is constant ($s = 2, 3, 4, 5$). The value of $\log c_s$ is indicated in the interior of the object.

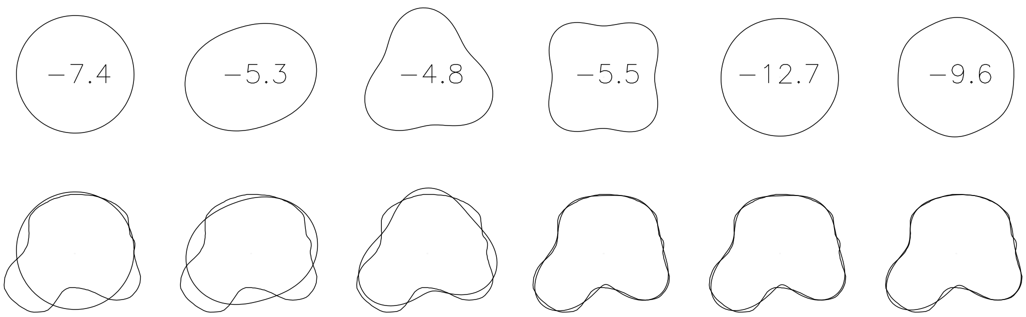

Figure 3: The upper row shows the objects of the form (2.3) for $s = 1, \ldots, 6$ (left to right) associated with the object K shown in the lower row. The values of $\log c_s$ are indicated in their interior. In the lower row, the reconstruction of K from the first s Fourier coefficients is also shown.

K, typically as the centre of mass, cf. Loncaric (1998) and Hobolth et al. (1999). In the latter paper it is used that with z equal to the centre of gravity the first phase amplitude of $r_K(\cdot; z)$ is approximately zero when K is a small deformation of a circle. In the Appendix it is shown that the centre of mass of K can in fact be characterized by the property that the first phase amplitude of $r_K(\cdot; z)^3$ is zero.

C

3 The generalized p-order model

We will now consider a random planar object K with normalized radius-vector function $(R(t))_{t\in[0,1]}$. Below we introduce the parametric statistical model to be used for R.

The starting point is the polar expansion (2.2) of the normalized radius-vector function. As argued in the previous section the first phase angle d_1 and phase amplitude c_1 play a special role as asymmetry parameters. We shall treat c_1 and d_1 as non-random nuisance parameters. The expansion of the normalized radius-vector function in polar form therefore becomes

$$R(t) = 1 + 2\sqrt{c_1}\cos(2\pi(t-d_1)) + 2\sum_{s=2}^{\infty}\sqrt{C_s}\cos(2\pi s(t-D_s)),\ t\in[0,1]. \qquad (3.1)$$

The remaining amplitudes C_s and angles D_s, $s \geq 2$, will be modelled by distributions on $\mathbf{R}_+$ and $[0,1/s]$, respectively.

The expansion (3.1) makes it possible to construct a variety of shape models. A *generalized $p-$order model* is a parametric model which satisfies that

$$C_s \sim \lambda_s Z_s, \quad D_s \sim U[0,1/s], \quad s \geq 2, \qquad (3.2)$$

where the error variables Z_s have mean 1 and $U[0,1/s]$ is the uniform distribution on the indicated interval. Furthermore, $C_s, D_s, s \geq 2$, are all independent and the expected phase amplitudes $\lambda_s = E(C_s)$ decrease as

$$\lambda_s^{-1} = \alpha_0 + \beta s^{2p}, \quad s \geq 2.$$

The parameters satisfy $\alpha_0 > -\beta 2^{2p}$ and $\beta > 0$ such that $\lambda_s > 0$ for all $s \geq 2$. We further assume $p > 1/2$, which implies that R has finite variance, as will be discussed below.

In order to facilitate a geometric interpretation of the regression parameters we use the reparametrization

$$\lambda_s^{-1} = \alpha + \beta(s^{2p} - 2^{2p}), \quad s \geq 2, \qquad (3.3)$$

where $\alpha > 0, \beta > 0, p > 1/2$. The parameter α determines the 'global' shape of the object. If α is high objects of circular shape are expected while a low value corresponds to an elongated or, in the extreme, a 'peanut-shell' shape. The reason is that under (3.3), α determines the expected phase amplitudes $\lambda_s = E(C_s)$ for small s and C_s governs the global shape for small s, cf. Section 2. As discussed below p determines the smoothness of the boundary of K. For fixed p the parameter β determines the 'local' shape of the object since it controls the behaviour of λ_s when s is high. Precisely, as $s \to \infty$, we have that $(\log s, \log \lambda_s^{-1})$ behaves as a line with slope $2p$ and intercept $\log \beta$. For small values of β rather irregular objects are expected while high values yield regular objects.

The random phase angles D_s determine the relative orientation of the $s-$fold symmetric objects associated with K, cf. Section 2. The uniform distribution on the angles implies that these objects do not have a 'preferred orientation'. A generalized $p-$order model is therefore expected to be appropriate for describing a population of objects which does not have a predominant non-circular shape. The shape variability of K is influenced by the variation of the error variables Z_s.

In Figure 4 simulations from the model (3.1)-(3.3) with exponentially distributed error variables, $p = 2$ and $c_1 = 0$ are shown. The values of α and β are typical for the objects studied in the data section. It is seen that in the corner corresponding to high values of α and β the simulated objects are smooth and 'circle'-like, while in the opposite corner the simulated objects are irregular.

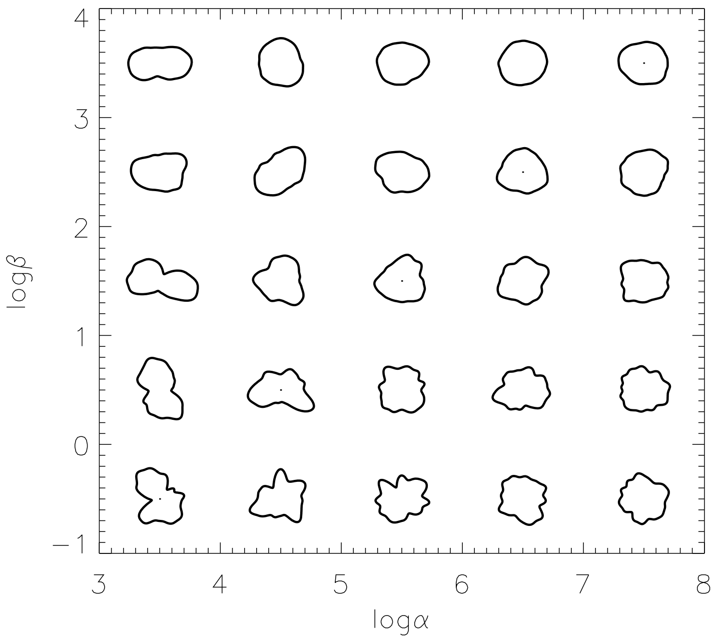

Figure 4: Simulated objects under the second-order model with $c_1 = 0$, exponentially distributed error variables and the indicated values of α and β.

To study the distribution of the radius-vector function let

$$R_1(t) = 2\sum_{s=2}^{\infty} \sqrt{C_s}\cos(2\pi s(t - D_s)), \quad t \in [0, 1],$$

contain all the random Fourier terms of $R(t)$. Using (3.2) and independence of the phase angles and amplitudes it follows that R_1 is a stationary process with covariance function

$$\sigma(t) = \operatorname{cov}(R_1(t), R_1(0)) = 2\sum_{s=2}^{\infty} \lambda_s \cos(2\pi st), \quad t \in [0, 1]. \tag{3.4}$$

The process R_1 has zero Fourier coefficients at phases 0 and 1. Similar constraints were used by Hobolth et al. (1999) and Kent et al. (2000) in a discrete time model.

Properties such as continuity and differentiability of R_1 (and hence also of R) are determined by the parameter p as follows from Cramér & Leadbetter (1967, Section 4.2 and 4.3).

Equation (3.4) gives the relation between the expected amplitudes and the covariance function. As an alternative to parametric specification of the λ_ss as in (3.3) one may suggest a simple parametric form of the covariance function σ, cf. e.g. Rue & Syversveen (1998). Since the amplitudes relate to the random geometry of the object we believe it is more natural to specify directly a parametric model for the expected amplitudes. Furthermore, the constraints on R_1 are easier to handle and interpret in the spectral domain.

In the shape literature a random object is often modelled by a multivariate normal distribution with a circulant covariance matrix or by a stationary Gaussian process in continuous time, cf. Grenander & Miller (1994), Hobolth et al. (1999), Rue & Hurn (1999), Hobolth & Jensen (2000), Kent et al. (2000). We now show that a Gaussian model is obtained by letting the error variables Z_s be exponentially distributed. This model will therefore be called the normal $p-$order model. Using (2.2) and (3.2) it follows that if Z_s is exponentially distributed then

$$R_1(t) = \sqrt{2}\sum_{s=2}^{\infty} A_s \cos(2\pi st) + \sqrt{2}\sum_{s=2}^{\infty} B_s \sin(2\pi st), \quad t \in [0,1],$$

where $A_s, B_s, s \geq 2$, are all mutually independent and $A_s \sim B_s \sim N(0, \lambda_s)$. This representation shows that R_1 is a stationary Gaussian process. By (3.3) and Rogers & Williams (1994, Theorem I.25.10), it follows that for the normal $p-$order model the sample paths of R_1, and hence also of R, are k times continuously differentiable where k is the integer satisfying $p \in]k - 1/2, k + 1/2]$. In particular, if p is an integer then $p = k$. In the normal first-order model, the sample paths of R are continuous while in the normal second-order model the sample paths are continuously differentiable.

The first- and second-order normal models have been studied in the literature (most often without the constraint $\lambda_0 = \lambda_1 = 0$). In particular, these models appear as limits of discrete time first- and second-order Markov models, cf. e.g. Grenander (1993, p. 476 and 484).

4 Some remarks on statistical inference

Maximum likelihood estimation of the parameters of the normal $p-$order model, based on a continuously observed normalized radius-vector function, has been discussed in an unpublished research report by two of us (A. Hobolth and J. Pedersen) from Laboratory for Computational Stochastics, University of Aarhus, 1999. (In this report, we also suggest the regression equation (3.3), which was later used in Hobolth et al. (1999).) For the first- and second-order normal models it is shown that with continuous observations β is determined with certainty from observable quantities. Thus, only α has to be estimated and a closed-form equation for the maximum likelihood estimate of α is given in the research report.

Although the likelihood-based estimates have several good properties in theory, they may perform poorly in practice. Thus, as expected from the interpretation of the parameter β as a local shape parameter, β is determined from differences between observations very close together at the boundary of the object under study. If only a digitized version of the object is available such differences cannot be determined accurately.

To avoid this obstacle one can use a so-called low-pass filter, cf. e.g. Bloomfield (1976). The idea is to determine the parameter estimates from the low frequency Fourier coefficients only which are robust against digitization effects. For the normal $p-$order model, the analysis is particularly simple. Recall that in this case the phase angles D_s are uniform in $[0, 1/s]$ and the phase amplitudes C_s are exponentially distributed with mean λ_s. Note in particular that the distribution of the phase angles does not depend on unknown parameters. Using the first S phase amplitudes the likelihood function becomes

$$L(\lambda_s; c_s) = \prod_{s=2}^{S} \lambda_s^{-1} e^{-\lambda_s^{-1} c_s}. \tag{4.1}$$

Defining the expected amplitudes by (3.3) the maximum likelihood estimates for (α, β, p) can be found by standard numerical methods. A likelihood function of the same form has been considered in Hobolth et al. (1999) and Kent et al. (2000).

If the normalized radius-vector function is only known at the data points $t = 0/n, 1/n, \ldots, (n-1)/n$, the phase amplitudes $c_s = (a_s^2 + b_s^2)/2$ can be approximated by using discretized versions of the integrals (2.1). The specific value of n is not important, just as long as it is reasonably high. That is, different values of n give approximately the same value of c_s.

5 Data analysis

The data set consists of 50 normal mantle cell nuclei and 50 cell nuclei from a mantle lymphoma (tumour in the mantle zone of a lymph node), cf. Figure 5. The nuclei from each of the groups were sampled from a microscopic section among those with sectioned boundary in focus. The normalized radius-vector function $r(t)$ with respect to the centre of mass was for each nucleus determined at $t = 0, 1/n, \ldots, (n-1)/n$. Unless otherwise stated we used $n = 100$. The nuclei are rather homogeneous in size (about $15\mu m$ in diameter), so the normalization factor was almost the same for all the nuclei.

5.1 Analysing each nuclear profile individually

First, each nuclear profile was analysed individually using the likelihood function (4.1). The choice of cut-off value S is important. If S is too small we are not using important shape information; if on the other hand S is too large the results will be influenced by digitization effects, see also Figure 3. Unless otherwise stated we used $S = 15$.

C

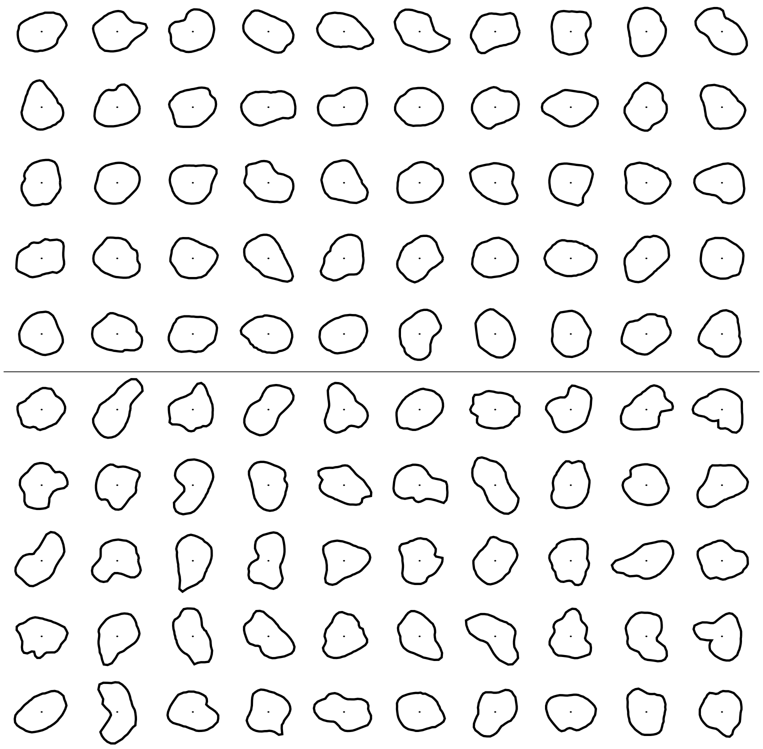

Figure 5: The 50 normal mantle cell nuclei (upper panel) and the 50 cell nuclei from a mantle lymphoma (lower panel).

For each object we found the estimates of (α, β, p). In both samples the estimates of p were close to 2 for all nuclei. For the normal sample the average was 2.07 with a standard deviation of 0.21 while for the lymphoma sample the average was 2.02 and the standard deviation 0.28. Therefore we fixed $p = 2$ and considered the normal second-order model only.

	$\log \hat{\alpha}$		$\log \hat{\beta}$		
	av.	s.d.	av.	s.d.	corr.
normal	5.35	0.84	2.26	0.72	0.27
lymphoma	4.94	1.11	1.09	0.81	0.03

Table 1: The average, standard deviation and correlation of $(\log \hat{\alpha}, \log \hat{\beta})$ for each sample.

