Continuous Medial Models in Two-Sample Statistics of Shape

by
Timothy B. Terriberry

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2006

Approved by:

Guido Gerig, Advisor

Sarang C. Joshi, Reader

Stephen M. Pizer, Reader

James N. Damon, Committee Member

Dinesh Manocha, Committee Member

Martin Styner, Committee Member

ABSTRACT
TIMOTHY B. TERRIBERRY: Continuous Medial Models in Two-Sample Statistics of Shape.
(Under the direction of Guido Gerig.)

In questions of statistical shape analysis, the foremost is how such shapes should be represented. The number of parameters required for a given accuracy and the types of deformation they can express directly influence the quality and type of statistical inferences one can make. One example is a medial model, which represents a solid object using a skeleton of a lower dimension and naturally expresses intuitive changes such as "bending", "twisting", and "thickening".

In this dissertation I develop a new three-dimensional medial model that allows continuous interpolation of the medial surface and provides a map back and forth between the boundary and its medial axis. It is the first such model to support branching, allowing the representation of a much wider class of objects than previously possible using continuous medial methods.

A measure defined on the medial surface then allows one to write integrals over the boundary and the object interior in medial coordinates, enabling the expression of important object properties in an object-relative coordinate system. I show how these properties can be used to optimize correspondence during model construction. This improved correspondence reduces variability due to how the model is parameterized which could potentially mask a true shape change effect.

Finally, I develop a method for performing global and local hypothesis testing between two groups of shapes. This method is capable of handling the nonlinear spaces the shapes live in and is well defined even in the high-dimension, low-sample size case. It naturally reduces to several well-known statistical tests in the linear and univariate cases.

ACKNOWLEDGMENTS

No scientific progress is possible without standing on the shoulders of giants, and it was my good fortune to have a group of professors and fellow students filled with many such shoulders. As the references will attest, I have made much use of their ideas, and this dissertation would not have been possible without them. I can think of no greater compliment one could pay a fellow scientist.

But there is another group of people who are essential to any student's progress: the accountants, administrators, assistants, and secretaries who keep the department running. And so, I extend my sincerest gratitude to Delphine Bull, Katrina Coble, Kelli Gaskill, Janet Jones, Sandra Neely, Deb O'Connor, Tammy Pike, and Karen Thigpen. These are the folks who make sure we get from here to there.

And finally, I would like to thank all the dear friends I made in Chapel Hill for helping to keep me sane throughout the process. I only meant to stay awhile.

Contents

List of Figures xi

List of Tables xiii

1 Introduction 1

1.1 Motivation and Goals . . . 1

1.1.1 A Continuous Medial Model . . . 2

1.1.2 Model Fitting With Explicit Correspondence Optimization . . . 3

1.1.3 Nonlinear Hypothesis Testing . . . 4

1.1.4 Thesis and Claims . . . 5

1.1.5 Overview . . . 5

2 Medial Properties 7

2.1 An Overview of the Medial Axis . . . 7

2.2 Shape Analysis Using the Medial Axis . . . 13

2.2.1 M-reps . . . 15

2.3 Mathematical Background . . . 18

2.3.1 Whitney Stratified Sets . . . 18

2.3.2 Skeletal Sets and Radial Vector Fields . . . 20

2.3.3 The Blum Medial Axis as a Skeletal Structure . . . 25

2.3.4 Geometry of the Boundary . . . 27

2.3.5 Integrals Over Skeletal Structures . . . 31

3 A Continuous Medial Model 37

3.1 Catmull-Clark Subdivision on the Medial Axis . . . 38

3.1.1 Ordinary Corner-Free Boundary . . . 39

3.1.2 Interior Patches . . . 42

3.2 Edge Patches . . . 46
3.2.1 Solving for $r_u(0, v)$ . . . 48
3.2.2 The Complete Control Curve . . . 50
3.2.3 Sympathetic Overfolding . . . 51
3.3 Branch Curve Patches . . . 56
3.3.1 Satisfying Branch Conditions Away From Fin Creation Points . . . 58
3.3.2 Satisfying Edge and Branch Conditions at a Fin Creation Point 59
3.3.3 Transition Region . . . 63
3.4 Summary and Conclusion . . . 64

4 Model Fitting With Explicit Correspondence Optimization 67
4.1 Approximating Medial Integrals . . . 69
4.1.1 Sampling the Medial Axis . . . 69
4.1.2 Moment Integrals . . . 72
4.1.3 Estimating Volume Overlap . . . 73
4.2 Single-Subject Model Fitting . . . 75
4.2.1 Template Alignment . . . 75
4.2.2 Model Deformation . . . 76
4.2.3 Local Nonlinear Optimization . . . 82
4.2.4 Constrained Optimization . . . 86
4.2.5 Results . . . 90
4.3 Population-Based Model Fitting . . . 93
4.3.1 The Objective Function for Correspondence . . . 98
4.3.2 Parameterizing the Correspondence Optimization . . . 100
4.3.3 Results . . . 102
4.4 Conclusion . . . 103

5 Nonlinear Hypothesis Testing 106
5.1 Introduction . . . 107
5.1.1 A Metric Space for M-reps . . . 109
5.1.2 One-sample Statistics in Nonlinear Spaces . . . 110
5.1.3 Two-sample Statistics . . . 112
5.2 Multivariate Permutation Tests . . . 113
5.2.1 The Univariate Case . . . 113
5.2.2 Partial Tests . . . 114
5.2.3 Multivariate Combination . . . 115

5.2.4 Relation to Other Testing Procedures 118
5.3 Experimental Data and Results 119
5.4 Conclusion 122

6 Conclusion **125**

A Computing $U^{\pm}$ and $S_{\mathbf{rad}}$ in Three Dimensions **130**
A.1 Derivatives of the Edge Control Curve 134

B Derivatives for Optimizing $U^{\pm}$ and $S_{\mathbf{rad}}$ **137**
B.1 Derivatives of the Edge Control Curve 140

Bibliography **143**

List of Figures

1.1 Different notions of correspondence 4

2.1 Examples of the medial axes of some generic shapes 10

2.2 The generic local singular structures for the Blum medial axis 11

2.3 The Whitney umbrella 22

2.4 An example of overfolding 28

3.1 Catmull-Clark subdivision masks 40

3.2 Control points around an extraordinary vertex 43

3.3 Subdivision near an extraordinary vertex 44

3.4 The control curve on an edge patch 47

3.5 An example of sympathetic overfolding 52

3.6 An example of a pair of edge patches 55

3.7 Two close-ups of the right portion of an edge patch 56

3.8 The cross-section of a branch curve 58

3.9 A fin point example 63

4.1 Population-based correspondence optimization 68

4.2 The best fit for the deformed ellipsoids 92

4.3 The worst fit for the deformed ellipsoids 92

4.4 An average fit for the deformed ellipsoids 92

4.5 Surface errors for the deformed ellipsoids 93

4.6 The best fits for the caudate data 94

4.7 The worst fits for the caudate data 94
4.8 Average fits for the caudate data 95
4.9 Surface errors for the caudate data 95
4.10 Clustering of corresponding points for the deformed ellipsoids 103
4.11 Clustering of corresponding points for the caudate data 104
5.1 Lateral ventricle examples 108
5.2 Example data and test statistics 116
5.3 Empirical distribution of our example 118
5.4 Genetic similarity results for local tests 122
5.5 Correspondence failure for MZ left ventricles 123

List of Tables

4.1 Surface errors for the deformed ellipsoids 91

4.2 Surface errors for the caudate data 93

5.1 Genetic similarity results for global tests 121

Chapter 1

Introduction

> . . . it appears to me possible. . . to reconstruct the skeleton of the one from our knowledge of the skeleton of the other, under the guidance of the same correspondence as is indicated in their external configuration.
>
> (Thompson, 1917)

1.1 Motivation and Goals

D'Arcy Thompson's words, written the better part of a century ago to describe his hand-made sketches of two different species of fish, eloquently sum up the goal of this work. This dissertation seeks to place a mathematical formalism and computational framework behind the process Thompson describes, and in so doing develops several new tools for the understanding of populations of shapes.

Given the widespread use of medical imaging, there is a great need for techniques to find and analyze the objects in them, a process greatly aided by an a priori understanding of the variability of shapes between subjects and over time. These can be used for treatment planning: for visualizations of minimally-invasive surgery (possibly even while the operation is in progress) or radiation therapy, which aims to target tumors without killing the surrounding healthy tissue.

However, many applications also require comparisons between multiple populations. Several neurological disorders such as schizophrenia and autism show systematic changes in various structures of the brain. Being able to prove a statistical difference in the shape of an organ between healthy and unhealthy patients can aid in diagnosis. Being able to localize these changes to specific areas and describe them in intuitive

ways, such as "bending", "twisting", or "thickness", helps researchers to understand the disease process. Furthermore, these tools help in understanding the effects of various drugs used to treat these diseases. They can also be used to analyze the processes of growth and early development.

However, describing shapes in these terms requires nonlinear models to capture what are inherently nonlinear processes. These models require new statistical methods that generalize those developed for linear spaces to nonlinear ones.

1.1.1 A Continuous Medial Model

One particularly powerful class of shape representations are *medial representations*, or m-reps. As a kind of skeleton, their description carries with it information about the solid shape that boundary descriptions do not and naturally decomposes into figural parts, connected at branches where multiple parts meet. They also naturally describe the intuitive types of shape change listed above.

The medial geometry provides an intrinsic link between the boundary and the medial axis, which allows one to use the appropriate representation for the desired analysis task. However, there are a number of limitations with current m-rep models. In the discrete version (Pizer et al., 1999) the connection with the boundary is tenuous: the link is given only at a coarse set of discrete points and the interpolation given by (Thall, 2004) to recover a dense sampling does not respect the intrinsic medial geometry. More recent work on interpolation that does respect this geometry shows a promising alternative (Han et al., 2006), but it sacrifices the uniqueness of the representation, requires expensive numerical integration, and in the end only approximately interpolates the original discrete model.

An alternative approach, cm-reps, consider the problems of designing a discrete computer representation and computing its associated continuous mathematical representation together as a coupled system. However, generalizations to three dimensions require an infinite number of boundary conditions to be satisfied. Existing techniques for enforcing them are limited to a single figure (Yushkevich, 2003; Yushkevich et al., 2005), and thus can represent only a limited class of objects.

In order to overcome these limitations, Chapter 3 describes a new continuous medial model constructed with subdivision surfaces. It is based on a novel technique, called a "control curve", which can enforce any arbitrary first- and second-order boundary condition on the entire length of the edges of a Catmull-Clark subdivision surface.

This technique allows the creation of the first continuous medial model that supports branching in three dimensions.

1.1.2 Model Fitting With Explicit Correspondence Optimization

The link between medial geometry and the geometry of the boundary provided by a continuous medial representation allows many operations on the boundary or interior of the shape to be expressed in terms of integrals in object-relative, medial coordinates. This forms the basis for the model-fitting framework outlined in Chapter 4. A method of sampling the continuous medial axis is given, and this sampling is used to numerically approximate medial integrals. Several examples of such integrals include the computation of volume overlap, which is a metric used to evaluate goodness-of-fit, as well as moments up through second order, which are used to initialize the fitting process by aligning a template to the target shape. The fitting process itself is governed by the constrained optimization of an objective function, also given by a medial integral, over several scales in succession. This chapter derives such multi-scale objective functions for both voxelized binary images and also triangulated surface meshes.

Finally, it shows how medial integrals can be used to tackle the problem of *correspondence*. An inherent problem in any shape representation used for statistical analysis, one must ensure that the parameters of the representation in some sense control the "same" features of the resulting shape. Otherwise, noise in the parameterization can overwhelm the size of any shape change effect, reducing or eliminating the power of the tests.

What Chapter 4 does *not* do is provide yet another means for determining the correspondence between two shapes. Instead, it takes the stance that correspondence is inherently specific to the problem domain. Consider the classic "box-bump" example in Figure 1.1 (Kotcheff and Taylor, 1998). Depending on whether the underlying physical model of the two shapes is that of multiple protrusions by varying amounts, or a single protrusion with a varying position, the "average shape" produced gives two very different results. Either result is valid, and the implicit correspondence given by a particular shape representation could just as easily produce either, but without some input governing the process whereby the representation is constructed results are unpredictable. One such governing process is to examine a sufficiently large number of examples from the population, and indeed many researchers have taken this approach.

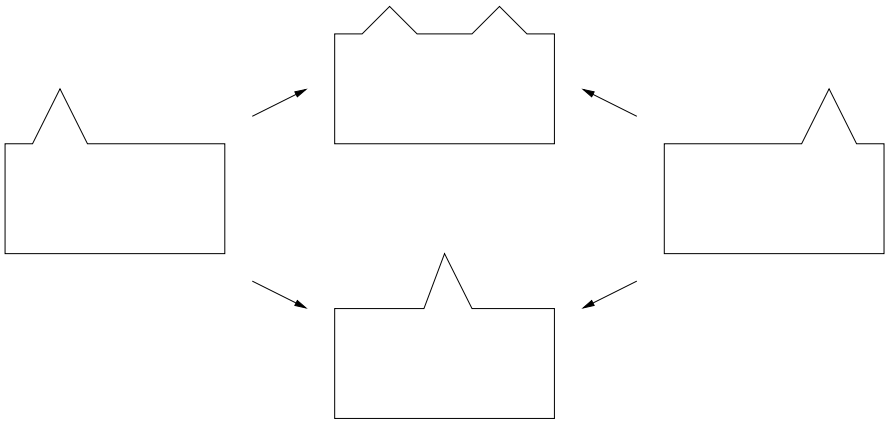


Figure 1.1: Different notions of correspondence can lead to different meanings of the average of a shape, as well as other statistics.

These methods of determining correspondence often focus solely on the boundary, since in many medical imaging modalities the interior of objects are relatively uniform in appearance and their features are poorly localized. Because a continuous representation affords an explicit link between the boundary and the medial locus, the implicit correspondence given by the parameterization of the latter is optimized to match an explicit correspondence given on the former. Any number of existing methods, as well as future, problem-specific ones, can determine the correspondence, providing increased flexibility to adapt to the problem at hand.

1.1.3 Nonlinear Hypothesis Testing

The recent development of tools like Principal Geodesic Analysis (Fletcher, 2004) can characterize a single statistical distribution in a nonlinear space. Chapter 5 in turn lays out a hypothesis testing framework for comparing two different distributions in nonlinear spaces that is designed specifically to address the challenges of shape analysis: a large number of parameters with a small number of subjects (high-dimension, low-sample-size), correlation among the parameters, and non-commensurate parameters measured in units that are not directly comparable. The resulting test is a multivariate permutation test that can simultaneously combine all of the parameters of the shape model—position, local thickness, etc.—in contrast with previous work that focused only on a single aspect at a time. It requires only the existence of a metric on the space, reduces to an ordinary permutation test in the univariate case, and converges to the classic Hotelling's T^2 test in the multivariate case as the sample sizes tend to infinity.

1.1.4 Thesis and Claims

Thesis: *Nonlinear growth and deformation of biological objects requires nonlinear shape models to effectively characterize these processes. A nonlinear hypothesis test based on multivariate permutation tests provides an effective means of detecting and localizing their effects between groups. A continuous medial model, which describes this variation in a natural way, can be used to effectively represent these shapes. The geometric link between the medial axis and the boundary such a model provides can be used to optimize the correspondence across a population of objects, reducing parameterization noise to increase the power of statistical tests.*

The main contributions are

1. A novel "control curve" formulation capable of enforcing first- and second-order boundary conditions at the edges of subdivision surfaces.

2. A new continuous medial model based on this formulation, the first such model capable of representing branching in three dimensions.

3. A method of approximating medial integrals driven by a uniform sampling of these continuous models.

4. Applications of these medial integrals. These include computing volume overlap to evaluate goodness-of-fit and computing second-order moments to align models to a common position, orientation and scale.

5. A multi-scale model fitting framework utilizing a highly constrained multivariate nonlinear optimization to fit these models to a target shape.

6. A new correspondence optimization method that works in tandem with the model fitting process to produce a group of models with a common parameterization.

7. A novel nonlinear hypothesis test for m-reps, which simultaneously considers all of the parameters of the shape model.

1.1.5 Overview

After this introduction, Chapter 2 provides background on the medial axis, previous research utilizing it for shape analysis, and the fundamental mathematics underlying it. Other necessary background material on subdivision surfaces, nonlinear optimization,

correspondence, Riemannian geometry, and hypothesis testing are embedded directly in the chapters that require them. The next three chapters form the bulk of the dissertation, covering each of the three topics outlined above. Chapter 3 defines the new medial representation, and shows how it can model branching in three dimensions. Chapter 4 gives a method for fitting these models to existing shapes. Chapter 5 describes the nonlinear hypothesis test used to compare shape populations. Finally, Chapter 6 summarizes the contributions made and discusses future work.

