Department of Engineering
University of Cambridge

Model-Based Three-Dimensional Freehand Ultrasound Imaging

冼顯明
Michael Hsien-Min Syn
Gonville & Caius College

May 1996

A dissertation submitted for the degree of Doctor of Philosophy at the University of Cambridge

Summary

A *3D freehand ultrasound* system augments a conventional clinical scanner with a position sensor on the hand-held probe. Such systems are safe, cheap, portable, and allow clinicians to scan using conventional techniques. Unfortunately the resulting freehand images are non-parallel, sometimes self-intersecting, and retain the noisy image artefacts inherent in conventional 2D ultrasound.

This dissertation proposes two model-based strategies for interpreting such images: an *organ shape model* is used for geometric reconstruction of scattered organ landmarks in the images, and the *Gompertz growth model* is used to register organ shape models to each other in a coherent and biologically justified way.

Both strategies are robust to noise and inaccuracies in the organ model meshes, and are intended to complement future work on the detection of tissue boundaries in ultrasound images. So a *model-based framework* to organise sparse and noisy cues about tissue boundaries, is a key element in any attempt at fully-automated interpretation of 3D freehand ultrasound images.

A biological model of organ growth is first developed using *Oster-Murray mechanisms*, whose *eigenmodes* describe the organ's modes of shape variation. An iterative procedure allows these idealised modes to be refined from organ examples. 3D freehand ultrasound images are then segmented by such organ models, for the purpose of organ *volume estimation*. However, an organ model can only be refined from the segmented organ shape if they both share a common shape parameterisation.

They are therefore *registered* to each other using their eigenmodes, which are proposed to represent *homologous* ('biologically corresponding') landmarks. The choice of registration solutions is restricted to biologically plausible ones using the Gompertz metric. Bayesian combination of the *likelihood* of eigenmode homology, with the *prior* constraint of Gompertzian growth, results in a *posterior* measure of homology which must be minimised for an optimal registration. The minimisation is efficiently performed using a multi-resolution implementation of the *highest confidence first* algorithm.

Keywords: medical imaging, 3D freehand ultrasound, volume estimation, registration

Contents

List of Figures

Acknowledgements

Thanks must go first and foremost to Richard Prager for enlightened guidance throughout three very enjoyable years; and to Lol Berman and David Lomas for a valuable clinical perspective. Many thanks also to Gonville and Caius College for support and shelter over the last 6 years, particularly to David Secher, Thomas Bligh, and Derek Ingram. The period of PhD research was made possible with a studentship granted by the EPSRC, complemented with full support from the Department of Engineering. The research presented in this dissertation was carried out under the auspices of EPSRC project GR/H74032 (Stradivarius).

Much gratitude to Gene Golub, Thomas Yu, Marcus Grote, and Sabine Vermeer of SCCM-Stanford for help with numerical methods and great hospitality in Stanford; and to colleagues at IPMI for discussions about the Oster-Murray model. Special thanks to fellow workers in the Fallside Lab: Chris, Rob, Rudy, Marty, George, Andrew, Patrick, Jon, Tat-Jen, and others, for nurturing and sharing an excellent research environment. Finally to family and friends, without whom these past six years would have been far less fulfilling.

Declaration

This dissertation is the result of my own original work, except where reference is made to the work of others. No part of it has been submitted for any other University degree or diploma.

Glossary

Growth model

$\mathbf{x}$	spatial position (Cartesian)
s	spatial position (arc-length)
t	time variable
$\mathbf{u}$	spatial displacement (due to ECM deformation)
c	(displacement of) cell concentration
S	(notates) organ surface
V	(notates and quantifies) organ volume
m	organ mass
$m^{\max}$	adult organ mass
D^{el}	elastic constant (for ECM deformation)
D^{dn}	diffusion constant (for cell diffusion)
$\mathbf{J}$	mass flux
$\mathbf{J}^{\mathrm{dn}}$	mass flux (due to cell diffusion)
$\mathbf{J}^{\mathrm{cv}}$	mass flux (due to cell convection)
p	hydrostatic pressure
$h_i(c_1, \ldots, c_n)$	reaction functions
h^{cell}	cell proliferation rate
h^{gz}	cell proliferation rate (Gompertz)
y^{gz}	Gompertz growth function
$a^{\mathrm{gz}}, b^{\mathrm{gz}}, k^{\mathrm{gz}}$	Gompertz function parameters
R^{gz}	Gompertz specific growth rate
W^{gz}	Gompertz growth energy
W^{ib}	inhibition energy during organ growth
$\mathbf{f}^{\mathrm{el}}(\mathbf{x}), f^{\mathrm{el}}(\mathbf{x})$	spatial eigenfunctions (for ECM deformation)
$f^{\mathrm{dn}}(\mathbf{x})$	spatial eigenfunctions (for cell diffusion)
$f^{\mathrm{cv}}(\mathbf{x})$	spatial eigenfunctions (for cell convection)
$g^{\mathrm{el}}(t)$	temporal modulation (for ECM deformation)
$g^{\mathrm{dn}}(t)$	temporal modulation (for cell diffusion)
$g^{\mathrm{cv}}(t)$	temporal modulation (for cell convection)

Finite element model

cs	centroid size of shape model
x, y, z	global Cartesian coordinate system
θ_1, θ_2	global spherical coordinate system
r, s, t	local coordinate system in a finite element
$\mathbf{J}$	Jacobian matrix of transformation between (x, y, z) and (r, s, t)
$\mathbf{M}$	mass matrix
$\mathbf{K}$	stiffness matrix
$\boldsymbol{\phi}_i$	the ith eigenmode
$\boldsymbol{\Phi}$	eigenmode matrix (columns are $\boldsymbol{\phi}_i$)
ω_i	natural-frequency of $\boldsymbol{\phi}_i$
$\boldsymbol{\Omega}^2$	eigenvalue matrix (diagonal elements are ω_i^2)
$\boldsymbol{\pi}_i$	spatial-frequency of $\boldsymbol{\phi}_i$
$h(r, s, t)$	polynomial interpolation function in an element
g_1, g_2, g_3, g_4	polynomial coefficients of $h(r, s, t)$ for a tetrahedral element
$\mathbf{G}$	geometry matrix
$\mathbf{H}$	displacement interpolation matrix
$\mathbf{B}$	strain-displacement interpolation matrix
$\mathbf{E}$	elasticity matrix

E	Young's modulus
ν	Poisson's ratio
W^{el}	elastic potential or energy of deformed body
$\mathbf{u}$	displacement components for body nodes
$\mathbf{u}^{\mathrm{n}}_j$	displacement components for body node j
$\mathbf{u}^{\mathrm{e}}_j(r,s,t)$	displacement function for element j
$\mathbf{u}^{\mathrm{e}}_j$	displacement components for nodes of element j
$\mathbf{r}^{\mathrm{V}}$	volume forces (per unit volume)
$\mathbf{r}^{\mathrm{S}}$	surface forces (per unit surface area)
$\mathbf{r}^{\mathrm{n}}$	point forces acting on body
$\mathbf{f}, \mathbf{f}(t)$	static and dynamic loading of body
$\mathbf{f}^{\mathrm{V}}$	volumetric component of $\mathbf{f}$
$\mathbf{f}^{\mathrm{S}}$	surface component of $\mathbf{f}$
$\mathbf{f}^{0}$	component of $\mathbf{f}$ due to initial loading
$\mathbf{f}^{\mathrm{n}}$	nodal components of $\mathbf{f}$
$\boldsymbol{\tau}, \boldsymbol{\epsilon}$	stress and strain vectors
$\overline{\boldsymbol{\tau}}, \overline{\boldsymbol{\epsilon}}$	stress and strain vectors (virtual)
$\tau_{xx}, \tau_{yy}, \tau_{zz}$	stress components (planar)
$\tau_{xy}, \tau_{yz}, \tau_{zx}$	stress components (shear)
$\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}$	strain components (planar)
$\epsilon_{xy}, \epsilon_{yz}, \epsilon_{zx}$	strain components (shear)

Non-linear finite element model

$\tilde{\mathbf{M}}$	mass matrix in relative coordinates
$\tilde{\mathbf{K}}$	stiffness matrix in relative coordinates
$\tilde{\mathbf{K}}^{\mathrm{nl}}$	non-linear stiffness matrix in relative coordinates
$\tilde{\boldsymbol{\phi}}_i$	the ith non-linear eigenmode in relative coordinates
q_i	amplitude of $\tilde{\boldsymbol{\phi}}_i$
$\mathbf{z}$	relative displacement components between body nodes
$\mathbf{f}(\mathbf{u})$	non-linear restoring forces
$\tilde{\mathbf{f}}(\mathbf{z})$	non-linear restoring forces in relative coordinates

Statistical model

$\mathbf{x}$	shape vector
$\boldsymbol{\phi}_i$	the ith principal component vector
$\boldsymbol{\Phi}$	principal component matrix (columns are $\boldsymbol{\phi}_i$)
λ_i	the ith principal variance (also Lagrange multiplier)
$\boldsymbol{\Lambda}$	principal variance matrix (diagonal elements are λ_i)
$\mathbf{S}$	sample covariance matrix of shape vectors $\mathbf{x}$
$\boldsymbol{\Sigma}$	population covariance matrix
σ_i	diagonal elements of $\boldsymbol{\Sigma}$

Homology map

$U^{\mathrm{pr}}, P^{\mathrm{pr}}$	prior energy, prior probability
$U^{\mathrm{lk}}, P^{\mathrm{lk}}$	likelihood energy, likelihood probability
$U^{\mathrm{ps}}, P^{\mathrm{ps}}$	posterior energy, posterior probability
G	stability measure of a homology
s_1, s_2	index of site on organ models

Symmetric eigenvalue problem

$\mathbf{0}$	null matrix or vector
$\mathbf{I}$	identity matrix
$\mathbf{A}$	real symmetric matrix whose eigenpairs are required
$R(\mathbf{u}), R(\mathbf{A})$	Rayleigh quotient of vector $\mathbf{u}$ or matrix $\mathbf{A}$
μ	origin shift
$\mathbf{q}_i$	ith orthonormalised vector from QR factorisation
$\mathbf{Q}$	matrix of column vectors $\mathbf{q}_i$
$\mathbf{R}$	upper-triangular factor from QR factorisation
$\hat{\mathbf{A}}$	QR transform of $\mathbf{A}$
$\mathcal{E}^n$	n-dimensional Euclidean space
$\mathcal{S}^n$	n-dimensional subspace
$\boldsymbol{\phi}_i$	ith eigenvector of $\mathbf{A}$
λ_i	ith eigenvalue of $\mathbf{A}$
$\boldsymbol{\psi}_i^{\mathrm{R}}$	ith eigenvector of (matrix) Rayleigh quotient $R(\mathbf{A})$
$\boldsymbol{\phi}_i^{\mathrm{R}}$	ith Ritz vector of $R(\mathbf{A})$
λ_i^{R}	ith Ritz value of $R(\mathbf{A})$
$\mathbf{r}_i$	residual error at ith iteration
$\mathbf{x}_i, \mathbf{y}_i, \mathbf{X}_i, \mathbf{Y}_i$	utility vectors and matrices during ith iteration
$\mathbf{K}^n(\mathbf{x})$	nth-order Krylov matrix formed by $\mathbf{A}$ and a vector $\mathbf{x}$
$\mathcal{K}^n(\mathbf{x})$	Krylov subspace spanning $\mathbf{K}^n(\mathbf{x})$
$\mathbf{T}_n$	tridiagonal matrix at nth Lanczos iteration
α_i, β_i	ith diagonal and off-diagonal elements of $\mathbf{T}_n$

Operators

$\dot{y}, \ddot{y}$	first and second time derivatives of y
$\|\mathbf{A}\|, \|\mathbf{A}\|_F$	2-norm and Frobenius norm of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}, \mathbf{A}^T$	inverse and transpose of matrix $\mathbf{A}$
$\nabla = (\frac{\partial}{\partial x}\ \frac{\partial}{\partial y}\ \frac{\partial}{\partial z})^T$	notates gradient and divergence operations
$\det \mathbf{A}$	determinant of matrix $\mathbf{A}$
span $\mathbf{A}$	subspace spanned by columns of matrix $\mathbf{A}$

Abbreviations

DOF	degree(s) of freedom
ECM	extra-cellular matrix
FEM	finite element method
SAT	symmetric axis transform
SLS	smoothed local symmetries
PDE	partial differential equation
PCA	principal component analysis
PDM	point distribution model
DFT	discrete Fourier transform
MRF	Markov random field
HCF	highest confidence first
ICM	iterative conditional modes
GNC	graduated non-convexity
MAP	maximum *a posteriori*
RR	Rayleigh-Ritz
MRI	magnetic resonance imaging
CT	computed tomography

Chapter 1

Introduction

This chapter introduces the concepts behind medical ultrasound imaging, and motivates it as a cheap, safe, and portable imaging modality, routinely used for non-invasive monitoring and diagnosis in many clinical applications.

During conventional 2D ultrasound scanning, the sonographer visualises anatomy by mentally integrating the estimated motion of the hand-held scanner probe with the sequence of ultrasound images. These images are usually noisy and distorted, and tissue regions are not easily discriminated from a single image, so an experienced sonographer overcomes these problems by employing prior knowledge of the expected anatomy to aid visualisation.

The aim of this dissertation is to reproduce this strategy using a *3D freehand ultrasound* system, which consists of a conventional scanner with a position sensor attached to its probe. The simple configuration enables the sonographer to operate the scanner normally, while the position information associated with each scan plane allows the set of ultrasound images to be accurately located in 3D, so that an anatomical model can be applied to their interpretation.

Chapter organisation

- **Section 1.1** contains a brief overview of the processes involved in producing 2D and 3D ultrasound images.
- **Section 1.2** surveys the relevant clinical and medical imaging literature, and describes the efficacy of 3D freehand ultrasound imaging.
- **Section 1.3** motivates the use of an organ shape model to constrain the interpretation of anatomy from noisy 3D ultrasound images.
- **Section 1.4** summarises the salient themes and contributions of the material presented in this dissertation.

1.1 Ultrasound imaging

Imaging in 2D

An ultrasound probe contains a piezo-electric transducer generating short sonic pulses, which propagate through tissue at an average speed of 1540 ms^{-1}. Back-scattered pulses are received at the same transducer and converted into an electrical signal, which is selectively amplified to compensate for attenuation in back-scattered pulses emerging from deeper tissue. The signal envelope can then be plotted against depth (determined by elapsed propagation time and approximate propagation speed) to give a 1D scan.

In 2D scanning, an array of such transducer elements produces a signal wave front, which is shaped by appropriately phased interference between each transducer element; the 2D image is then constructed from the array of 1D scans. Chervenak et al. (ch.1–2)[25] provide a comprehensive reference to the physics of the imaging process.

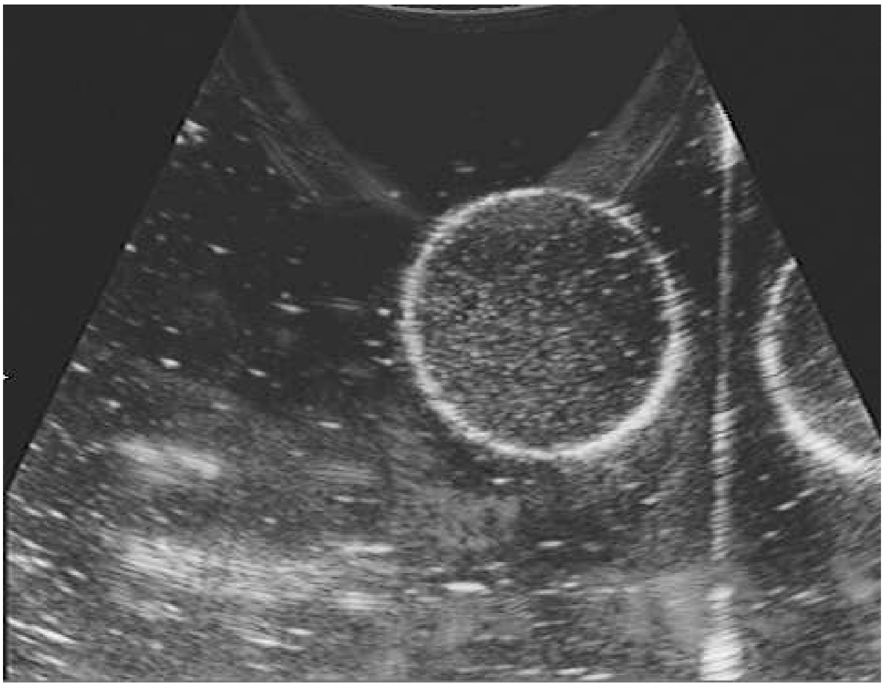


Figure 1.1: 2D ultrasound image of a balloon in a water bath, showing a reflection against the side of the bath on the right, and speckling artefacts throughout

Three important causes of error in ultrasound imaging are: speckle, distortion, and refraction. Ultrasound image 'textures' can be thought of as being formed from a multitude of point scatterers in the tissue[1], and the coherent nature of back-scattered pulses causes interference effects – *speckling* artefacts – throughout Figure 1.1.

As mentioned previously, depth is only approximately determined by propagation time since the propagation speed varies with tissue type and temperature, inevitably *distorting* the 2D image reconstructed by the scanner. The differing speeds of neighbouring tissue types also cause *refraction* during beam propagation, which leads to the incorrect placement of structures in the ultrasound image[2].

[1]For example, air bubbles inside the balloon in Figure 1.1 give rise to visible texturing.

[2]Chapter 5.6 describes how image distortion may give rise to volume estimation errors in ultrasound imaging.

Imaging in 3D

The acquisition rate of 3D ultrasound imaging systems is limited by the finite propagation speed of ultrasound pulses, and real-time systems suitable for clinical use are not currently available. For example, the experimental "Explososcan" approach uses a grid of transducers to transmit a wide-angle pulse, but real-time 3D reconstruction is feasible only if information from transducer elements is processed in parallel (Wells [166]).

If the requirement for real-time imaging is relaxed, then 3D ultrasound imaging is achieved either by a special probe which mechanically sweeps over a pre-defined acquisition volume (Merz et al. [92]), or by attaching some form of position sensor to a standard probe. Early implementations of the latter approach utilised a mechanical arm to carry and track the probe (Ohbuchi et al. [106]); alternatively, Trobaugh et al. [160] describe an accurate but expensive system which uses multiple cameras to track light-emitting diodes on the probe[3].

Currently the most affordable and flexible approach to 3D ultrasound imaging is to monitor the probe's six degrees-of-freedom (DOF) in translation and rotation, using a small electromagnetic sensor. Probe motion is usually defined relative to the axes of the probe's scan plane, and Figure 1.2 shows the common labels given to these axes.

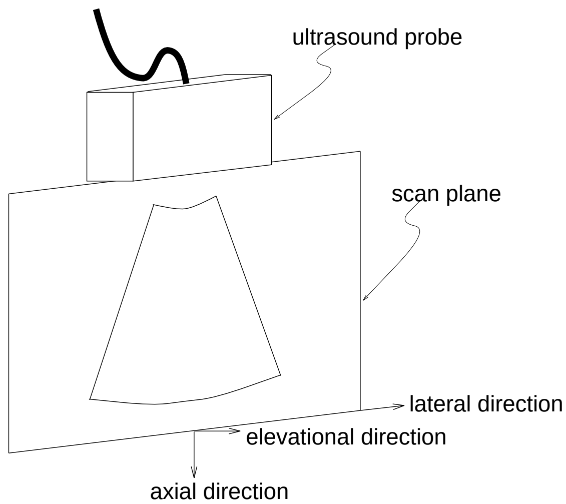


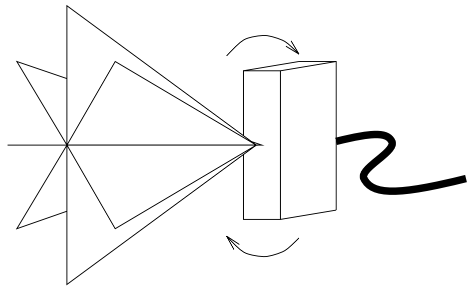
Figure 1.2: Naming conventions for scanning directions of an ultrasound probe

Probe movement can be restricted to a single DOF for cone, fan, or prism-shaped sweeps (see Figure 1.3), since there are advantages to each swept volume. For example, the prismatic volume gives parallel images which are easy to render in 3D; and the fan-shaped and cone-shaped sweeps require only a small acoustic window from which to view an anatomical volume.

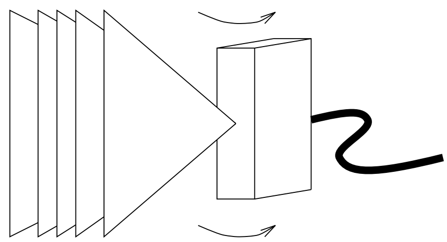
If probe movement is completely unrestricted, then *3D freehand imaging* allows the sonographer to operate the probe conventionally while its six DOFs are continuously monitored and recorded. This is the approach taken by the Stradivarius project (Dance et al. [37]), whose equipment configuration is summarised in Figure 1.4. All ultrasound data used in this dissertation is derived from this equipment.

Section 1.2 now places the problems posed by freehand imaging in the context of current research in 3D ultrasound, after which Section 1.3 justifies this dissertation's approach to addressing these problems.

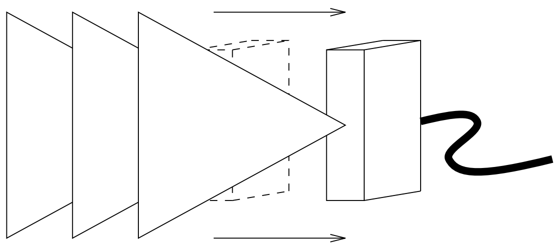
[3]This requires that the probe not be occluded from view of the cameras.



(a) Cone-shaped sweep by rotating in axial direction



(b) Fan-shaped sweep by rotating in lateral direction



(c) Prism-shaped sweep by translating in elevational direction

Figure 1.3: Sweeping through an anatomical volume by restricting probe motion to a single axis

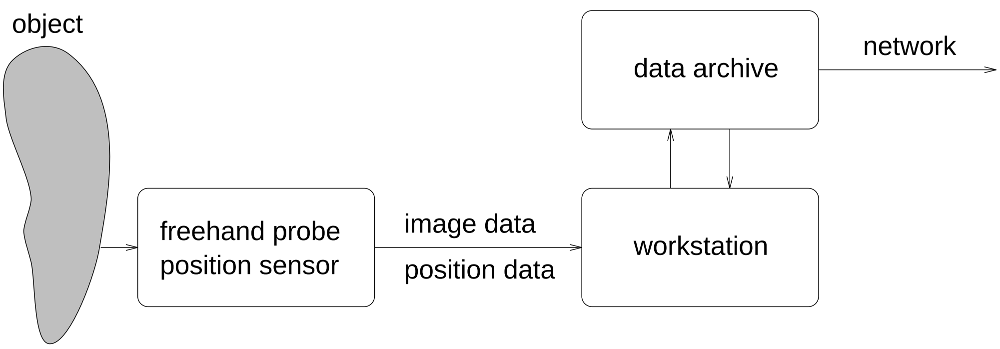


Figure 1.4: Hardware configuration for the Stradivarius project

1.2 Survey and motivation

We now survey relevant work by other researchers in 3D ultrasound imaging. Much of it is currently directed at visualisation (discussed further in Chapter 5) and noise-reduction of ultrasound images, while methods of tissue discrimination are still largely unreliable.

A number of publications describe the clinical potential of 3D ultrasound imaging for: the urethra (Ng et al. [105]), vascular structure (Klein et al. [74]), gallbladder (Fine et al. [40]), breast (Moskalik et al. [95]), kidney (Rankin et al. [121]), and uterus (Balen et al. [4]). Applications in obstetrics (Steiner et al. [141]) and echocardiography (Salustri & Roelandt [127]) attract particular interest, due to their clinical importance.

Specialist applications for 3D freehand ultrasound are proposed by Scott [131], who intends to guide a biopsy needle into breast lesions by matching the needle's (sensed) position with a 3D reconstruction of the lesion; and by Cena et al. [23, 168], who aim to monitor scar healing in burns victims, and to reconstruct skin grafts for plastic surgery.

Ultrasound image processing

Speckle suppression is one of the most important elements of noise reduction in ultrasound images, and adaptive filters for this purpose are described by Karaman et al. [70] for 2D images, and by Bamber et al. [6] for 3D volume rendered images. Pasterkamp et al. [111] average out speckle in flowing blood to increase contrast for 3D segmentation and visualisation of the arterial lumen; and Ashton & Parker [2] subsample ultrasound images to achieve Gaussian noise statistics, which are more easily estimated and filtered.

The frequency spectrum of speckle is similar to the scanner's modulation transfer function, so speckle noise is passed straight through the imaging system, potentially obscuring diagnostically important information. Speckle patterns become uncorrelated if a region is viewed from different look directions, however, and the signal-to-noise ratio is expected to increase in proportion to the square root of the number of compounded views (Shankar [133]).

Compounding can be performed spatially or temporally, at the expense of spatial resolution. Ohbuchi et al. [106] perform spatial compounding of 2D image planes during incremental volume reconstruction of 3D ultrasound images; and temporal compunding is commonly implemented by image persistence on the ultrasound system monitor.

Metcalfe & Evans [93] examine a range of ultrasound scanners and transducers, and find that scanner parameters require different settings[4] to optimise image contrast when scanning kidney, uterus, and liver. Reliable discrimination of tissues in ultrasound images therefore requires texture features which are invariant to operator settings; it also requires the texture features to be specialised for different tissue types. For example, Muzzolini et al. [98, 99] discriminate textures by statistical classification, Basset et al. [7] analyse prostate images using co-occurence matrices, and Wu et al. [169] classify liver textures using fractal features.

Nastar & Ayache [101] suggest that ultrasound image processing can be improved by analysing radial 1D scanlines directly in polar coordinates, before they are interpolated to give the 2D Cartesian display. Unfortunately, access to pre-interpolated images is not available for most scanners.

Rohling & Gee [122] provide an overview of technical issues in 3D freehand ultrasound imaging, as do more general reviews by Rankin et al. [121] and Sohn [135]. Ayache [3] surveys the problems involved in automatic interpretation of medical images, in the context of current research in machine vision.

[4] Individual ultrasonographers also have different preferences for scanner parameters.

Harris et al. [60] examine ultimate limits in the resolution of ultrasound images: higher ultrasonic frequency and intensity increase spatial resolution, but both are limited by safety considerations, non-linear attenuation effects, and tissue inhomogeneity. Joint optimisation of ultrasonic frequency with probe design is proposed, to give a compact beam which achieves a good frame rate at good spatial resolution.

Although 3D ultrasound imaging has great clinical potential, this brief survey shows that its utility is significantly limited by image noise problems. We should therefore justify persevering with a 3D freehand generalisation of 2D ultrasound imaging.

Motivation for 3D freehand ultrasound

3D medical imaging is primarily motivated by the fact that human anatomy is defined in 3D, since isolated 2D slices or projections cannot capture the complex relationships between anatomical structures of clinical interest. For example, inferring 3D structure from parallel 2D image slices in *magnetic resonance imaging* (MRI) and *computed tomography* (CT) is a routine task for clinicians, and segmented information from these modalities is very useful for biometric and morphological studies, surgery planning, and manufacturing implants etc.

2D ultrasound imaging is cheaper, safer, and more portable than either CT and MRI, and is also routinely used for organ examination. For example, ventricular volume and wall motion determine the degree of impairment after heart attacks, and indicate response to therapy; similarly, changes in tumour volume determine the nature and intensity of further treatment. Ultrasound imaging is especially useful in obstetric practice, since it is the only imaging modality safe for use in foetal monitoring (see Chapter 5.6 for a discussion of 3D foetal imaging).

3D freehand ultrasound is a straightforward extension of 2D ultrasound, which still allows the sonographer to scan in a manner consistent with normal examination. The flexibility of freehand scanning is important, because some parts of the anatomy are visible only from small acoustic windows (e.g. viewing the spleen through ribs), and regions shadowed by strongly echogenic structures (e.g. bone-tissue interfaces) need to be viewed from multiple directions. Unfortunately, this flexibility results in non-parallel and (potentially) intersecting scan planes.

1.3 Model-based interpretation

Reconstructing 3D anatomy from a sequence of freehand images is therefore a difficult task, particularly since the images are noisy, occluded, and distorted. A sonographer performs this reconstruction by employing a mental model of expected anatomy to guide the interpretation of each image, which has its position estimated by the hand controlling the probe. This suggests that we require a similar model-based strategy for interpreting 3D freehand ultrasound images.

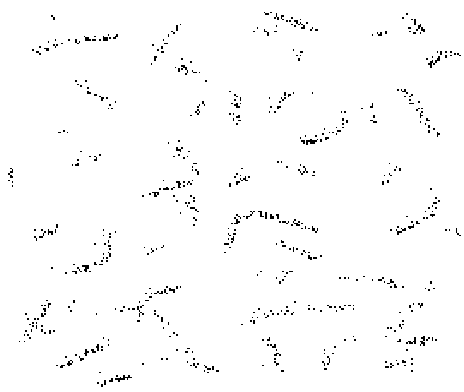


Figure 1.5: Model-based interpretation of a noisy image is a more reliable strategy than purely data-driven interpretation (adapted from Cham [24])

Figure 1.5 offers an instructive example: the perceptual primitives are edge elements, which the human visual system automatically clusters together from the noisy speckle. Other cues, such as symmetry and geometric saliency of locally-curved edges, give further indication of the global organisation of primitives (there may be many). For simple images, such a *data-driven* strategy for image interpretation should conclude with a single most probable segmentation.

However, ultrasound images are much noisier and so require additional *model-based* knowledge to be brought to bear on their visual interpretation (Chervenak et al. (ch.5)[25]). To demonstrate: if told that Figure 1.5 shows a key with axis orientated top-right to bottom-left, the key's silhouette soon becomes apparent as the most probable segmentation of the image.

Unfortunately, perception of 3D shape from freehand ultrasound images is complicated by the requirement that image features be integrated over non-parallel scan planes. The way to organise such information over all the planes, while tolerating image noise, distortion, and occlusion, as well as errors in frame position measurement (Gosling et al. [56]), is to incorporate knowledge of expected shape and shape variation into an organ model.

Note that fully-automated segmentation in 3D freehand ultrasound requires proprioceptive accuracy, solutions to imaging noise and distortion, and methods of reliable tissue discrimination. Such problems are beyond the scope of this dissertation, and are currently being addressed by co-workers in the Stradivarius project: Scott [131]; Dance [36]; Rohling & Gee [122].

Instead, the focus of this dissertation is on applying a model-based framework to 3D freehand ultrasound imaging. Such a framework must incorporate knowledge of organ shape, while allowing for an expected range of variations in shape. It must also be able to learn from new examples of the organ, thus refining its expectation of organ shape variation.

There should be a strong geometric basis to model representation, since segmentation is usually motivated by the analysis of organ biometry and morphology. The organ's intrinsic coordinate system must not be parameterised in an *ad hoc* fashion, however, since learning from new examples – each sampled as a 'shape vector' – requires them all to be described in a mutually consistent shape parameterisation.

1.4 Prospectus

Themes and contributions

The salient contributions in this dissertation are biological models of the mechanisms responsible for organ growth and development:

- Oster-Murray mechanisms are proposed to describe an organ's *cellular growth* patterns;
- the Gompertz model is proposed to describe the net energy required for *organ growth.*

One important theme is the use of eigenfunctions of the Oster-Murray mechanisms. These eigenfunctions describe all possible patterns of mass flux during organ growth, while the eigenfunctions' shapes are also strongly reflective of the underlying organ shape.

We require an organ shape model to guide the interpretation of 3D freehand ultrasound images: a computer application makes this available through a geometric tool for operator-assisted image segmentation. We also require organ shape models to learn from examples: in order to compare two organ shapes (so as to learn about their differences), a method of registration is required to describe them in a mutually consistent shape coordinate system.

Since eigenfunctions intrinsically constitute such a coordinate system, two models can be fitted to each other by geometrically matching their eigenfunctions. The optimal match is computed using the Gompertz metric to ensure its biological plausibility.

Thesis organisation

An introduction to content and organisation is provided on the first page of each chapter, with a brief summary at the end of each chapter. The chapters and appendices are as follows:

- **Chapter 2: Biological Morphogenesis and Growth**
  Three important mechanisms from the Oster-Murray model are introduced. Patterns of biological growth and deformation are described by the mechanisms' eigenfunctions.

- **Chapter 3: Modelling Shape**
  Finite element models are used to represent organ shape, since Oster-Murray eigenfunctions can be approximately computed for such models to give finite element eigenmodes.

- **Chapter 4: Modelling Variation in Shape**
  Organ shape models can therefore initialise their knowledge of shape variation using eigenmodes. This knowledge is refined by incorporating new examples of organ shape.

- **Chapter 5: Experiments: 3D Volume Estimation**
  A computer application is described which uses these shape models to perform operator-assisted organ volume estimation from 3D ultrasound images.

- **Chapter 6: Likelihood of Homology**
  Two similar organ models can be fitted to each other by comparison of their eigenmodes. Such a fit is shown to be biologically justified.

- **Chapter 7: Coherence of Homology**
  The map of corresponding points (*homologies*) between two fitted organ models can be made globally coherent, by ensuring its consistency with Gompertzian growth.

- **Chapter 8: Experiments: 3D Registration**
  The highest confidence first (HCF) algorithm computes an optimal homology map using eigenmode features from Chapter 6, and the Gompertz constraint from Chapter 7.

- **Chapter 9: Discussion and Conclusions**
  Material presented in this dissertation is reviewed and discussed, and suggestions are made for possible future research.

- **Bibliography**
  Citations made in this dissertation are listed in alphabetical order.

- **Appendix A: Finite Element Method**
  Procedures are described for assembling a finite element model's mass and stiffness matrices.

- **Appendix B: The Symmetric Eigenproblem**
  Techniques are described for computing finite element eigenmodes from mass and stiffness matrices.

Chapter 2

Biological Morphogenesis and Growth

This chapter examines three mechanisms of biological shape formation (*morphogenesis*):

(i) the *deformation* of the cellular matrix in which cells are embedded;

(ii) cell proliferation, and the resultant *diffusion* of cells into areas of lower concentration;

(iii) the *convection* of embedded cells when the cellular matrix deforms.

These processes form an integral part of the Oster-Murray model of pattern formation, in which tissue boundaries are fixed so that pattern emergence can be analysed. However, the analysis in this chapter allows tissue boundaries to change shape as a consequence of mass flux in the organ. Variations in organ shape can, in this way, be modelled from a biological perspective (Syn & Prager [146]).

The spatial *eigenfunctions* of these three mechanisms, when computed for some underlying organ shape, form a complete description of all potential patterns of mass flux. In other words, these eigenfunctions are a description of organ shape variation. Organ tissues have material properties which cannot be easily established, so their spatial uniformity is assumed during this chapter's analysis, with the consequence that all three growth mechanisms are found to have identical spatial eigenfunctions.

Chapter organisation

- **Section 2.1** introduces the classic Turing model of biological pattern formation.
- **Section 2.2** introduces the more general Oster-Murray model of biological pattern formation.
- **Section 2.3** provides the biological background to a proposed growth model, which applies the Oster-Murray mechanisms to biological growth (rather than pattern formation).
- **Section 2.4** analyses the patterns of mass flux during cellular matrix deformation.
- **Section 2.5** analyses the patterns of mass flux during cell proliferation and diffusion.
- **Section 2.6** analyses the interaction between the deforming matrix and the cell population.

2.1 The Turing mechanism

Turing [161] studied pairs of reaction-diffusion equations, generally formulated as

$$\frac{\partial c_i}{\partial t} = D_i \nabla^2 c_i + h_i(c_1, \ldots, c_n) \qquad\qquad i = 1, \ldots, n \tag{2.1}$$

Each chemical *morphogen* has a concentration c_i, a constant diffusivity D_i, and a rate of diffusion determined by the concentration gradient $\nabla^2 c_i$. Reactions between morphogens cause each one to change concentration by h_i.

Under certain conditions, it was found that uniform solutions to these equations were unstable to small peturbations, resulting in stable but non-uniform patterns; such *symmetry-breaking* processes are thought to underlie biological pattern formation.

For example, Figure 2.1(a) shows pattern emergence in Young's [172] system, with (black) areas of high concentration clustering as the system evolves; and Figure 2.1(b) shows the final patterns achieved with different system parameters, all of which are reminiscent of animal coat patterns. These suggest that relatively simple systems can give rise to the complex patterns and forms seen in biological morphogenesis (Levin & Segel [80]).

(a) Pattern evolution

(b) Final patterns for different system parameters

Figure 2.1: Pattern emergence in Young's activator-inhibitor system (Prusinkiewicz [117])

Morphogen gradients

Meinhardt & Klingler [91] advanced similar ideas to Turing, emphasising the combination of short-range activation (reaction) and long-range inhibition (diffusion), and pointing out that gradients of morphogens do not require pre-arranged sources and sinks. Gierer's [48] identification of morphogens in *hydra* confirmed the use of such gradients in controlling head and foot regeneration.

Wolpert [167] proposed that cells react to threshold levels of morphogen concentration. This allows control of cell differentiation at desired locations, and was demonstrated in experiments on embryo chicken wings.

More recently, Hunding & Engelhardt [67] used Turing's mechanism to model stripe prepatterning and head formation in insect embryogenesis; morphogen gradients were again the source of positional information. They argued that a control system which is able to extract reliable positional information from a morphogenetic gradient, is also prone to yield pattern formation by symmetry-breaking processes.

2.2 The Oster-Murray model

Turing-type models show that a chemical pre-pattern can be established, which then activates the differentiation and growth of biological structures. Thus morphogenesis is treated as a slave process which gives a pre-determined result once the chemical pre-pattern has been established.

Oster & Murray [107] contend instead, that the pattern formation process is mainly determined by mechanical forces which arise during shape changes in growing tissue. The Oster-Murray approach differs from the chemical pre-pattern approach, in that pattern formation and morphogenesis proceed simultaneously using an array of biological mechanisms.

One important benefit of simultaneous development is that the mechanisms can be self-correcting, whereas Turing's process is effectively an open-loop system (Murray et al. [97]). In addition, Goodwin et al. [53] show that the close coupling of different mechanisms drastically reduces the number of degrees-of-freedom the coupled system has, when converging to a stable state. This results in a morphogenetic process which is robust to noisy parameters.

Mechano-chemical processes

The Oster-Murray approach considers the coordinated patterning of *mesenchymal* and *epithelial* cells, which are two types of early embryonic cells (Moore [94]; Trinkhaus [159]). Where standard Turing-type models involve the interaction of hypothetical morphogens, the Oster-Murray model is centred on the cell as the fundamental unit of biological growth. The model incorporates the effects of:

(i) cell convection;

(ii) short and long range diffusion of cells;

(iii) cell proliferation rate;

(iv) contact inhibition and guidance by neighbouring cells;

(v) haptotaxis (cell movement up an adhesive gradient);

(vi) galvanotaxis (cell movement influenced by electric potentials);

(vii) chemotaxis (cell movement influenced by chemical concentration gradients).

In other words, the cell flux term $\mathbf{J}$ in the general cell-conservation equation[1]

$$\frac{\partial c}{\partial t} = -\nabla \cdot \mathbf{J} + h^{\text{cell}} \tag{2.2}$$

is no longer merely the localised diffusion in equation (2.1), but includes the mechano-chemical processes itemised above. Note that c now refers to cell rather than morphogen concentration; and that h^{cell} is a function describing cell proliferation, rather than a reaction expression as before in equation (2.1).

Murray et al. [97] and Murray (ch.17.3)[96] analyse a linearised model of these processes, and find a large class of spatially inhomogeneous solutions similar to Turing-type patterns. A numerical simulation by Savic [128] of animal coat patterning using a mechanical model, also gives very similar results to Meinhardt [90] who uses a reaction-diffusion model.

[1] See equations (2.9)–(2.11) for the derivation of the cell-conservation equation.

2.3 The model of biological growth

Murray [96] analyses the interaction of the Oster-Murray mechanisms during the formation of feathers, scales, and teeth, by fixing tissue boundary conditions so that temporally unstable waves within the boundaries can settle into stable patterned solutions.

However, we are interested in modelling growth and not merely pattern formation. Therefore, in our proposed growth model the tissue boundaries are no longer fixed, so that the Oster-Murray mechanisms can be analysed for their effects on tissue growth.

In the model, cell *proliferation* causes an increase in organ mass, which is primarily redistributed by three mechanisms: cell *diffusion*, matrix *deformation*, and matrix *convection* of attached cells; these mechanisms are analysed in Sections 2.4–2.6.

The significance of our growth model is that cells are recognised as the fundamental units of biological growth and shape change. Only the extra-cellular matrix and mesenchymal cells participate in the model; epithelial cells are important in shaping organ boundaries, but are not modelled explicitly.

The extra-cellular matrix

The extra-cellular matrix (ECM) forms part of the tissue within which cells move, and is approximately linear elastic for small strains. Murray (ch.17.2)[96] assumes ECM to be isotropic in its elastic properties, but presents a possible mechanism for inclusion of anisotropic effects. Figure 2.2 plots the stress-strain behaviour of ECM.

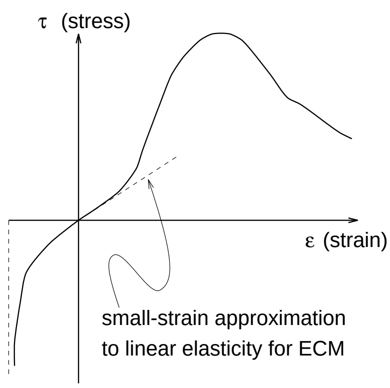


Figure 2.2: Stress-strain behaviour of extra-cellular matrix (ECM) (adapted from Murray [96])

Inertial effects in cell-ECM interaction can be ignored because it occurs in a very low Reynolds number[2] regime (Murray (ch.17.2)[96]; Purcell [118]).

The time scale of incremental organ development is very long (hours) and the spatial scale is very small (millimetres), so *viscous effects* within the ECM can also be neglected, since the viscous component of stress[3] is proportional to $\frac{\partial \boldsymbol{\epsilon}}{\partial t}$.

ECM is formed from the secretions of fibroblast cells, which are of mesenchymal origin. It is implicitly assumed in our growth model that these secretions act to relieve elastic strain in the ECM as growth progresses.

[2]The Reynolds number indicates the relative importance of inertial and viscous effects, with the latter highly dominant at cellular level.

[3]Stress $\boldsymbol{\tau}$ and strain $\boldsymbol{\epsilon}$ are defined in Appendix A.

Mesenchymal cells

Mesenchymal cells move by exerting traction forces on the (fibrous) ECM, and on the surface of other cells. These traction forces cause ECM deformation, which in turn causes convection of other cells embedded in the ECM.

Cells proliferate by subdivision, and the resulting growth in cell numbers disturbs local cell concentrations, to form concentration gradients which are equalised by cell diffusion.

Epithelial cells

Epithelial cells are not mobile, but spread or thicken in response to forces exerted by mesenchymal cells. They can be thought of as enforcing shape control over the growing mesenchymal cell population.

2.4 Analysis of ECM deformation

The *wave equation* governs the ECM's elastic behaviour, and on the assumption of uniform mass and stiffness properties[4] (hence constant D^{el}) it is

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = -D^{\text{el}} \nabla (\nabla \cdot \mathbf{u}) \tag{2.3}$$

which has a separation of variables solution for ECM displacement $\mathbf{u}$

$$\mathbf{u} = \mathbf{f}^{\text{el}}(\mathbf{x}) g^{\text{el}}(t) \tag{2.4}$$

In the absence of damping effects[5], $g^{\text{el}}(t)$ is a sinusoidal modulation, and the *elastic eigenfunctions* $\mathbf{f}^{\text{el}}(\mathbf{x})$ are approximated on an organ mesh by FEM eigenmodes (as will be shown in Chapter 4). So the spatial behaviour of any ECM deformation is a sum of vector displacement fields $\mathbf{f}^{\text{el}}(\mathbf{x})$.

Appendix A.7 shows that ECM deformation $\mathbf{u}$ gives rise to a pressure distribution p, governed by a scalar analogy to the wave equation (2.3)

$$\frac{\partial^2 p}{\partial t^2} = -D^{\text{el}} \nabla^2 p \tag{2.5}$$

which also has a separation of variables solution

$$p = f^{\text{el}}(\mathbf{x}) g^{\text{el}}(t) \tag{2.6}$$

In other words, there is a scalar eigenfunction $f^{\text{el}}(\mathbf{x})$ associated with every vector eigenfunction $\mathbf{f}^{\text{el}}(\mathbf{x})$; the latter describes the ECM's deformation, while the former describes the associated change in ECM pressure.

This association is important in Section 2.5, when a comparison is made between (scalar) cell-concentration displacement and (vector) ECM displacement.

[4] Appendix A.2 shows that D^{el} is proportional to stiffness over mass.

[5] Section 2.3 notes that viscous damping can be ignored for the ECM, but a Rayleigh damping model is used if such effects need to be incorporated. Rayleigh damping assumes the damping matrix to be a linear combination of the mass and stiffness matrices (see Appendix A).

2.5 Analysis of cell proliferation and redistribution

The model used for cell proliferation is the Gompertz function y^{gz} from equation (7.1) in Chapter 7. We require the *rate of cell proliferation* h^{gz}

$$\begin{aligned} h^{\mathrm{gz}}(t) &= \frac{d}{dt}\Big(y^{\mathrm{gz}}\Big) \\ &= \frac{d}{dt}\Big(ae^{-be^{-kt}}\Big) \end{aligned} \tag{2.7}$$

where superscripts are omitted from a^{gz}, b^{gz}, and k^{gz} for clarity.

Cell proliferation causes local gradients in cell concentration c, and such gradients give rise to local cell diffusion

$$\mathbf{J}^{\mathrm{dn}} = -D^{\mathrm{dn}}\nabla c \tag{2.8}$$

where D^{dn} is the diffusion coefficient, and $\mathbf{J}^{\mathrm{dn}}$ is the cell flux due to diffusion.

These expressions for diffusive flux $\mathbf{J}^{\mathrm{dn}}$ and cell proliferation $h^{\mathrm{gz}}(t)$ can now be incorporated into the general cell-conservation[6] equation (2.2)

$$\frac{\partial c}{\partial t} = h^{\mathrm{gz}}(t) + D^{\mathrm{dn}}\nabla^2 c \tag{2.12}$$

Equation (2.12) governs cell proliferation and diffusive redistribution, and has the solutions

$$c = y^{\mathrm{gz}}(t) + f^{\mathrm{dn}}(\mathbf{x})g^{\mathrm{dn}}(t) \tag{2.13}$$

Here $g^{\mathrm{dn}}(t)$ is a modulation function describing exponential decay, and the *diffusion eigenfunctions* $f^{\mathrm{dn}}(\mathbf{x})$ are scalar fields describing the spatial behaviour of cell-concentration c.

Comparing elastic and diffusion eigenfunctions

We wish to compare the spatial behaviour of the diffusing cells with the spatial behaviour of ECM deformation, but:

- cell-concentration c is a scalar field which has diffusion eigenfunctions $f^{\mathrm{dn}}(\mathbf{x})$;
- ECM deformation $\mathbf{u}$ is a vector field which has elastic eigenfunctions $\mathbf{f}^{\mathrm{el}}(\mathbf{x})$.

However, we know from Section 2.4 that ECM deformation $\mathbf{u}$ is associated with a scalar pressure field p which has the eigenfunctions $f^{\mathrm{el}}(\mathbf{x})$; we can therefore compare pressure ('mass-concentration') directly with cell-concentration.

[6] Let the organ volume V be bounded by its surface S. Conservation of cell mass implies that the rate of change of cell mass in V is equal to the rate of flow of cell mass across S into V, including the rate of cell mass production in V

$$\frac{\partial}{\partial t}\int_V c\, dV = -\int_S \mathbf{J}\cdot d\mathbf{S} + \int_V h^{\mathrm{cell}}(t)\, dV \tag{2.9}$$

Applying the divergence theorem to the surface integral and assuming c is continuous, equation (2.9) becomes

$$\int_V \left(\frac{\partial c}{\partial t} + \nabla\cdot\mathbf{J} - h^{\mathrm{cell}}(t)\right) dV = 0 \tag{2.10}$$

The integrand must be zero since organ volume V is arbitrary, which gives the general cell-conservation expression from equation (2.2)

$$\frac{\partial c}{\partial t} + \nabla\cdot\mathbf{J} = h^{\mathrm{cell}}(t) \tag{2.11}$$

Substituting for p from equation (2.6) to equation (2.5) gives

$$f^{\mathrm{el}}(\mathbf{x})\ddot{g}^{\mathrm{el}}(t) = D^{\mathrm{el}}\nabla^2 f^{\mathrm{el}}(\mathbf{x})g^{\mathrm{el}}(t) \tag{2.14}$$

and separating the spatial and temporal components gives

$$D^{\mathrm{el}}\frac{\nabla^2 f^{\mathrm{el}}}{f^{\mathrm{el}}}(\mathbf{x}) = \frac{\ddot{g}^{\mathrm{el}}}{g^{\mathrm{el}}}(t) = -k^{\mathrm{el}} \qquad (k^{\mathrm{el}} \text{ arbitrary constant}) \tag{2.15}$$

Doing the same with cell growth and redistribution in equations (2.12) and (2.13) gives

$$D^{\mathrm{dn}}\frac{\nabla^2 f^{\mathrm{dn}}}{f^{\mathrm{dn}}}(\mathbf{x}) = \frac{\dot{g}^{\mathrm{dn}}}{g^{\mathrm{dn}}}(t) = -k^{\mathrm{dn}} \qquad (k^{\mathrm{dn}} \text{ arbitrary constant}) \tag{2.16}$$

Principal modes of ECM mass flux and cellular mass flux

Kreysig (ch.11.3)[77] shows that the same boundary conditions will constrain k^{el} and k^{dn} to give the same spatial eigenvectors, albeit with different eigenvalues[7]. Since both ECM and cellular distributions are constrained by the same epithelial boundaries, this means that f^{el} in equation (2.15) and f^{dn} in equation (2.16) are identical scalar eigenfunctions. Morever:

- ECM pressure (scalar field) is associated with ECM deformation (vector field);
- cell-concentration (scalar field) is similarly associated with cell diffusion (vector field);

so the vector eigenfunctions of both are also identical. In other words, the elastic eigenfunctions $\mathbf{f}^{\mathrm{el}}(\mathbf{x})$ can be used to describe the principal spatial modes of both ECM mass flux and cellular mass flux.

2.6 Analysis of cell-ECM interaction

The Oster-Murray model considers convection of cells by the deforming ECM to be a major mode of cell transport. Convection occurs concurrently with the mechanisms described in the Sections 2.4–2.5, but inertial effects in cell-ECM interaction are ignored for reasons mentioned in Section 2.3.

A model is now developed of interaction between the deforming ECM and the passively transported cell population, using the mass flux due to convection

$$\mathbf{J}^{\mathrm{cv}} = c\frac{\partial \mathbf{u}}{\partial t} \tag{2.17}$$

Incorporate this into the general cell-conservation equation (2.11) to give

$$\frac{\partial c}{\partial t} = -\nabla\cdot(c\frac{\partial \mathbf{u}}{\partial t}) \tag{2.18}$$

and again postulating a separation of variables solution for c

$$c = f^{\mathrm{cv}}(\mathbf{x})g^{\mathrm{cv}}(t) \tag{2.19}$$

[7]The respective eigenvalues determine the temporal scaling of the sinusoidal modulation g^{el}, and the exponentially decaying modulation g^{dn}.

As before, the objective is to analyse the *convection eigenfunctions* $f^{\text{cv}}(\mathbf{x})$, which describe the principal spatial modes of mass flux due to cell-ECM interaction. Substituting equation (2.19) into equation (2.18) gives

$$f^{\text{cv}} \frac{\partial g^{\text{cv}}}{\partial t} = -g^{\text{cv}} \frac{\partial g^{\text{el}}}{\partial t} \nabla \cdot \left(f^{\text{cv}} \mathbf{f}^{\text{el}} \right) \tag{2.20}$$

and rearranging to put temporal terms on the left and spatial terms on the right

$$\begin{aligned} \frac{1}{g^{\text{cv}}} \frac{\dot{g}^{\text{cv}}}{\dot{g}^{\text{el}}} &= -\frac{1}{f^{\text{cv}}} \nabla \cdot \left(f^{\text{cv}} \mathbf{f}^{\text{el}} \right) \\ &= k^{\text{cv}} \qquad (k^{\text{cv}} \text{ arbitrary constant}) \end{aligned} \tag{2.21}$$

The temporal solution is found by separation of variables

$$\frac{\partial g^{\text{cv}}}{\partial t} = k^{\text{cv}} g^{\text{cv}} \frac{\partial g^{\text{el}}}{\partial t} \tag{2.22}$$

$$\Rightarrow g^{\text{cv}}(t) = a^{\text{cv}} e^{k^{\text{cv}} g^{\text{el}}(t)} \qquad (a^{\text{cv}} \text{ arbitrary constant}) \tag{2.23}$$

The spatial solution[8] is given by

$$\nabla \cdot (f^{\text{cv}} \mathbf{f}^{\text{el}}) = -k^{\text{cv}} f^{\text{cv}} \tag{2.28}$$

Principal modes of convective mass flux

Inspection of equation (2.17) shows that convective mass flux $\mathbf{J}^{\text{cv}}$ is directed along $\frac{\partial \mathbf{u}}{\partial t}$. But

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{f}^{\text{el}}(\mathbf{x}) \frac{\partial g^{\text{el}}(t)}{\partial t} \tag{2.29}$$

so we can see that $\mathbf{J}^{\text{cv}}$ must be parallel to ECM deformation, i.e. the principal modes of convection are identical to the principal modes of ECM deformation $\mathbf{f}^{\text{el}}(\mathbf{x})$. In fact, we have now shown that the elastic eigenfunctions $\mathbf{f}^{\text{el}}(\mathbf{x})$ describe the:

(i) principal spatial modes of ECM mass flux due to ECM deformation;

(ii) principal spatial modes of cellular mass flux due to cell diffusion;

(iii) principal spatial modes of cellular mass flux due to ECM convection of cells.

[8] Analytically, the convection eigenfunctions $f^{\text{cv}}(\mathbf{x})$ are found by utilising the standard identity

$$\nabla \cdot \left(f^{\text{cv}} \mathbf{f}^{\text{el}} \right) = f^{\text{cv}} \left(\nabla \cdot \mathbf{f}^{\text{el}} \right) + (f^{\text{cv}} \nabla) \cdot \mathbf{f}^{\text{el}} \tag{2.24}$$

so from equation (2.28)

$$\left(\mathbf{f}^{\text{el}} \cdot \nabla \right) f^{\text{cv}} = -\left(\nabla \cdot \mathbf{f}^{\text{el}} + k^{\text{cv}} \right) f^{\text{cv}} \tag{2.25}$$

The left hand side of this is just $||\mathbf{f}^{\text{el}}||$ times the spatial derivative of f^{cv}, projected along eigenmodes $\mathbf{f}^{\text{el}}$. Using a change of variable to s, the arc-length along each convection path,

$$\delta \mathbf{x} = \frac{\mathbf{f}^{\text{el}}}{||\mathbf{f}^{\text{el}}||} \delta s \tag{2.26}$$

and applying it in equation (2.25) gives

$$\frac{d}{ds} \ln f^{\text{cv}} = -\frac{\nabla \cdot \mathbf{f}^{\text{el}} + k^{\text{cv}}}{||\mathbf{f}^{\text{el}}||} \tag{2.27}$$

which can be solved to give the spatial distribution of $f^{\text{cv}}(\mathbf{x})$ along each convection path.

2.7 Summary

A model of biological growth is developed, which incorporates the effects of ECM deformation, cell-ECM interaction, and cell diffusion resulting from cell proliferation. This growth model is general enough to apply across organs of interest in ultrasound imaging, almost all of which are of mesenchymal origin. It is also specific enough to incorporate the cell as the fundamental unit of biological growth.