The estimates of (α, β) under the second-order model are shown for each nucleus in Figure 6 and summarized in Table 1. The estimates of the local shape parameter β are on average lowest in the lymphoma sample. This was to be expected from the geometric interpretation of β given in Section 3. A t-test for identical βs, based on the distribution of $\log \hat{\beta}$, shows a significant difference between the two samples (p-value less than 0.05%). On average the estimates of the global shape parameter α are also lowest in the lymphoma sample, but the difference is not as significant (p-value close

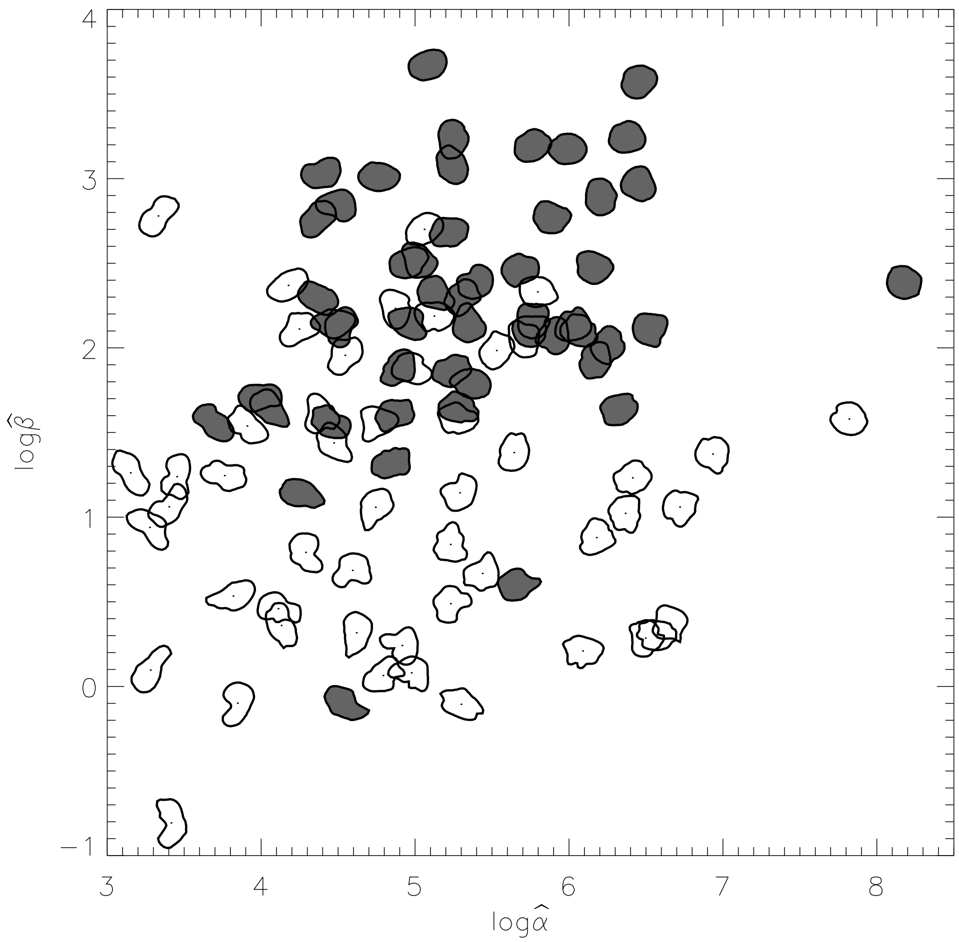

Figure 6: The estimates of (α, β) under the normal second-order model. The hatched nuclei are from the normal mantle cells while the white nuclei are from cells in the mantle lymphoma.

to 5%). Furthermore we see that the estimates of α from the lymphoma sample vary over a somewhat larger range than the estimates from the normal sample.

We also investigated how the choice of cut-off value S influences the analysis. Since the estimate of α is determined by the first few amplitudes the estimate of this parameter only changes slightly when S is larger than 8, say. From Figure 7 it is seen that the estimate of β does change with S, but the changes are rather small.

The number of data points n should be high compared to S, but otherwise the specific choice is less important. In Figure 8 we see that for $S = 15$ the estimates are stable, and the analysis is robust to the specific choice of $n \geq 50$.

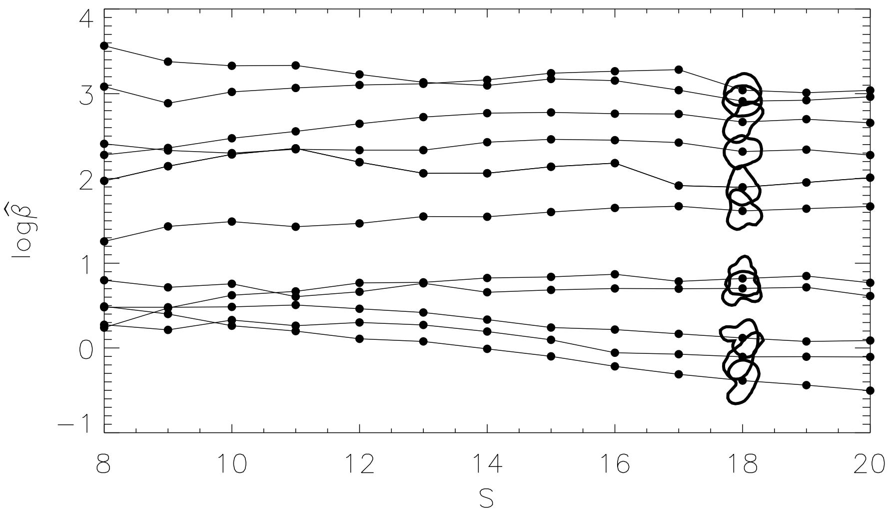

Figure 7: The estimates of β as a function of S for the 11 nuclei shown.

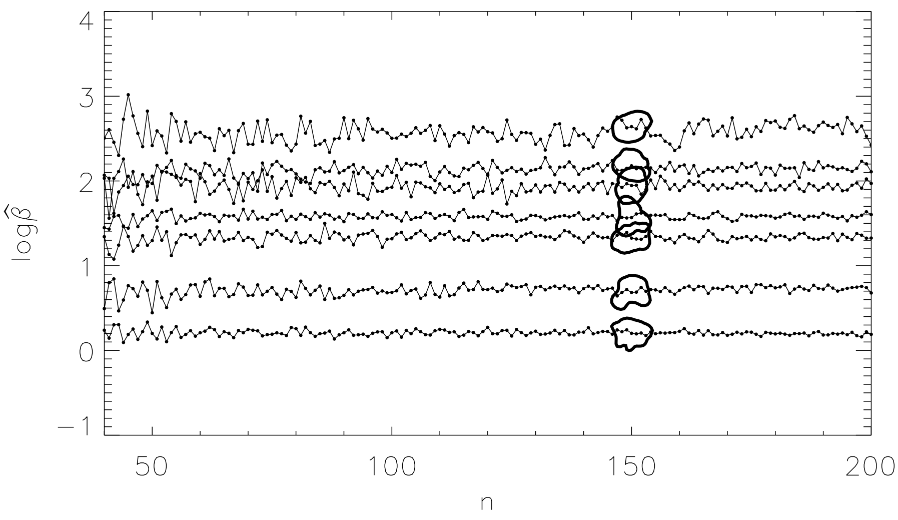

Figure 8: The estimates of β as a function of n for the 7 nuclei shown.

5.2 Analysing the profiles under an iid-assumption

We now investigate whether the profiles within each of the groups can be regarded as independent and identically distributed realizations from a normal $p-$order model. Let the indices (i, j) denote the jth nucleus ($j = 1, \ldots, N = 50$) in the normal sample ($i = 1$) or the lymphoma sample ($i = 2$) and let c_{sij} be the corresponding phase amplitudes of the normalized radius-vector function. If we let $\mathrm{Exp}(\lambda)$ be the notation used for the exponential distribution with mean λ, we then want to investigate whether

$$C_{sij} \sim \text{Exp}(\lambda_{si}), \quad j = 1, \ldots, N, \tag{5.1}$$

for each $s = 2, \ldots, S$ and $i = 1, 2$.

We will examine (5.1) by considering the more general model

$$C_{sij} \sim \Gamma(\gamma_{si}, \rho_{si}, \delta_{si}), \quad j = 1, \ldots, N,$$

where $\Gamma(\gamma, \rho, \delta)$ is the notation used for the generalized gamma distribution with density

$$f(y) = \frac{\delta y^{\delta\gamma - 1}}{\Gamma(\gamma)\rho^{\delta\gamma}} \exp\{-(\frac{y}{\rho})^{\delta}\}, \quad y > 0.$$

Here, $\gamma, \delta > 0$ are shape parameters while $\rho > 0$ is a scale parameter. The ordinary gamma distribution is obtained for $\delta = 1$, the Weibull distribution for $\gamma = 1$, while the exponential distribution corresponds to $\delta = \gamma = 1$.

The class of generalized gamma distributions is in fact rather flexible. When $\delta < 1$ (> 1) the tails are heavier (lighter) than the exponential tails. When $\delta\gamma \leq 1$ the density $f(y)$ is strictly decreasing in y. Moreover $\lim_{y \to 0} f(y)$ exists and is finite if and only if $\delta\gamma \geq 1$. When $\delta\gamma > 1$ the density has a mode.

Plots of the empirical survival functions of c_{sij} for fixed s and i showed that the distributions of the phase amplitudes had somewhat heavier tails than expected under (5.1) (the estimated values of δ were less than 1). In each sample the tendency was only significant for a few high values of s. Thus it seems reasonable to consider exponentially distributed error variables, at least for low frequencies. The same conclusion was obtained by testing (5.1) by Bartlett tests.

Assuming that the phase amplitudes c_{sij} are $\text{Exp}(\lambda_{si})$-distributed, the next step in the analysis is to fit a $p-$order model within each group,

$$\lambda_{si}^{-1} = \alpha_i + \beta_i \left(s^{2p_i} - 2^{2p_i}\right), \quad s = 2, \ldots, S, \quad i = 1, 2. \tag{5.2}$$

The likelihood function is given by

$$L(\lambda_{si}; c_{sij}) = \prod_{j=1}^{N} \prod_{s=2}^{S} \lambda_{si}^{-1} \exp(-\lambda_{si}^{-1} c_{sij}) = \{\prod_{s=2}^{S} \lambda_{si}^{-1} \exp(-\lambda_{si}^{-1} \bar{c}_{si\cdot})\}^{N},$$

where $\bar{c}_{si\cdot} = N^{-1} \sum_{j=1}^{N} c_{sij}$ is the average of the amplitudes within the ith group at phase s and λ_{si} is given by (5.2). As expected the estimated value of p is close to 2 in both samples (2.0 in the normal and 1.8 in the lymphoma sample), and again we consider the second-order model. The estimated regression lines are shown in Figure 9 and in Table 2 the estimates and approximate standard errors and correlation coefficients based on the observed information are summarized.

As in the previous subsection we observe a significant difference between the two samples in the value of β. The difference in α is not as significant.

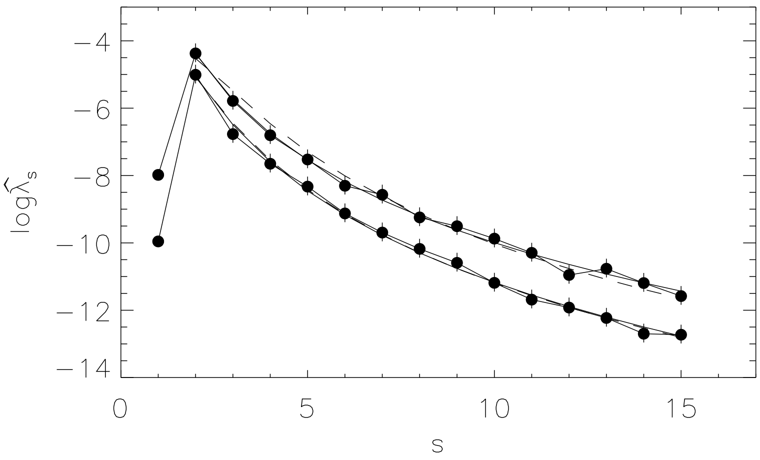

Figure 9: The estimated regression $\hat{\lambda}_s = [\hat{\alpha} + \hat{\beta}(s^{2\hat{p}} - 2^{2\hat{p}})]^{-1}$ in the normal p-order model (solid) and the estimated regression under the normal second-order model (dashed) is shown together with the average phase amplitudes as a function of s for the normal sample (lower curve) and the lymphoma sample (upper curve). The vertical lines are the 95% confidence limits.

	$\log\hat{\alpha}$			$\log\hat{\beta}$			
	est.	conf.int.	s.e.	est.	conf.int.	s.e.	corr
normal	5.08	4.81-5.35	0.14	1.97	1.89-2.05	0.04	-0.10
lymphoma	4.52	4.27-4.77	0.13	0.82	0.74-0.90	0.04	-0.13

Table 2: The estimates and approximate confidence intervals, standard errors and correlation of $(\log\hat{\alpha}, \log\hat{\beta})$.

5.3 Simulations from the normal second-order model

In the normal second-order model truncated at $S = 15$ we have

$$C_s \sim \mathrm{Exp}(\lambda_s), \quad s = 2, \dots, 15, \quad \text{independent},$$

with

$$\lambda_s^{-1} = \alpha + \beta(s^4 - 2^4). \tag{5.3}$$

In order to investigate the model more closely we made the following simulation study. For each sample we calculated λ_s according to (5.3) with (α, β) replaced by the average estimated value from Table 1 and simulated $C_s \sim \mathrm{Exp}(\lambda_s)$, $s = 2, \dots, 15$. From the values of C_s we calculated the maximum likelihood estimates of α and β. This procedure was repeated 500 times for each of the samples and the results are shown in Figure 10 and summarized in Table 3.

When we compare Figures 6 and 10 it is seen that the variation in $\log\hat{\alpha}$ is almost the same for the observed and simulated data for both groups. The variation range

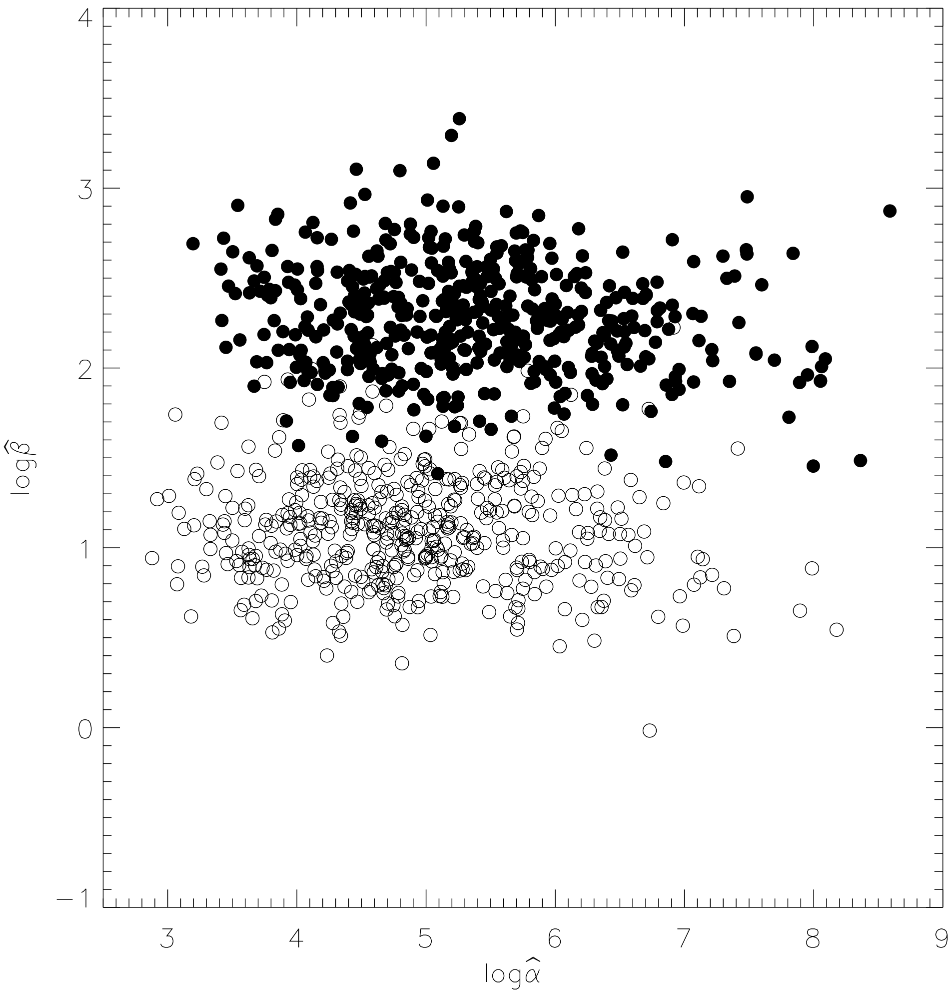

Figure 10: Simulated distribution of $(\hat{\alpha}, \hat{\beta})$ under the normal second-order model is shown for the normal sample ($\bullet$) and the lymphoma sample ($\circ$).