Chapter 2

Medial Properties

> We are apt to think of mathematical definitions as too strict and rigid for common use, but their rigour is combined with all but endless freedom. ... By means of these large limitations, by this controlled and regulated freedom, we reach through mathematical analysis to mathematical synthesis... as for instance, when we learn that, however we hold our chain, or however we fire our bullet, the contour of the one or the path of the other is always mathematically homologous.
>
> (Thompson, 1917)

This chapter outlines the properties and mathematics of the Blum medial axis as it has been used in research up to this point. Additional background material, being fairly localized in its application, is not presented here, but rather directly in the chapters that require it. Chapter 3 contains a review of Catmull-Clark subdivision surfaces, Chapter 4 describes some necessary nonlinear optimization techniques, and Chapter 5 briefly reviews hypothesis testing and one-sample statistics in Riemannian symmetric spaces.

2.1 An Overview of the Medial Axis

The medial axis was introduced by (Blum, 1967) primarily as a tool for 2D shape analysis. Extensions to 3D were discussed in (Blum, 1973) and further developed in (Nackman, 1982), and the relationship between the medial axis and the associated boundary were explored in (Vermeer, 1994; Gelston and Dutta, 1995; Teixeira, 1998;

Giblin and Kimia, 2004). Extending the definition to higher dimensions is straightforward; doing so, (Damon, 2003; Damon, 2004) provides an analysis of the differential geometry in any dimension.

A precise definition may be formulated in several different, equivalent ways. One approach is in terms of a *maximally inscribed ball*.

Definition 2.1 *Given a point $m \in \mathbb{R}^n$, a **closed ball** of radius r around m is the set $\overline{B}_r(m) = \{x \in \mathbb{R}^n : \|x - m\| \leq r\}$.*

Definition 2.2 *Given a closed, connected set $\Omega \subseteq \mathbb{R}^n$, a **maximally inscribed ball** in Ω is a closed ball $B \subseteq \Omega$ such that there is no closed ball B' with $B \subsetneq B' \subseteq \Omega$.*

Definition 2.3 *The **Blum medial axis** of a closed, connected set $\Omega \subseteq \mathbb{R}^n$ is the set $M = \{m \in \mathbb{R}^n : \overline{B}_r(m)$ is maximally inscribed for some $r\}$.*

By itself, the Blum medial axis is insufficient to describe the shape of an object, as many objects may have the same medial axis. Thus, we also define an augmented structure.

Definition 2.4 *The **augmented Blum medial axis** of a closed, connected set $\Omega \subseteq \mathbb{R}^n$ is the set $\mathcal{M} = \{(m, r) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} : \overline{B}_r(m)$ is maximally inscribed$\}$.*

This *is* a complete description of the object. The set Ω may be recovered from $\mathcal{M}$—at least in an abstract mathematical sense—as the union of the maximally inscribed balls. Authors in various fields have used the terms *medial scaffold*, *central set*, *internal medial locus*, and others to refer to the augmented or unaugmented versions of the medial axis. Where the notation, M or $\mathcal{M}$, makes the definition in question clear, this dissertation will omit the word "augmented".

There are many equivalent, alternative ways to define the medial axis. One is based on the *symmetry set*, which replaces maximally inscribed balls with *bitangent spheres*: closed balls that are tangent to the boundary $\mathcal{B} = \partial\Omega$ of the object in (at least) two different places (Giblin and Brassett, 1985). This requires that the object have a well-defined tangent plane (almost) everywhere on its boundary, which in practice is a reasonable restriction. The full symmetry set is larger than the medial axis, containing the closure of all such bitangent spheres. The closure is necessary to include points like the end points (such as the A_3 point in Figure 2.1(a)), which in the generic case are only tangent in one place. The medial axis is then defined as the subset of those spheres in the symmetry set that are also entirely contained inside the object.

A third approach defines the medial axis as the shocks of the *grassfire flow* (Blum, 1973). The analogy likens the object to a field of grass whose entire boundary is simultaneously set ablaze. As the fire's front spreads from the boundary at a uniform speed, it meets at places called *shocks* from two different directions at once, where it extinguishes itself. The augmented medial axis can then be defined as the location of the shocks together with the times when they were formed. This can be modeled by the differential equation

$$\frac{\partial \mathcal{C}}{\partial t} = -\alpha \mathcal{N} \ . \tag{2.1}$$

Here $\mathcal{C}$ denotes the location of the fire front at time t, $\mathcal{N}$ is the outward facing normal to the front, and α is a positive constant controlling the speed of the flow. This analytic approach was shown to be equivalent to the geometric definition by (Calabi and Hartnett, 1968).

A fourth approach is based on the Euclidean distance transform.

Definition 2.5 *The* ***signed Euclidean distance transform*** *of a set Ω is the function $D_\Omega : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ given by*

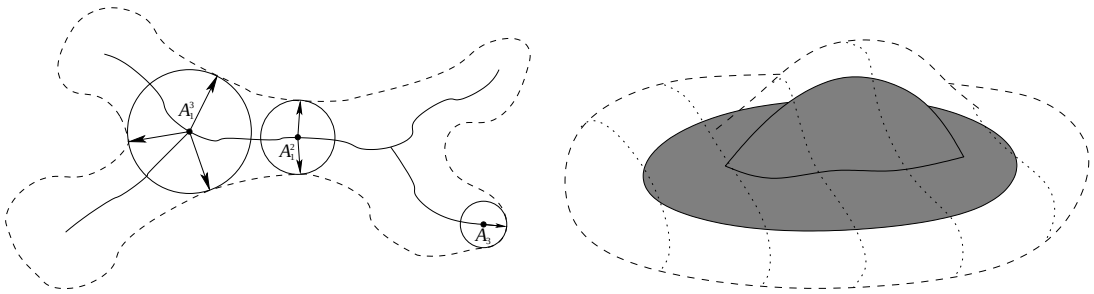
$$D_\Omega(x) = \begin{cases} \inf_{x' \in \Omega} d(x, x'), & x \notin \Omega \\ -\inf_{x' \in \mathbb{R}^n \setminus \Omega} d(x, x'), & x \in \Omega \end{cases} , \tag{2.2}$$

where $d(\cdot, \cdot)$ is the Euclidean distance function.

Then the medial axis is given by the "ridges" of the graph of D_Ω that lie inside Ω. These are given by the closure of the set of singular points, where two distinct points $x' \in \mathcal{B}$ both give the infimum in (2.2).

It is helpful to get a mental picture of what the set M looks like, so Figure 2.1 illustrates an example for two and three dimensions. In two dimensions, the centers of the inscribed balls form a set of curves that meet at *branch points*. In three dimensions, they are typically a set of 2D manifolds with boundary, called *sheets*, which meet along *branch curves*. We will call the boundary of these sheets *edges* and points along them *edge points*, to avoid confusion with the boundary of the object, $\mathcal{B}$. The vectors from the center of a ball to the points of tangency on the boundary have been called in various places *spokes*, *sails*, or *oars*. In this work we shall adopt the term spokes.

These illustrations raise the question of whether or not it is possible to classify how the medial axis is structured. Indeed, in their original work (Blum and Nagle, 1978)



(a) A 2D example.

(b) A 3D example.

Figure 2.1: Examples of the medial axes of some generic shapes. Most points are smooth points, like point A_1^2, with a bitangent circle. Examples of singular points in 2D include branch points, like point A_1^3, with a tritangent circle, and edge points, like point A_3, with an osculating circle. The A_k^n notation is explained in detail in the text.

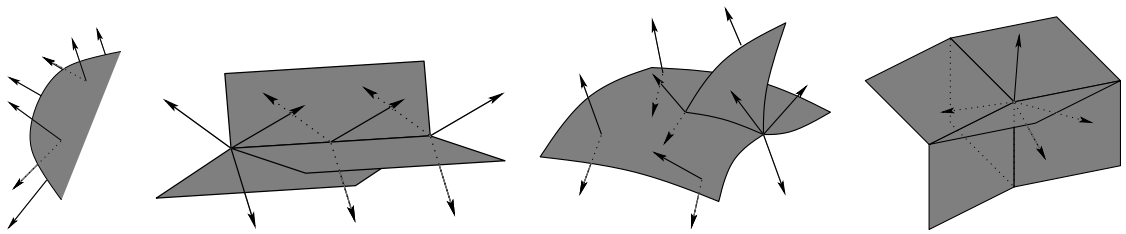
give such a classification for two dimensions, while (Giblin and Kimia, 2004) give a rigorous classification in three dimensions for a *generic* boundary. We shall adopt the notation of the latter, where each point A_k^n is classified by the number of unique points of tangency of the maximally inscribed ball with the boundary (n), along with the order of contact (A_k). The only two possible orders of contact of a maximally inscribed ball $\overline{B}_r(m)$ at a point on the boundary $x \in \mathcal{B}$ are

- A_1 contact: $\overline{B}_r(m)$ is tangent to $\mathcal{B}$ at x.
- A_3 contact: $\overline{B}_r(m)$ is tangent to $\mathcal{B}$ at x, r is one of the principal radii of curvature of $\mathcal{B}$ at x, and the corresponding principal curvature is a local extremum (also called a *ridge point* of $\mathcal{B}$).

There are three other orders of contact, A_2, A_4, and D_4, but at such places the boundary must penetrate the surface of the ball, so it can never be maximally inscribed. Thus no ball on the medial axis will make such contact with $\mathcal{B}$ (but points on the symmetry set might).

This notation is extended directly for contact at more than one point, e.g., $A_1A_1 = A_1^2$ is normal bitangent contact. In 2D, the possible generic cases are

1. Curves (one dimensional manifolds) of A_1^2 points.
2. A_1^3 branch points (zero dimensional manifolds), where three A_1^2 curves meet.
3. A_3 edge points (zero dimensional manifolds), where an A_1^2 curve ends.



(a) Edge. (b) Branch curve. (c) Fin point. (d) Six-junction.

Figure 2.2: The generic local singular structures for the Blum medial axis in $\mathbb{R}^3$.

These can be extended to three dimensions, where two additional cases arise:

1. Sheets (two dimensional manifolds) of A_1^2 points.
2. A_1^3 branch curves (one dimensional manifolds), where three A_1^2 sheets meet.
3. A_3 edge curves (one dimensional manifolds), where an A_1^2 sheet ends.
4. A_1A_3 fin points (zero dimensional manifolds), where an A_1^3 branch curve ends.
5. A_1^4 six-junction points (zero dimensional manifolds), where four A_1^3 branch curves connecting six A_1^2 sheets meet.

These are illustrated in Figure 2.2.

In general, if we want to restrict our attention to genus zero objects—those homeomorphic to the n-dimensional sphere—then we require some restrictions on how the strata are connected to each other. For example, in 2D, it is enough to require that the (undirected) graph of the connections between the medial sheets forms a tree (e.g., contains no cycles). In higher dimensions, this is still sufficient, but not necessary. Rather, the situation is more complex. Not only may the connectivity graph no longer be a tree, but neither is it sufficient to simply describe *which* sheets are connected to each other. Rather, *how* they are connected along branch curves also becomes important. A complete characterization of the topologies of the medial axes of genus zero objects in 3D is given by (Damon, 2005b).

There are many variations on the definition of a medial axis, including (Brady and Asada, 1984; Leyton, 1987), which have different properties at branches and end-points, different geometric relationships with the associated boundary, and different generic structures. There are also numerous methods for computing (or approximating) the

medial axis of a given shape. Different approaches are based on Voronoi skeletons (Székely et al., 1994; Sheehy et al., 1996; Sherbrooke et al., 1996; Amenta et al., 2001; Leymarie and Kimia, 2003), morphological thinning (Borgefors et al., 1999; Manzanera et al., 1999), core tracking (Pizer et al., 1998), and shock detection (Kimia et al., 1990; Siddiqi et al., 2002; Torsello and Hancock, 2006); see (Pizer et al., 2003) for an overview and comparison of most of these. However, the problem is in general ill-posed, as small perturbations in the boundary of the shape can introduce large branches on the medial axis, and the branching structure can vary drastically between similar shapes. It is not clear how meaningful statistics can be performed over shape models with such widely varying topology.

This fundamental difficulty has led researchers to a different approach (Golland et al., 1999; Pizer et al., 1999). In a sense that will be made precise in Section 2.3.5, the effect of these extra branches on the actual shape is very small. Thus, instead of trying to extract the medial axis corresponding to an object, we start with a medial axis with a fixed branching topology and deform it until it matches our target object within a desired tolerance. Starting from a desired tolerance, (Styner et al., 2003) give an automatic method for computing a suitable topology for a collection of objects based on pruning their Voronoi diagrams. The specific incorporation of tolerance defines the scale of the features being modeled, and varying it produces multi-scale methods with all of their advantages.

This reversal of the original relationship, from an object describing a medial axis to a medial axis describing an object, is the critical idea that makes medial representations effective image analysis tools. It replaces an inherently unstable and ill-defined problem—computing the medial axis of an object from its boundary—with a well-defined and stable one: computing the boundary of an object from its medial axis. Proving such stability, in the sense that small changes in the medial axis lead to small changes in the boundary, is beyond the scope of this dissertation, but the derivatives of the boundary in terms of changes to the medial axis given in Appendix B for use in Chapter 4 provide a direct means of computing and bounding such changes. One cannot write derivatives for the reverse, the change in the medial axis for a change in the boundary, since this can introduce topological changes.

2.2 Shape Analysis Using the Medial Axis

As a descriptor of *shape*, the medial axis can be used to provide a detailed quantitative and qualitative analysis that simpler object descriptors, such as volume, surface area, pose, etc., cannot. To this end, the medial axis has been applied to a number of shape analysis tasks by various researchers, as summarized below. The medial axis is certainly not the only shape descriptor found in the literature. Others include landmarks and point distribution models (PDMs) (Cootes et al., 1992; Bookstein, 1997), spherical harmonics (SPHARMs) (Brechbühler et al., 1995; Shenton et al., 2002), deformation fields (e.g., from a standard template or atlas) (Christensen et al., 1993; Martin et al., 1994; Machado and Gee, 1998), or distance transforms (Golland et al., 2002; Leventon et al., 2000). A complete discussion of all of these is beyond the scope of this chapter.

A medial representation, however, has a number of compelling features. Their branching topology provides a natural organization of shapes into connected parts or figures. There is evidence that the human visual system makes use of medial geometry to understand (the 2D projection of) objects (Marr and Nishihara, 1978; Psotka, 1978; Biederman, 1987; Burbeck et al., 1996). These symmetric relationships provide natural ways to measure intuitive object properties such as local thickness, bending, narrowing, and expansion. And finally, with a relatively small number of model parameters these local properties can be used to statistically analyze coherent global shape changes, such as bending, twisting, and tapering.

Näf et al. use the medial axis to identify places where the local bone thickness is large enough to support a prosthetic hip replacement and to describe the sulco-gyral foldings of the cortical surface in the brain (Näf et al., 1996). They construct the medial axis using the Voronoi skeleton and, as is typical of such methods, apply considerable attention to heuristics to prune away the insignificant portions of the result caused by sampling artifacts and noise.

Bouix et al. apply the medial axis to estimate the local width of the hippocampus, a part of the brain that has an important role in memory and learning (Bouix et al., 2005). They compare their results with previous studies on volume loss in males during early adulthood and localize this loss to the head and tail regions. They construct the medial axis using the average outward flux of the Euclidean distance transform to detect the shocks of the grassfire flow (Siddiqi et al., 2002). However, this method operates only on voxelized objects and produces a voxel representation of the medial axis. In order to overcome the limitations of working with discrete surfaces, they project all of

the points into a single 2D plane. This is possible *only* because the medial axis of the hippocampus is dominated by a single, relatively flat medial sheet, with the remaining sheets contributing less than 1% to the total volume. However, in order to perform a statistical analysis, there still remains the problem of identifying *corresponding* points between different subjects. Bouix et al. propose two different approaches to solving this problem. In the first, they rigidly align the projected axes in the plane. However, the variations of that projected shape mean that only a small subset of the points will be available across all subjects, and the correspondence between points aligned in this matter is only weakly related to their true anatomical structure. The second approach uses nonlinear deformations to warp the axis to a common template shape. This might come much closer to producing anatomical correspondences, but the variability in the shape of the medial axis is eliminated. It might be possible to analyze this variability independently by examining the deformation fields, but they have not yet explored methods of doing so.

Zhang et al. use the medial axis for articulated shape matching (Zhang et al., 2005). They also use Siddiqi et al.'s shock detection algorithm to construct the medial axis and classify the voxels according to Giblin and Kimia's taxonomy directly via (Malandain et al., 1991). This gives an automatic partition of the object into "parts": volumetric regions associated with a single medial sheet that are connected via articulated joints along the branch curves. A directed acyclic graph (DAG) represents this connectivity, and the *graph spectra* of these graphs are used to index and search a database of shapes for a similar object. The graph spectra are computed from the first several eigenvalues of a graph's adjacency matrix, which are insensitive to "minor perturbations of graph structure" (Shokoufandeh et al., 2005), which can account for minor topological changes in the medial axis. Major topological instabilities are still reported to hinder performance, e.g., with four-legged animals. Also, the associated object is reconstructed directly via the union of enclosed balls, which does not give an *explicit* link between points on the medial axis and the associated points of tangency on the boundary. This does not hinder their method, which does not require such a link to place an object into a broad category, but it makes the approach unsuitable when one wishes analyze variations in *corresponding* points in multiple objects from the same category. Yushkevich also argues in (Yushkevich, 2003) (pg. 35) that such a reconstruction cannot (directly) produce the symmetry relationships between these boundary points, relationships that contain important information about an object.