Material properties of organs are assumed to be spatially uniform, and under this assumption it is shown that the elastic eigenfunctions of the organ describe the principal modes of mass flux due to all three growth mechanisms.

The next two chapters show how these eigenfunctions can be approximated as FEM eigenmodes by using finite element organ models; the eigenmodes are also shown to have appealing properties in characterising organ shape.

Chapter 3

Modelling Shape

Chapter 2 shows that elastic eigenfunctions describe the patterns of organ shape variation. However, there are few organ geometries that can be described in a coordinate system for which eigenfunctions can be analytically derived.

The *finite element method* (FEM) is a technique for solving such problems numerically. The organ model is divided into geometric elements so that the eigenfunction problem consists of elemental problems, each of which is then soluble in its own coordinate system. Such finite element approximations to elastic eigenfunctions are known as *FEM eigenmodes.*

This chapter is concerned with the problem of *shape representation*, since the choice of representation determines the efficiency with which organ models can be manipulated. Eigenmodes are shown to characterise the shape and symmetry of the underlying organ model at different spatial scales; and these properties of shape representation are shown to be resilient to the choice of finite element mesh, which is important when comparing different organ models.

Chapter organisation

- **Section 3.1** reviews the problem of shape representation, and surveys approaches taken by other researchers.
- **Section 3.2** introduces finite element models and their finite element eigenmodes.
- **Section 3.3** examines the eigenproblem associated with finite element equilibrium. Eigenmodes derived from this eigenproblem are shown to have desirable properties in shape representation.
- **Section 3.4** demonstrates these properties with eigenmode examples, which are repeated for six different model meshes to show eigenmode resilience to noisy meshing.
- **Section 3.5** concludes that eigenmodes are well suited to the task of shape representation, due to their symmetry, scale-ordering, stability, and object-centred description. In comparison to other representations, eigenmodes characterise model symmetries in a very general way.

3.1 Shape representation

Blum [13] recognised three aspects to the problem of biological shape representation:

(i) the *morphological* and *taxonomic* problem of formally describing cell, organ, or organism shape;

(ii) the *neuro-physiological* and *psychological* problem of how an organism describes and characterises other organisms' shapes;

(iii) the *developmental* problem of how organisms encode shape information during morphogenesis and growth.

D'Arcy Thompson [156] explained biological form in terms of mechanical processes and constraints, producing a detailed comparison of *related forms* using Cartesian transformations. Under his theory of transformations, *deformation* is defined to be the mapping between related forms.

Cartesian transformations found few applications due to their analytical unwieldiness, but gave rise to simpler methods of allometric growth analysis (Huxley [68]), culminating in Bookstein's [16] development of *morphometrics* as a statistical method for analysing the "covariances of biological form".

Morphometrics applies multivariate statistical methods to 2D *biological landmarks*[1]. Landmark triplets are converted to *shape coordinate* pairs, which are manipulated and visualised on a geometric plane; deformations between sets of shape coordinates are then interpolated using a thin-plate spline underlying the shape coordinate plane. The spline's closed-form elastic eigenfunctions are known as *principal warps*, and these are used to quantify anatomical shape differences.

Bookstein suggests extending landmark sets by searching on *ridge curves* of locally-maximal surface curvature, while Bookstein & Green [17] augment landmarks by incorporating local edge information. Landmark data is suited to biometric analysis, but only partially fulfills the function of taxonomic representation, and barely addresses the other aspects of shape representation enumerated by Blum.

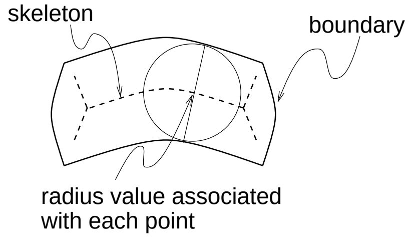


Figure 3.1: Example of the symmetric axis transform

Blum's proposed solution was the *symmetric axis*, derived by a reversible shape transform to give a *skeleton* and a radius value associated with each point on the skeleton. As a representation of biological development and morphology, the *symmetric axis transform* (SAT) is a simple model of growth by isotropic mass accretion around the skeleton. Figure 3.1 shows an example of an object's boundary and skeleton.

[1]Bookstein (ch.3)[16] discusses the difficult question of landmark identification: "landmarks are the points at which one's explanations of biological processes are grounded".

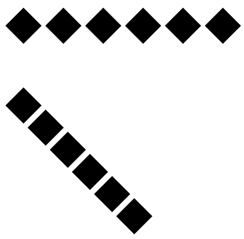


Figure 3.2: The Mach illusion: the choice of object-centred coordinate system affects local shape perception

Marr & Nishihara [87] use the Mach illusion in Figure 3.2 to demonstrate the visual importance of the skeleton: local shapes are perceived either as diamonds or squares relative to the global axis. They examine the psychology of 3D shape representation and recognition, and find that the SAT is an appropriate representation when used in *generalised cones*, which are defined in an *object-centred coordinate system*, and allow a modular organisation of volumetric primitives. Brady & Scott [19] extend the SAT to represent rotational and mirror symmetries, using *smoothed local symmetries* (SLS).

Scale-space representations

One important aspect of shape representation is the characterisation of shape at different spatial scales. A *scale-space* adds an extra dimension to a shape representation by spanning a range of spatial scales (Koenderink [75]; Lindeberg [83]). For example, Kimia et al. [72, 73] model the evolution of boundary curves using a reaction-diffusion equation, which allows *shocks* (shape singularities) to subdivide a shape hierarchically as scale-space is traversed.

Gaussian scale-space is computed by propagating the linear diffusion equation over an initial image or curve, using a Gaussian kernel convolution[2]. Alternatively, Boomgaard & Smeulders [164] propose propagation by a set of non-linear morphological operators, and conclude that both shocks and SAT skeletons can be computed in *morphological scale-space*.

Deformable representations

Another important aspect of shape representation, especially in medical imaging, is the ability to deform to fit a range of variations. For example, Amit et al. [1] employ *deformable templates* which model local probabilities of deformation under diffusion processes. Kass et al. [71] introduce *active contours* which search for boundary fragments, and interpolate them by minimising the contour's bending energy. Fourier *harmonic surfaces* (Ballard & Brown (ch.9.2.3)[5]; Staib & Duncan [139]) and *superquadrics* (Solina & Bajcsy [136]; Terzopoulos & Metaxas [154]) are similarly used as deformable shape models in 3D shape recovery and boundary finding.

Cohen & Cohen [30] apply the active contour to 3D problems by tracking dynamic equilibrium using a FEM model, which allows arbitrary shape parameterisation (unlike a sphere or superquadric). Nastar & Ayache [100, 101] present similar work in medical imaging.

Pentland et al. [112, 113, 114] use superquadrics and elastic FEM shape models in shape recovery for 3D and 3D time-series data. FEM eigenmodes are used to deform FEM shape models, in the same way that Fourier harmonics deform harmonic surfaces.

[2] Koenderink & van Doorn [76] note that receptive field families derived from the diffusion equation bear great resemblance to (mammalian) receptive fields observed in neuro-physiological experiments.

3.2 Finite element elastic model

The growth model in Chapter 2 shows that elastic eigenfunctions describe the principal modes of mass flux during organ growth. These eigenfunctions are the spatial solutions of the wave equation (2.3) describing the ECM's elastic behaviour, so any ECM deformation can be described as an eigenfunction expansion, because of the important properties of eigenfunction orthogonality and completeness (Kreysig (ch.4.8)[77]).

A numerical solution to the wave equation is achieved by a finite element sampling of the organ model. The model is treated as an elastic body whose displacements at n sample points are described by a displacement vector $\mathbf{u}$. The elastic equilibrium implied by the wave equation then takes the form of a discretely sampled equilibrium system (see Appendix A.2)

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{0} \tag{3.1}$$

which is of order $3n$, since there are 3 displacement components at each sample point. Appendix A shows how the body's mass matrix $\mathbf{M}$ and stiffness matrix $\mathbf{K}$ are assembled for FEM equilibrium[3].

Just as the wave equation can be expanded into eigenfunctions, the equilibrium equation (3.1) can be expanded into orthogonal FEM eigenmodes $\boldsymbol{\phi}_i$, which are sampled and linearised approximations to the wave equation's eigenfunctions $\mathbf{f}^{\mathrm{el}}$. Equation (2.4) suggests a separation of variables solution to equation (3.1)

$$\mathbf{u} = \boldsymbol{\phi}_i e^{j\omega_i t} \tag{3.2}$$

where ω_i is the natural-frequency of vibration associated with $\boldsymbol{\phi}_i$ (note that $j = \sqrt{-1}$ here). Applying this simple harmonic solution to equation (3.1) results in the eigenproblem

$$\mathbf{K}\boldsymbol{\phi}_i = \omega_i^2 \mathbf{M}\boldsymbol{\phi}_i \tag{3.3}$$

Given the mass matrix $\mathbf{M}$ and stiffness matrix $\mathbf{K}$ of an organ model, this eigenproblem is solved to give the eigenmodes $\boldsymbol{\phi}_i$ and eigenvalues ω_i^2. Eigenpairs $(\boldsymbol{\phi}_i, \omega_i^2)$ are usually collected into the matrices $(\mathbf{\Phi}, \mathbf{\Omega}^2)$ and ranked in increasing order of eigenvalue[4]

$$\mathbf{\Phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_{3n}) \tag{3.6}$$

$$\mathbf{\Omega}^2 = \begin{pmatrix} \omega_1^2 & & & \\ & \omega_2^2 & & \\ & & \ddots & \\ & & & \omega_{3n}^2 \end{pmatrix} \tag{3.7}$$

[3] $\mathbf{M} > \mathbf{0}$ and $\mathbf{K} \geq \mathbf{0}$ are both symmetric. $\mathbf{K}$ is positive semi-definite because an unsupported body can freely undergo rigid-body rotation and translation.

[4] Eigenmodes are conventionally $\mathbf{M}$-orthonormalised

$$\mathbf{\Phi}^T \mathbf{M} \mathbf{\Phi} = \mathbf{I} \tag{3.4}$$

so that

$$\begin{aligned} \mathbf{\Phi}^T \mathbf{K} \mathbf{\Phi} &= \mathbf{\Phi}^T (\mathbf{M}\mathbf{\Phi}\mathbf{\Omega}^2) \\ &= \left((\mathbf{\Omega}^2)^T \mathbf{\Phi}^T \mathbf{M}\mathbf{\Phi}\right)^T \\ &= \mathbf{\Omega}^2 \end{aligned} \tag{3.5}$$

3.3 Finite element eigenmodes

By collecting eigenpairs into the eigenmode matrix $\mathbf{\Phi}$ and the eigenvalue matrix $\mathbf{\Omega}^2$, equation (3.3) can be rewritten as

$$\mathbf{K\Phi} = \mathbf{M\Phi\Omega}^2 \tag{3.8}$$

This is the *generalised eigenproblem* for the matrix pair $(\mathbf{M}, \mathbf{K})$. It occurs frequently in finite element analysis of structures, and there are powerful numerical techniques for computing $\mathbf{\Phi}$ and $\mathbf{\Omega}^2$. Appendix B introduces some of these techniques.

The transition has now been made from a continuous elastic model defined by the wave equation (2.3), to a finite element elastic model defined by the FEM equilibrium equation (3.1). This equilibrium equation is described in terms of the model's deformation $\mathbf{u}$, but can also be described by projecting $\mathbf{u}$ onto the complete set of eigenmodes $\mathbf{\Phi}$. This latter approach is known as *modal analysis*, and is the natural approach to analysing structural deformation, e.g. Chapter 4.4 uses modal analysis to determine a model's response to random excitation.

The eigenmodes are an approximation to the elastic eigenfunctions, which were previously shown to describe the organ's shape variation; Chapter 4 will examine eigenmodes in this context. However, eigenmodes also have a number of properties that make them useful in shape representation, which is the concern of this chapter. These properties are now characterised by closer examination of the generalised eigenproblem in equation (3.8).

Rayleigh quotient characterisation

The Rayleigh quotient $R(\mathbf{u})$ is defined for symmetric $\mathbf{M}$ and $\mathbf{K}$

$$R(\mathbf{u}) = \frac{\mathbf{u}^T\mathbf{K}\mathbf{u}}{\mathbf{u}^T\mathbf{M}\mathbf{u}} \tag{3.9}$$

where $\mathbf{u} \neq \mathbf{0}$ can also be a matrix. For an undamped and unforced vibrating body[5], $R(\mathbf{u})$ is physically interpreted as the ratio of the body's elastic potential energy to kinetic energy.

The *minimax* (Parlett (ch.10.2)[109]) or *Rayleigh quotient* characterisation (Bathe (ch.2.8)[8]) uses $R(\mathbf{u})$ in an alternative definition of the generalised eigenproblem[6]

$$\omega_i^2 \quad = \quad \min_{\mathcal{S}^i} \max_{\mathbf{u} \in \mathcal{S}^i} R(\mathbf{u}) \tag{3.10}$$

This characterisation is useful because it avoids explicit reliance on previously extracted eigenmodes, so allowing us to concentrate on eigenvalue properties.

Certain modes of vibration are locally maximum in the Rayleigh quotient, i.e. they have high elastic energy and low kinetic energy. Out of these maxima, the one with the lowest Rayleigh quotient is the eigenmode $\boldsymbol{\phi}_i$ required by equation (3.10), whose eigenvalue ω_i^2 is simply $R(\boldsymbol{\phi}_i)$. Successive eigenvalues are non-decreasing (Bathe (ch.2.8)[8]), and all are non-negative since $\mathbf{M} > \mathbf{0}$ and $\mathbf{K} \geq \mathbf{0}$.

There is also a graphical interpretation using a matrix $\mathbf{A}$ with eigenvalues $(\omega_1^2, \omega_2^2, \omega_3^2)$, which has the Rayleigh quotient $R(\mathbf{u}) = (\mathbf{u}^T\mathbf{A}\mathbf{u})/(\mathbf{u}^T\mathbf{u})$. The unit sphere is mapped by $\mathbf{A}^{\frac{1}{2}}$ to an ellipsoid with axes of radius $(\omega_1, \omega_2, \omega_3)$; the aim is to search the ellipsoid for these axes, whose radii give us the eigenvalues. The first eigenvalue is simply found by searching the entire

[5]From equation (3.2), the *natural-frequency* of vibration for each eigenmode is ω_i.

[6]See Appendix B for the definition of an i-dimensional subspace $\mathcal{S}^i$.

ellipsoid for the smallest radius

$$\omega_1 = \min_{\mathbf{u}\in\mathcal{S}^3} \|\mathbf{A}^{\frac{1}{2}}\mathbf{u}\| \qquad (\|\mathbf{u}\| = 1) \tag{3.11}$$

In order to find the next axis ω_2, we examine the intersection of the ellipsoid with all planes (notated by $\mathcal{S}^2$) passing through its origin. Each intersection locus is an ellipsoid with a maximum radius on its major axis, and out of all possible intersection loci the major axis is smallest at the desired plane, as shown in Figure 3.3.

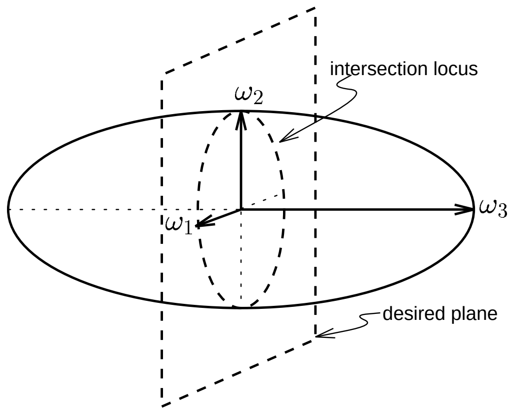


Figure 3.3: Graphical interpretation of the Rayleigh quotient characterisation

In this plane, the major axis has radius ω_2

$$\omega_2 = \min_{\mathcal{S}^2} \max_{\mathbf{u}\in\mathcal{S}^2} \|\mathbf{A}^{\frac{1}{2}}\mathbf{u}\| \qquad (\|\mathbf{u}\| = 1) \tag{3.12}$$

whence the minimax characterisation, which generalises to higher dimensional examples.

Repeated eigenmodes

It follows from the Rayleigh quotient characterisation that structures which possess mirror and rotational symmetries, will also have local minima of $R(\mathbf{u})$ with identical values. Such *repeated eigenmodes* have identical eigenvalues, and identical eigenmode shapes about their respective planes or axes of symmetry.

Repeated eigenmodes[7] span an *eigenspace* uniquely defined with respect to other eigenmodes. Within this eigenspace they are mutually orthogonal and uniquely determined with respect to each other. However, they are not individually uniquely determined with respect to eigenmodes outside the eigenspace.

We intend to use the eigenmode representation to characterise organ symmetry, so the representation must cope with potential mirror and rotational symmetries. Such symmetries are very unlikely in multiplicity 3 (e.g. sphere or cube), so we need only consider how to deal with 2D eigenspaces. Where an eigenmode defines a vector component at each finite element node, a pair of repeated eigenmodes (a 2D eigenspace) defines a plane with a unique normal component. By employing these components, repeated or single eigenmodes can both be treated in the same way when used as shape features[8].

[7] The block Lanczos algorithm is specifically designed to detect repeated eigenmodes (see Appendix B).

[8] In practice, errors in eigenmode computation make it difficult to distinguish between repeated eigenmodes slightly separated in natural-frequency, and genuinely separate eigenmodes.

You cannot hear the shape of a drum

Having established that eigenmodes can be used as shape features, even with mirror or rotationally symmetric organ models, we now discuss the properties of eigenvalues and eigenmodes in shape representation.

Kac [69] examined the problem of determining an elastic membrane's shape from its eigenvalue spectrum: area, circumference, and connectivity (for polygonal boundaries) were established. However, Gordon et al. [54] showed that two differently shaped domains can in fact have the same eigenvalue spectrum, which suggests that elastic eigenvalues are probably not the shape representation we require[9].

In general, however, even if an eigenvalue representation is augmented with eigenmodes, the geometry of the underlying elastic body cannot be uniquely characterised by the eigenpairs. This is simply because the generalised eigenproblem in equation (3.8) is posed in $(\mathbf{M}, \mathbf{K})$; the resulting eigenpairs are therefore determined by the body's mass and stiffness distribution, rather than the body's geometry. Also, the eigenmode transform $(\boldsymbol{\Phi}, \boldsymbol{\Omega}^2)$ is only invertible to the single matrix $\mathbf{M}^{-1}\mathbf{K}$, and not to the matrix pair $(\mathbf{M}, \mathbf{K})$.

It should be recognised that this reduction in dimensionality is, in fact, a desirable property for *shape features*, since their purpose is to distill shape information into more representative quantities. We now show that eigenmodes are ideal shape features in this respect.

Symmetry of eigenmodes

Eigenmode shapes tend to emphasise the various symmetries of the underlying organ shape; such behaviour is highly desirable in a shape representation, and the consequences are discussed later in Section 3.5. We first examine why elastic eigenmodes might generally be symmetric.

The first six eigenmodes of a freely vibrating structure are those of rigid-body motion, since there is no elastic deformation involved, and hence no cost in elastic energy. These rigid-body modes account for the translational or rotational motion of the body about its centre of mass.

There are no supports or physical boundary conditions in our equilibrium analysis, so after an initial impulse acts upon the body, there are no external forces acting upon the freely vibrating body. Therefore, non-rigid eigenmodes cannot result in any net translation of the centre of mass, nor can they exert any net rotational moment.

This does not, in itself, necessitate any symmetry in the body's choice of deformation during free vibration. If the deformation is an eigenmode, however, it is also a local minimum in the Rayleigh quotient, and symmetric deformations allow for a minimal penalty in elastic energy, while obeying the physical constraints set out above[10]. Each eigenmode therefore tends to emphasise symmetries in the organ, where the dominant symmetry in each eigenmode depends on the *spatial scale* of the eigenmode.

Eigenmode features require global support[11] and cannot easily cope with occluded part structures. This is irrelevant in model-based imaging since organ shape models must always be fully constructed. In cases where partially occluded models require shape analysis, however, wavelet modes in Chapter 6.4 are a possible solution, since they have localised support.

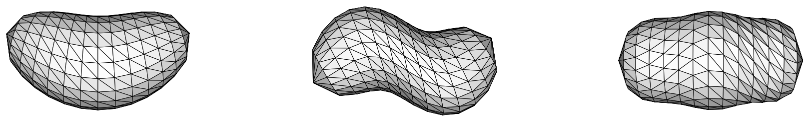
[9] Elastic eigenvalues have properties similar to moment invariants when used in shape representation. Although both are invariant to linear transformation and scaling, they are global features and their representation of actual geometry is weak. Even so, moment invariants are found to be useful in shape discrimination (Hu [64]).

[10] A less persuasive argument is that most simply-connected biological structures map conformally from an ellipsoid volume, and the ellipsoid's eigenmodes should retain their symmetric features (see examples in Figure 3.4) under such a mapping (Kreysig (ch.16.1)[77]).

[11] In other words, the shape of each eigenmode is dependent on the global shape of the organ.

Scale-ordering of eigenmodes

The eigenmodes of a spherical membrane are not only harmonic in time, but also harmonic in spatial distribution[12] (Ballard & Brown (ch.9.2.3)[5]). The latter implies a spatial scaling of eigenmodes, which can also be seen in other elastic models such as the volumetric ellipsoid in Figure 3.4.



(a) $\boldsymbol{\phi}_7$ (b) $\boldsymbol{\phi}_{11}$ (c) $\boldsymbol{\phi}_{15}$

Figure 3.4: Eigenmodes of an ellipsoid volume are ordered in spatial scale

The Rayleigh quotient characterisation predicts that higher eigenmodes have higher elastic energies[13], which accords with the scale-ordering in Figure 3.4, since deformations of high spatial-frequency (i.e. smaller spatial scale) are more energetically expensive.

Also, Appendix A.8 shows that each eigenmode's elastic energy is determined by its *modal amplitude* and *modal stiffness*. Since higher eigenmodes are of smaller spatial scale, they require a larger amplitude for the same physical displacement; and equation (3.5) shows that modal stiffness is equal to the natural-frequency ω_i^2, which again increases with higher eigenmodes. In other words, both modal amplitude and modal stiffness increase for successive eigenmodes, which must therefore have successively higher elastic energies.

Figure 3.5 shows two eigenmodes of a uniform elastic string, which in this case are also Fourier harmonics. In fact, elastic eigenmodes can be thought of as analogous harmonics for finite element shape models, and it is the harmonic nature of eigenmodes which explains their increasing spatial-frequency[14]. This natural ordering in spatial scale is another desirable property of eigenmodes as shape features.

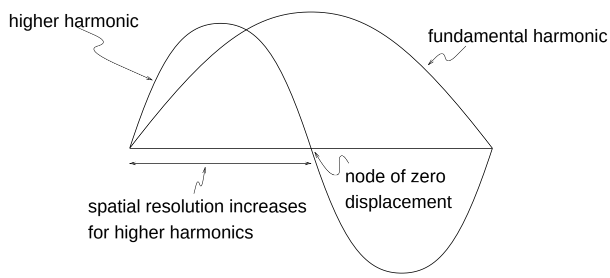


Figure 3.5: Fourier harmonics of a uniform elastic string are ordered in spatial scale

Higher spatial-frequency modes are unimportant in determining a structure's dynamic response (Bathe [8]). There is also an increased possibility of (spatial) aliasing in eigenmodes of high spatial-frequency, due to Nyquist sampling considerations for the model's mesh. Organ deformations described using an eigenmode expansion can therefore be truncated in principled order of eigenmode importance and accuracy.

[12] The nth spherical harmonic is temporally modulated by $e^{j\omega_n t}$, and spatially modulated by $e^{jn\theta_1}$ and $e^{jn\theta_2}$, where (θ_1, θ_2) are the spherical coordinates.

[13] Recall that the Rayleigh quotient is the ratio of elastic potential energy to kinetic energy.

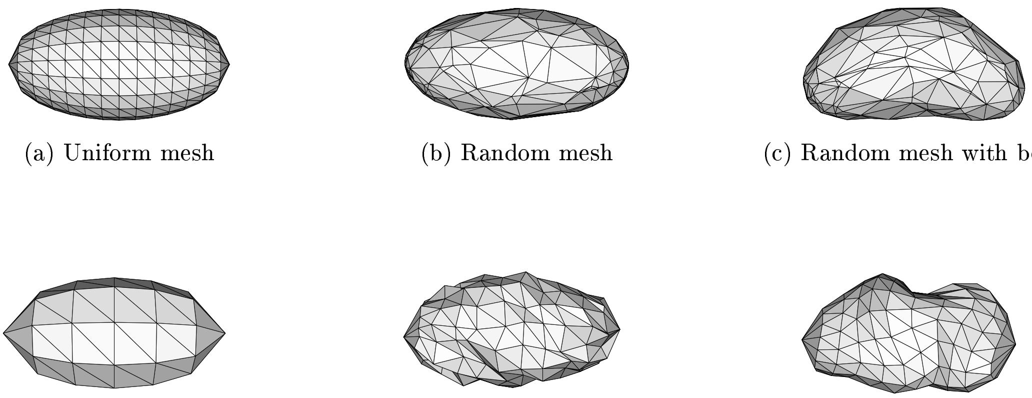
[14] Chapter 6 utilises this analogy in eigenmode spectrum analysis (cf. Fourier spectrum analysis).

3.4 Eigenmodes as shape features

We have shown that eigenmodes are shape features with excellent properties in symmetry and spatial scaling; morever, eigenmodes of large spatial scale are better sampled than other eigenmodes of smaller spatial scale.

These properties are now demonstrated in Figures 3.7–3.12, which show eigenmodes $\boldsymbol{\phi}_7$ to $\boldsymbol{\phi}_{12}$ of the six ellipsoid models in Figure 3.6. Each ellipsoid model has a different finite element mesh, to show that eigenmodes of large spatial scale are relatively resilient to the choice of mesh (Syn & Prager [147]).

- Figure 3.6(a) is a uniform volumetric mesh.
- Figure 3.6(b) is a randomly sampled volumetric mesh.
- Figure 3.6(c) is a randomly sampled volumetric mesh, with global bending.
- Figure 3.6(d) is a sparsely sampled volumetric mesh.
- Figure 3.6(e) is a uniform ellipsoid mesh with added Gaussian noise of small variance.
- Figure 3.6(f) is a uniform ellipsoid mesh with added Gaussian noise of large variance.



(a) Uniform mesh (b) Random mesh (c) Random mesh with bending

(d) Sparse mesh (e) Noise of high spatial-frequency (f) Noise of low spatial-frequency

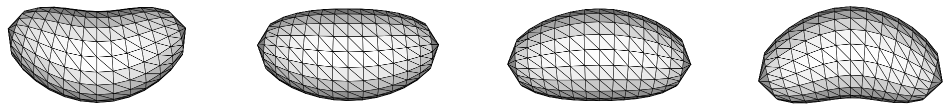
Figure 3.6: Six ellipsoid shape models

Note that $\boldsymbol{\phi}_1$ to $\boldsymbol{\phi}_6$ are the rigid-body modes of translation and rotation; the first non-rigid mode shown in Figures 3.7–3.12 is therefore $\boldsymbol{\phi}_7$. Four snapshots are shown of each eigenmode, to capture its shape as it ranges over positive and negative amplitudes.

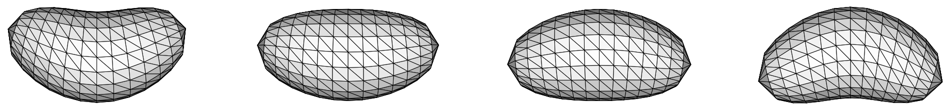
It should be clear from comparing equivalent eigenmodes between each figure, that the eigenmode shapes are relatively unaffected by the choice of mesh. As mentioned before, this is a consequence of the eigenmode's spatial scale being much larger than the spacing of the underlying mesh.

Eigenmode features are seen to be ordered in spatial scale: $\boldsymbol{\phi}_7$ has fundamental spatial-wavelength, so its bending shape spans the entire ellipsoid; $\boldsymbol{\phi}_{12}$ has half the fundamental spatial-wavelength, so it describes an S-shaped bending.

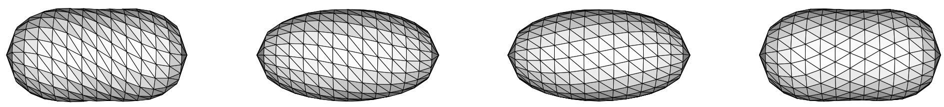
Ellipsoid volume with uniform tessellation



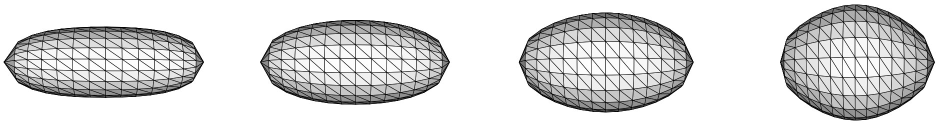
(a) Snapshots of 7th eigenmode



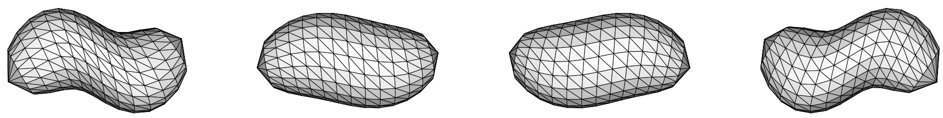
(b) Snapshots of 8th eigenmode



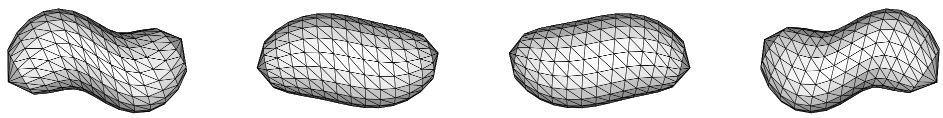
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode

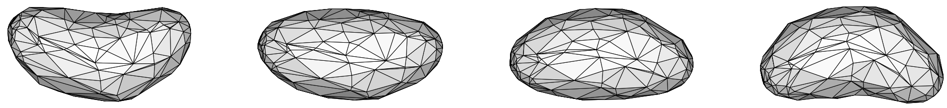


(f) Snapshots of 12th eigenmode

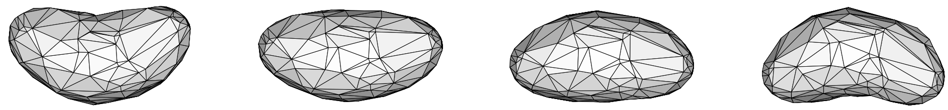
Figure 3.7: Eigenmodes of uniformly tessellated ellipsoid

Figure 3.7 shows the eigenmodes of an ellipsoidal volume tessellated uniformly with tetrahedral elements. The ellipsoid's symmetry gives rise to pairs of repeated eigenmodes in Figures 3.7(a,b) and 3.7(e,f).

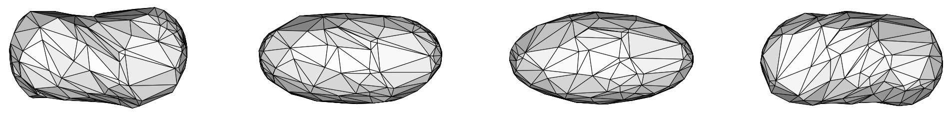
Ellipsoid volume with non-uniform tessellation



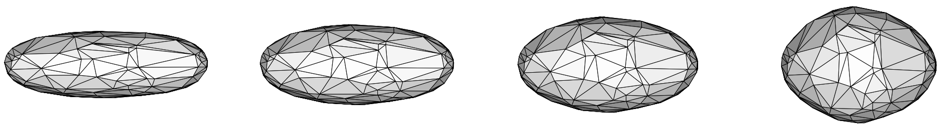
(a) Snapshots of 7th eigenmode



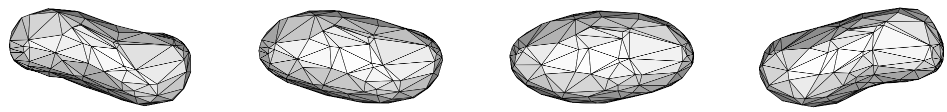
(b) Snapshots of 8th eigenmode



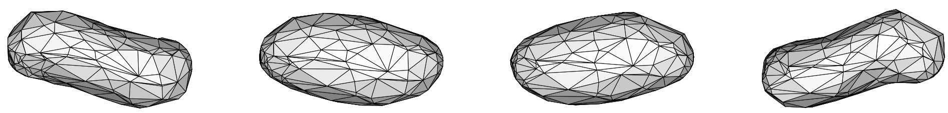
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode

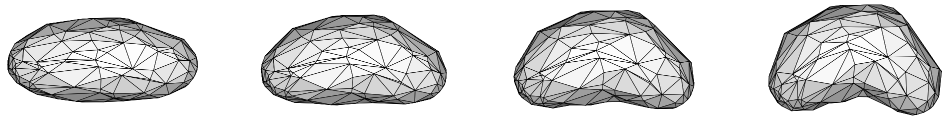


(f) Snapshots of 12th eigenmode

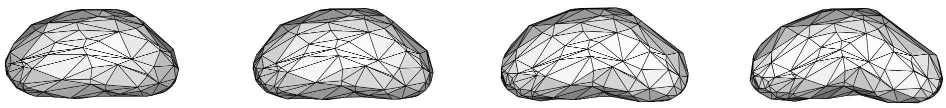
Figure 3.8: Eigenmodes of non-uniformly tessellated ellipsoid

Figure 3.8 shows the eigenmodes of an ellipsoid volume sampled randomly both internally and on the surface, with a higher sampling density near the poles. This is the quality of meshing typical of manual sampling, or automated sampling algorithms operating on segmented MRI volume images.

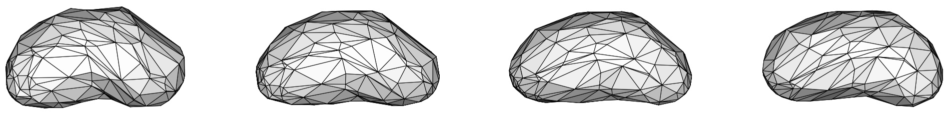
Bent ellipsoid volume with non-uniform tessellation



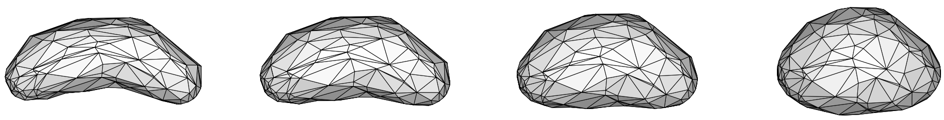
(a) Snapshots of 7th eigenmode



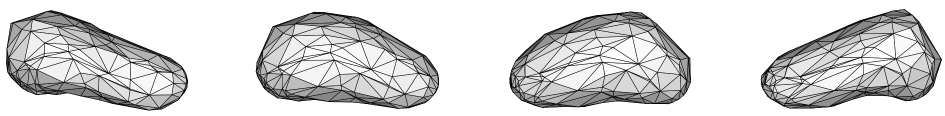
(b) Snapshots of 8th eigenmode



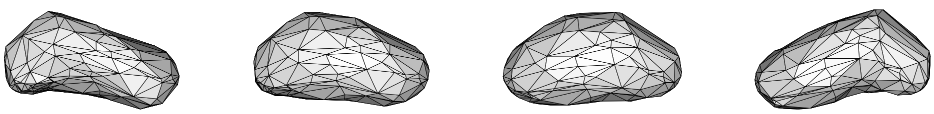
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode

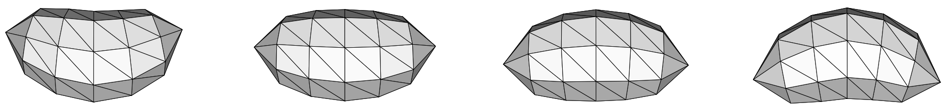


(f) Snapshots of 12th eigenmode

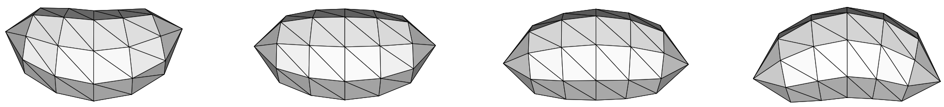
Figure 3.9: Eigenmodes of non-uniformly tessellated ellipsoid with bending

Figure 3.9 shows the eigenmodes of a bent version of the ellipsoid shown in Figure 3.8. This demonstrates the stability of eigenmodes features to large-scale shape changes, e.g. due to ultrasound image distortion from refraction effects (see Chapter 5.6).

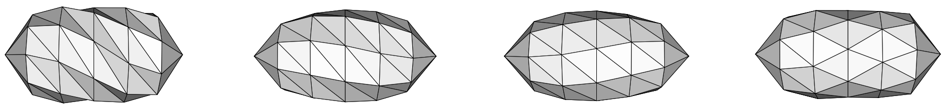
Ellipsoid volume with sparse tessellation



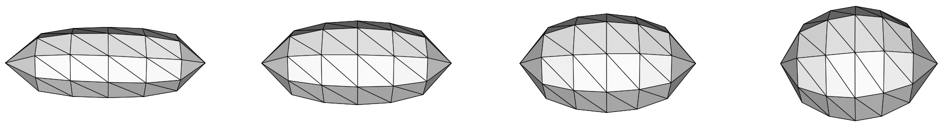
(a) Snapshots of 7th eigenmode



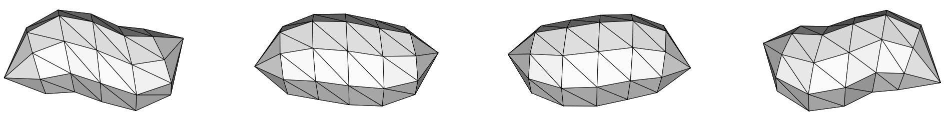
(b) Snapshots of 8th eigenmode



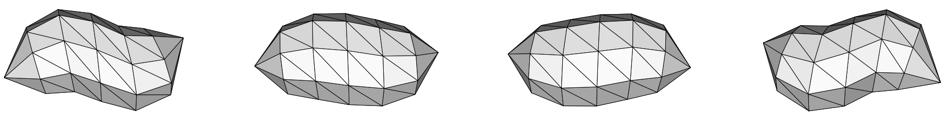
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode

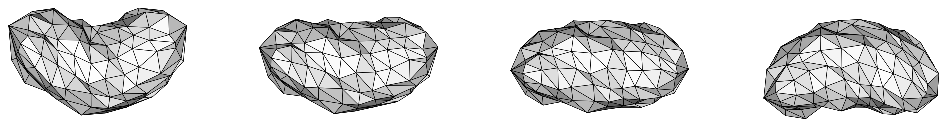


(f) Snapshots of 12th eigenmode

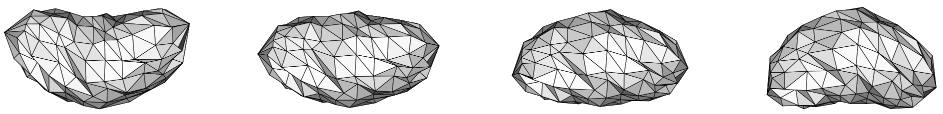
Figure 3.10: Eigenmodes of sparsely tessellated ellipsoid

Figure 3.10 shows the eigenmodes of a sparsely sampled ellipsoid. Eigenmode features are stable for models with different sampling densities, e.g. when comparing and validating segmented models derived from ultrasound and MRI.

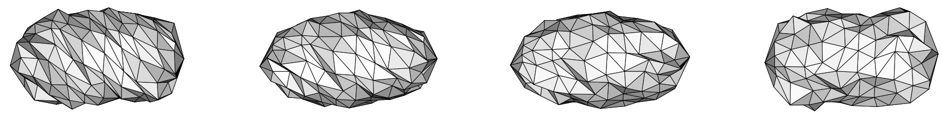
Ellipsoid volume with noise of high spatial-frequency



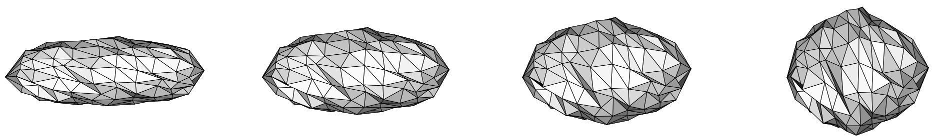
(a) Snapshots of 7th eigenmode



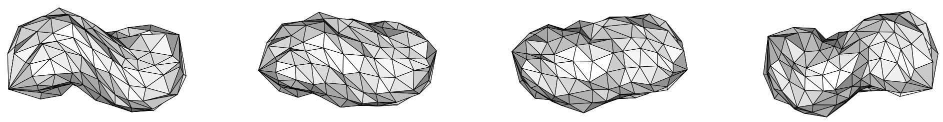
(b) Snapshots of 8th eigenmode



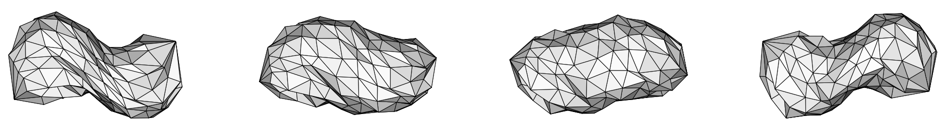
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode

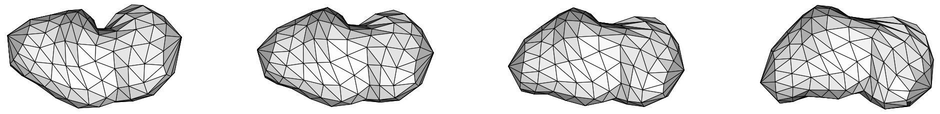


(f) Snapshots of 12th eigenmode

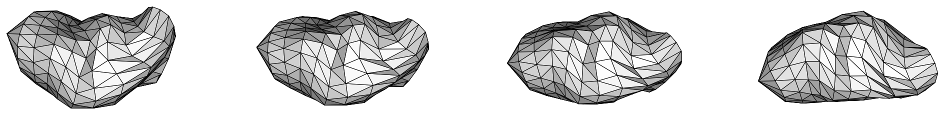
Figure 3.11: Eigenmodes of ellipsoid with noise of high spatial-frequency

Figure 3.11 shows the eigenmodes of a noisily sampled ellipsoid, where additional Gaussian noise with a standard deviation of one element width is added to each element node. This demonstrates that eigenmodes are relatively resilient to noise of high spatial-frequency, e.g. errors in manual landmark sampling.

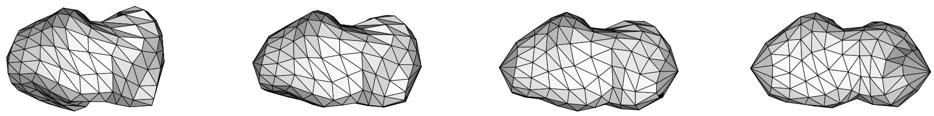
Ellipsoid volume with noise of low spatial-frequency



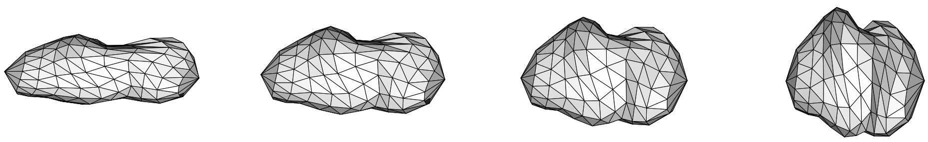
(a) Snapshots of 7th eigenmode



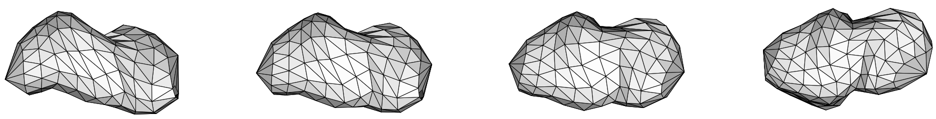
(b) Snapshots of 8th eigenmode



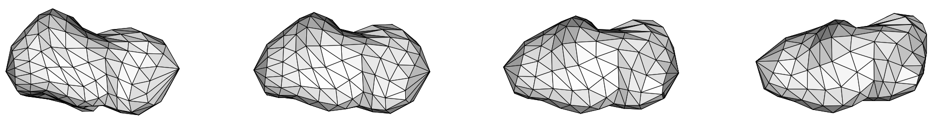
(c) Snapshots of 9th eigenmode



(d) Snapshots of 10th eigenmode



(e) Snapshots of 11th eigenmode



(f) Snapshots of 12th eigenmode

Figure 3.12: Eigenmodes of ellipsoid with noise of low spatial-frequency

Figure 3.12 shows the eigenmodes of an ellipsoid with noise of large spatial correlation added to each node. Such noise simulates the (normal) inter-patient variability in biological structures being scanned.

3.5 Generalised symmetries

We have demonstrated the eigenmode properties of symmetry, scale-ordering and stability. Eigenmodes of large spatial scale are ideal for shape representation because of these properties. Morever, eigenmodes are sensitive to the *global* shape of the organ model because of their global support.

Scale-ordering is a consequence of the harmonic nature of eigenmodes. Each eigenmode has an associated vector field describing its *spatial wavelength* (discussed below), and this spatial wavelength defines the spatial scale at which the eigenmode's symmetries are dominant. Eigenmode stability is in turn a consequence of scale-ordering, since eigenmodes of large spatial scale are relatively unaffected by an underlying mesh of small spatial scale.

It is worth noting that Sclaroff & Pentland [130] suggest elastic eigenmodes to be the *generalised symmetries* of a 2D boundary model, although no particular characterisation is provided to justify such a sweeping label. Still, the idea of generalised symmetry seems appropriate because of the harmonic behaviour mentioned above.

Let us now compare the eigenmode representation to others previously reviewed in Section 3.1, in order to show that eigenmodes are indeed more generalised symmetries than either the SAT or SLS.

Comparison with other shape representations

Marr & Nishihara's [87] requirements of an object-centred coordinate system are fulfilled, since the eigenmode fields $\boldsymbol{\phi}_i$ are always defined relative to the underlying organ shape, independent of pose and location. In addition, eigenmodes are ranked in order of accuracy, significance and spatial localisation, so that shape symmetries are exhibited in global-to-local order. Eigenmode symmetries also cope naturally with mirror and rotational symmetries in the organ shape.

The SAT (and variant) representations suffer in comparison because of possible skeletal branching in the presence of incremental shape changes, as shown in Figure 3.13. However, both eigenmode and SAT representations are susceptible to occlusion effects, since both are computed with global support. The wavelet modes developed in Chapter 6.4 may be a solution to this, since they are shape features with localised support.

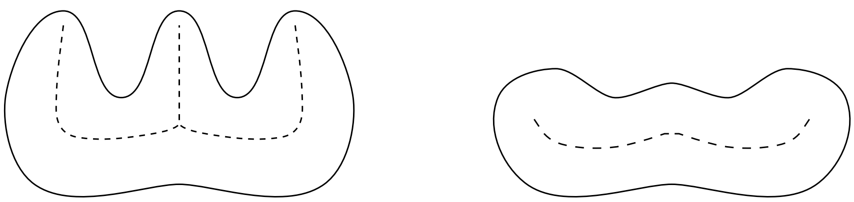
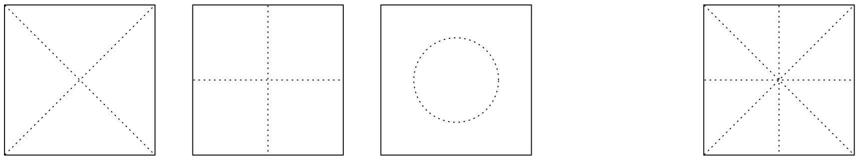


Figure 3.13: Example of SAT sensitivity to boundary shape change (adapted from Blum [13])

Pizer et al. [115, 116] examine SAT behaviour at different spatial resolutions in order to sensitise the computed skeleton to shape symmetries at different spatial scales. The resulting representation is the multi-resolution *core*, computed directly from intensity images; cores therefore have the significant advantage of not requiring prior segmentation of shapes in the image. However, appropriate techniques have not been developed to analyse and manipulate the resulting 3D skeletons in the core.

Blum [13] computes the SAT using the grassfire algorithm[15], which solves Laplace's equation within the shape, defined with appropriate boundary conditions. Pizer et al. [116] and Brady & Scott [19] note that the SAT and SLS can also be computed as the solution to other partial differential equations (PDEs), again given appropriate boundary conditions.

In fact, Brady & Scott speculate that eigenfunctions of these PDEs represent more generalised symmetries than the SLS, because the SLS imposes a rigid distinction between rotational and mirror symmetries. This is illustrated in Figure 3.14(a), which compares mirror and rotational symmetries highlighted by elastic eigenfunctions, with the SLS in Figure 3.14(b).



(a) Elastic eigenfunctions (b) SLS

Figure 3.14: Comparing elastic eigenfunctions and smoothed local symmetries (SLS) for an unclamped plate (adapted from Brady & Scott [19])

We therefore conclude that elastic eigenmodes represent an organ model's symmetries in a more general way than either the SAT or SLS. In comparison with other representations surveyed in Section 3.1: eigenmodes span a range of spatial scales (cf. scale-space representations), and describe modes of organ deformation (cf. deformable representations). Eigenmodes can also be computed for any shape model, and are relatively invariant to the choice of shape meshing.

In the context of Blum's [13] classification of the shape representation problem:

(i) The *morphological* and *taxonomic* aspect:
Chapter 6 shows that eigenfunction extrema fulfill the function of biological landmarks, and can in fact be used to compute biological fits between organ models.

(ii) The *neuro-physiological* and *psychological* aspect:
Bruce & Green [20] suggest that psychological representations of shape exist at multiple spatial scales, and depend strongly on symmetric cues. Eigenmodes fulfill both criteria by characterising model symmetries in a scale-ordered fashion[16].

(iii) The *developmental* aspect:
Chapter 2 shows that elastic eigenfunctions are a description of principal modes of mass flux in our model of biological growth.

[15] An analogy is that of placing a uniform distribution of electric charge along the boundary, so that ridges of stationary potential describe the skeleton.

[16] Although eigenmodes can only be computed for shape models, Pizer et al.'s [115, 116] *core* representation is a variant which is computed directly from intensity images.

Spatial scale of eigenfunctions

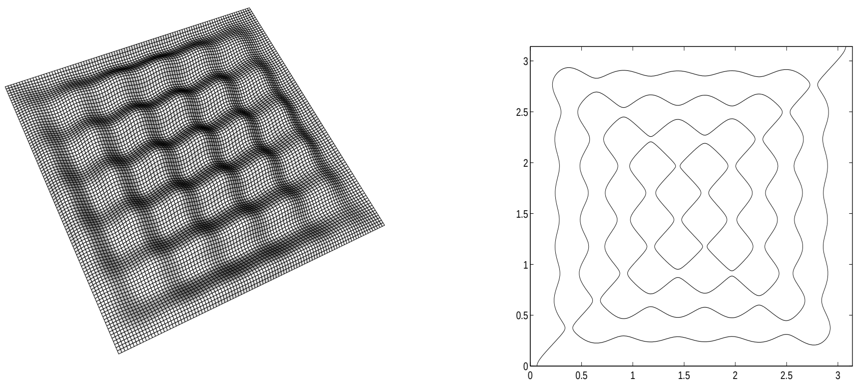
It was suggested earlier that *spatial wavelength* decreases with successive eigenfunctions. This was explained to be a consequence of the harmonic nature of elastic eigenfunctions, which have a non-decreasing number of node points[17]. Recall that eigenfunction nodes are points with zero displacement.

We are interested in defining spatial wavelength, since this is precisely what we mean when we speak of an eigenfunction's "spatial scale". However, spatial wavelength is not a simple scalar quantity when dealing with 2D or 3D models, particularly those with complex geometry. For example, an eigenfunction may have a small spatial wavelength along one axis of a plate, but a very large spatial wavelength along the other; in this case, it is clear that spatial scale has two components.

In the worst case of an arbitrary eigenfunction shape, spatial wavelength has to be specified everywhere on the model. For example, if we choose our model to be a square plate of side π, then one possible set of eigenfunctions for the scalar wave equation (2.15) is[18]

$$\boldsymbol{\phi}_i = \sin 2ix \sin y + \mu \sin x \sin 2iy \tag{3.13}$$

The eigenfunction $\boldsymbol{\phi}_6$ shown in Figure 3.15 is confusing, because it has a single nodal curve (i.e. it should have fundamental wavelength), although its spatial wavelength appears to be $\frac{\pi}{6}$.



(a) Eigenfunction $\boldsymbol{\phi}_6$ (b) Nodal curve of zero displacement

Figure 3.15: Example of eigenfunction which has only a single nodal curve ($\mu = 0.9$)

We therefore need to salvage the notion of an eigenfunction's spatial scale being a simple quantity derived from spatial wavelength. Fortunately Chapter 6.4 shows that, for the elastic eigenfunctions of uniform elastic models, spatial wavelength is a vector field which exists only at a single magnitude, although its vector components may point in many directions. An elastic eigenfunction's spatial scale should therefore be defined as being this magnitude.

[17] This is predicted by Sturm-Liouville theory, which studies the behaviour of scalar eigenfunctions occurring in many physical problems (Courant & Hilbert (ch.6)[35]).

[18] Equation (3.13) describes incomplete sets of eigenfunctions, whereas the elastic eigenfunctions are the only complete set. In fact, the eigenfunctions in equation (3.13) turn out to be linearly assembled from elastic eigenfunctions (Courant & Hilbert (ch.6.7)[35]).

3.6 Summary

The finite element procedure allows the elastic eigenfunctions from Chapter 2 to be approximately computed for any organ model. These approximations are the finite element eigenmodes, which are used to model organ shape variation in Chapter 4. Eigenmodes form a complete and orthogonal basis set, and are useful for modal analysis of organ deformations.

This chapter shows that eigenmodes are also ideal for shape representation. They can be thought of as the generalised symmetries of the organ model, being symmetric at multiple spatial scales. Eigenmodes of large spatial-scale are particularly useful, because they are relatively invariant to noise in the underlying organ mesh.

Chapter 4

Modelling Variation in Shape

Chapter 2 introduces a biological growth model in which organ shape variation is described by elastic eigenfunctions, and Chapter 3 shows that these eigenfunctions can be linearly approximated by finite element eigenmodes.

Eigenmodes are *model-based modes* of shape variation, since they are dependent only on the shape and material properties of the underlying finite element organ model. This chapter discusses the use of these modes, and introduces a non-linear representation for cases where shape variation cannot be adequately described by linearised modes.

Another common approach to modelling organ shape variation, is the statistical analysis of organ shape examples. This analysis gives rise to the principal components of shape variation, which are *data-driven modes* of shape variation.

A common framework is developed to combine model-based and data-driven modes of shape variation. The result is a shape model whose knowledge of organ shape variation is initialised with eigenmodes, but can be refined from new organ examples.

Before this common framework can be applied, however, new organ examples must be scaled to the correct size relative to previous examples. There is an inevitable covariation between an organ's size and its shape, so an incorrect change in its size will distort its shape. The *centroid size* metric attempts unbiased estimation of size differences in the presence of this covariation.

Chapter organisation

- **Section 4.1** introduces a *centroid size* metric for size-normalisation of finite element organ models. This removes an undesired source of variation in datasets of organ shape.
- **Section 4.2** discusses the use of model-based modes of shape variation, and introduces a non-linear extension to these model-based modes.
- **Section 4.3** discusses data-driven modes of shape variation, which are derived by statistical analysis of a training set of organ shapes. Data-driven modes have the advantage of being specific to the training data, but lack the descriptive and analytic framework afforded by model-based modes.
- **Section 4.4** examines the statistical distribution of model-based modes under noisy excitation of eigenmodes; data-driven analysis is applied to this distribution.
- **Section 4.5** uses this analysis to implement a model-based framework, which incorporates data-driven analysis by iteratively updating itself with new organ shape data.