	$\log\hat{\alpha}$		$\log\hat{\beta}$		
	av.	s.d.	av.	s.d.	corr.
normal	5.36	1.01	2.28	0.30	-0.14
lymphoma	4.89	0.95	1.10	0.30	0.03

Table 3: The average, standard deviation and correlation of $(\log\hat{\alpha}, \log\hat{\beta})$ for each sample.

of $\log\hat{\beta}$ is smaller in the simulation study than in the samples. One explanation is that the local shape variability in the data is somewhat higher than predicted from the normal model, i.e. the assumption that the error variables are exponential is not appropriate at high phases. Another reason might be that a well located ‘blob’ results in many high phase amplitudes.

C

6 Perspectives

The generalized p-order model is expected to be useful for describing a population which does not have a predominant non-circular shape. Let us stress, though, that the Fourier expansion (3.1) of the normalized radius-vector function makes it possible to construct a variety of shape models. If, for instance, one considers a population with a dominant triangular shape, it would be natural to use a model where on average C_3 is the highest amplitude. A more challenging task is to model objects with a dominant elliptical shape. An ellipse has vanishing amplitudes at odd phases and decreasing amplitudes at even phases. Thus, to model elliptical shape one should probably let the odd and even amplitudes decrease at different rates. Moreover, the even phase angles should have approximately the same values. Elliptical models were studied by Hobolth & Jensen (2000). We leave a concrete model as a topic for future research.

Acknowledgements

The authors want to thank Professor Flemming Sørensen, Institute of Pathology, University of Aarhus, for kindly providing the data. This work was supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation. The second author is supported by a grant from the Danish Natural Science Research Council.

Appendix

Characterization of asymmetry and centre of mass

Let $x = (x_1, x_2)$ denote a generic point in $\mathbf{R}^2$ and let $||x|| = (x_1^2 + x_2^2)^{\frac{1}{2}}$.

Proposition *Let $z = (z_1, z_2)$ be an interior point of a compact subset K of $\mathbf{R}^2$. Let K be star-shaped with respect to z and let the radius-vector function $r_K(t; z)$ be continuously differentiable in t.*

(i) *We have*

$$\int_K \frac{x_1 - z_1}{||x - z||^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t; z) \cos(2\pi t) dt \tag{A.1}$$

$$\int_K \frac{x_2 - z_2}{||x - z||^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t; z) \sin(2\pi t) dt. \tag{A.2}$$

(ii) *If z is the centre of mass of K then*

$$\int_0^1 r_K(t; z)^3 \cos(2\pi t) dt = \int_0^1 r_K(t; z)^3 \sin(2\pi t) dt = 0. \tag{A.3}$$

Conversely, if z is such that (A.3) *is satisfied then z is the centre of mass of K.*

(iii) *Let* $r_K(t;z) = 1 + 2\sqrt{c_s}\cos(2\pi s(t - d_s))$, *where* $s \geq 2, 0 \leq c_s \leq 1/4$ *and* $d_s \in [0, \frac{1}{s}[$. *Then* z *is the centre of mass of* K.

Proof. Let $F : [0,1]^2 \to \mathbf{R}^2$ be defined by

$$F(v,t) = (z_1, z_2) + v r_K(t;z)(\cos(2\pi t), \sin(2\pi t))\,.$$

Then F is onto K and $\left|\det\left(F'(v,t)\right)\right| = 2\pi v r_K(t;z)^2$. In order to prove (A.1) note that if $x = (x_1, x_2) \in K$ is such that $x = F(v,t)$ then

$$\frac{x_1 - z_1}{||x - z||^2} = \frac{\cos(2\pi t)}{v r_K(t;z)},$$

and from the transformation theorem we get

$$\int_K \frac{x_1 - z_1}{||x - z||^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t;z)\cos(2\pi t) dt.$$

The result (A.2) is proved similarly.

The same kind of arguments show that

$$\begin{aligned} &\left(\int_K (x_1 - z_1) dx_1 dx_2, \int_K (x_2 - z_2) dx_1 dx_2\right) \\ &= \frac{2\pi}{3}\left(\int_0^1 r_K(t;z)^3 \cos(2\pi t) dt, \int_0^1 r_K(t;z)^3 \sin(2\pi t) dt\right). \end{aligned} \tag{A.4}$$

The left-hand side is zero if and only if z is the centre of mass of K. Therefore (ii) is an immediate consequence of (A.4).

To prove (iii) one has to show that $r_K(t;z) = 1 + 2\sqrt{c_s}\cos(2\pi s(t - d_s))$ satisfies the condition (A.3). This follows from elementary calculations, and is left to the reader.

References

Bloomfield, P. (1976). *Fourier Analysis of Time Series: an Introduction.* New York: Wiley.

Cramér, H. & Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes.* New York: Wiley.

Gardner, R. (1995). *Geometric Tomography.* New York: Cambridge University Press.

Grenander, U. (1993). *General Pattern Theory.* Oxford: Oxford University Press.

Grenander, U. & Miller, M.I. (1994). Representations of knowledge in complex systems (with discussion). *J. R. Statist. Soc.* B **56**, 549-603.

Hansen, M.B., Møller, J. & Tøgersen, F.Aa. (2000). Bayesian contour detection in a time series of ultra-sound images through dynamic deformable template models. *Research Report* **17**, Centre for Mathematical Physics and Stochastics, University of Aarhus.

Hobolth, A. & Jensen, E.B.V. (2000). Modelling stochastic changes in curve shape, with an application to cancer diagnostics. *Adv. Appl. Prob. (SGSA)* **32**, 344-362.

Hobolth, A., Kent, J.T. & Dryden, I.L. (1999). On the relationship between edge and vertex modelling. *Research Report* **7**, Laboratory for Computational Stochastics, University of Aarhus.

Hurn, M.A., Steinsland, I. & Rue, H. (1999). Parameter estimation for a deformable template model. *Preprint Statistics* **5/1999**, Norwegian University of Science and Technology, Trondheim, Norway.

Jensen, E.B.V. (1998). *Local Stereology.* Singapore: World Scientific.

Kent, J.T., Mardia, K.V. & Walder, A.N. (1996). Conditional cyclic Markov random fields. *Adv. Appl. Prob.* **28**, 1-12.

Kent, J.T., Dryden, I.L. & Anderson, C.R. (2000). Using circulant symmetry to model featureless objects. *Biometrika* **87**, 527-544.

Lestrel, P.E. (ed.) (1997). *Fourier Descriptors and their Applications in Biology.* New York: Cambridge University Press.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognition* **31**, 983-1001.

Rogers, L.C.G. & Williams, D. (1994). *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations.* 2nd Edition. Chichester: Wiley.

Rue, H. & Hurn, M.A. (1999). Bayesian object recognition. *Biometrika* **86**, 649-660.

Rue, H. & Syversveen, A.R. (1998). Bayesian object recognition with Baddeley's delta loss. *Adv. Appl. Prob.* **30**, 64-84.

Stoyan, D. & Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics.* Chichester: Wiley.

Zahn, C.T. & Roskies, R.Z. (1972). Fourier descriptors for plane closed curves. *IEEE Trans. Computers* **C-21**, 269-281.

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2001).
A deformable template model,
with special reference to elliptical templates.
Research report no. 14, Laboratory for Computational
Stochastics, University of Aarhus.

D

A deformable template model, with special reference to elliptical templates

Asger Hobolth, Jan Pedersen and Eva B. Vedel Jensen
Laboratory for Computational Stochastics
University of Aarhus

Abstract

This paper suggests a high-level continuous image model for planar star-shaped objects. Under this model, a planar object is a stochastic deformation of a star-shaped template. The residual process, describing the difference between the radius-vector function of the template and the object, is allowed to be non-stationary. Stationarity is obtained by a time change. A parametric model for the residual process is suggested and straightforward parameter estimation techniques are developed. The deformable template model makes it possible to detect pathologies as demonstrated by an analysis of a data set of cell nuclei from a benign and a malignant tumour, using stochastic deformations of ellipses.

Keywords: Deformable template model, ellipse, Fourier analysis, non-Gaussian errors, non-stationarity, shape, time change.

1 Introduction

In high-level image modelling, the objects of an image are modelled directly. A very powerful approach is the deformable template model suggested by Ulf Grenander and the group around him, cf. e.g. [3, 4, 5]. The basic idea is to model the observed object as a stochastic deformation of a template, and the challenging task is to model the deformation mechanism.

Deformable template models for featureless objects have attracted a lot of attention in the statistical literature recently, cf. e.g. [4, 5, 7, 8, 11, 15, 16], and the focus has mainly been on circular templates. In the present paper we suggest a deformable template model for a random star-shaped planar object K which is useful in the case of non-circular templates. The radius-vector function $R = \{R(t)\}_{t\in[0,1]}$ of K is modelled as $R(t) = r(t) + X(t)$ where $r(t)$ is the deterministic radius-vector function of the template and $X(t)$ is a random residual process. For non-circular templates it is not natural to assume that $X(t)$ is stationary, as will be demonstrated in a simulation study. We therefore introduce a time change $\gamma(t)$ such that $X_0(t) = X(\gamma^{-1}(t))$ is stationary. This is a generalization of the approach described in [17], p. 90.

Modelling of $X_0(t)$ is based on a Fourier expansion. For elliptical templates it is assumed that the Fourier coefficients of $X_0(t)$ at the phases $s = 0, 1, 2$ are small. The remaining Fourier coefficients are modelled as normal variables with mean zero and variance λ_s at phase s given by the regression equation

$$\lambda_s^{-1} = \alpha + \beta\left(s^{2p} - 3^{2p}\right), \quad s \geq 3,$$

where α, β, p are unknown parameters. We discuss how the parameters influence the random geometry of the object and consider various choices of time changes. In [6] elliptical templates were also studied, but the approach deviates significantly from ours as will be discussed in Sections 3 and 4. The papers [7, 8, 11, 18], are based on Fourier expansions of either the tangent-angle function or the radius-vector function. The statistical models proposed in these papers describe a circular rather than an elliptical shape.

In Section 2 we define the general model. Various distributional results will be provided when X_0 is Gaussian and extensions to the non-Gaussian case will be discussed. In Section 3 we specialize to elliptical templates. The suggested model is used in the analysis of a data set concerning cancer diagnostics in Section 4. We conclude with some perspectives concerning Bayesian object recognition.

2 A deformable template model

Let a random planar object K be star-shaped relative to $z \in K$ such that K is determined by the radius-vector function $R = \{R(t)\}_{t\in[0,1]}$ with respect to z, where

$$R(t) = \max\{u : z + u(\cos 2\pi t, \sin 2\pi t) \in K\}, \quad t \in [0,1].$$

For a detailed description of the radius-vector function, see [8] and references therein. We suppose that the radius-vector function of K is on the form

$$R(t) = r(t) + X(t), \quad t \in [0,1], \tag{2.1}$$

where $r = \{r(t)\}_{t\in[0,1]}$ is the radius-vector function of the template and $X = \{X(t)\}_{t\in[0,1]}$ is a residual process which is periodic and has mean zero. Furthermore, we assume that there exists an increasing transformation γ of $[0,1]$ onto $[0,1]$ such that $\{X(\gamma^{-1}(t))\}_{t\in[0,1]}$ is stationary. In particular, the correlation between $X(\gamma^{-1}(t_1))$ and $X(\gamma^{-1}(t_2))$ depends on $t_2 - t_1$ only. We say that X is $\gamma-$stationary. An obvious choice of $\gamma(t)$ is the distance travelled on the boundary of the template between the points with index 0 and t, $t \in [0,1]$. With this choice of γ, the correlation between $X(t_1)$ and $X(t_2)$ only depends on the distance along the template between the points indexed by t_1 and t_2.

In the stochastic process literature γ is referred to as a time change, cf. e.g. [14], and we will use the same terminology here.

We can rewrite (2.1) as

$$R_0(t) = r_0(t) + X_0(t), \quad t \in [0,1],$$

where $R_0(t) = R(\gamma^{-1}(t))$ and similarly for the other quantities. Note that $X_0 = \{X_0(t)\}_{t\in[0,1]}$ is stationary in the ordinary sense. Let us suppose that the residual process is Gaussian. Let

$$r_0(t) = a_0 + \sqrt{2}\sum_{s=1}^{\infty} a_s \cos(2\pi st) + \sqrt{2}\sum_{s=1}^{\infty} b_s \sin(2\pi st)$$

D

and

$$X_0(t) = A_0 + \sqrt{2}\sum_{s=1}^{\infty} A_s \cos(2\pi st) + \sqrt{2}\sum_{s=1}^{\infty} B_s \sin(2\pi st) \tag{2.2}$$

be the Fourier expansions of r_0 and X_0. Since X_0 is Gaussian, A_0 and $A_s, B_s, s \geq 1$, are all mutually independent, $A_0 \sim N(0, \lambda_0)$ and $A_s \sim B_s \sim N(0, \lambda_s), s \geq 1$. It follows that the Fourier expansion of R_0 has the same distributional properties as those of X_0, except that zero mean-values are substituted by the relevant Fourier coefficients from the template.

For a polar Fourier expansion

$$R_0(t) = \sqrt{C_0} + 2\sum_{s=1}^{\infty} \sqrt{C_s}\cos(2\pi s(t - D_s)), \quad t \in [0,1],$$

where $C_s \geq 0$ and $D_s \in [0, 1/s)$, we have under the Gaussian assumption that C_0 and $(C_s, D_s), s \geq 1$, are all independent. Furthermore, the observed phase amplitude

$$C_s = \begin{cases} (a_0 + A_0)^2 & s = 0 \\ ((a_s + A_s)^2 + (b_s + B_s)^2)/2 & s \geq 1, \end{cases}$$

follows a non-central χ^2-distribution with mean

$$EC_s = c_s + \lambda_s, \quad s \geq 0,$$

where c_s is the sth phase amplitude of the template. Finally, the conditional distribution of D_s given C_s is given by, cf. the Appendix,

$$2\pi s D_s \mid C_s = c \sim vM(2\pi s d_s, 2\frac{\sqrt{cc_s}}{\lambda_s}), \quad s \geq 1, \tag{2.3}$$

where $d_s \in [0, 1/s)$ is the sth phase angle of the template. Here, $vM(\mu, \kappa)$ is the notation used for the von Mises distribution with mean direction $\mu \in [0, 2\pi)$ and concentration parameter $\kappa \geq 0$. For $\kappa = 0$ we get the uniform distribution on $[0, 2\pi)$ while for $\kappa > 0$ large the distribution is concentrated around the mean direction. Other properties of this distribution are described in [12], p. 36.