Golland et al. use "fixed-topology skeletons" for 2D shape classification. The medial

axis of an object is estimated by starting with a graph that models the connectivity of the branches, embedding this graph into a particular image, an then deforming it using active contour methods (or "snakes") to locate the ridges of the Euclidean distance transform (Golland and Grimson, 2000). It is precisely because the topology is fixed a priori that sensitivity to noise is reduced and statistical comparisons among populations of objects are possible. Classification is done using two features (sampled uniformly in arclength): the angle formed between three consecutive samples and the radius at each sample, using both Fisher linear discriminants and linear Support Vector Machines (Golland et al., 1999). They apply their method to the corpus callosum to distinguish between schizophrenics and healthy controls, and while the cross-validation results generally yielded below 70% accuracy, this may be because the classes are not truly separable. While unsuitable for diagnosis by itself, such a classification could support a diagnosis in conjunction with other evidence.

2.2.1 M-reps

Similar to Golland's "fixed-topology skeletons", (Pizer et al., 1999) introduce a sampled medial representation called (discrete) *m-reps*, which was later extended to three dimensions in (Joshi et al., 2001). The idea is that instead of using a complete representation of the medial axis, a discrete set of samples called *medial atoms* are used instead. Each medial atom $\underline{m} = (m_0, r, n_0, n_1)$ represents a smooth point on the medial axis m_0, the radius of the maximally inscribed ball at that point r, and the two unit spoke vectors pointing towards the tangent points on the boundary n_0 and n_1 (Fletcher, 2004). In 3D, this constitutes 8 intrinsic parameters: three for m_0, one for r, and two each for n_0 and n_1. This gives sufficient information to reconstruct two points on the boundary per medial atom: $m_0 + rn_0$ and $m_0 + rn_1$. In place of the spoke directions, earlier work used a local coordinate frame $F = (b, n, b^\perp)$ and an object angle θ, where $b = \frac{n_0+n_1}{\|n_0+n_1\|}$ is the bisector of the spokes, n is the normal vector $n = \mathcal{N}$, and $b^\perp = b \times n$. The spokes are arranged symmetrically at an angle θ from b in the b-n plane. Both representations have the same number of intrinsic parameters if one uses a unit quaternion, requiring three parameters, to represent the rotation that gives the coordinate frame F, and one can easily convert between the two representations. However, the coordinate frame is degenerate when the spoke directions are nearly parallel, which occurs when the radius is changing slowly. Any rotation of the frame around the normal direction will produce similar, nearly parallel spoke directions. Representing the two

spoke directions explicitly avoids this degeneracy.

At the edges of the medial sheets, *end atoms* are used that contain one extra parameter, an elongation factor $\eta \geq 1$. This is used to add a third boundary point along the crest, at the point $m_0 + \eta rb$. Conceptually, this can be thought of as adding an extra sample point on the edge of the medial axis itself, where there is only one point of tangency to the boundary, but with multiplicity three. However, coupling it with the interior atom just inside the edge in this manner helps avoid the instabilities that might occur from estimating it independently, though this does limit the models to shapes in which the medial sheet remains flat past the end atoms, instead of allowing it to bend right up until the edge.

The medial atoms are typically sampled on a coarse mesh, so the boundary points that can be reconstructed directly are too sparse relative to the voxel spacing in associated volumetric imagery. Thus one needs a method of producing a finer sampling of the boundary. Up until now, research has focused on the method described by (Thall, 2004), which uses a subdivision surface that approximately interpolates the explicitly constructed boundary points, with normals approximately agreeing with the discrete spoke directions. Branching structures are supported using a tree-structure hierarchy of *figures*, which represent separate medial sheets. These are joined via *hinge atoms*, which are positioned relative to their parent, and around which the boundary is blended between the two figures. For details we refer the reader to (Han et al., 2004; Thall, 2004). In 3D, a tree-structure is not sufficient to describe every possible branching topology, but it can represent a much broader class of objects than a single sheet. However, figures can also be used to represent *indentations*, which in the manner of constructive solid geometry are subtracted from the interior of their parent. This is a departure from strictly medial geometry, but its close relation with the more general symmetry set can provide additional flexibility in representation.

The primary drawback of Thall's approach is that it does not respect the underlying medial geometry that produced the explicitly sampled boundary points. The true medial axis of the reconstructed boundary may not have the same branching structure as the original m-rep model, and as described in Chapter 3 of (Yushkevich, 2003), attempting to interpolate the medial axis from such a boundary without adjusting the branching structure can produce an illegal axis that overfolds on itself. If one follows the boundary normals inwards into the shape (for example, to sample image *profiles* near the boundary), they will cross at points on the Blum medial axis. The introduction of additional small branches may move these crossings quite close to the object boundary,

even though the local thickness implied by the m-rep radius suggests one can go quite deep into the object before encountering them. Additionally, if one does not force the boundary to be orthogonal to the spokes at the explicitly sampled boundary points, the true medial axis will not even contain the discrete samples used to generate it. However, (Thall, 2004) shows that enforcing this orthogonality exactly with subdivision surfaces produces fine-scale ripples in the boundary that are not plausibly implied by the coarser underlying shape model. These are precisely the boundary features that generate additional branches in the true medial axis of the reconstruction. Only by relaxing exact orthogonality to fall within a relatively loose tolerance, often set as high as fifty or sixty degrees, can these ripples be eliminated. This method of interpolation was always meant as a temporary stopgap to be used until an interpolation that did respect the medial geometry could be found.

An alternative approach is to interpolate the medial axis itself and reconstruct the boundary from the medial information at every point. This avoids all of the problems described above, but is quite difficult to do in practice due to the interrelations between the medial parameters and the boundary conditions that must hold at the singular points (the edge curves, branch curves, etc.). The first example of such an approach was the *cm-reps* (for "continuous m-reps") described by (Yushkevich, 2003), a B-spline based representation limited to a single medial sheet. The work in this dissertation is heavily based on this approach, but we defer a more detailed discussion to Chapter 3.

More recently, (Han et al., 2006) have attempted to interpolate discrete single-figure m-reps directly, and in as yet unpublished work, multi-figure m-reps. In its current form, it only approximates the spoke directions provided by the original m-rep, but like Yushkevich's cm-reps, overcomes most of the difficulties associated with interpolating the boundary on its own. However, instead of producing a medial axis, it produces a (restricted form of a) *skeletal structure*, foregoing the strict Blum requirement that the spoke directions be normal to the boundary. The next section shall define skeletal structures in more detail. Removing the Blum restriction introduces an ambiguity into the shape representation, because the medial axis describing an object is unique, but the skeletal structure is not.

Some argue that this is a potential advantage due to the highly constrained nature of the parameters of the Blum medial axis. As we will show in the next section, the spoke directions of the Blum medial axis can be derived entirely from m and r. A skeletal structure, on the other hand, must model them independently using additional parameters. Both representations must satisfy numerous conditions to avoid

singularities in the reconstructed boundary. With its additional degrees of freedom, a skeletal structure can more accurately represent a shape without violating these conditions. However, when performing statistics, one generally wants to eliminate unnecessary degrees of freedom. Thus we have taken the philosophical stance that for shape statistics, one wants a unique representation, meaning the medial axis. This is still an area of active research. Only time will tell the real practical merits of the two approaches.

2.3 Mathematical Background

This section covers the necessary mathematical background needed to work with the medial geometry in this thesis in more precise terms, including the differential geometry of the medial axis and the computation of integrals in medial coordinates.

First, let us define the term *generic*. Topologically speaking, a property holds in the generic case if it holds for all the elements of a residual subset of a Baire space. A Baire space, such as $\mathbb{R}^n$, is a topological space where any countable intersection of open and dense sets is dense[1]. Thus, a Baire space is "large" in a way that, for example, a set of isolated points is not. A residual subset is one that contains a countable intersection of open and dense sets, and thus in a Baire space it is itself dense. In layman's terms, this means that something holds *generically* if it can be made true for any object by an arbitrarily small perturbation. Similarly, a generic property of an object is one that does not disappear under any sufficiently small perturbation. Being able to represent generic objects means being able to come arbitrarily close to any degenerate configuration, which is usually sufficient.

2.3.1 Whitney Stratified Sets

With this introduction, we can state that for a generic object Ω, the medial axis M forms a *Whitney stratified set* (Damon, 2003), which has the following definition.

Definition 2.6 *A subset M of a topological space X is **locally closed** if every point $x \in M$ has a neighborhood $N_x \subseteq X$ such that $M \cap N_x$ is closed in N_x.*

Lemma 2.7 *Any open set M is locally closed, as M itself is an (open) neighborhood containing every point in M, and M is always closed in itself.*

[1] Recall that a dense set is one whose closure is the whole space.

Definition 2.8 *A collection of subsets $\{M_\alpha\}_{\alpha \in I}$, where I is an index set, of a topological space X is* ***locally finite*** *if, for every $x \in X$, there exists a neighborhood of x that intersects only a finite number of the subsets M_α.*

Definition 2.9 *A closed set $M \subseteq \mathbb{R}^n$ is a* ***stratified set*** *if, given a partially ordered index set I, M can be written as the union of a locally finite collection of disjoint, locally closed smooth submanifolds $M_\alpha \subseteq \mathbb{R}^n, \alpha \in I$, called* ***strata****, which satisfy the axiom of the frontier: $\overline{M}_\alpha \cap M_\beta \neq \emptyset$ if and only if $M_\beta \subseteq \overline{M}_\alpha$ and only if $\beta \leq \alpha$. Here $\overline{M}_\alpha$ is the closure of M_α in $\mathbb{R}^n$.*

Lemma 2.10 *The strata $\{M_\alpha\}$ also form a partially ordered set, with $M_\beta \leq M_\alpha$ if and only if $M_\beta \subseteq \overline{M}_\alpha$.*

The resulting strata are smooth manifolds "without boundary"; the boundaries are separated into their own strata of lower dimension. Throughout this chapter we will take *smooth* to mean differentiable up to order C^μ, with μ fixed throughout as either ∞ (the normal definition of smooth), or some positive integer. The definitions and theorems are valid in either case.

Definition 2.11 *A* ***Whitney stratified set*** *is a stratified set satisfying the two Whitney conditions. For any pair of strata $M_\beta < M_\alpha$, suppose $\{x_i\} \subseteq M_\alpha$ is a sequence converging to $y \in M_\beta$, and $\{y_i\} \subseteq M_\beta$ is a sequence also converging to y. Let τ be the limit of the sequence of tangent spaces $T_{x_i} M_\alpha$ and let ℓ be the limit of the sequence of secant lines ℓ_i connecting x_i and y_i. Then the two Whitney conditions are*

1. *$T_y M_\beta \subseteq \tau$, and*
2. *$\ell \subseteq \tau$.*

The Whitney conditions are technical requirements that ensure that in some sense each stratum M_α locally "looks the same" at every point. As it turns out, the second condition actually implies the first. In 3D, as long as the local structure of a set is given by one of the cases illustrated in Figure 2.2, it will admit a stratification that satisfies the Whitney conditions. An excellent, brief introduction to stratified sets is provided by (Gunnells, 2003), and the reader interested in a more in-depth introduction is referred to (Mather, 1970).

2.3.2 Skeletal Sets and Radial Vector Fields

The foregoing has prepared the presentation of a number of constructs introduced by Damon that will be indispensable in describing the structure of a medial axis. Most of these definitions are taken directly from (Damon, 2003), with minor adaptations to our notation.

The Whitney stratification allows us to break up the points in M into two subsets. One set, M_{reg}, contains the points in the strata of codimension one (dimension $n-1$): the curves in 2D and the sheets in 3D. These are the highest dimensional strata; a medial axis never has strata of codimension zero (dimension n). Strictly speaking, there is no requirement that these strata be *connected*; however, the stratification can always be refined into connected components, and we assume henceforth that this is done. This document refers to these connected, codimension 1 strata individually as *medial sheets*, regardless of the dimension they appear in.

The other set, M_{sing}, contains all the points in strata of higher codimension (lower dimension): the branch points and end points in 2D, and branch curves, edge curves, etc. in 3D. In other words, M_{reg} contains the "smooth points" of M, while M_{sing} contains the singular points. It is helpful to single out the subset $\partial M \subseteq M_{\text{sing}}$, defined to be the points of M_{sing} where M is locally an $n-1$ dimensional manifold with boundary, as well as its closure, $\overline{\partial M}$. ∂M is the set containing all of the *edges* or A_3 *edge points* of the medial sheets in M_{reg}. $\overline{\partial M}$ also contains combination edge and branch points, such as the A_1A_3 fin points.

We add one final definition so that we can talk about the local geometry of the strata in the neighborhood of a point.

Definition 2.12 *The* ***local neighboring components*** *of a singular point $m_0 \in M_{sing}$ are the portions of the medial sheets M_α adjacent to m_0, in the sense that they are part of the intersection with a small closed ball around m_0, $\overline{B}_\epsilon(m_0) \cap M_{reg}$, where $\epsilon > 0$ is taken sufficiently small so that this is well-defined.*

The Whitney conditions ensure that this can be done. Although this dissertation uses the same notation M_α for a full sheet and for the local neighboring component, the latter is assumed to be a single connected component inside the closed ball. Thus, a sheet M_α that wraps around and abuts itself in a branch curve would create three distinct local neighboring components at the points along that curve.

We are now prepared to review what Damon refers to as a *skeletal set*.

Definition 2.13 *An $(n-1)$-dimensional compact*[2] *Whitney stratified set $M \subset \mathbb{R}^n$ is a **skeletal set** if it satisfies the following conditions:*

1. *Given a point $m_0 \in M_\beta$, where $M_\beta \subseteq M_{sing}$, for each local neighboring component M_α, there is a unique limiting tangent space $T_{m_0}M_\alpha$ from all sequences of points in M_α. By the first Whitney condition, $T_{m_0}M_\beta \subset T_{m_0}M_\alpha$.*

2. *For any point $m_0 \in M_{sing}$, for some sufficiently small neighborhood $N_{m_0} \in \mathbb{R}^n$, the set $M \cap N_{m_0}$ can be written as the union of a set of smooth $n-1$ dimensional manifolds M_γ with boundaries and corners, called the **local manifold components** for m_0, where any pair of manifolds intersect only at these boundary facets.*

3. *If $m_0 \in \overline{\partial M}$, the local manifold components M_γ of m_0 that intersect ∂M do so in an $n-2$ dimensional facet.*

The first condition prevents situations like that found in the Whitney umbrella, which is illustrated in Figure 2.3. The "peak" of the umbrella forms a *pinch point*, which does not have a single, well-defined tangent plane. It is a well-known, but non-trivial fact that such a point cannot occur on the medial axis of a generic object.

The second condition ensures that M contains no degenerate pieces. For example, the medial axis of a circle in 2D is a single point, which is *not* part of a 1D manifold with boundary. Similarly, degenerate curves are produced by tubes in 3D, etc. Such degeneracies are not generic, in that a small perturbation of the shape makes them non-tubular. It may still be useful to create specialized shape models to handle them in order to avoid instabilities in *how* such a perturbation is chosen, as in (Damon, 2006). However, this dissertation focuses on the general case.

The third condition ensures that two pieces do not meet in a degenerate way. For example, if two medial sheets meet in 3D, they must do so along one-dimensional branch curve and not just at a zero-dimensional point.

The skeletal set, however, is just one part of the medial axis. We also need to define how the points on the medial axis relate to the boundary of the object. For the Blum medial axis, this is done in terms of the spokes, the vectors that lie normal to the boundary, but we start with a more general structure called the *radial vector field*, S. This vector field is defined over the set M and, like a number of other constructs we will use, is *multivalued*. That is, it has multiple values at each point and not necessarily the same number of them at every point (cf. the arrows in Figure 2.1(a)).

[2]Recall that in $\mathbb{R}^n$, a set is compact if and only if it is closed and bounded.