4.1 Covariances between size and shape

The aim of this chapter is a model of organ shape variation which can learn from new examples of organ shape. These new examples must be normalised in size, because examples which vary in size will also seem to vary in shape (see Figure 4.4). Size normalisation is also important in Chapter 5 when fitting an organ shape model to 3D ultrasound data, and in Chapter 8 when fitting organ shape models to each other.

Bookstein (ch.5.5)[16] shows the *centroid size* metric to be the only size variable invariant to independent, identical, and isotropic noise in landmark positions; these noise assumptions are reasonable, although somewhat idealised. There are a number of equivalent definitions of centroid size, and the one used below is the sum of all inter-landmark squared-distances.

Although it is possible to apply this metric directly to the finite element nodes of organ models, it is more accurate to generalise the metric so that it applies to finite element faces[1]. This generalisation is important in finite element models which have unevenly positioned nodes.

Assuming that the model's surface is composed of triangular elements, an expression is first found for the centroid size of a pair of triangles. This expression involves integration over the triangle pair, so an appropriate triangular coordinate system is required.

Coordinates in a triangle

Figure 4.1 shows a triangle with node positions $(\mathbf{a}, \mathbf{b}, \mathbf{c})$. Consider a small strip $(\mathbf{p}, \mathbf{q})$ parallel to the baseline $(\mathbf{a}, \mathbf{b})$, and let $\mathbf{r}$ denote a position in this strip. Then

$$\mathbf{q} - \mathbf{p} = (1-\mu)(\mathbf{b} - \mathbf{a}) \tag{4.1}$$

$$\mathbf{r} = \mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}) \tag{4.2}$$

$$dA = (h\, d\mu)(\|\mathbf{q} - \mathbf{p}\|\, d\lambda) \tag{4.3}$$

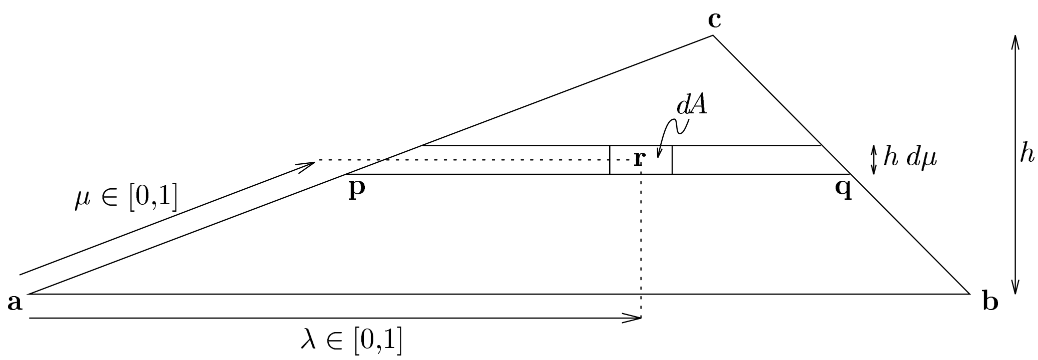


Figure 4.1: Coordinate parameters of triangle in 3D space

Centroid size integrated over two triangles

The metric can now be integrated over two triangles T_1 and T_2, respectively defined by their nodal coordinates $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ and $(\mathbf{d}, \mathbf{e}, \mathbf{f})$. If $\mathbf{r}$ is a point within T_1 and $\mathbf{s}$ is a point within T_2,

[1] Surface meshing is almost always of higher quality than internal sampling, so integrating over surface faces gives a very good measure of the model's size, for both volumetric and surface model meshes.

then integration over the triangle pair gives[2]

$$\begin{aligned}\int_{T_1}\int_{T_2} \|\mathbf{r}-\mathbf{s}\|^2 \, dA_1 dA_2 &= \int_0^1\int_0^1\int_0^1\int_0^1 h_1\|\mathbf{q_1}-\mathbf{p_1}\|\, h_2\|\mathbf{q_2}-\mathbf{p_2}\|\, \|\mathbf{r}-\mathbf{s}\|^2 \, d\lambda_1 d\mu_1 d\lambda_2 d\mu_2 \\ &= (\mathbf{a}\cdot\mathbf{b}\times\mathbf{c})(\mathbf{d}\cdot\mathbf{e}\times\mathbf{f})\left(\frac{3c_1+3c_2-4c_3}{18}\right) \qquad (4.4)\end{aligned}$$

where

$$c_1 = \|\mathbf{a}\|^2+\|\mathbf{b}\|^2+\|\mathbf{c}\|^2+\|\mathbf{d}\|^2+\|\mathbf{e}\|^2+\|\mathbf{f}\|^2 \qquad (4.5)$$

$$c_2 = \mathbf{a}\cdot\mathbf{b}+\mathbf{b}\cdot\mathbf{c}+\mathbf{c}\cdot\mathbf{a}+\mathbf{d}\cdot\mathbf{e}+\mathbf{e}\cdot\mathbf{f}+\mathbf{f}\cdot\mathbf{d} \qquad (4.6)$$

$$c_3 = (\mathbf{a}+\mathbf{b}+\mathbf{c})\cdot(\mathbf{d}+\mathbf{e}+\mathbf{f}) \qquad (4.7)$$

Centroid size generalised to model surface

Summing equation (4.4) over all triangle pairs (T_i, T_j) on the model surface, gives the centroid size metric cs

$$cs = \sum_i \sum_{j\geq i} \int_{T_i}\int_{T_j} \|\mathbf{r}-\mathbf{s}\|^2 \, dA_1 dA_2 \qquad (4.8)$$

Size-normalisation is achieved by forcing models to have the same centroid size, i.e. scaling all coordinates in proportion to $\sqrt{cs}$. This should only be done in cases where the models are expected to have similar centroid sizes, e.g. fitting a model to 3D ultrasound data in Chapter 5; another example in Chapter 5 is sizing organ models for computer display in a graphics window.

There should be no need for size normalisation in cases where organ models are defined in true coordinates, e.g. meshes directly segmented from MRI. However, size normalisation is performed in Chapter 8 before fitting two organ models to each other using their eigenmodes. This is because the assumption of eigenmode similarity implies that the two organs are proximal in 'growth space' (see Chapter 6.1), so that they should have similar centroid sizes.

Growth eigenfunctions are used to traverse the 'growth space' from one organ to the other. However, if the two models are significantly different in size, then eigenmodes – i.e. linearised eigenfunctions – can no longer be used in traversal. A centroid size normalisation is therefore forced, so that a fit can be approximated even for organs of different size.

Figure 4.2 shows twelve liver models with normalised centroid sizes; these models are segmented from MRI scans of foetal cadavers at different stages of growth (Pasapula [110]).

4.2 Model-based modes of variation

Chapter 3 shows that the principal modes of shape variation for an organ model are approximated by finite element eigenmodes. The eigenmodes depend only on the geometry and material properties of the organ model, and are therefore *model-based modes* of variation.

Note that there is a distinction between shape variation and shape deformation, terms sometimes used interchangeably in the literature.

[2]Equation (4.4) was algebraically integrated by computer, but it can be verified to have some expected properties:

(i) symmetric under permutations of $(\mathbf{a},\mathbf{b},\mathbf{c})$ or $(\mathbf{d},\mathbf{e},\mathbf{f})$ and their components;

(ii) symmetric under swapping $(\mathbf{a},\mathbf{b},\mathbf{c})$ with $(\mathbf{d},\mathbf{e},\mathbf{f})$.

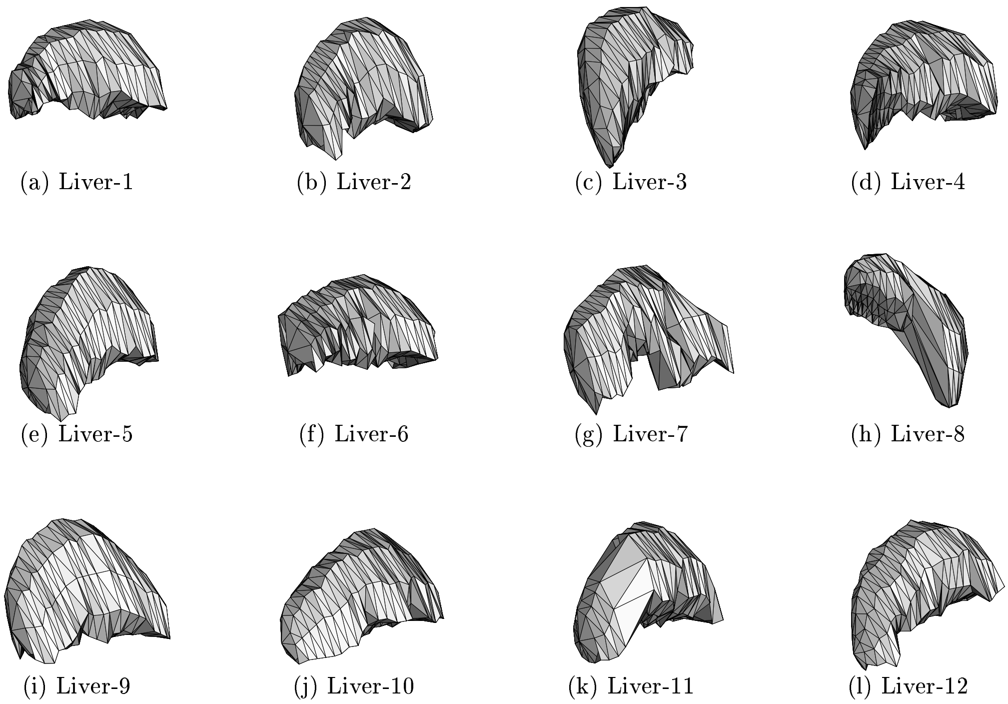


Figure 4.2: Size-normalised volumetric liver models

- In our growth model, cell proliferation causes patterns of mass redistribution and ECM deformation. Eigenmodes describe these patterns, so they are the modes of *organ shape variation*, with each eigenmode describing a pattern at a particular spatial scale.
- *Organ shape deformation* occurs when external forces act upon the organ[3], whose response is approximated by the ECM's elastic properties. In this case, the deformation can be divided into constituent eigenmodes, where each eigenmode again exists at a particular spatial scale.

Model-based modes of variation are intended to aid in interpretation of 3D ultrasound images. This is done by comparing the interpretation of organ shape in the image, to the expected organ shape. The plausibility of the interpretation is then determined by the shape difference between the two, where the difference should be made up of the model-based modes.

There are bound to be many sources of shape difference, however. Model-based modes can account for differences due to organ shape variation and organ shape deformation; and Section 4.3 shows how other differences can be included by statistical examination of organ shape data.

There is a *modal stiffness* associated with each eigenmode during organ shape deformation, therefore eigenmodes are ranked by the probability that each participates in a deformation. However, there is no such ranking for eigenmodes in organ shape variation; their relative probabilities are not easily deducible from our growth model.

In order to incorporate both shape variation and shape deformation into model-based modes, the assumption is made that both are ranked by modal stiffness. In other words, eigenmodes are

[3]Deformation is a factor in ultrasound imaging, because organs are supported and shaped by neighbouring structures and gravitational forces.

weighted by the elastic energy each requires for a given amplitude of variation or deformation. Sections 4.3–4.5 show how this assumption can be iteratively refined.

Relative weighting is important because shape differences can be accounted for by many combinations of eigenmodes, and the appropriate combination can only be chosen by knowing the relative likelihood of each eigenmode in the combination.

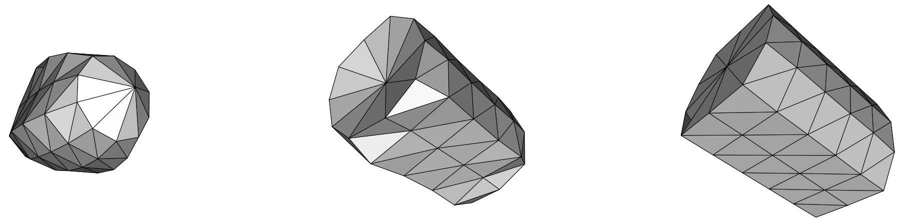
Size of modal subspace

Another issue to be examined is the number of model-based modes of variation to be used. Model-based modes are computed as elastic eigenmodes, and every additional eigenmode required increases computational expense. 'Higher' eigenmodes are also computed with less accuracy (see Appendix B), and are less well represented on the finite element mesh due to their smaller spatial wavelength. 'Lower' eigenmodes should therefore be used in preference to 'higher' ones.

A set of selected eigenmodes can be thought of as forming a subspace of shape variation. Since eigenmodes are ordered in decreasing order of spatial scale, the subspace size is determined by the 'highest' eigenmode, which has the smallest desired spatial wavelength; the choice of this wavelength is dictated by the resolution to which the expected shape variations are defined. The two important determinants of this resolution are therefore the quantity and quality of information about the expected shape variation.

For example, Chapter 5.2 performs volume estimation by fitting an organ shape model to landmarks selected in a 3D ultrasound image. If very few landmarks are selected, then the subspace used in fitting should be kept small, since landmark information is insufficient for fitting to a high spatial resolution. Similarly, if landmarks are badly selected, then fitting should only proceed at low spatial resolution.

This is demonstrated in Figure 4.3, which shows an ellipsoid model deforming to fit a box[4]. The size of the modal subspace used determines the spatial resolution at which the box is reconstructed.



(a) Using 21 modes (b) Using 60 modes (c) Using 180 modes

Figure 4.3: Deformation from ellipsoid to box using different numbers of eigenmodes

Since a modal subspace consists of all eigenmodes below a certain spatial-frequency, it can be thought of as a form of 'low-pass' filter during reconstruction. Section 4.4 shows that this low-pass subspace is also the most significant in a statistical model of organ shape variation.

Elastic eigenmodes are linearised approximations to elastic eigenfunctions, i.e. they are derived under the assumption that eigenmode amplitudes will be kept small. Eigenmode deformations must therefore be kept within some linearisation limit, for a given approximation error. Low-pass eigenmodes have large spatial scale, so their linearisation limits are correspondingly large; a *low-pass subspace* therefore maximises the range of (linearised) shape variations that can be represented by model-based modes of variation.

[4]Note the Gibbs phenomenon in Figure 4.3(b), where the deformed ellipsoid 'overshoots' the box's edges.

Non-linear eigenmodes

The range of shape variation in some complex structures (e.g. in foetal imaging), is too large to be accurately described by eigenmodes which are linearised under a small-amplitude assumption. Non-linearities in shape variation, such as articulated motion in foetal limbs, are particularly difficult to model.

Biomechanical models of such structures can be solved using numerical methods, but computational expense scales unreasonably for large systems with many degrees-of-freedom (Setio et al. [132]). Szemplinkska-Stupnicka [152] generalises modal analysis to such situations using *non-linear eigenmodes*, and Setio et al. [132] show that non-linear eigenmodes can be used to estimate solutions for large systems with *non-linear stiffness*.

Non-linear modal analysis allows complex shape variations to be decoupled into independent non-linear eigenmodes, where each eigenmode describes a component of non-linear shape variation at a certain spatial scale. The cost and spatial resolution of the analysis is then determined by the number of eigenmodes used.

Consider the equilibrium of a non-linearly elastic body, whose response is dictated by non-linear restoring forces $\mathbf{f}(\mathbf{u})$

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} + \mathbf{f}(\mathbf{u}) = \mathbf{0} \tag{4.9}$$

The behaviour of non-linear elastic elements depends on *relative displacements* between node points, so a coordinate transformation $\mathbf{Z}$ is used to map from nodal displacements $\mathbf{u}$ to (a vector of) relative displacements $\mathbf{z}$. Components of $\mathbf{z}$ are assembled from differences between nodal displacement components $\mathbf{u}_i^{\mathrm{n}}$ and $\mathbf{u}_j^{\mathrm{n}}$ to give $\mathbf{z} = \mathbf{Z}^{-1}\mathbf{u}$.

The relative mass matrix $\tilde{\mathbf{M}}$ and relative stiffness matrix $\tilde{\mathbf{K}}$ are then given by

$$\begin{aligned}
\tilde{\mathbf{M}} &= \mathbf{Z}^T\mathbf{M}\mathbf{Z} && (4.10)\\
\tilde{\mathbf{K}} &= \mathbf{Z}^T\mathbf{K}\mathbf{Z} && (4.11)\\
\tilde{\mathbf{f}}(\mathbf{z}) &= \sum_r \alpha_r \mathbf{z}^r && (4.12)
\end{aligned}$$

The relative restoring forces vector $\tilde{\mathbf{f}}(\mathbf{z})$ is a polynomial model, where α_r weights the desired polynomial components. The equilibrium system (4.9) is now in relative coordinates

$$\tilde{\mathbf{M}}\ddot{\mathbf{z}} + \tilde{\mathbf{K}}\mathbf{z} + \tilde{\mathbf{f}}(\mathbf{z}) = \mathbf{0} \tag{4.13}$$

Non-linear modal analysis of equation (4.13) is performed by projection onto non-linear eigenmodes $\tilde{\boldsymbol{\phi}}_i(q_i)$, each defined as a function of modal amplitude q_i

$$\left(\tilde{\mathbf{K}} + \tilde{\mathbf{K}}^{\mathrm{nl}}(q_i)\right)\tilde{\boldsymbol{\phi}}_i = \omega_i^2 \tilde{\mathbf{M}}\tilde{\boldsymbol{\phi}}_i \tag{4.14}$$

Equation (4.14) is a non-linear eigenproblem and is not, in general, soluble using standard linear methods. Setio et al. [132] introduce a Newton-Raphson procedure for approximating non-linear relative eigenmodes and eigenvalues $(\tilde{\boldsymbol{\phi}}_i, \omega_i^2)$ as a function of modal amplitude q_i. The iteration is initialised with linearised eigenmodes and eigenvalues, and non-linear stiffness parameters $\tilde{\mathbf{K}}^{\mathrm{nl}}$ are estimated from the non-linear system's forced harmonic response.

The intention of this basic analysis is to show one method (there may be many) for extending modal analysis to non-linear shape variation. Non-linear modal analysis is not commonly used, however, due to the difficulties involved in determining non-linear material properties.

4.3 Data-driven modes of variation

An alternative approach to deriving modes of shape variation from a growth model, is to compute them from a dataset of organ shapes. These are *data-driven modes* of variation.

If a dataset is available wherein each shape is sampled by the same landmark configuration, then *principal component analysis* (PCA) is an attractive method for analysing shape variation in the dataset (Cootes et al. [33]). Each sampled shape is a vector $\mathbf{x}_i$, and the mean shape vector is $\overline{\mathbf{x}}$; a *sample covariance matrix* $\mathbf{S}$ is then assembled from n such samples in the dataset

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \tag{4.15}$$

The covariance matrix is very useful for shape analysis, because it contains information about the variability of landmarks; it also contains information about the extent to which variation in one landmark is related to variation in others. All these variations and covariations constitute a 'shape space', within which all samples in the dataset range.

PCA extracts principal component vectors $\boldsymbol{\phi}_i$ from this shape space, and the aim is to have most of the shape space represented using a small number of vectors. This is achieved by having each principal component vector $\boldsymbol{\phi}_i$ maximise its projected variance $\boldsymbol{\phi}_i^T\mathbf{S}\boldsymbol{\phi}_i$, subject to the magnitude normalisation $\boldsymbol{\phi}_i^T\boldsymbol{\phi}_i = 1$.

The first principal component $\boldsymbol{\phi}_1$ is found using a Lagrange multiplier λ_1 to perform the maximisation (Therrien [155])

$$\begin{aligned} \frac{d}{d\boldsymbol{\phi}_1}\left(\boldsymbol{\phi}_1^T\mathbf{S}\boldsymbol{\phi}_1 - \lambda_1(\boldsymbol{\phi}_i^T\boldsymbol{\phi}_i - 1)\right) &= 2\mathbf{S}\boldsymbol{\phi}_1 - 2\lambda_1\boldsymbol{\phi}_1 \\ &= \mathbf{0} \end{aligned} \tag{4.16}$$

which gives a standard eigenproblem for $\boldsymbol{\phi}_1$

$$\mathbf{S}\boldsymbol{\phi}_1 = \lambda_1\boldsymbol{\phi}_1 \tag{4.17}$$

The second principal component $\boldsymbol{\phi}_2$ is found in a similar fashion, with the additional constraint of mutual orthogonality $\boldsymbol{\phi}_1^T\boldsymbol{\phi}_2 = 0$. It turns out that the principal components are identically the eigenvectors of $\mathbf{S}$, so the required data-driven modes of variation are given by

$$\mathbf{S}\boldsymbol{\phi}_i = \lambda_i\boldsymbol{\phi}_i \tag{4.18}$$

The projected variance of each principal component is given by its eigenvalue

$$\begin{aligned} \boldsymbol{\phi}_i^T\mathbf{S}\boldsymbol{\phi}_i &= \boldsymbol{\phi}_i^T\lambda_i\boldsymbol{\phi}_i \\ &= \lambda_i \end{aligned} \tag{4.19}$$

A modal subspace of data-driven modes can therefore be chosen to account for any desired proportion of variation in the shape space.

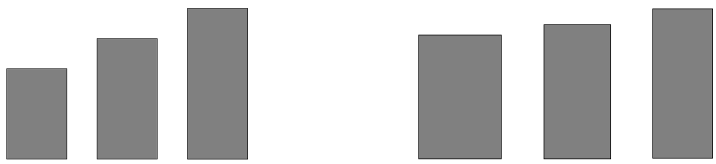
Point distribution model

The mean shape vector and its principal components of variation constitute a *point distribution model* (PDM), which Cootes et al. [33] apply in an active contour framework for segmenting 2D and 3D images. Grenander et al. [58] describe a similar statistical model of boundary shape, which changes position and shape to conform to image data. Sozou et al. [137, 138] describe non-linear implementations of the PDM.

The advantage of the PDM is that its data-driven modes of variation are specific to the training set. Some classes of object are more obviously suited to this approach, e.g. in manufactured objects Gaussian distributions of landmark positions are expected to result from cumulative errors in the manufacturing process. However, Cootes & Taylor [33] also demonstrate that PDMs (of hands and heart ventricles) can be used in segmenting biological structures which may have non-uniform (and non-linear) distributions of landmark variation.

In the PDM, the sample covariance matrix **S** in equation (4.15) is assembled from landmarked shapes which have been normalised for translation, rotation, and scaling[5]. Section 4.1 has already remarked on the danger of size-normalisation corrupting shape information, so training sets for a PDM must be assembled with care in this respect.

The danger of inappropriate size normalisation is illustrated in Figure 4.4(b) which shows that the size-normalised modes of variation are no longer entirely vertical, as was the case before normalisation in Figure 4.4(a).



(a) Original training set

(b) Size-normalised training set

Figure 4.4: Shape and size are not trivially separable: size normalisation can corrupt shape information in training sets

Training sets

PDMs can be constructed more easily than other application specific models, such as Yuille et al.'s [173] deformable templates for face recognition. Assumptions implicit in PDM training are that landmark covariances can be accounted for by a small number of significant principal components, and that shape statistics are stationary.

The training set must be large enough to ensure a statistically significant sampling of landmark variations, and landmarks have to be chosen accurately, consistently, and in homologous ('biologically corresponding') locations across training examples. Although the training framework is easily extensible to 3D, obtaining training examples and choosing landmarks is significantly more difficult, particularly for 3D freehand ultrasound images.

One is faced with a 'bootstrap' problem:

(i) how to segment a 3D ultrasound image without a constraining model;

(ii) and how to obtain training examples to build the model, without first segmenting the 3D ultrasound image.

This problem is important, because a sufficiently large and representative 3D training set is required for each organ being modelled.

Another problem with PDMs is that the landmark configuration for a training set is defined *a priori*. Information from new (or degenerate) landmark configurations cannot be incorporated, without updating the configurations of all other training examples.

[5]The weighted sum of landmark square distances is minimised; see Bookstein (ch.7.1)[16] for an analysis of why this *Procrustes metric* may be systematically misleading cf. the centroid size metric.

Representation of prior knowledge

All models of shape variation (e.g. PDM or deformable templates) incorporate implicit assumptions about the range of allowed shapes, which are manifested in the parameterisation or representation of the model's shape and shape variation. For example, the PDM limits shape variation to a small number of standard deviations for each principal component. It is important, however, that these assumptions do not unduly distort the space of permitted shape variation.

With this warning in mind, a model is desired which can cope with new shape examples, the space of which can be constrained *a priori* either:

(i) by a PDM training set, which *includes* all likely organ shape variations;

(ii) or by a model of biological growth, which *predicts* the likely types of shape variations.

An analogy to this choice is the use of Gaussian lookup tables, versus the Gaussian expressed as a function of mean and variance. The former is efficient for reference purposes, but the latter model lends itself to analysis and (ultimately) an explanation of form[6]. After all, the gathering of experimental data is almost always motivated by the desire to explain the form of the data using an analytic model[7].

For example, our organ growth model in Chapter 2 allows an analytic prediction of modes of organ shape variation; if it is a complete and accurate model, it should accommodate the organ shape variations seen in experimental datasets.

Model refinement

Unfortunately, the growth model is far from complete; although it has proved useful in predicting modes of shape variation, it also needs to refine itself from training data[8]. This suggests a solution to the 'bootstrap' problem for 3D ultrasound images:

(i) model-based modes of variation constrain the fitting of an organ model to the 3D ultrasound image, resulting in an organ segmentation;

(ii) each segmented example allows the organ model to refine its mean shape and its modes of variation (which are now data-driven)[9].

The mean shape is initialised, or 'bootstrapped', in (i) using a finite element organ model, and its modes of variation are similarly bootstrapped using eigenmodes. The linear form of the mean shape vector and modes of shape variation allows efficient refinement from data in (ii).

Section 4.4 develops a common framework for combining model-based and data-driven modes of variation. Once this framework is established, Section 4.5 derives the iterative formulae for model refinement.

[6] The central limit theorem for Gaussian distributions (Kreysig (ch.24.6)[77]).

[7] Bookstein & Green [18] also suggest that organ data can be examined in the inverse sense using an analytic growth model, for the *causes* of deformation and variation.

[8] Pentland & Horowitz [113] use a Kalman filter to track 3D time-series model deformation: this can potentially be extended to estimate and refine model parameters.

[9] The FEM interpolation functions (see Appendix A) of an organ shape model allow integration of landmark information at arbitrary locations cf. previous criticism of the PDM.

4.4 Statistical distribution of eigenmodes

The aim now is to develop a common framework for model-based and data-driven modes of shape variation. Such a framework allows an organ model's (data-driven) modes of variation to be bootstrapped using (model-based) eigenmodes, and then refined iteratively. Martin et al. [88] and Cootes & Taylor [32] have previously examined a more limited fusion of the advantages of *a priori* physical models with statistical models.

The strategy is to derive the statistical distribution of the organ model's modes of variation, under the assumption that each of its eigenmodes are excited to the same amplitude. This statistical distribution has a covariance matrix whose principal components can be found as shown in Section 4.3. A direct relationship can then be established between the principal components and the organ model's eigenmodes.

Uniform eigenmode excitation

The statistical distribution of $\boldsymbol{\phi}_i$ under uniform excitation of all eigenmodes is first examined. Consider a peturbation $\mathbf{f}$ to the elastic system in equation (3.1), which results in a steady-state displacement of the system

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{f} \tag{4.20}$$

Let $\mathbf{f}' = \mathbf{\Phi}^T\mathbf{f}$ be the projection of the peturbation vector $\mathbf{f}$ onto the eigenmode bases $\mathbf{\Phi}$. The components of $\mathbf{f}'$ give the individual peturbations of each eigenmode.

Eigenmodes are conventionally $\mathbf{M}$-orthonormal, so from equations (3.4)–(3.5) in Chapter 3.2, the mass matrix $\mathbf{M}$ and the stiffness matrix $\mathbf{K}$ obey the following identities

$$\mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi} = \mathbf{I} \tag{4.21}$$

$$\mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = \mathbf{\Omega}^2 \tag{4.22}$$

Equation (4.22) shows that ω_i^2 is the modal stiffness, therefore the elastic strain energy W_i^{el} for an eigenmode $\boldsymbol{\phi}_i$ of amplitude q_i (cf. Appendix A.8) is

$$W_i^{\mathrm{el}} \propto \omega_i^2 q_i^2 \tag{4.23}$$

If each eigenmode is peturbed to the same overall amplitude $q_i\|\boldsymbol{\phi}_i\|$, then each modal energy will be

$$W_i^{\mathrm{el}} \propto \frac{\omega_i^2}{\|\boldsymbol{\phi}_i\|^2} \tag{4.24}$$

Excitation covariance

From equation (4.24), modal variance is of the form

$$\sigma_i^2 = \frac{\omega_i^2}{\|\boldsymbol{\phi}_i\|^2} \tag{4.25}$$

but since the eigenmodes in $\mathbf{\Phi}$ are independent and orthogonal, their (diagonal) covariance matrix due to $\mathbf{f}'$ is simply

$$\mathbf{\Sigma}_{\mathbf{f}'} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Omega}^2 \tag{4.26}$$

From the identities in equations (4.21)–(4.22)

$$\mathbf{\Sigma}_{\mathbf{f}'} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}(\mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi})$$

$$
\begin{aligned}
&= (\mathbf{\Phi}^{-1}\mathbf{K}\mathbf{\Phi}) \\
&= (\mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi})(\mathbf{\Phi}^{-1}\mathbf{K}\mathbf{\Phi}) \\
&= \mathbf{\Phi}^T\mathbf{M}\mathbf{K}\mathbf{\Phi}
\end{aligned} \tag{4.27}
$$

Since $\mathbf{f}' = \mathbf{\Phi}^T\mathbf{f}$

$$
\begin{aligned}
\mathbf{\Sigma}_{\mathbf{f}'} &= \mathbf{\Phi}^T\mathbf{\Sigma}_{\mathbf{f}}\mathbf{\Phi} \\
&= \mathbf{\Phi}^T\mathbf{M}\mathbf{K}\mathbf{\Phi} \\
\Rightarrow \quad \mathbf{\Sigma}_{\mathbf{f}} &= \mathbf{M}\mathbf{K}
\end{aligned} \tag{4.28}
$$

Finally from equation (4.20) the covariance of the system's displacement $\mathbf{u}$ is

$$
\begin{aligned}
\mathbf{\Sigma}_{\mathbf{u}} &= \mathbf{K}^{-1}\mathbf{\Sigma}_{\mathbf{f}}\mathbf{K}^{-T} \\
&= \mathbf{K}^{-1}\mathbf{M}
\end{aligned} \tag{4.29}
$$

Equivalence of statistical and model-based systems

Recall that FEM equilibrium in (3.1) leads to the generalised eigenproblem

$$\mathbf{K}\mathbf{\Phi} = \mathbf{M}\mathbf{\Phi}\mathbf{\Omega}^2 \tag{4.30}$$

where $\mathbf{\Phi}$ is the matrix of eigenmodes, which are *model-based modes* of variation.

If $\mathbf{\Phi}$ is interpreted as a matrix of principal component vectors, or *data-driven modes* of variation, then from equation (4.18) the principal components have a covariance matrix $\mathbf{\Sigma}$ given by

$$\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda} \tag{4.31}$$

where $\mathbf{\Lambda}$ is the matrix of principal variances.

By comparing equation (4.31) with equation (4.30), and using $\mathbf{\Sigma} = \mathbf{K}^{-1}\mathbf{M}$ from equation (4.29), it can be seen that[10]

$$\mathbf{\Omega}^2 = \mathbf{\Lambda}^{-1} \tag{4.32}$$

Therefore the principal variances in $\mathbf{\Lambda}$ are inversely proportional to the squares of the natural frequencies of the elastic system

$$\lambda_i = \frac{1}{\omega_i^2} \tag{4.33}$$

We have now shown that there is a simple relationship in equation (4.29), between the elastic system governed by $(\mathbf{M}, \mathbf{K})$ and the statistical system governed by $\mathbf{\Sigma}$. The modes of variation for both systems are in fact identical, i.e. the model-based eigenmodes of $(\mathbf{M}, \mathbf{K})$ are identical to the data-driven principal components of $\mathbf{\Sigma}$.

In addition, the low-pass modal subspace discussed in Section 4.2 is now seen to be the most significant statistical subspace, since eigenmodes of low spatial-frequency (low ω_i) are also principal components with significant variance (high λ_i).

The result is that the elastic eigenmodes and eigenvalues can be used to bootstrap a set of principal components and principal variances, which can then be iteratively refined from data using the formulae derived next in Section 4.5.

[10] In this case it is assumed that the matrix system $(\mathbf{M}, \mathbf{K})$ has been deflated (see Appendix B.3) to exclude the rigid-body nullspace; or equivalently that principal component analysis is performed after translation and rotation-normalisation of samples. This means that $\mathbf{K}$ is full-rank and hence invertible.

4.5 Refining models of shape variation

It remains now to derive formulae for updating a data-driven model of shape variation using new shape examples. An organ model is first bootstrapped using a finite element shape model and its elastic eigenmodes, with an initial weighting of (say) $n = 10$.

Having been updated n times, the organ model is described by its mean shape $\overline{\mathbf{x}}_n$, and its sample covariance matrix $\mathbf{S}_n$; its modes of variation are simply the principal components of $\mathbf{S}_n$.

Updating $\overline{\mathbf{x}}_{n+1}$

From the definition of $\overline{\mathbf{x}}_n$

$$\begin{aligned}
\overline{\mathbf{x}}_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \\
&= \frac{1}{n}\left(\sum_{i=1}^{n-1}\mathbf{x}_i + \mathbf{x}_n\right) \\
&= \frac{n-1}{n}\,\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbf{x}_i + \frac{1}{n}\mathbf{x}_n \\
\Rightarrow \overline{\mathbf{x}}_n &= \frac{1}{n}\left((n-1)\overline{\mathbf{x}}_{n-1} + \mathbf{x}_n\right) \qquad (4.34)
\end{aligned}$$

the iterative rule for updating the mean shape vector from a new example $\mathbf{x}_{n+1}$ is given by

$$\overline{\mathbf{x}}_{n+1} = \frac{1}{n+1}\left(n\overline{\mathbf{x}}_n + \mathbf{x}_{n+1}\right) \qquad (4.35)$$

Updating $\mathbf{S}_{n+1}$

From the definition of $\mathbf{S}_n$

$$\begin{aligned}
\mathbf{S}_n &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i - \overline{\mathbf{x}}_n\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i^2 - n\overline{\mathbf{x}}_n^2\right) \\
\Rightarrow \frac{n-1}{n}\mathbf{S}_n &= \frac{1}{n}\left(\sum_{i=1}^{n}\mathbf{x}_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\mathbf{x}_i\right)^2\right) \qquad (4.36)
\end{aligned}$$

and from the definition of $\mathbf{S}_{n+1}$

$$\begin{aligned}
\mathbf{S}_{n+1} &= \frac{1}{n}\sum_{i=1}^{n+1}\left(\mathbf{x}_i - \overline{\mathbf{x}}_{n+1}\right)^2 \\
&= \frac{1}{n}\left(\sum_{i=1}^{n+1}\mathbf{x}_i^2 - (n+1)\overline{\mathbf{x}}_{n+1}^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n+1}\mathbf{x}_i^2 - \frac{1}{n+1}\left(\sum_{i=1}^{n+1}\mathbf{x}_i\right)^2\right) \qquad (4.37)
\end{aligned}$$

Subtracting equation (4.36) from equation (4.37) gives

$$\mathbf{S}_{n+1} - \frac{n-1}{n}\,\mathbf{S}_n \;=\; \frac{1}{n}\,\mathbf{x}^2_{n+1} - \frac{n+1}{n}\,\overline{\mathbf{x}}^2_{n+1} + \overline{\mathbf{x}}^2_n \tag{4.38}$$

whence the iterative rule for updating the sample covariance matrix from a new example $\mathbf{x}_{n+1}$

$$\mathbf{S}_{n+1} \;=\; \frac{n-1}{n}\,\mathbf{S}_n + \frac{1}{n}\,\mathbf{x}^2_{n+1} - \frac{n+1}{n}\,\overline{\mathbf{x}}^2_{n+1} + \overline{\mathbf{x}}^2_n \tag{4.39}$$

4.6 Summary

Elastic eigenmodes describe an organ's modes of shape variation, according to the growth model developed in Chapter 2. These modes are used for volume estimation by fitting organ models to 3D ultrasound images in Chapter 5, so it is important that they are able to improve their representation of shape variation from training data.

The framework developed in this chapter allows an organ model to be 'bootstrapped' using elastic eigenmodes, so that it can immediately be used to segment 3D ultrasound images. After each segmentation is completed, the organ model iteratively refines its modes of variation using the newly segmented examples. These refined modes should converge to being the best statistical estimate of linear shape variation in the training data.

This chapter also develops a metric for normalising the size of the organ model relative to (landmarks in) the 3D ultrasound image. This minimises the risk that modes of variation will be forced beyond their linearisation limits, when fitting the model to the 3D image. The size metric also finds other applications in Chapter 8, when fitting two organ models to each other.

Chapter 5

Experiments: 3D Volume Estimation

Chapter 4 proposes a framework for organising knowledge of expected organ shape and expected variation in shape, where the prior expectations can be refined by incorporating new organ shape examples.

This framework of iterative refinement is applied in this chapter to volume estimation from 3D ultrasound images. A computer application 'bootstraps' itself from a finite element organ model and its eigenmodes, and volume is estimated by fitting the organ model to operator-selected landmarks in the 3D ultrasound image. The organ model then refines its expectations of typical organ shape using the segmented example, so that it can perform future fitting more accurately (Syn et al. [148, 150]). The clinical utility of this operator-assisted approach is validated by examining the accuracy of the volume estimates.

Segmented ultrasound images are useful for diagnostic or monitoring purposes, such as volume estimation and visualisation of 3D anatomy. However, automated segmentation is not usually possible, because tissue boundaries in the images cannot be reliably detected. As more reliable tissue segmentation techniques are developed, an organ shape model can be used to integrate segmentation cues from multiple freehand images, each with a different orientation and quality of information. The ultimate aim is to complement our *model-based* framework with *data-driven* techniques of tissue segmentation.

Chapter organisation

- **Section 5.1** describes the `stradview` application, which comprises the `geomwish` interface for manipulating organ shape models, and the `stradwish` interface for manipulating 3D ultrasound image sequences. These interfaces allow an operator to select corresponding landmarks between a model and an image.

- **Section 5.2** shows how these corresponding landmarks cause the organ shape model to deform to fit the 3D ultrasound image. The volume of the fitted model is then estimated.

- **Section 5.3** discusses the facilities available in `stradview` for operator-assisted and automated refinement of an organ segmentation.

- **Section 5.4** presents results which validate this method of volume estimation against measured and MRI-estimated volumes.

- **Section 5.5** shows segmented livers which have been refined using active contour searches.

- **Section 5.6** discusses volume rendering and volume estimation in 3D ultrasound imaging.

5.1 The `stradview` user interface

The `stradview` user interface for display and manipulation of 3D ultrasound images consists of the `geomwish` and `stradwish` components (described below). It is portable to a wide range of workstation platforms, and can be operated remotely across a computer network.

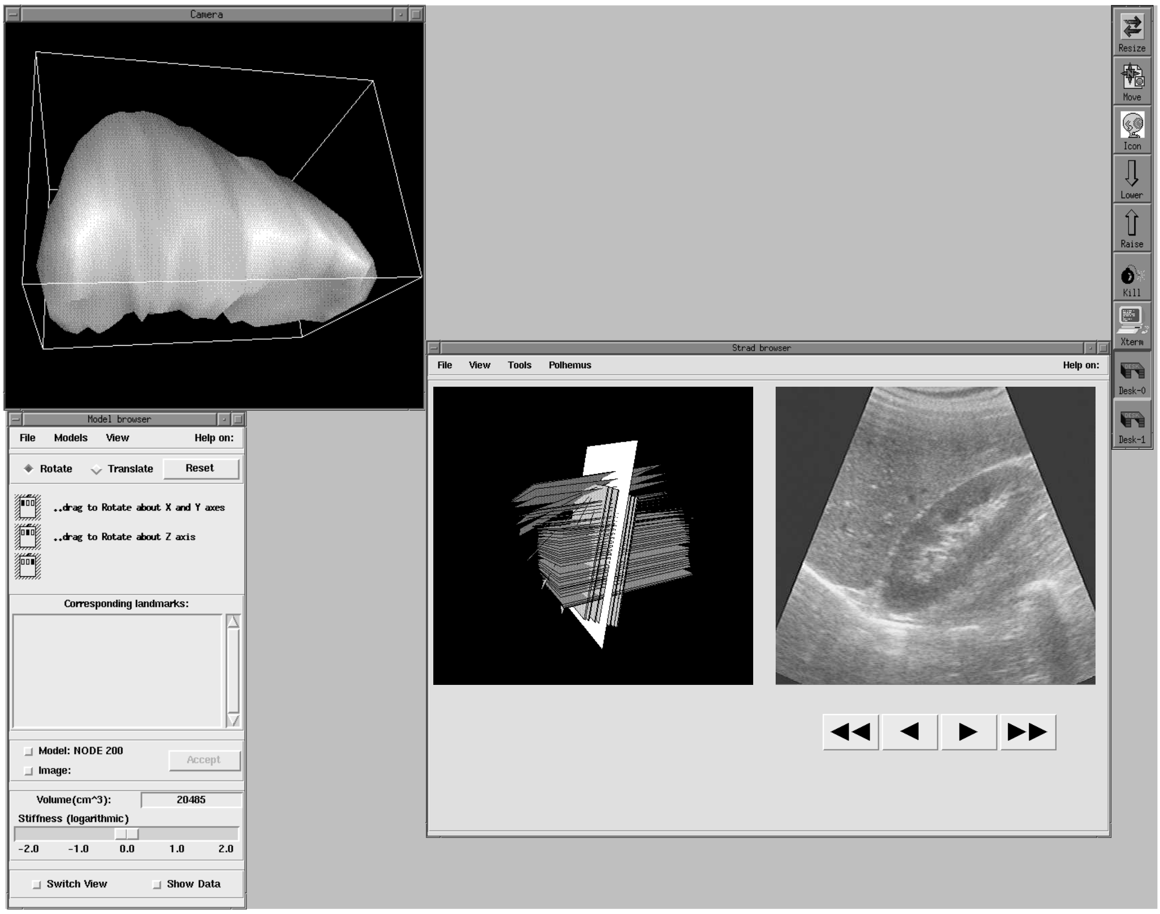


Figure 5.1: Typical screen display of `stradview` user interface for 3D ultrasound images

The `geomwish` interface for model manipulation

The `geomwish` application shown on the left of Figure 5.1 consists of a *camera window* (top left) for display and manipulation of mesh models, and a *command window* (bottom left) for controlling various model parameters.

The Geomview [47] graphics engine drives the camera window using the X Window System [119] for remote display management across computer networks, and has specific support for hardware graphics acceleration on Silicon Graphics workstations, as well as provision for stereo viewing displays.

The `tcl/tk` graphics toolkit and scripting language (Ousterhout [108]) is used to construct the command window, and to control the behaviour of the camera window.

The stradwish interface for image manipulation

The `stradwish` application, shown on the right of Figure 5.1, controls the acquisition of 3D ultrasound image sequences, and provides browsing and editing facilities for the image sequence. It has an internal window showing 3D outlines of ultrasound image frames (the current frame is highlighted as shown), which allows 3D visualisation and manipulation of the frame sequence.

The `tcl/tk` graphical scripting language is used to construct the `stradwish` window, and also to control and interrogate the 3D ultrasound imaging hardware.

5.2 Volume estimation

Chapter 1.2 notes that ultrasound imaging is routinely used for volume estimation. For example, ventricular volume determines the degree of impairment after heart attacks, and indicates response to therapy. Similarly, changes in tumour volume after radiotherapy or chemotherapy determine the nature and intensity of further treatment.

Volume estimation can be achieved by interaction between the `geomwish` and `stradwish` applications: the operator locates an anatomical landmark by browsing through the ultrasound images in `stradwish`, and the corresponding location is located by directly manipulation of an organ model in the `geomwish` camera window.

The organ model first estimates its appropriate size and orientation; it then fits itself to the 3D ultrasound image, where the required shape change is determined by the operator's choice of corresponding landmarks. The volume of the fitted shape is reported by `geomwish`.

Pose and size normalisation

Three correspondences between the model and the 3D image are required to size and orientate the model correctly. If both landmark triplets form exactly similar triangles, then centroid size normalisation (described in Section 4.1) will correctly scale the triangles, so that a rigid-body transformation can be found between them.

The two landmark triplets are never exactly similar, however, since the user will not be able to pick three correspondences in exactly similar configuration, nor will the mesh model be an exactly scaled version of the 3D organ shape.

Nevertheless, a rotation transformation can be estimated by first scaling two landmarks in each triplet to have the same length baseline. The third landmark in one triplet is then temporarily scaled to force both triangles to be similar, from which a rigid-body rotation can be computed. Figure 5.2 illustrates this.

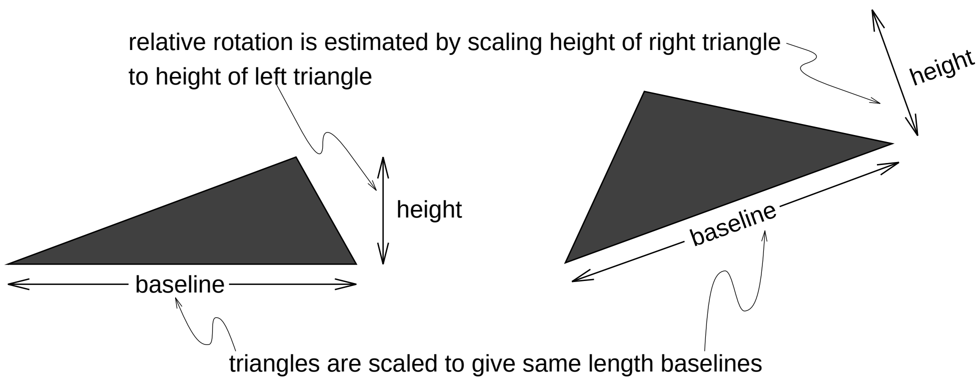


Figure 5.2: Relative rotation is estimated by approximating the triangles to be exactly similar

Given the two 3×3 matrices of triangle coordinates, $\mathbf{L_1}$ and $\mathbf{L_2}$, we wish to estimate the rotation matrix $\mathbf{Q}$ which maps $\mathbf{L_2} = \mathbf{L_1}\mathbf{Q}$. Accumulated round-off errors make computing the rotation using $\mathbf{Q} = \mathbf{L_1}^{-1}\mathbf{L_2}$ unsafe, so Golub & van Loan (ch.12.4.1)[51] describe a more robust approach using the error measure[1] $\|\mathbf{L_2} - \mathbf{L_1}\mathbf{Q}\|_F$.

The *singular value decomposition* is computed for $\mathbf{L_1}^T\mathbf{L_2}$

$$\mathbf{U}^T(\mathbf{L_1}^T\mathbf{L_2})\mathbf{V} = \mathbf{D} \tag{5.1}$$

where $\mathbf{D}$ is the diagonal matrix of singular values, and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. The rotation transformation which minimises $\|\mathbf{L_2} - \mathbf{L_1}\mathbf{Q}\|_F$ is then given by $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$.

If the singular value decomposition is ill-conditioned, $\mathbf{Q}$ describes a reflection transformation with negative determinant, rather than a rotation transformation with positive determinant; in such cases the triangle landmarks are peturbed with some noise and a rotation estimated again. This gives rise, overall, to a reliable method for size and pose estimation requiring only the first three correspondences.

Model fitting

Figure 5.3(a) shows an ellipsoid model with three correspondences, after size and rotation normalisation. Each red line connects a landmark on the surface of the ellipsoid, to a red dot indicating its corresponding landmark.

This represents a situation typical in 3D ultrasound segmentation: some few anatomical landmarks are recognised in the ultrasound images, from which one wishes to estimate the shape (and hence volume) of the entire organ. Since an organ shape model incorporates prior knowledge of the organ's expected shape, it is used to interpolate the segmentation boundary between the landmarks.

Sonographers visualise 3D anatomical structure in a similar way: knowledge of anatomical variation (i.e. modes of shape variation[2]) allows the sonographer to choose the appropriate fit of a prototypical shape model to the ultrasound images. In other words, top-down (model-based) knowledge of organ shape is used to frame bottom-up (data-driven) information about anatomical landmarks.

The fitting process involves deforming the organ model, so that its landmarks approach the corresponding landmarks in the image as closely as possible. The appropriate deformation is computed by employing (three times) as many modes of variation as there are landmarks.

This deformation describes an *exactly-constrained* least-squares problem, which is solved using the QR factorisation (Golub & van Loan (ch.5.3.4)[51]). The problem is posed as

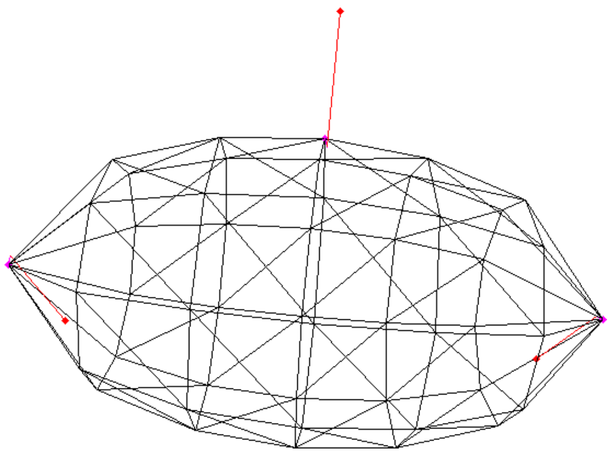
$$\min_{\mathbf{e}} \|(\mathbf{L_1} - \mathbf{L_2}) - \mathbf{\Phi}\mathbf{e}\| \tag{5.2}$$

where $\mathbf{L_1}$ and $\mathbf{L_2}$ are column vectors stacked with landmark positions, and $\mathbf{e}$ is the vector of modal amplitudes to be found. Each column of $\mathbf{\Phi}$ contains a mode of shape variation between the landmark positions.

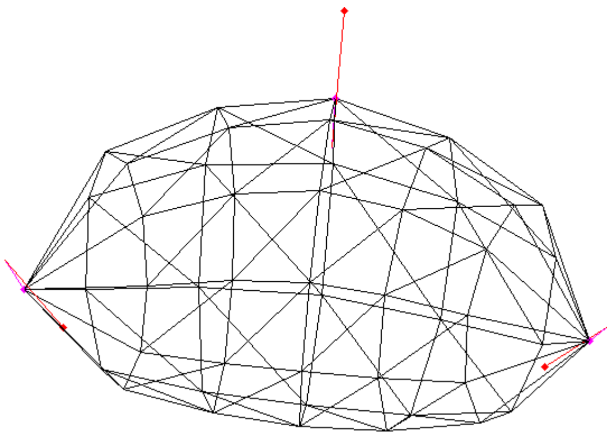
In cases where there is noise in landmark specification, however, a more reliable approach is used which employs all available modes of shape variation. This gives rise to an *under-constrained* least-squares problem, for which an additional global constraint of minimal de-

[1] The *Frobenius norm* $\|\mathbf{A}\|_F^2$ is defined as the sum of squares of the elements of $\mathbf{A}$.

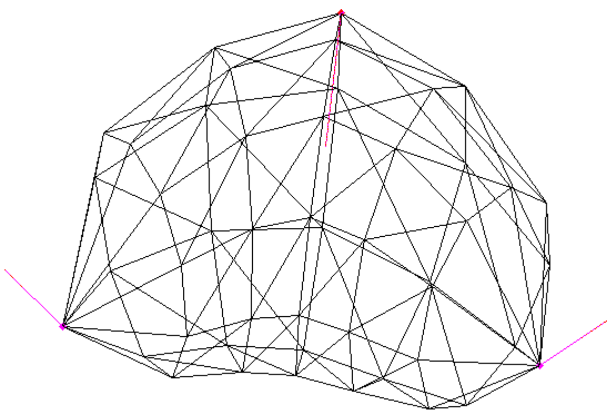
[2] As described in Chapter 4, modes of shape variation are initialised with elastic eigenmodes, and subsequently refined from each new segmentation example.



(a) After scale and rigid-body normalisation between model and landmarks



(b) Model fitting with low confidence, hence high stiffness



(c) Model fitting with high confidence, hence low stiffness

Figure 5.3: Fitting a model to 3D landmark data with 3 correspondences

formation energy $\|\mathbf{\Omega e}\|^2$ is required (see Appendix A.8)

$$\min_{\mathbf{e}} \|(\mathbf{L_1} - \mathbf{L_2}) - \mathbf{\Phi e}\|^2 + \eta \|\mathbf{\Omega e}\|^2 \tag{5.3}$$

The relative weighting η between constraints is specified *a priori*[3]. Differentiating to find the minimum in equation (5.3) gives

$$(\mathbf{\Phi \Phi}^T + \eta \mathbf{\Omega}^2)\mathbf{e} = \mathbf{\Phi}^T(\mathbf{L_1} - \mathbf{L_2}) \tag{5.4}$$

whence $\mathbf{e}$ by inversion, or more safely by least-squares. So the deformation required is described by the modes of shape variation in $\mathbf{\Phi}$, whose amplitudes are the components of $\mathbf{e}$.

The smoothness or stiffness parameter η is determined[4] by the accuracy with which corresponding landmarks are located:

- A high stiffness value implies less confidence in landmark positions, resulting in a more reluctant fit. This is illustrated in Figure 5.3(b) .
- A low stiffness value implies high confidence in landmark positions, resulting in a more confident fit. This is illustrated in Figure 5.3(c).

Volume computation

Once the organ model is fitted to the ultrasound image, its volume is computed using the divergence theorem[5] (Kreysig (ch.9.6)[77])

$$\int\!\!\int\!\!\int_V \nabla\cdot\Gamma \, dV = \int\!\!\int_S \Gamma\cdot\mathbf{n}\, dS \tag{5.5}$$

where $\mathbf{n}$ is the unit outward normal on the organ model's surface S.

By using $\Gamma = (x/3, y/3, z/3)^T$ to set $\nabla\cdot\Gamma = 1$, the organ volume enclosed in V is given by the right-hand surface integral. This surface integral can be broken into a sum over surface elements of the organ model

$$\int\!\!\int\!\!\int_V dV = \sum_j \int\!\!\int_{S_j} \Gamma\cdot\mathbf{n}\, dS_j \tag{5.6}$$

Each jth elemental integral is evaluated numerically via Gaussian quadrature sampling of the element's interpolation functions (see Appendix A.5).

5.3 Editing facilities

An approximate organ segmentation can be made either by model fitting in 3D as described above, or by an operator sketching outlines in each 2D ultrasound image. The segmentation can be improved using automated search techniques, which are limited to a search space close to the initial segmentation; such techniques usually seek intensity edges, and require operator intervention if incorrect edges are found.

[3]If uniform Gaussian uncertainty in landmarks is assumed, then η is inversely proportional to the Gaussian's standard deviation (Blake & Zisserman [12]).

[4]The `geomwish` command window has a sliding bar for interactive adjustment of the stiffness η; this single parameter can, in principle, be experimentally established for a certain user segmenting a certain organ type.

[5]If the organ model consists of tetrahedral elements, its volume can also summed by computing elemental volumes using the scalar triple product (Kreysig (ch.6.9)[77]).

Intensity-based segmentation algorithms assume that different tissues are differentiated by their ultrasound image intensities. Although tissue interfaces do produce large ultrasound echoes, there are various sources of signal loss and distortion as described in Chapter 1.1. These noise sources can obscure tissue boundaries in the image.

More robust automated segmentation algorithms require 2D and 3D texture features computed with regional support at multiple spatial scales, which can accurately discriminate between tissues in the image (Muzzolini et al. [98]). However, there is currently no consensus on the appropriate computational characterisation of tissue 'texture' in ultrasound images.