If the template is a circle, then $a_s = b_s = 0 = c_s, s \geq 1$. Therefore, in this case, C_s and D_s are independent, C_s follows an exponential distribution with mean λ_s and D_s is uniformly distributed on $[0, 1/s), s \geq 1$.

The distribution of C_s can for $s \geq 1$ be approximated by a $(c_s + \lambda_s)\chi^2(f_s)/f_s$-distribution where

$$f_s = 2\left(1 + \frac{c_s^2}{\lambda_s^2 + 2c_s\lambda_s}\right),$$

cf. e.g. [9]. Note that for a circular template, $c_s = 0$, $f_s = 2$ and the result is exact. If $c_s \gg \lambda_s$, then f_s will be large and the distribution of C_s is concentrated around c_s.

The distributional results obtained above for (C_s, D_s) in the Gaussian case motivate extensions of our model to the non-Gaussian case. Instead of the $(c_s+\lambda_s)\chi^2(f_s)/f_s$-distribution one might use a generalized gamma distribution as a model for C_s, cf.

[10], Section 8.4. Under a Gaussian assumption the phase angles are von Mises distributed as indicated in (2.3). An extension is here to consider a von Mises distribution of the type $vM(2\pi s d_s, \kappa_s\sqrt{c})$ where the parameter $\kappa_s > 0$ is arbitrary, allowing for larger and smaller variation than in the Gaussian case.

3 Elliptical templates

From now on we consider the special case where the template is an ellipse.

Let us start by introducing some notation for an ellipse. Assume the centre z is located at the origin, let the lengths of the the axes be denoted by $a \geq b$ and the eccentricity be $\epsilon = (1 - b^2/a^2)^{1/2}$. If the major axis of the ellipse has an angle of $2\pi\theta$ relatively to the first axis, $\theta \in [0, 1/2)$, then the boundary of the ellipse is given by

$$(x(t), y(t)) = r(t)\left(\cos(2\pi t), \sin(2\pi t)\right), \quad t \in [0, 1],$$

where the radius-vector function is

$$r(t) = \frac{ab}{\sqrt{a^2\sin^2(2\pi(t-\theta)) + b^2\cos^2(2\pi(t-\theta))}}. \tag{3.1}$$

The boundary length between the points with indices 0 and t is

$$\begin{aligned} l(t) &= \int_0^t \sqrt{(x'(u))^2 + (y'(u))^2}du \\ &= 2\pi ab \int_{-\theta}^{t-\theta} \frac{\left(a^4\sin^2(2\pi u) + b^4\cos^2(2\pi u)\right)^{1/2}}{\left(a^2\sin^2(2\pi u) + b^2\cos^2(2\pi u)\right)^{3/2}}du. \end{aligned} \tag{3.2}$$

The time change γ will be taken to be either the relative boundary length, $\gamma(t) = l(t)/l(1)$, or the identity. In both cases one easily shows that at odd phases the time changed radius-vector function $r_0(t) = r(\gamma^{-1}(t))$ has vanishing Fourier coefficients. In fact, if the eccentricity is not too large then the elliptical shape is mainly determined by the Fourier coefficients at the phases $s = 0$ and $s = 2$ for these two choices of time change.

Recall that the time changed radius-vector function R_0 of the random object K is $R_0(t) = r_0(t) + X_0(t)$. We want to specialize the general Gaussian model for X_0 such that K has a pronounced elliptical shape. As noted above this should be reflected in the Fourier coefficients at phases $s = 0$ and $s = 2$. We assume that X_0 has small Fourier coefficients when $s = 0$ and $s = 2$ such that at these phases R_0 is described almost entirely by the terms from the ellipse. Moreover, typically K will be rather symmetrical with respect to the centre z, and arguing as in Section 2 of [8] this implies that the Fourier coefficients of R_0 at phase $s = 1$ are small.

We hence consider the Gaussian model (2.2) with variances $\lambda_0, \lambda_1, \lambda_2$ small. The remaining variances are modelled by the simple regression model

$$\lambda_s^{-1} = \alpha + \beta(s^{2p} - 3^{2p}), \quad s \geq 3. \tag{3.3}$$

The parameter α determines the 'global' deviation from the template while β, p determine the 'roughness' of the boundary. The reason is that when $s \geq 3$ is small then λ_s is mainly determined by α. A small value of α gives a large value of λ_s which typically implies a high 'global' fluctuation in $X_0(t)$. Similarly when s is large λ_s is merely determined by β and p. Large values of these parameters yield small variances λ_s such that the boundary of K will be rather smooth. In [8] it is discussed how p relates to continuity and differentiability of the trajectories of X_0, see in particular Section 3 of that paper. The regression model is called a p-order model because it appears as the limit of discrete time p-order Markov models, cf. [8].

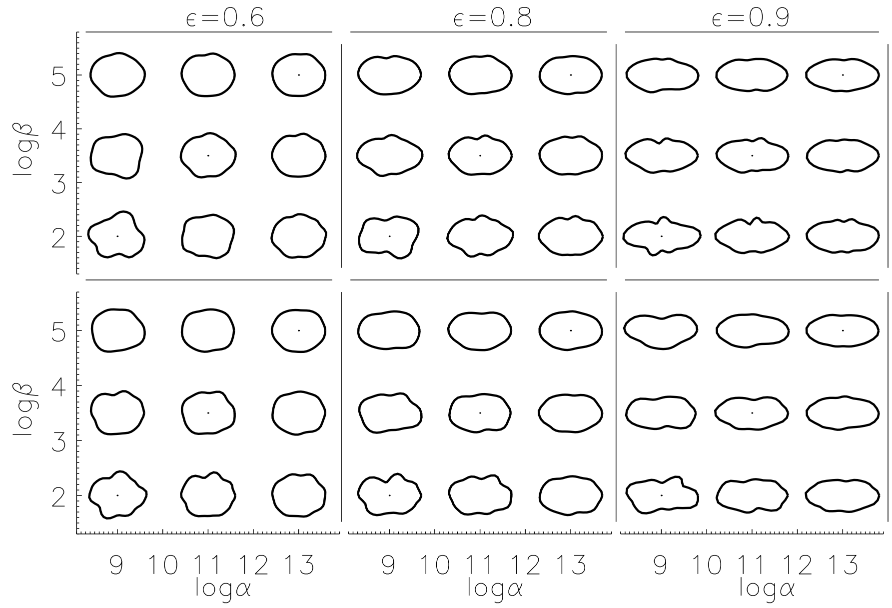

Figure 1: Simulations from the model (2.2) with $A_s = B_s = 0, s \leq 2$, λ_s given by (3.3) and an elliptical template with unit perimeter. In the first three rows the time change is $\gamma(t) = t$ while in the last three rows $\gamma(t) = l(t)$. The values of ϵ, α and β are as indicated and $p = 2.5$.

In the statistical shape literature a Gaussian model is commonly used, cf. e.g. [4, 5, 6, 7, 8, 11, 15, 16, 17]. The papers [7, 8, 11] considered Gaussian models with Fourier coefficients at phase $s = 0, 1$ close to zero. The additional constraint on $s = 2$ here is due to the choice of template. Compared to [7, 8, 11] we have also introduced the time change γ. In [6] a template ellipse was considered but the constraints on the Fourier coefficients were not incorporated.

An effective way of checking that a model has the right properties is to inspect random samples from the model. In Figure 1 we show simulations using both $\gamma(t) = t$ and $\gamma(t) = l(t)/l(1)$. All templates were scaled such that the perimeter was $l(1) = 1$. For $\epsilon = 0$ the two time changes are identical and therefore yield the same model. However, at high eccentricities it is apparent that $\gamma(t) = t$ results in some undesirable small 'blobs' in K near the minor axis of the template ellipse. In the following we will therefore mainly use $\gamma(t) = l(t)/l(1)$ as our time change.

4 Data analysis

The data consists of 27 profiles of cell nuclei from a malignant tumour and 27 cell nuclei from a benign tumour of human skin, cf. Figure 2 and [6], where the data has previously been analysed. The cell nuclei from the benign tumour seem to be small deformations of ellipses with varying eccentricities while the nuclei from the malignant tumour are larger deformations of ellipses. Our model should be able to capture this difference.

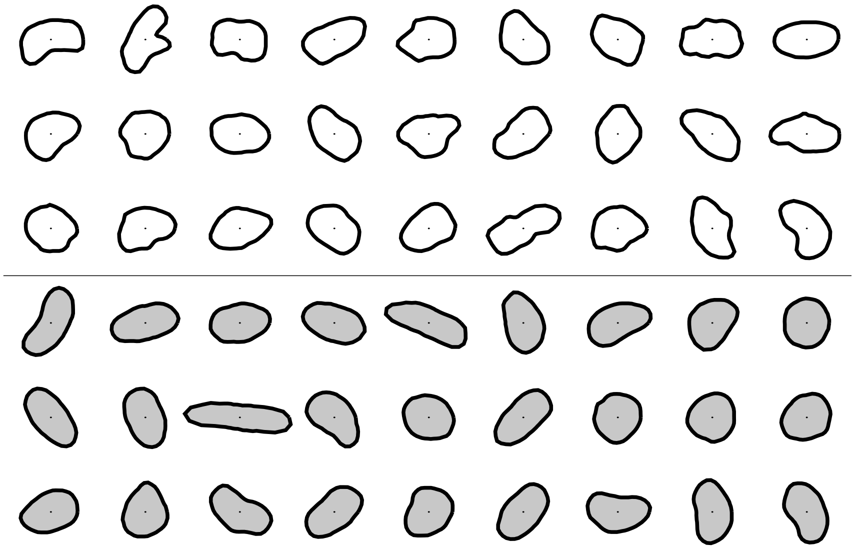

Figure 2: The upper panel is profiles of cell nuclei from a malignant tumour while the lower is from a benign tumour. The cell nuclei have been scaled so that they have approximately the same size.

For each profile we chose z as the centre of mass and calculated the radius-vector function R at the points $t = 0, 1/n, \ldots, (n-1)/n$, where $n = 50$. We tried several different ways of fitting the ellipse. The elegant method described in [2] was implemented, the least squares method described in [4] was also used, but we ended up fitting the ellipse using the Fourier coefficients at the phases $s = 0, 2$ only. However, the three methods resulted in almost the same template ellipse.

Since we are only interested in the shape we scaled the resulting residual process $X(t) = R(t) - r(t)$ by the perimeter of the ellipse. Finally we calculated the Fourier coefficients of the normalized time changed process $X_0(t) = X(\gamma^{-1}(t))$, where we used $\gamma(t) = l(t)/l(1)$. The Fourier coefficients are

$$A_s = \sqrt{2}\int_0^1 X_0(t)\cos(2\pi st)dt = \sqrt{2}\int_0^1 X(t)\cos(2\pi s\gamma(t))\gamma'(t)dt, \qquad (4.1)$$

and the expression for B_s is similar.

It remains to fit the regression model (3.3), based on $A_s, B_s, s \geq 3$. The Fourier coefficients at high phases are poorly determined due to digitization effects, cf. e.g. [7, 8, 11]. We therefore considered the well-determined Fourier coefficients $A_3, B_3, \ldots, A_S, B_S$ only, where S is a reasonable cut-off value. In practice it turned out that relatively few Fourier coefficients are well-determined and we used $S = 11$. Since the Fourier coefficients are zero mean Gaussian it follows that the likelihood function for a profile is

$$L(\alpha, \beta, p) = \prod_{s=3}^{S} \frac{1}{2\pi\lambda_s} \exp\left(-\frac{A_s^2 + B_s^2}{2\lambda_s}\right), \tag{4.2}$$

where λ_s is given by (3.3).

For each profile we found the estimates of (α, β, p) by maximising (4.2). For the malignant sample the average of p was 2.72 with a standard deviation of 0.68, while for the benign sample the average was 2.49 and the standard deviation 0.79. We therefore fixed $p = 2.5$. The estimates of (α, β) under the p-order model with $p = 2.5$ are shown in Figure 3 and summarized in Table 1. The estimates of the local shape parameter β are on average significantly lower in the malignant sample (p-value for identical βs in the two samples is less than 0.01 %). This was to be expected from the simulations and geometric interpretation of β given in Section 3.

	$\log \hat{\alpha}$		$\log \hat{\beta}$		
	av.	s.d.	av.	s.d.	corr.
benign	11.59	1.42	4.68	0.53	0.43
malignant	10.65	1.10	3.66	0.88	0.45

Table 1: The average, standard deviation and correlation of $(\log \hat{\alpha}, \log \hat{\beta})$ for each sample.

On average the estimates of the global shape parameter α are also lowest in the malignant sample, and again the difference is significant (p-value close to 1%). Furthermore the variance of $\log \beta$ is significantly larger in the malignant sample.

5 Perspective

The deformable template model considered in the present paper can be embedded in the marked point process framework described in [1] and thereby used in Bayesian object recognition. In this set-up the observed digital image y depends on the true scene of interest x through the likelihood $f(y|x)$. Inference for x is made using the posterior

$$p(x|y) \propto f(y|x)p(x),$$

where $p(x)$ is the prior distribution of x. The scene x is represented as a finite set of m objects, $x = \{x_1, \ldots, x_m\}$, where m is unknown. Each object x_i is specified as a marked point where the point gives the location and the mark determines the object. The basic distribution is the Poisson object process, where the number of objects

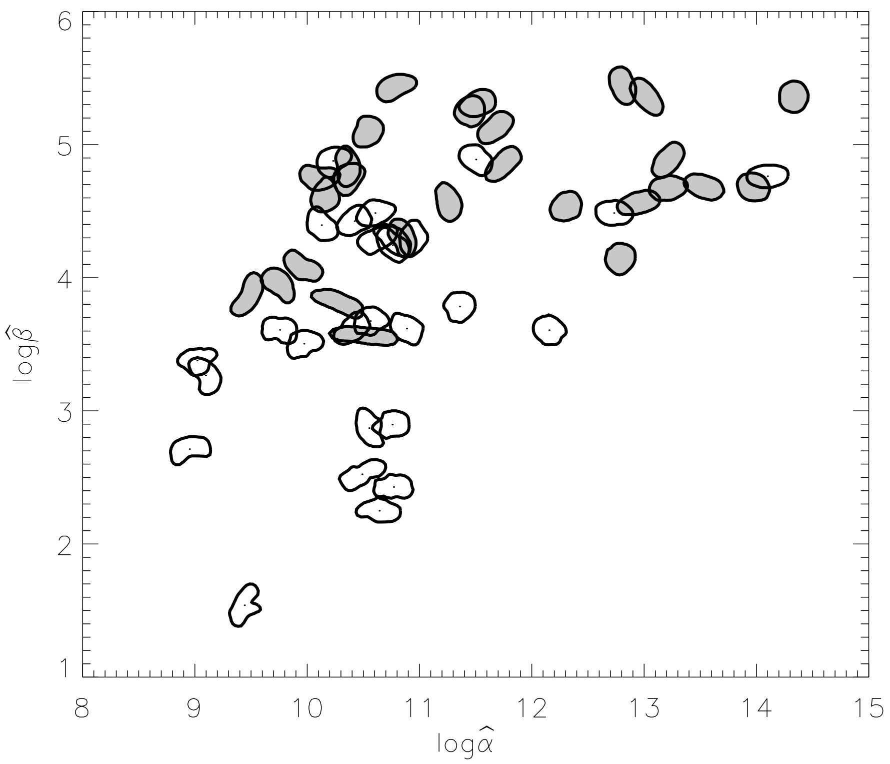

Figure 3: The estimates of (α, β) when p is fixed at $p = 2.5$ The hatched nuclei are from the benign sample while the white nuclei are from the malignant sample.

are Poisson distributed and, conditional on this number, the locations of the objects are independent and uniformly distributed. The distribution of the marks is often given by a deformable template model. Examples include [13] where mushrooms in a growing bed are analysed. Here, the mushrooms are modelled as discrete circles and the deformation is through scaling only. In [16] cells in a confocal microscopy image are located and discrete circles of different sizes are used, but a residual process similar to the one presented in this paper is also added. A third example is provided by [15] where the template is a discrete circle or a discrete ellipse with fixed eccentricity. The residual process, which determines the deformation of the template edges, is a discrete first-order Markov process. Our analysis can be viewed as a detailed investigation of the deformable template model with an ellipse as template. In particular we found that a time change is needed if the eccentricity is high and the radial representation is used.