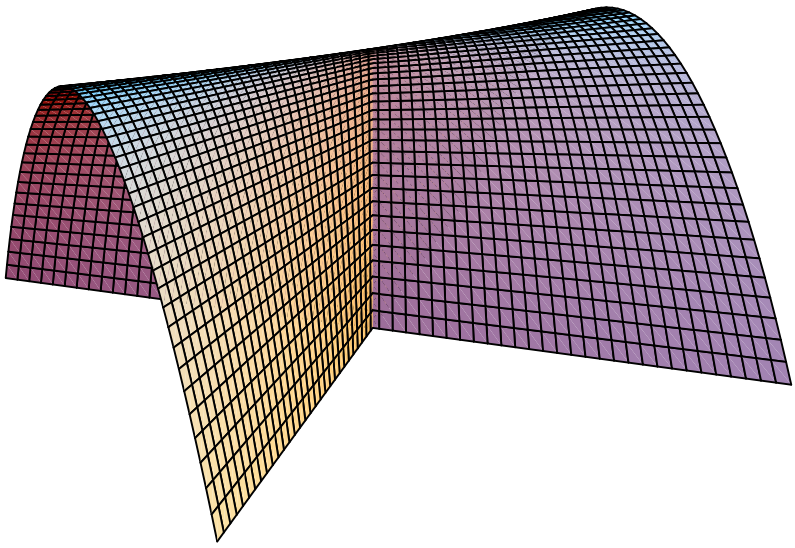


Figure 2.3: The Whitney umbrella is an example of a stratified set that fails the first condition required of a skeletal set. It is given by the equation $(x, y, z) = (uv, u, 1 - v^2)$ over the domain $(u, v) \in [-1, 1]^2$.

The purpose of the radial vector field is to give a mapping from a skeletal set to the object boundary. At regular points on the skeletal set, there are always two radial vectors, which correspond precisely to two points of tangency on the bitangent spheres. One corresponds to the "top" of the medial sheet, and the other corresponds to the "bottom". However, smoothly extending the vector field to singular points may require one, three, or even more vector values at a single point.

We adopt a slightly non-standard definition of *smooth* at edge points to overcome a problem of parameterization. In order to have a closed boundary, at edge points the top vector field must continuously join with the bottom vector field on the corresponding smooth manifold component, $\overline{M}_\alpha$. However, like the function $\sqrt{x}$ evaluated at $x = 0$, on a Blum medial axis the resulting vector fields are not differentiable on $\overline{M}_\alpha$, because the slope lies perpendicular to the manifold. Damon uses an *edge coordinate parameterization* to overcome this limitation.

Definition 2.14 *An **edge coordinate parameterization** at an edge closure point $m_0 \in \overline{\partial M}$ lying on the boundary of the closure of a local neighboring component $\overline{M}_\alpha$ consists of an open neighborhood $N_{m_0} \in \overline{M}_\alpha$, an open neighborhood $\tilde{N}_0 \in \mathbb{R}^{n-1}_{\geqq 0}$, and a differentiable homeomorphism $\phi : \tilde{N}_0 \to N_{m_0}$ such that both ϕ restricted to $(\tilde{N}_0 \cap \mathbb{R}^{n-1}_+)$*

and to $(\tilde{N}_0 \cap \mathbb{R}^{n-2})$ *are diffeomorphisms.*

The $(\tilde{N}_0 \cap \mathbb{R}^{n-1}_+)$ restriction corresponds to the parameterization of the local piece of the neighboring manifold M_α, while $(\tilde{N}_0 \cap \mathbb{R}^{n-2})$ corresponds to the parameterization along the edge curve. Then we say a scalar or vector field is smooth at an edge closure point if the composition with an edge parameterization is smooth. As a trivial example, $\sqrt{x}$ is smooth on $x \in \mathbb{R}_{\geq 0}$ under the edge parameterization $\phi(u) = u^2$. In general, any smooth function on $\overline{M}_\alpha$ is smooth under an edge parameterization, but as the $\sqrt{x}$ example illustrates, the converse is not true.

Describing how smooth extensions work at non-edge singular points requires the introduction of *local complementary components*.

Definition 2.15 *Given* $m_0 \in M_{sing}$ *and a closed ball* $\overline{B}_\epsilon(m_0) \subset \mathbb{R}^n$*, let* C_i *be the connected components of* $\overline{B}_\epsilon(m_0) \setminus M$*. Again* $\epsilon > 0$ *is chosen to be sufficiently small so that these are well defined (up to a homeomorphism). These are the* ***local complementary components*** *of* m_0*.*

Definition 2.16 *Given* $m_0 \in M_{sing}$ *with local complementary components* C_i *and local neighboring components* M_α*, let* $\partial C_i \triangleq \{M_\alpha : M_\alpha \cap \overline{C}_i \neq \emptyset\}$*. These are the local neighboring components adjacent to* C_i*.*

Definition 2.17 *We say a vector* u ***points into*** *a local complementary component* C_i *from a point* m *if there exists an* $\epsilon > 0$ *such that* $m + tu \in C_i$ *for all* $0 < t < \epsilon$*.*

The idea will be to ensure that every local complementary component has exactly one radial vector pointing into it at each singular point m_0. This, along with smoothness constraints, ensures that the radial vectors do not poke through one of their neighboring strata, at least locally. Of course if the vector is too long, it may poke through a non-local piece of the medial axis, but in so doing it must of necessity cross other radial vectors. Later, we will show how to detect such crossings. But first, we give Damon's complete definition of a radial vector field.

Definition 2.18 *Given a skeletal set* $M \subset \mathbb{R}^n$*, a* ***radial vector field*** S *on* M *is a nowhere zero multivalued vector field satisfying the following conditions.*

1. *(Behavior at smooth points) Let* $\mathcal{N}$ *be a smooth unit normal vector field over each of the medial sheets in* M_{reg}*. Then for each smooth point* $m_0 \in M_{reg}$*,* S *has exactly two values,* S^+ *and* S^-*, such that* $S^+ \cdot \mathcal{N}(m_0) > 0$ *and* $S^- \cdot \mathcal{N}(m_0) < 0$*. Moreover, in a neighborhood of* m_0 *in* M_{reg}*,* S^+ *and* S^- *each form a smooth vector field.*

2. *(Behavior at non-edge singular points) Let $m_0 \in M_{sing} \setminus \partial M$ be a singular point with local neighborhood component M_α. Then S^+ and S^- on M_α each extend smoothly to a value of $S(m_0)$. If there exists a neighborhood N_{m_0} such that $N_{m_0} \cap \overline{M}_\alpha \cap \partial M = \emptyset$, then that value $S(m_0) \notin T_{m_0} M_\alpha$. Conversely, for each value of $S(m_0)$, there is exactly one local complementary component C_i of M at m_0 such that the value $S(m_0)$* ***locally points into*** *C_i in the following sense. Let $M_\alpha \in \partial C_i$ be a local neighboring component adjacent to C_i and let N_{m_0} be an open neighborhood chosen sufficiently small so that $N_{m_0} \cap M_\alpha \subset \overline{C_i}$. Then for all $m \in N_{m_0} \cap M_\alpha$, either $\mathcal{N}$, $-\mathcal{N}$, or both point into C_i. Let S' be a restriction of S to the values where the corresponding normal direction points into C_i, along with the value of $S(m_0)$ under consideration. Then $S(m_0)$ smoothly extends to the values of S' on M_α. Furthermore $S'(m_0)$ points into C_i and for all $m \in N_{m_0} \cap M_\alpha$, $S'(m)$ points into C_i.*

3. *(Tangency behavior at edge points) Let $m_0 \in \partial M$ be an edge point and $M_\alpha \subseteq M_{reg}$ be the stratum whose edge it is, i.e., $m_0 \in \overline{M}_\alpha$. Then S has exactly one value $S(m_0)$ such that $S(m_0) \in T_{m_0} \overline{M}_\alpha$, and $S(m_0)$ points away from M. This value extends smoothly into M_α (under some edge parameterization).*

The radial vector field corresponds to the spokes linking the centers of the maximally inscribed balls to their points of tangency with the boundary in the Blum medial axis. In addition to the constraints in its definition, two *local initial conditions* must be satisfied:

1. (Local Separation Property) For each a non-edge point $m_0 \notin \partial M$, let M_α be the local neighboring components, and let C_i be the local complementary components contained in the ball $B_\epsilon(m_0)$. Then for all $m \in \partial \overline{M}_\alpha$ on the boundary of the closures of the neighboring component with radial vector $S(m)$ locally pointing into a given C_i, the set $\{m + tS(m) : t \in \mathbb{R}_{\geq 0}\} \cap B_\epsilon(m_0) \subset C_i$.

2. (Local Edge Property) For each edge closure point $m_0 \in \partial \overline{M}$, there exists a neighborhood $N_{m_0} \subset M$ and an $\epsilon > 0$ such that for each choice of smooth vector fields S over W, the function $\psi(m, t) = m + tS(m)$ is one-to-one on $N_{m_0} \times [0, \epsilon]$.

These two conditions are necessary to account for the fact that a different ϵ value can be used at every point to determine if a vector points into a particular component. As one approaches a given point m_0, these choices of ϵ can converge to zero, with the result

that in a neighborhood around m_0, for any fixed choice of ϵ one can find a spoke that pierces the skeletal set before traveling a distance ϵ along it.

A skeletal set together with a radial vector field satisfying the local initial conditions is called a *skeletal structure*. It is a far more general structure than a Blum medial axis, since the radial vectors are not explicitly required to be normal to the boundary, nor even the same length on each side. Several theorems presented later will apply to all skeletal structures, not just the Blum case.

2.3.3 The Blum Medial Axis as a Skeletal Structure

A skeletal structure, as defined, does not have a radius associated with each point, but we can partition the radial vector field into two parts: $S = rU$, where $r = \|S\|$ is a positive function (obeying smoothness constraints similar to those on S), and U is a unit radial vector field. On the smooth parts of M, we will use $U^{\pm}$ to denote the two unit radial vectors on either side. In general, r is multivalued, but for a Blum medial axis, all of the values at a single point must be equal, e.g., all the spoke values must be the same length. Also, r is allowed to be zero in Definition 2.4, but in a skeletal structure it must always be positive. The only places that the radius would go to zero are sharp, non-smooth corners or creases. These are not generic, as they make the boundary undifferentiable at those points, so they are intentionally disallowed.

On the Blum medial axis the radial vectors correspond to the points where the maximally inscribed balls are tangent to the boundary of the object. This gives rise to relation between r and U described by the *compatibility one-form*[3]. It is defined as $\eta_U \triangleq \omega_U + dr$, where $\omega_U(v) \triangleq U \cdot v$ is the one-form dual of U, and dr is the one-form defined by the directional derivative of r. Like U, η_U is multivalued. Lemma 6.1 of (Damon, 2003) states that if $\eta_U \equiv 0$ at a point m_0, then $U(m_0)$ is orthogonal to the boundary at $m_0 + rU(m_0)$. Here $\eta_U \equiv 0$ means that for all $v \in T_{m_0}M$, $\eta_U(v) = 0$. The tangent space at singular points is chosen to correspond to the limiting tangent space of a local neighboring component for which $U(m_0)$ is a smooth extension. There can be more than one such local neighboring component for a particular value of $U(m_0)$ and hence more than one tangent space and more than one value of η_U. $\eta_U \equiv 0$ holds only if $\eta_U(v) = 0$ holds for *all* choices of η_U and limiting tangent spaces at m_0. Damon terms this the *partial Blum condition*. By itself it is sufficient to describe the geometry of the boundary of an object, even when r or dr has multiple distinct values at a point.

[3] A *one-form* is a linear function ω^1 mapping a vector to a scalar, $\omega^1(v) : \mathbb{R}^n \to \mathbb{R}$, e.g., it has "one slot" to fill. One forms are the duals of vectors.

We can use the partial Blum condition to derive an explicit formula for U in terms of r on the smooth points M_{reg}. The immediate implication is that the component of U in the tangent plane is $-\nabla r$, where $\nabla r \in T_{m_0}M$ is the Riemannian gradient[4]. This vector lies in the tangent space $T_{m_0}M$ and is independent of the parameterization of r and M. Since $U(m_0)$ is a unit vector, this leaves only two possible choices for the component normal to the tangent plane, one for each side:

$$U^{\pm} = -\nabla r \pm \sqrt{1 - \|\nabla r\|^2} \cdot \mathcal{N} \tag{2.3}$$

This definition can then be smoothly extended to points in M_{sing}. It is independent of dimension and generalizes earlier formulas given by (Blum and Nagle, 1978) for two dimensions, and (Nackman, 1982) and (Giblin and Kimia, 2004) for three.

On the Blum medial axis, where r and thus ∇r are single-valued, the vector $(U^+ - U^-)$ points in the normal direction $\mathcal{N}$ of M, and the bisector $\frac{1}{2}(U^+ + U^-)$ lies in the tangent space $T_{m_0}M$. Furthermore, at edge points, by Definition 2.18(3) one must have $U(m_0) \in T_{m_0}M$, implying the normal component vanishes. Hence, at these points we have

$$\|\nabla r\| = 1 \ . \tag{2.4}$$

In the complementary component C_i of a non-edge singular point $m_0 \in M_{\text{sing}}$, each of the local neighboring components $M_\alpha \in \partial C_i$ must contribute a spoke $S(m_0)$ pointing into C_i by smooth extension, and all of these $S(m_0)$ values must be equal. Using (2.3), we can write this as a constraint on ∇r in each of the local neighboring components. For example, on branch curves in 3D, given three local components M_k oriented so that $S^{(k)-}(m_0) = S^{(k\oplus 1)+}(m_0)$, the restriction becomes (Yushkevich et al., 2003)

$$\nabla r^{(k\oplus 2)} - \nabla r^{(k\oplus 1)} = \mathcal{N}^{(k)} \cdot \sqrt{1 - \|\nabla r^{(k)}\|^2} \ , \tag{2.5}$$

where $\oplus$ denotes addition modulo 3. At fin creation points, $\|\nabla r^{(k)}\|$ goes to 1 for one of the sheets, causing the angle between the other two to go to π, and the constraint in (2.5) to disappear.

Enforcing the two conditions in (2.4) and (2.5) will become the primary focus of Chapter 3.

[4]The vector dual of the directional derivative on manifolds.

2.3.4 Geometry of the Boundary

Given a skeletal structure (M, S), we have only guaranteed that the radial vectors S do not cross in a small neighborhood of the space around M. In order to ensure that the boundary associated with a skeletal structure is smooth, we must ensure that no such crossings occur anywhere along the entire length of the spokes. These can be classified into *local intersections* and *non-local intersections* (Hoffman and Vermeer, 1994). Non-local intersections are caused when two spokes separated by a non-trivial distance intersect, e.g., when one end of an object wraps around and abuts another. These are difficult to describe mathematically but for most objects are relatively easy to avoid during the modeling process by using geometric priors. Local intersections, on the other hand, occur when two spokes that originate from within an arbitrarily small neighborhood of each other on M cross somewhere along their length (not necessarily close to M), causing the boundary to fold over on itself, as illustrated in Figure 2.4. These can be detected by using a *radial shape operator*, S_{rad}, on non-edge points and a similar *edge shape operator*, S_E, on edge points.

The radial shape operator is not an ordinary shape operator in the differential geometric sense. The traditional shape operator measures how the unit normal $\mathcal{N}$ changes for an infinitesimal step along the surface of a manifold; it is a linear map that can be expressed as a self-adjoint matrix for any choice of orthonormal basis. Its eigenvalues are called the principal radial curvatures.

The radial and edge shape operators, in contrast, give the components of the change in the radial vectors U for such an infinitesimal step on M and ∂M, respectively. They are linear operators but are not in general self-adjoint. The eigenvalues of S_{rad} are termed the *principal radial curvatures* κ_{ri}, and the generalized eigenvalues of S_E the *principal edge curvatures* κ_{Ei}. Their definition now proceeds.

Definition 2.19 *Given a non-edge point $m_0 \in M \setminus \partial M$ of M and a choice of smooth values of S on a stratum M_α or the closure of a local neighboring component $\overline{M}_\alpha$ containing m_0 with tangent space $T_{m_0}M$, the* ***radial shape operator*** *$S_{rad} : T_{m_0}M \to T_{m_0}M$ is defined as*

$$S_{rad}(v) = -\operatorname{proj}_U\left(\frac{\partial U}{\partial v}\right) . \tag{2.6}$$

Here, proj_U denotes projection onto $T_{m_0}M$ along U, which is in general not orthogonal projection. Like U, the radial shape operator is multivalued. At smooth points, it

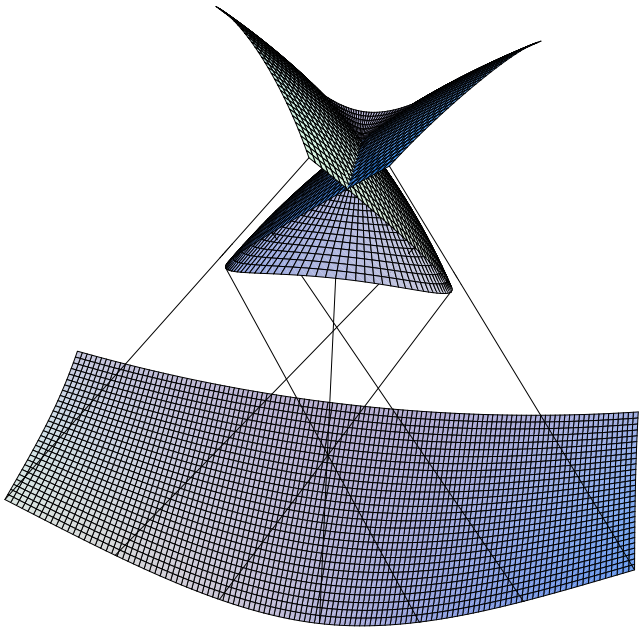


Figure 2.4: An example of overfolding, causing a swallowtail singularity in the boundary. The top surface is one half of the boundary, the bottom is the medial sheet, and the lines connecting them denote the spoke field. The radial shape operator detects this overfolding everywhere in the bottom segment of the swallowtail, where the spokes exhibit *retrograde motion*: a step forward along the medial axis causes a step backwards along the boundary.

will have two values, one for each value of U, and at non-edge singular points $m_0 \in M_{\text{sing}} \setminus \partial M$ each local neighboring component M_α will contribute two values via the smooth extension of U onto m_0.