Editing facilities in `stradview` allow improvements in approximate organ segmentations to be either computed automatically using these intensity-based algorithms, or specified by the operator.

2D image segmentation

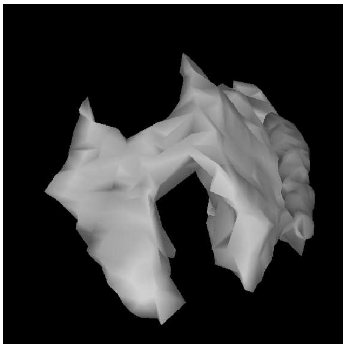


Figure 5.4: Rendered view of a manually segmented thyroid (with portion of carotid artery)

Figure 5.4 shows an early experiment by Gosling et al. [57] in thyroid volume estimation, where the ultrasound images were manually segmented, and the selected landmark points triangulated into an *alpha hull*. This approach is extremely laborious since a few hundred landmarks need to be located in order to achieve a clinically acceptable volume estimate.

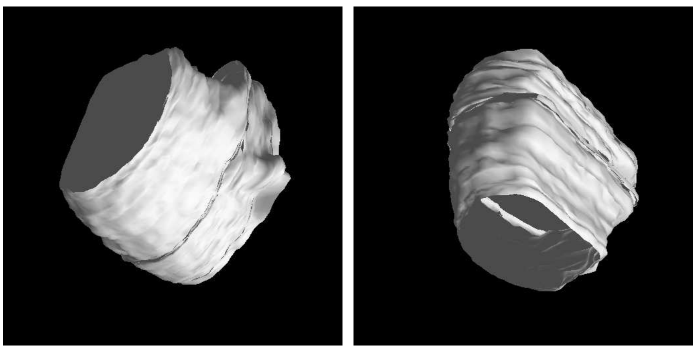
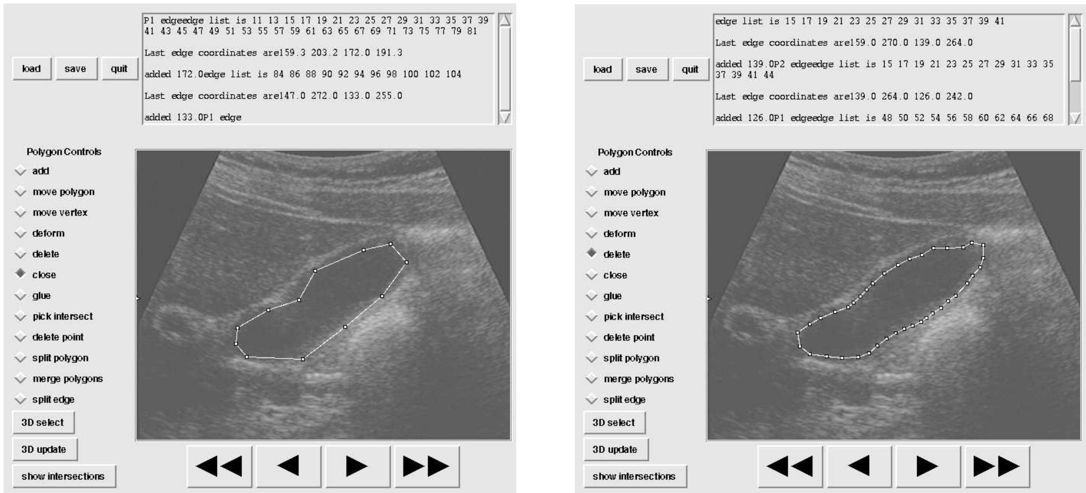


Figure 5.5: Two views of an automated simulated annealing segmentation of a pig's heart

Figure 5.5 shows another early result by Syn et al. [142], computed using an intensity-based segmentation of a pig's heart in a water bath. 2D segmented contours are propagated and connected between nearby frames to construct the 3D hull, with each 2D contour optimised using a computationally expensive simulated annealing search.

Manual segmentation in Figure 5.4 is labour intensive but supervised, and automatic segmentation in Figure 5.5 is computationally intensive but completely unsupervised. Semi-automatic segmentation is a popular compromise, whereby images are manually initialised, then automatically refined by active contours. This is illustrated in Figure 5.6 using facilities from Dance's [36] `abcwish` application.



(a) Manual sketch of gallbladder section (b) Refined segmentation using active contour

Figure 5.6: Typical screen displays of `abcwish` user interface for 2D segmentation

The particular difficulty in freehand imaging is that consecutive image frames are not constrained to be parallel; in fact, standard scanning practice for many organs is to acquire images along at least two orthogonal axes. The `stradwish` window in Figure 5.1 gives a good idea of typical scan trajectory.

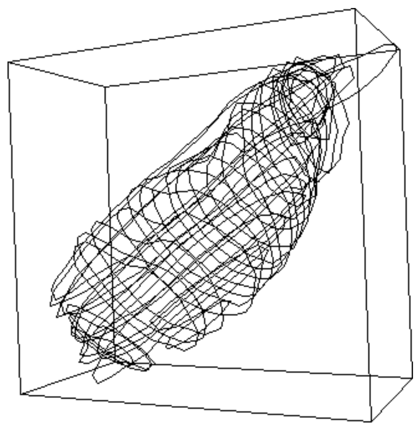


Figure 5.7: A set of 2D segmentation contours produced by `abcwish`

Figure 5.7 shows the set of contours resulting from refined segmentation by `abcwish` of the images in a 3D gallbladder scan; each frame is manually initialised as shown in Figure 5.6(a). Even though gallbladder images have very good intensity contrast, some frames suffer signal loss, leading to contour 'leakage' at the top right of Figure 5.7. Such leakage can potentially be alleviated by imposing gallbladder shape information on the active contours, which merely minimise bending energy in the `abcwish` implementation (i.e. they prefer circular shapes).

A 3D gallbladder segmentation is achieved by connecting the non-parallel gallbladder sections in Figure 5.7 together. Methods for connecting parallel MR and CT segmentations are well established; these methods make certain assumptions about how segmented contours branch and merge between the parallel frames, then apply (some variant of) *Delaunay triangulation* to link contours between frames (Boissonat & Geiger [15]).

However, freehand ultrasound scans have non-parallel and self-intersecting planes, and sometimes contain outlying frame positions due to proprioceptive errors. It is therefore difficult to make reasonable yet reliable assumptions about the way contours link between frames. Dance [36] addresses this issue using a sophisticated triangulation algorithm, which considers the probable ways in which contours can be connected, given their relative spatial positions.

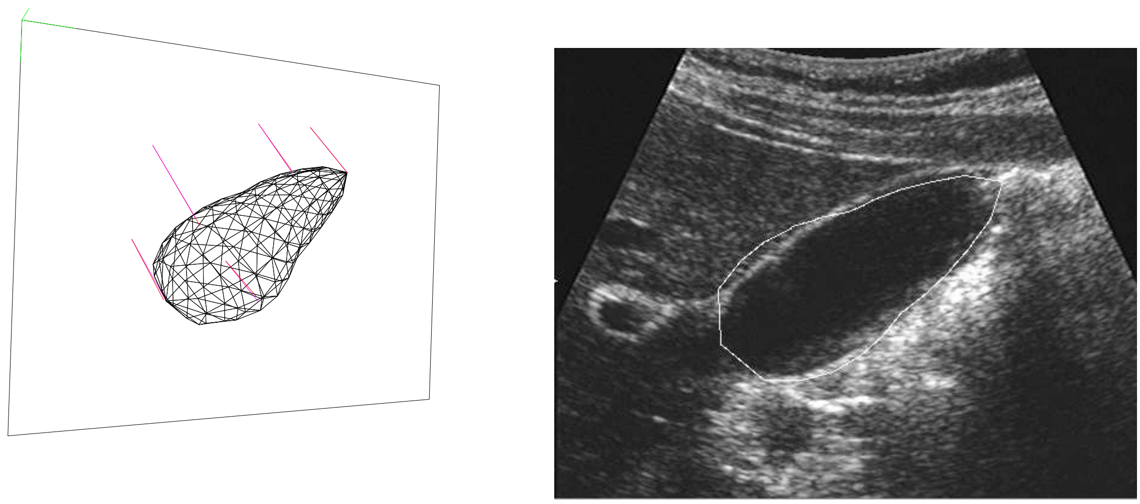
In summary, the `abcwish` approach is driven bottom-up by intensity data, and constrained top-down using a simple elastic spline (i.e. the active contour). Each spline separately performs a 2D segmentation, and the set of segmented contours is connected in 3D by making further assumptions about the proximity and distribution of the contours.

Operator refinement of 3D model-based segmentation

Although such methods are in common currency in MR and CT imaging, they are too *ad hoc* for routine use in 3D freehand sequences, particularly since intensity-based boundary estimation is unreliable in ultrasound imaging.

The 3D model-based framework described in Section 5.2 offers stronger and more specific guidance during segmentation. It directly applies knowledge of typical organ shape and expected shape variation, and also offers the ability to iteratively refine its expectations by incorporating each new segmentation example

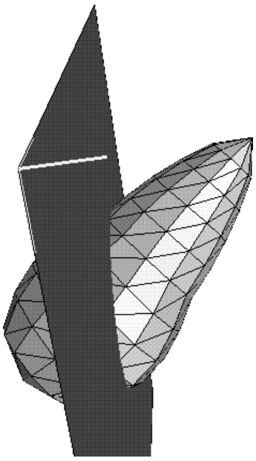
Working in 3D removes the artificial separation of first constraining an active contour segmentation within each frame, then constraining the connection of segmentations between frames using some best-fitting 3D triangulation. An added bonus of 3D model-based segmentation is that fewer landmarks need to be selected; many of the boundary landmarks in 2D-to-3D segmentation are required primarily to drive the triangulation.



(a) Model fitted to 5 landmarks (b) Intersection of model with current scan frame

Figure 5.8: Model-based gallbladder segmentation using 5 landmark correspondences

Figure 5.8 shows the gallbladder segmentation achieved in `stradview`, after adding a further 2 correspondences to the 3 required for initial pose and size normalisation (recall Section 5.2). Note that the tripod marker on the top left of the intersection plane in Figure 5.8(a) corresponds to the top left of the image frame in Figure 5.8(b).

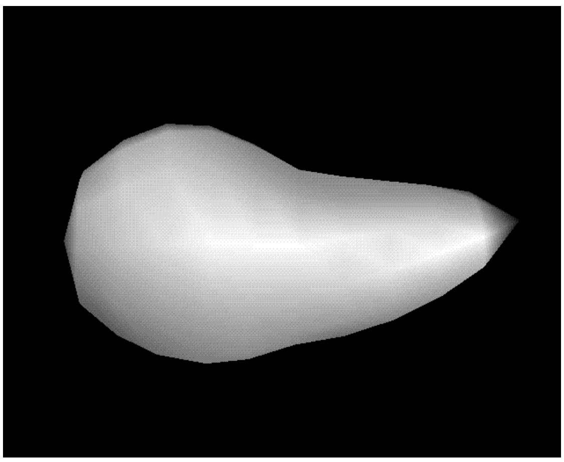


(a) Position of new scan frame

(b) Intersection contour gives segmentation

Figure 5.9: Gallbladder segmentation seen in another scan frame

Figure 5.9 shows the gallbladder segmentation in another frame, scanned across the longitudinal axis; this segmentation contour is formed by intersecting the new scan plane with the deformed gallbladder model. The overall quality of the segmentation depends on the choice of landmark correspondences, and the accuracy of the organ shape model. However, it can be seen that with only 5 correspondences in this example, a good estimate of boundary shape is achieved, even in a cross-section very different to the one in Figure 5.8.



(a) Rendered view of gallbladder segmentation

(b) Intersection contour refined further

Figure 5.10: Further refinement of gallbladder segmentation with 6th correspondence

Figure 5.10 shows an improved segmentation, achieved by selecting a 6th landmark in the new cross-section.

Automated refinement of 3D model-based segmentation

The model-based segmentation can still be refined further by adding more landmarks, or by allowing the operator to 'drag' segmentation contours in the `stradwish` window. An automated method of refining the 3D segmentation is preferred, however, since the search space in which refinement occurs can be easily controlled by restricting the energy of deformation during search.

One possible refinement algorithm applies simulated annealing to solve the optimisation problem of minimising the model's energy of deformation, while at the same time maximising the fit of the 3D model to intensity edges in each 2D ultrasound image. This algorithm has the benefit of finding the globally optimal fit over the search space, but is computationally expensive. Another unattractive feature is that additional parameters need to be specified for the annealing schedule (Syn et al. [142]).

Figure 5.11 shows the result of applying simulated annealing search to a 3D ultrasound image of a liver.

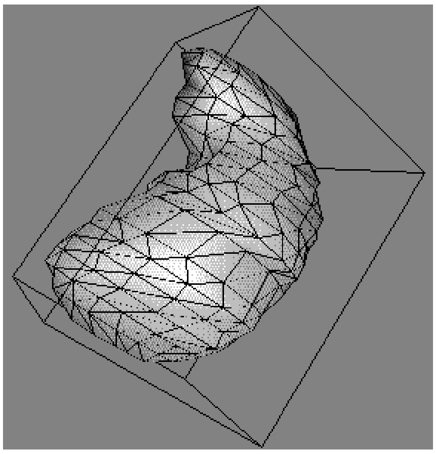


Figure 5.11: Automated refinement of liver segmentation using simulated annealing search

Figure 5.12 shows the result of applying a 3D active contour search to a 3D ultrasound scan of a liver. More results are shown in Section 5.5 of this chapter.

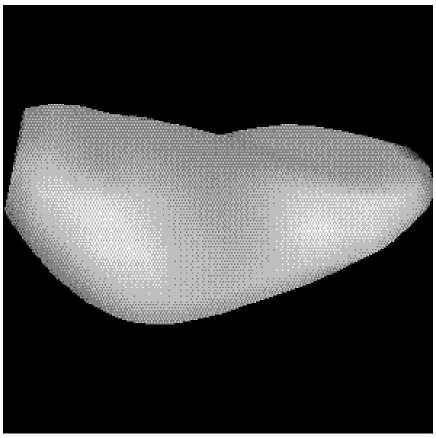


Figure 5.12: Automated refinement of liver segmentation using active contour search

Both refinement algorithms improve the accuracy of volume estimates achieved using the `stradview` paradigm of operator-selection of landmark correspondences. They also enable the model to incorporate new shape variations with greater accuracy.

However, the refined 3D contours can be significantly deformed compared to the organ model's original mesh, and it is usually desirable to resample a deformed contour with a better distributed mesh. Such a resampling can potentially be achieved with the non-rigid registration algorithm described in Chapter 8, further improving the accuracy of the refined modes of shape variation[6].

The segmentation algorithms described in this section are motivated by the need to measure organ volumes to clinically reliable accuracies; they are not intended for strict geometric reconstruction of organ shape[7].

5.4 Results: validation of volume estimates

The accuracy of organ volume estimates is now examined. Estimation is performed using only operator-selection of landmarks, since the segmentation refinement algorithms described previously are not yet reliable enough for routine clinical use,

Model-based volume estimation is first validated by comparing `stradview` estimates to 'gold standard' voxel count estimates. These volume estimates are performed on MRI scans, for which the gold standard can be achieved; the use of MR images also excludes errors due to ultrasound image distortion.

The second validation test is performed on 3D ultrasound images, which include (as yet unquantified) errors in probe proprioception and image distortion (Rohling & Gee [122]). This test compares `stradview` estimates for a balloon in a water bath, to the amount of water actually decanted from the balloon.

Validation using MR liver images

Volumes of foetal livers in 3D MR scans were estimated using `stradview` (Syn et al. [151]), then compared to volumes previously determined by voxel count (Pasapula [110]).

Image	Actual (cm^3)	Estimated (cm^3)	Error (%)
Liver-1	13.2	14.4	9
Liver-2	20.2	21.7	7
Liver-3	24.9	26.2	5
Liver-4	28.3	31.1	10
Liver-5	28.1	29.5	5
Liver-6	16.8	19.0	13

Table 5.1: Volume estimation of foetal livers using ellipsoid model

Table 5.1 shows volume estimates from fitting a simple ellipsoid model to each 3D scan using 5–8 landmarks. There is a small but consistent over-estimate due to the use of an ellipsoid

[6]A deformed 3D contour has a deformed parameterisation imposed on its geometry, and hence on the 'shape vector' which the organ model learns from.

[7]Section 5.6 discusses sources of distortion in ultrasound imaging, which need to be addressed if geometrically faithful reconstruction from 3D freehand ultrasound is to be achieved.

shape model, but estimates can be made arbitrarily accurate (compared to the gold standard) by using a better shape model, and by selecting more landmarks to improve the model's fit.

Figure 5.13 shows one of the MR liver image sequences used. The liver models in Figure 4.2 in Chapter 4 are manually segmented from these sequences.

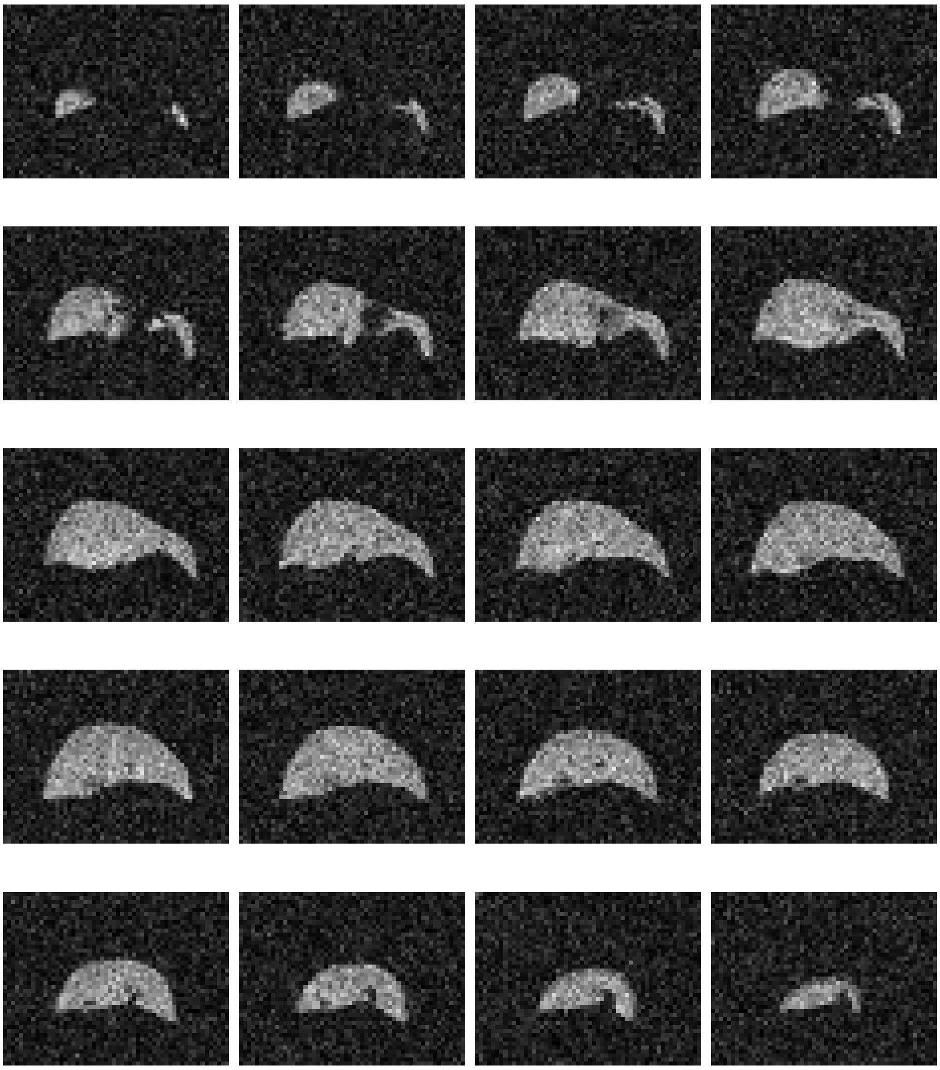


Figure 5.13: MR image sequence of foetal liver (top-left to bottom-right)

Validation using water-filled balloons in ultrasound

Given that model-fitting results in acceptably accurate volume estimates, the next step is to validate the volume estimates achieved with 3D ultrasound images. It should be noted that volume estimation using 2D ultrasound is routinely performed to clinically acceptable accuracy, and a number of publications report that 3D ultrasound also gives accurate volume estimates.

Gilja et al. [49] validate the accuracy of 3D ultrasound volume estimations *in vivo* by comparing 18 human kidney volumes with MRI estimates. Ultrasound volumes are computed by manual segmentation of 81 kidney sections, and a good correlation is reported between the two. A mean underestimate of 10% of average kidney volume is noted in the ultrasound estimates, compared to the MRI estimates.

Hodges et al. [63] validate blood conduit volumes using 3D ultrasound reconstructions of phantoms *in vitro*, and vein bypass grafts *in vivo*. Hughes et al. [65] similarly validate volume measurements of balloons, kidneys, and livers *in vitro* using manual segmentation of freehand ultrasound images.

Factors absent in the above validations are: the use of model-fitting in `stradview`, and freehand reconstructive errors due to probe proprioception and image distortion. The latter errors are also absent from the MR liver volume test performed previously. The importance of these factors in volume estimates from 3D ultrasound images are now examined.

Figure 5.14 is a rendered view of a balloon in a water bath, whose volume was computed by fitting to an ellipsoid model, then compared to the volume of water actually decanted from the balloon.

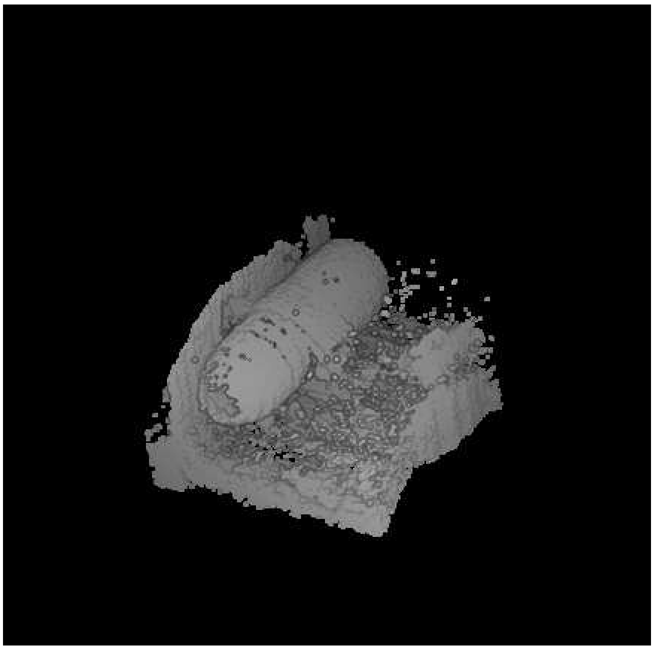


Figure 5.14: Intensity thresholded and volume rendered image of a balloon in a water bath

As mentioned previously, a model-based estimate can be made arbitrarily accurate depending on the number of correspondences specified, and on how awkwardly the balloon is shaped (due to being pinched in the middle by an anchoring strap) compared to the ellipsoid model. For three balloons of approximately 19 cm^3 volume each, volume estimates within 10% were achieved using fewer than 8 landmarks.

Model deformation could not be well controlled beyond that number of landmarks, because of the resolution of the ellipsoid mesh, and the fact that only 75 eigenmodes were used to compute the deformation. In any case, it is unrealistic to expect ultrasound volume estimates of greater accuracy, since errors due to distortion in image formation are probably in the order of 5%; this issue is discussed in Section 5.6.

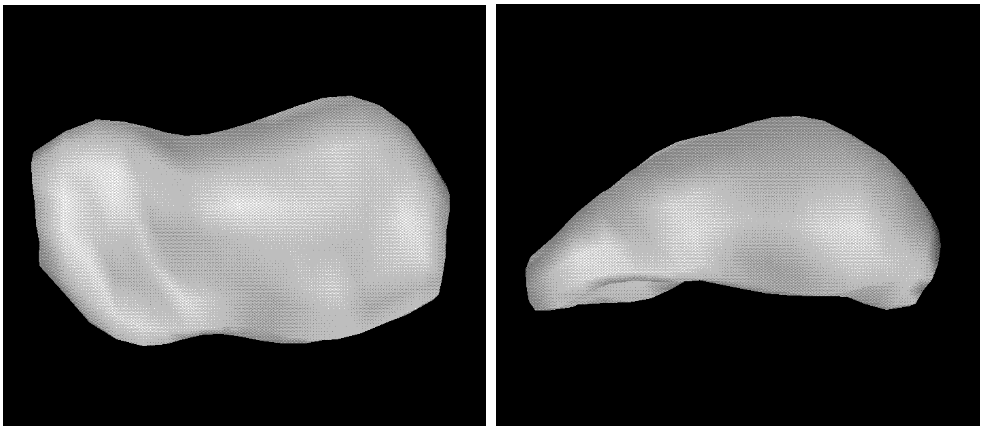
The validation tests in this section suggest that model fitting in 3D freehand ultrasound gives volume estimates of clinically acceptable accuracy, even in the presence of image distortion and position measurement errors.

5.5 Results: segmented livers

Section 5.3 shows that model-based segmentations can be improved using active contour searches. Such searches are suited to echocardiographic images which, like gallbladder images, tend to give good intensity contrast at the organ walls. For example, Neveu et al. [104] use a deformable superquadric with non-parallel image frames, and Coppini et al. [34] use a deformable spherical surface; neither superquadric nor spherical models are specifically heart shape models, however.

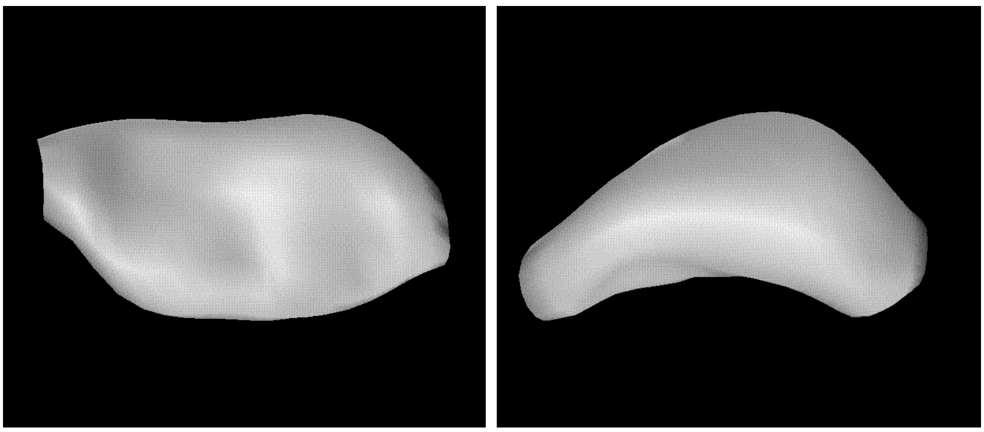
A specific organ shape model allows more accurate initial segmentation using `stradview`. It also allows a search space to be more accurately defined around the initial segmentation, thus improving active contour searches.

To demonstrate this, a liver model was constructed from MR images (recall Section 5.4), and initial segmentations performed on 3D ultrasound images using `stradview`. Figures 5.15–5.18 show four results of active contour refinements to these initial segmentations. These liver models are used later in Chapter 8 for experiments in non-rigid registration.



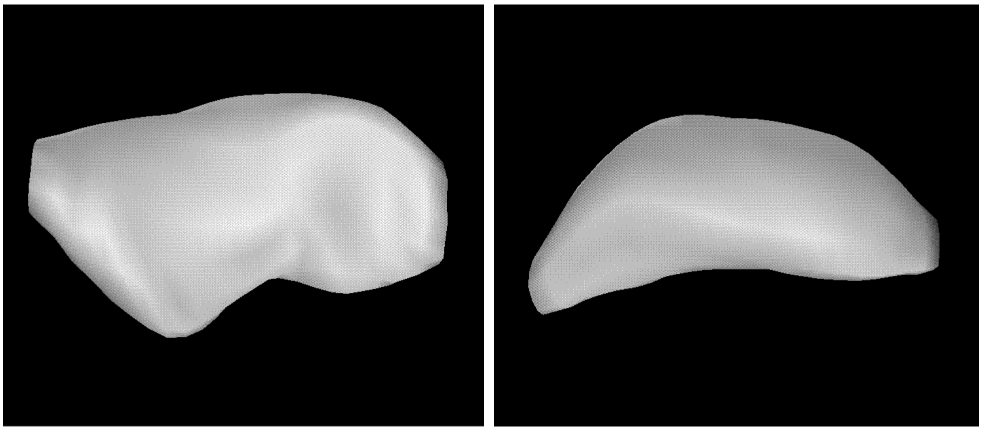
(a) First view (b) Second view

Figure 5.15: Two views of the Liver-A model



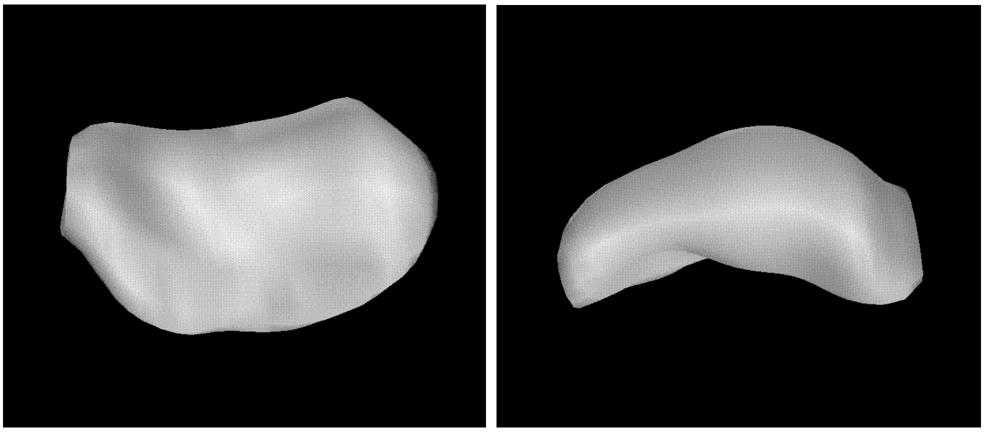
(a) First view (b) Second view

Figure 5.16: Two views of the Liver-B model



(a) First view (b) Second view

Figure 5.17: Two views of the Liver-C model



(a) First view (b) Second view

Figure 5.18: Two views of the Liver-D model

5.6 Discussion

User interface

The initial intention for `stradview` was a proof-of-concept interface for visualising and experimenting with shape models and modes of shape variation. Although volume estimation using this interface has been validated to clinically acceptable accuracy, its ease-of-use can be further improved.

One weakness is its two-stage visualisation process: the sonographer visualises 3D structure during freehand scanning, but is then required to do it again when utilising `stradview`. Visualisation in ultrasonography depends very much on *active perception* (Blake & Yuille [11]) of organ structure, i.e. image information is continually integrated with feedback from the probe's position and orientation, and areas with ambiguous information are actively reinforced with additional images.

Unfortunately, the interactivity of clinical sonography is restricted to a less intuitive windowing interface when using `stradview`. The ideal situation would be for the workstation to take a 'live' feed of video images from the scanner[8], and display this on a touch-sensitive screen in front of the sonographer.

The `geomwish` application would then allow touch-manipulation of the organ model on this screen and, most importantly, touch-selection of landmarks between the model and the 'live' ultrasound image. This would provide a real-time volume estimation facility, which also quantifies the segmented organ shape for biometric analysis and archiving purposes.

Volume estimation

Fitting 3D organ shape models is undoubtedly more accurate than conventional techniques of volume estimation in ultrasound imaging. Typically, the sonographer attempts to find the central plane of a thyroid lobe (say), then estimates volume from the long axis by assuming an ellipsoidal lobe. In fact, Rahmouni et al. [120] find that prostate volume estimation by ellipsoid fitting in trans-rectal ultrasonography, gives significantly worse results than contour segmentation in MRI, when compared to the wet weight of the excised specimen.

Fitting a shape model is similar in many ways to fitting 2D splines; as for a spline, accuracy in shape reproduction is determined by the control points. The number and accuracy of control points required depends on the initial shape of the model, and on whether the required deformation is expected by the model's 'database' of typical shape variation.

The most significant source of error in these control points comes from distortions in the ultrasound image. Ultrasound scanners reconstruct images using an assumed average speed of sound (in human tissue) of 1540 ms^{-1}, whereas in reality the ultrasound beam passes through multiple tissues, each with different propagation speeds. The reconstructed image will therefore be a distorted representation of true tissue geometry.

Figure 5.19 depicts a standard example of distortion, in which the ultrasound image reconstruction of a (spherical) ball in a water bath gives an ellipsoidal cross-section.

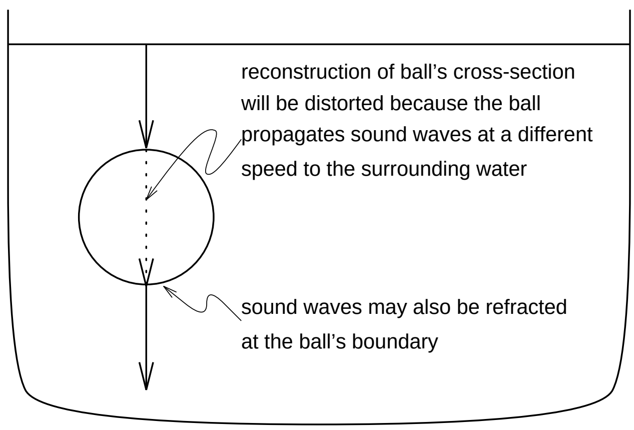


Figure 5.19: Sphere in a water bath gives an ellipsoidal cross-section in ultrasound imaging, for differing propagation speeds of water and ball material

[8] This is technically feasible with the use of a radio frequency (RF) output board provided for the Stradivarius project. The board allows full access to the ultrasound image stream during scanning.

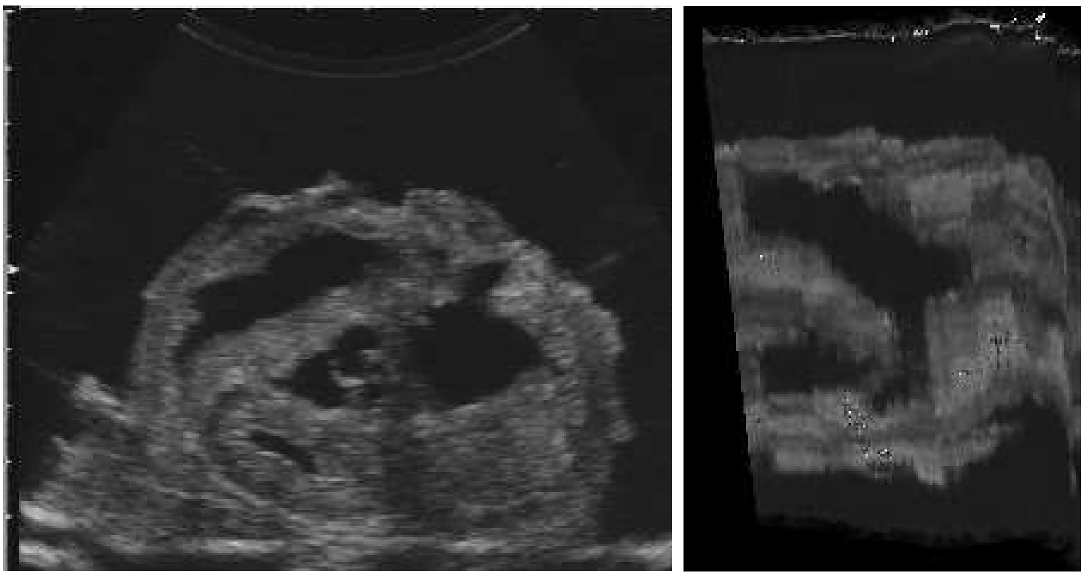
The speed of sound in water also varies with temperature, from 1535 ms^{-1} at 40°C to 1450 ms^{-1} at 10°C (Cho et al. (ch.15)[26]), although this is only significant when performing validation tests in a water bath, using an ultrasound scanner meant for human tissue. The difference between the scanner's assumed speed, and the speed in water at 15°C, can account for the 5–10% volume error in the balloon validation test in Section 5.4.

Volume estimates may also be distorted by a number of other errors inherent in ultrasound image reconstruction (Rohling & Gee[122]), such as non-uniform beams from transducers, frequency and amplitude shifts, multi-path reflections etc. In particular, 'smearing' in ultrasound images due to beam width, finite pulse length, and multi-path phenomena, causes uncertainties when placing landmarks on tissue boundaries. There are also a number of unquantified factors, regarding systematic and random proprioceptive errors in the Stradivarius 3D ultrasound equipment (Gosling et al.[56]).

Volume rendering of 3D ultrasound data

Given that 2D and 3D segmentation of ultrasound images faces many obstacles, the most popular approach to clinical utilisation of 3D ultrasound scans is to render the set of 2D planar images as a single 3D volume image.

Belohlavel et al. [9] suggest that, for 3D trans-oesophageal echocardiography using a fan or cone-shaped acquisition geometry, 3D visualisation of filtered and segmented colour Doppler flow images improves comprehension of anatomic relationships. However, Fuchs et al. [42] review current techniques for interactive visualisation of 3D medical data, and suggest that they cannot yet replace slice-by-slice presentation in routine clinical practice.



(a) Original view (b) Reconstructed view

Figure 5.20: Reconstruction of orthogonal view from 3D volume rendering of a pig's heart

Figure 5.20(a) shows an image from a 3D ultrasound scan of a pig's heart in a water bath. A 3D volume rendering was performed, and Figure 5.20(b) shows an image reconstructed from an orthogonal cross-section in the 3D volume. The internal structure of the heart's chambers can be clearly seen, but the overall clarity of the 3D volume image is degraded because of the rendering process, which involves resampling, interpolating, and spatial compounding of 2D images from different views.

The important characteristics of a good ultrasound image are: clear detail, obvious tissue texture, visualisation of detail at depth, and low noise (Chervenak et al. (ch.1)[25]). As the quality and resolution of acquisition hardware improves, clinicians may be able to perform diagnoses directly using 3D ultrasound data, given appropriate visualisation tools (Nelson & Elvins [103]).

Figure 5.21 shows an early experiment in assessing femoral anteversion from a rendered view of a manually segmented femur (Gosling et al. [55]). Bone cross-sections can also be automatically segmented using the techniques described in Section 5.3.

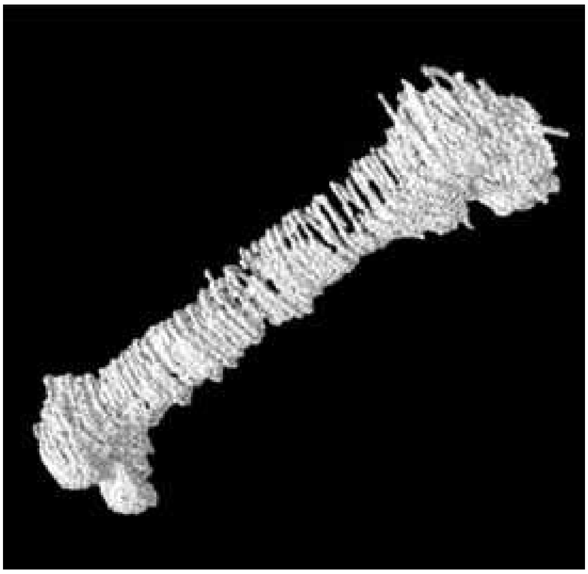


Figure 5.21: 3D rendered view of femur outline

The most important application of 3D rendering and visualisation is expected to be in the field of obstetrics[9] (Sohn [135]), which uniquely offers the possibility of validating measures made *in vivo* during pregnancy, with those after delivery. These measures include: foetal volume, uterine volume, placental volume, and placental position. Ultrasound is also the only imaging modality which is safe, cost-effective, and portable enough for routine use in foetal monitoring.

The tissue masses involved in obstetric imaging are relatively large, and the weakly echogenic amniotic fluid gives a clear delineation to foetal images. This makes volume estimation more accurate, since foetal images are more likely to be accurately segmentated and rendered.

Sakas et al. [125, 126] review the state of the art in volume rendering for visualisation of 3D foetal ultrasound images, and find that morphological and filtering operations can aid surface segmentation for clinical visualisation. Nelson et al. [102] introduce an electronic paintbrush tool which facilitates the segmentation of the foetus from the amniotic background.

The smoothing effects of 2D filtering then 3D rendering will lose vital textural information, but 3D visualisation is potentially useful for viewing large-scale geometric properties. In fact, Steen & Oldstad [140] suggest that direct 3D volume rendering of 2D image features can give useful structural detail not available in conventional ultrasound imaging.

[9] Conventional ultrasound imaging has already revolutionised the practice of obstetric medicine.

5.7 Summary

Volume estimation is one of the most important clinical applications envisaged for 3D freehand ultrasound, and is conventionally performed by fitting simple geometric models to landmarks in 2D ultrasound images. This chapter presents the `stradview` interface, which generalises the conventional approach by employing 3D models of actual organ shape.

Current techniques for discriminating between ultrasound tissues are unreliable, so tissue segmentation is performed by an operator, who selects landmarks in the 3D ultrasound image using `stradview`. A 3D segmentation of the image is then achieved by fitting an organ shape model to the landmarks.

This chapter shows that model-based segmentation is a convenient and accurate method of volume estimation. The initial segmentation can be refined either manually or automatically in `stradview`; future methods for reliable tissue discrimination can also be integrated into the model-based framework for fully-automated segmentation.

Chapter 6

Likelihood of Homology

Chapter 3 shows that elastic eigenmodes are ideal shape features for organ models, because the set of eigenmodes emphasises a model's (global) symmetries at multiple spatial scales. These properties of shape representation are relatively invariant to the choice of model mesh, which is an important consideration when comparing organ models.

Similar organ models have similar sets of eigenmodes, and shape differences between the organ models are more pronounced at smaller spatial scales. Since eigenmodes are ordered in decreasing spatial scale, the corresponding eigenmodes of two organ models are also ordered in decreasing similarity.

Two organ models can therefore be fitted (*registered*) to each other by matching their eigenmodes. Eigenmodes of large spatial scale are used first, because they are the most similar and the most accurately computed.

Such a registration must be biologically accurate; for example, the anatomical landmarks (e.g. fingertips and knuckles) of two hand shapes must be matched to each other. This chapter argues that eigenmode displacement extrema are landmarks suitable for determining registrations, where such 'biologically corresponding' landmarks are known as *homologous* landmarks.

Eigenmodes are global features, and cannot be used to match partially constructed shape models. This chapter suggests a wavelet representation for eigenmodes, which is less sensitive to missing shape information in the organ model.

Chapter organisation

- **Section 6.1** examines biological morphogenesis in a parameter space traversed by growth eigenfunctions. Similar organ shapes lie close together in parameter space, so their eigenmodes (linearised eigenfunctions) are also similar.
- **Section 6.2** discusses methods of normalising eigenmodes, which may have arbitrary magnitude and direction.
- **Section 6.3** proposes that eigenmodes of fundamental spatial wavelength can be approximated by a set of axes. These axes are used to estimate the organ model's orientation.
- **Section 6.4** shows that eigenmodes can be summed in such a way that their displacements are localised in space. This localisation makes *wavelet modes* less sensitive to missing shape information in the rest of the organ model.

6.1 Homologous eigenmodes

Chapter 5.3 notes that a biologically correct registration is required for an organ model and an organ example to share a common shape parameterisation, in which the former can learn about new shape information in the latter. The aim of this section is to show that eigenmodes can be used to perform this registration in a manner justified by our biological growth model in Chapter 2.

Organ models have eigenmodes which are ordered in decreasing spatial scale. Since the shape differences between two organ models are accentuated at small spatial scales, their eigenmodes are also ranked in decreasing order of similarity to each other.

Eigenmodes are particularly useful for registration because they are global features, i.e. a single eigenmode characterises the entire shape of the model. So for each spatial scale, only a single pair of eigenmode features needs to be matched during registration. The matching criteria are difficult to establish, however, since the problem of matching eigenmodes is little different from that of matching the shape models themselves. Conventional criteria for the latter include the matching of curvature extrema, axes of inertia, or ridge curves (Ayache [3]).

We establish appropriate criteria for eigenmode matching, using our growth model from Chapter 2. Biological growth is first assumed to occur in a low-dimensional parameter space, which is traversed by a small number of growth eigenfunctions. Homologous points between two organs are then defined by tracking points on one organ, as parameter space is traversed to the other organ.

The map of all homologous points defines the biologically correct registration of the two organs, where such a map is usually approximated by interpolation between a number of *homologous landmarks*. It is argued that eigenfunction extrema – i.e. points of locally maximal and minimal shape change during traversal – should be employed as homologous landmarks.

Biological morphogenesis is a robust process

Chapter 2.2 notes that closely-interacting growth mechanisms result in robust morphogenesis. Goodwin et al. [53] explain that close interaction significantly reduces the degrees-of-freedom of the growth system, so that it reaches equilibrium in a very stable way; and Gell-Mann [46] and Lewin [81] observe this to be an important *emergent property* of biological systems which consist of simple but closely coupled mechanisms.

Goodwin et al. [53] illustrate this emergence of robust morphogenesis using a thought experiment, which involves populations of red and white cells in a sphere which has two mechanisms of cell growth. Recall from Chapter 3.3 that spherical symmetry requires each growth mechanism to have eigenfunctions which are also spherically symmetric.

Mechanism-A encourages red cells to the surface, and white cells to the interior; and the first eigenfunction of mechanism-B encourages red cells to the north, and white cells to the south. If the cell population remains constant, then the interaction of the combined mechanisms results in a circular flow of white cells from the interior southwards.

If the sphere grows large enough, the second eigenfunction of mechanism-B becomes active. This second eigenfunction has two possible symmetries: red poles and a white interior, or white poles and a red interior. There is already an interior core of white cells present, however, which biases the second eigenfunction towards the former symmetry.

This is the *symmetry breaking* behaviour essential for biological pattern formation (recall Chapter 2.1). The overall growth system on the sphere therefore only gives rise to a restricted set of stable growth patterns.

Morphogenesis in parameter space

The progress of robust morphogenesis can be visualised in a low-dimensional parameter space[1], where the number of parameters reflects the low number of degrees-of-freedom in robust morphogenesis. There is a large *basin of attraction* in this space, so a wide range of growth trajectories will settle in the vicinity of the basin. Figure 6.1 depicts this basin of attraction as the domain of stable adult forms[2].

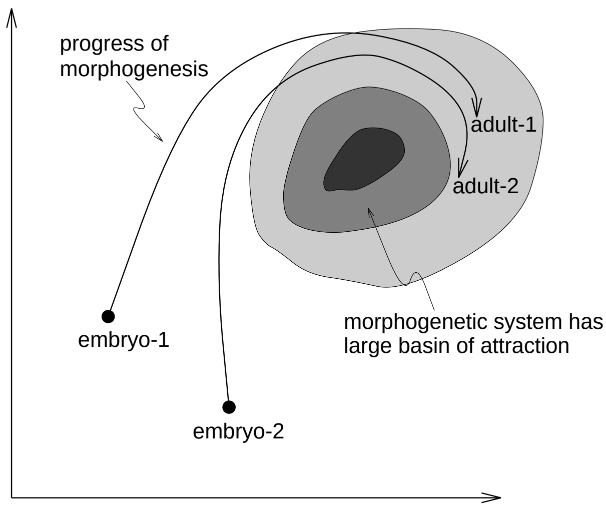


Figure 6.1: Progress of morphogenesis in parameter space

An adult organ is an instantiation in this space of possible adult shapes, where the range of normal organ shapes is caused by noise during morphogenesis[3]. Such noise is manifested in the initial embryonic conditions, and in genetic shape and control parameters. Shape models of adult organs also suffer from noise in the shape modelling process, and from ultrasound image distortions.

Biological homology

Coordinates in parameter space are transformed to physical space under the action of the growth system; Figure 6.2 illustrates the growth system acting as a 'transfer function' from adult-1 and adult-2, to body-1 and body-2 respectively.

When moving a cursor between adult-1 to adult-2 in parameter space, body-1 transforms ('morphs') into body-2 in physical space. Any markers placed on body-1 will shift to new positions on body-2, and such corresponding positions are defined to be *biologically homologous* under the given growth system.

Eigenfunctions (taken about some origin) of the growth system form local axes in parameter space. Any transformation path from adult-1 to adult-2 in Figure 6.2, can therefore be described by the same (linear) sum of independent (non-linear) eigenfunctions[4]. This implies that a point on body-1 always transforms to the same homologous point on body-2 and vice-versa, regardless of transformation path.

[1]Such parameter spaces are generally useful for visualising non-linear system dynamics (Drazin [39]).

[2]Chapter 7.3 develops a statistical control model for adult growth in organs.

[3]Cootes et al.'s [33] point distribution model in Chapter 4.3 basically estimates the axes of parameter space centred on some mean shape, using a training set which is distributed about the basin of attraction.

[4]In physical space, one body transforms to another by superposition of eigenfunction displacement fields.

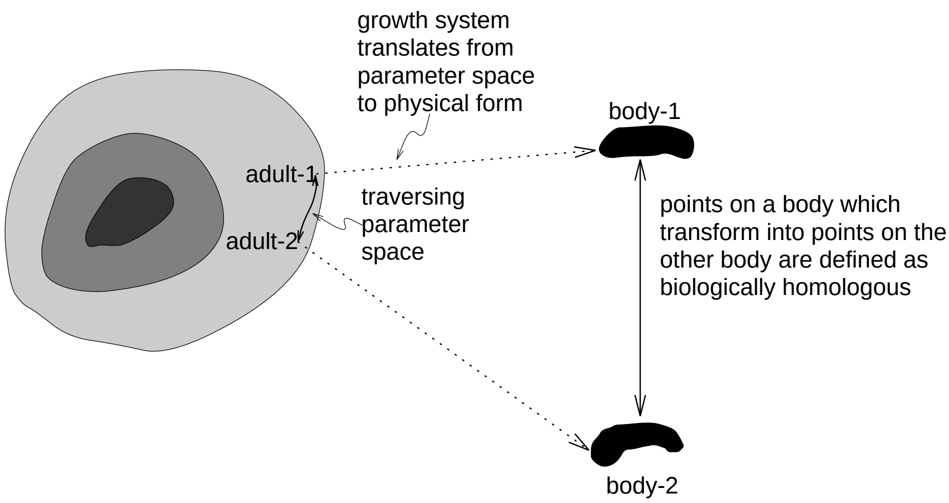


Figure 6.2: Biological homology is defined by traversing parameter space ('morphing')

Eigenmodes indicate homologies

Eigenmodes are approximations to the eigenfunctions of shape change, linearised about a particular instantiation of shape. Figure 6.3 shows two similar (i.e. proximal in parameter space) organ shapes, which lie within the linearisation distance[5] of some of each other's eigenmodes. A pair of eigenfunction axes are also shown for each organ.

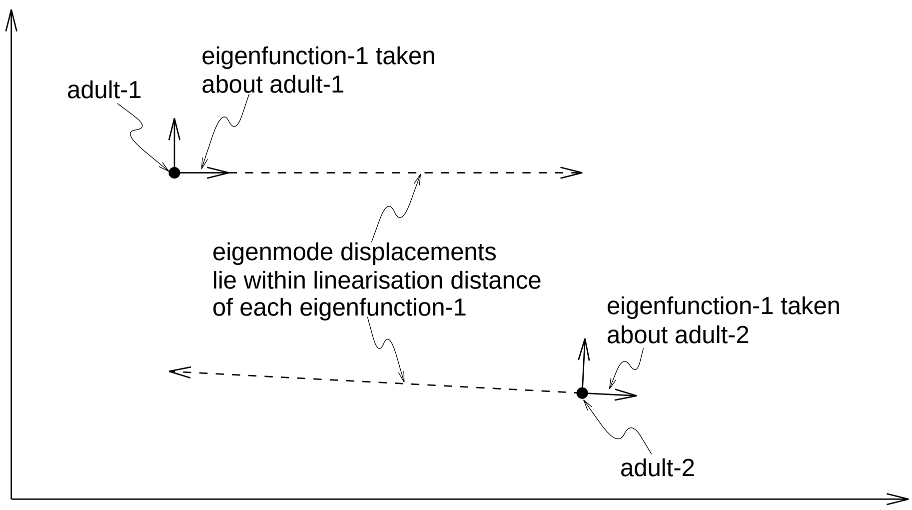


Figure 6.3: Component of shape change along eigenfunction-1 of each adult

Eigenfunction-1 is drawn nearly parallel for adult-1 and adult-2, to illustrate that their eigenfunctions are very similar. Eigenmodes with large linearisation distance (i.e. those of large spatial scale) are more likely to be similar than eigenmodes with smaller linearisation distance, so the eigenmodes of the two adults are ranked in decreasing order of similarity, as mentioned at the beginning of this section.

[5]Eigenmodes of larger spatial scale have larger linearisation distances. Linearisation limits occur (in physical space) where the linearly amplified eigenmode no longer approximates the non-linear eigenfunction well.

Eigenmodes are also harmonic fields with nodes and anti-nodes of minimal and maximal displacement. Given that "landmarks are the points at which one's explanations of biological processes are grounded" (Bookstein (ch.3.3)[16]), these eigenmode extrema are proposed as ideal biological landmarks. Bookstein defines three principal types of biological landmarks:

(i) discrete juxtapositions of tissues;

(ii) maxima of curvature or other local morphogenetic processes;

(iii) extremal points.

Eigenmode displacement extrema fall into (ii) and (iii) in Bookstein's classification. In a hand, for example, the important landmarks are the finger-tips and inter-digital gaps; the elastic eigenmodes of each hand also have local displacement extrema at these locations[6].

Morever, each hand has eigenmodes at different spatial scales, which give rise to displacement extrema at multiple spatial wavelengths; therefore, eigenmode extrema form biological landmarks at multiple spatial scales. The more similar the two hands are, the more eigenmodes fall within their linearisation distance, and the higher the spatial resolution at which homologies can be defined.

6.2 Eigenmode normalisation

We have shown that organ models can be registered to each other using their eigenmodes. However, eigenmode vectors must first be normalised in magnitude and direction, since they are computed with arbitrary scaling from the generalised eigenproblem in equation (3.8).

The organ models being compared may have finite element meshes of different density and distribution, so the conventional procedure of scaling the eigenmode vector's 2-norm is inapplicable. Normalising the mean displacement component at surface sample points is also inappropriate, because eigenmode fields have near-zero net displacement (see Chapter 3.3 for a discussion on eigenmode symmetry and net moments).

One possible procedure is to scale the largest single component at any mesh node, to some fraction (normally 10%) of the model's (centroid) size, where this fraction is a functional definition of the eigenmode's linearisation limit[7]. Eigenmode sign is then chosen by constraining the net volume change associated with the normalised eigenmode to be the larger positive one.

Eigenmodes of large spatial-frequency (preferentially used as shape features) are vector displacement fields with a small number (1–3) of phase changes and local displacement maxima. The largest single component, on which normalisation depends, therefore lies in a smoothly varying displacement maximum of large spatial span. This means that the proposed normalisation is fairly reliable, even with noise in spatial sampling and eigenmode computation.

Section 6.3 shows that normalised eigenmodes are used to compute model pose; the procedure described here avoids the circular problem of having to normalise eigenmodes by first assuming model pose.

[6] A more complete and accurate growth model could potentially give eigenmode extrema at the knuckles, i.e. landmarks of type (i).

[7] For a simple pendulum, the 10% eigenmode linearisation limit is equivalent to an angular peturbation of $\theta = 6^\circ$, which is safely within the bounds of the $\sin\theta = \theta$ linearisation approximation.