Acknowledgements

The data has kindly been supplied by Professor Flemming Sørensen, Institute of Pathology, University of Aarhus. This work is supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation. The second author is supported by a grant from the Danish Natural Science Research Council.

Appendix

Let $Y_1 \sim N(a, \lambda)$ and $Y_2 \sim N(b, \lambda)$ be independent random variables and let

$$a = l \cos \mu,\ b = l \sin \mu,\ Y_1 = R \cos \Theta,\ Y_2 = R \sin \Theta.$$

We show that

$$\Theta | R = r \sim vM(\mu, \frac{lr}{\lambda}).$$

From the change of variables formula it follows that the density function of (R, Θ) is

$$f_{R,\Theta}(r, \theta) = \frac{r}{2\pi\lambda} \exp\left(\frac{-r^2 - l^2 + 2lr\cos(\theta - \mu)}{2\lambda}\right),$$

and by integrating with respect to θ we get

$$f_R(r) = \frac{r}{\lambda} \exp\left(\frac{-r^2 - l^2}{2\lambda}\right) I_0\left(\frac{lr}{\lambda}\right),$$

where I_0 denotes the modified Bessel function of the first kind and order 0. The result now follows immediately.

References

[1] A. J. Baddeley and M. N. M. van Lieshout. Stochastic geometry models in high-level vision. In K. V. Mardia and G. K. Kanji, editors, *Statistics and Images, Advances in Applied Statistics, a supplement to the Journal of Applied Statistics*, volume 20, chapter 11, pages 231 – 261. Carfax publishing, 1993.

[2] A. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:476 – 480, 1999.

[3] U. Grenander. *General Pattern theory.* New York: Oxford University Press, 1993.

[4] U. Grenander and K. M. Manbeck. A stochastic model for defect detection in potatoes. *Journal of Computer Graphics and Statistics*, 2:131 – 151, 1993.

[5] U. Grenander and M. I. Miller. Representations of knowledge in complex systems (with discussion). *Journal of the Royal Statistical Society, B*, 56:549 – 603, 1994.

[6] A. Hobolth and E. B. V. Jensen. Modelling stochastic changes in curve shape, with an application to cancer diagnostics. *Advances in Applied Probability (SGSA)*, 32:344 – 362, 2000.

[7] A. Hobolth, J. T. Kent, and I. L. Dryden. On the relation between edge and vertex modelling. Research Report no. 7, Laboratory for Computational Stochastics, University of Aarhus. Submitted, 1999.

[8] A. Hobolth, J. Pedersen, and E. B. V. Jensen. A continuous parametric shape model. Research Report no. 13, Laboratory for Computational Stochastics, University of Aarhus. Submitted, 2000.

[9] J. L. Jensen. A large deviation-type approximation for the "Box class" of likelihood ratio criteria. *Journal of the American Statistical Association*, 86:437 – 440, 1991.

[10] N. L. Johnson and S. Kotz. *Continuous univariate distributions - 1.* Boston: Houghton Mifflin Company, 1970.

[11] J. T. Kent, I. L. Dryden, and C. R. Anderson. Using circulant symmetry to model featureless objects. *Biometrika*, 87:527 – 544, 2000.

[12] K. V. Mardia and P. E. Jupp. *Directional Statistics.* Chichester: Wiley, 2000.

[13] K. V. Mardia, W. Qian, D. Shah, and K. M. A. de Souza. Deformable template recognition of multiple occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:1035 – 1042, 1997.

[14] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Vol. 2.* Chichester: Wiley, 2000.

[15] H. Rue and M. A. Hurn. Bayesian object recognition. *Biometrika*, 86:649 – 660, 1999.

[16] H. Rue and A. R. Syversveen. Bayesian object recognition with Baddeley's delta loss. *Advances in Applied Probability (SGSA)*, 30:64 – 84, 1998.

[17] D. Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields.* Chichester: Wiley, 1994.

[18] C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, 21:269 – 281, 1972.

E

Hobolth, A and Jensen, E.B.V. (2001).
A note on design-based versus model-based variance estimation in stereology.
Research report no. 17, Laboratory for Computational Stochastics, University of Aarhus.

E

A note on design-based versus model-based variance estimation in stereology

ASGER HOBOLTH AND EVA B. VEDEL JENSEN
Laboratory for Computational Stochastics
University of Aarhus

Abstract

Recently, systematic sampling on the circle and the sphere has been studied by Gual-Arnau and Cruz-Orive (2000) from a design-based point of view. In this note, it is shown that their mathematical model for the covariogram is in a model-based statistical setting a special case of the p-order shape model, suggested in Hobolth *et al.* (1999, 2000) for planar objects without landmarks. Benefits of this observation include an alternative variance estimator, applicable in the original problem of systematic sampling. In a wider perspective, the paper contributes to the discussion concerning design-based versus model-based stereology.

Keywords: covariogram, circulant matrix, Fourier series, planar objects, shape, stationarity, stereology, systematic sampling.

1 Introduction

In stereology, the aim is typically to make inference about a population of spatial objects from geometric samples of the objects such as line and plane sections. The objective is not to reconstruct the objects, but instead to make inference about quantitative properties such as volume or surface area. If a typical object from the population can be regarded as a realization of a stochastic process R, then the quantitative property of interest can be expressed as a function f of R. Using a geometric sampling design ϕ, independent of R, a predictor $\hat{f}(R,\phi)$ of $f(R)$ can often be constructed, based on reasoning from stochastic geometry, which is design-unbiased, i.e.

$$\mathrm{E}(\hat{f}(R,\phi)|R) = f(R).$$

It is part of the methodology of design-based stereology to construct a design-unbiased estimator $\hat{\sigma}^2_R(\phi)$ of the conditional variance

$$\sigma^2_R = \mathrm{Var}(\hat{f}(R,\phi)|R).$$

The estimator $\hat{\sigma}^2_R(\phi)$ thus satisfies

$$\mathrm{E}(\hat{\sigma}^2_R(\phi)|R) = \sigma^2_R.$$

Usually $\hat{\sigma}^2_R(\phi)$ is based on the empirical covariogram.

In most cases it is of interest to make statements about the population of objects and not only about the sampled objects. A relevant quantity is here the prediction error

$$\mathrm{E}(\hat{f}(R,\phi) - f(R))^2.$$

Using that $\hat{f}(R,\phi)$ and $\hat{\sigma}_R^2(\phi)$ are design-unbiased, the prediction error can be rewritten as

$$
\begin{aligned}
&\mathrm{E}(\hat{f}(R,\phi) - f(R))^2 && (1.1)\\
&= \mathrm{Var}\big(\hat{f}(R,\phi) - f(R)\big)\\
&= \mathrm{Var}\big(\mathrm{E}(\hat{f}(R,\phi) - f(R)|R)\big) + \mathrm{E}\big(\mathrm{Var}(\hat{f}(R,\phi) - f(R)|R)\big)\\
&= \mathrm{E}\big(\mathrm{Var}(\hat{f}(R,\phi)|R)\big)\\
&= \mathrm{E}\sigma_R^2 = \mathrm{E}\hat{\sigma}_R^2(\phi). && (1.2)
\end{aligned}
$$

Therefore, $\hat{\sigma}_R^2(\phi)$ or an average of such estimators for a sample of objects can be regarded as an unbiased estimator of the prediction error.

In the present paper we propose the alternative of using a likelihood-based method of estimating the prediction error. The discussion is centred around the example where $R = \{R(2\pi t) \in \mathbb{R} : 0 \le t \le 1\}$ is a 2π periodic stochastic process and

$$f(R) = \int_0^1 R(2\pi t)dt$$

is the quantity of interest. Based on $n \ge 2$ equally spaced measurements

$$\{R(2\pi(\phi + j/n)) : j = 0, \ldots, n-1\}$$

of the stochastic process R, with ϕ uniformly distributed in $[0, 1/n]$, Gual-Arnau and Cruz-Orive (2000) have recently suggested a design-unbiased estimator of the conditional variance of

$$\hat{f}(R,\phi) = \frac{1}{n}\sum_{j=0}^{n-1} R(2\pi(\phi + j/n)). \qquad (1.3)$$

In this paper we suggest a parametric model for R with a covariance structure similar to that of Gual-Arnau and Cruz-Orive (2000). The prediction error is estimated by inserting the maximum likelihood estimates of the model parameters into a closed form parametric expression for the prediction error. The proposed estimator of the prediction error is optimal under the suggested model for R.

The paper is organised as follows. In Section 2 we recall the design-based variance estimation of Gual-Arnau and Cruz-Orive (2000), while the likelihood-based variance estimation is carried out in Section 3. In Section 4 it is shown that the proposed model is a special case of the p-order shape model, suggested in Hobolth *et al.* (1999, 2000) for planar objects without landmarks. It is also pointed out that a similar discussion about estimation procedures has taken place in the geostatistical community during the last decade.

2 Design-based variance estimation

Let $R = \{R(2\pi t) \in \mathbb{R} : 0 \le t \le 1\}$ be a 2π periodic stochastic process, which is of bounded variation, square integrable and piecewise continuous, and let $\phi \sim \mathrm{U}[0, 1/n]$

be independent of R. If we define $\hat{f}(R,\phi)$ as in (1.3), then Cruz-Orive and Gual-Arnau (2000) treat the problem of estimating the conditional variance $\mathrm{Var}(\hat{f}(R,\phi)|R=r)$. In particular they show that, cf. Gual-Arnau and Cruz-Orive (2000, Corollary 2.1),

$$\mathrm{Var}(\hat{f}(R,\phi)|R=r) = \sum_{k\in\mathbb{Z}\setminus\{0\}} c_{kn}, \tag{2.1}$$

where

$$c_k = \int_0^1 g(t)e^{-2\pi ikt}dt,\ k\in\mathbb{Z},$$

are the Fourier coefficients of the covariogram

$$g(t) = \int_0^1 r(2\pi h)r(2\pi(h+t))dh,\ 0\le t\le 1.$$

Here and throughout the paper we use periodic extensions of the functions (i.e. $r(2\pi(x+k)) = r(2\pi x),\ k\in\mathbb{Z}$). Note that c_k is real and $c_k = c_{-k}$ because $g(1-t)=g(t)$. The covariogram

$$g(t) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi ikt} = c_0 + 2\sum_{k=1}^{\infty} c_k\cos(2\pi kt), \tag{2.2}$$

is modelled by a polynomium of order $2p,\ p\in\mathbb{N}$. The fact that $g(t)=g(1-t)$ causes restrictions on the coefficients of the polynomium. Gual-Arnau and Cruz-Orive (2000, p. 635) show that in fact the polynomium only depends on two real parameters β_0,β, and that the Fourier coefficients of g take the form

$$c_0 = \beta_0 - \sum_{k\in\mathbb{Z}\setminus\{0\}} c_k,\quad c_k = \frac{(2p)!}{k^{2p}}\beta,\ k\in\mathbb{Z}\setminus\{0\}, \tag{2.3}$$

where $c_0,\beta>0$. Unbiased estimators of $g(0)$ and $g(1/n)$ are obtained by

$$\begin{aligned}\hat{g}(0) &= \frac{1}{n}\sum_{j=0}^{n-1} r(2\pi(\phi+j/n))^2,\\ \hat{g}(1/n) &= \frac{1}{n}\sum_{j=0}^{n-1} r(2\pi(\phi+j/n))r(2\pi(\phi+(j+1)/n)),\end{aligned}$$

and using the formula for the Bernoulli polynomium, cf. e.g. Abramovitz and Stegun (1970, p. 805),

$$B_{2p}(t) = \frac{(-1)^{p-1}(2p)!}{(2\pi)^{2p}}2\sum_{k=1}^{\infty}\frac{\cos(2\pi kt)}{k^{2p}},\quad 0\le t\le 1,\quad p\in\mathbb{N},$$

an unbiased estimator of $\mathrm{Var}(\hat{f}(R,\phi)|R=r)$ given by

$$\frac{\hat{g}(0)-\hat{g}(2\pi/n)}{n^{2p}}\frac{1}{1-B_{2p}(1/n)/B_{2p}}, \tag{2.4}$$

is obtained, where $B_{2p} = B_{2p}(0)$ is the Bernoulli number of order $2p$. Note that

$$2n(\hat{g}(0)-\hat{g}(2\pi/n)) = \sum_{j=0}^{n-1}\big(r(2\pi(\phi+j/n)) - r(2\pi(\phi+(j+1)/n)\big)^2,$$

and thus the estimator is based on first-order differences.

3 Model-based setting

Now we recast the models and estimation procedures of Gual-Arnau and Cruz-Orive (2000) in terms of a stationary, random periodic process R with mean μ and covariance function

$$\sigma(t) = \sum_{k\in\mathbb{Z}}\lambda_k e^{2\pi ikt} = \lambda_0 + 2\sum_{k=1}^{\infty}\lambda_k\cos(2\pi kt), \quad 0\le t\le 1.$$

Note that the λ_k's are real because $\sigma(1-t)=\sigma(t)$. If we make a Fourier expansion of the random covariogram

$$G(t) = \int_0^1 R(2\pi h)R(2\pi(h+t))dh = C_0 + 2\sum_{k=1}^{\infty}C_k\cos(2\pi kt)$$

then $EC_k = \lambda_k,\ k\ge 1$, and $EC_0 = \lambda_0+\mu^2$. Accordingly, the covariogram model (2.3) corresponds to a covariance function $\sigma(t)$ with

$$\lambda_0 = \beta_0 - \sum_{k\in\mathbb{Z}\backslash\{0\}}\lambda_k,\ \lambda_k = \frac{(2p)!}{k^{2p}}\beta,\ k\in\mathbb{Z}\backslash\{0\}. \tag{3.1}$$

Note that

$$\sigma(0) = \lambda_0 + \sum_{k\in\mathbb{Z}\backslash\{0\}}\lambda_k = \beta_0,$$

which means that β_0 determines the variance and β/β_0 the correlation structure.

In a model-based setting, the aim is to estimate the error involved in using

$$\hat{f}(R,\phi) = \frac{1}{n}\sum_{j=0}^{n-1}R(2\pi(\phi+j/n))$$

as a predictor of $f(R) = \int_0^1 R(2\pi t)dt$. In terms of model parameters the prediction error is given by, cf. (1.1), (2.1) and (3.1),

$$\begin{aligned} \mathrm{E}(\hat{f}(R,\phi) - f(R))^2 &= \mathrm{E}(\mathrm{Var}(\hat{f}(R,\phi)|R)) \\ &= \sum_{k \in \mathbb{Z}\setminus\{0\}} \lambda_{kn} \\ &= \beta \sum_{k \in \mathbb{Z}\setminus\{0\}} \frac{(2p)!}{(kn)^{2p}} \\ &= (-1)^{p-1}(2\pi)^{2p} B_{2p} \frac{1}{n^{2p}} \beta. \end{aligned} \tag{3.2}$$

Note that the prediction error will be the same if we fix $\phi = 0$, say.