For a given (not necessarily orthonormal) basis of $T_{m_0}M$, $\mathbf{v} = \{v_1, \ldots, v_{n-1}\}$, S_{rad} can be represented by an $(n-1)\times(n-1)$ matrix $S_{\mathbf{v}}$. The precise details of calculating $S_{\mathbf{v}}$ for a parameterized surface in 3D are given in Appendix A. Because U is not necessarily normal and the projection is not orthogonal, $S_{\mathbf{v}}$ will not in general be self-adjoint, so its eigenvalues need not be real. Nevertheless, these eigenvalues, which are independent of the choice of basis, form what are termed the principal radial curvatures κ_{ri}. If the partial Blum condition is satisfied in a local neighborhood, then the eigenvalues *are* real and have a one-to-one correspondence with the principal curvatures of the boundary,

κ_i, given by

$$\kappa_i = \frac{\kappa_{ri}}{1 - r\kappa_{ri}} \ , \qquad\qquad \kappa_{ri} = \frac{\kappa_i}{1 + r\kappa_i} \ . \tag{2.7}$$

This holds so long as $\frac{1}{r}$ is not an eigenvalue of S_{rad}, which corresponds to a non-smooth singularity of the boundary.

We can also define radial analogues of the mean curvature and Gaussian curvature, namely the *mean radial curvature* H_{rad} and the *Gaussian radial curvature* K_{rad}. They are defined exactly like their classical counterparts, except using S_{rad} instead of the traditional shape operator:

$$H_{\text{rad}} \triangleq \frac{1}{2}\operatorname{trace}(S_{\text{rad}}) \ , \qquad\qquad K_{\text{rad}} \triangleq \det\left(S_{\text{rad}}\right) \ . \tag{2.8}$$

Again, these numbers are the same irrespective of the choice of basis.

Every operation in this dissertation involving S_{rad} in three dimensions can be defined in terms of these two numbers. The principal radial curvatures κ_{ri} can be recovered via

$$\kappa_{ri} = H_{\text{rad}} \pm \sqrt{{H_{\text{rad}}}^2 - K_{\text{rad}}} \ . \tag{2.9}$$

And the usual relations hold:

$$H_{\text{rad}} = \frac{\kappa_{r1} + \kappa_{r2}}{2} \ , \qquad\qquad K_{\text{rad}} = \kappa_{r1}\kappa_{r2} \ . \tag{2.10}$$

We now turn to the edge shape operator, S_E.

Definition 2.20 *Given an edge point $m_0 \in \overline{\partial M}$ lying in the boundary of the closure of a local neighboring component $\overline{M}_\alpha$ with tangent space $T_{m_0}M$, a choice of smooth values of U on $\overline{M}_\alpha$, and a unit normal vector field $\mathcal{N}$ on $\overline{M}_\alpha$, the* ***edge shape operator*** *$S_E : T_{m_0}M \to T_{m_0}\overline{\partial M} \oplus \langle \mathcal{N} \rangle$ is defined as*

$$S_E(v) = -\operatorname{proj}'_U(\frac{\partial U}{\partial v}) \ . \tag{2.11}$$

Now proj'_U denotes projection onto $T_{m_0}\overline{\partial M} \oplus \langle \mathcal{N} \rangle$ along U, which again is not in general orthogonal. The edge shape operator is unusual in that it changes bases from the source to the target. However, this is precisely what is needed to overcome the non-differentiability of U on the edge of $\overline{M}_\alpha$. The matrix representation $S_{E\mathbf{v}}$ starts with

a (not necessarily orthonormal) basis $\tilde{\mathbf{v}} = \{v_1, \ldots, v_{n-2}\}$ of $T_{m_0}\overline{\partial M}$ and adds a vector v_{n-1} that points in the $U(m_0)$ direction under the edge coordinate parameterization to form a basis $\mathbf{v} = \{v_1, \ldots, v_{n-1}\}$ of $T_{m_0}M$. The basis of the destination space is then taken as $\begin{pmatrix} \tilde{\mathbf{v}} \\ \mathcal{N} \end{pmatrix}$. Then, letting $\mathbf{I}_{n-2,1}$ denote the $(n-1) \times (n-1)$ diagonal matrix with a one in the first $(n-2)$ diagonal elements and zeros everywhere else, the principal edge curvatures κ_{Ei} are defined as the generalized eigenvalues[5] of $(S_{E\mathbf{v}}, \mathbf{I}_{n-2,1})$.

Suppose we are given a skeletal structure (M, S) satisfying the following conditions:

1. (Radial Curvature Condition) For all points $m_0 \in M \setminus \partial M$, $r < \frac{1}{\kappa_{ri}}$ for all real, positive principal radial curvatures κ_{ri}.

2. (Edge Curvature Condition) For all points $m_0 \in \overline{\partial M}$, $r < \frac{1}{\kappa_{Ei}}$ for all real, positive principal edge curvatures κ_{Ei}.

3. (Compatibility Condition) For all singular points $m_0 \in M_{\text{sing}}$, $\eta_U \equiv 0$.

Theorem 2.5 of (Damon, 2003) states that if the associated boundary $\mathcal{B}$ has no non-local intersections, it is an embedded manifold that is smooth at all points except those corresponding to points in M_{sing}. At points corresponding to the singular points M_{sing}, it has G^1 continuity, i.e., it has a well-defined tangent plane.

The compatibility condition is just the partial Blum condition restricted to singular points. If we add the partial Blum condition everywhere and restrict r to be single-valued, we obtain the following lemma.

Lemma 2.21 *Suppose (M, S) is a skeletal structure satisfying the radial and edge curvature conditions and the partial Blum condition and further suppose that the associated radius function r is single-valued. Then if the associated boundary $\mathcal{B}$ has no non-local intersections, it is an embedded manifold that is smooth at all points except those corresponding to points in M_{sing}, where it has G^1 continuity. Furthermore, if every ball defined by M is contained in the interior of the object, M is the Blum medial axis of $\mathcal{B}$.*

This is the main result of this section. Given a generative model for a medial axis, it tells us what conditions need to be checked to ensure the associated boundary is legal, in the sense that it could be the boundary of a real, physical object. We start with a single-valued r and derive U and hence S from (2.3), ensuring that the partial

[5]The generalized eigenvalues of a pair of matrices $(\mathbf{A}, \mathbf{B})$ are the values λ such that $\mathbf{A} - \lambda\mathbf{B}$ is singular.

Blum condition holds by construction. The most difficult part is then ensuring that the behavior of U at singular points is sufficient to make it a valid radial vector field, which is the subject of Chapter 3. Then all that remains is to verify the radial and edge curvature conditions, and check for non-local intersections. The first is done as the model evolves to fit a particular shape in Chapter 4, and the assumption that our shapes are sufficiently well-behaved makes checking the second unnecessary.

Because the edge shape operator requires an edge coordinate parameterization, it is more difficult to deal with than the radial shape operator. However, since almost all of the points on the medial axis are smooth points, in general it is substantially less useful. The edge curvature condition is technically required for a smooth boundary, but so long as a sufficient margin of safety is kept in the radial curvature conditions, it cannot be violated. Therefore the remainder of this work will not devote a great deal of effort to its analysis.

2.3.5 Integrals Over Skeletal Structures

The ability to compute integrals over an object Ω and its boundary $\mathcal{B}$ using medial coordinates is an invaluable tool. Trivial examples include computing the volume, center of mass, or second order moment of an object defined by its medial axis. These are later used to align two or more objects. More complex examples include the geometry to image match function used to fit a model to a given shape, as in Chapter 4. Because the model is defined via its medial axis, the medial coordinate system provides a natural way to parameterize these integrals. Damon shows how to define a medial measure dM that makes M into a measure space and consequently how to write formulas for these integrals (Damon, 2005a).

One conceptual difficulty that arises is that integration must take place on "both sides" of M. To do this, Damon introduces the "double" of M, denoted $\tilde{M}$.

Definition 2.22 *The **double** of a skeletal structure (M, S) is the set $\tilde{M} = \{(m_0, S_0) \in \mathbb{R}^n \times \mathbb{R}^n : m_0 \in M$ and S_0 is a value of S at $m_0\}$.*

The vectors S_0 serve to label each "side" of the smooth strata in M_{reg}. As the notation indicates, points of $\tilde{M}$ should be thought of as a point on the medial axis together with a direction, and not as the associated point on the boundary, $m_0 + S_0$.

The definition of a topology on $\tilde{M}$ proceeds from a system of neighborhoods.

Definition 2.23 *Let $(m_0, S_0) \in \tilde{M}$ be a point in the double of M, W' be an open neighborhood of m_0 in $\mathbb{R}^n$, and let C_i be the unique local complementary component of*

m_0 into which S_0 points. Then the ***abstract neighborhood*** *of (m_0, S_0) is given by the set $W = \{(m, S') \in \tilde{M} : m \in W' \cap \partial C_i$ and S' points into $C_i\}$.*

Then $\tilde{M}$ can be decomposed into a Whitney stratified set so that the canonical projection $\pi : \tilde{M} \to M$ given by $\pi(m_0, S_0) = m_0$ is continuous and smooth with respect to this neighborhood system.

The definition of the skeletal integral of a multivalued function f on M now follows from assuming that f has one value corresponding to each value of S at a point $m_0 \in M$, so that f lifts to a well-defined function $\tilde{f} : \tilde{M} \to \mathbb{R}$ given by $\tilde{f} = f \circ \pi$. Then Proposition 2.2 of (Damon, 2005a) states that there is a unique regular positive Borel measure dM on $\tilde{M}$ such that the integral of f on M is given by

$$\int_{\tilde{M}} \tilde{f}\, dM \ . \tag{2.12}$$

The measure dM is called the *medial measure*, and it is given by $dM = \rho\, dA$, where dA is the ordinary $(n-1)$-dimensional differential unit of area on M, and $\rho = \pm U^{\pm} \cdot \mathcal{N}$ on points in M_{reg}. Values on M_{sing} are defined by smooth extension from the local neighboring components. At edge points ∂M it is exactly zero, and the total measure of the whole set M_{sing} is also zero, meaning that so long as f is bounded on M_{sing}, the singular points may be ignored for purposes of integration. For a generic skeletal structure, ρ is multivalued, but on the Blum medial axis it is actually single-valued over M_{reg}, thanks to the symmetry of $U^{\pm}$.

The factor ρ corrects for the failure of S to be orthogonal to M. Thus, areas where S is nearly tangent to M are "small" with respect to this measure. For example, adding a small bump to the boundary of an object can create a spurious branch on the medial axis, but all of the spokes will be nearly tangent to this branch, making their contribution to any integral small.

This integral has the usual linearity and positivity properties, e.g., $\int_{\tilde{M}} \tilde{f} + \tilde{g}\, dM = \int_{\tilde{M}} \tilde{f}\, dM + \int_{\tilde{M}} \tilde{g}\, dM$ and if $\tilde{f} \geq 0$, $\int_{\tilde{M}} \tilde{f}\, dM \geq 0$. We say that f is measurable and integrable on M if $\tilde{f}$ is measurable and integrable on $\tilde{M}$. Damon calls this a *skeletal integral* and, in the special case where (M, S) satisfies the partial Blum condition, a *medial integral*.

This shows how to compute integrals of functions defined on $\tilde{M}$, but typically interesting functions are defined on Ω or $\mathcal{B}$. To compute these, associate a *normal line bundle* N with $\tilde{M}$ to use as the domain for integration. A vector bundle on a manifold is a set of vector spaces associated with the points on the manifold that together also

have a manifold structure. A line bundle is a rank one vector bundle, meaning that each of these vector spaces has dimension one. In our case, the vector space associated with (m_0, S_0) is spanned by S_0. The word "normal" appears by convention, even though this vector space is not in general orthogonal to M. Locally N looks like $\tilde{M} \times \mathbb{R}$, with a point in N denoted (m_0, tS_0) for $t \in \mathbb{R}$. From this domain, the *radial vector flow* mapping in the opposite direction of the grassfire flow, from the medial axis to the boundary, yields the object interior:

Definition 2.24 *The* ***radial vector flow*** $\tilde{\psi} : N \to \mathbb{R}^n$ *is the mapping given by* $\tilde{\psi}(m_0, tS_0) = m_0 + tS_0$.

Unlike the grassfire flow, the rate of flow at each point is not uniform but rather is controlled by the radius function r. Thus the flow reaches the boundary everywhere at time $t = 1$. This is denoted by the special function $\psi_1 : \tilde{M} \to \mathcal{B}$, where $\psi_1(m_0, S_0) \triangleq m_0 + S_0$.

Now given a Borel measurable and Lebesgue integrable function $g : \Omega \to \mathbb{R}$, let

$$\tilde{g}(m_0, S_0) = \int_0^1 g(m_0 + tS_0) \cdot \det\left(\mathbf{I} - t \cdot rS_{\text{rad}}\right) \, dt \tag{2.13}$$

if the integral is defined. Then Theorem 6 of (Damon, 2005a) says that $\tilde{g}$ is defined for almost all $(m_0, S_0) \in \tilde{M}$, it is integrable on $\tilde{M}$, and

$$\int_\Omega g \, dV = \int_{\tilde{M}} \tilde{g} \cdot r \, dM \ . \tag{2.14}$$

This converts an integral over the object Ω into a skeletal integral over $\tilde{M}$. The expression $\det\left(\mathbf{I} - t \cdot rS_{\text{rad}}\right)$ is derived from the Jacobian of the radial vector flow given in (Damon, 2003).

If (M, S) satisfies the partial Blum condition, a similar integral may be defined on the boundary $\mathcal{B}$. If $h : \mathcal{B} \to \mathbb{R}$ is Borel measurable and integrable, then Theorem 1 of (Damon, 2005a) states that

$$\int_{\mathcal{B}} h \, d\mathcal{B} = \int_{\tilde{M}} \tilde{h} \cdot \det\left(\mathbf{I} - rS_{\text{rad}}\right) \, dM \ , \tag{2.15}$$

where $\tilde{h} = h \circ \psi_1$. This converts an integral over the boundary $\mathcal{B}$ into a medial integral over $\tilde{M}$.