Volume change in non-rigid eigenmodes

The sign of volume change is a macroscopic property of the eigenmode displacement field, which has been found to be robust enough for use in direction normalisation. However, it is possible, when comparing two noisy organ models, for a sparsely sampled eigenmode to give volume changes of contradictory signs. This is the case, for instance, in higher eigenmodes of the foetal liver models in Figure 4.2, because their sparsely sampled meshes fall below the minimum Nyquist spatial-frequency needed to specify eigenmodes without aliasing effects.

Note that non-rigid eigenmodes always cause a change in model volume; displacement fields which do not do so for any eigenmode amplitude are, by definition, *rigid-body displacements*. Note also that symmetric eigenmodes acting on symmetric models will give the same volume change for positive and negative eigenmode displacement.

An eigenmode's displacement field has a 'shape', which is visualised by varying modal amplitude between its positive and negative linearisation limits[8]. Therefore, an ideal registration algorithm would not require direction normalisation, since it should fit eigenmode 'shapes' (i.e. the range of displacements) to each other, rather than fitting absolute eigenmode displacements. However, it is found that where reliable direction normalisation is achievable (i.e. with well sampled organ models), the quality of fitting is better with absolute displacements, because fitting 'shape' is less discriminating[9].

Normalising higher eigenmodes

A standard constraint used in finite element modelling, is to ensure that the largest single nodal component has a positive x-direction (Hitchens (ch.8)[62]). This is pointless when the initial pose is undetermined, although higher eigenmodes can be direction-normalised after initial pose is estimated using the (normalised) fundamental modal axes (see Section 6.3).

An organ model's higher eigenmodes are defined on the same mesh as its lower eigenmodes, so the former can also be normalised with respect to the latter using their 2-norms. This is a preferable normalisation scheme for higher eigenmodes, since it does not depend on reliably choosing a largest component.

Modal projection

Syn & Prager [143] use *modal projection* to derive an object-centred set of projected coordinates. Figure 6.4 illustrates how eigenmode components at each node are projected onto fundamental eigenmodes[10], to give a coordinate set which is invariant to translation and rotation in the organ model.

At each of the organ model's nodes, components of the three fundamental eigenmodes form local 3D axes. The component of the third fundamental eigenmode is chosen to form a right-handed trio, with respect to the components of the first two fundamental eigenmodes.

Components of higher (than fundamental) eigenmodes are projected onto these axes, to give *projection coordinates* at each node. These components are also chosen to be right-handed with respect to the components of first two fundamental eigenmodes.

[8] Figure 3.6 shows snapshots of ellipsoid eigenmodes at different modal amplitudes.

[9] Take a simple example of two sine waves of fundamental frequency: fitting 'shape' means that a lobe on one can fit to either lobe on the other; fitting absolute displacements matches positive lobe with positive lobe, and negative lobe with negative lobe.

[10] Section 6.3 defines fundamental eigenmodes as having fundamental spatial-wavelength. They are the most stable when compared between organ models, and provide local coordinate axes which change most slowly (at fundamental spatial-frequency) across the body's span.

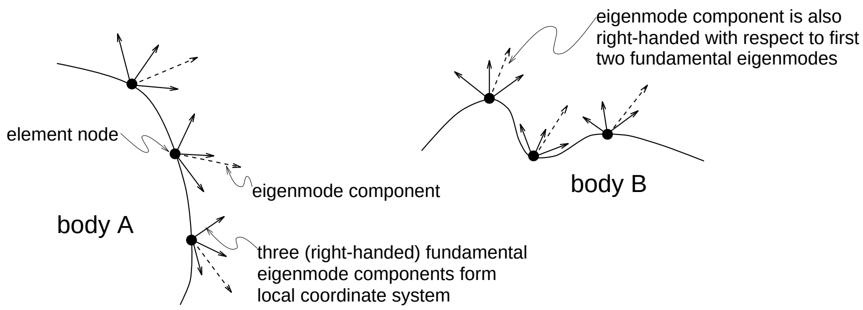


Figure 6.4: Modal projection of eigenmode components

The local axes at each node define object-centred coordinate systems, so projection coordinates can be compared directly between organ models, since they are invariant to rotation in either model.

The major disadvantage of modal projection – and the reason it is not used comprehensively in this dissertation – is that it requires dense sampling of the body mesh to work effectively. It also requires a larger number of eigenmodes, since the three fundamental eigenmodes are reserved for use as coordinate axes. Morever, modal projection leads to an increased search space, since an initial pose estimation is not required.

Without modal projection, eigenmodes require normalisation in both magnitude and direction, so that they can be used to compute organ model registrations. A method of pose estimation is also required, since registration algorithms have large search spaces, and work more efficiently if both models are initialised in approximately the same pose.

6.3 Pose estimation using fundamental eigenmodes

The first three non-rigid eigenmodes of an organ shape model are of fundamental spatial-frequency: they are the *fundamental eigenmodes*. Each fundamental eigenmode has a fundamental symmetry[11], and three fundamental symmetries define the model's intrinsic rigid-body coordinate system.

This section suggests that each fundamental symmetry can be approximated by a *fundamental modal axis*, which is approximately given by the bending axis of the eigenmode field. Three such axes approximately define the model's pose, and rigid-body registration between organ models is achieved by matching these axes (Syn et al. [149]).

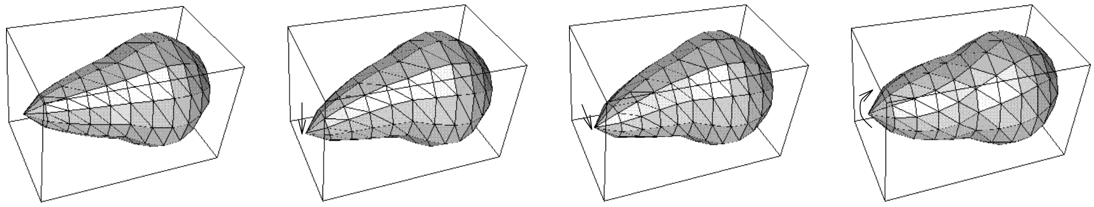
Fundamental eigenmodes of gallbladder models

Figure 6.5 shows two gallbladder models in the same pose. The fundamental eigenmodes give almost identical displacement fields for both models, and there is a single axis about which each eigenmode bends or twists. Figures 6.7–6.9 indicate these bending axes more clearly.

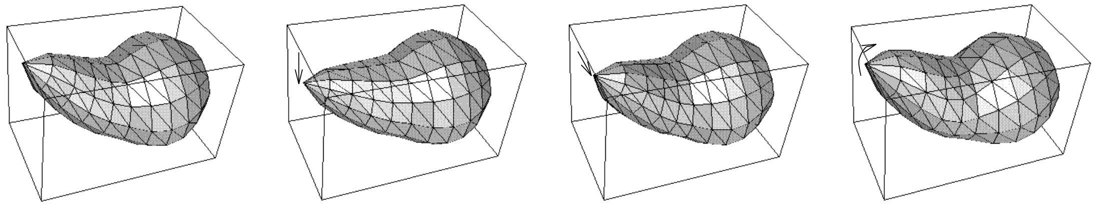
Fundamental eigenmodes of liver models

Figure 6.6 shows the fundamental eigenmodes of the (worst case) Liver-1 and Liver-11 models, originally shown in Figure 4.2. Arrows indicate the gross modal action, from which each bending axis can be estimated. Figures 6.10–6.12 indicate these bending axes more clearly.

[11]The eigenmodes of an organ model reflect its generalised symmetries (see Chapter 3.5).

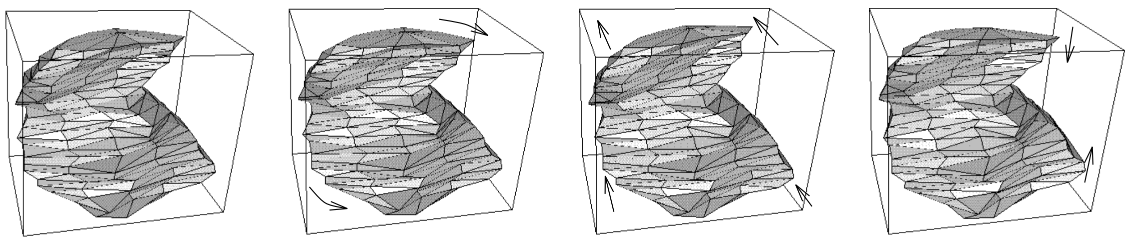


(a) Gallbladder (b) First eigenmode (c) Second eigenmode (d) Third eigenmode

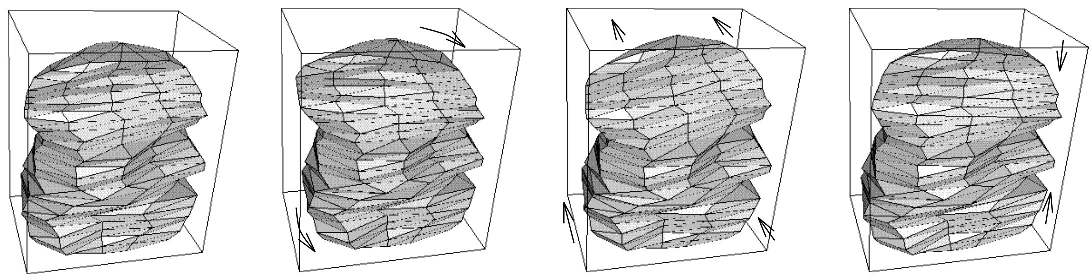


(e) Bent gallbladder (f) First eigenmode (g) Second eigenmode (h) Third eigenmode

Figure 6.5: Fundamental eigenmodes of two volumetric gallbladder models in the same pose



(a) Liver-1 (b) First eigenmode (c) Second eigenmode (d) Third eigenmode



(e) Liver-11 (f) First eigenmode (g) Second eigenmode (h) Third eigenmode

Figure 6.6: Fundamental eigenmodes of Liver-1 and Liver-11 models

Fundamental modal axes

Figures 6.5–6.6 illustrate that each fundamental eigenmode has a bending axis which defines an approximate plane of fundamental symmetry. Unfortunately, a bending axis cannot be computed by taking moments of each eigenmode component, because eigenmode displacement fields have zero net rotational moment (see Section 3.3).

This is somewhat dependent on the organ model's mesh, since a sparsely sampled mesh may not accurately sample the eigenmode field, thus resulting in a non-zero residual moment; a residual moment can also be computing by using only boundary components of the eigenmode field. Nevertheless, residual moments are an unreliable indicator of the eigenmode's bending axis.

Instead, vector moments are taken of each eigenmode component, at its point of action on each element node. The bending axis maximises the projection of the moment vector distribution, so it is the primary eigenvector of the covariance matrix of moment vectors. Bending axes are computed from the first two fundamental eigenmodes, and then mutually orthogonalised to give a right-handed trio, which forms the set of fundamental modal axes for the organ model.

Figures 6.7–6.9 show each of the first three fundamental eigenmodes of a gallbladder model, superposed at a range of amplitudes. Surface normals track local displacement directions and magnitudes over the range of modal amplitudes; the approximate position is also indicated of the bending axis of each fundamental eigenmode.

Figures 6.10–6.12 show the surface normals and bending axes for the Liver-9 model. Again, the colours of the surface normals range between red and green, to track movement of the Liver-9 model's surface with changing modal amplitude.

Fundamental symmetries

The order of fundamental eigenmodes reflects the dominant fundamental symmetries of the gallbladder and Liver-9 models.

In Figures 6.7–6.9 (cf. Figure 6.5), the first two gallbladder eigenmodes have bending axes which are mutually perpendicular, defining the two dominant planes of mirror symmetry along the major axis. The third gallbladder eigenmode has a plane of minor symmetry which is perpendicular to the long axis.

In Figures 6.10–6.12 (cf. Figure 6.6), the first liver eigenmode is perpendicular to the long axis and is the dominant plane of symmetry. The two other eigenmodes reveal planes of fundamental symmetry which are less immediately apparent.

Accuracy of modal axes

As mentioned above, the directional accuracy of a bending axis is dependent to some extent upon the sampling distribution of the finite element nodes, irrespective of the accuracy of the eigenvector computation routine. In the worst case of the liver models in Figure 6.6, for example, there is an error of approximately 30°, which is barely accurate enough for estimating initial rigid-body registration.

When viewing animated eigenmodes on a computer screen, we find that the human eye does a very good job of estimating the orientation of bending axes, because the displacement field is visually integrated continuously over the organ model's surface. A dense eigenmode field can similarly be achieved using finite element interpolation of nodal components, but at the cost of increased computation.

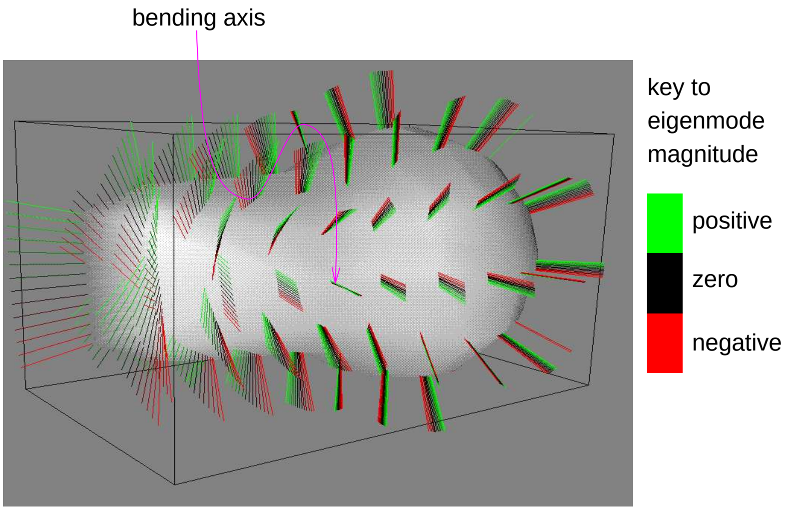


Figure 6.7: Surface normals of gallbladder model at range of amplitudes of first eigenmode

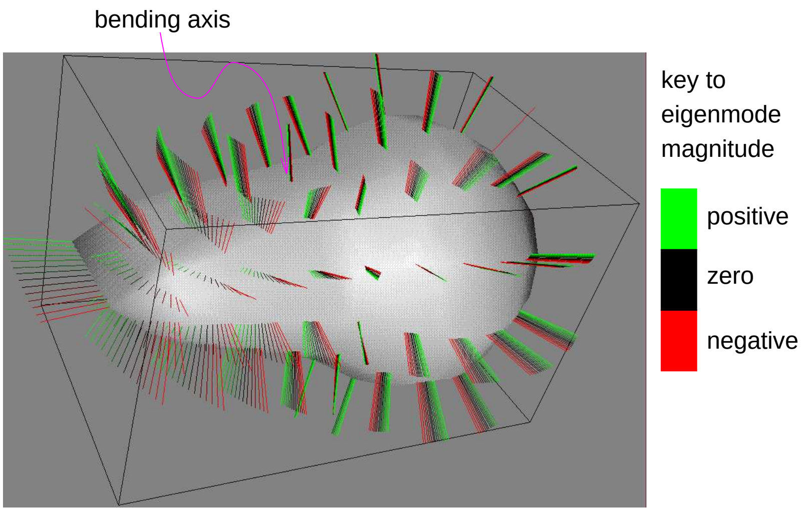


Figure 6.8: Surface normals of gallbladder model at range of amplitudes of second eigenmode

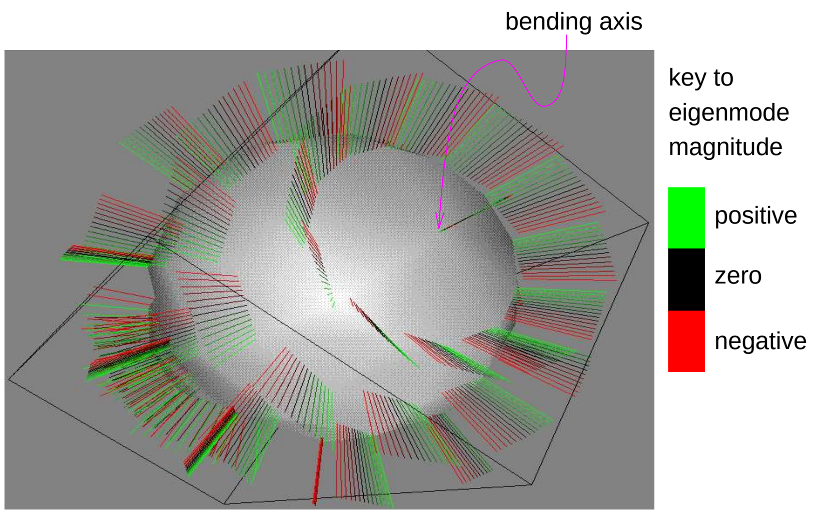


Figure 6.9: Surface normals of gallbladder model at range of amplitudes of third eigenmode

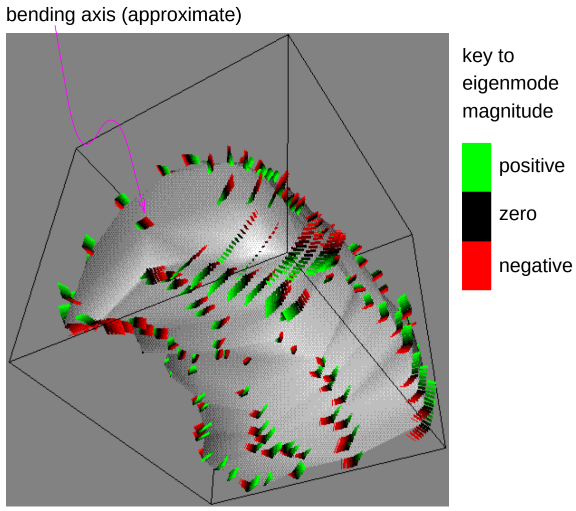


Figure 6.10: Surface normals of Liver-9 model at range of amplitudes of first eigenmode

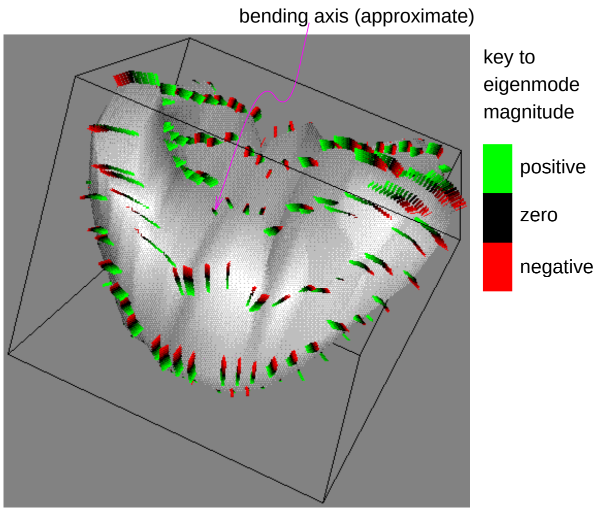


Figure 6.11: Surface normals of Liver-9 model at range of amplitudes of second eigenmode

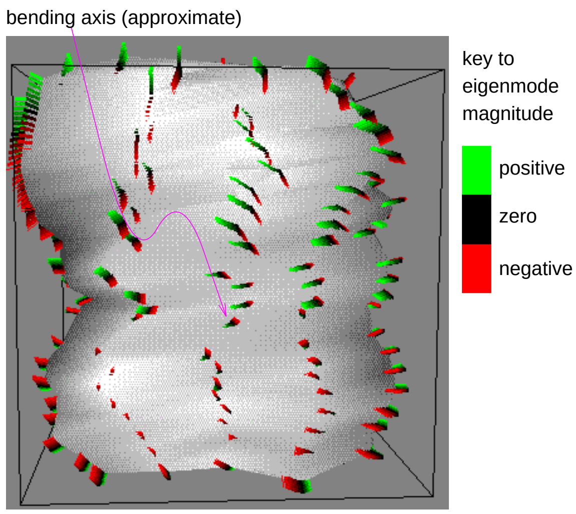


Figure 6.12: Surface normals of Liver-9 model at range of amplitudes of third eigenmode

6.4 Wavelet modes

Eigenmodes are computed with global support, so they are sensitive to shape changes anywhere on the organ model. Although global sensitivity can amplify the effects of irregularities in shape meshing[12], *mass lumping* is an effective remedy to this problem (see Appendix B.5).

A more important concern is that occlusion or signal loss in ultrasound imaging may lead to a highly uncertain segmentation at certain parts of the organ boundary. In such cases, a more suitable shape feature would be localised in sensitivity.

Unfortunately, localised support means that multiple local features are required to characterise organ shape. During registration, for example, additional work needs to be done to ensure that the correct local feature on one organ is matched to the correct local feature on another. The advantage of global sensitivity, in comparison, is that only a single eigenmode from one organ needs to be matched to a single eigenmode from another.

This section develops a *wavelet mode* representation for finite element models, in which wavelet displacements are localised in spatial support, while retaining some desirable shape properties of elastic eigenmodes. Since elastic eigenmodes span space globally but exist only at isolated natural-frequencies, any attempt to reduce the former must have consequences on the latter. The starting point is therefore to quantify the relationships between spatial position, spatial-frequency, and natural-frequency.

Joint localisation in space and frequency

Gabor [44] showed that the *uncertainty principle* in information theory limits a signal's specificity simultaneously in time and frequency. The theoretical lower limit on the time-bandwidth product is achieved by Gaussian-modulated sinusoids, also known as *Gabor wavelets*.

Daugman [38] extends this treatment to 2D signals, and shows that there is a more general definition of joint uncertainty (or joint localisation) in 2D spatial position and 2D spatial-frequency. Again Gabor wavelets give the best joint localisation, and are defined in this case as the product of an elliptical Gaussian with a 2D sinusoid.

These uncertainty relations show that the localised representation we require must suffer increased spread in spatial-frequency, in order to achieve a localisation of signal extent. Morever, the optimal tradeoff in joint localisation for this representation, can be achieved by employing design principles analogous to those used for Gabor wavelets.

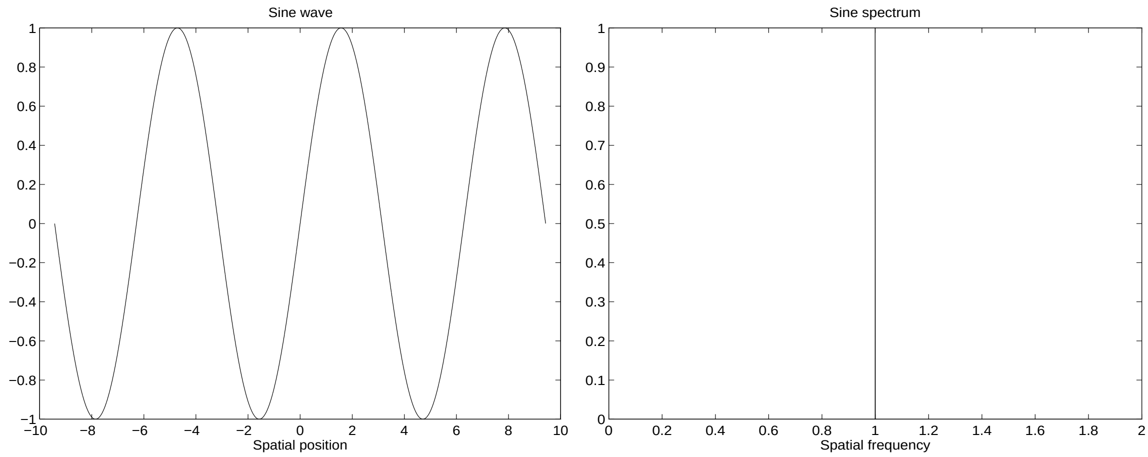
Eigenmodes ('generalised harmonics') are analogous to sinusoidal signals in that both are completely localised at a single spatial-frequency (see next), and both have global extent. The strategy is for an envelope function to localise (by *modulation*) the harmonic signal in space, while increasing uncertainty (by *convolution*) in spectral locality[13]. An envelope function is desired which can achieve the optimal tradeoff between the former and the latter.

Gabor wavelets employ a Gaussian envelope because it decays smoothly and rapidly away from the central lobe, while retaining the same form in both spatial and spectral domains. Figure 6.13 shows how this modulation localises signal extent by dampening the sinusoid away from the Gaussian's main lobe while, at the same time, smearing the spectral peak.

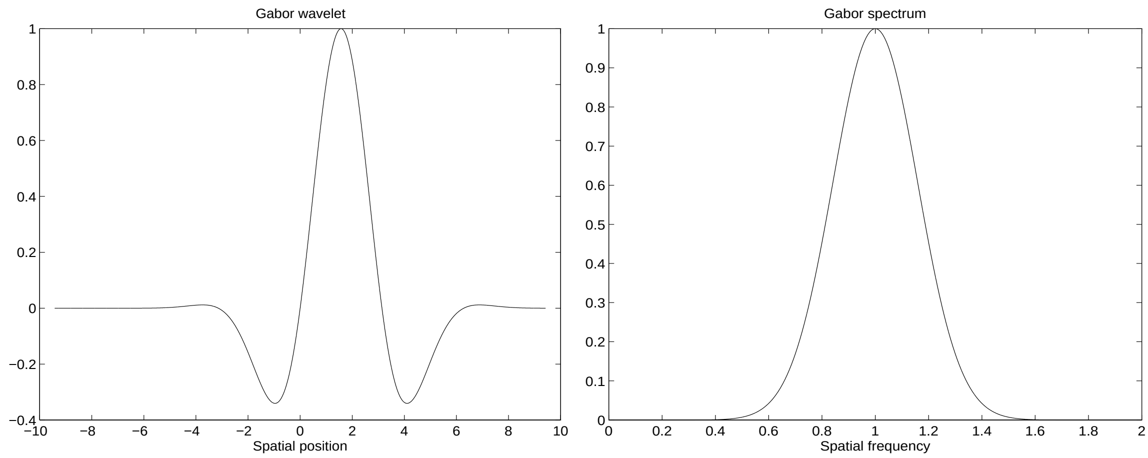
This modulation can be thought of as the design procedure for Gabor wavelets. In generalising this procedure to eigenmodes, however, it is unclear whether spatial-frequency or natural-frequency should be used in the uncertainty tradeoff. The spectral localisation of eigenmodes is first examined, so as to clarify this matter.

[12]See Figures 3.7– 3.12 for examples of noisy meshing.

[13]Multiplication in the spatial domain is equivalent to convolution in the Fourier domain.



(a) Sinusoidal signal has global extent (b) Sinusoidal spectrum at a single peak



(c) Gaussian modulation localises spatial support (d) Gaussian modulation causes spectral smearing

Figure 6.13: Gabor wavelets achieve optimal joint localisation in position and frequency

Spectral localisation of eigenmodes

The wave equation (2.3) for uniform elastic bodies

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = -D^{\mathrm{el}}\nabla(\nabla\cdot\mathbf{u}) \tag{6.1}$$

has the variable-separable solutions (see equation (2.15))

$$\mathbf{u} = \mathbf{f}_i(\mathbf{x})e^{j\omega_i t} \tag{6.2}$$

where $\mathbf{f}_i^{\mathrm{el}}$ (superscript omitted for clarity) is the ith elastic eigenfunction, and ω_i is its natural-frequency. Note that $j = \sqrt{-1}$ here.

We wish to examine the eigenfunctions in the spatial-frequency domain $\boldsymbol{\pi}$, so applying the Fourier transform to equation (6.1) gives (Kreysig (ch.11.14)[77])

$$\omega_i^2\hat{\mathbf{f}}_i = -\frac{k^{\mathrm{el}}}{D^{\mathrm{el}}}\boldsymbol{\pi}_i(\boldsymbol{\pi}_i\cdot\hat{\mathbf{f}}_i) \tag{6.3}$$

where the substitution $\nabla(\nabla\cdot\mathbf{f}_i) = -\frac{k^{\mathrm{el}}}{D^{\mathrm{el}}}\mathbf{f}_i$ is made from equation (2.15).

This shows that the Fourier spectrum of an eigenfunction $\hat{\mathbf{f}}_i(\boldsymbol{\pi})$, is parallel to $\boldsymbol{\pi}_i$

$$\omega_i^2\hat{\mathbf{f}}_i = -\frac{k^{\mathrm{el}}}{D^{\mathrm{el}}}\|\boldsymbol{\pi}_i\|\|\hat{\mathbf{f}}_i\|\boldsymbol{\pi}_i \tag{6.4}$$

which also means that the spatial-frequency vector $\boldsymbol{\pi}_i$ has a constant amplitude proportional to natural frequency ω_i. In other words, each elastic eigenfunction $\mathbf{f}_i$ is localised at a spatial-frequency of constant three-dimensional magnitude, so we expect its Fourier spectrum to show a spherical shell centred at the origin.

Fourier spectrum of eigenmodes

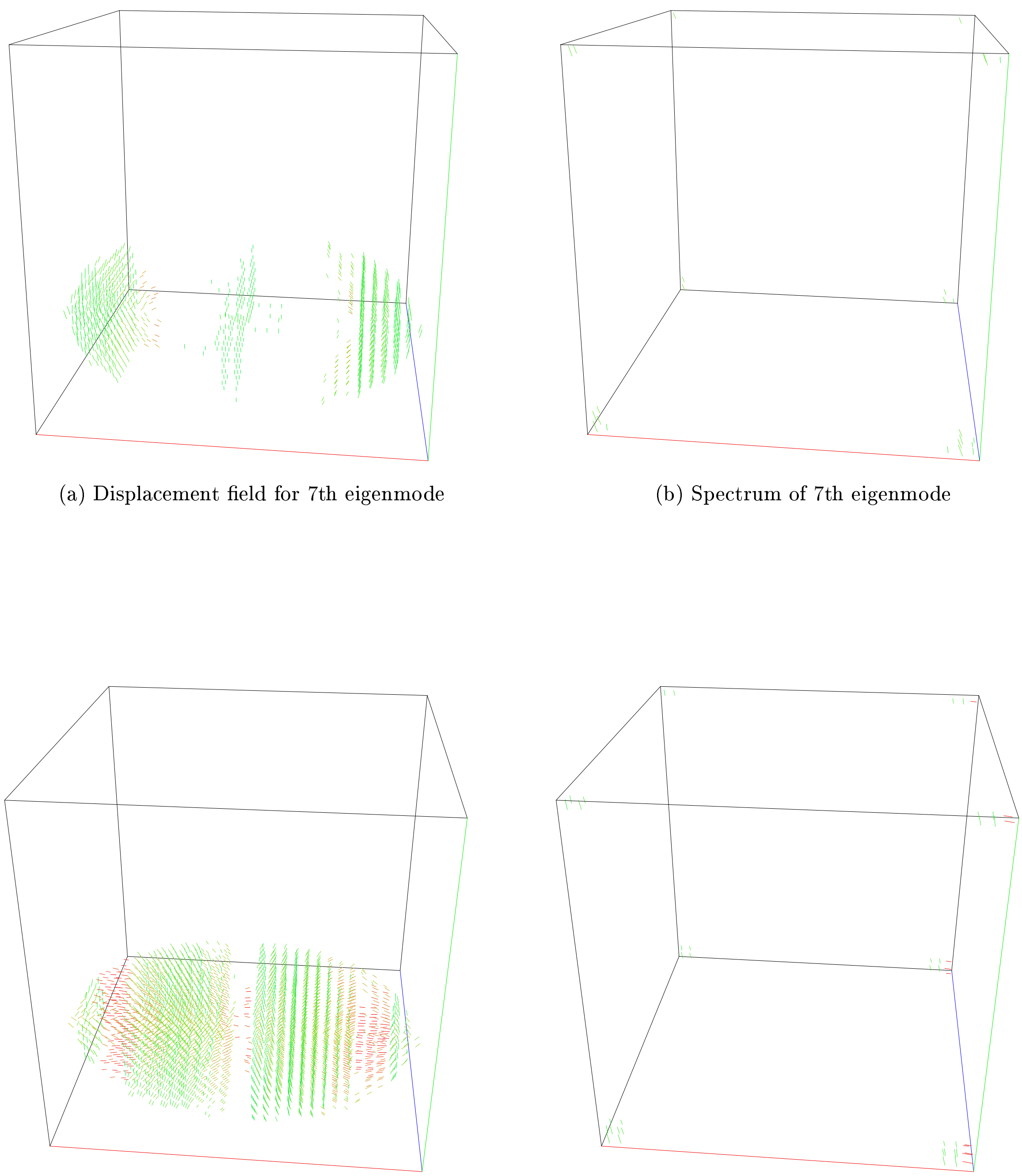
Let us now examine the Fourier spectra of the 7th and 12th eigenmodes of a volumetric ellipsoid (first shown in Figure 3.7).

Figure 6.14(a) shows the eigenmode's displacement field, which is resampled onto a regular grid using FEM interpolation functions. The (x,y,z)-axes are shown in (red,green,blue) respectively to highlight the orientation of the displacement field. In this case, the displacement components are predominantly green, since they act primarily along the y-axis.

Figure 6.14(b) shows the discrete Fourier transform (DFT) of Figure 6.14(a). Equation (6.4) implies that the 7th eigenmode's spectrum should be localised at a constant radial magnitude of fundamental frequency, with frequency components acting radially. This localisation is obscured by the 'windowing' effect[14] , however, even on the relatively fine 64×64×64 grid used in Figure 6.14.

Figure 6.14(c) shows the 12th eigenmode, which is of higher (than fundamental) spatial-frequency and has displacement components along both x and y-axes. The eigenmode spectrum in Figure 6.14(d) is localised at a higher (than fundamental) magnitude of spatial-frequency, and also has components acting along both x and y-axes. However, this localisation is again difficult to see due to spectral smearing.

[14] The DFT inevitably spans a grid which is defined beyond the boundary of the displacement field, giving rise to artefacts at the DFT grid's high-frequency corners. This can be interpreted as a 'windowing' effect (Candy (ch.3.4)[21]), in which the eigenmode's displacement field is masked by the ellipsoid's boundary function, which transforms to a 'smearing' convolution of the eigenmode spectrum.



(a) Displacement field for 7th eigenmode

(b) Spectrum of 7th eigenmode

(c) Displacement field for 12th eigenmode

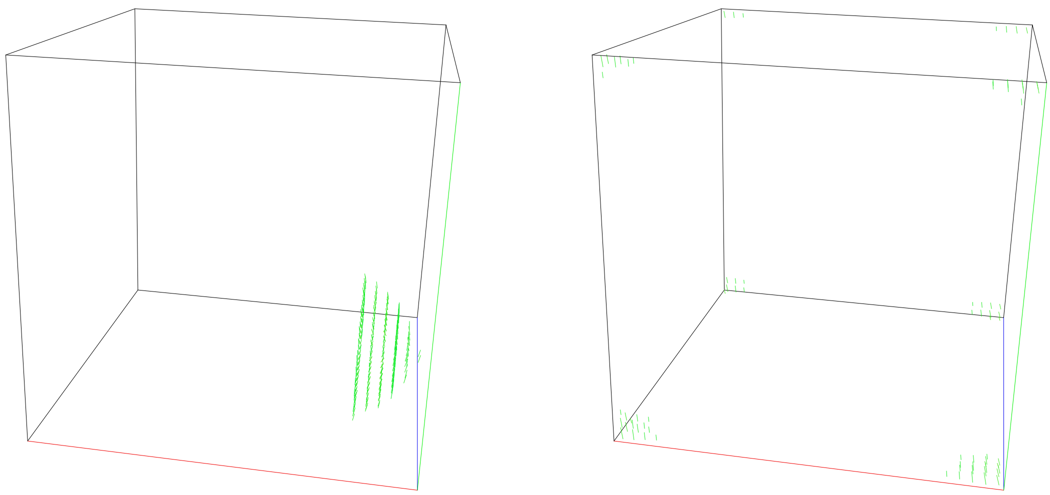
(d) Spectrum of 12th eigenmode

Figure 6.14: Fourier spectrum for eigenmode displacement fields of volumetric ellipsoid

Fourier spectrum of Gaussian-modulated eigenmodes

Figure 6.14 is intended to demonstrate that elastic eigenmodes are localised in spatial-frequency, as predicted by equation (6.4). We now proceed to localise an eigenmode's spatial extent by Gaussian modulation, and to examine the effect on its Fourier spectrum.

Figure 6.15(a) shows the displacement field for the 7th eigenmode, after modulation by a 3D Gaussian envelope centred on one end of the major axis. As expected, the displacement field is localised to the unmasked (right-hand) half of the ellipsoid. Also expected is the loss of spectral localisation observed in Figure 6.15(b) (cf. Figure 6.14(b)) near the fundamental frequency; however, this may again be the result of DFT windowing.



(a) Displacement field (b) Fourier spectrum

Figure 6.15: Fourier spectrum for Gaussian-modulated displacement field of 7th eigenmode

These results are frustrating, because they do not show clearly that Gaussian modulation generalises the Gabor wavelet design procedure to eigenmode harmonics. The resolution of the DFT grid cannot be increased much further than $64 \times 64 \times 64$, and it is not clear if this would result in clearer DFT spectra anyway.

Fortunately, this turns out to be a flawed approach to generalised wavelet design, because the eigenmode field is defined in the model's intrinsic coordinate system, whereas the Gaussian is still defined in 3D Cartesian coordinates.

Figure 6.16 illustrates that this inconsistency in coordinate systems does not correctly localise spatial support. If the Gaussian envelope is centred on the left pole of the discus, it dampens eigenmode displacements at the equator, while only partially doing so for the right pole. However, if the Gaussian is defined in the discus' intrinsic coordinate system, instead of in Cartesian space, the region of the right pole would be considered furthest from the left pole.

(a) Discus model (b) 7th eigenmode

Figure 6.16: Gaussian modulation does not correctly localise spatial support in (b) to left pole

Eigenmode spectrum of Gaussian-modulated eigenmodes

Since the revised aim is now to use a body-centric Gaussian envelope, we should also reconsider the use of the (Cartesian) Fourier transform to examine spectral behaviour. After all, the windowing artefacts observed in Figures 6.14–6.15 are due to resampling from a sparse eigenmode field defined only at element nodes, to a densely arrayed Cartesian grid which spans beyond the model. The consequent loss of sparsity also imposes computational limits on DFT resolution.

Elastic eigenmodes form a unique transform basis and are, by definition, localised in natural-frequency; Fourier analysis in equation (6.4) also shows that each eigenmode is localised in 3D spatial-frequency, at a magnitude directly related to its natural-frequency[15]. Therefore, the elastic eigenmodes form a natural and body-centred transform basis for spectral analysis, and displacement fields can be projected onto eigenmodes to give an *eigenmode spectrum*[16].

Figure 6.17 shows the distribution of natural-frequencies for an ellipsoidal model. The boundary conditions and material properties of the model's wave equation determine the spectrum's discrete and unevenly distributed sampling. The high-frequency components of the spectrum are more sensitive to the finite element model's tessellation[17], as previously discussed in Section 3.3 on eigenmode accuracy.

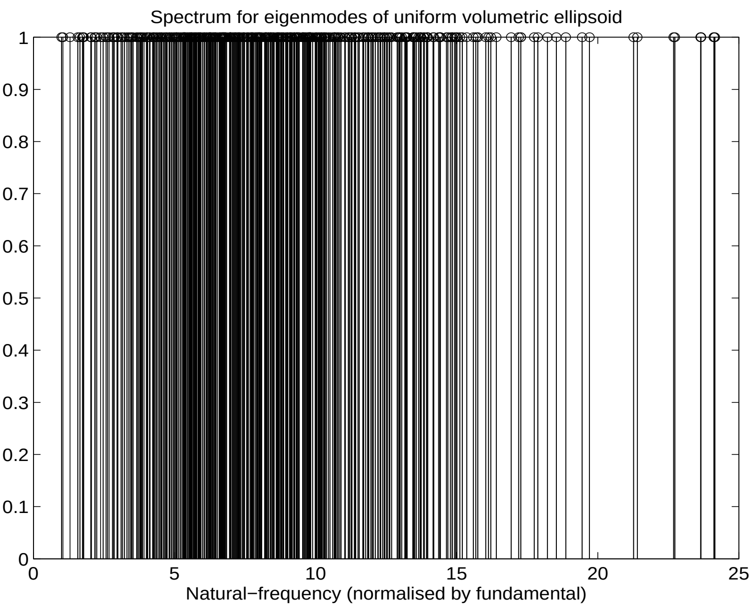
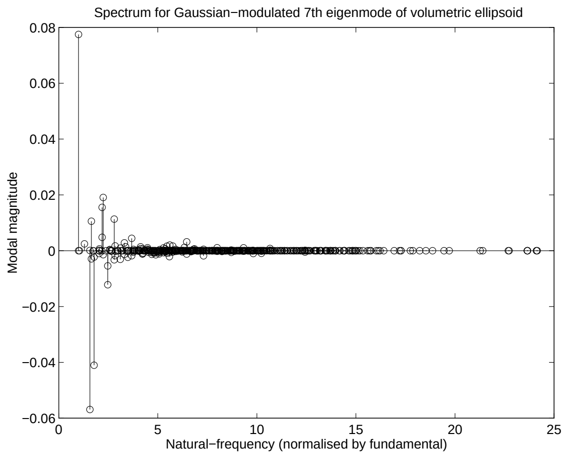


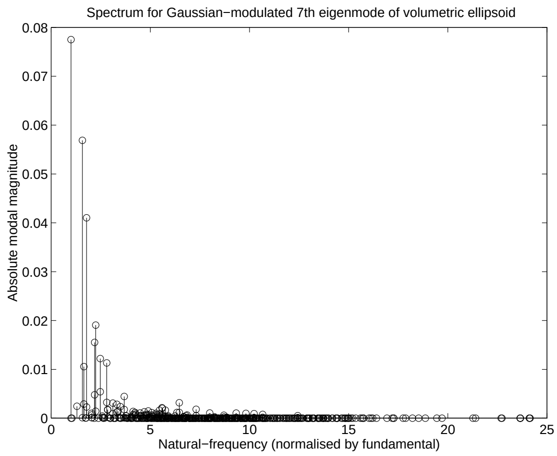
Figure 6.17: Eigenmode spectrum of uniform volumetric ellipsoid

The Gaussian-modulated 7th and 12th eigenmodes are now projected onto their eigenmode spectra, as shown in Figures 6.18–6.19. These projections are instructive, because Figure 6.18(b) shows the spectral envelope to be approximately Gaussian in shape, and centred on the 7th (fundamental) eigenmode. Figure 6.19(b) similarly shows an approximately Gaussian envelope, this time centred on the 12th eigenmode.

[15]Recall that this holds true only if uniform elasticity is assumed.
[16]On a uniform Cartesian grid, the elastic eigenmode bases are identically the Fourier transform bases.
[17]Mass lumping is a partial remedy to this, as mentioned at the very beginning at this section.

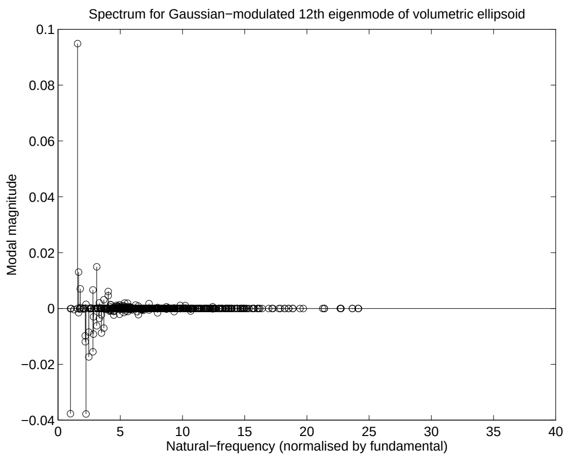


(a) Signed magnitudes

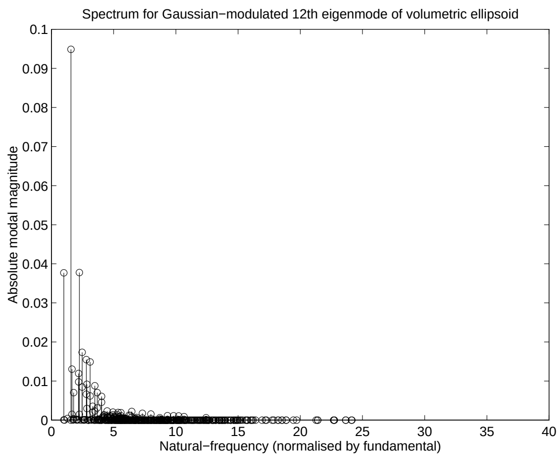


(b) Absolute magnitudes

Figure 6.18: Eigenmode spectrum of Gaussian-modulated 7th eigenmode



(a) Signed magnitudes



(b) Absolute magnitudes

Figure 6.19: Eigenmode spectrum of Gaussian-modulated 12th eigenmode

Gabor wavelet modes

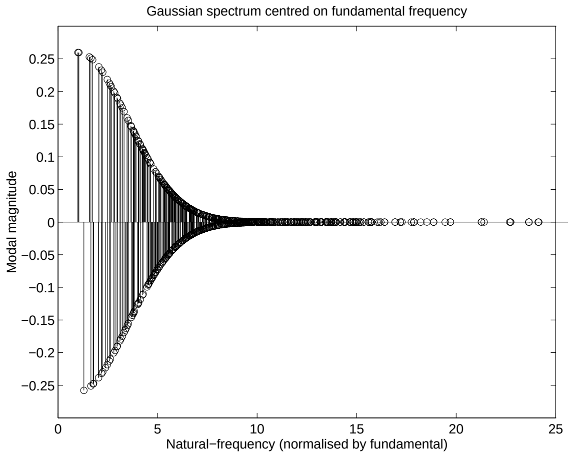
Figure 6.18(b) and Figure 6.19(b) suggest that *wavelet modes* should be designed directly by Gaussian modulation in the eigenmode spectrum. This is the generalisation of the Gabor wavelet design procedure we are seeking, since the Fourier transform used to characterise joint uncertainty in space and frequency for regular Cartesian arrays, is a particular case of the eigenmode transform for arbitrary finite element meshes.

In other words, Cartesian coordinates and frequency in regular wavelets, generalise to model-centred coordinates and natural-frequency in finite element models. This means that wavelet design is dependent on the number of eigenmodes available, and is thus limited by computational resources and the Nyquist sampling resolution of the finite element mesh.

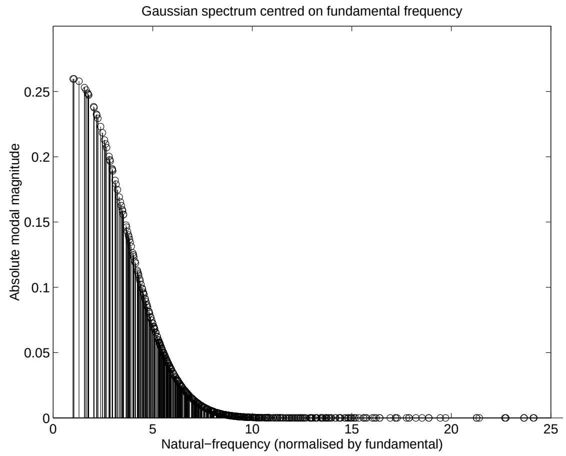
The design procedure for a *Gabor wavelet mode* is therefore to perform Gaussian modulation in the eigenmode spectrum. However, this only determines the *magnitude* of each constituent eigenmode, and not its *sign*, i.e. eigenmodes can interfere constructively or destructively depending on their relative phase, although their relative magnitudes are fixed.

Daugman [38] addresses this in the design of 2D Gabor wavelets, by minimising the "effective spread" of each wavelet's energy distribution. Since we are designing in the eigenmode spectral domain, an overall constraint of minimal effective spread is imposed in the spatial domain, where spatial spread is quantified by the mean-square displacement of the wavelet mode. Fortunately, eigenmodes are mutually orthogonal and linearly separable, so each eigenmode's sign can be individually determined in consecutive order.

Figure 6.20 shows the eigenmode spectrum of a Gabor wavelet mode. Figure 6.21 shows the associated wavelet displacement field, which has a large spatial spread due to the narrow Gaussian spectrum, and large spatial-wavelength because the spectrum is centred on the fundamental natural-frequency.

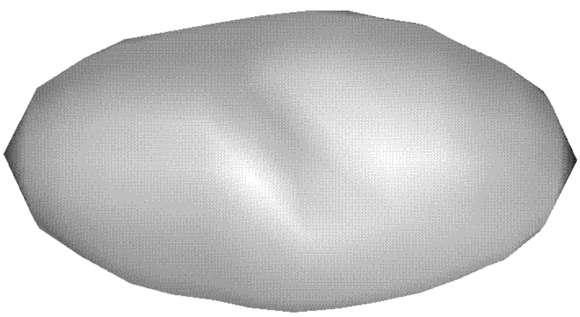


(a) Signed magnitudes

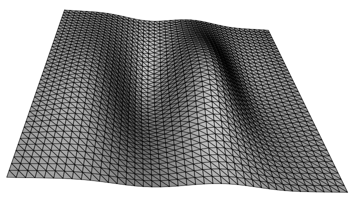


(b) Absolute magnitudes

Figure 6.20: Gabor wavelet mode with a narrow Gaussian as its eigenmode spectrum



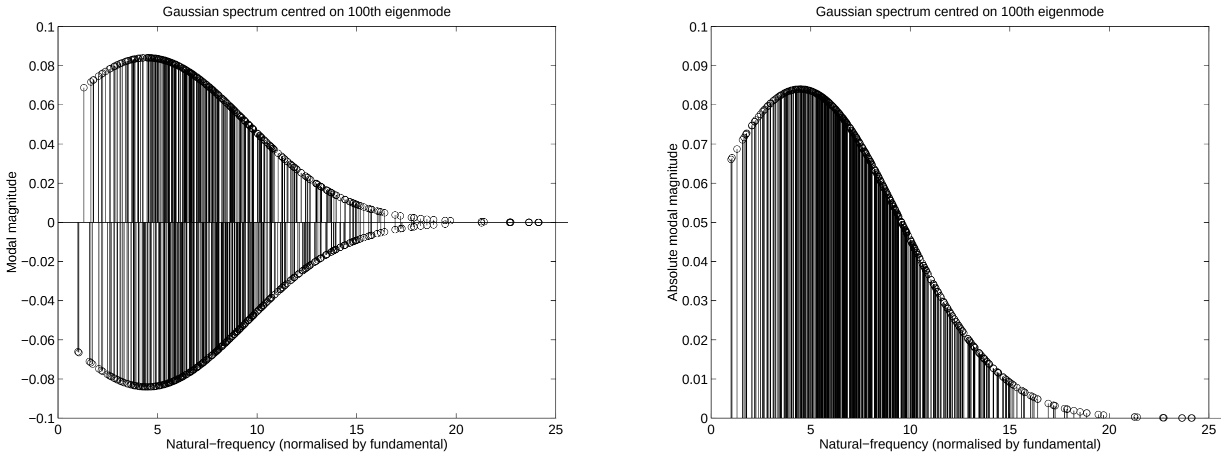
(a) Gabor wavelet mode



(b) A similar 2D Gabor wavelet

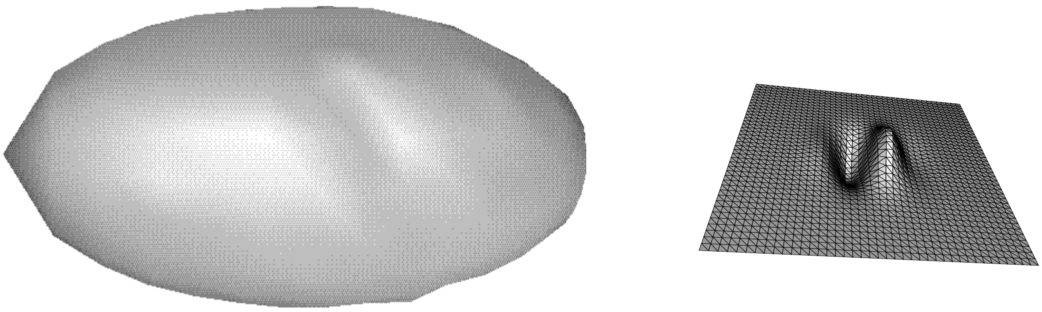
Figure 6.21: Gabor wavelet mode has wider spatial spread

Figure 6.22 shows the eigenmode spectrum of another Gabor wavelet mode. Figure 6.23 shows the associated wavelet displacement field, which has a narrow spatial spread due to the wide Gaussian spectrum, and smaller spatial-wavelength because the Gaussian spectrum is centred on a higher natural-frequency.



(a) Signed magnitudes (b) Absolute magnitudes)

Figure 6.22: Gabor wavelet mode with a wide Gaussian as its eigenmode spectrum



(a) Gabor wavelet mode (b) A similar 2D Gabor wavelet

Figure 6.23: Gabor wavelet mode has higher spatial-frequency and narrower spatial spread

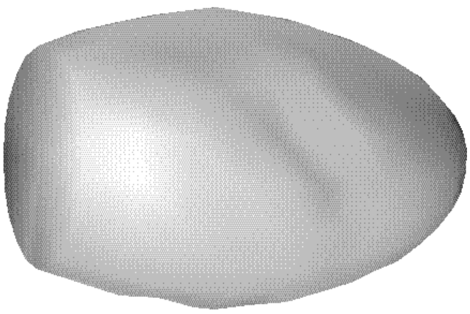


Figure 6.24: Gabor wavelet mode for truncated ellipsoid

Figure 6.24 shows a sharply truncated ellipsoid with an equivalent Gabor wavelet mode to Figure 6.23. It can be seen that both Gabor wavelet modes have similar displacement fields, and this is as expected, since the localised spatial support of the wavelet leaves it relatively unaffected by the truncation of the ellipsoid.

The set of Gabor wavelet modes is vastly overcomplete, since there is a wide choice for both the Gaussian spectrum's position and extent. In practical terms, however, this choice is limited by the number of eigenmodes available. Recall also that the high-frequency eigenmodes are less accurately computed, and require greater computational expense due to more problematic convergence[18].

The width of the Gaussian should be large enough that it is adequately sampled by the eigenmode spectrum[19], but not so large that it relies heavily on high-frequency eigenpairs. There is also a tradeoff with the position of the Gaussian spectrum, since it is undesirable to have either tail prematurely truncated by the lowest and highest natural-frequencies, or to have a displacement field at higher spatial-frequency than the Nyquist sampling limit of the finite element mesh.

6.5 Summary

Chapter 2 shows the elastic eigenfunctions to be the principal patterns of growth. This chapter argues further that the displacement extrema of eigenfunctions are biological landmarks, so that organ models can be registered to each other by matching these landmarks.

Registrations can be computed more efficiently if the organ models are approximately aligned beforehand. This chapter shows that alignment can be performed using each organ model's fundamental eigenmodes, which describe a set of axes defining the model's pose.

Eigenmodes are sensitive to the global shape of the organ model, which may be problematic when the boundary of the organ is uncertain, e.g. from segmenting occluded ultrasound images. This chapter proposes Gabor wavelet modes as a novel solution. These wavelet modes are computed from eigenmodes, so they have desirable properties as shape features; more importantly, wavelet modes are only locally sensitive to organ shape. Unfortunately the method proposed here requires significant computation for Gabor wavelet modes to be accurately represented.

[18]The resolution of the Gabor wavelet modes shown in Figures 6.20–6.24 is fundamentally limited by the sampling density of the underlying (434 node) finite element mesh, which requires 1302 eigenpairs to be computed to tolerable accuracy. Larger meshes were attempted, but the noisiness of the higher eigenmodes gives rise to significant high-frequency artefacts in the wavelet displacement field. See Appendix B for issues involved in solving large eigenproblems.

[19]It remains unclear how the discrete and uneven spacing of eigenmode natural-frequencies affects the wavelet displacement field.

Chapter 7

Coherence of Homology

Chapter 2 employs Oster-Murray mechanisms in a model of biological growth, where the eigenfunctions of these mechanisms describe the modes of organ shape change. Other chapters in this dissertation explore the eigenfunctions' properties in representing organ shape and organ shape variation. Chapter 6, in particular, shows that eigenfunction extrema are biological landmarks which can be matched between organ models to find their registration.