We can use the procedure suggested by Gual-Arnau and Cruz-Orive (2000) for obtaining an unbiased estimator of the prediction error in the model-based setting. Another approach is to estimate β in the parametric model for R by maximum likelihood estimation. Suppose for instance that the process R is Gaussian. Then the vector

$$R_n = \big(R(2\pi\phi), R(2\pi(\phi + 1/n)), \ldots, R(2\pi(\phi + (n-1)/n))\big)^T$$

follows a multivariate normal distribution with mean $(\mu, \ldots, \mu)^T = \mu 1_n^T$ and an $n \times n$ circulant covariance matrix

$$\Sigma = \mathrm{circ}\big(\sigma(0), \sigma(1/n), \ldots, \sigma((n-1)/n)\big).$$

The covariance matrix can be diagonalised by the complex $n \times n$ discrete Fourier transform matrix W with entries $w_{jk} = e^{2\pi i jk/n}/n,\ 0 \le j, k \le n-1$, cf. e.g. Wei (1990, Chapter 10). Let w_k denote the $(k+1)$'th coloumn of W so that $W = [w_0, \ldots, w_{n-1}]$ and let $W^* = \overline{W}^T$ denote the complex conjugate of W. Then

$$W^* \Sigma W = \mathrm{diag}(\tilde{\lambda}_0, \ldots, \tilde{\lambda}_{n-1})$$

is a diagonal matrix with

$$\tilde{\lambda}_j = w_j^* \Sigma w_j = \sum_{k \in \mathbb{Z}} \lambda_{j+nk}, \quad j = 0, \ldots, n-1,$$

on the diagonal. Note that only the parameter β is present in the expression of $\tilde{\lambda}_j, j = 1, \ldots, n-1$, while both β and β_0 are present in the expression of $\tilde{\lambda}_0$. Similarly we find that

$$\begin{aligned} (R_n - \mu 1_n)^* \Sigma^{-1} (R_n - \mu 1_n) &= (R_n - \mu 1_n)^* n W W^* \Sigma^{-1} n W W^* (R_n - \mu 1_n) \\ &= \frac{(\hat{f} - \mu)^2}{\tilde{\lambda}_0} + \sum_{j=1}^{n-1} \frac{\hat{\lambda}_j}{\tilde{\lambda}_j} \\ &= \frac{(\hat{f} - \mu)^2}{\tilde{\lambda}_0} + \beta \sum_{j=1}^{n-1} \frac{\hat{\lambda}_j}{\tilde{\kappa}_j}, \end{aligned}$$

where

$$\hat{f} = \hat{f}(R, \phi), \quad \hat{\lambda}_j = \hat{\lambda}_j(R, \phi) = w_j^* R_n R_n^* w_j, \ j = 0, \dots, n-1,$$

and

$$\tilde{\kappa}_j = \tilde{\lambda}_j / \beta = \sum_{k \in \mathbb{Z}} \frac{(2p)!}{(j + nk)^{2p}}, \ j = 1, \dots, n-1. \tag{3.3}$$

Thus the sufficient statistic is given by

$$T = (\hat{f}, \hat{f}^2, \sum_{j=1}^{n-1} \hat{\lambda}_j / \tilde{\kappa}_j).$$

According to the Rao-Blackwell theorem any function of T is the minimum variance estimator of its mean value. Furthermore it follows from the theory of exponential families that T is complete, and hence

$$\hat{\beta} = \frac{1}{n-1} \sum_{j=1}^{n-1} \frac{\hat{\lambda}_j}{\tilde{\kappa}_j}$$

is the unique unbiased estimator of β with minimum variance. Using the real discrete Fourier transform matrix similar calculations show that $\hat{\beta}$ follows a $\beta\chi^2(n-1)/(n-1)$ distribution, and therefore we can supply the point estimate of β with a confidence interval. This is an important option which does not exist in a design-based setting.

The likelihood-based estimator of β is a weighted sum of the squared length of the discrete complex Fourier coefficients $\hat{\lambda}_j$. It is clear from (3.3) that the weights $\tilde{\kappa}_j$ depend crucially on the order $2p$ of the polynomium. Below we discuss how p relates to the smoothness of the sample paths, and may be considered as a third parameter in the model.

We estimate the prediction error by, cf. (3.2),

$$(-1)^{p-1}(2\pi)^{2p} B_{2p} \frac{1}{n^{2p}} \hat{\beta}. \tag{3.4}$$

It is worth noticing that for $n = 2$ and $n = 3$ this estimator actually coincides with the estimator (2.4) of Gual-Arnau and Cruz-Orive (2000). Note also that in a design-based setting, (3.4) is an unbiased estimator of $\mathrm{Var}(\hat{f}(R, \phi) | R = r)$ under the covariogram model (2.3).

4 Discussion

4.1 The p-order shape model

The model (3.1) is a special case of the p-order model suggested in Hobolth *et al.* (1999, 2000) which appears to be very natural for modelling the shape of planar objects K without landmarks. In this setting K is assumed to be star-shaped with respect to a fixed point $z \in K$, and $R(2\pi t)$ is the radius-vector function evaluated at $2\pi t$, i.e. the distance from z to the boundary of K along a line with angle $2\pi t$ relative to a fixed axis. In the p-order model the λ_k's are determined by

$$\lambda_0 \geq 0, \quad \lambda_k^{-1} = \tilde{\alpha} + \tilde{\beta} k^{2p}, \quad k \in \mathbb{Z} \backslash \{0\}, \tag{4.1}$$

where $\tilde{\alpha} \geq 0$, $\tilde{\beta} > 0$, $p > 1/2$. Note that in this model p is a parameter and not a fixed integer as in (2.3). For $\tilde{\alpha} = 0$ and $1/\tilde{\beta} = (2p)!\beta$ we get the model (3.1). In Hobolth *et al.* (2000) it is discussed how the parameters $(\tilde{\alpha}, \tilde{\beta}, p)$ relates to the shape of the object. The parameter p determines the smoothness of the object boundary. In the Gaussian case the sample paths are k times continuously differentiable, where k is the integer satisfying $p \in]k - 1/2, k + 1/2[$. For fixed p, $\tilde{\alpha}$ determines the global shape while $\tilde{\beta}$ determines the local shape. Furthermore, it can be argued that λ_1 relates to asymmetry of K relative to $z \in K$, so the regression model (4.1) should for geometrical reasons only be considered for $|k| \geq 2$. In Hobolth *et al.* (2000) it is demonstrated how the three parameters can be estimated using maximum likelihood. Based on the observed information it is also possible to determine confidence intervals of the parameters.

In geometric examples, R is typically a power (2 or 3) of the radius-vector function. In such cases, a Gaussian assumption may not be appropriate. Hobolth *et al.* (2000) provide tools for analysing non-Gaussian processes in this context.

4.2 Covariogram versus likelihood-based methods

The estimation procedure of Gual-Arnau and Cruz-Orive (2000) is based on the empirical covariogram, while we suggest a likelihood-based method. In the geostatistical community a discussion of the two procedures have taken place during the last decade, and has resulted in a move towards the adoption of likelihood-based methods (Diggle *et al.*, 1998, p. 305). We refer the interested reader to the recent monograph Stein (1999) and references therein for more information on parameter estimation using covariogram- or likelihood-based methods. We believe that a corresponding discussion is needed among the stereologists and we hope with this paper to have contributed in a constructive manner to such a discussion.

Acknowledgement

We are very grateful to Jan Pedersen for his valuable comments on this work. This work was supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation.

References

Abramovitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions.* Dover, New York.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Appl. Statist.* **47**, 299-350.

Gual-Arnau, X. and Cruz-Orive, L.M. (2000). Systematic sampling on the circle and on the sphere. *Adv. Appl. Prob. (SGSA)* **32**, 628-647.

Hobolth, A., Kent, J.T. and Dryden, I.L. (1999). On the relation between edge and vertex modelling. *Research Report* **7**, Laboratory for Computational Stochastics, University of Aarhus. To appear in *Scand. J. Statist.*

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2000). A continuous parametric shape model. *Research Report* **13**, Laboratory for Computational Stochastics, University of Aarhus. Submitted.

Stein, M.L. (1999). *Interpolation of Spatial Data.* Springer, New York.

Wei, W.W.S (1990). *Time Series Analysis.* Addison-Wesley, Redwood City, California.

F

Hobolth, A (2002).
The spherical deformation model.
Research report no. 18, Laboratory for Computational Stochastics, University of Aarhus.

F

The spherical deformation model

Asger Hobolth
University of Aarhus

Abstract

Miller *et al.* (1994) describe a model for representing spatial objects with no obvious landmarks. Each object is represented by a global translation and a normal deformation of a sphere. The normal deformation is defined via the orthonormal spherical-harmonic basis. In this paper we analyse the spherical deformation model in detail and describe how it may be used to summarise the shape of star-shaped three-dimensional objects with few parameters. It is of interest to make statistical inference about the three-dimensional shape parameters from continuous observations of the surface and from a single central section of the object. We use maximum likelihood based inference for this purpose and demonstrate the suggested methods on real data.

Keywords: deformable templates, shape, spherical harmonic, stationarity.

1 Introduction

Miller *et al.* (1994) describe a model for representing three-dimensional amoebae in optical sectioning microscopy. Each amoeba is represented by a global translation and a normal deformation of a sphere. The Gaussian model for the normal deformation is defined via the spherical-harmonic basis and possesses rotational symmetry. The purpose of this paper is to give a more detailed investigation of the spherical deformation model, and to use the model to explore shape variability. Here shape refers to the geometrical properties of an object which are invariant under translation, rotation and isotropic scaling.

Shape modelling of planar objects with no obvious landmarks has attracted much attention recently. Grenander and Miller (1994) propose a model where a planar object is represented by n vertices around its perimeter, and is described by deforming an n-sided regular polygon using edge transformations. The edge model of Grenander and Miller (1994) is analysed in detail in Kent *et al.* (2000), who use the model to explore shape variability. In Hobolth *et al.* (2002) the object is described using vertex transformations, and it is concluded that Fourier analysis of the standardized radius-vector function is an efficient way of exploring shape variability for star-shaped planar objects. In this paper the three-dimensional spherical counterpart of the planar circulant Fourier analysis is investigated. It is also demonstrated how to perform statistical inference about the three-dimensional shape parameters from central sections of the object.

The paper is organised as follows. In Section 2 the two samples of data are described. The first data sample consists of the surfaces of five neurons from the human hippocampus, and the second data sample consists of ten central sections of neurons from the human hippocampus. Section 3 is concerned with the geometry of the spherical-harmonic basis. In Section 4 the spherical deformation model is specified, and in Section 5 the model is fitted to the first data sample. The relation between the circulant and the spherical methods is outlined in Section 6, and in Section 7 the relation is used to fit the spherical deformation model to the second data sample. The paper is concluded with a discussion on the use of the spherical deformation model.

2 The data

Optical sectioning microscopy (OSM) is a widely used tool to get access to the 3D structure of biological specimens. In OSM a series of focal planes spanning the specimen are acquired. This paper is devoted to the study of neurons from the human hippocampus. Hippocampus is the part of the brain concerned with recent memory and the ability to learn new facts and skills. A hippocampus neuron can be viewed as a small deformation of a sphere with a typical diameter of 10-15 μm. The focal planes of a 70 μm thick glycol methacrylate section were viewed in an Olympus BX 50 light microscope. The microscope is mounted with a 100X oil objective with a small focal depth (NA 1.4), a motorized microscope stage, and a 3 CCD video camera connected to a personal computer. We collected the images of 5 neurons where the focal planes were 0.5 μm apart, and the magnification was chosen so that the central section of each neuron had a diameter of around 400 pixels. Due to the substantial amount of optical distortion the boundary of each neuron was traced manually when the focal depth was at the top or bottom of the neuron while a simple thresholding was sufficient to obtain the boundary at the remaining focal depths. We also collected the boundary of the central section of 10 neurons. The data is shown in Figure 1 and Figure 2.

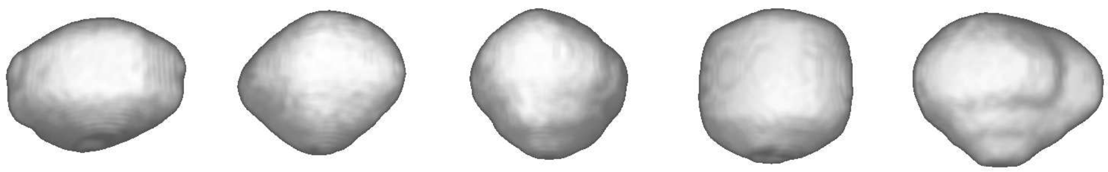

Figure 1: Surfaces of five neurons from the human hippocampus.

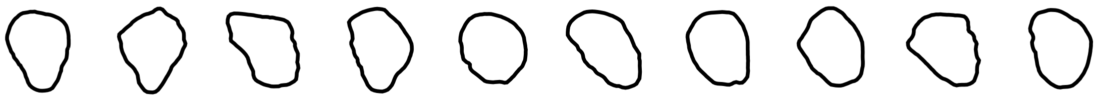

Figure 2: Central sections of ten neurons from the human hippocampus.

3 Spherical harmonics

Consider a solid object $K \subset \mathbb{R}^3$ and suppose that K is star-shaped relative to $z \in K$, i.e. every point on the surface of K is visible from z. Using spherical coordinates the surface is determined by

$$\{z + r(\theta, \phi)\omega(\theta, \phi) : \ 0 \leq \theta < 2\pi, \ 0 \leq \phi \leq \pi\},$$

where $\omega(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)$ is the vector on the unit sphere with polar longitude θ and polar latitude ϕ, and $r(\theta, \phi)$ is the distance from z to the surface of K in direction $\omega(\theta, \phi)$. It is useful to express the radius-vector function $r(\theta, \phi)$ in terms of the spherical harmonics

$$\{\varphi_n^m(\theta, \phi) : \ n \in \mathbb{N}_0, \ m = -n, \ldots, n\},$$

which constitute an orthonormal basis on the sphere. The spherical harmonics are given by

$$\varphi_n^m(\theta, \phi) = \begin{cases} k_n^{|m|} P_n^{|m|}(\cos\phi) \cos m\theta, & m = -n, \ldots, -1 \\ k_n^0 P_n^0(\cos\phi), & m = 0 \\ k_n^m P_n^m(\cos\phi) \sin m\theta, & m = 1, \ldots, n, \end{cases}$$

where

$$k_0^0 = \frac{1}{\sqrt{2\pi}}, \ k_n^0 = \sqrt{\frac{2n+1}{4\pi}}, \ k_n^m = \sqrt{\frac{2n+1}{2\pi} \frac{(n-m)!}{(n+m)!}}, \ n \in \mathbb{N}, \ m = 1, \ldots, n, \quad (3.1)$$

are normalizing constants and P_n^m are the associated Legendre functions of the first kind.

Now consider the Fourier-Legendre series expansion

$$r(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(\theta, \phi)$$

of the radius-vector function, where the Fourier-Legendre coefficients are given by

$$a_n^m = \int_0^{2\pi} \int_0^{\pi} r(\theta, \phi) \varphi_n^m(\theta, \phi) \sin\phi d\phi d\theta. \quad (3.2)$$

The mean radius-vector length is determined by

$$\overline{r} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} r(\theta, \phi) \sin\phi d\phi d\theta = \frac{1}{\sqrt{2\pi}} a_0^0,$$

and hence a_0^0 can be used as a measure of the size of the object. To remove size one may consider the standardized radius-vector function $r(\theta, \phi)/\overline{r}$.