A useful expansion of these integral formulas can be obtained in terms of r via the

elementary symmetric polynomials of the principal radial curvatures κ_{ri}. The elementary symmetric polynomials σ_{rj} are the sums of all combinations of a fixed number of the radial curvatures, e.g.,

$$\sigma_{r0} \triangleq 1 \ , \tag{2.16}$$

$$\sigma_{r1} \triangleq \sum_{1 \le i < n} \kappa_{ri} \ , \tag{2.17}$$

$$\sigma_{r2} \triangleq \sum_{1 \le i < j < n} \kappa_{ri} \kappa_{rj} \ , \tag{2.18}$$

$$\vdots$$

$$\sigma_{rn-1} \triangleq \prod_{1 \le i < n} \kappa_{ri} \ . \tag{2.19}$$

Then we have

$$\det\left(\mathbf{I} - rS_{\text{rad}}\right) = \sum_{j=0}^{n-1} (-1)^j \sigma_{rj} r^j \ . \tag{2.20}$$

In three dimensions σ_{r1} and σ_{r2} have particularly simple forms, namely $\sigma_{r1} = 2H_{\text{rad}}$ and $\sigma_{r2} = K_{\text{rad}}$. Thus

$$\det\left(\mathbf{I} - rS_{\text{rad}}\right) = 1 - 2rH_{\text{rad}} + r^2 K_{\text{rad}} \ . \tag{2.21}$$

This can be used to expand $\tilde{g}(m_0, S_0)$ to obtain exact formulas for simple functions g. Letting

$$m_j(g)(m_0, S_0) \triangleq \int_0^1 g(m_0 + tS_0) \cdot t^j \, dt \ , \tag{2.22}$$

we can expand $\tilde{g}$ into

$$\tilde{g}(m_0, S_0) = \sum_{j=0}^{n-1} (-1)^j m_j(g) \sigma_{rj} r^j \ . \tag{2.23}$$

Chapter 4 makes use of the following three functions to align multiple shapes: the volume $V(\Omega)$, the center of mass $\overline{\Omega}$, and second order moment tensor $\mathbf{I}_\Omega$. These are

given by

$$V(\Omega) = \int_\Omega 1 \, dV \ , \tag{2.24}$$

$$\overline{\Omega} = \frac{1}{V(\Omega)} \int_\Omega x \, dV \ , \tag{2.25}$$

$$\mathbf{I}_\Omega = \int_\Omega x x^T \, dV - V(\Omega) \overline{\Omega} \, \overline{\Omega}^T \ , \tag{2.26}$$

where the integration is over all $x \in \Omega$. For the volume, the function to integrate is $g = 1$. Let $\delta_0 = \tilde{g}$. Then

$$\delta_0 = \sum_{j=0}^{n-1} (-1)^j \frac{1}{j+1} \sigma_{rj} r^j \ . \tag{2.27}$$

This leads to the three-dimensional volume formula

$$V(\Omega) = \int_{\tilde{M}} r \, dM - \int_{\tilde{M}} r^2 H_{\text{rad}} \, dM + \frac{1}{3} \int_{\tilde{M}} r^3 K_{\text{rad}} \, dM \ . \tag{2.28}$$

A similar formula for the center of mass may be obtained via integrating the function $g(x) = x$ and dividing by the volume. In this case letting $\delta_1 = \tilde{g}$ yields

$$\delta_1 = \sum_{j=0}^{n-1} (-1)^j \left(\frac{1}{j+1} m_0 + \frac{1}{j+2} S_0 \right) \sigma_{rj} r^j \ . \tag{2.29}$$

In three dimensions again, this gives

$$\begin{aligned} \overline{\Omega} = \frac{1}{V(\Omega)} \Bigg(\int_{\tilde{M}} \left(m_0 + \frac{1}{2} S_0 \right) r \, dM - \int_{\tilde{M}} \left(m_0 + \frac{2}{3} S_0 \right) r^2 H_{\text{rad}} \, dM \\ + \int_{\tilde{M}} \left(\frac{1}{3} m_0 + \frac{1}{4} S_0 \right) r^3 K_{\text{rad}} \, dM \Bigg) \ . \end{aligned} \tag{2.30}$$

Finally, the second order moment tensor may be obtained via $g(x) = xx^T$. Defining $\delta_2 = \tilde{g}$ yields

$$\delta_2 = \sum_{j=0}^{n-1} (-1)^j \left(\frac{1}{j+1} m_0 m_0^T + \frac{1}{j+2} (m_0 S_0^T + S_0 m_0^T) + \frac{1}{j+3} S_0 S_0^T \right) \sigma_{rj} r^j \ . \tag{2.31}$$

The full formula in three dimensions is now

$$\begin{aligned}\mathbf{I}_\Omega = &\int_{\tilde{M}} \left(m_0 m_0^T + \frac{1}{2}(m_0 S_0^T + S_0 m_0^T) + \frac{1}{3} S_0 S_0^T \right) r \, dM \\ &- \int_{\tilde{M}} \left(m_0 m_0^T + \frac{2}{3}(m_0 S_0^T + S_0 m_0^T) + \frac{1}{2} S_0 S_0^T \right) r^2 H_{\text{rad}} \, dM \\ &+ \int_{\tilde{M}} \left(\frac{1}{3} m_0 m_0^T + \frac{1}{4}(m_0 S_0^T + S_0 m_0^T) + \frac{1}{5} S_0 S_0^T \right) r^3 K_{\text{rad}} \, dM \\ &- V(\Omega) \overline{\Omega} \, \overline{\Omega}^T \ . \quad (2.32)\end{aligned}$$

The last term centers the tensor around the center of mass, i.e., subtracting it off is equivalent to replacing x with $x - \overline{\Omega}$ in the definition of g. Separating the contribution of this term out, however, allows the computation of all of these integrals simultaneously in a single pass.

Chapter 3

A Continuous Medial Model[1]

> . . . if the difficulties of description and representation could be overcome, it is by means of such coordinates in space that we should at last obtain an adequate and satisfying picture of the processes of deformation and the direction of growth.
>
> (Thompson, 1917)

This chapter describes a new, continuous medial model, which is based on subdivision surfaces. The goal in creating such a model is to obtain a system of coordinates over the entire interior of the object (and some of the exterior) that is directly comparable with other objects. Previously, Yushkevich also developed a model based on cubic B-splines, which is well suited for 2D shapes, where the number of points on the boundary of the medial axis is finite (Yushkevich et al., 2003; Yushkevich, 2003). Its extension to 3D, however, has several issues.

The primary issue is enforcing the condition $\|\nabla r\| = 1$ at the boundary. Since there are infinitely many boundary points, but only a finite number of control point values to use as free parameters, this cannot be enforced explicitly by constraining control point values. Instead, Yushkevich chooses the value of the radius at the boundary control points to be a large, negative number, which ensures that the condition holds along some curve in the region, which can be solved for implicitly. Unfortunately, this means that the boundary for every shape model appears at a different location in the parameter space, and due to its unconstrained, irregular shape, it is difficult to obtain

[1] This chapter is based on work done in collaboration with Guido Gerig and previously published in (Terriberry and Gerig, 2006).

a mapping between the domains of different shapes. This makes the representation unsuitable for performing shape statistics across a population.

Later work solved this problem by interpolating a potential function ρ across the medial surface instead of the radius (Yushkevich et al., 2005). The radius is recovered from ρ by numerically solving a differential equation whose boundary conditions also ensure $\|\nabla r\| = 1$. This allows the boundary condition to be enforced explicitly on a rectangular domain and thereby solves the problem of an implicit boundary curve, at the expense of losing an explicit function for the radius.

However, both methods also exhibit a problem with boundary conditions at branch points. In order to handle branching, two constraints must be enforced where medial sheets meet: a condition on the derivatives of r to ensure the spoke directions $U^{\pm}$ are the same and the condition that the values of r themselves are equal. The first method cannot choose where the derivative condition holds true and thus has little chance of getting the radii (or even the edge of the medial sheets themselves) to line up. The second method, with only a single free parameter, the potential function value at that point, yields an over-constrained problem. A solution to these problems has not, as of the time of this writing, been found.

We take a different approach. To move efficiently from a discrete grid of medial control points to a continuous surface and to provide *local control*, we use Catmull-Clark subdivision surfaces with what we shall term an *ordinary, corner-free boundary* to model the medial sheets. These surfaces model the four-dimensional quantity $\mathcal{M} = (m, r)$, where m is the medial sheet location, and r the radius. The boundary conditions are then enforced by modifying the patches at the edges of the sheet to use an interpolating spline. This reduces the degree of the polynomial equation for the boundary conditions from 12 to 2, yielding an explicit and efficient solution. The solutions to these equations are used to construct a "control curve" that replaces the outer layer of control points, effectively providing the infinite number of free parameters needed to enforce the condition everywhere. These two key ideas—interpolating splines and control curves—are what makes this approach possible. The rest of the chapter lays out the details.

3.1 Catmull-Clark Subdivision on the Medial Axis

First, we present a brief review of Catmull-Clark subdivision surfaces. These are a generalization of B-spline knot insertion to meshes of arbitrary topology (Catmull and

Clark, 1978). They were initially created as a graphics primitive to represent the boundary of closed objects. A continuous surface, called the *limit surface*, is created from an initial, discrete mesh by recursive subdivision. At each step, a new vertex is added to every face and every edge, and these are then connected in a new mesh, as illustrated in Figure 3.1. After the first level of subdivision, all of the faces are quadrilaterals, and every new vertex will have valence four.

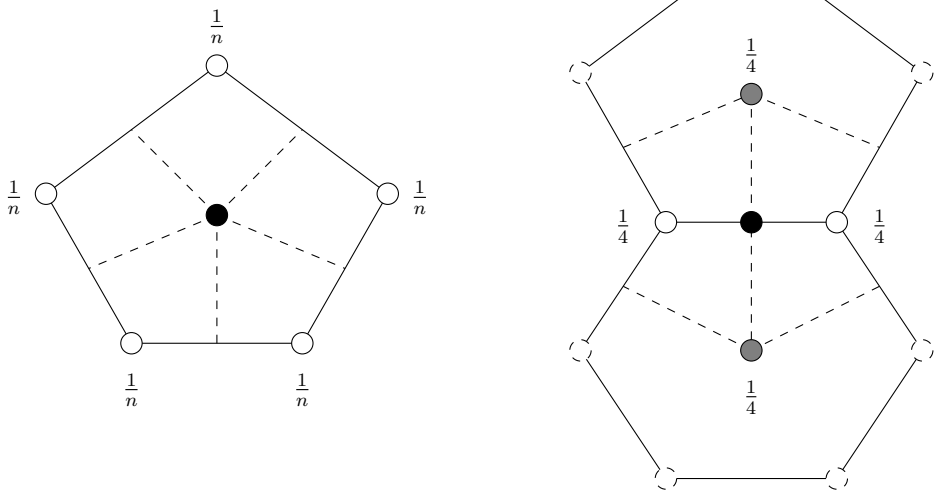
Catmull-Clark surfaces as originally developed are restricted to closed surfaces, so the rules on the interior of the mesh are presented first. Any (interior) vertex of valence four is thus called an *ordinary* vertex, while the remaining (interior) vertices are *extraordinary*. After two subdivision steps, every face has at most one extraordinary vertex. Away from an extraordinary vertex, the limit surface behaves exactly like a B-spline patch, and thus the surface and its derivatives can be evaluated efficiently at arbitrary points. Stam showed how, with some one-time setup, the limit surface could also be efficiently evaluated near extraordinary vertices (Stam, 1999). Catmull-Clark surfaces are everywhere C^2 continuous, except at extraordinary vertices, where they are C^1.

This level of continuity is sufficient for our purposes. C^1 continuity ensures that ∇r, and thus the spoke field, is continuous everywhere. Since the spoke field is normal to the boundary, a continuous spoke field ensures a well-defined tangent plane and thus G^1 continuity of the reconstructed boundary. C^2 continuity everywhere except on a set of measure zero ensures that we can compute S_{rad} almost everywhere, so we can check for illegalities and compute medial integrals.

Subdivision surfaces are also appealing because a control point only influences a small local region of the surface surrounding it, instead of the whole surface. This is the local control property alluded to earlier. It will become advantageous later when computing derivatives describing how the medial axis changes when a control point is moved.

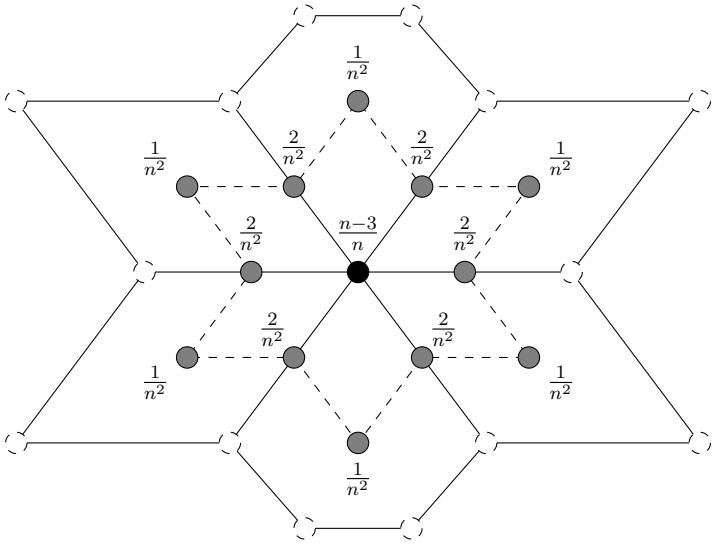
3.1.1 Ordinary Corner-Free Boundary

Blum medial sheets are surfaces with edges and creases, and there are several different strategies in the literature for handling such meshes. Initially, (Hoppe et al., 1994) proposed a set of rules for triangular subdivision surfaces that are provably C^1. They have a straightforward adaptation to quadrilateral subdivision surfaces—including Catmull-Clark surfaces—as described by, e.g., (DeRose et al., 1998; Warren and Schaefer, 2004).



(a) F-mask.

(b) E-mask.



(c) V-mask.

Figure 3.1: Catmull-Clark subdivision masks. The F-mask creates a new vertex at the center of each existing face, and the E-mask creates a new vertex splitting each existing edge. Finally the V-mask updates all existing vertices in the previous subdivision level to their new positions. The black vertex in each mask is the vertex being created or updated, the white vertices are other pre-existing vertices, and the gray vertices are other new vertices created by the subdivision. Only vertices with a solid border contribute to the mask; those with a dashed border are shown merely for context. The value of n in the F-mask and the V-mask is the valence of the center vertex.

The latter also describe a simple method of implementing these rules, but neither proved that the rules produce C^1 limit surfaces. A more complete set of rules that are provably C^1 everywhere was proposed by (Biermann et al., 2000), addressing problems with extraordinary edge vertices and convex and concave corners.

For simplicity, we require that the edges of the mesh contain only ordinary vertices. Unlike interior vertices, they will not have valence four, but the term may also describe vertices that *could* be made ordinary by some extension of the mesh past the edge. For reasons made clear in Section 3.2, we also disallow corners, i.e., no two adjacent edges on a branch or edge curve can belong to the same face. Besides being a practical requirement, mesh corners are also undesirable for many of the objects we wish to model. The result is that all vertices on the edge of the mesh will have valence three, and any two adjacent edges will belong to different patches.

There is one exception where corners are allowed, and in fact required, and that is at the fin creation points at the end of a branch curve. Two of the three medial sheet pieces joined at a branch curve flatten into a single medial sheet, and the end of this crease is modeled with a *dart vertex*, which blends smoothly into the joined sheet. The remaining fin sheet ends in an edge curve, which intersects the branch curve in a corner. This is the sole exception to the exclusion of corners.

We call the result a mesh with ordinary, corner-free boundary. Under these mesh restrictions, the subdivision rules due to both (Hoppe et al., 1994) and (Biermann et al., 2000) are equivalent. They are implemented here as follows. Each vertex is assigned a dimension corresponding to the type of structure it belongs to; interior points have dimension two, edge and crease points have dimension one, and corners have dimension zero. Then, during subdivision, the subdivision masks are modified to exclude vertices of a higher dimension.

This effectively causes the edge of the mesh to degenerate to a one-dimensional B-spline curve, unaffected by the other control points in the interior of the mesh. This makes it straightforward to join two medial sheets along a branch curve, simply by sharing control points everywhere on that curve except at a fin creation point (see Figure 2.2(c)). Corners are not modified at all by the subdivision process, while dart vertices are treated as interior points and require no special rules. Hence we set the corner at a fin creation point equal to the limit point of the associated dart vertex.

Now, in order to enforce the boundary conditions required to ensure that the reconstructed boundary forms a closed surface, we need to subdivide the mesh into *patches* and then modify patches near the edge and branch curves. To construct the patches,

we first perform two levels of subdivision. This ensures that each face is a quadrilateral and falls into exactly one of three disjoint classes: patches adjacent to an edge or branch curve, patches with exactly one extraordinary vertex, or ordinary interior patches.

3.1.2 Interior Patches

The ordinary interior patches behave exactly like uniform cubic B-spline surfaces, with the local vertices from the subdivided mesh used directly as control points. We denote these $P_{i,j}$, where each is a tuple (m, r) of control points for $\mathcal{M}$, and use $\mathbf{P}$ to refer to the 4×4 matrix of control points for a single patch. We write $m(\mathbf{P})$ or $r(\mathbf{P})$ to refer to each component of the control point matrix $\mathbf{P}$, and similarly for an individual control point $P_{i,j}$.

The local portion of the medial sheet is now defined by the equation

$$\mathcal{M}(u, v) \triangleq \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \mathbf{B}\mathbf{P}\mathbf{B}^T \begin{bmatrix} 1 & v & v^2 & v^3 \end{bmatrix}^T . \tag{3.1}$$

Here $\mathbf{B}$ is the B-spline basis matrix:

$$\mathbf{B} \triangleq \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} . \tag{3.2}$$

The reconstructed boundary is then given by

$$\mathcal{B}^{\pm}(u, v) \triangleq m(u, v) \pm r(u, v) \cdot U^{\pm}(u, v) , \tag{3.3}$$

where $U^{\pm}$ is defined as in (2.3). For notational simplicity we drop the parameters (u, v) where they are unnecessary.

Since m and r are cubic polynomials, their derivatives are simple to calculate, and with these derivatives $U^{\pm}$ can be computed as shown in Appendix A. The case is similar for patches with an extraordinary vertex. For these, we evaluate the patch using Stam's method (Stam, 1999), which we will briefly summarize here.