The task remains of finding an appropriate metric for the 'cost' of biological growth. Such a metric is useful during registration, when it can be employed in deciding the plausibility of a suggested landmark fit. It is also useful for interpolating the registration between landmarks in a coherent and biologically justified way, by choosing the pattern of interpolations with least 'cost'.

This chapter develops a suitable metric using the *Gompertz model* of organ growth. This energy metric interprets the cost of Gompertzian growth by its associated mass change, and is derived from a function describing cell proliferation, which can be incorporated into equation (2.7) of our growth model in Chapter 2.5.

Chapter organisation

- **Section 7.1** introduces the Gompertz model of cell proliferation during organ growth.
- **Section 7.2** introduces the Gompertz energy metric, which describes the cost of organ growth. This metric has a simple and intuitively appealing interpretation; it also has linear form when describing small growth changes near adult mass.
- **Section 7.3** justifies the use of Gompertzian growth, and shows that it results from a simple statistical model of biological growth control.
- **Section 7.4** proposes that the Gompertz energy metric can be used to coherently interpolate registrations between landmarks. A simple Bayesian model combines the *likelihood* of landmark similarity, with the *prior* constraint that Gompertz energy is minimised. The resulting *posterior* measure of biological homology can then be used to register organ models at all points on and between landmarks.

7.1 Gompertz growth function

Medawar [89] measured the growth of the embryo chicken heart, which was found to follow the Gompertz growth function

$$y^{\mathrm{gz}} = ae^{-be^{-kt}} \tag{7.1}$$

where $a^{\mathrm{gz}}, b^{\mathrm{gz}}$ and k^{gz} are constants (superscripts omitted for clarity) and t is the time variable. The time derivative of y^{gz} is

$$\begin{aligned} \dot{y}^{\mathrm{gz}} &= akbe^{-kt}e^{-be^{-kt}} \\ &= y^{\mathrm{gz}}kbe^{-kt} \qquad (7.2) \end{aligned}$$

Equation (7.1) and equation (7.2) are used to define the *specific growth rate* R^{gz}

$$\begin{aligned} R^{\mathrm{gz}} &= \frac{\dot{y}^{\mathrm{gz}}}{y^{\mathrm{gz}}} \\ &= kbe^{-kt} \qquad (7.3) \\ \ln R^{\mathrm{gz}} &= \ln kb - kt \qquad (7.4) \end{aligned}$$

It is useful to employ equation (7.1) again

$$be^{-kt} = \ln a - \ln y^{\mathrm{gz}} \tag{7.5}$$

for substitution into equation (7.3)

$$R^{\mathrm{gz}} = k\ln a - k\ln y^{\mathrm{gz}} \tag{7.6}$$

7.2 Gompertz growth energy

Medawar [89] determined *growth energy* W^{gz} for the embryo chicken heart[1], by examining the specific amount of *growth inhibitor* required to cause tissue growth to cease. Growth was inhibited at a series of different ages, and making the assumption that growth energy is directly proportional to the amount of inhibitor applied, Medawar found a linear trend when plotting the logarithm of W^{gz} against time t

$$\ln W^{\mathrm{gz}} = \ln kb - kt \tag{7.7}$$

Medawar also found a linear relationship between W^{gz} and the logarithm of tissue mass m

$$W^{\mathrm{gz}} = k\ln a - k\ln m \tag{7.8}$$

Comparing equation (7.7) with equation (7.4), and also equation (7.8) with equation (7.6), it can be seen that:

(i) Gompertz growth energy W^{gz} is a measure of specific growth rate R^{gz};

(ii) the growth of tissue mass m must conform to the Gompertz function.

[1] Medawar collected growth energy data over a range of 6–18 days, which is equivalent in 'physiological time' to a high proportion of the growth period. Meaningful physical interpretation of equations (7.7)–(7.8) starts to fail near the origin of the time and mass axes.

Growth energy as a function of mass

Equation (7.8) can be rewritten, using $a = m^{\max}$ to represent the adult mass at which the growth rate slows to a halt

$$W^{\text{gz}} = k \ln \frac{m^{\max}}{m} \tag{7.9}$$

Figure 7.1 illustrates the form of Gompertz growth energy in equation (7.9). The growth rate slows to zero when the organ reaches adult mass, and the adult growth rate can be linearly approximated as shown.

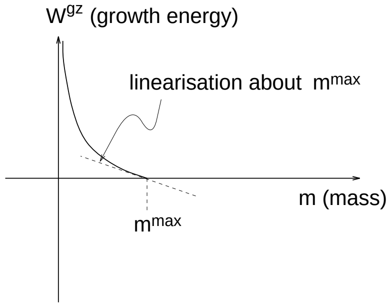


Figure 7.1: Plot of Gompertz growth energy against organ mass

Interpretation of growth energy metric

In Medawar's experiments W^{gz} represents the the growth potential the organ has left to resist the applied growth inhibitor, which acts by stopping mesenchymal cells (fibroblasts) from redistributing themselves after proliferation (cf. Chapter 2.5). Medawar suggests that this prevents further cell proliferation, hence halting tissue differentiation and growth[2].

Now consider the change in W^{gz} due to a small change in tissue mass δm, linearised about adult mass $m^{\max}$ as shown in Figure 7.1

$$\begin{aligned} \left.\frac{dW^{\text{gz}}}{dm}\right|_{m^{\max}} &= -\frac{k}{m^{\max}} \\ \delta W^{\text{gz}} &= -\frac{k}{m^{\max}}\,\delta m \end{aligned} \tag{7.10}$$

Equation (7.10) is a metric quantifying the change in growth energy which results from small changes δm in the tissue mass of fully-formed organs; energetic change is negative since it describes the work done against inhibitory mechanisms. Its linear form[3] is particularly useful, since the energy required for any pattern of (adult) growth can be easily computed from the associated mass change.

7.3 Justification of Gompertzian growth

The Gompertz energy metric developed in Sections 7.1–7.2 derives from Medawar's [89] experiments on organ growth. It is more commonly applied in population and mortality studies (Lestienne [79]); in fact, Gompertz [52] originally formulated equation (7.1) in an actuarial context (see also Causton [22]).

[2]Recall that y^{gz} is employed as the cell proliferation model for equation (2.7) in Chapter 2.5.

[3]Using the non-linear $\delta W^{\text{gz}} = -\frac{k}{m}\delta m$ would incur significant computational cost for numerical integration. In any case, the important thing to note about the metric is that energetic cost increases with δm.

The Gompertz growth function is also used in modelling organ growth. For example, Fujita [43] uses 60 Gompertzian growth segments to model the anatomical growth of mammalian brains; and Luecke & Wosilait [85] model the growth of human foetuses and 16 foetal organs using the Gompertz function.

The logistic, Bertalanffy, and Gompertz functions are widely used in modelling biological growth, since they all approximately sigmoidal in shape. All three have been justified as biologically plausible tumour growth models (Xu [170]; Vaidya & Alexandro [163]), but Xu & Ling [171] and Ling & He [84] find the Gompertz function to be the expectational model for all three, given a uniform distribution of parameters for a generalised model encompassing all three[4].

Statistical model of growth control

To complement these experimental reports of Gompertzian organ growth, we also develop a statistical interpretation of Gompertz energy using a simple model of organ growth control.

Recall that Medawar calculated growth energy W^{gz} by the amount of chemical inhibitor used. Although each chemical molecule represents a constant amount of inhibitory energy, the *probability* of inhibition P is determined by the kinetic energy of the inhibitory molecules, which can be shown to follow a Boltzmann distribution (van Laarhoven & Aarts (ch.2.1)[165])

$$P(W^{\mathrm{ib}}) = \frac{1}{T} e^{-W^{\mathrm{ib}}/T} \tag{7.11}$$

where W^{ib} is the *inhibition energy* (cf. Medawar's growth energy W^{gz}). T is a temperature parameter which reflects the hostility of the inhibitory control system.

The cell proliferation rate y^{gz} is directly determined (up to some constant a) by inhibitory challenges (Shymko & Glass [134]), which meet or exceed the cellular growth energy W^{gz}

$$\begin{aligned} y^{\mathrm{gz}} &= a \int_{W^{\mathrm{gz}}}^{\infty} P(W^{\mathrm{ib}})\, dW^{\mathrm{ib}} \\ &= \frac{a}{T} \int_{W^{\mathrm{gz}}}^{\infty} e^{-W^{\mathrm{ib}}/T}\, dW^{\mathrm{ib}} \\ &= a e^{-W^{\mathrm{gz}}/T} \end{aligned} \tag{7.12}$$

In comparison, the Gompertz growth energy in equation (7.7)

$$\begin{aligned} \ln W^{\mathrm{gz}} &= \ln kb - kt \\ W^{\mathrm{gz}} &= kb e^{-kt} \end{aligned} \tag{7.13}$$

can be substituted into the Gompertz function in equation (7.1) to give

$$\begin{aligned} y^{\mathrm{gz}} &= a e^{-b e^{-kt}} \\ &= a e^{-\frac{b}{k} W^{\mathrm{gz}}} \end{aligned} \tag{7.14}$$

Comparing this to equation (7.12), it can be seen that the Gompertz growth energy metric is justified by our simple statistical model of growth control.

[4] Xu & Ling [171] incorporate the effects of spatial organisation, surface roughness, and competition for nutrition. This results in a self-limiting growth model which can reasonably be used as a simple approximation to other forms of tissue growth. Morever, it is a simple model of spheroidal growth *in vitro* which is more relevant to general organ growth, since there are no complicating mechanisms of nutrition or proliferation as would be the case for tumour growth *in vivo*.

7.4 Local coherence of homology

Sections 7.1–7.3 show that the Gompertz energy metric quantifies the energetic cost of organ growth, which is important as a means of discriminating between possible landmark fits during organ model registration. The form of Gompertz energy in equation (7.10) is particularly useful, since it shows the energetic cost of a growth pattern to be directly linked to the associated mass change.

Recall that Chapter 6 shows eigenmode extrema to be biological landmarks which establish homologous locations between two organ models; this section employs a simple Bayesian framework to integrate landmark fitting with the Gompertz energy metric. In this framework, landmark fits give the *likelihood* of homology, and the energy cost of the landmark fit is a *prior* constraint on the possible range of fits (Syn & Prager [145, 146]).

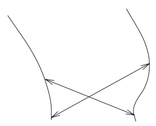
Likelihood energy

The likelihood of two points, one on each organ model, being similar is established by testing the (normalised) eigenmode components at these points. Similarity is quantified by the squared difference between these vector components, which is a measure of local *likelihood energy* U^{lk}.

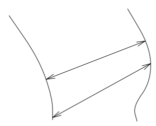
Provided normalised eigenmodes are used, this is found to be an efficient implementation for fitting eigenmode extrema without having to compute surface curvature. Another advantage is that this metric exists continuously over the organ model, instead of just at eigenmode extrema.

Prior energy

The difficulty with matching points between organ models using likelihood energy U^{lk} alone, is that local coherence may not be maintained. Figure 7.2(a) illustrates an incoherent outcome, where points are matched between organ models without reference to neighbouring matches. Figure 7.2(b) illustrates a more desirable outcome, where two matches are mutually coherent.



(a) Incoherent matches



(b) Coherent matches

Figure 7.2: Matching points using only likelihood energy may not give coherent local patterns

The purpose of the Gompertz energy metric is to ensure a coherent pattern of local matches. It does this by encouraging patterns which are most biologically plausible, i.e. those which entail minimal mass change. Mass change can be computed by the volume change associated with the set of *homology vectors*, i.e. the vectors linking matching points.

When there are only a few matching points, however, homology vectors have to be estimated for the remaining unmatched points. This is done by projecting the existing homology vectors onto eigenmodes, then using the eigenmodes' components at unmatched points as estimates for the missing homology vectors. The interpolated volume change then gives the Gompertz metric of local *prior energy* U^{pr}.

The aim now is to combine likelihood energy U^{lk} with prior energy U^{pr} in a common framework, so that registration can be performed using a single combined measure of homology, which explicitly enforces coherent local patterns of homology vectors between organ models.

Markov neighbourhoods in finite element models

Markov neighbourhoods offer such a framework, because energies can be translated to probabilities, which can then be combined with each other using Bayes' theorem. The first step is to show that finite element organ models form Markov neighbourhoods.

In the finite element method, the displacement of a point within an element is solely a function of the displacements of the element's nodes (see Appendix A). Consider the special case when this point coincides with an element node: the node is shared between adjacent elements, so its displacement is only influenced by the nodal displacements of all these adjacent elements.

The neighbourhood formed by such a group of adjacent elements is Markovian, so each nodal component of an eigenmode exists within a Markov neighbourhood, also known as a *Markov random field* (MRF).

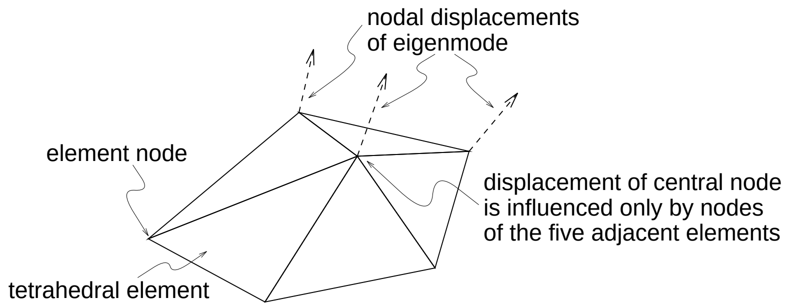


Figure 7.3: Five tetrahedral finite elements in a Markov neighbourhood

MRF-Gibbs equivalence

The *MRF-Gibbs equivalence* (Chou & Brown [27]) maps energy values in a MRF into probability values. Likelihood and prior probabilities for the ith MRF site on one organ model, being homologous to the jth site on another model, are given by

$$P^{\mathrm{pr}}(i,j) = A_1 e^{-U^{\mathrm{pr}}(i,j)/T_1} \tag{7.15}$$
$$P^{\mathrm{lk}}(i,j) = A_2 e^{-U^{\mathrm{lk}}(i,j)/T_2} \tag{7.16}$$

Bayes' theorem allows the *prior probability* P^{pr} and *likelihood probability* P^{lk} to be combined in the MRF to give the *posterior probability* P^{ps}

$$P^{\mathrm{ps}} = \frac{P^{\mathrm{pr}}P^{\mathrm{lk}}}{Z} \tag{7.17}$$

where Z is a normalisation constant. The posterior probability P^{ps} describes the net probability that a site on one organ model is homologous to a site on another organ model. It is determined, through P^{lk}, by eigenmode component similarity between the sites; it is also determined, through P^{pr}, by the coherence of homology vectors in its Markov neighbourhood.

Bayesian combination in equation (7.17) shows that *posterior energy* U^{ps} is the sum of U^{pr} and U^{lk}. However, this summation can only be performed after appropriate normalisation of the unknowns in equations (7.15)–(7.17), i.e. A_1, A_2, T_1, T_2, and Z.

Normalising prior and likelihood energies

One normalisation constraint for equations (7.15) and (7.16) is

$$A_1 = A_2 = 1 \tag{7.18}$$

which translates to a constant offset to U^{pr} and U^{lk} in each MRF. A probabilistic constraint also applies

$$\sum_{j \in \text{MRF}} P(i,j) = 1 \tag{7.19}$$

when matching the ith site on one organ model, to the central site of an MRF in another organ model.

Dropping the normalisation offset Z from equation (7.17) gives the posterior energy U^{ps} (scaled by some T_3)

$$U^{\text{ps}} = U^{\text{pr}}/T_1 + U^{\text{lk}}/T_2 \tag{7.20}$$

By using the change of variable $T' = -\frac{1}{T}$ over an MRF of n sites, the probabilistic constraint in equation (7.19) translates to[5]

$$\begin{aligned} \sum_{i=1}^{n} e^{U_i T'} &= 1 \\ &= (e^{U_1})^{T'} + (e^{U_2})^{T'} + \ldots + (e^{U_n})^{T'} \end{aligned} \tag{7.23}$$

This can be used to find T_1 and T_2, after which equation (7.20) gives the posterior energy U^{ps} as the weighted (Bayesian) sum of the likelihood energy U^{lk} and the prior energy U^{pr}.

7.5 Summary

This chapter introduces the Gompertz function as a model of cell proliferation. The Gompertz growth function is associated with an energy metric which describes the cost of organ growth, where this cost is approximately given by the (small) change in mass during organ growth.

Chapter 6 shows that a point on one organ model can be matched to a point on another organ model, by comparing their eigenmode components. However, if points are matched without reference to other matches in the neighbourhood, an incoherent pattern of matches will result.

The Gompertz energy metric encourages local coherence in these patterns by choosing the most biologically probable patterns. This constraint is combined with eigenmode matching using a Bayesian framework, to give an overall probability for any suggested homology between organ models.

[5] T' can be found by computing the root of the monotonic function

$$y = (e^{U_1})^{T'} + (e^{U_2})^{T'} + \ldots + (e^{U_n})^{T'} - 1 \tag{7.21}$$

using a simple Newton-Raphson iteration

$$\begin{aligned} T' &\rightarrow T' - \frac{y}{\frac{dy}{dT'}} \\ &= T' - \frac{(e^{U_1})^{T'} + (e^{U_2})^{T'} + \ldots + (e^{U_n})^{T'} - 1}{U_1(e^{U_1})^{T'} + U_2(e^{U_2})^{T'} + \ldots + U_n(e^{U_n})^{T'}} \end{aligned} \tag{7.22}$$

which gives values for T_1 and T_2 in equation (7.20).

Chapter 8

Experiments: 3D Registration

Chapter 6 shows that eigenmode extrema are biological landmarks which can be used to register organ models. Chapter 7 then uses a Bayesian framework to show how eigenmode similarity can be combined with the Gompertz energy metric, to give a posterior probability for a match between any point on an organ model, with any point on another organ model.

The space of possible homologies between two organ models is large, so an optimisation algorithm is required to find the map of homologies which minimises these posterior probabilities. Such a *homology map* defines the non-rigid registration of the two organ models.

This chapter introduces the *highest confidence first* (HCF) algorithm, which considers local evidence in order of global importance when performing optimisation. Thus, the homology map is initially filled in with the most probable matches, with less probable matches waiting for neighbourhood support before making commitments.

A globally optimal registration is arrived at by steadily guiding the HCF solution to increasing spatial resolutions, using eigenmodes in decreasing order of spatial scale.

Chapter organisation

- **Section 8.1** outlines characteristics of the optimisation problem, and presents the solution strategy.
- **Section 8.2** describes the highest confidence first (HCF) algorithm, which tackles the global optimisation problem by scheduling the propagation of local evidence in order of importance.
- **Sections 8.3** describes a scale-ordered implementation of HCF, which guides the optimisation solution to a global registration by using eigenmodes of gradually improving spatial resolution.
- **Sections 8.4** illustrates the relative importance of *likelihood* features (i.e. eigenmode similarity) in defining homologies, compared to the *prior* constraint (i.e. Gompertz energy).
- **Sections 8.5** illustrates HCF registration results achieved using a *posterior* combination of eigenmode features and the Gompertz energy metric.
- **Sections 8.6** shows six sets of registration results, performed on all pair combinations of the Liver-A, Liver-B, Liver-C, and Liver-D models.
- **Section 8.7** discusses the HCF registration algorithm and the registration results presented in this chapter.

8.1 Mapping homologies

Registration of biological organs is a difficult task which, practically speaking, can only be achieved by an experienced anatomist. It requires a significant amount of spatial reasoning, as well as anatomical and medical knowledge about the range of normal and abnormal organ shapes. Anatomists always perform this correspondence task using biological landmarks, cf. Bookstein's [16] work in landmark morphometry referred to in Chapter 3.

Relevant current research in medical imaging is mainly directed towards methods of *non-rigid registration* of 2D and 3D brain images. This is motivated by the desire to map standard atlases to patient data, for the purpose of automatic labelling of brain structures (Gee et al. [45]; Christensen et al. [29]). Automated segmentation and archiving then makes statistical analysis and training from large datasets feasible.

Betting & Feldmar [10] report a technique for *rigid-body registration* of MR or CT images to the actual position of a patient, which can be very useful in neurosurgery or laparoscopy. Rigid-body registration is also important for integrating functional or anatomical information in multiple imaging modalities (Hill et al. [61]).

Eigenmodes have previously been used by Sclaroff & Pentland [130] in registration (using *modal matching*) of images in a video database, leading to applications in object recognition and classification. An "affinity table" is compiled of the strongest correspondences between two objects of similar pose, from which the overall map of correspondences is derived.

Syn & Prager [143] apply a rotation-invariant extension[1] of this algorithm with some success to an artificial dataset of 3D ellipsoidal models, but it fails with more complex organ models such as the foetal liver models from Figure 4.2.

Optimisation strategy

Chapter 7 develops a posterior measure which describes the probability of homology between two sites, conditioned by a local constraint of biological plausibility. These local probabilities have to be integrated over the entire organ model, in order to arrive at a global map of the most probable homologies at all sites. This resultant *homology map* describes the non-rigid registration between two organ models.

Fundamental eigenmodes have only one phase change over the organ model, so a non-rigid registration using a fundamental eigenmode can give two locally optimal fits[2]. Either fit will be of low spatial resolution because there are few displacement extrema on fundamental eigenmodes. As registration proceeds using higher eigenmodes, however, the non-rigid registration gradually improves in spatial resolution due to the increasing spatial-frequency of eigenmode extrema.

Chapter 6.3 shows that fundamental modal axes give a rigid-body registration, which is performed before embarking on non-rigid registration. Provided the fundamental modal axes are correctly normalised, this initial rigid-body registration should lie close to the correct non-rigid registration for the fundamental eigenmodes.

The homology map is then refined further using higher eigenmodes. The number of similar eigenmodes two organ models have in common, determines the spatial resolution to which the homology map can be determined.

[1] See Chapter 6.2 for a description of modal projection.

[2] Two optima of fit between a pair of organ models occur, because the two phases of each fundamental eigenmode can be matched in two ways.

8.2 Highest confidence first

The optimisation problem outlined above has the following characteristics:

(i) initialisation within the domain of a local optimum;

(ii) graduated presentation of increasingly localised optima;

and the optimisation strategy is to:

(i) initialise the homology map using a fundamental eigenmode;

(ii) refine the homology map using higher eigenmodes, while taking into account the registration given by previous eigenmodes.

The *highest confidence first* (HCF) algorithm is implemented to meet these criteria. HCF is an efficient algorithm which embodies the principles of graceful degradation and "least commitment" (Marr [86]), with a small and predictable run-time (Chou & Brown [27]). It performs an approximate *maximum a posteriori* (MAP) optimisation using local homology estimates, with global scheduling to rank commitment decisions in order of importance.

The experiments in this chapter, first presented in Syn et al. [145, 149], represent the first application of the HCF algorithm to non-rigid registration of organ models; previous applications have been in stereo correspondence of image features (Lim [82]) and in image labelling (Chou & Brown [27]). HCF is itself an improved adaptation of the *iterative conditional modes* (ICM) algorithm (see Lim [82] for a comparison of ICM and HCF).

State-of-the-art methods for non-rigid registration of brain images similarly apply Bayesian combination of evidence: Christensen et al. [29] employ a viscous fluid prior, and Gee et al. [45] measure local likelihoods using the local intensity cross-correlation between the brain images. Both approaches lead to complex optimisation landscapes, which are solved using iterative methods.

Augmented search space

The problem specification in the Section 8.1 shows that the local optimisation landscape for each site, consists of the posterior energy of its homology to the set of possible sites on the other body. The most probable homology is the one with the smallest posterior energy.

The task of the HCF algorithm, therefore, is to decide which local optimisations to perform first, and to propagate information about these decisions to other sites which have yet to decide. This allows confident sites to commit to homologies early on, while equivocal sites wait for neighbourhood support before making a commitment.

Chou & Brown [27] define an *augmented search space* as the set of possible homologies or *commitments* for a site, augmented with the additional option of being *uncommitted*. After a site has committed to a homology, it can still change its commitment but cannot nullify it. The search space can be pruned to a set of more probable homologies (e.g. on the basis of Euclidean distance) to reduce computational effort.

The *augmented posterior energy* is a modified measure which only considers committed sites in the MRF; from now on all references to posterior energy imply this augmented measure. An uncommitted site does not therefore participate in the computation of its neighbours' posterior energy; however, it always takes the states of active neighbours into account when making a commitment.

Local stability measures

The order in which sites have their commitments updated is determined by a *priority queue*, which has the least stable site at the top of the queue (also known as a *heap*).

The stability measure for a committed site is the difference in posterior energy between the current commitment s_1 and a commitment (different from the current one) s_2, which gives the smallest posterior energy. This represents the largest reduction in local posterior energy which can be achieved by a change in commitment

$$G = \min_{s_2 \neq s_1} U^{\mathrm{ps}}(s_2) - U^{\mathrm{ps}}(s_1) \tag{8.1}$$

The stability measure for an uncommitted site is the negative difference in posterior energy between the sites which have the two smallest posterior energy values. This indicates the depth of the energetic 'well' which the best commitment sits in, and is always negative so that HCF does not converge until all sites are committed

$$\begin{aligned} G &= \min_{s_2 \neq s_1} -\Big(U^{\mathrm{ps}}(s_2) - U^{\mathrm{ps}}(s_1)\Big) \\ \text{where} \quad U^{\mathrm{ps}}(s_1) &= \min_{s} U^{\mathrm{ps}}(s) \end{aligned} \tag{8.2}$$

Note that s, s_1 and s_2 are sites on another organ model, which lie within the search space of the site whose stability is being computed.

8.3 Implementation

All sites are initialised as being uncommitted to begin with, so that the augmented local posterior energies are essentially determined by local likelihood. Therefore, sites with strong likelihood evidence of one commitment over others (i.e. highly unstable sites), will be visited early in the construction of the homology map. Sites without strong likelihood evidence will delay making an initial commitment, and will take their neighbourhood configuration into account when they finally do so, thus ensuring locally coherent commitments.

A serial implementation of HCF still has the weakness that the first commitment made can strongly influence subsequent commitments. Chou et al. [28] present *local HCF* as a parallel implementation, which addresses this potential instability by integrating initial commitments over the network of sites. Parallelisation also promises a major performance improvement over serial HCF, if executed on a suitable hardware platform.

Graduated spatial localisation

HCF terminates when the top of the stability queue, which represents the least stable site, is positive. The homology map thus obtained is used to update the estimate of rigid-body registration, thus improving the accuracy of local prior energy measures during the subsequent HCF run, which employs the next higher eigenmode as the likelihood feature.

Successive runs are initialised using the homology map from the previous run, which gives a principled method of propagating posterior estimates from all previous runs as prior information in the current run. Morever, the graduated spatial localisation achieved by using consecutive eigenmodes of increasing spatial-frequency, allows a stable descent into a global registration optimum of gradually increasing spatial resolution.

This approach is similar to Blake & Zisserman's (ch.3.5)[12] *graduated non-convexity* (GNC), where a non-convex function is approximated by a series of functions, initialised as a convex function with a single global minimum, whose shape gradually approaches the true cost function. As illustrated by Figure 8.1, these approximating functions are initialised at the convex minimum, then successively minimised in the vicinity of the previous minimum, so that the true global minimum is approached without interference from spurious local minima.

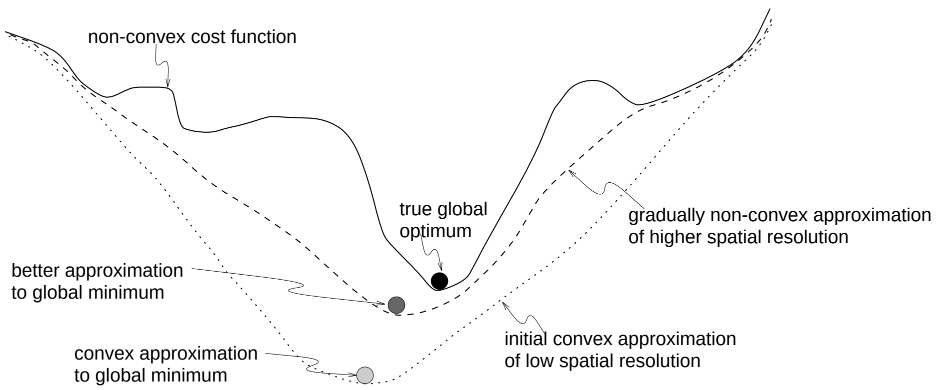
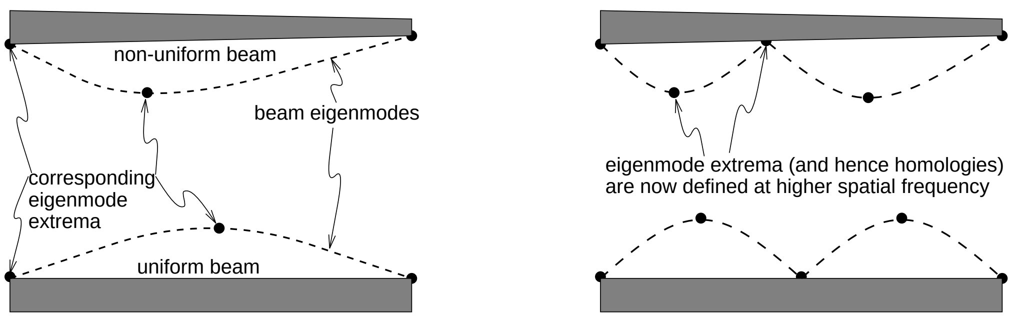


Figure 8.1: Successive approximations to non-convex function using graduated non-convexity (adapted from Blake & Zisserman [12])

Figure 8.2(a) shows that a fundamental eigenmode, which has few displacement extrema, will give rise to a registration of low spatial-frequency. An HCF run using this eigenmode terminates with a homology map which is locally determined with strong likelihood at these extrema[3]. Where evidence of homology is weaker, the Gompertz prior propagates information from these extrema, to give an interpolation throughout the homology map which accords with the Gompertz growth model.



(a) Fundamental eigenmode

(b) Higher eigenmode

Figure 8.2: Registration of uniform beam to non-uniform beam proceeds with graduated spatial refinement, by employing successive eigenmodes of increasing spatial-frequency

Figure 8.2(b) shows that, after inheriting the fundamental eigenmode's homology map, registration with a higher eigenmode offers an increased number of extrema for improving spatial resolution in the homology map.

[3]Recall that a posterior energy is computed for all possible homologies to a site, and that the stability of the site is determined by the *depth* of the energetic well. Therefore, a displacement extremum which commits to another (appropriately normalised) displacement extremum gives the greatest local stability.

Computing prior energy

Local prior energies in a MRF are usually defined with respect to *cliques*, which are totally connected subgraphs in the MRF. Cliques allow the prior energy to be built up piecewise from different configurations of spatial dependency; examples of energy functions assembled from (heuristically assigned) clique functions are found in Chou & Brown [27].

In contrast, the Gompertz energy metric is directly dependent on local mass change, and does not need to be assembled from clique functions. It is an augmented prior measure (i.e. computed only from committed sites) and can be computed from any neighbourhood topology and (non-)commitment configuration. A least-squares fit is locally performed from the eigenmodes to the committed homology vectors, and mass change is approximated as the local volume change associated with the set of eigenmode amplitudes.

The Gompertz prior encourages homology vectors in each element's neighbourhood not only to minimise displacement, but also to form a coherent cluster of commitments. This coherence propagates from committed sites through overlapping MRFs, giving supporting information to undecided sites about the appropriate commitment to make.

In cases where the mesh is both noisily and sparsely sampled, the determination of eigenmode amplitudes is made over-constrained by fitting fewer eigenmodes to the homology vectors. This reduces the ability of local prior energy to discriminate between similar commitment patterns, and dampens the local variability of the homology map.

8.4 Results: relative importance of likelihood features

Gee et al. [45] note that the exact form of the prior function is unimportant, so long as it penalises 'unlikely' homologies and encourages 'likely' ones. As demonstrated in this section, the quality of the homology map is determined very much by the quality of the likelihood features. Note that the wireframe organ models are scaled up by 10% to show the red homology vectors more clearly in Figures 8.3–8.4.

Registration using only prior energy

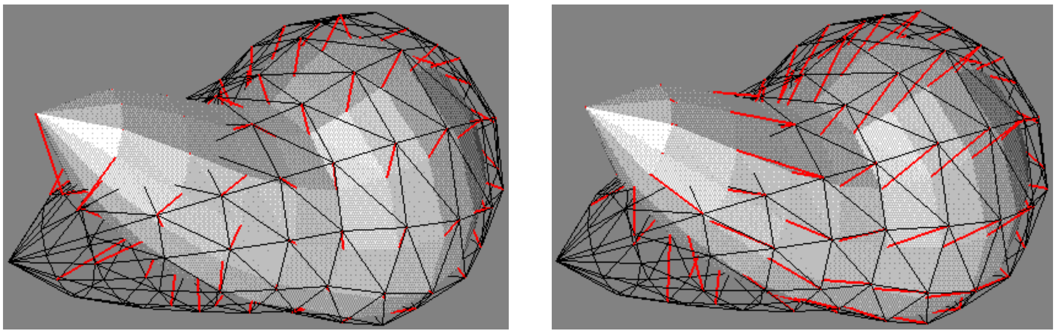
Figure 8.3(a) shows the registration achieved using only a simple Euclidean prior metric in comparing homologous sites. The Euclidean metric performs relatively well in this situation, even without neighbourhood information, because of the lack of noise, good initial pose estimation, and the coherent relative spacing of the two models' element nodes.

Figure 8.3(b) shows the registration achieved using only the Gompertz prior, which is generally more reliable because it seeks neighbourhood support in deciding the biological plausibility of each suggested homology.

Registration using only likelihood energy

Figure 8.4 shows the homologies obtained by using only likelihood features in HCF estimation, i.e. there is no prior constraint on a site when choosing its commitment. Figure 8.4(a) shows that not all sites commit to the correct side of the other model, and that the homology map has poor spatial localisation since only the first eigenmode is used. Figure 8.4(b) shows that spatial localisation is improved by complementing the first with the second fundamental eigenmode.

The two gallbladder models have 18 non-rigid eigenmodes in common, where the 18th eigenmode has a spatial wavelength of approximately 1/4 the length of the model. Figure 8.4(c) shows that with gradual spatial localisation using all 18 eigenmodes, all sites including the



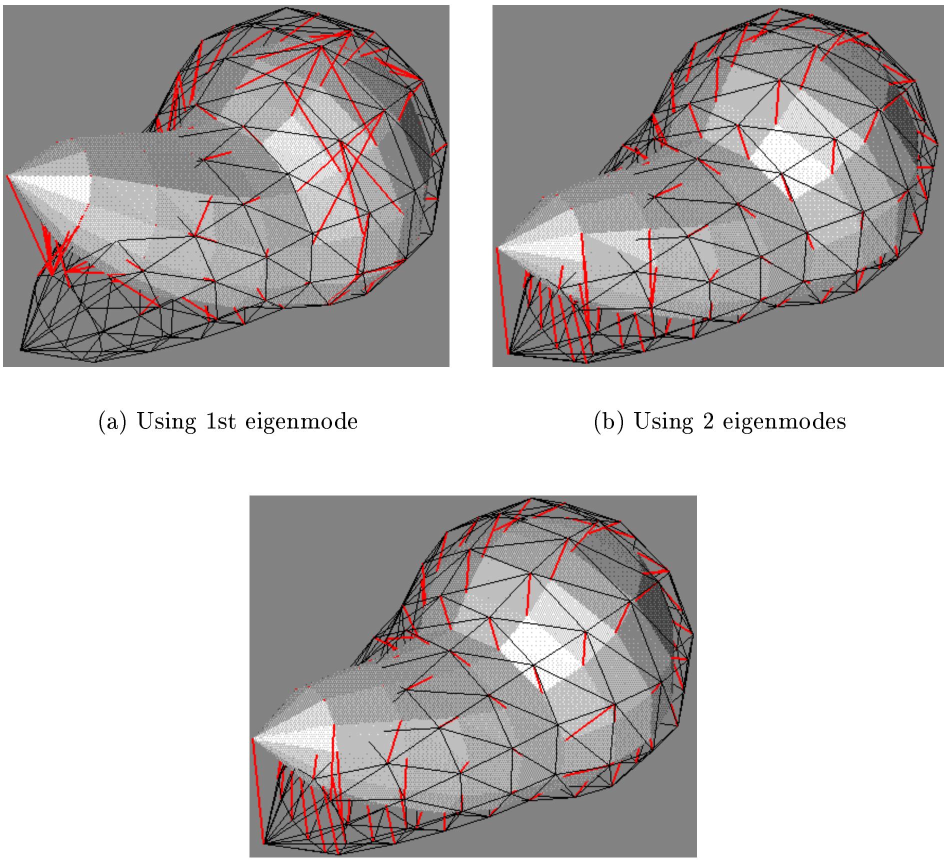
(a) Simple Euclidean prior

(b) Gompertz prior

Figure 8.3: Registration using only prior energy

awkward ones near the gallbladder tip have committed with good spatial resolution. The hidden part of the homology map to the rear of the models, also has improved localisation compared to Figure 8.4(b). The improved pose estimation given by the homology maps results in a better underlying rigid-body registration.

Thus Figure 8.4 shows that eigenmode likelihood is important for determining homologies, whereas Figure 8.3 shows that the Gompertz prior is well suited to maintaining local coherence.



(a) Using 1st eigenmode

(b) Using 2 eigenmodes

(c) Using 18 eigenmodes

Figure 8.4: Registration using only eigenmode likelihood

8.5 Results: registration using posterior energy

Rigid-body registration using fundamental modal axes

Figure 8.5 shows the rigid-body registration of two gallbladder models using fundamental modal axes.

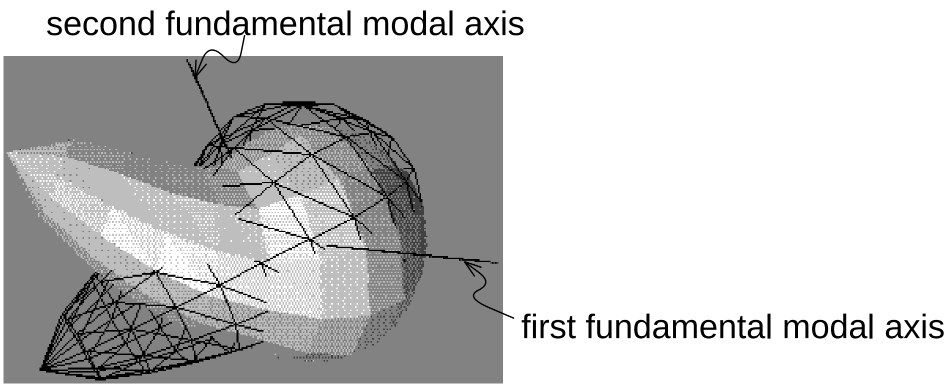


Figure 8.5: Rigid-body registration of gallbladder models by aligning the fundamental modal axes

Figure 8.6 shows the refinement in pose estimation derived from the first and second modal homology maps in Figure 8.7.

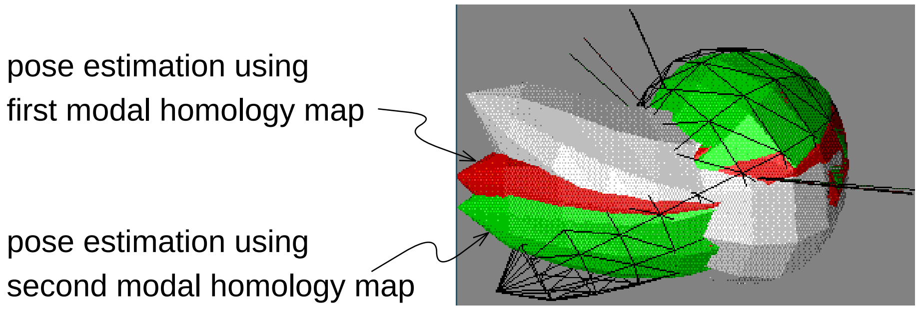


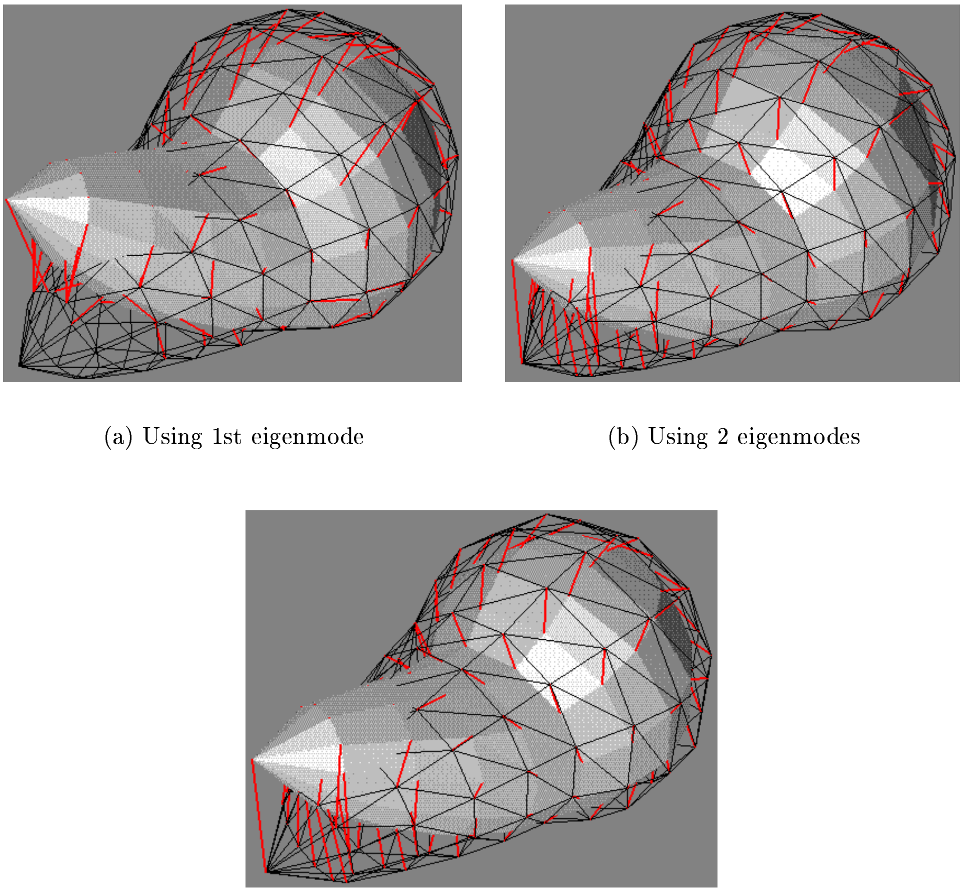
Figure 8.6: Refinement of underlying rigid-body registration given by successive homology maps

Registration of gallbladder models

Figure 8.7 shows the successively refined HCF registration between two gallbladder models. The first three eigenmodes are of fundamental spatial wavelength (cf. Figure 6.5), so that the registration achieved in Figure 8.7(b) should be localised to approximately half the length of the gallbladder model. In fact the presence of the Gompertz prior and the regular spacing of the two meshes, means that the homology map is almost perfectly recovered[4].

The effect of the Gompertz prior can be seen most clearly in Figure 8.7(a), when compared to Figure 8.4(a) which has a messy homology map on the main body of the gallbladder.

[4]Standard FEM modelling practice suggests 8 elements per eigenmode wavelength, so the registration resolution of the gallbladder models according to this criterion is limited to half the model's length.



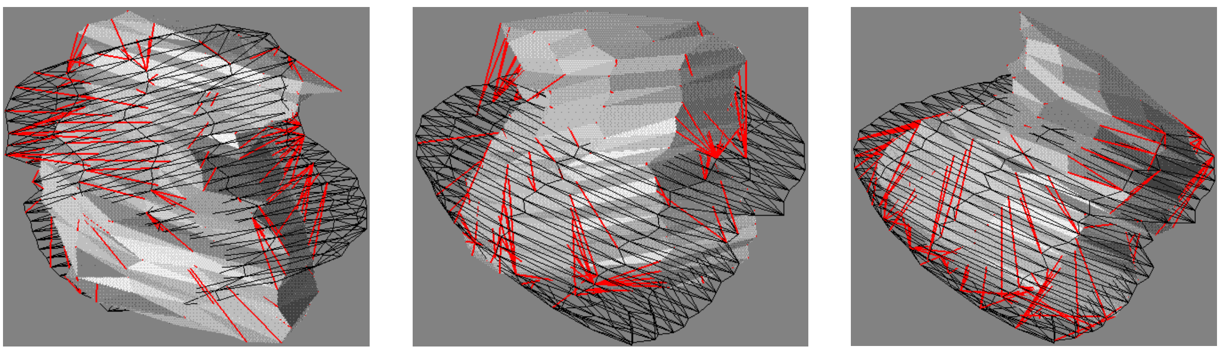
(a) Using 1st eigenmode (b) Using 2 eigenmodes

(c) Using 18 eigenmodes

Figure 8.7: HCF estimate and refinement of gallbladder registration

Registration of worst case liver models

Figure 8.8 shows successive homology maps for two manually sampled foetal liver models from Figure 4.2. The final pose estimate in Figure 8.8(c) is approximately correct, but the homology map computed is almost incoherent. This is due to the sparse and uneven sampling of the liver models, which gives noisy eigenmode features and restricts the choice of homologous sites between models.

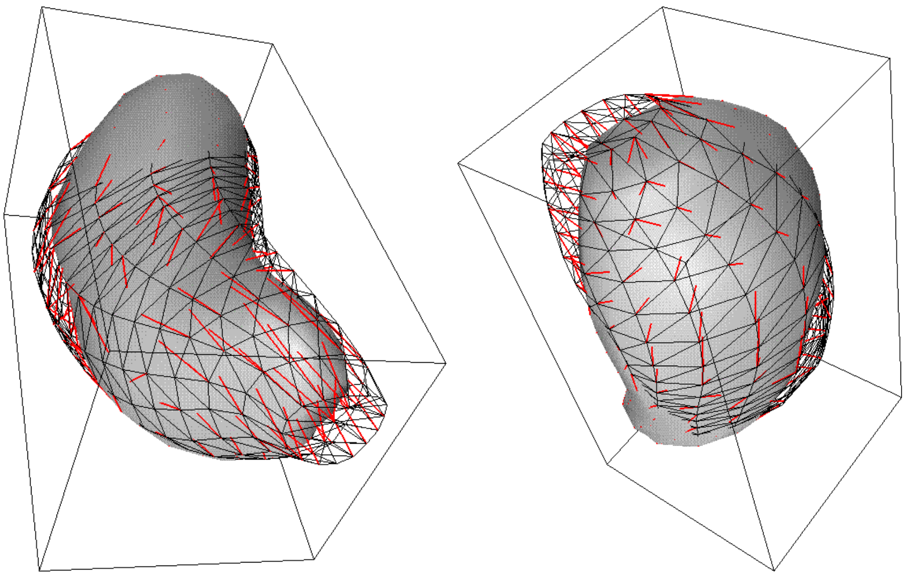


(a) Using 1st eigenmode (b) Using 2 eigenmodes (c) Using 3 eigenmodes

Figure 8.8: HCF estimate and refinement of liver registration

8.6 Results: registration of liver models

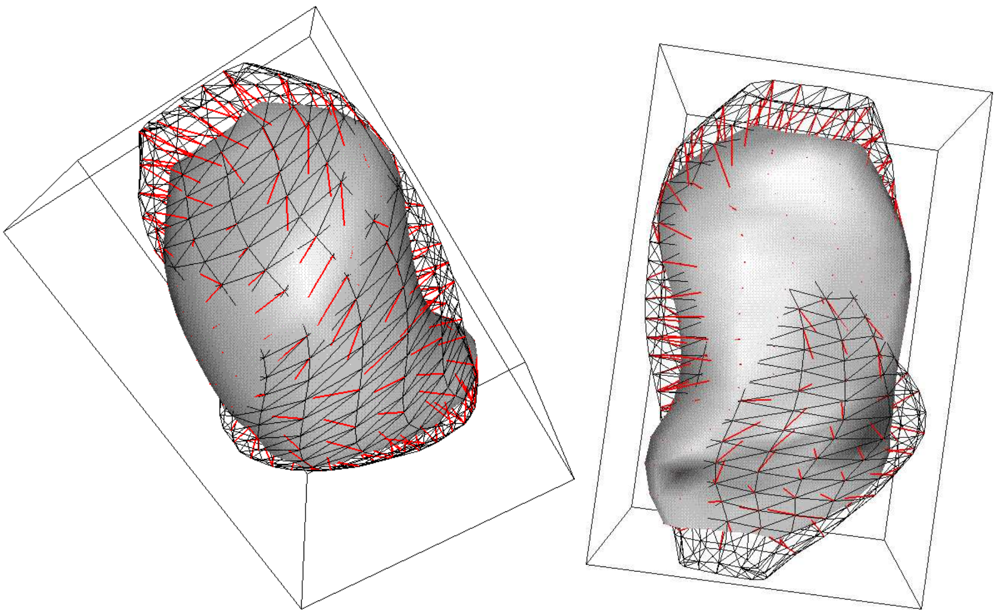
Figures 8.9–8.14 show homology maps for HCF registration between each pair combination of the Liver-A, Liver-B, Liver-C, and Liver-D models originally segmented in Chapter 5.5.



(a) First view

(b) Second view

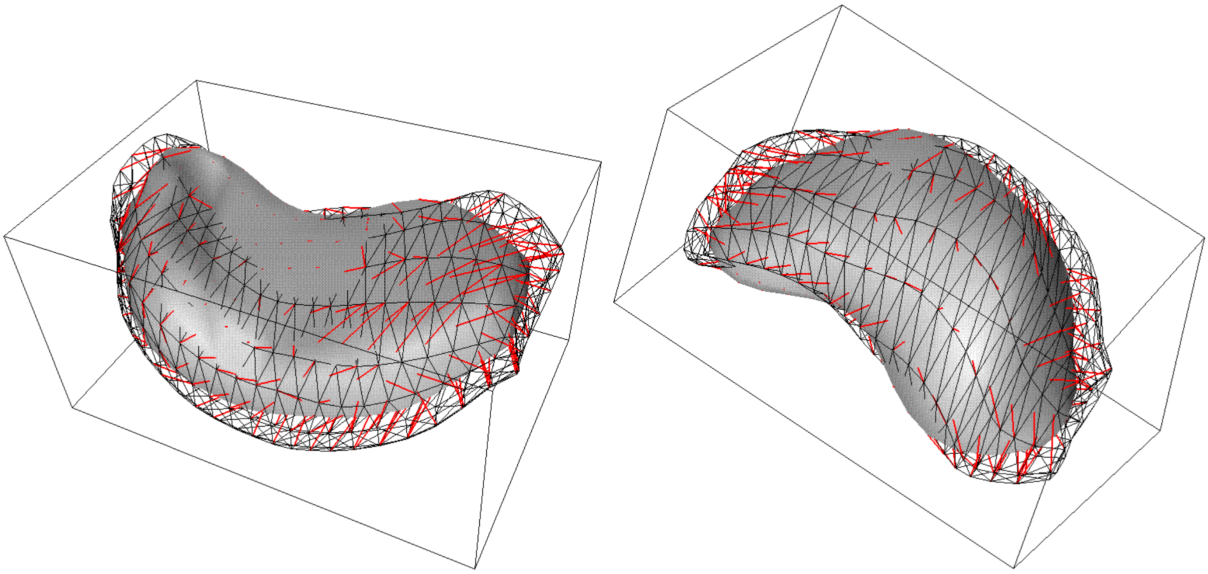
Figure 8.9: HCF registration of Liver-A (solid) with Liver-B (wireframe scaled by 110%)



(a) First view

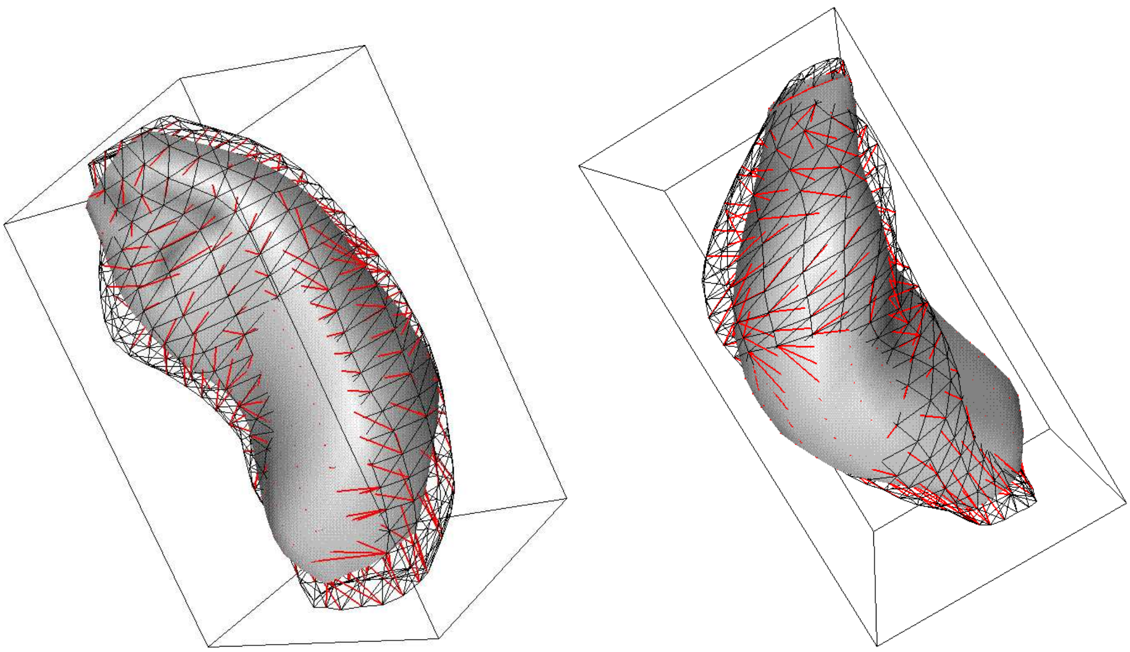
(b) Second view

Figure 8.10: HCF registration of Liver-A (solid) with Liver-C (wireframe scaled by 110%)



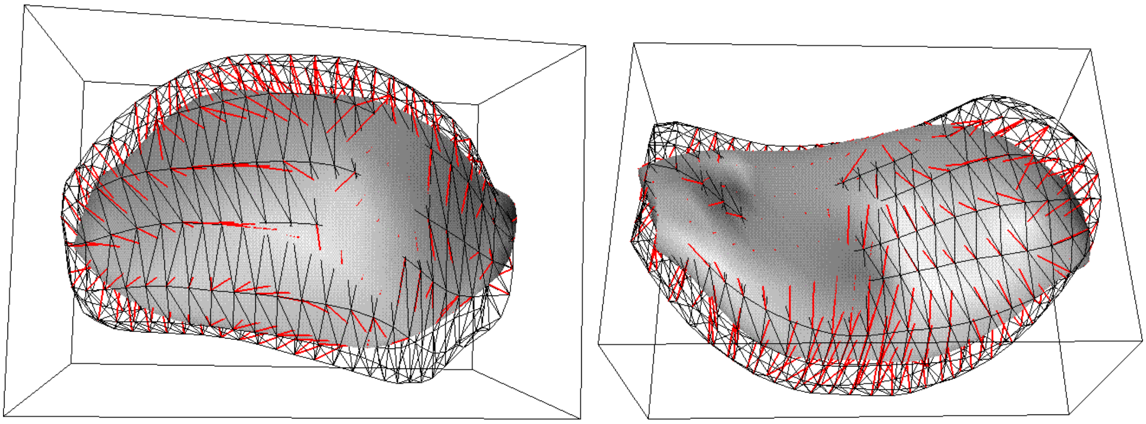
(a) First view (b) Second view

Figure 8.11: HCF registration of Liver-A (solid)with Liver-D (wireframe scaled by 110%)



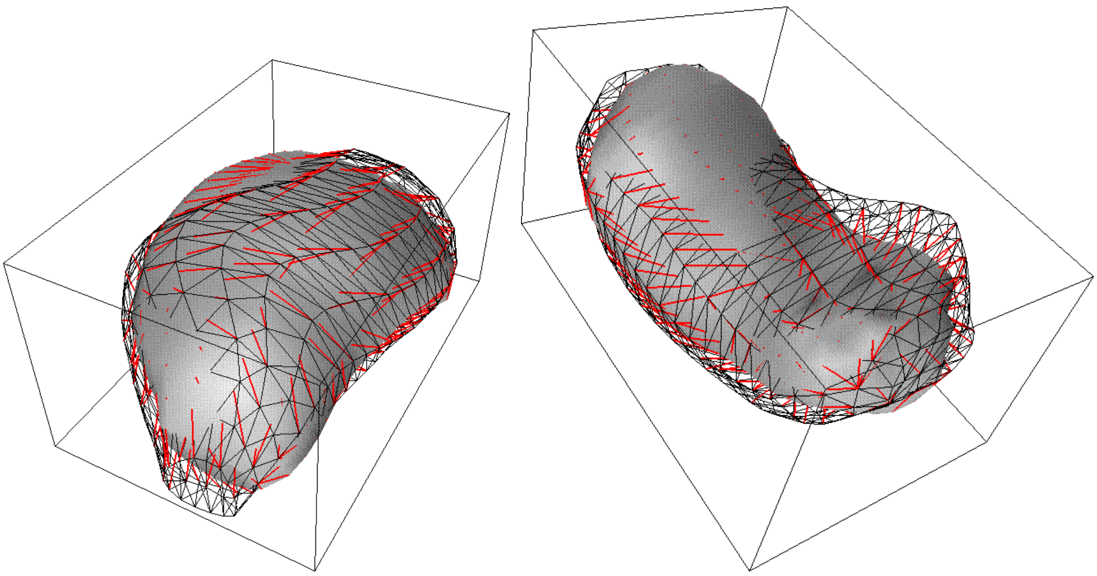
(a) First view (b) Second view

Figure 8.12: HCF registration of Liver-B (solid)with Liver-C (wireframe scaled by 110%)



(a) First view (b) Second view

Figure 8.13: HCF registration of Liver-B (solid) with Liver-D (wireframe scaled by 110%)



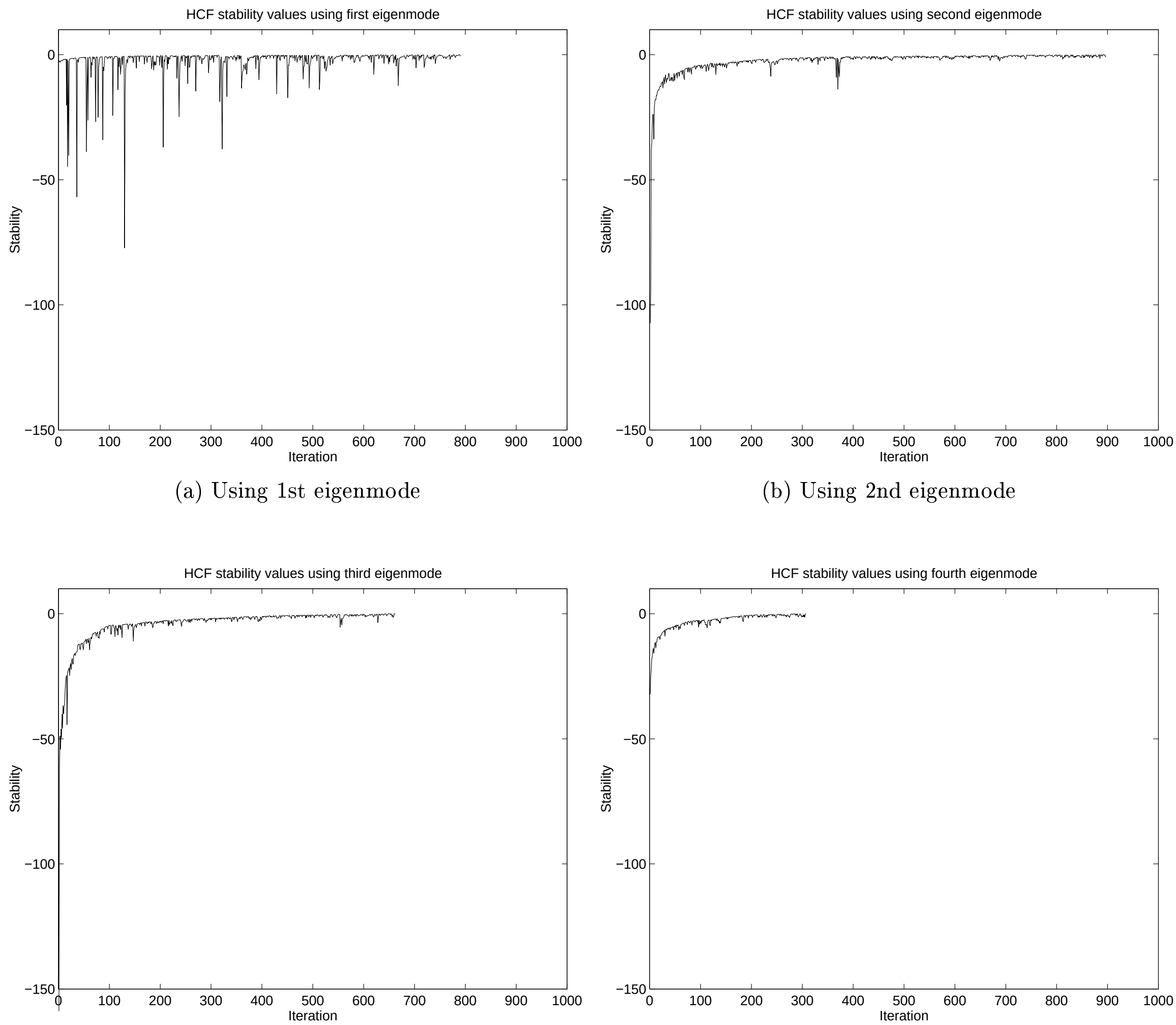
(a) First view (b) Second view

Figure 8.14: HCF registration of Liver-C (solid) with Liver-D (wireframe scaled by 110%)

Four non-rigid eigenmodes are used for HCF registration of the liver models in Figures 8.9–8.14, since eigenmode similarity degrades beyond this[5]. The spatial resolution of the homology map is therefore limited to that of the fourth eigenmode, whose extrema are spaced at approximately half the liver's width in each direction. The local coherence of the homology map, although enforced by the Gompertz prior, is also limited by the relative sparsity of finite element mesh nodes.