The choice of centre implies constraints on the radius-vector function. If z is the centre of mass of the object then the constraints can be written explicitly, as shown in the appendix. In this case the constraints involve the fourth power of the radius-vector function, but assuming that the object is a small deformation of a sphere, a first-order Taylor expansion leads to the approximate constraints $a_1^m \approx 0,\ m = -1, 0, 1$.

To understand the geometry of the Fourier-Legendre coefficients a_n^m with index $n \geq 2$ it is useful to consider objects given by

$$r(\theta, \phi) = 1 + a_n^m \varphi_n^m(\theta, \phi). \tag{3.3}$$

In general these objects are quite complicated. An example is given in Figure 3.

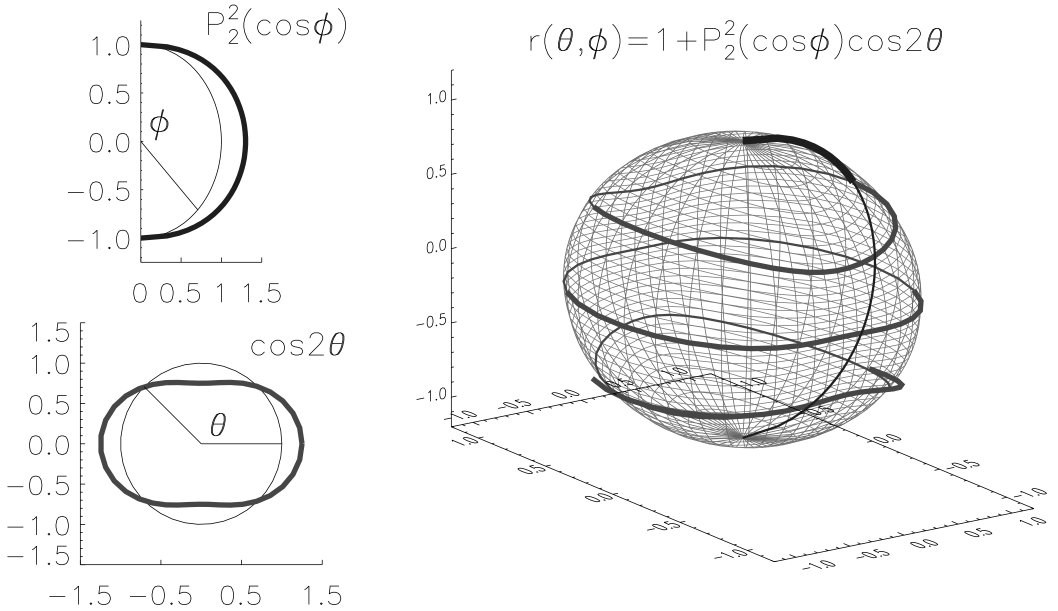

Figure 3: Illustration of the geometry of an object given by (3.3) with $m = -2$ and $n = 2$. To the left the Legendre function $P_2^2(\cos\phi)$, $0 \leq \phi \leq \pi$, and the trigonometric function $\cos 2\theta$, $0 \leq \theta < 2\pi$, are wrapped around the semicircle and the circle. To the right the resulting surface is shown. The four accentuated curves correspond to fixed $\phi = \pi/2, \pi, 3\pi/2$ and $\theta = 0$.

4 The spherical deformation model

Consider the Fourier-Legendre series expansion of the standardized radius-vector function

$$r(\theta, \phi) = 1 + \sum_{n=1}^{\infty} \sum_{m=-n}^{n} a_n^m \varphi_n^m(\theta, \phi).$$

As argued in the previous section the three Fourier-Legendre coefficients corresponding to $n = 1$ are approximately zero if z is the centre of mass and the object is a small deformation of a sphere. Alternatively one could choose the centre as the point satisfying $a_1^m = 0,\ m = -1, 0, 1$ (assuming the object is star-shaped relative to this point). In any case these coefficients are treated as non-random nuisance parameters. The remaining coefficients

$$a_n^m \sim N(0, \lambda_n^m),\ n \geq 2,\ m = -n, \dots, n,$$

are modelled as independent Gaussian random variables with mean zero (the average shape is a sphere) and variance λ_n^m. We also suppose that we have *stationarity on the sphere*, in the sense that the covariance between two points on the sphere depends only on the angle between the points. Stationarity is obtained by assuming

$$\lambda_n^m = \lambda_n \geq 0,\ n \geq 2,\ m = -n, \dots, n, \tag{4.1}$$

and the covariance becomes

$$\begin{aligned} \mathrm{Cov}\big(r(\theta_1, \phi_1), r(\theta_2, \phi_2)\big) &= \sum_{n=2}^{\infty} \lambda_n \sum_{m=-n}^{n} \varphi_n^m(\theta_1, \phi_1)\varphi_n^m(\theta_2, \phi_2) \\ &\overset{(*)}{=} \sum_{n=2}^{\infty} \lambda_n (k_n^0)^2 P_n(\cos\psi_{12}), \end{aligned} \tag{4.2}$$

where $\cos\psi_{12} = \omega(\theta_1, \phi_1) \cdot \omega(\theta_2, \phi_2)$, and we at $(*)$ have used the *addition theorem*, cf. Müller (1966, Theorem 2).

The covariance is thus determined by the variances λ_n, and to proceed further we seek a parametric model for the variances. Miller *et al.* (1994) use a model induced from Poisson's equation for pressure fields acting on thin membranes. The potential associated with the model is given by

$$E(r) = \frac{1}{2} \int_0^{2\pi} \int_0^{\pi} |Lr(\theta, \phi)|^2 \sin\phi d\phi d\theta,$$

where

$$L = \frac{\partial^2}{\partial \theta^2} + \frac{\cos\theta}{\sin\theta} \frac{\partial}{\partial \theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial \phi^2}$$

is the Laplacian operator on the surface of the sphere. Since $\varphi_n^m(\theta, \phi)$ is an eigenfunction of the Laplacian operator with eigenvalue $\eta_n = -n(n+1)$ the potential corresponds to a model where the variances decrease as $1/\lambda_n = \eta_n^2 = (n(n+1))^2$.

Grenander and Miller (1998, Section 5.3) suggest obtaining more general models by introducing polynomials $p(L) = a_0 I + a_1 L + \cdots + a_d L^d$ of the basic operator L. The bi-harmonic operator is for example obtained by choosing $p(L) = L^2$. For such models the variances decrease as $1/\lambda_n = p(\eta_n)^2$.

In this paper we suggest letting the variances decrease according to

$$1/\lambda_n = \alpha + \beta n^p,\ n \geq 2,\ p > 2,\ \beta > 0,\ \alpha > -\beta 2^p. \tag{4.3}$$

There are several reasons why we believe this is a good model. The parameter p makes the model very flexible with regard to the smoothness of the radius-vector function. From Stein (1999, Chapter 2) it follows that the degree of mean square differentiability is determined by the behaviour of the covariance function and its derivatives near the origin. By making repeated use of the relation

$$P'_{n+1}(x) = (2n+1)P_n(x) + P'_{n-1}(x), \quad -1 \leq x \leq 1,$$

and using $P_0(x) = 1$, $P_1(x) = x$ and $P_n(1) = 1$ it can be shown that the radius-vector function is k times mean square differentiable when $2(k+1) < p \leq 2(k+2)$. When $2 < p \leq 4$ the radius-vector function is mean square continuous. Note that in the model used by Miller *et al.* (1994) the variances decrease according to $p = 4$, while for the bi-harmonic operator the variances decrease according to $p = 8$.

For fixed p the value of β determines the 'local' shape of the object since the variances with large indices are determined by β. The third parameter in the model is most easily interpreted when making the reparametrization $\tilde{\alpha} = \alpha + \beta 2^p$, in which case $\tilde{\alpha}$ controls the first few variances, and thereby the 'global' shape of the object. In Figure 4 simulations of central sections $\{r(\theta, \pi/2) : 0 \leq \theta < 2\pi\}$ from the model (4.3) with fixed $p = 4$ and varying $\tilde{\alpha}$ and β are shown.

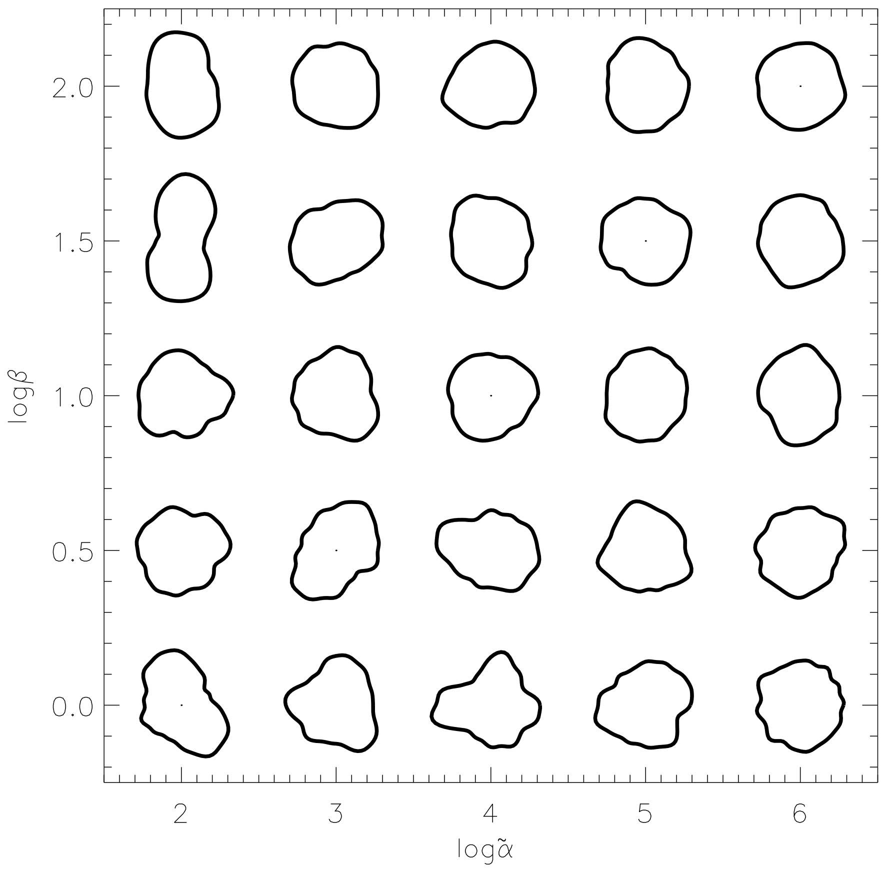

Figure 4: Simulated central sections under the model (4.3) with $p = 4$ and indicated values of $\tilde{\alpha}$ and β.

$\boxed{\mathbf{F}}$

5 Application I: Observing the whole surface

Statistical inference in the spherical deformation model is straightforward when the observations of the object takes the form of (approximately) continuous measurements of the radius-vector function. The data from Figure 1 is used to illustrate the methods. As explained in Section 2 we begin by tracing the boundary of the neuron in each section and thus represent the neuron by a binary volume (1 for a neuron voxel and 0 for a background voxel). From the centre of mass of the neuron equally spaced radii (in terms of ϕ and θ) are traced to the boundary. A boundary point is defined as the voxel location just before the first zero is obtained. The number of boundary points on the object should be chosen high enough to capture the shape of the neuron, but since the neuron is only represented as a binary volume not much extra information is gained by choosing a very high number.

The Fourier-Legendre coefficients $a_{n,k}^m$ for each neuron $k = 1, \ldots, 5 = K$ are obtained from discrete versions of the integrals in (3.2) where the standardized radius-vector function is used. The variances $\lambda_{n,k}$ are calculated from the Fourier-Legendre coefficients

$$\hat{\lambda}_{n,k} = \frac{1}{2n+1} \sum_{m=-n}^{n} (a_{n,k}^m)^2 \sim \lambda_{n,k} \chi^2(2n+1)/(2n+1), \ n \geq 2.$$

Suppose the objects are independent and identically distributed. Then the common variances λ_n are estimated by averaging the variances of the single neurons

$$\hat{\lambda}_n = \frac{1}{K} \sum_{k=1}^{K} \hat{\lambda}_{n,k} \sim \lambda_n \chi^2((2n+1)K)/((2n+1)K), \ n \geq 2.$$

In Figure 5 the estimated log-variances are displayed.

In order to avoid effects due to the discretization of the neuron we choose to proceed with variances $\lambda_{n,k}$ of indices less than $n^\star = 10$ only. The variances with indices above $n^\star$ are judged as being too noisy. The cutting off of the high frequency components in the spectrum is well known in the analysis of time series, where it is referred to as low-pass filtering since only the low frequency components are used for subsequent analysis. The (correspondingly truncated) model (4.3) is fitted to the variances $\hat{\lambda}_2, \ldots, \hat{\lambda}_{n^\star}$ using maximum likelihood. The maximum likelihood estimates of $(\tilde{\alpha}, \beta, p)$ are $(24.9, 2.2, 3.9)$, and Figure 5 displays the log-variances under the fitted model. Carrying out a goodness of fit test of the (truncated) model (4.3) under the (truncated) model (4.1) assuming stationarity we obtain a likelihood ratio testor equal to 6.4 on $(n^* - 1) - 3 = 6$ degrees of freedom. Under the $\chi^2(6)$ approximation of the testor the 'p-value' is 0.38, and so one can consider the model (4.3) as a reasonable way of describing the decrease of the variances.

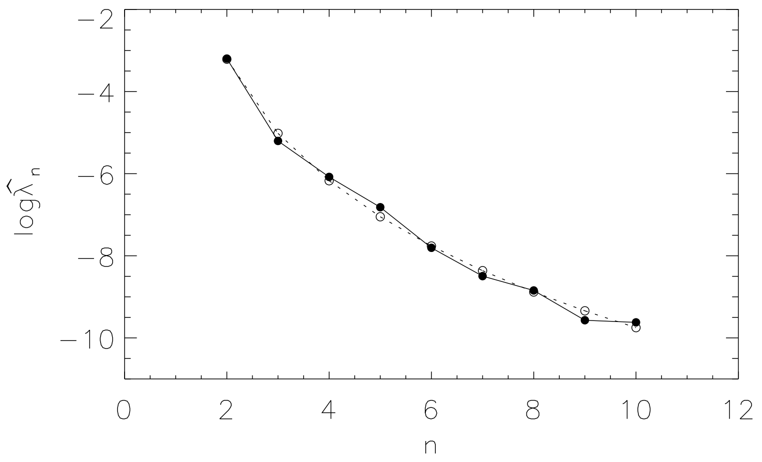

Figure 5: Plot of the estimated log-variances of the stationary model (4.1; solid line) and the regression model (4.3; dotted line) for the 3D surfaces.