A patch with an extraordinary vertex of valence k is affected by $2k + 8$ control points, as illustrated in Figure 3.2. We arrange these in a single column vector $\mathbf{P}_0$. A new set of $2k + 17$ control points $\bar{\mathbf{P}}_1$ can be produced by subdivision, allowing $\frac{3}{4}$ of the patch to be evaluated as ordinary B-splines, as illustrated in Figure 3.3. The nine

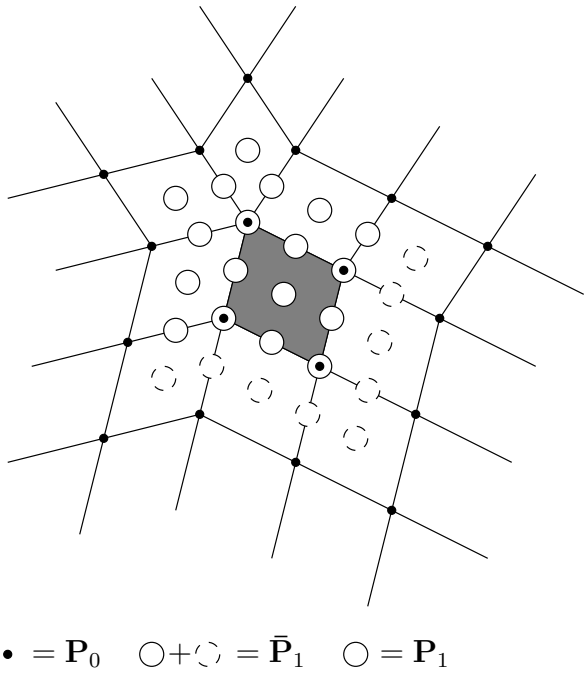


Figure 3.2: Control points around an extraordinary vertex. The different circles indicate membership in each of the sets as labeled. These sets are described in the text.

points in $\bar{\mathbf{P}}_1$ away from the extraordinary vertex can be discarded, to obtain a new set of $2k + 8$ points $\mathbf{P}_1$, which describe the remaining quadrant of the patch.

Both $\bar{\mathbf{P}}_1$ and $\mathbf{P}_1$ can be written as linear combinations of $\mathbf{P}_0$, related by the subdivision matrices $\mathbf{A}$ and $\bar{\mathbf{A}}$:

$$\mathbf{P}_1 \triangleq \mathbf{A}\mathbf{P}_0 \ , \qquad\qquad \bar{\mathbf{P}}_1 \triangleq \bar{\mathbf{A}}\mathbf{P}_0 \ . \tag{3.4}$$

The precise definition of these subdivision matrices can be derived from the Catmull-Clark weights and is given in (Stam, 1999). This subdivision can be repeated indefinitely to get a parametric surface representation arbitrarily close to the extraordinary vertex with, e.g.,

$$\mathbf{P}_n = \mathbf{A}\mathbf{P}_{n-1} = \mathbf{A}^n\mathbf{P}_0 \ , \qquad\qquad \bar{\mathbf{P}}_n = \bar{\mathbf{A}}\mathbf{P}_{n-1} = \bar{\mathbf{A}}\mathbf{A}^{n-1}\mathbf{P}_0 \ . \tag{3.5}$$

The matrix exponentiation is easily computed if the eigendecomposition of $\mathbf{A}$ is available:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \ , \tag{3.6}$$

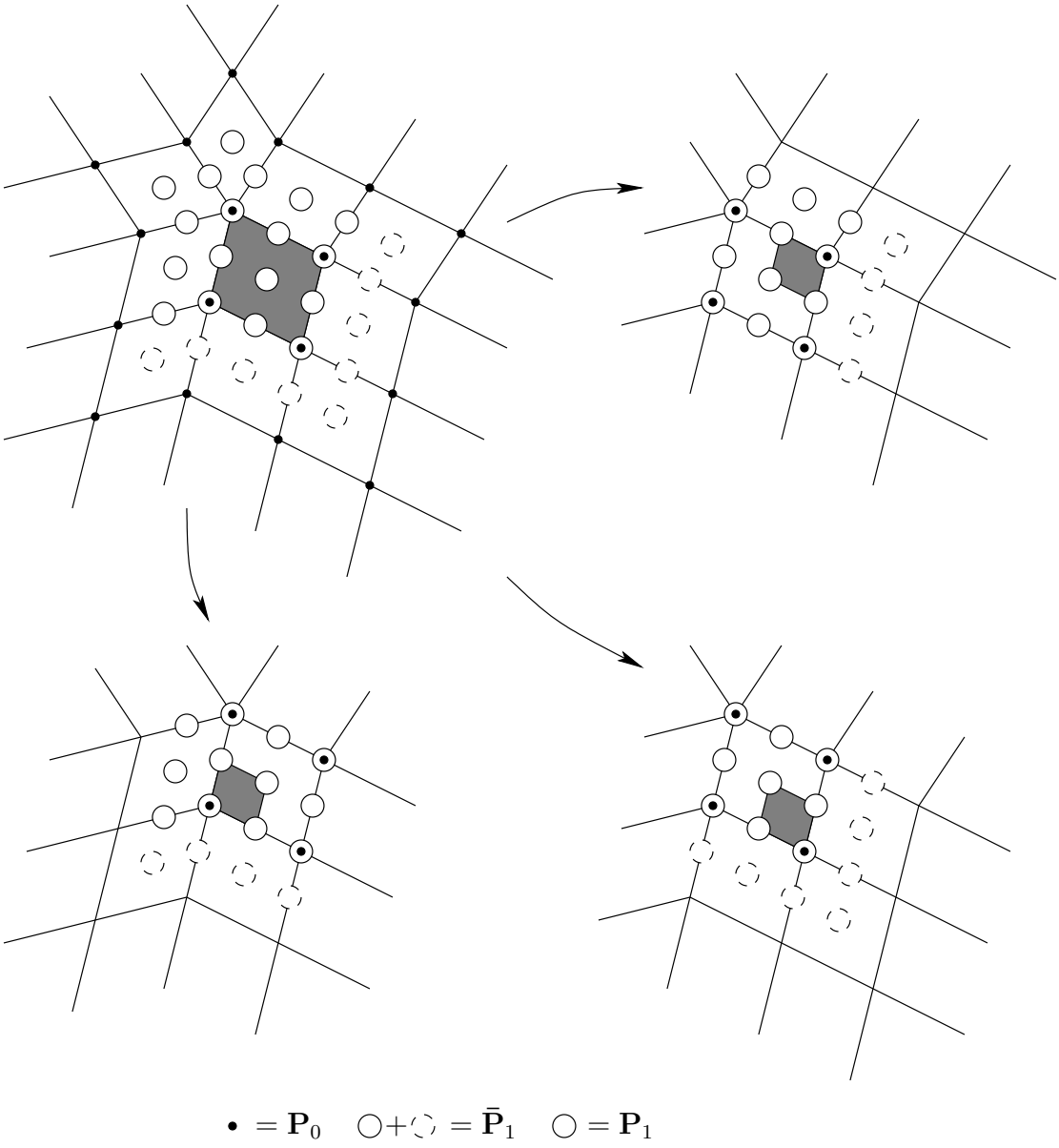


Figure 3.3: Subdivision near an extraordinary vertex. Groups of 16 control points from the set $\bar{\mathbf{P}}_1$ can be used to construct B-spline patches over three of the four quadrants of the patch near the extraordinary vertex.

where $\mathbf{V}$ is a matrix of orthogonal eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix of real eigenvalues, sorted in descending order. Then the new control points can be written

$$\bar{\mathbf{P}}_n = \bar{\mathbf{A}}\mathbf{A}^{n-1}\mathbf{P}_0 = \bar{\mathbf{A}}\mathbf{V}\mathbf{\Lambda}^{n-1}\mathbf{V}^{-1}\mathbf{P}_0 \ . \tag{3.7}$$

The product on the right, $\mathbf{V}^{-1}\mathbf{P}_0$, is constant regardless of what portion of the patch is being evaluated. The product on the left, $\bar{\mathbf{A}}\mathbf{V}$, depends solely on the valence, n. Thus, if $\mathbf{V}^{-1}$, $\mathbf{\Lambda}$, and $\bar{\mathbf{A}}\mathbf{V}$ are pre-computed for each valence, then after one matrix multiply per patch, some exponentiation, and one additional matrix multiply per quadrant each point can be evaluated via a normal B-spline. For precise details, we refer the reader to (Stam, 1999), which contains some corrections[2] over a previous version of the work appearing in (Stam, 1998).

This allows the patch and its derivatives to be evaluated everywhere except at the extraordinary vertex without the computational expense and memory needed to compute n levels of subdivision. At the extraordinary vertex itself, the values of m and r are easily obtained from the first eigenvector, whose corresponding eigenvalue is 1. All of the other eigenvalues are less than one, so setting

$$\mathbf{\Lambda}^{\infty} = \begin{bmatrix} 1 & & \\ & 0 & \\ & & \ddots \end{bmatrix} \tag{3.8}$$

will return the limit point. To compute first derivatives within a patch, Stam suggests

[2] Even in this version, there are several mistakes in its Appendix A. The re-shuffling of the eigenvalues described gives precisely the same order as before. The correct order is

$$\begin{aligned} \mu_4 = \lambda_1^- \ , \quad & \mu_5 = \lambda_{n-1}^+, \\ \mu_6 = \lambda_2^+ \ , \quad & \mu_7 = \lambda_{n-2}^-, \\ & \vdots \end{aligned}$$

This also means that μ_{l+3} in the expressions for eigenvectors $\mathbf{u}_{2l+2}$ and $\mathbf{u}_{2l+3}$ should be μ_{2l+2} instead. Additionally, when Stam calls for "the same reshuffling as above" to compute $\mathbf{V}^{-1}$, the reshuffling (even if the above were correct) is not precisely the same. Obviously it is now done on rows instead of columns, but in addition, for the resulting real left eigenvectors to form a true inverse, the formulas must be modified slightly, as follows:

$$\begin{aligned} \mathbf{u}_{2l+2}^{-1} &= \mathbf{k}_{l+3}^{-1} + \mathbf{k}_{2N-l+2}^{-1} \ , \\ \mathbf{u}_{2l+3}^{-1} &= i(\mathbf{k}_{l+3}^{-1} - \mathbf{k}_{2N-l+2}^{-1}) \ . \end{aligned}$$

We have thus far been unsuccessful in contacting Stam to offer these corrections, but we would be remiss if we did not provide them to you here.

simply using a large value for n in the exponent. Another possibility is reparameterizing the surface based on the eigenbasis functions corresponding to the second and third eigenvectors. This is, however, much more complex, as it requires computing derivatives of the inverse characteristic map.

Unlike the first derivatives, the second derivatives at the extraordinary vertex are inherently unstable, as the curvature diverges when $n \neq 4$. This means that S_{rad} cannot be computed, so the radial curvature condition cannot be checked. This does not pose a problem for evaluating the spoke field itself, as these curvatures remain integrable (Reif and Schröder, 2001). While theoretically it would be nice to be able to compute S_{rad} for an extraordinary vertex, as a practical matter it is not a serious hindrance. However, should a fully G^2 solution be desired, there are several known approaches in the subdivision surface literature. One strategy is to modify the subdivision masks, as described by (Prautzsch and Umlauf, 1998), but this produces surfaces whose curvature is zero at the extraordinary point (and no local subdivision algorithm can do otherwise). Alternatively, the patches can be replaced by Bézier patches of a higher degree, as described by (Loop, 2004), but this requires biseptic polynomials. Many other alternatives exist with differing restrictions, requirements, and constructions; see (Loop, 2004) for a recent summary.

3.2 Edge Patches

Having described the general case in the interior, we now turn to the challenge of patches along the edge of the medial sheet. On these patches, the four control points that lie outside of the surface, beyond the edge, represent free parameters that can be modified without affecting the surface defined by the interior patches. However, with the four parameters of m and r per control point, this gives only sixteen free parameters to enforce the requisite boundary condition at all of the points on the edge, of which there are infinitely many. With a straightforward formulation, finding a configuration which satisfies this edge condition even just at the control points requires solving for a root of a twelfth degree polynomial. To find a valid solution at every point on the curve, a more creative approach had to be derived.

The general approach taken here is two-fold. First, the spline is converted to an interpolating basis in one of the variables. This breaks the dependence between u and v, meaning that the solution for a control point that enforces the edge condition at one specific point does not depend on the solution at other points, reducing the resulting

polynomial to a quadratic. Second, to extend this solution from just the control points to the entire edge curve, the four free control points are replaced with a control curve. The control curve provides one free parameter for every point on the edge, which is fixed by the boundary condition. The details of these two steps follow.

Without loss of generality, assume the edge of the medial axis lies along the line $u = 0$, as illustrated in Figure 3.4. Adding $P_{0,j}$ points out past the boundary, where $P_{0,j} \triangleq 2P_{1,j} - P_{2,j}$, gives a full set of B-spline control points that produce the Catmull-Clark limit surface of this patch.

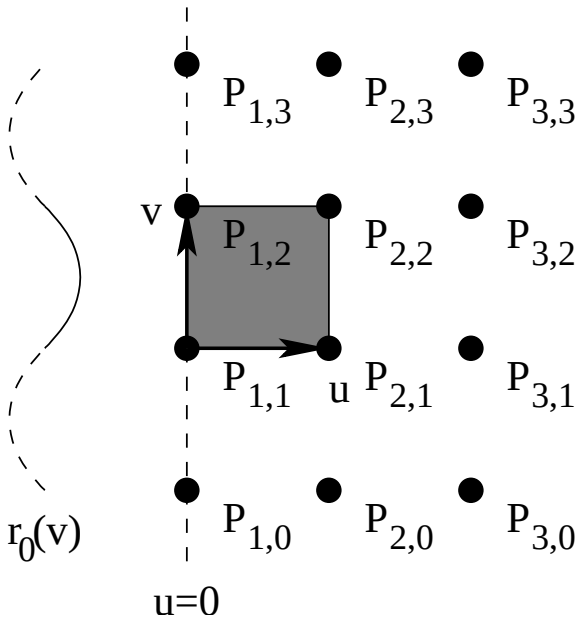


Figure 3.4: The control curve, $r_0(v)$, on an edge patch.

The medial sheet m is computed us usual, but in order to enforce the edge condition (2.4) along this edge, we compute r using a *control curve* $r_0(v)$ instead of a few isolated control points. This lets us ensure the condition holds at every point on the edge, which would not be possible with only a finite number of control point values as free parameters. This curve does not represent a set of parameters explicitly defined by the modeler, but is implicitly defined by the boundary condition.

Unfortunately, the use of such a curve makes evaluation fundamentally asymmetric with respect to the two variables u and v. We must first expand the v terms, obtaining new control points for a spline in u. Then we use a change-of-basis to convert the spline in u to an *interpolating spline*, which passes through its control points. Finally, we replace the control point on the left with the point from our control curve and

perform interpolation in the u direction. Because of this asymmetry, we cannot, for example, add a control curve to both the $u = 0$ and $v = 0$ edges of the same patch, which is why we disallow corners on the edge of the mesh.

The particular interpolating spline we use is the Catmull-Rom spline, since it is C^1 and has third-order, meaning that when used to approximate an arbitrary function, it matches the first three terms of its Taylor series approximation, making any error terms at least cubic. It is defined as in (3.1), but replacing the basis matrix on the left with

$$\mathbf{C} \triangleq \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{bmatrix} . \tag{3.9}$$

The change-of-basis may be accomplished by replacing $\mathbf{BP}$ in (3.1) by $\mathbf{CP'}$ and using the new set of control points

$$\mathbf{P'} \triangleq \mathbf{C}^{-1}\mathbf{BP} , \tag{3.10}$$

producing

$$\mathcal{M}(u, v) = \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \mathbf{CP'B}^T \begin{bmatrix} 1 & v & v^2 & v^3 \end{bmatrix}^T . \tag{3.11}$$

Now, we can replace the row $P'_{0,j}$ with our control curve, and the result will still pass through the limit surface at either end of the patch. Furthermore, $P'_{0,j}$ has no influence on the derivative with respect to u on the right edge of the patch. Hence, whatever the value of the control curve, the right edge retains C^1 continuity. We now derive an equation for the control curve and show that we also retain C^1 continuity along the top and bottom edges.