The registration results in Figures 8.9–8.14 seem very reasonable overall. However, anatomical accuracy can only really be validated by an experienced clinician.

Profile of HCF runs



(a) Using 1st eigenmode (b) Using 2nd eigenmode

(c) Using 3rd eigenmode (d) Using 4th eigenmode

Figure 8.15: Stability values during HCF registration

[5]The liver meshes are constructed with 3 internal nodes per cross-section, giving rise to significant coupling between opposite sides of each model. This means that eigenmodes with greater than fundamental spatial-frequency are not well sampled by the meshes.

Figure 8.15(a) shows the stability values during HCF registration using the first eigenmode. All sites are initialised as uncommitted, so stability values fluctuate over a wide range of values: low stabilities are due mostly to committed sites, while high stabilities correspond to previously uncommitted sites reaching the top of the stability queue. Figure 8.16 shows only the (high) stability values for these uncommitted sites, at the time they first make a commitment.

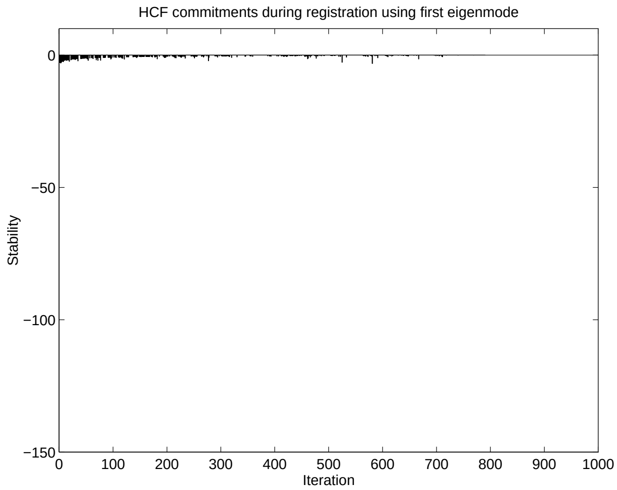


Figure 8.16: HCF commitment behaviour during registration using first eigenmode

All sites are committed after the first HCF pass, and Figures 8.15(b)–8.15(d) show that stability values converge more smoothly during subsequent passes with the second, third, and fourth eigenmodes.

Each liver model has approximately 500 nodes, so the HCF runs make only a small number of passes over each node, since each run requires fewer than 1000 iterations. Note that the search space of possible homologies (for each site on one model) is restricted to the 250 closest nodes (on the other model) at the start of each HCF run.

8.7 Discussion

Initial rigid-body registration

The eigenmode normalisation scheme described in Section 6.2 relies on volume change as a reliable macroscopic property of the eigenmode field. Eigenmode directions and magnitudes are determined from this, so as to give a unique initial pose estimate for each organ model. A more robust scheme is desirable, however, when the model's (and hence the eigenmodes') spatial sampling is unreliably sparse and noisy.

Without a unique normalisation of the first two fundamental modal axes, there are four possible rigid-body registrations between two organ models. The best one can be chosen by comparing a 'best fit' metric, computed by summing the distances between each surface element (on one model) and its nearest neighbour (on the other model).

The initial pose estimation process should be made separate from subsequent non-rigid registration, since pose information is in principle available from patient orientation during ultrasound scanning. In fact, ultrasonographers routinely use such information, derived from knowledge of standard scanning procedure or from annotation of image sequences, in order to visualise 3D structures in ultrasound images.

Registration without normalisation of eigenmode directions

The Necker cube illusion (Fischler & Firschein (ch.8)[41]) strongly suggests that the human visual system performs local optimisation in fitting a model to the image, while actively peturbing each local optimum to seek alternative optima. This is consistent with a scheme in which a number of hypothesised registrations are computed and then compared.

In this scenario, the homology map is initialised using only the prior energy measure, since the eigenmode likelihood features are not assumed to be direction-normalised. This is done for each of the four possible rigid-body registrations given by the fundamental modal axes.

For every pair of corresponding eigenmodes, displacement components are compared at homologous sites on each model, which determines whether the two eigenmodes are signed consistently. This allows eigenmode signs to be normalised with respect to each other, and the pair of eigenmodes can then be used to refine the homology map.

This sign-normalisation and homology refinement process is repeated for successive eigenmodes. Four final homology maps result from each of the four initial rigid-body registrations, so the final homology map which has the lowest net posterior energy is chosen as the best-fitting one. This strategy has not been implemented, but may be useful in situations where direction normalisation of eigenmodes cannot be reliably achieved.

Projected modal components do not require direction normalisation, and can also be usefully employed when there are enough eigenmodes available (see Chapter 6.2 and comments below about eigenmode features).

Deterministic versus statistical algorithms

The deterministic local-search approach of HCF estimation is also in accordance with the Necker cube illusion. HCF is preferable to stochastic algorithms such as simulated annealing (van Laarhoven & Aarts [165]), which are computationally expensive although globally optimal. HCF also has the advantage of an explicitly defined schedule for updating sites, while simulated annealing requires additional parameters to control the annealing schedule. Note that stochastic algorithms do not give repeatable results so the quality of solutions needs to be assessed over a number of runs.

Gee et al. [45] make the point that Bayesian matching algorithms are highly dependent upon the quality of the likelihood features used, and that the prior measure has just to behave as a reasonable penalty function; eigenmodes are powerful likelihood features, so our need for a globally optimal stochastic algorithm is much reduced. Also, the scale-ordered application of eigenmodes guides HCF registration to the global optimum with gradually improving spatial localisation, and without interference from spurious registration solutions.

Comments on Gompertz prior

The Gompertz prior has a number of attractive properties:

(i) it is experimentally and statistically motivated as a metric of biological growth energy;

(ii) it encourages coherent (i.e. consistent with Gompertzian growth) local patterns of commitment, propagated by HCF over the entire homology map;

(iii) it is locally computable for any clique configuration in a MRF.

The Gompertz energy metric is well suited to the needs of HCF registration, has a straightforward intuitive interpretation, and is simply computable on finite element organ models.

In comparison, most penalty functions employed in medical imaging are mechanically inspired, e.g. membrane or thin-plate models. The viscous fluid prior used by Christensen et al. [29] is intended to improve on these when applied to brain image registration, and has produced particularly impressive results. However:

(i) viscous fluids specifically allow large non-linear strains (typical in matching brain sulci), and have little biological basis as penalty functions;

(ii) the viscous fluid prior cannot be computed for arbitrary clique commitment configurations.

Comments on eigenmode features

FEM eigenmodes are computed using the block Lanczos algorithm described in Appendix B. The implementation used in this dissertation employs a slow but robust reorthogonalisation scheme, which means that computing more than a few eigenpairs of large eigensystems is expensive; morever, higher eigenmodes are computed with less accuracy due to roundoff errors during orthogonalisation against lower eigenmodes.

These computational limitations are important drawbacks, because densely and accurately sampled organ model meshes are likely to have a larger number of similar eigenmodes, each of which is also more accurately sampled on the meshes. Given sufficiently many similar eigenmodes, projected modal components (described in Section 6.2) can be used as likelihood features. The projected components are defined relative to lower eigenmodes and are therefore rotation invariant, so that there is no need for pose estimation before registration. However, it is unclear how the number of similar eigenmodes can be established *a priori*.

The Gabor wavelet modes developed in Chapter 6.4 also seem an excellent class of shape feature, which have many of the properties of eigenmodes, as well as resilience to missing shape information. However, registration using wavelet modes is currently impractical due again to the very high computational cost of producing wavelet modes to sufficient spatial resolution.

Comments on HCF registration

Registration by expert anatomists is performed by selection of corresponding landmarks, and we similarly employ eigenmode extrema as scale-ordered sets of landmarks, which have a straightforward biological interpretation (in the context of our growth model) as locations of maximal and minimal growth. The results in this chapter, particularly in Section 8.4, indicate the validity of this argument; the quality of registration results are as expected, however, given the intrinsically excellent shape and symmetry properties of elastic eigenmodes.

Interpolating homologies between the landmarks is an ill-posed – and usually irrelevant – problem in clinical practice, but the Gompertz prior is well qualified to perform the interpolation, on both experimental and theoretical grounds. The Gompertz prior acts in an intuitive fashion by penalising local patterns of growth which result in excessive mass flux, and is approximately quantified by the volume change due to local growth patterns. The quality of discrimination between growth patterns is determined by the organ model's mesh, since a sparsely sampled mesh limits the range of possible homologies.

The efficient and deterministic HCF algorithm is well suited to computing a global registration from local posterior evidence, which combines the likelihood of eigenmode similarity, together with a prior constraint of local consistency with Gompertzian growth. The scale-ordered HCF implementation described in this chapter refines registration results while avoiding local optima, by employing eigenmodes of increasing spatial frequency as the likelihood features.

8.8 Summary

The HCF algorithm efficiently and deterministically propagates local evidence of biological homology, to form a global estimate of non-rigid registration. Local evidence is presented to the algorithm in increasing order of spatial-frequency, so that registration is safely conducted to a globally optimal solution at gradually increasing spatial resolutions.

Each local posterior measure is computed by Bayesian combination of the likelihood of homology, together with a prior measure of the coherence of local commitments. The prior measure is computable for arbitrary commitment configurations, so that sites can be initialised as uncommitted, when there is no evidence on which to commit.

Chapter 9

Discussion and Conclusion

This dissertation is motivated by the need to extract organ shapes from 3D freehand ultrasound images. A model-based strategy is proposed for this purpose, which employs an organ shape model to geometrically organise fragmented cues about tissue boundaries in the images. The organ model incorporates prior knowledge about the expected range of organ shapes, and is required to refine this idealised knowledge by learning from organ shape examples.

Chapters 2–4 develop a representation of organ shape which fulfills these requirements. Chapter 5 then demonstrates the model-based strategy, using organ shape models to estimate organ volumes from 3D freehand ultrasound images. In order for the organ model to learn from organ shape examples, both are first parameterised in a common shape coordinate system using a method developed in Chapters 6–8.

This chapter discusses the approaches to organ shape modelling taken in this dissertation, assesses their limitations, and suggests areas of future research to address these limitations.

Chapter organisation

- **Section 9.1** reviews and discusses the material presented in this dissertation.
- **Section 9.2** suggests promising avenues of future research.
- **Section 9.3** concludes this dissertation.

9.1 Discussion

Summary

An organ growth model is developed using the Oster-Murray mechanisms of substrate *deformation*, cell *diffusion*, cell *convection*, and cell *proliferation*. The mechanisms' eigenfunctions describe the growth patterns of the modelled organ. The organ's material properties are not usually known, however, so they are assumed to be uniform. Under this assumption it turns out that the mechanisms have identical eigenfunctions.

These eigenfunctions are approximated by *elastic eigenmodes* using a finite element model of the organ, so 'bootstrapping' the organ model with elastic eigenmodes gives it an idealised knowledge of organ shape variation. This knowledge is used to segment 3D ultrasound images, in conjunction with both automated and operator-assisted means of ultrasound tissue detection. This knowledge is also refined by incorporating the segmented organ shapes, so as to make it more representative of inter-patient variation.

Elastic eigenmodes make excellent shape features and are ordered in spatial scale. They represent *generalised symmetries* of the underlying organ model, because eigenmodes emphasise the organ model's dominant symmetry at each spatial scale. Eigenmodes of large spatial scale are resilient to noisy and uneven finite element meshing, and are the most accurately and easily computed. They form a 'low-pass' subspace which excludes shape variations of small spatial scale, while at the same time describing the most statistically important modes of shape variation in the growth model. 'Lower' eigenmodes are therefore preferentially used as shape features and as modes of shape variation.

Similar organ models have similar eigenmodes, and the eigenmodes' displacement extrema form good biological landmarks. Therefore, organ models can be *registered* to each other by choosing them to be *homologous* at these landmarks. A complete registration requires homologies to be specified everywhere on the organs, however, not merely at landmarks.

The *Gompertz energy* metric is introduced to interpolate the registration between landmarks. A simple Bayesian combination of the *likelihood* of eigenmode similarity, with the *prior* constraint of conformity with Gompertzian organ growth, gives a *posterior* measure of homology at all points on and between landmarks.

The *highest confidence first* (HCF) algorithm then registers two organ models by maximising the posterior probability of all homologies. HCF is an efficient and deterministic algorithm, which is designed to seek the globally optimal registration using a multi-resolution implementation.

Review

A model-based strategy is employed in two respects:

(i) tissue landmarks are organised into a coherent shape using an organ shape model;

(ii) eigenmode homologies are organised into a coherent homology map using the Gompertz model.

An organ shape model is employed in (i) to help interpret noisy, non-parallel, and potentially self-intersecting ultrasound images. Given that tissue boundary landmarks are (by some means) detected in the images, the organ's boundary is interpolated by the shape model, which has prior knowledge of the expected range of organ shapes. In other words, (i) is a model-based strategy for reconstructing organ anatomy from sparse and noisy boundary cues in 3D freehand ultrasound images.

The Gompertz model of organ growth is similarly employed in (ii) to fit organ shape models to each other using homologous landmarks, which are in this case the eigenmode displacement extrema. A 'biologically correct' (with respect to the Gompertz model) interpolation is achieved between these landmarks, so that organ shapes can be parameterised for biometric analysis.

The methods and assumptions used for both approaches are developed and justified, and results are presented to demonstrate their validity. Before concluding this dissertation, however, the following general topics deserve some further discussion:

(i) organ shape models and their representation of shape variation;

(ii) volume estimation using organ shape models;

(iii) registration of organ shape models;

(iv) limitations of our methods and assumptions, and possibilities for future improvement.

Organ shape models

Even if our proposed growth model incorporates an insufficient number of mechanisms to fully reflect the true biology of organ growth, the use of growth eigenfunctions is well justified. Given accurate shape and material properties for the organ model, the set of eigenfunctions describes all the possible patterns of organ growth. Unfortunately, the material properties are not easily established[1], but their uniformity means that deformation, diffusion, and convection in the model give identical patterns of mass flux described by the elastic eigenfunctions.

Elastic eigenfunctions can be approximated as elastic eigenmodes using conventional methods of finite element analysis. In fact, the elastic eigenmodes are successfully used by Pentland et al. [112, 113, 114] in modelling and tracking object deformations, and by Sclaroff & Pentland [130] in object registration. Their success is due to elastic eigenmodes having excellent properties as shape features, and this dissertation provides a thorough characterisation of these properties. More importantly, this dissertation also justifies the use of elastic eigenmodes in modelling organ shape variation, as discussed in the previous paragraph.

Organ shape models are initialised using eigenmodes of large spatial scale as the principal components of organ shape variation. This 'low-pass' subspace filters (noisy) high-frequency shape variations when interpolating ultrasound image landmarks, so the successful volume estimates reported in Chapter 5 are quite expected. However, there is no particular reason why the statistical principal components of organ shape variation should be of large spatial scale, although Chapter 4.4 shows that this is true if our growth model is assumed.

The issue, therefore, is whether our proposed growth model is a reasonable 'bootstrap' for use in volume estimation. If it is, then Chapter 4.5 shows how the bootstrapped organ model can be refined after each volume estimation. If it is not, then an adequate refinement may require an impractical number of iterations.

Fortunately, the important principal components of organ shape variation are generally of large spatial scale (Cootes et al. [31]; Taylor [153]). Figure 9.1 illustrates that human visual perception is intrinsically sensitive to particular large-scale variations in organ shape. Experienced sonographers and anatomists are similarly sensitive to expected modes of large-scale shape variation when visualising organ anatomy.

[1] In fact, the only important Oster-Murray mechanism omitted from our growth model is long-range cell diffusion (see Chapter 2.2). A more sophisticated growth model would include the actual distribution of material properties in the organ, which means that the eigenfunctions of deformation, diffusion and convection are no longer identical. The relative mass flux due to each mechanism would also need to be established in this case.

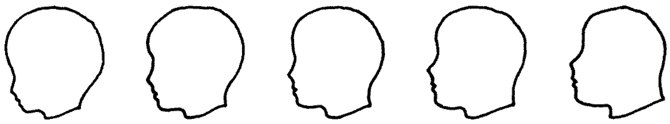


Figure 9.1: Cardiodal variation in head shape (adapted from Bogin [14])

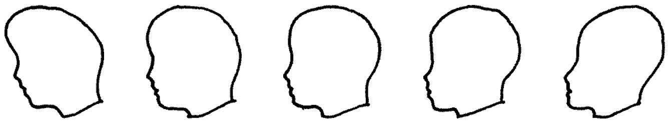
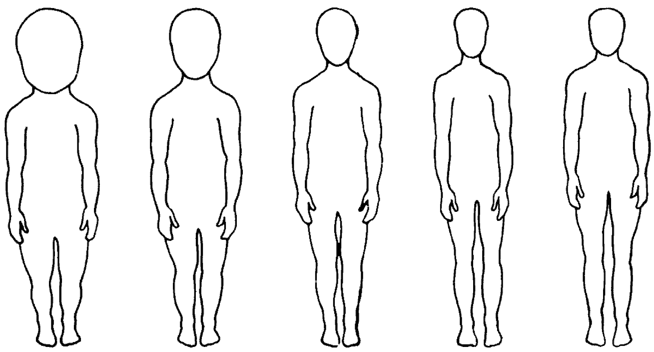


Figure 9.2: Affine variation in head shape (adapted from Bogin [14])

Figure 9.1 applies a cardiodal ('heart shaped') mode of shape variation to the sequence of head outlines, which is usually perceived as increasing in age from left to right[2]. In contrast, Figure 9.2 applies an affine shear to the head outlines; no biological significance is perceived in this component of shape variation (Bogin (ch.3)[14]).

Figures 9.1–9.3 illustrate that important modes of biological shape variation tend to be of large spatial scale. Just as human perception is intrinsically sensitive to the examples chosen for Figures 9.1–9.2, so are sonographers sensitised by prior experience to the range of expected organ shapes. It is this experience which our organ shape models aim to acquire, so as to interpret 3D ultrasound images in a biologically justified manner.



(a) Birth (b) 2 years (c) 6 years (d) 12 years (e) 25 years

Figure 9.3: Size-normalised variation in typical male body shape (adapted from Bogin [14])

Figure 9.3 shows five male outlines which, although identical in height, are still perceived as increasing in maturity from left to right. In this particular case, biological landmarks on the body (e.g. hands, shoulders, chin etc.) give sufficient *within-sample* shape information to indicate maturity, even though the absolute positions of the landmarks are not available for *between-sample* shape comparisons.

Most organs have few anatomical landmarks, however, so size normalisation is required to minimise corruption of information about organ shape variation in a dataset. The generalised centroid size metric developed in Chapter 4.1 successfully normalises organ shape examples, so that homologies between the shapes can be established using biological landmarks, to provide an accurate estimate of the modes of shape variation in a given dataset.

[2]Todd et al. [158] report that human subjects also perceive maturation in a range of bird and dog heads subjected to cardiodal transformation; it is also reported for transformed Volkswagen Beetle cars.

Volume estimation

There is scope for a number of engineering improvements to the Stradivarius project hardware (see Figure 1.4). The ultrasound video information is currently digitised at 5–10 frames per second, and at a significantly poorer image quality than the scanner's video display unit. The calibration, consistency, and accuracy of position measurements from the probe's sensor could also be improved; a mechanical arm may in fact be more appropriate in certain applications, where precision is more important than the flexibility of freehand scanning (e.g. burns assessment, needle biopsy).

3D freehand ultrasound imaging is motivated primarily as a refinement to conventional 2D ultrasound, which does not require a change in conventional scanning techniques. Its full utility is not realised in current applications, however, because the 3D images are stored for secondary analysis, which only occurs after clinical examination is undertaken in the standard way.

A major improvement in useability can be realised by implementing `stradview` as a 'real-time' application. A touch-sensitive screen to display the organ shape model next to a 'live' 3D ultrasound image sequence, would allow both to be touch-manipulated for volume estimation during clinical examination. With improved quality in the digitised images, and more sophisticated tissue detection algorithms, the organ segmentation can be further refined and analysed for biometric studies. With improved computer network capacity, volume estimation can even be performed at a remote site by specialist sonographers, i.e. "telemedicine".

Registration

Figure 9.3 serves to illustrate the importance of biological landmarks in registering organ shape models. Chapter 6.1 shows that eigenmode displacement extrema are excellent biological landmarks, and Chapter 8 develops a reliable method for organ model registration using these landmarks. As mentioned previously, our successful registrations may simply be due to the excellent shape representation properties of elastic eigenmodes. It is unclear if eigenmode extrema will still be successful landmarks if non-uniform material properties are used in the organ model, i.e. if eigenmodes are no longer the uniform elastic eigenmodes.

One weakness of eigenmodes, highlighted in Chapter 6.2, is the need for direction normalisation on which pose estimation[3] and registration of organ models both depend. Modal projection is suggested as the ideal solution, but it requires densely sampled organ models and a larger number of accurate eigenmodes. Dense sampling gives rise to very large eigenproblems which may not be soluble in a reasonable time on conventional workstations, unless powerful memory management and eigensolution methods are used.

Densely sampled models are also difficult to construct, since manual segmentation of organs (even from relatively clean MRI scans) is required. Deriving a well-distributed finite element mesh from the segmentation is largely still a heuristic exercise[4]. However the dataset of organ models is segmented and meshed, automatic registration is essential if the modes of shape variation are to be analysed from datasets large enough to be statistically significant.

One important area of future work is therefore to perform scanning, organ segmentation, and organ registration on a sufficiently large pool of patients, in order to validate the utility of automatic registration for organ model refinement. This and other areas of possible future work are now discussed.

[3] There is no reason in principle, why patient (and hence anatomical) pose cannot be made available with 3D ultrasound data.

[4] Methods of automatic mesh generation are also important in other engineering applications (Shroeder & Shephard [129]).

9.2 Future work

The work presented in this dissertation has some limitations which remain to be addressed. Two such areas are:

- Eigenmode computation
  Eigenmodes can be computed for densely sampled organ models using more advanced implementations of the Lanczos method, e.g. employing selective (Grimes et al. [59]) rather than complete reorthogonalisation (Appendix B). Provided organ models can also be accurately sampled and tessellated (see previous discussion on registration), this would allow the use of modal projection and wavelet modes as described below.

- Tissue discrimination
  Improved (multi-scale texture-based) methods of tissue discrimination in ultrasound images are very much desired (Muzzolini et al. [98]). Development of future clinical tools requires the model-based approach of this dissertation to be complemented by data-driven tissue discrimination techniques. Only then can the ultimate goal of fully-automated ultrasound image segmentation be envisaged.

Interesting issues which remain to be fully explored include:

- HCF registration
  Given sufficiently many high-quality eigenmodes, modal projection can be used in organ model registration, without the need for (unreliable) eigenmode normalisation during pose estimation. The performance and reliability of the HCF algorithm can also be improved by parallelisation (Chou et al. [28]).
  Gaussian interpolation functions (Sclaroff & Pentland [130]) may allow scattered tissue boundaries (e.g. from texture discrimination in ultrasound images) to be tentatively connected together to form an approximate organ mesh, whose mass and stiffness matrices are used to derive approximate eigenmodes. Noisy boundary fragments can in this way (potentially) be registered to an organ model, for an organ segmentation.

- Wavelet modes
  High-quality eigenmodes also allow high-quality wavelet modes. Methods of computing Gabor wavelet modes and other wavelet forms in the spatial-frequency domain should be explored, while the Gabor wavelet modes developed in this dissertation for the natural-frequency domain, can be further examined for their properties in shape representation. Our wavelet generalisation of the elastic eigenmodes may also be useful in localised contact problems, e.g. in simulating foetal delivery through the pelvis (Lapeer [78]).

- Biometric analysis
  Datasets of organ models should be registered by automated and manual means: firstly to validate automated registrations against expert anatomists, and secondly to check the variability of manual registrations between anatomists. The registered datasets can then be examined to see if the principal components of shape variation are best represented using linear or non-linear modes. The appropriateness of elastic eigenmodes in bootstrapping an organ model's modes of shape variation can thus be validated.
  Expert knowledge of organ shape variation is currently acquired in an *ad hoc* way by individual sonographers. Registration allows such knowledge to be pooled and quantified so that biometric analysis can be performed on a common knowledge base, which can then be incorporated in simulated scans for standardisation of sonographic training.

9.3 Conclusion

This dissertation introduces 3D freehand ultrasound as a flexible and affordable improvement to conventional clinical ultrasound imaging. Unfortunately, volume rendering of freehand images is of little clinical utility, due mainly to the inherent noise and artefacts in the images, and to widely and inconsistently spaced frame positions.

Morever, current methods of noise reduction and tissue discrimination are insufficiently reliable for automated segmentation of the images. Even if tissues can be reliably segmented (e.g. by an operator, or by future methods of texture analysis), the segmented information from multiple non-parallel and self-intersecting frames still needs to be geometrically reconstructed.

Two model-based strategies are successfully developed for this purpose: an organ shape model to interpolate scattered tissue boundary landmarks, and the Gompertz growth model to register organ shape models. The purpose of the latter is to allow organ shape models to correctly incorporate novel organ examples, since interpolation in the former relies on accurate knowledge of expected organ shape variation.

These strategies are validated by experiments on both volume estimation from 3D freehand ultrasound images and registration of 3D organ models. A number of innovations emerge from devising these strategies:[5]

(i) Oster-Murray mechanisms are used to model biological growth in an organ model, whose patterns of shape variation are described by the mechanisms' eigenfunctions.

(ii) A finite element sampling of the organ enables eigenfunctions to be linearly approximated by eigenmodes, which are computed using a memory efficient implementation of the state-of-the-art block Lanczos algorithm.

(iii) Displacement extrema on an organ model's eigenmodes are utilised as biological landmarks, so that organ models can be registered where these landmarks are homologous.

(iv) The Gompertz model is used to derive an energy metric which quantifies the biological plausibility of suggested homologies.

(v) Bayesian combination of (iii) and (iv) results in an overall measure of homology. Registration is performed by optimising this measure over two organ models, using an efficient multi-resolution implementation of the HCF algorithm.

In conclusion, this dissertation successfully develops a model-based framework for organising sparse and noisy cues about tissue boundaries in 3D ultrasound images. This framework is intended to complement future work on tissue boundary analysis in ultrasound images, so as to achieve fully-automated 3D ultrasound image intepretation.

[5] Computer codes implemented for assembling finite element organ models, and for computing finite element eigenmodes, are available from the author. Codes are also available for graphical visualisation and manipulation of organ models, eigenmodes, and wavelet modes.

Bibliography

[1] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86(414):376–387, 1991.

[2] E. Ashton and K. Parker. Multiple resolution Bayesian segmentation of ultrasound images. *Ultrasonic Imaging*, 17(4):291–304, 1996.

[3] N. Ayache. Medical computer vision, virtual reality and robotics. In *Proc. British Machine Vision Conference*, pages 1–25, Birmingham, 1995. BMVA Press.

[4] F. Balen, C. Allen, J. Gardener, N. Siddle, and W. Lees. 3-dimensional reconstruction of ultrasound images of the uterine cavity. *British Journal of Radiology*, 66(787):588–591, 1993.

[5] D. Ballard and C. Brown. *Computer Vision*. Prentice-Hall, 1982.

[6] J. Bamber, R. Eckersley, P. Hubregtse, N. Bush, D. Bell, and D. Crawford. Data processing for 3D ultrasound visualization of tumour anatomy and blood flow. In R. Robb, editor, *Visualization in Biomedical Computing*, volume 1808 of *Proc. SPIE*, pages 651–663, 1992.

[7] O. Basset, Z. Sun, J. Mestas, and G. Gimenez. Texture analysis of ultrasonic images of the prostate by means of cooccurence matrices. *Ultrasonic Imaging*, 15(3):267–285, 1993.

[8] K-J. Bathe. *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, 1982.

[9] M. Belohlavek, D. Foley, J. Seward, and J. Greenleaf. 3D echocardiography: reconstruction algorithm and diagnostic performance of resulting images. In R. Robb, editor, *Visualization in Biomedical Computing*, volume 2359 of *Proc. SPIE*, pages 170–179, 1994.

[10] F. Betting and J. Feldmar. 3D-2D projective registration of anatomical surfaces with their projections. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, Computational Imaging and Vision, pages 113–125. Kluwer Academic, 1995.

[11] A. Blake and A. Yuille, editors. *Active Vision*. MIT Press, 1992.

[12] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[13] H. Blum. Biological shape and visual science. *Journal of Theoretical Biology*, 38:205–287, 1973.

[14] B. Bogin. *Patterns of Human Growth*. Cambridge University Press, 1988.

[15] J-D. Boissonat and B. Geiger. Three-dimensional reconstruction of complex shapes based on the Delaunay triangulation. Technical Report 1697, INRIA, BP 93–06902 Sophia Antipolis Cedex (France), April 1992.

[16] F. Bookstein. *Morphometric Tools for Landmark Data.* Cambridge University Press, 1st edition, 1991.

[17] F. Bookstein and W. Green. Edge information at landmarks in medical images. In R. Robb, editor, *Visualization in Biomedical Computing*, volume 1808 of *Proc. SPIE*, pages 242–258, 1992.

[18] F. Bookstein and W. Green. Hinting about causes of deformation: visual explorations of a novel inverse problem in medical imaging. In F. Bookstein, J. Duncan, N. Lange, and D. Wilson, editors, *Mathematical Methods in Medical Imaging III*, volume 2299 of *Proc. SPIE*, pages 2–15, 1994.

[19] M. Brady and G. Scott. Parallel algorithms for shape representation. In I. Page, editor, *Parallel Architectures and Computer Vision*, pages 97–118. Clarendon Press, 1988.

[20] V. Bruce and P. Green. *Visual Perception.* Lawrence Erlbaum Associates, 2nd edition, 1990.

[21] J. Candy. *Signal Processing.* McGraw-Hill, 1988.

[22] D. Causton. *A Biologist's Mathematics.* Contemporary Biology. Edward Arnold, 1977.

[23] B. Cena, B. Backman, K. Booth, and F. Wood. Computerised integration of multimodality imaging in scar assessment. In *Proc. Australia and New Zealand Burns Assiciation Conference*, August 1995.

[24] T-J. Cham. Visual symmetry: towards a role in perceptual organisation. First year PhD report, June 1994. Dept. of Engineering, University of Cambridge.

[25] F. Chervenak, G. Isaacson, and S. Campbell, editors. *Ultrasound in Obstetrics and Gynecology.* Little, Brown & Co., 1993.

[26] Z-H. Cho, J. Jones, and M. Singh. *Foundations of Medical Imaging.* John Wiley & Sons, 1993.

[27] P-B. Chou and C. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

[28] P-B. Chou, P. Cooper, M. Swain, C. Brown, and L. Wixson. Probabilistic network inference for cooperative high and low level vision. In *Markov Random Fields: Theory and Application*, pages 211–243. Academic Press, 1993.

[29] G. Christensen, R. Rabbitt, M. Miller, S. Joshi, U. Grenander, T. Coogan, and D. Van Essen. Topological properties of smooth anatomic maps. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, Computational Imaging and Vision, pages 101–112. Kluwer Academic, 1995.

[30] L. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2D and 3D images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[31] T. Cootes, A. Hill, C. Taylor, and J. Haslam. The use of active shape models for locating structures in medical iamges. In *Lecture Notes in Computer Science*, volume 687, pages 33–47. Springer-Verlag, 1993.

[32] T. Cootes and C. Taylor. Combining point distribution models with shape models based on finite-element analysis. In *Proc. British Machine Vision Conference*, pages 419–428, York, 1994. BMVA Press.

[33] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[34] G. Coppini, R. Poli, and G. Valli. Recovery of the 3D shape of the left ventricle from echocardiographic images. *IEEE Trans. Medical Imaging*, 14(2):301–317, 1995.

[35] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience, 1st English edition, 1953.

[36] C. Dance. *Segmentation and reconstruction in 3D freehand ultrasound imaging.* PhD thesis, Dept. of Engineering, University of Cambridge, 1996.

[37] C. Dance, R. Prager, and L. Berman. The Stradivarius project final report, 1996. Dept. of Engineering, University of Cambridge.

[38] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal Optical Society of America A*, 2(7):1160–1169, 1985.

[39] P. Drazin. *Non-linear Systems.* Cambridge University Press, 1992.

[40] D. Fine, S. Perring, J. Herbetko, C. Hacking, J. Fleming, and K. Dewbury. Three-dimensional (3D) ultrasound imaging of the gallbladder and dilated biliary tree: reconstruction from real-time B-scans. *British Journal of Radiology*, 64:1056–1057, 1991.

[41] M. Fischler and O. Firschein. *Intelligence: the Eye, the Brain, and the Computer.* Addison-Wesley, 1987.

[42] H. Fuchs, M. Levoy, and S. Pizer. Interactive visualization of 3D medical data. *Computer*, 22(8):46–51, August 1989.

[43] S. Fujita. Synthetico-analytic approach to mechanism of shaping of the brain through simulation with 3D computer graphics. In S. Ishizaka, editor, *Science on Form: Proceedings of the First International Symposium for Science on Form*, pages 339–352, Tokyo, 1986. KTK Scientific Publishers.

[44] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.

[45] J. Gee, L. Le Briquer, and C. Barillot. Probabilistic matching of brain images. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, Computational Imaging and Vision, pages 113–125. Kluwer Academic, 1995.

[46] M. Gell-Mann. *The Quark and the Jaguar.* W.H. Freeman, 1994.

[47] The Geometry Center (`software@geom.umn.edu`), University of Minnesota, 1300 South Second Street, Suite 500, Minneapolis, MN 55454. *Geomview/OOGL Release 1.5*, 1994.

[48] A. Gierer. Hydra as a model for the development of biological form. *Scientific American*, December 1974.

[49] O. Gilja, A. Smievoll, N. Thune, K. Matre, T. Hausken, S. Odegaard, and A. Berstad. In vivo comparison of 3D ultrasonography and magnetic resonance imaging in volume estimation of human kidneys. *Ultrasound in Medicine and Biology*, 21(1):25–32, 1995.

[50] G. Golub, R. Underwood, and J. Wilkinson. The Lanczos algorithm for the symmetric $Ax = \lambda Bx$ problem. Technical Report STAN-CS-72-270, Computer Science Dept., Stanford University, 1972.

[51] G. Golub and C. van Loan. *Matrix Computations.* Johns Hopkins University Press, 2nd edition, 1989.

[52] B. Gompertz. On the nature of function expressive of the law of human mortality, and on a new method of determining the value of life contingencies. *Philosophical Transactions of the Royal Society*, 115:513–585, 1825.

[53] B. Goodwin, S. Kauffman, and J. Murray. Is morphogenesis an intrinsically robust process? *Journal of Theoretical Biology*, 163:135–144, 1993.

[54] C. Gordon, D. Webb, and S. Wolpert. You cannot hear the shape of a drum. *Bulletin of the American Mathematical Society*, 27(1):134–138, 1992.

[55] J. Gosling, D. Lomas, R. Prager, and L. Berman. Virtual reality technology in the assessment of femoral neck anteversion. In *Proc. Radiological Society of North America*, 1993.

[56] J. Gosling, R. Prager, and L. Berman. Proprioception accuracy in free-hand three-dimensional ultrasound imaging. Technical Report CUED/F-INFENG/TR 169, Dept. of Engineering, University of Cambridge, 1994.

[57] J. Gosling, M. Syn, R. Prager, and L. Berman. Estimation for thyroid volumes using three-dimensional free-hand ultrasound. In *Proc. 12th Biosignal*, Brno, Czech Republic, June 1994.

[58] U. Grenander, Y. Chow, and D. Keenan. *Hands. A Pattern Theoretic Study of Biological Shapes.* Springer-Verlag, 1991.

[59] R. Grimes, J. Lewis, and H. Simon. A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. *SIAM Journal on Matrix Analysis and Applications*, 15(1):228–272, January 1994.

[60] R. Harris, D. Follett, M. Halliwell, and P. Wells. Ultimate limits in ultrasonic imaging resolution. *Ultrasound in Medicine and Biology*, 17(6):547–558, 1991.

[61] D. Hill, D. Hawkes, M. Gleeson, T. Cox, A. Strong, W-L. Wong, C. Ruff, N. Kitchen, D. Thomas, A. Sofat, J. Crossman, C. Studholme, A. Gandhe, S. Green, and G. Robinson. Accurate frameless registration of MR and CT images of the head: applications in planning surgery and radiotherapy. *Radiology*, 191:447–454, 1994.

[62] D. Hitchens, editor. *A Finite Element Dynamics Primer.* NAFEMS, 1992.

[63] T. Hodges, P. Detmer, D. Burns, K. Beach, and D. Strandness. Ultrasonic three-dimensional reconstruction: in vitro and in vivo volume and area measurement. *Ultrasound in Medicine and Biology*, 20(8):719–729, 1994.

[64] M-K. Hu. Visual pattern recognition by moment invariants. *IEEE Trans. Information Theory*, 8:179–187, 1962.

[65] S. Hughes, T. D'Arcy, D. Maxwell, W. Chiu, A. Milner, J. Saunders, and R. Sheppard. Volume estimation from multiplanar 2D ultrasound images using a remote electromagnetic position and orientation sensor. *Submitted to Ultrasound in Medicine and Biology*, 1995.

[66] T. Hughes. *The Finite Element Method.* Prentice-Hall, 1987.

[67] A. Hunding and R. Engelhardt. Early biological morphogenesis and non-linear dynamics. *Journal of Theoretical Biology*, 173:401–413, 1995.

[68] J. Huxley. *Problems of Relative Growth.* Methuen, 1932.

[69] M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, 1966.

[70] M. Karaman, M. Kutay, and G. Bozdagi. An adaptive speckle suppresion filter for medical ultrasonic imaging. *IEEE Trans. Medical Imaging*, 14(2), 1995.

[71] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. In *Proc. 1st International Conference on Computer Vision*, pages 259–268, 1987.

[72] B. Kimia, A. Tannenbaum, and S. Zucker. Toward a computational theory of shape: an overview. In *Lecture Notes in Computer Science*, volume 427, pages 402–407. Springer-Verlag, 1990.

[73] B. Kimia, A. Tannenbaum, and S. Zucker. On the evolution of curves via a function of curvature. *Journal of Mathematical Analysis and Applications*, 163:438–458, 1992.

[74] H-M. Klein, R. Gunther, M. Verlande, W. Schneider, D. Vorwerk, J. Kelch, and M. Hamm. 3D surface reconstruction of intravascular ultrasound images using personal computer hardware and a motorized catheter control. *Cardiovascular and Interventional Radiology*, 15:97–101, 1992.

[75] J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[76] J. Koenderink and A. van Doorn. Receptive field families. *Biological Cybernetics*, 63:291–297, 1990.

[77] E. Kreysig. *Advanced Engineering Mathematics.* John Wiley & Sons, 6th edition, 1988.

[78] R. Lapeer. 3D simulation of a childbirth process using a finite element model. First year PhD report, June 1996. Dept. of Engineering, University of Cambridge.

[79] R. Lestienne. On the thermodynamical and biological interpretation of the Gompertzian mortality rate distribution. *Mechanisms of Ageing and Decelopment*, 42:197–214, 1988.

[80] S. Levin and L. Segel. Pattern generation in space and aspect. *SIAM Review*, 27(1):45–67, March 1985.

[81] R. Lewin. *Complexity: Life at the Edge of Chaos.* Dent, 1993.

[82] K-G. Lim. *Area based stereo.* PhD thesis, Dept. of Engineering, University of Cambridge, 1994.

[83] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[84] Y-B. Ling and B. He. Entropic analysis of biological growth models. *IEEE Trans. Biomedical Engineering*, 40(12):1193–1200, 1993.

[85] R. Luecke, W. Wosilait, and J. Young. Mathematical representation of organ growth in the human embryo/fetus. *International Journal of Bio-Medical Computing*, 39:337–347, 1995.

[86] D. Marr. *Vision.* W.H. Freeman, 1982.

[87] D. Marr and H. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. Royal Society of London B*, 200:269–294, 1978.

[88] J. Martin, A. Pentland, and R. Kikinis. Shape analysis of brain structures using physical and experimental modes. Technical Report 276, M.I.T. Media Laboratory, Perceptual Computing, January 1994.

[89] P. Medawar. The growth, growth energy, and ageing of the chicken's heart. *Philosophical Transactions of the Royal Society B*, 129:332–355, 1940.

[90] H. Meinhardt. *Models of Biological Pattern Formation.* Academic Press, 1982.

[91] H. Meinhardt and M. Klingler. A model for pattern formation on the shells of molluscs. *Journal of Theoretical Biology*, 126:63–89, 1987.

[92] E. Merz, F. Bahlmann, G. Weber, and D. Macchiella. 3-dimensional ultrasonography in prenatal-diagnosis. *Journal of Perinatal Medicine*, 23(3):213–222, 1995.

[93] S. Metcalfe and J. Evans. Optimization of ultrasound scanner characteristics: a preliminary study. *British Journal of Radiology*, 66:609–613, 1993.

[94] K. Moore. *The Developing Human.* W.B. Saunders, 3rd edition, 1982.

[95] A. Moskalik, P. Carson, C. Meyer, J. Fowlkes, J. Rubin, and M. Roubidoux. Registration of three-dimensional compound ultrasound scans of the breast for refraction and motion correction. *Ultrasound in Medicine and Biology*, 21(6):769–778, 1995.

[96] J. Murray. *Mathematical Biology*, volume 19 of *Biomathematics Texts.* Springer-Verlag, 1989.

[97] J. Murray, P. Maini, and R. Tranquillo. Mechano-chemical models for generating biological pattern and form in development. *Physics Reports*, 171(2):59–84, 1988.

[98] R. Muzzolini, Y-H. Yang, and R. Pierson. Multiresolution texture segmentation with application to diagnostic ultrasound images. *IEEE Trans. Medical Imaging*, 12(1):108–123, 1993.

[99] R. Muzzolini, Y-H. Yang, and R. Pierson. Texture characterisation using robust statistics. *Pattern Recognition*, 27(1):119–134, 1994.

[100] C. Nastar. Analytical computation of the free vibration modes: application to non-rigid motion analysis and animation in 3D images. Technical Report 1935, INRIA-Rocquencourt, 1993.

[101] C. Nastar and N. Ayache. Non-rigid motion analysis in medical images: a physically based approach. In *Lecture Notes in Computer Science*, volume 687, pages 17–32. Springer-Verlag, 1993.

[102] T. Nelson, T. Davidson, and D. Pretorius. Interactive electronic scalpel for extraction of organs from 3D US data. *Radiology*, 197(P):191, November 1995.

[103] T. Nelson and T. Elvins. Visualization of 3D ultrasound data. *IEEE Computer Graphics and Applications*, November 1993.

[104] M. Neveu, D. Faudot, and B. Derdouri. Recovery of 3D deformable models from echocardiographic images. In F. Bookstein, J. Duncan, N. Lange, and D. Wilson, editors, *Mathematical Methods in Medical Imaging III*, volume 2299 of *Proc. SPIE*, pages 367–376, 1994.

[105] K. Ng, J. Gardener, D. Rickards, W. Lees, and E. Milroy. Three-dimensional imaging of the prostatic urethra – an exciting new tool. *British Journal of Urology*, 74(5):604–608, 1994.

[106] R. Ohbuchi, D. Chen, and H. Fuchs. Incremental volume reconstruction and rendering for 3D ultrasound imaging. In R. Robb, editor, *Visualization in Biomedical Computing*, volume 1808 of *Proc. SPIE*, pages 312–323, 1992.

[107] G. Oster and J. Murray. Cell traction models for generating pattern and form in morphogenesis. *Journal of Mathematical Biology*, 19(3):265–279, 1984.

[108] J. Ousterhout. *Tcl and the Tk Toolkit*. Addison-Wesley, 1994.

[109] B. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, 1980.

[110] C. Pasapula. 3D shape analysis using point based distribution models on fetal cadaveric livers. Internal report, Guy's and St. Thomas' Hospital, April 1995.

[111] G. Pasterkamp, C. Borst, A-F. Moulaert, C. Bouma, D. van Dijk, M. Kluytmans, and B. ter Haar Romeny. Intravascular ultrasound image subtraction: a contrast enhancing technique to facilitate automatic three-dimensional visualization of the arterial lumen. *Ultrasound in Medicine and Biology*, 21(7):913–918, 1995.

[112] A. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, 4:107–126, 1990.

[113] A. Pentland and B. Horowitz. Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):730–742, 1991.

[114] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):715–729, 1991.

[115] S. Pizer and C. Burbeck. Cores as the basis for object vision in medical images. In *Image Perception*, volume 2166 of *Proc. SPIE*, pages 191–198, 1994.

[116] S. Pizer, W. Oliver, and S. Bloomberg. Hierarchical shape description via the multi-resolution symmetric axis transform. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:505–511, 1987.

[117] P. Prusinkiewicz. Visual models of morphogenesis. *Artificial Life*, 1:67–74, 1994.

[118] E. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45(1):3–11, January 1977.

[119] V. Quercia and T. O'Reilly. *X Windows System User's Guide*. O'Reilly & Associates, 1993.

[120] A. Rahmouni, A. Yang, C. Tempany, T. Frenkel, J. Epstein, P. Walsh, P. Leichner, C. Ricci, and E. Zerhouni. Accuracy of in vivo assessment of prostatic volume by MRI and transrectal ultrasonography. *Journal of Computer Assisted Tomography*, 16(6):935–940, 1992.

[121] R. Rankin, A. Fenster, D. Downey, P. Munk, M. Levin, and A. Vellet. Three-dimensional sonographic reconstruction: techniques and diagnostic applications. *American Journal of Roentgenology*, 161(4):695–702, 1993.

[122] R. Rohling and A. Gee. Issues in 3D free-hand medical ultrasound imaging. Technical Report CUED/F-INFENG/TR 246, Dept. of Engineering, University of Cambridge, January 1996.

[123] H. Rutishauser. Simultaneous iteration method for symmetric matrices. In J. Wilkinson and C. Reinsch, editors, *Handbook for Automatic Computation (Linear Algebra)*, pages 202–211. Springer-Verlag, 1971.

[124] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* Manchester University Press, 1992.

[125] G. Sakas, L-A. Schreyer, and M. Grimm. Preprocessing and volume rendering of 3D ultrasonic data. *IEEE Computer Graphics and Applications*, July 1995.

[126] G. Sakas and S. Walter. Extracting surfaces from fuzzy 3D-ultrasound data. In *Computer Graphics*, Proc. SIGGRAPH. ACM, August 1995.

[127] A. Salustri and J. Roelandt. Ultrasonic three-dimensional reconstruction of the heart. *Ultrasound in Medicine and Biology*, 21(3):281–293, 1995.

[128] D. Savic. Model of pattern formation in animal coatings. *Journal of Theoretical Biology*, 172:299–303, 1995.

[129] W. Schroeder and M. Shephard. Geometry-based fully automatic mesh generation and the Delaunay triangulation. *International Journal for Numerical Methods in Engineering*, 26:2503–2515, 1988.

[130] S. Sclaroff and A. Pentland. Modal matching for correspondence and recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.

[131] M. Scott. *Statistical texture analysis of ultrasound images.* PhD thesis, Dept. of Engineering, University of Cambridge, 1996.

[132] S. Setio, H. Setio, and L. Jezequel. Modal analysis of nonlinear multi-degree-of-freedom structures. *The International Journal of Analytical and Experimental Modal Analysis*, 7(2):75–93, April 1992.

[133] P. Shankar. Speckle reduction in ultrasound B-scans using weighted averaging in spatial compounding. *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, 33(6):754–758, 1986.

[134] R. Shymko and L. Glass. Cellular and geometric control of tissue growth and mitotic instability. *Journal of Theoretical Biology*, 63:355–374, 1976.

[135] C. Sohn. Challenges remain in 3D ob/gyn ultrasound. *Diagnostic Imaging*, November 1991.

[136] F. Solina and R. Bajcsy. Recovery of parametric models from range images: the case for superquadrics with global deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):131–146, 1990.

[137] P. Sozou, T. Cootes, C. Taylor, and E. Di Mauro. A non-linear generalisation of PDMs using polynomial regression. In *Proc. British Machine Vision Conference*, pages 397–406, 1994.

[138] P. Sozou, T. Cootes, C. Taylor, and E. Di Mauro. Non-linear point distribution modelling using a multi-layer perceptron. In *Proc. British Machine Vision Conference*, pages 107–116, 1995.

[139] L. Staib and J. Duncan. Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, 1992.

[140] E. Steen and B. Oldstad. Volume rendering of 3D medical ultrasound data using direct feature mapping. *IEEE Trans. Medical Imaging*, 13(3):517–525, 1994.

[141] H. Steiner, A. Staudach, D. Spitzer, and H. Schaffer. Three-dimensional ultrasound in obstetrics and gynaecology: technique, possibilities and limitations. *Human Reproduction*, 9(9):1773–1778, 1994.

[142] M. Syn, J. Gosling, R. Prager, L. Berman, and J. Crowley. Tracking the inter-frame deformation of structures in 3D ultrasound imaging. In R. Robb, editor, *Visualization in Biomedical Computing*, volume 2359 of *Proc. SPIE*, pages 170–179, 1994.

[143] M. Syn and R. Prager. Mesh models for three-dimensional ultrasound imaging. Technical Report CUED/F-INFENG/TR 210, Dept. of Engineering, University of Cambridge, December 1994.

[144] M. Syn and R. Prager. Model-based three-dimensional ultrasound imaging. Technical Report CUED/F-INFENG/TR 180, Dept. of Engineering, University of Cambridge, October 1994.