6 Relation to the circular deformation model

Suppose we want to estimate the spherical variances λ_n of (4.1) from continuous measurements of the central sections

$$\{r(\theta,\pi/2) : 0 \leq \theta < 2\pi\}$$

only. This task is of considerable interest since the optical distortion complicates the reconstruction of the boundary at the top and bottom of the neurons. Whereas the spherical harmonics constitute the orthonormal basis on the sphere, the most convenient orthonormal basis on the circle is the Fourier basis

$$\{\frac{1}{\sqrt{2\pi}}, \ \frac{1}{\sqrt{\pi}}\cos n\theta, \ \frac{1}{\sqrt{\pi}}\sin n\theta : \ n \in \mathbb{N}\}.$$

Similar to the spherical case we can write the radius-vector function in terms of the Fourier basis

$$r(\theta,\pi/2) = \frac{b_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty}(b_n^c \frac{1}{\sqrt{\pi}}\cos n\theta + b_n^s \frac{1}{\sqrt{\pi}}\sin n\theta),$$

where the Fourier coefficients are given by

$$\begin{aligned} b_0 &= \int_0^{2\pi} r(\theta,\pi/2)\frac{1}{\sqrt{2\pi}}d\theta, \\ b_n^c &= \int_0^{2\pi} r(\theta,\pi/2)\frac{1}{\sqrt{\pi}}\cos n\theta d\theta, \ b_n^s = \int_0^{2\pi} r(\theta,\pi/2)\frac{1}{\sqrt{\pi}}\sin n\theta d\theta, \ n \in \mathbb{N}. \quad (6.1) \end{aligned}$$

As in Section 2 we remove size by considering the standardized radius-vector function $r(\theta,\pi/2)/\overline{r}_{\pi/2}$, where

$$\overline{r}_{\pi/2} = \frac{1}{2\pi}\int_0^{2\pi} r(\theta,\pi/2)d\theta = \frac{1}{\sqrt{2\pi}}b_0.$$

Furthermore the choice of centre implies constraints on the radius-vector function. As in 3D the centre could be chosen as the point fulfilling $b_1^c = b_1^s = 0$. Another possibility is to choose the centre of mass in which case we have $b_1^c \approx 0$ and $b_1^s \approx 0$, provided that the central section is a small deformation of a circle. The last result is derived in the appendix of Hobolth *et al.* (2000), where a careful treatment of the geometry of the radius-vector function in the plane can also be found.

We now determine the distribution of the remaining Fourier coefficients. From

$$\begin{aligned} b_n^s &= \int_0^{2\pi} r(\theta,\pi/2)\frac{1}{\sqrt{\pi}}\sin n\theta d\theta = \int_0^{2\pi}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_l^m \varphi_l^m(\theta,\pi/2)\frac{1}{\sqrt{\pi}}\sin n\theta d\theta \\ &= \sqrt{\pi}\sum_{l=n}^{\infty} k_l^n P_l^n(0) a_l^n \sim N\big(0, \pi\sum_{l=n}^{\infty}(k_l^n P_l^n(0))^2\lambda_l^n\big), \quad n \geq 2, \end{aligned}$$

and a similar calculation of b_n^c it follows that if we assume stationarity on the sphere (4.1) then the variances of b_n^c and b_n^s are equal and given by

$$\kappa_n = \sum_{l=n}^{\infty}\frac{2l+1}{2}\frac{(l-n)!}{(l+n)!}P_l^n(0)^2\lambda_l, \quad n \geq 2. \tag{6.2}$$

Note that the relation between the circular variances κ_n and the spherical variances λ_n is linear. Despite the linearity, the relation is rather complicated. Since $P_{n+1+2s}^n(0) = 0$, $s \in \mathbb{N}_0$, we have that κ_n depends linearly on λ_{n+2s}, $s \in \mathbb{N}_0$, only. Furthermore in our application the spherical variances λ_n are decreasing rapidly, and therefore the most important term when calculating κ_n is the first, which equals a constant times λ_n.

The covariance function in terms of the circular variances κ_n is given by

$$\mathrm{Cov}(r(\theta_1,\pi/2), r(\theta_2,\pi/2)) = \sum_{n=2}^{\infty}\kappa_n\frac{1}{\pi}\cos(n(\theta_2-\theta_1)),$$

and we have *stationarity on the circle* in the sense that the covariance between two points on the circle depends only on the angle between the points. This is of course a direct consequence of the stationarity on the sphere.

7 Application II: Observing the central section

Consider the ten central sections of neurons from the human hippocampus shown in Figure 2. The Fourier coefficients are obtained from discrete versions of the integrals in (6.1), where the standardized radius-vector function is used. Estimates of the circular variances $\kappa_{n,k}$ for each of the ten objects $k = 1,\ldots,10 = K$ are calculated from the Fourier coefficients

$$\hat{\kappa}_{n,k} = \frac{(b_{n,k}^c)^2 + (b_{n,k}^s)^2}{2} \sim \kappa_{n,k}\chi^2(2)/2, \quad n \geq 2.$$

Suppose the objects are independent and identically distributed. Then the common variances κ_n are estimated by averaging the variances of the individual neurons

$$\hat{\kappa}_n = \frac{1}{K}\sum_{k=1}^{K} \hat{\kappa}_{n,k} \sim \kappa_n \chi^2(2K)/2K, \quad n \geq 2.$$

The assumption of equal variances can be verified by a Bartlett test. Another possibility is to use the fact that a $\chi^2(2)/2$ distribution is an exponential distribution with mean one and consider the empirical survival function for each n. In Figure 6 the estimated log-variances are displayed (upper solid lines). As in Section 4 we choose to truncate at $n^\star = 10$ since the variances with indices larger that $n^\star$ are subject to noise due to digitization effects.

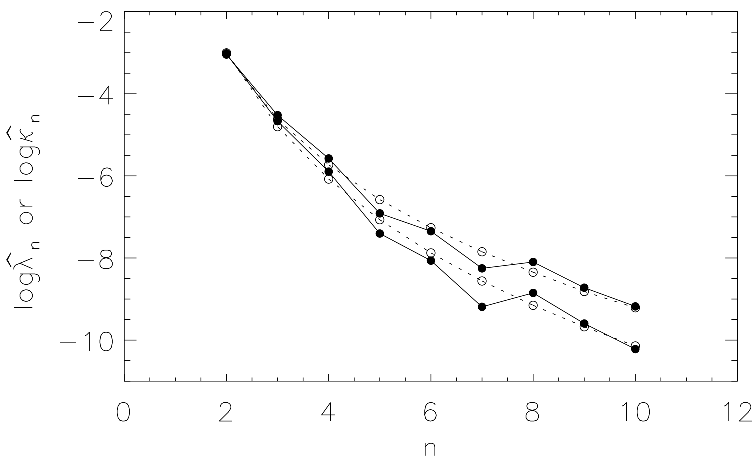

Figure 6: Plot of the estimated log-variances of the stationary model (4.1; solid lines) and the regression model (4.3; dotted lines) for the 2D central sections. Both the estimated circular variances $\hat{\kappa}_n$ (top) and the spherical variances $\hat{\lambda}_n$ (bottom) are plotted.

Using the linear relationship (6.2) it is possible to find approximate estimates of the spherical variances λ_n. The estimates are only approximate because the sum in (6.2) needs to be truncated. In Figure 6 the estimated spherical variances are displayed where we have truncated at $l = 15$ (lower solid lines). The similarity of the decrease of the variances in Figure 5 and Figure 6 is worth noticing. Note that the relation between the spherical variances and the circular variances is similar to the relation between mean squares and variance components in ANOVA estimation of variance components from balanced data, cf. e.g. Searle, Casella and McCulloch (1992, Chapter 4). In particular Satterthwaite's procedure (Satterthwaite, 1946) can be applied to approximate the distribution of the spherical variance estimates by $\chi^2(p)/p$ distributions, and thereby to determine confidence intervals for the spherical variances.

The linear relationship (6.2) also makes it possible to fit the (truncated) regression model (4.3). We fitted the model using maximum likelihood, and obtained the maximum likelihood estimates $(\hat{\tilde{\alpha}}, \hat{\beta}, \hat{p}) = (20.2, 0.9, 4.4)$. Figure 6 displays the circular and spherical log-variances under the fitted model (upper and lower dotted lines). Carrying out a goodness of fit test of the regression model (4.3) under the model (4.1) assuming stationarity we obtain a likelihood ratio testor equal to 3.7 on 6 degrees of freedom. Under the $\chi^2(6)$ approximation of the testor the 'p-value' is 0.72, and so we again judge the model as being a reasonable way of describing the decrease of the variances.

In order to validate the regression model a simulation study was carried out. Figure 7 shows random samples from the regression model (4.3), where the maximum likelihood estimates from the central sections was used for (α, β, p). The random samples show the same shape variability as that seen in the observed central sections in Figure 2.

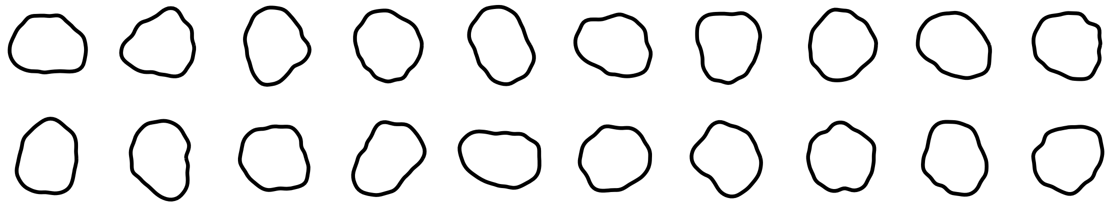

Figure 7: Random samples from the regression model (4.3) with the fitted parameters from the ten central sections of neurons from the human hippocampus.

8 Discussion

In this paper we have analysed the spherical deformation model using two particular types of observations from an object, namely continuous observations of the surface or continuous observations of a central section. It should be emphasized that the spherical deformation model can also be analysed using a finite number of random or systematic observations of the surface. Suppose $r_i = r(\theta_i, \phi_i)$, $i = 1, \dots, n$, is a set of measurements of the radius-vector function of an object, and assume that the vector $(r_1, \dots, r_n)$ follows a multivariate normal distribution with mean $\mu \in \mathbb{R}^n$ and $n \times n$ covariance matrix K. Under rotational symmetry (or stationarity) the entries $K_{ij} = \mathrm{Cov}(r_i, r_j)$ of the covariance matrix K should only depend on the angle ψ_{ij} between the two vectors $\omega(\theta_i, \phi_i)$ and $\omega(\theta_j, \phi_j)$ on the unit sphere. Among the many possible families of covariance functions which could be proposed, this paper has mainly been concerned with the three-parameter regression model given by (4.2) and (4.3).

In Joshi *et al.* (1997) a stochastic model for representing non-spherical objects is introduced in order to describe the shape of the cortical and hippocampal surfaces of macaque and human brains. First a cortical or hippocampal template $\mathcal{M}$ is established. Second a complete orthonormal basis φ_n, $n \in \mathbb{N}$, of the template has to be determined. Joshi *et al.* (1997) suggest choosing the basis functions to correspond to eigenfunctions associated with a differential operator L, derived from thin elastic shell theory. Thus $L\varphi_n = \eta_n \varphi_n$, $n \in \mathbb{N}$, where η_n is the eigenvalue associated with the eigenfunction φ_n. A cortical or hippocampal surface is then determined by $\{x + U(x) : x \in \mathcal{M}\}$, where the Gaussian random field $\{U(x) : x \in \mathcal{M}\}$ is given by

$$U(x) = \sum_{n=1}^{\infty} a_n \varphi_n(x), \; x \in \mathcal{M}.$$

Joshi *et al.* (1997) consider two models for the Gaussian random variables a_n, $n \in \mathbb{N}$. In both models the random variables are independent with zero means, but in the first model the variances λ_n are chosen to be the inverse of the squared eigenvalues $1/\lambda_n = \eta_n^2$ and in the second model the variances are estimated empirically from a sample of surfaces. The relation to the spherical deformation model is obtained by letting $\mathcal{M} = S^2$ and letting the operator L be the Laplacian operator with eigenfunctions φ_n, $n \in \mathbb{N}$, equal to the spherical harmonics. In the spherical deformation model the independence assumption is reasonable on ground of rotational symmetry, but in general the assumption seems rather arbitrary and should be justified.

Acknowledgement

Thanks go to Eva B. Vedel Jensen, John T. Kent and Jan Pedersen for valuable suggestions for improving the paper, and to Jens R. Nyengaard of the Stereological Research Laboratory at Aarhus University for providing the data.

References

Grenander, U. and Miller, M.I. (1994). Representations of knowledge in complex systems (with discussion). *Journal of the Royal Statistical Society, Series B*, **56**, 549-603.

Grenander, U. and Miller, M.I. (1998). Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics.* **5**, 617-694.

Hobolth, A., Kent, J.T. and Dryden, I.L. (2002). On the relation between edge and vertex modelling in shape analysis. To appear in *Scandinavian Journal of Statistics.*

Hobolth, A., Pedersen, J. and Jensen, E.B.V. (2000). A continuous parametric shape model. *Research report* **13**, Laboratory for Computational Stochastics, University of Aarhus.

Joshi, S.C., Miller, M.I. and Grenander, U. (1997). On the Geometry and Shape of Brain Submanifolds. *International Journal of Pattern Recognition and Artificial Intelligence.* **11**, 1317-1343.

Kent, J.T., Dryden, I.L. and Anderson, C.R. (2000). Using circulant symmetry to model featureless objects. *Biometrika*, **87**, 527-544.

Miller, M.I., Joshi, S., Maffitt, D.R., McNally, J.G. and Grenander, U. (1994). Membranes, mitochondria and amoeba: shape models. In *Advances in Applied Statistics*, vol. II, ed. K. Mardia, Carfax Publishing, Abingdon, Oxfordshire, England, 141-163.

Müller, C. (1966). *Spherical harmonics.* Springer-Verlag, Berlin.

Satterthwaite, F.E. (1946). An approximate distribution of variance components. *Biometrics Bull.* **2**, 110-114.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance components.* Wiley, New York.

Stein, M.L. (1999). *Interpolation of Spatial Data.* Springer, New York.

Appendix: Characterizing the centre of mass

Proposition *Suppose $K \subset \mathbb{R}^3$ is a star-shaped object with respect to $z \in K$, and let $r(\theta, \phi; z)$ be the corresponding radius-vector function. If z is the centre of mass of K then*

$$\int_0^{2\pi} \int_0^{\pi} r(\theta, \phi; z)^4 \varphi_1^m(\theta, \phi) \sin\phi d\phi d\theta = 0, \quad m = -1, 0, 1. \tag{A.1}$$

Conversely, z is the centre of mass if (A.1) is fulfilled.

Proof. Let $F : [0, 2\pi] \times [0, \pi] \times [0, 1] \to \mathbb{R}^3$ be defined by

$$F(\theta, \phi, v) = z + vr(\theta, \phi; z)\omega(\theta, \phi)$$

Then F is onto K and elementary calculations show that the absolute value of the Jacobian determinant is $v^2 r(\theta, \phi; z)^3 \sin\phi$. Now if $x = (x_1, x_2, x_3) = F(v, \theta, \phi) \in K$ then

$$x_1 - z_1 = vr(\theta, \phi; z) \cos\theta \sin\phi = vr(\theta, \phi; z)\varphi_1^{-1}(\theta, \phi),$$

and by applying the transformation theorem we obtain

$$\int_K (x_1 - z_1) dx_1 dx_2 dx_3 = \int_0^{2\pi} \int_0^{\pi} \int_0^1 vr(\theta, \phi; z)\varphi_1^{-1}(\theta, \phi) v^2 r(\theta, \phi; z)^3 \sin\phi dv d\phi d\theta, \tag{A.2}$$

which is zero if z is the centre of mass. The two remaining equations in (A.1) are obtained by replacing $(x_1 - z_1)$ in (A.2) by $(x_2 - z_2)$ and $(x_3 - z_3)$, and making similar calculations. □

By making a first-order Taylor expansion we get

$$r(\theta, \phi)^4 \approx \overline{r}^4 + 4\overline{r}r(\theta, \phi),$$

and it follows from the proposition that $a_1^m \approx 0$, $m = -1, 0, 1$ if z is close to the centre of mass. The Taylor expansion is adequate when the objects are small deformations of a sphere.