[145] M. Syn and R. Prager. Bayesian registration of models using FEM eigenmodes. Technical Report CUED/F-INFENG/TR 213, Dept. of Engineering, University of Cambridge, September 1995.

[146] M. Syn and R. Prager. A biological growth metric for 3D shape registration. Technical Report CUED/F-INFENG/TR 225, Dept. of Engineering, University of Cambridge, July 1995.

[147] M. Syn and R. Prager. FEM eigenmodes as shape features. Technical Report CUED/F-INFENG/TR 211, Dept. of Engineering, University of Cambridge, May 1995.

[148] M. Syn and R. Prager. A model-based approach to 3D freehand ultrasound imaging. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*, Computational Imaging and Vision, pages 361–362. Kluwer Academic, 1995.

[149] M. Syn, R. Prager, L. Berman, J. Gosling, and D. Lomas. Shape registration in 3D freehand ultrasound: eigenmodes as shape features. *Radiology*, 197(P):192, November 1995.

[150] M. Syn, R. Prager, L. Berman, J. Gosling, and D. Lomas. Shape segmentation in 3D freehand ultrasound: an iterative algorithm for defining tissue boundaries. *Radiology*, 197(P):162, November 1995.

[151] M. Syn, R. Prager, L. Berman, and D. Lomas. Model-based 3D freehand ultrasound imaging. *Submitted to Radiology*, 1996.

[152] W. Szemplinkska-Stupnicka. Non-linear normal modes and the generalized Ritz method in the problems of vibrations of non-linear elastic continuous systems. *International Journal of Non-Linear Mechanics*, 18(2):149–165, 1983.

[153] C. Taylor. Personal communication, June 1995.

[154] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: deformable superquadrics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):703–714, 1991.

[155] C. Therrien. *Decision Estimation and Classification*. Wiley, 1989.

[156] D. Thompson. *On Growth and Form*. Cambridge University Press, 1961.

[157] S. Timoshenko and J. Goodier. *Theory of Elasticity*. Engineering Societies Monographs. McGraw-Hill, 3rd edition, 1970.

[158] J. Todd, L. Mark, R. Shaw, and J. Pittenger. The perception of human growth. *Scientific American*, 242(2):132–44, 1980.

[159] J. Trinkhaus. *Cells into Organs*. Prentice-Hall Inc., 2nd edition, 1984.

[160] J. Trobaugh, D. Trobaugh, and W. Richard. Three-dimensional imaging with stereotactic ultrasonography. *Computerized Medical Imaging and Graphics*, 18(5):315–323, 1994.

[161] A. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B*, 237:37–72, 1952.

[162] R. Underwood. An iterative block Lanczos method for the solution of large sparse symmetric eigenproblems. Technical Report STAN-CS-75-496, Computer Science Dept., Stanford University, 1975.

[163] V. Vaidya and F. Alexandro. Evaluation of some mathematical models for tumor growth. *International Journal of Biomedical Computing*, 13:19–36, 1982.

[164] R. van den Boomgaard and A. Smeulders. The morphological structure of images: the differential equations of morphological scale-space. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(11):1101–1113, 1994.

[165] P. van Laarhoven and E. Aarts. *Simulated Annealing: Theory and Applications*. D. Reidel Publishing Company, 1987.

[166] P. Wells. *Advances in Ultrasound Techniques and Instrumentation*. Churchill Livingstone Inc., 1993.

[167] L. Wolpert. Pattern formation in biological development. *Scientific American*, 239(4):124–138, October 1978.

[168] F. Wood, K. Currie, B. Backman, and B. Cena. The current difficulties and the possible future directions in scar assessment. In *Proc. International Symposium for Hypertrophic Scar*, Hong Kong, June 1995.

[169] C-M. Wu, Y-C. Chen, and K-S. Hsieh. Texture features for classification of ultrasonic liver images. *IEEE Trans. Medical Imaging*, 11(2):141–152, 1992.

[170] X-L. Xu. The biological foundation of the Gompertz model. *International Journal of Biomedical Computing*, 20:35–39, 1987.

[171] X-L. Xu and Y-B. Ling. A study on the expectational model for tumor growth. *International Journal of Biomedical Computing*, 22:135–141, 1988.

[172] D. Young. A local activator-inhibitor model of vertebrate skin patterns. *Mathematical Biosciences*, 72:51–58, 1984.

[173] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.

Appendix A

Finite Element Method

The *finite element method* (FEM) is a numerical technique for solving partial differential equations. It partitions the continuous solution domain into a finite number of *elements*, so that solutions are sampled at element nodes, then propagated within each element using an *interpolation function*.

This appendix applies the FEM to the structural equilibrium of elastic bodies (Bathe [8]; Hughes [66]); computer codes implemented from this appendix are available from the author. Variables used are mostly consistent with Bathe:

x, y, z	global coordinate system
r, s, t	local coordinate system within element
$\mathbf{u}$	displacement components for nodes of whole body
$\mathbf{u}_i^{\mathrm{n}}$	displacement components for body node i
$\mathbf{u}_j^{\mathrm{e}}(r, s, t)$	displacement function for element j
u_x, u_y, u_z	displacement components for element j
$\mathbf{u}_j^{\mathrm{e}}$	displacement components for nodes of element j

Appendix organisation

- **Section A.1** considers the *principle of virtual work*, which embodies the equilibrium between external forces acting on an elastic body, and its internal stresses and strains. The equilibrium is interpolated over a finite element model of the elastic body.
- **Section A.2** shows the equivalence between this continuous equilibrium, and the discretely sampled equilibrium of a finite element body.
- **Section A.3** uses this equivalence to derive an expression for each element's *stiffness matrix*.
- **Section A.4** does likewise for each element's *mass matrix*.
- **Section A.5** shows how the integral expressions for mass and stiffness matrices can be numerically evaluated.
- **Section A.6** derives linear interpolation functions for the tetrahedral elements used for all FEM models in this dissertation.
- **Section A.7** shows that a (scalar) pressure distribution accompanies any (vector) elastic deformation.
- **Section A.8** derives an expression for the energy of elastic deformation.

A.1 Equilibrium of elastic body

Consider the equilibrium of a body with the following forces acting upon it:

$\mathbf{r}^{\mathrm{V}}$	volume forces (per unit volume)
$\mathbf{r}^{\mathrm{S}}$	surface forces (per unit surface area)
$\mathbf{r}^{\mathrm{n}}$	point forces acting on the body

The displacements of the body from its unloaded state, due to the action of these forces, are described by the vector $\mathbf{u}$. The corresponding plane and shear components of *strain* are

$$\boldsymbol{\epsilon} = (\ \epsilon_{xx} \quad \epsilon_{yy} \quad \epsilon_{zz} \quad \gamma_{xy} \quad \gamma_{yz} \quad \gamma_{zx} \ )^T \tag{A.1}$$

and the *stress* components are

$$\boldsymbol{\tau} = (\ \tau_{xx} \quad \tau_{yy} \quad \tau_{zz} \quad \tau_{xy} \quad \tau_{yz} \quad \tau_{zx} \ )^T \tag{A.2}$$

Strain is defined as the ratio of displacement to original length in the direction of displacement, and stress is the force experienced per unit area.

Principle of virtual work

Analysis of elastic body response requires the wave equation (2.3), which describes elastic equilibrium subject to appropriate boundary and compatibility conditions. An equivalent but more straightforward differential formulation employs the *principle of virtual work* (Bathe (ch.4.2)[8]). This states that a body in equilibrium requires the total internal virtual work done to equal the total external virtual work done

$$\int_V \overline{\boldsymbol{\epsilon}}^T \boldsymbol{\tau} \, dV = \int_V \overline{\mathbf{u}}^T \mathbf{r}^{\mathrm{V}} \, dV + \int_S \overline{\mathbf{u}}^T \mathbf{r}^{\mathrm{S}} \, dS + \sum_i \overline{\mathbf{u}}^T \mathbf{r}_i^{\mathrm{n}} \tag{A.3}$$

where the overline signifies a virtual strain or displacement. For equation (A.3) to hold, the displacements must meet the following conditions:

(i) they should be compatible and continuous between elements;

(ii) they must satisfy the displacement boundary conditions;

(iii) they must satisfy constitutive relationships (i.e. stresses can be evaluated from strains).

Finite element formulation

A finite element discretisation of the body requires a system of interpolation functions. These functions are collected in a *displacement interpolation matrix* $\mathbf{H}$, which relates displacements within the jth element $\mathbf{u}_j^{\mathrm{e}}(r,s,t)$ to the displacements of its bounding nodes $\mathbf{u}_j^{\mathrm{e}}$

$$\mathbf{u}_j^{\mathrm{e}}(r,s,t) = \mathbf{H}_j(r,s,t)\mathbf{u}_j^{\mathrm{e}} \tag{A.4}$$

Strains can similarly be interpolated using a *strain-displacement interpolation matrix* $\mathbf{B}$

$$\boldsymbol{\epsilon}_j(r,s,t) = \mathbf{B}_j(r,s,t)\mathbf{u}_j^{\mathrm{e}} \tag{A.5}$$

where

$$\mathbf{B}(r,s,t) = \begin{pmatrix} \frac{\partial}{\partial x} & & \\ & \frac{\partial}{\partial y} & \\ & & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \\ & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & & \frac{\partial}{\partial z} \end{pmatrix} \mathbf{H}(r,s,t) \tag{A.6}$$

The stress-strain relationship for the jth element (with an initial stress $\boldsymbol{\tau}_j^0$) is

$$\boldsymbol{\tau}_j = \mathbf{E}_j \boldsymbol{\epsilon}_j + \boldsymbol{\tau}_j^0 \tag{A.7}$$

where $\mathbf{E}$ is the *elasticity matrix* relating local stresses to resultant local strains. Over a volumetric element

$$\mathbf{E} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1 & \frac{\nu}{1-\nu} & \frac{\nu}{1-\nu} & & & \\ \frac{\nu}{1-\nu} & 1 & \frac{\nu}{1-\nu} & & & \\ \frac{\nu}{1-\nu} & \frac{\nu}{1-\nu} & 1 & & & \\ & & & \frac{1-2\nu}{2(1-\nu)} & & \\ & & & & \frac{1-2\nu}{2(1-\nu)} & \\ & & & & & \frac{1-2\nu}{2(1-\nu)} \end{pmatrix} \tag{A.8}$$

where E is the *Young's modulus*, and ν is the *Poisson ratio*. Summing for all N elements of the body, equation (A.3) can be rewritten in finite element form[1]

$$\begin{aligned} \sum_{j=1}^{N} \int_{V_j} \bar{\boldsymbol{\epsilon}}_j^T \boldsymbol{\tau}_j \, dV_j \quad &= \quad \sum_{j=1}^{N} \int_{V_j} \left(\mathbf{H}_j \bar{\mathbf{u}}_j^{\mathrm{e}}\right)^T \mathbf{r}_j^{\mathrm{V}} \, dV_j + \sum_{j=1}^{N} \int_{S_j} \left(\mathbf{H}_j \bar{\mathbf{u}}_j^{\mathrm{e}}\right)^T \mathbf{r}_j^{\mathrm{S}} \, dS_j \\ &\quad + \sum_i \left(\bar{\mathbf{u}}_i^{\mathrm{n}}\right)^T \mathbf{r}_i^{\mathrm{n}} \end{aligned} \tag{A.9}$$

Elastic potential at equilibrium

Assuming a linear elastic continuum and applying the stress-strain equation (A.7), gives the the total elastic potential of the body (Bathe (ch.4.2)[8])

$$W^{\mathrm{el}} = \frac{1}{2} \int_V \boldsymbol{\epsilon}^T \mathbf{E} \boldsymbol{\epsilon} \, dV - \int_V \mathbf{u}^T \mathbf{r}^{\mathrm{V}} \, dV - \int_S \mathbf{u}^T \mathbf{r}^{\mathrm{S}} \, dS - \sum_i \mathbf{u}^T \mathbf{r}_i^{\mathrm{n}} \tag{A.10}$$

For stationary potential $\delta W^{\mathrm{el}} = 0$, this becomes

$$\int_V \delta(\boldsymbol{\epsilon}^T \mathbf{E} \boldsymbol{\epsilon}) \, dV = \int_V \delta(\mathbf{u}^T \mathbf{r}^{\mathrm{V}}) \, dV + \int_S \delta(\mathbf{u}^T \mathbf{r}^{\mathrm{S}}) \, dS + \sum_i \delta(\mathbf{u}^T \mathbf{r}_i^{\mathrm{n}}) \tag{A.11}$$

Providing displacements satisfy boundary conditions, so will the corresponding strains in equation (A.11). Stationarity of W^{el} can now be seen to be equivalent to equation (A.3), with $\delta\boldsymbol{\epsilon} \equiv \bar{\boldsymbol{\epsilon}}$ and $\delta\mathbf{u} \equiv \bar{\mathbf{u}}$. The principle of virtual work therefore implies an equilbrium state where net elastic potential is stationary (minimised).

[1] Note that there are different $\mathbf{H}$ for surface and volume interpolations.

A.2 Equilibrium of discrete elastic system

The *discrete* FEM equilibrium equation for an undamped and unforced structure

$$\mathbf{M\ddot{u}} = -\mathbf{Ku} \tag{A.12}$$

differs from the *continuous* wave equation

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = -D^{\mathrm{el}}\nabla(\nabla\cdot\mathbf{u}) \tag{A.13}$$

in that equation (A.12) has the right-hand term $\mathbf{u}$, where equation (A.13) has $\nabla(\nabla\cdot\mathbf{u})$.

The limiting behaviour of a one-dimensional finite element system is examined in Figure A.1, as the elements decrease in size[2]. Nodes are connected elastically to each other at natural length d, and elements have mass density ρ.

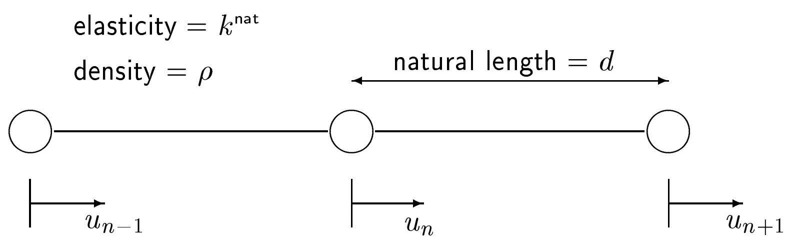


Figure A.1: Discrete one-dimensional finite element system

Consider the displacements u_{n-1}, u_n and u_{n+1} of three adjacent elements. The mass belonging to each node is ρd, with half the elemental contribution on either side. Then for natural elasticity k^{nat} the equilibrium equation is

$$\rho d\ddot{u}_n = \frac{k^{\mathrm{nat}}(u_{n-1} + u_{n+1} - 2u_n)}{d} \tag{A.14}$$

In the limit as $d \mapsto 0$ this gives

$$\rho\ddot{u} = k^{\mathrm{nat}}\nabla^2 u \tag{A.15}$$

which is the one-dimensional version of the wave equation (A.13), with

$$D^{\mathrm{el}} = \frac{k^{\mathrm{nat}}}{\rho} \tag{A.16}$$

Generalising this result to three dimensions allows equation (A.9), which describes the elastic equilibrium of finite elements, to be manipulated to look like equation (A.12). This similarity in form is exploited in Sections A.3–A.4, so as to derive expressions for the mass $\mathbf{M}$ and stiffness $\mathbf{K}$ matrices.

[2]It is simplest to have all elements at the same mass and natural length, so that in the limit the system tends straightforwardly to a continuum.

A.3 Stiffness matrix

Substituting the strain-displacement interpolation relation from equation (A.5) into equation (A.7), and factoring out the initial stress state $\boldsymbol{\tau}^0$ later in equation (A.26), gives

$$\boldsymbol{\tau}_j = \mathbf{E}_j \mathbf{B}_j \mathbf{u}_j^{\mathrm{e}} \tag{A.17}$$

The displacement and strain distributions within the jth element are only affected by displacements at the element nodes. So $\mathbf{u}$ can be used in place of $\mathbf{u}_j^{\mathrm{e}}$ to give

$$\boldsymbol{\tau}_j = \mathbf{E}_j \mathbf{B}_j \mathbf{u} \tag{A.18}$$

Substituting equation (A.5) for $\overline{\boldsymbol{\epsilon}}_j$, and substituting equation (A.17) for $\boldsymbol{\tau}_j$, in the left-hand side of equation (A.9)

$$\begin{aligned} \sum_{j=1}^{N} \int_{V_j} \overline{\boldsymbol{\epsilon}}_j^T \boldsymbol{\tau}_j \, dV_j &= \sum_{j=1}^{N} \int_{V_j} \left(\mathbf{B}_j \overline{\mathbf{u}}_j^{\mathrm{e}}\right)^T (\mathbf{E}_j \mathbf{B}_j \mathbf{u}) \, dV_j \\ &= \left(\sum_{j=1}^{N} \int_{V_j} (\overline{\mathbf{u}}_j^{\mathrm{e}})^T \mathbf{B}_j^T \mathbf{E}_j \mathbf{B}_j \, dV_j \right) \mathbf{u} \end{aligned} \tag{A.19}$$

Equating the right-hand sides of equation (A.19) and equation (A.9) gives

$$\begin{aligned} \left(\sum_{j=1}^{N} \int_{V_j} (\overline{\mathbf{u}}_j^{\mathrm{e}})^T \mathbf{B}_j^T \mathbf{E}_j \mathbf{B}_j \, dV_j \right) \mathbf{u} &= \sum_{j=1}^{N} \int_{V_j} (\overline{\mathbf{u}}_j^{\mathrm{e}})^T \mathbf{H}_j^T \mathbf{r}_j^{\mathrm{V}} \, dV_j \\ &\quad + \sum_{j=1}^{N} \int_{S_j} (\overline{\mathbf{u}}_j^{\mathrm{e}})^T \mathbf{H}_j^T \mathbf{r}_j^{\mathrm{S}} \, dS_j + \sum_{i} (\overline{\mathbf{u}}_i^{\mathrm{n}})^T \mathbf{r}_i^{\mathrm{n}} \end{aligned} \tag{A.20}$$

Comparing this with the steady-state equilibrium equation (A.12) for an elastic body under a static load $\mathbf{f}$

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{A.21}$$

gives an expression for the symmetric *stiffness matrix* $\mathbf{K}$

$$\mathbf{K} = \sum_{j=1}^{N} \int_{V_j} \mathbf{B}_j^T \mathbf{E}_j \mathbf{B}_j \, dV_j \tag{A.22}$$

A.4 Mass matrix

Equation (A.20) also gives an expression for the load $\mathbf{f}$, composed of body, surface, and point components[3], minus initial loading $\mathbf{f}^0$ (which causes initial stresses $\boldsymbol{\tau}_j^0$)

$$\mathbf{f} = \mathbf{f}^{\mathrm{V}} + \mathbf{f}^{\mathrm{S}} + \mathbf{f}^{\mathrm{n}} - \mathbf{f}^0 \tag{A.23}$$

where

$$\mathbf{f}^V = \sum_{j=1}^{N} \int_{V_j} \mathbf{H}_j^T \mathbf{r}_j^{\mathrm{V}} \, dV_j \tag{A.24}$$

$$\mathbf{f}^S = \sum_{j=1}^{N} \int_{S_j} \mathbf{H}_j^T \mathbf{r}_j^{\mathrm{S}} \, dS_j \tag{A.25}$$

$$\mathbf{f}^0 = \sum_{j=1}^{N} \int_{V_j} \mathbf{B}_j^T \boldsymbol{\tau}_j^0 \, dV_j \tag{A.26}$$

$$\mathbf{f}^{\mathrm{n}} = \mathbf{r}^{\mathrm{n}} \tag{A.27}$$

For a time-varying load $\mathbf{f}(t)$, inertial effects are included using *d'Alembert's principle*[4]. Assuming that the element accelerations are in the same directions as the element displacements, and considering only body forces $\mathbf{f}^V$ for ease of notation, the inertia force is incorporated into the load vector

$$\begin{aligned} \mathbf{f}(t) &= \sum_{j=1}^{N} \int_{V_j} \mathbf{H}_j^T \left(\mathbf{f}_j^V - \rho_j \mathbf{H}_j \ddot{\mathbf{u}} \right) dV_j \\ &= \sum_{j=1}^{N} \int_{V_j} \mathbf{H}_j^T \mathbf{f}_j^V \, dV_j - \left(\sum_{j=1}^{N} \int_{V_j} \mathbf{H}_j^T \rho_j \mathbf{H}_j \, dV_j \right) \ddot{\mathbf{u}} \end{aligned} \tag{A.28}$$

where ρ is the mass density.

Comparing this to the body's dynamic equilibrium equation (A.12)

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f}(t) \tag{A.29}$$

gives the symmetric *mass matrix* $\mathbf{M}$

$$\mathbf{M} = \sum_{j=1}^{N} \int_{V_j} \mathbf{H}_j^T \rho_j \mathbf{H}_j \, dV_j \tag{A.30}$$

[3]Point forces are chosen to act at element nodes with no loss of generality.
[4]This treats the ma term in $f = ma$ as an *inertial* force, so that $f - f^{\mathrm{inert}} = 0$.

A.5 Numerical integration

Elemental mass and stiffness matrices from equations (A.30) and (A.22) contain integral terms

$$\mathbf{K}_j = \int_{V_j} \mathbf{B}_j^T \mathbf{E}_j \mathbf{B}_j \, dV_j \tag{A.31}$$

$$\mathbf{M}_j = \int_{V_j} \mathbf{H}_j^T \rho_j \mathbf{H}_j \, dV_j \tag{A.32}$$

which require numerical evaluation. One method of evaluation uses the *Gaussian quadrature formula* (Kreysig (ch.18.5) [77]), in which the integral

$$I = \int_{x_1}^{x_2} a(x)\, dx \tag{A.33}$$

is computed from n samples of $a(x)$ using the change of variable $(0 \leq r \leq 1)$

$$x(r) = x_1(1-r) + x_2 r \tag{A.34}$$

The integral is approximated by

$$\begin{aligned} I &= (x_2 - x_1) \int_0^1 a(x(r))dr \\ &\approx \sum_{i=1}^{\mathrm{n}} w_i a(r_i) \end{aligned} \tag{A.35}$$

where *weights* w_i and *abscissae* r_i are taken from Gaussian quadrature tables for an n-sample approximation. Equation (A.35) gives an exact integration for polynomials of degree $(2n-1)$.

A.6 Interpolation matrix for linear tetrahedron

Tetrahedral elements are used throughout this dissertation, because volumetric meshes for finite element models are relatively easily computed using (some form of) volumetric triangulation (Syn & Prager [144]). A tetrahedral element has four nodes, and is spanned by three *local coordinate* spatial parameters (r, s, t), varying from 0 to 1; thus its four nodal coordinates are $(0,0,0), (1,0,0), (0,1,0)$, and $(0,0,1)$.

A linear *interpolation function* $h(r,s,t)$ spanning the tetrahedron is completely specified by four constants (g_1, g_2, g_3, g_4), determined in this case using constraints at the four nodes

$$h(r,s,t) = \begin{pmatrix} 1 & r & s & t \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} \tag{A.36}$$

$$\Rightarrow \begin{pmatrix} h(0,0,0) \\ h(1,0,0) \\ h(0,1,0) \\ h(0,0,1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} \tag{A.37}$$

Equation (A.37) is written in terms of the *geometry matrix* $\mathbf{G}$

$$\mathbf{h} = \mathbf{G}\mathbf{g} \tag{A.38}$$

where $\mathbf{h}$ contains nodal values of $h(r,s,t)$, and $\mathbf{g}$ contains the polynomical coefficients of $h(r,s,t)$. Substituting $\mathbf{g}$ back into equation (A.36) gives

$$\begin{aligned} h(r,s,t) &= \begin{pmatrix} 1 & r & s & t \end{pmatrix} \mathbf{G}^{-1}\mathbf{h} \\ &= \begin{pmatrix} 1 & & & \\ -r & r & & \\ -s & & s & \\ -t & & & t \end{pmatrix} \mathbf{h} \end{aligned} \tag{A.39}$$

Comparing this to equation (A.4) shows that the displacement interpolation matrix $\mathbf{H}$ for a linear tetrahedral element is

$$\mathbf{H}(r,s,t) = \begin{pmatrix} 1 & & & \\ -r & r & & \\ -s & & s & \\ -t & & & t \end{pmatrix} \tag{A.40}$$

Iso-parametric elements

Along one edge of the tetrahedral element (r,s) are held constant, so the interpolation function varies linearly with t. Since this linear function is uniquely determined by the nodal constraints at both ends of the edge, there will be continuity along any interface between linear elements. Note that this does not hold true if global coordinates (x,y,z) are used to define the interpolation, instead of local coordinates (r,s,t).

Iso-parametric elements use local coordinate interpolation functions to map local properties to global coordinates, so as to maintain inter-element continuity. For example, the displacement of a point (r,s,t) within an element, can be described in terms of its global coordinate components (u_x,u_y,u_z)

$$\begin{pmatrix} u_x(r,s,t) \\ u_y(r,s,t) \\ u_z(r,s,t) \end{pmatrix}^T = \mathbf{H}(r,s,t)\mathbf{u}_j^{\mathrm{e}} \tag{A.41}$$

since element node displacements $\mathbf{u}_j^{\mathrm{e}}$ are also defined in global coordinates. This is useful in equation (A.3) which is presented in global coordinates and involves derivatives with respect to global coordinates.

Jacobian

The relationship established between local and global coordinates gives a Jacobian matrix $\mathbf{J}$, the determinant of which (the *Jacobian*) is used to effect a change of variables. For local coordinate components (r,s,t) in a tetrahedral element

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial s} \\ \frac{\partial}{\partial t} \end{pmatrix} = \mathbf{J} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \tag{A.42}$$

$$\Rightarrow \mathbf{J} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} & \frac{\partial z}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} & \frac{\partial z}{\partial t} \end{pmatrix} \tag{A.43}$$

The change of variables required during numeric integration can be then achieved with

$$dV = dx\,dy\,dz = \det \mathbf{J}\;dr\,ds\,dt \tag{A.44}$$

The Jacobian $det\,\mathbf{J}$ is efficiently computed for linear tetrahedra, by applying equation (A.42) to equation (A.41). It has a constant value $6V$, where V is the tetrahedral volume.

A.7 Hydrostatic pressure

Timoshenko & Goodier (ch.1.7)[157] define the *hydrostatic pressure* p of an element to be the negative mean of its principal stresses

$$p = -\frac{\tau_{xx} + \tau_{yy} + \tau_{zz}}{3} \tag{A.45}$$

The stress-strain relation for planar components (see equation (A.7)) is

$$\boldsymbol{\tau} = \mathbf{E}\boldsymbol{\epsilon} \tag{A.46}$$

$$\begin{pmatrix} \tau_{xx} \\ \tau_{yy} \\ \tau_{zz} \end{pmatrix} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1 & \frac{\nu}{1-\nu} & \frac{\nu}{1-\nu} \\ \frac{\nu}{1-\nu} & 1 & \frac{\nu}{1-\nu} \\ \frac{\nu}{1-\nu} & \frac{\nu}{1-\nu} & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \end{pmatrix} \tag{A.47}$$

where E is the Young's modulus and ν is the Poisson ratio. This means that hydrostatic pressure is directly dependent upon the local principal strain components[5]

$$p = -\frac{E}{(1-2\nu)} \frac{(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz})}{3} \tag{A.48}$$

For an infinitesimal element, the strain vector $\boldsymbol{\epsilon} = \nabla \cdot \mathbf{u}$. Let us examine the vector components of $\mathbf{u} = \begin{pmatrix} u_x & u_y & u_z \end{pmatrix}^T$ in the wave equation (A.13)

$$\begin{aligned} \frac{\partial^2}{\partial t} \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} &= -D^{\mathrm{el}} \nabla \left(\nabla \cdot \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} \right) \\ &= -D^{\mathrm{el}} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} \right) \end{aligned} \tag{A.49}$$

Since u_x exists only in the x-direction

$$\frac{\partial^2}{\partial y \partial x}(u_x) = \frac{\partial^2}{\partial z \partial x}(u_x) = 0 \tag{A.50}$$

and similarly for u_y and u_z, equation (A.49) becomes

$$\frac{\partial^2}{\partial t} \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = -D^{\mathrm{el}} \begin{pmatrix} \frac{\partial^2}{\partial x^2} \\ \frac{\partial^2}{\partial y^2} \\ \frac{\partial^2}{\partial z^2} \end{pmatrix} \cdot \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} \tag{A.51}$$

[5]The sum of principal strains is also known as the *dilatation* $\theta = \epsilon_{xx} + \epsilon_{yy} + \epsilon_z$. This gives a more intuitive explanation of pressure, as the result of local compression or negative dilatation.

By taking the mean of the spatial derivatives of the three vector components, the scalar wave equation for hydrostatic pressure p is given by

$$\frac{\partial^2 p}{\partial t} = -D^{\mathrm{el}}\nabla^2 p \tag{A.52}$$

This shows that there is a (scalar) pressure distribution associated with any (vector) displacement of an elastic body. Morever, there is a scalar pressure eigenfunction associated with each displacement eigenfunction of the (vector) wave equation (A.13).

A.8 Energy of elastic deformation

There is a potential energy W^{el} associated with the deformation of an elastic body. It is a measure of the work required to deform the body to a desired configuration, so the natural approach to computing the energy is to examine the eigenmodes involved in deformation.

The spectral distribution of W^{el} is determined by projecting the deformation $\mathbf{u}$ onto each eigenmode $\boldsymbol{\phi}_i$. Each projected displacement has *modal amplitude* $\|\boldsymbol{\phi}_i^T\mathbf{u}\|$, and from equation (3.5) *modal stiffness* is ω_i^2 (Bathe (ch.4.2.5)[8]).

The body's net elastic energy is therefore the sum of *modal energy* contributions from each of its $3n$ independent modes of elastic deformation[6]

$$W^{\mathrm{el}} = \sum_{i=1}^{3n} \frac{1}{2}\omega_i^2\|\boldsymbol{\phi}_i^T\mathbf{u}\|^2 \tag{A.53}$$

[6]Cf. the single degree-of-freedom expression $W^{\mathrm{el}} = \frac{1}{2}ku^2$, where k is stiffness and u is displacement.

Appendix B

The Symmetric Eigenproblem

The real symmetric generalised eigenproblem

$$\mathbf{K}\mathbf{\Phi} = \mathbf{M}\mathbf{\Phi}\mathbf{\Lambda}$$

arises in Section 3.2 from consideration of a body's elastic equilibrium. The matrix $\mathbf{\Phi}$ consists of orthogonal eigenmodes $\boldsymbol{\phi}_i$, and the (diagonal) matrix $\mathbf{\Lambda}$ consists of eigenvalues $\lambda_i = \omega_i^2$, where the natural-frequency of vibration of each eigenmode is ω_i. The mass matrix $\mathbf{M}$ and stiffness matrix $\mathbf{K}$ are assembled as described in Appendix A.

There is considerable literature on numerical methods used to solve large eigenproblems[1] in structural mechanics, since eigenmodes determine an elastic structure's dynamic response. This appendix describes the eigensolution methods used in this dissertation, and the presentation follows that of Parlett [109] and Saad [124]. Golub & van Loan [51] is also useful as the standard reference in numerical linear algebra.

Throughout this appendix the variable i indexes matrix elements, while n indicates the iteration count or a subspace dimension.

Appendix organisation

- **Section B.1** introduces the QR factorisation, which is used in orthonormalising vectors, as well as solving small eigenproblems.
- **Section B.2** presents the Rayleigh-Ritz procedure for computing optimal eigenpairs from an approximately invariant subspace.
- **Section B.3** presents the power method and subspace method for solving large eigenproblems.
- **Section B.4** presents the Lanczos method, which can be seen as a more efficient generalisation of the basic power method.
- **Section B.5** addresses computational issues which arise in the implementation of a generalised eigensolver.

[1] A 1000 node structure would be considered large in the context of currently available workstations, with each 3000×3000 double-precision matrix requiring 70 megabytes for uncompressed storage.

B.1 Orthogonality

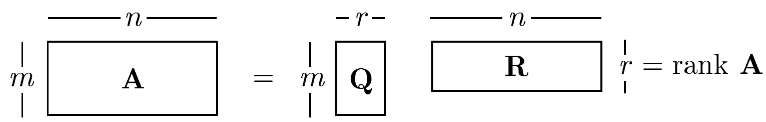
A real matrix $\mathbf{A}$ is defined as *orthogonal* if

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I} \tag{B.1}$$

where the columns and rows of $\mathbf{A}$ are also mutually orthogonal.

QR factorisation

The unique QR factorisation of a (non-null) rectangular $m \times n$ matrix $\mathbf{A}$



is the matrix formulation of the *Gram-Schmidt procedure* (Golub & van Loan (ch.5.2)[51]) applied to the orthonormalisation of the columns of $\mathbf{A}$. The factorisation results in an upper-triangular matrix $\mathbf{R}$, and the columns of $\mathbf{Q}$ form an orthonormal basis for *span* $\mathbf{A}$ (defined later).

There are a number of QR factorisation algorithms other than the straightforward translation of Gram-Schmidt, e.g. *Householder QR* applies Householder reflection transformations to an upper-triangular reduction (Golub & van Loan (ch.5.2)[51]).

QR algorithm for small eigenvalue problems

The QR transform algorithm repeatedly applies a similarity transform to a tridiagonal matrix[2] while preserving its tridiagonal form during each iteration, until convergence to a diagonal matrix is achieved. This algorithm can therefore be used to solve small eigenvalue problems which crop up in the Rayleigh-Ritz procedure in Section B.2.

Consider the QR factorisation

$$\mathbf{A} - \mu\mathbf{I} = \mathbf{Q}\mathbf{R} \tag{B.2}$$

for some origin shift μ. The *QR transform* for $\mathbf{A}$ is

$$\begin{aligned} \hat{\mathbf{A}} &\equiv \mathbf{R}\mathbf{Q} + \mu\mathbf{I} \\ &= \mathbf{Q}^T(\mathbf{A} - \mu\mathbf{I})\mathbf{Q} + \mu\mathbf{I} \\ &= \mathbf{Q}^T\mathbf{A}\mathbf{Q} \qquad \text{(since } \mathbf{Q}^T = \mathbf{Q}^{-1}\text{)} \end{aligned} \tag{B.3}$$

The close relationship between the convergence of this iteration and the power iteration (see later) are examined in Parlett (ch.8.5)[109]. The choice of shift μ to give good convergence to diagonal form is also discussed.

[2]Matrix tridiagonalisation for small problems can be achieved using Householder reflections (Parlett (ch.7.4) [109]). The Lanczos iteration described in Section B.4 achieves this efficiently for large eigenproblems.

B.2 Approximations from a subspace

A *subspace* $\mathcal{S}$ (of n-dimensional Euclidean space $\mathcal{E}^n$) is a subset which is closed under the operation of taking linear combinations (Parlett (ch.1.2)[109]). Any small set of vectors which generates $\mathcal{S}$ is a *spanning set*, and there are infinitely many spanning sets for each $\mathcal{S}$.

If these vectors form the columns of some matrix $\mathbf{A}$, then $\mathbf{Ax}$ denotes some linear combination of these vectors, and $\mathbf{A}$ can be said to *span* $\mathcal{S}$

$$\mathcal{S} = \operatorname{span} \mathbf{A} \tag{B.4}$$

The minimal set of spanning vectors forms a *base*, and the number of vectors in a base gives the dimension of $\mathcal{S}$. An n-dimensional base is notated by $\mathcal{S}^n$.

Invariant subspaces

Consider the eigenvector $\boldsymbol{\phi}_i$ of $\mathbf{A}$, which gives rise to the subspace $\mathcal{S}^1 = \operatorname{span} \boldsymbol{\phi}_i$ with the properties:

(i) $\mathcal{S}^1$ is mapped into itself by $\mathbf{A}$ (i.e. $\mathbf{A}\mathcal{S}^1 \subset \mathcal{S}^1$);

(ii) the image under $\mathbf{A}$ of any $\boldsymbol{\phi}_i$ in $\mathcal{S}^1$ is simply a fixed multiple of $\boldsymbol{\phi}_i$ (i.e. $\mathbf{A}\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i$ and λ_i depends on $\mathcal{S}^1$ alone and not $\boldsymbol{\phi}_i$).

Subspaces satisfying (ii) are called *eigenspaces*. Subspaces satisfying (i) are called *invariant*, and have a basis of eigenvectors.

Rayleigh-Ritz procedure

If the subspace $\mathcal{S}^n$ at hand turns out to be almost invariant under $\mathbf{A}$, it should contain good approximations of some eigenvectors of $\mathbf{A}$. The *Rayleigh-Ritz procedure* (RR) computes these approximations optimally[3], for a given initial $\mathbf{A}_0$ whose (full-rank) columns are a basis for $\mathcal{S}^n$.

The basic procedure is as follows:

(i) Orthonormalise $\mathbf{A}_0$ to give a full-rank $\mathbf{Q}$.

(ii) Form $\mathbf{AQ}$.

(iii) Form the (matrix) Rayleigh quotient $R(\mathbf{Q}) = \mathbf{Q}^T(\mathbf{AQ})$ (previously defined in Section 3.3).

(iv) Compute the eigenpairs of $R(\mathbf{Q})$ from $R\boldsymbol{\psi}_i^{\mathrm{R}} = \lambda_i^{\mathrm{R}} \boldsymbol{\psi}_i^{\mathrm{R}}$. The λ_i^{R} are the *Ritz values*.

(v) The *Ritz vectors* $\boldsymbol{\phi}_i^{\mathrm{R}} = \mathbf{Q}\boldsymbol{\psi}_i^{\mathrm{R}}$, together with the Ritz values, are the best approximations to eigenpairs of $\mathbf{A}$ which can be derived from $\mathcal{S}^n$ alone.

(vi) Compute $\|\mathbf{r}_i\|$, where $\mathbf{r}_i = \mathbf{A}\boldsymbol{\phi}_i^{\mathrm{R}} - \lambda_i^{\mathrm{R}}\boldsymbol{\phi}_i^{\mathrm{R}} = (\mathbf{AQ})\boldsymbol{\psi}_i^{\mathrm{R}} - \lambda_i^{\mathrm{R}}\boldsymbol{\phi}_i^{\mathrm{R}}$, to give residual error bounds. Each interval $[\lambda_i^{\mathrm{R}} - \|\mathbf{r}_i\|, \lambda_i^{\mathrm{R}} + \|\mathbf{r}_i\|]$ contains an eigenvalue of $\mathbf{A}$, and extra work (Parlett (ch.11.5)[109]) needs to be done if intervals overlap, since it is possible that two or more different λ_i^{R} are approximating the same eigenvalue of $\mathbf{A}$.

[3] Parlett (ch.11)[109] explains the sense in which RR is optimal.

B.3 Vector and subspace methods

Eigenvectors of rank one matrices

Consider the rank one matrix $\mathbf{A} = \boldsymbol{\phi}_1\boldsymbol{\phi}_1^T$. Its eigenvector[4] is found by first computing

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} \\ &= \boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\mathbf{x} \qquad (\mathbf{x} \neq \mathbf{0}) \end{aligned} \tag{B.5}$$

for some arbitrary $\mathbf{x}$. If $\mathbf{y} = \mathbf{0}$ then $\mathbf{x}$ is an eigenvector of eigenvalue 0 (i.e. $\boldsymbol{\phi}_i$ for $i > 1$). Otherwise since

$$\begin{aligned} \mathbf{A}\mathbf{y} &= \mathbf{A}\boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\mathbf{x} \\ &= \boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\mathbf{x} \\ &= \boldsymbol{\phi}_1^T\boldsymbol{\phi}_1\mathbf{y} \end{aligned} \tag{B.6}$$

then $\mathbf{y}$ is an eigenvector with eigenvalue $\boldsymbol{\phi}_1^T\boldsymbol{\phi}_1$. This eigenvalue can be computed by dividing a component of $\mathbf{A}\mathbf{y}$ by the corresponding (non-zero) component of $\mathbf{y}$.

Power iteration

For large n, the normalised matrix $\mathbf{A}^n/\|\mathbf{A}^n\|$ is close to a rank one matrix (Parlett (ch.4.2)[109]). The *power iteration* utilises this fact, and estimates the rank one eigenvector as shown above:

(i) Initialise with $\mathbf{x}_0$.

(ii) Compute $\mathbf{y}_n = \mathbf{A}\mathbf{x}_{n-1}$.

(iii) Normalise to give $\mathbf{x}_n = \mathbf{y}_k/\|\mathbf{y}_k\|$.

(iv) Test for convergence by comparing $\mathbf{x}_n$ and $\mathbf{A}\mathbf{x}_n$.

Subspace deflation

Parlett (ch.5.2)[109] discusses a number of methods of *subspace deflation*, which restricts a subspace to be orthogonal to previous eigenvectors, in order that the power iteration can proceed to find other eigenvectors[5].

The simplest technique is *deflation by restriction* which uses a starting vector $\mathbf{x}_0$ orthogonal to already computed eigenvectors. Loss of orthogonality during the power iteration will occur due to round-off errors, and this is remedied by occasional reorthogonalisation using the QR factorisation.

Subspace iteration

Subspace iteration is a straightforward generalisation of the power iteration, using iterations of a p-dimensional subspace and occasional orthonormalisation. The advantage over power iteration comes from an improved convergence to eigenvalue clusters, which makes up for any extra work required to deal with a bigger subspace than necessarily required. Another significant advantage is the ability to detect multiple eigenvales (depending on choice of p).

[4]If $\mathbf{A}$ is known *a priori* to be rank one, then any non-null column will give the dominant eigenvector.
[5]The accumulated round-off errors in orthogonalising higher eigenmodes against lower ones, mean that higher eigenmodes are computed with worse accuracy.

The difficulty is that eigenvalue distribution is not usually known *a priori*, and efficiency is sensitive to the value of p chosen. Parlett (ch.14.4)[109] discusses convergence and termination criteria for subspace iteration.

The basic algorithm is as follows:

(i) Initialise with $\mathbf{X}_0$ of dimension p.

(ii) Compute $\mathbf{Y}_n = \mathbf{A}\mathbf{X}_{n-1}$.

(iii) Test each column for convergence as in power iteration.

(iv) Orthonormalise using QR factorisation $\mathbf{Y}_n = \mathbf{Q}_n\mathbf{R}_n$.

(v) Set $\mathbf{X}_n = \mathbf{Q}_n$.

More refined implementations incorporate the application of RR to $\mathbf{X}_n$, since the columns of $\mathbf{X}_n$ are not the optimal approximations to target eigenvectors (Parlett (ch.14)[109]).

B.4 Lanczos method

Krylov subspaces

A *Krylov subspace* $\mathcal{K}^n(\mathbf{x})$ is spanned by the column vectors in a *Krylov matrix* $\mathbf{K}^n(\mathbf{x})$

$$\mathbf{K}^n(\mathbf{x}) = (\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{n-1}\mathbf{x}) \tag{B.7}$$

$$\mathcal{K}^n(\mathbf{x}) = \operatorname{span}\mathbf{K}^n(\mathbf{x}) \tag{B.8}$$

The power iteration only utilises $\mathbf{A}^n\mathbf{x}$ and not $\mathcal{K}^n(\mathbf{x})$, so by applying RR to the subspace $\mathcal{K}^n$, we obtain better and more cost-effective approximations. See Parlett (ch.12)[109] for the convergence analysis.

There is an important theoretical limitation on the power iteration and Krylov subspace methods: if $\mathbf{x}$ is orthogonal to any eigenvectors of $\mathbf{A}$, then their projections will be zero and they cannot be detected using $\mathcal{K}^n$. However roundoff errors during computation usually make this an irrelevant consideration.

Tridiagonal form

Consider the *Lanczos basis* $\mathbf{Q}_n$ defined by Gram-Schmidt orthonormalisation of a full-rank Krylov matrix $\mathbf{K}^n(\mathbf{x})$

$$\mathbf{K}^n(\mathbf{x}) = \mathbf{Q}_n\mathbf{R}_n \tag{B.9}$$

This QR factorisation is very efficient for Krylov subspaces, yielding a three-term recurrence relation for columns of $\mathbf{Q}_n = (\mathbf{q}_1, \ldots, \mathbf{q}_n)$. For full-rank $\mathbf{K}^n(\mathbf{x})$, a tridiagonal matrix $\mathbf{T}_n$ is given by (Parlett (ch.12.6)[109])

$$\begin{aligned}\mathbf{T}_n &= \mathbf{Q}_n^T\mathbf{A}\mathbf{Q}_n \\ &= \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{n-1} \\ & & & \beta_{n-1} & \alpha_n \end{pmatrix}\end{aligned} \tag{B.10}$$

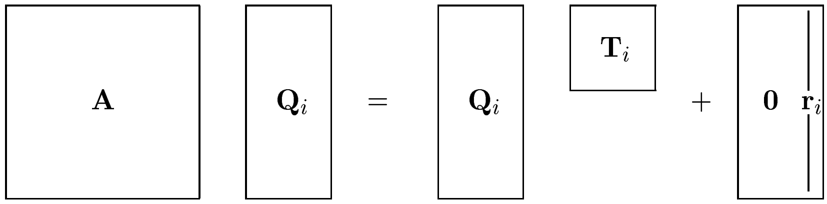
from which

$$\alpha_i = \mathbf{q}_i^T \mathbf{A}\mathbf{q}_i \tag{B.11}$$

$$\beta_i = \mathbf{q}_{i+1}^T \mathbf{A}\mathbf{q}_i \tag{B.12}$$

$$\mathbf{A}\mathbf{q}_i = \mathbf{q}_{i-1}\beta_{i-1} + \mathbf{q}_i\alpha_i + \mathbf{q}_{i+1}\beta_i \tag{B.13}$$

For each $i < n$



so that the residual $\mathbf{r}_i$ in the last column on the right is given by

$$\mathbf{r}_i = \mathbf{A}\mathbf{q}_i - \mathbf{q}_{i-1}\beta_{i-1} - \mathbf{q}_i\alpha_i \tag{B.14}$$

$$= \mathbf{q}_{i+1}\beta_i \tag{B.15}$$

Lanczos iteration

For $\mathbf{q}_0 = \mathbf{0}$, and some initial $\mathbf{q}_1$ approximating the desired eigenvector:

(i) Compute $\mathbf{r}_i$ from equation (B.14).

(ii) $\beta_i = \|\mathbf{r}_i\|$ (since $\beta_i = \|\mathbf{q}_{i+1}\beta_i\| = \|\mathbf{r}_i\|$).

(iii) If $\beta_i = 0$ then the residual is zero, so that $\mathbf{A}\mathbf{Q}_i = \mathbf{Q}_i\mathbf{T}_i$. Therefore span $\mathbf{Q}_i$ is the smallest invariant subspace containing $\mathbf{x}$.

(iv) If $\beta_i > 0$ then $\mathbf{q}_{i+1} = \mathbf{r}_i/\|\mathbf{r}_i\|$, go to (i).

The Lanczos iteration can also be interpreted as the application of RR to the sequence of Krylov subspaces $\mathcal{K}^i(\mathbf{x})$. At each step the subspace dimension increases by one, and the best approximate eigenpairs $(\lambda_i^{\mathrm{R}}, \boldsymbol{\psi}_i^{\mathrm{R}})$ in the subspace are computed economically and straighforwardly[6] from the tridiagonal $\mathbf{T}_i$, which is effectively the (matrix) Rayleigh quotient $R(\mathbf{Q}_i)$. The Ritz pairs $(\lambda_i^{\mathrm{R}}, \mathbf{Q}_i\boldsymbol{\psi}_i^{\mathrm{R}})$ are then tested for convergence.

The efficiency of this algorithm, together with superior convergence properties, makes the Lanczos iteration currently the best algorithm for solving a few inner or outer eigenpairs of large symmetric eigenproblems (Golub & van Loan (ch.9)[51]; Grimes et al. [59]).

Lanczos vectors $\mathbf{q}_i$ lose mutual orthogonality as the number of iterations increase, so occasional reorthogonalisation is required; *selective orthogonalisation* is a more efficient modification (Parlett (ch.13)[109]). The ability to automate this orthogonalisation, by maintaining a reasonable preset level of linear independence, means that the Lanczos method can be used as a 'black box' eigensolver, requiring no delicate parameters to be set by the user.

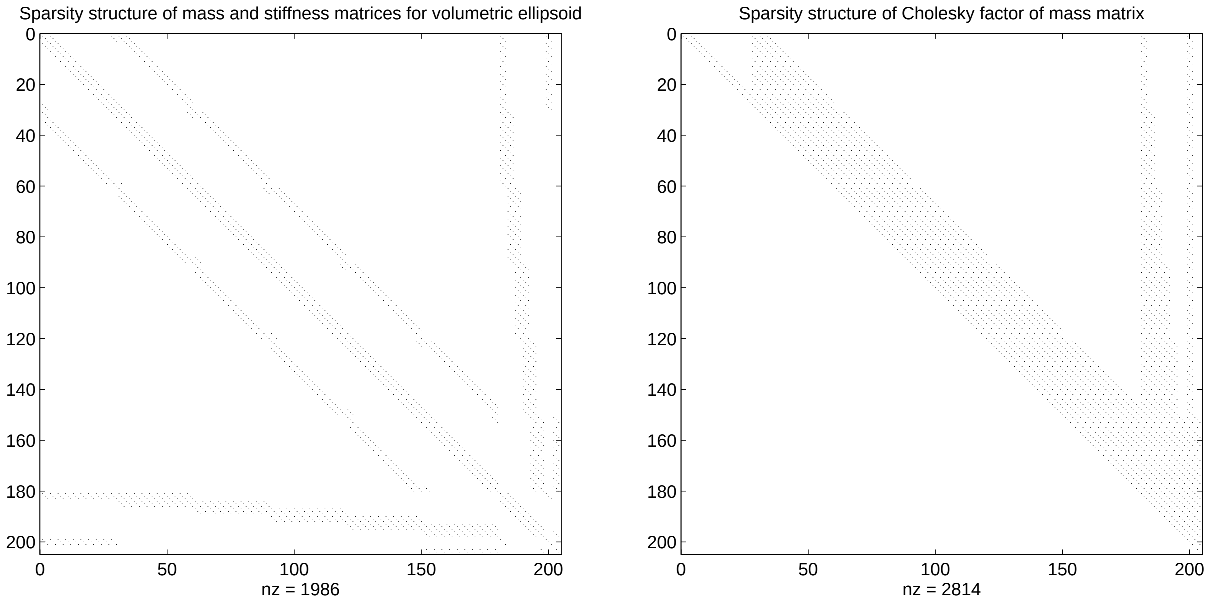
The block generalisation of the Lanczos method (Parlett (ch.13.10)[109]) operates with a block of eigenvectors, which allows the detection of multiple eigenvalues (cf. subspace iteration).

[6] See previously for the QR transform algorithm.

B.5 Implementation

Sparse indexing

The sparsity structure of the mass and stiffness matrices of an ellipsoidal volume are shown in Figure B.5(a).



(a) Sparsity of mass and stiffness matrices

(b) Sparsity of Cholesky factor of mass matrix

Figure B.1: Sparsity plots for a finite element ellipsoidal volume (nz = number of non-zero elements)

Dense storage of the matrix structure is very expensive, both in memory usage as well as in computation. Sparse indexing technology (Saad (ch.2)[124]) allows efficient storage of sparse matrices, so that more matrix elements can be kept in core memory. When properly designed, a sparse indexing system also allows for efficient computation of matrix-vector products by exploiting the fact that many of the elemental multiplications are guaranteed to be zero.

Bathe (app.2.3)[8] recommends renumbering nodes in order to reduce matrix bandwidth, which directly affects the computational cost of solution methods. A sparse matrix-vector product routine should incorporate this consideration.

Standard eigensolver for generalised eigenproblem

Given a standard eigensolver for

$$\mathbf{A\Phi} = \mathbf{\Phi\Lambda} \tag{B.16}$$

the naive approach to solving the generalised eigenproblem

$$\mathbf{K\Phi} = \mathbf{M\Phi\Lambda} \tag{B.17}$$

is to invert the mass matrix to form the standard eigenproblem (B.16), which destroys the sparsity structure of the mass and stiffness matrices.

The *Cholesky factorisation* for $\mathbf{M} > \mathbf{0}$ is given by

$$\mathbf{M} = \mathbf{L}^T \mathbf{L} \tag{B.18}$$

and Figure B.5(b) shows that sparsity is preserved for the upper-triangular Cholesky factor $\mathbf{L}$. Given this factorisation, equation (B.17) can be solved in standard form[7]

$$\mathbf{L}^{-T} \mathbf{K} \mathbf{L}^{-1} \mathbf{\Phi} = \mathbf{\Phi} \mathbf{\Lambda} \tag{B.19}$$

The power iteration, subspace iteration and Lanczos algorithms only require the matrix $\mathbf{A}$ to be specified during the formation of a matrix-vector product

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} \\ &= (\mathbf{L}^{-T} \mathbf{K} \mathbf{L}^{-1}) \mathbf{x} \end{aligned} \tag{B.20}$$

which allows the user to exploit special properties of the matrix $\mathbf{A}$, e.g. with sparse indexing.

An efficient method of computing equation (B.20) is to use sparse triangular-solve and matrix-vector product routines as follows

$$\begin{aligned} \mathbf{L}\mathbf{w} &= \mathbf{x} && \text{(giving } \mathbf{w}\text{)} \\ \mathbf{z} &= \mathbf{K}\mathbf{w} && \text{(giving } \mathbf{z}\text{)} \\ \mathbf{L}^T \mathbf{y} &= \mathbf{z} && \text{(giving } \mathbf{y}\text{)} \end{aligned} \tag{B.21}$$

Mass lumping

A *lumped-mass* approximation to the *consistent-mass* representation derived in Appendix A.4, gives a diagonal mass matrix instead of a densely populated one, by lumping the element's distributed mass at nodal positions.

There are a number of lumping schemes, the simplest of which Hughes[66] calls *special lumping*. This discards the off-diagonal elements of each element's mass matrix, and scales the remaining ones to give the same total elemental mass. This scheme has the advantage of ensuring that the mass matrix is positive-definite.

Special lumping is a very good approximation for a uniformly tessellated structure with uniformly distributed mass and stiffness, especially when only the low-frequency response is required for dynamic analysis (Hitchens (ch.1)[62]). Mass lumping also tends to compensate for the FEM's tendency to over-estimate natural frequencies.

The FEM converges monotonically to give the true eigenmode shapes as the mesh gets finer (Bathe (ch.4)[8]), but the FEM tessellation is frequently formed from a sparse and noisy mesh with non-uniform spacing. The eigenmode displacement fields resulting from mass lumping are bound to be less accurate, but are at the same time more invariant to changes in mesh density, which is an advantage when eigenmodes are being used as shape features[8].

A diagonal mass matrix is trivial to invert, hence equations (B.21) can be replaced with a single matrix-vector product using $\mathbf{M}^{-1}\mathbf{K}$. This reduces computation time substantially, since code profiling suggests that up to 65% of computing cycles are spent in the two sparse triangular-solve routines in equations (B.21).

[7] Using $\mathbf{L}^{-T}\mathbf{K}\mathbf{L}^{-1}$ preserves symmetry.

[8] The inertial effects of distributed mass are effectively muted by mass lumping, so that frequency response is determined mainly by the stiffness matrix, which can be thought of as a shape 'connectivity matrix'.

Software

The eigensolvers used to generate the results in this dissertation, are based on the following publically available software:

(i) **power method**: the `rsg()` subroutine from the `eispack` eigensystem solution package.

(ii) **subspace method**: the `simitz()` subroutine based on Rutishauser [123].

(iii) **block Lanczos**: the `minval()` subroutine based on Underwood [162] and Golub et al. [50].

(iv) **sparse matrices**: the `sparskit` subroutine package described in Saad (ch.2)[124].

The first three items are available on `netlib`, together with utility subroutines for basic linear algebraic manipulation. See the `sci.math.num-analysis` newsgroup for ftp sites which contain the `netlib` collection. The fourth item is available via anonymous access to `ftp.cs.umn.edu`.