3.2.1 Solving for $r_u(0, v)$

As outlined in Section 2.3.3, in order to ensure that the top and bottom spokes meet, we must have

$$\|\nabla r\| = 1 . \tag{3.12}$$

The gradient here is the Riemannian gradient, which on a two-dimensional manifold is given by

$$\nabla r \triangleq \begin{bmatrix} m_u & m_v \end{bmatrix} \mathbf{I}_m^{-1} \begin{bmatrix} r_u \\ r_v \end{bmatrix} , \tag{3.13}$$

where $\mathbf{I}_m$ is the metric tensor of m:

$$\mathbf{I}_m \triangleq \begin{bmatrix} E_m & F_m \\ F_m & G_m \end{bmatrix} \triangleq \begin{bmatrix} m_u \cdot m_u & m_u \cdot m_v \\ m_v \cdot m_u & m_v \cdot m_v \end{bmatrix} . \tag{3.14}$$

With some algebra, the entire expression can be rewritten as

$$\left\| \frac{(r_u m_v - r_v m_u) \times (m_u \times m_v)}{\|m_u \times m_v\|^2} \right\| = \frac{\|r_u m_v - r_v m_u\|}{\|m_u \times m_v\|} = 1 . \tag{3.15}$$

Hence,

$$\|r_u m_v - r_v m_u\|^2 = \|m_u \times m_v\|^2 , \tag{3.16}$$

which expands to

$$r_u^2 G_m - 2 r_u r_v F_m + r_v^2 E_m = E_m G_m - F_m^2 . \tag{3.17}$$

Now, we hold $m_u(0, v)$, $m_v(0, v)$, and $r_v(0, v)$ fixed and solve for $r_u(0, v)$. The key observation is that using a spline interpolative in u means that $r(0, v)$, and hence $r_v(0, v)$, is completely determined by the control point values $r(P'_{1,j})$ along the edge. In particular, they do *not* depend on the values of our unknown control curve. This is the crucial reason for choosing an interpolating spline in the first place. Solving (3.17) for r_u produces

$$r_u = \frac{1}{G_m} \left(r_v F_m \pm \sqrt{(G_m - r_v^2)(E_m G_m - F_m^2)} \right) . \tag{3.18}$$

There are two possible solutions, one of which corresponds to spokes along the crest pointing outward from the medial sheet, and the other of which corresponds to spokes pointing inwards. The latter is clearly illegal for a Blum medial axis, so we want to identify which solution corresponds to the former.

Along the crest, the spokes always lie in the (m_u, m_v) plane, so we can construct

the vector $((m_u \times m_v) \times m_v)$, which should lie in the same plane, perpendicular to the crest's tangent vector and pointing outwards. Then, we can test that

$$-r\nabla r \cdot ((m_u \times m_v) \times m_v) \geq 0 \ . \tag{3.19}$$

By applying Lagrange's formula to expand the cross products and employing some simplifying algebra, this can be reduced to

$$-r(r_v F_m - r_u G_m) \geq 0 \ . \tag{3.20}$$

However, plugging our two solutions for r_u into (3.20) yields

$$-r\left(r_v F_m - \left(r_v F_m \pm \sqrt{(G_m - r_v^2)(E_m G_m - F_m^2)}\right)\right) \geq 0 \ , \tag{3.21}$$

which reduces to

$$\pm r\sqrt{(G_m - r_v^2)(E_m G_m - F_m^2)} \geq 0 \ , \tag{3.22}$$

whose sign clearly only depends on the radius and the square root term. Since we always require a positive radius, the plus solution is always the solution pointing outwards.

3.2.2 The Complete Control Curve

Now, given $r_u(0, v)$, it is a simple matter to solve for the value of the control curve $r_0(v)$ that produces this derivative. Let $r_1(v)$, $r_2(v)$, and $r_3(v)$ be the continuous curves given by the three lines of ordinary control points in the v direction:

$$r_i(v) \triangleq \begin{bmatrix} P'_{i,0} & P'_{i,1} & P'_{i,2} & P'_{i,3} \end{bmatrix} \mathbf{B}^T \begin{bmatrix} 1 & v & v^2 & v^3 \end{bmatrix}^T \ , \qquad i = 1 \ldots 3 \ . \tag{3.23}$$

Then

$$r_u(0, v) = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{C} \begin{bmatrix} r_0(v) & r_1(v) & r_2(v) & r_3(v) \end{bmatrix}^T \ , \tag{3.24}$$

and solving for $r_0(v)$ yields

$$r_0(v) = \frac{1}{C_{1,0}} \left(r_u(0, v) - C_{1,1} r_1(v) - C_{1,2} r_2(v) - C_{1,3} r_3(v)\right) \ , \tag{3.25}$$

where $C_{i,j}$ is the (0-indexed) i,jth element of $\mathbf{C}$. It is important to emphasize that this formulation only works for interpolating splines, since otherwise $r_u(0,v)$ would not be independent of $r_0(v)$ in the right-hand side of (3.25).

Now $r_0(v)$ is a function of $m_u(0,v)$, $m_v(0,v)$, and $r_v(0,v)$ only. As these are all first derivatives of functions obtained from B-spline patches, they are C^1 across patches, and hence so is $r_0(v)$. This ensures that $r(u,v)$ is C^1 with neighboring patches along the top and bottom edges, as well as the right edge.

Finally, we complete the interpolation in the u direction to obtain the final value of r:

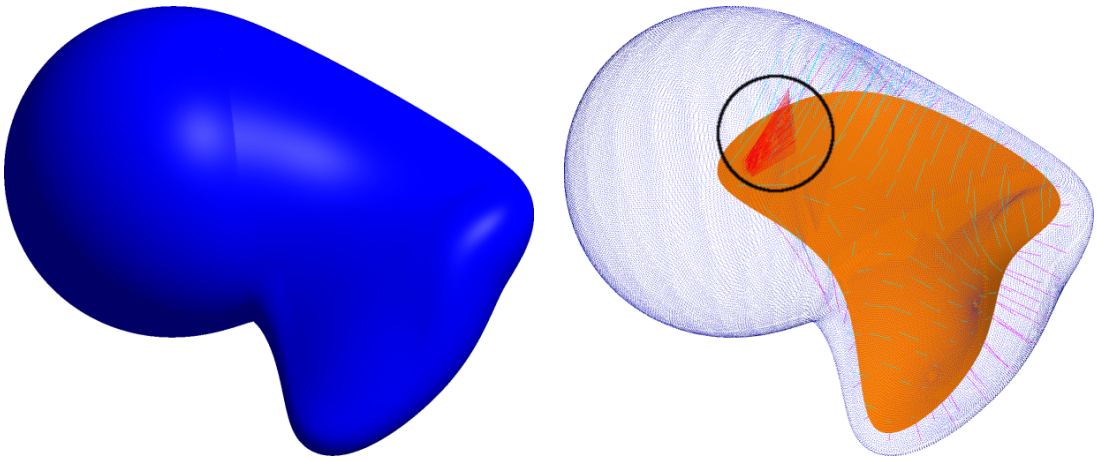
$$r(u,v) = \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \mathbf{C} \begin{bmatrix} r_0(v) & r_1(v) & r_2(v) & r_3(v) \end{bmatrix}^T . \tag{3.26}$$

3.2.3 Sympathetic Overfolding

Due to the cubic nature of our functions, modification to r on the left edge of the patch can produce a sympathetic crease on the right edge, which can cause overfolding, especially when r is increasing quickly as one approaches the edge of the medial axis. An example is illustrated in Figure 3.5. Our initial hypothesis was that this is due to the fact that our choice of spline, the Catmull-Rom spline, is only C^1, and that raised the question of whether or not another spline would give better behavior. This turns out not to be the case, but the better spline developed to answer the former question gave the mechanism needed to solve the real problem.

In general there at not enough free parameters to create a C^2 interpolating spline with only one piecewise cubic segment between control points. Enforcing the necessary constraints produces a basis matrix whose first column is all zeros. The resulting spline is thus quadratic and completely independent of $r_0(v)$, so we cannot enforce our boundary condition. Under the traditional formulation, the local C^2 interpolating spline with smallest degree is quartic, and its support spans six control points (Blu et al., 2003).

A cubic, interpolating C^2 spline *can* be created by subdividing the interval into multiple piecewise cubic segments. Each segment interpolates just part of the interval $[0,1]$, but all three use the same set of control points. Using equal-sized segments and following the procedure outlined by (Keys, 1981), one can derive a spline with the necessary continuity. Adding more segments introduces extra free parameters, which can be fixed by optimizing the order of convergence. Three segments are sufficient to achieve third-order convergence, meaning that any quadratic can be represented exactly.



(a) A crease on the boundary. (b) The medial axis and spoke field.

Figure 3.5: An example of sympathetic overfolding, created even when $\zeta = \frac{11}{8}$. The cluster of red spokes near the crease indicate the places where overfolding was detected.

The approximation order cannot be increased further by adding more segments; the additional segments become duplicates of one another.

The resulting third-order spline is given by the following three basis matrices, used over the intervals $[0, \frac{1}{3}]$, $[\frac{1}{3}, \frac{2}{3}]$, and $[\frac{2}{3}, 1]$, respectively:

$$\begin{aligned}
\mathbf{B}^{(0)} &\triangleq \frac{1}{12}\begin{bmatrix} 0 & 12 & 0 & 0 \\ -6 & 0 & 6 & 0 \\ 6 & -12 & 6 & 0 \\ 7 & -21 & 21 & -7 \end{bmatrix}, \\
\mathbf{B}^{(1)} &\triangleq \frac{1}{12}\begin{bmatrix} 1 & 9 & 3 & -1 \\ -15 & 27 & -21 & 9 \\ 33 & -93 & 87 & -27 \\ -20 & 60 & -60 & 20 \end{bmatrix}, \\
\mathbf{B}^{(2)} &\triangleq \frac{1}{12}\begin{bmatrix} -7 & 33 & -21 & 7 \\ 21 & -81 & 87 & -27 \\ -21 & 69 & -75 & 27 \\ 7 & -21 & 21 & -7 \end{bmatrix}.
\end{aligned} \tag{3.27}$$

Other splines are possible for different choices of the segment end points.

Unfortunately, because each segment has its own basis matrix, there is no longer a

simple change of basis formula to convert a B-spline to the new spline; each segment has its own formula. Yet if the segments do not all use the same set of control points, the spline is no longer C^2. Enforcing the constraint that all three change of basis formulas produce the same control points again restricts the curve to a quadratic. However, the advantage to this formulation over one using just one segment is that it is no longer independent of $r_0(v)$, so the boundary condition can be enforced.

By increasing the support size from four to six control points (while retaining the cubic basis functions), a fourth-order C^2 interpolating spline can be obtained, still using three segments. Using equally spaced segment endpoints again, the spline is given by the following three basis matrices:

$$\begin{aligned}
\mathbf{B}^{(0)} &\triangleq \frac{1}{24}\begin{bmatrix} 0 & 0 & 24 & 0 & 0 & 0 \\ 2 & -16 & 0 & 16 & -2 & 0 \\ 0 & 12 & -24 & 12 & 0 & 0 \\ -6 & 23 & -36 & 30 & -14 & 3 \end{bmatrix}, \\
\mathbf{B}^{(1)} &\triangleq \frac{1}{216}\begin{bmatrix} -5 & 24 & 170 & 44 & -21 & 4 \\ 63 & -360 & 414 & -252 & 171 & -36 \\ -135 & 756 & -1458 & 1296 & -567 & 108 \\ 81 & -441 & 918 & -918 & 441 & -81 \end{bmatrix}, \\
\mathbf{B}^{(2)} &\triangleq \frac{1}{24}\begin{bmatrix} 3 & -16 & 58 & -36 & 19 & -4 \\ -9 & 44 & -130 & 156 & -77 & 16 \\ 9 & -42 & 102 & -132 & 81 & -18 \\ -3 & 14 & -30 & 36 & -23 & 6 \end{bmatrix}.
\end{aligned} \tag{3.28}$$

However, because the spline is now fourth-order, it can represent any cubic curve exactly, which means that a single change of basis matrix must exist. Solving for it produces

$$\mathbf{A} \triangleq \frac{1}{6}\begin{bmatrix} 27 & -44 & 31 & -8 \\ 8 & -5 & 4 & -1 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ -1 & 4 & -5 & 8 \\ -8 & 31 & -44 & 27 \end{bmatrix}. \tag{3.29}$$

Using six control points instead of four requires slightly more computation, but because

all six are computed from one set of four control points, the locality of the method remains unchanged. That is, the six control points produced here do not depend on any more model parameters than the original four did.

Either of these splines gives C^2 continuity on the right edge, regardless of the control point value on the left edge, but the fourth-order spline does not limit the input to quadratic curves. However, we have tested both of them and found that the increased order of continuity does nothing to help avoid the sympathetic overfolding. Therefore our initial hypothesis was incorrect.

The real problem appears to be related to the rate at which the parameterization approaches the left edge of the medial sheet. We have observed less overfolding when $m_u(0, v)$ is small. A simple way to control this value is to modify our choice of $P_{0,j}$. If we also require that the spline interpolate the same limit point at $u = 0$, we will need to replace $P_{1,j}$ with a modified $\tilde{P}_{1,j}$, as well, which will affect the adjacent patch slightly. This gives rise to the one-parameter family of control points,

$$P_{0,j} = (6 - 4\zeta)P_{1,j} - (5 - 4\zeta)P_{2,j} \ , \qquad \tilde{P}_{1,j} = \zeta P_{1,j} + (1 - \zeta)P_{2,j} \tag{3.30}$$

The original Catmull-Clark control points are given by $\zeta = 1$. A choice of $\zeta = \frac{6}{5}$ makes $P_{0,j} = \tilde{P}_{1,j}$, which gives some improvement, but better results are obtained by a choice of $\zeta = \frac{11}{8}$, which makes $P_{0,j} = \frac{1}{2}(P_{1,j} + P_{2,j})$. This is the strategy we have adopted in this work, and as Figure 3.5 shows, the disparity in radii near the edge now needs to be quite severe to cause overfolding. This is the approach we will take throughout the rest of this work, due to its simplicity.

However, we continue to describe a more theoretically pleasing solution that completely eliminates the problem at the cost of some additional complexity. Further increasing ζ gives only marginally better results, until $\zeta = \frac{3}{2}$ makes $P_{0,j} = P_{2,j}$ and forces $m_u(0, v)$ to be identically zero, regardless of the choice of control points. Unfortunately, this poses a problem as we use $m_u(0, v)$ to compute the normal vector and tangent plane of the medial axis along the crest.

If we switch to an alternate parameterization, $\hat{u} = u^2$, and differentiate with respect to $\hat{u}$ instead of u, we can recover a well-defined tangent plane. However, unless $r_u(0, v)$ is also identically zero, ∇r will diverge to infinity. Thus we have two constraints to satisfy: that $r_u(0, v) = 0$ and that $r_{\hat{u}}(0, v)$ is chosen in accordance with (3.18). However, we can obtain the extra free parameter we need to enforce both of these conditions simultaneously from our fourth-order interpolating spline.

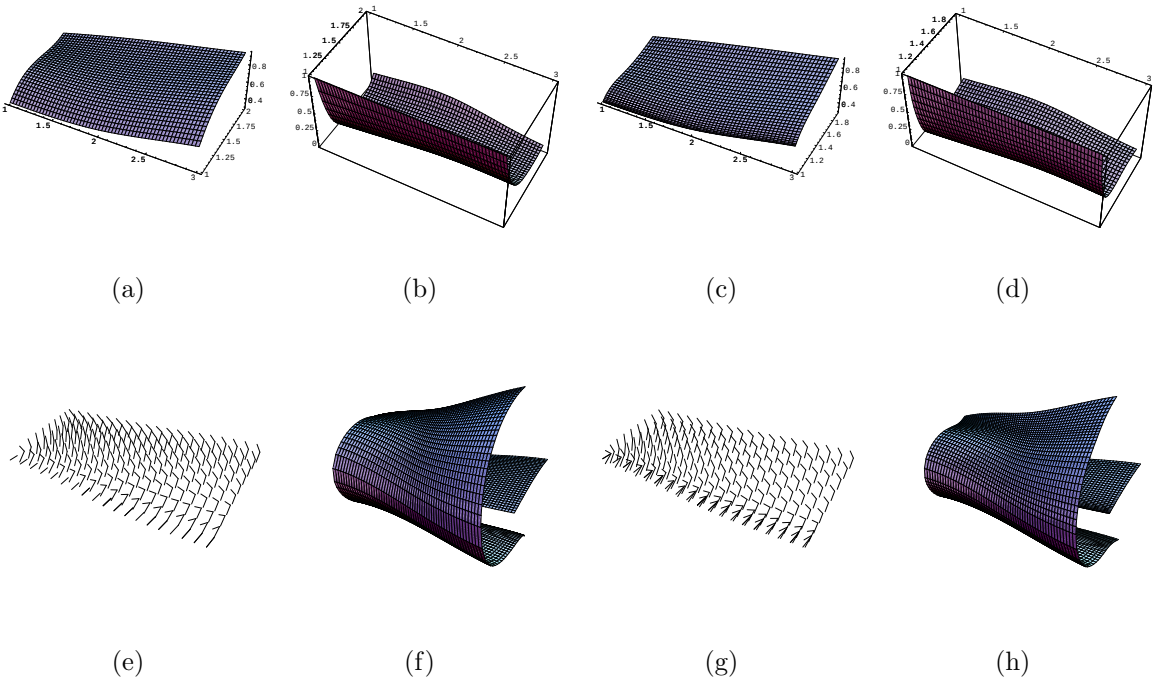


Figure 3.6: An example of a pair of adjacent edge patches evaluated with two different methods. On the left, we use the Catmull-Rom spline with $\zeta = 1$, the value corresponding to ordinary Catmull-Clark subdivision. On the right, we use our fourth-order cubic spline and $\zeta = \frac{3}{2}$. Both sides were derived using the same set of original control points. Pictured are (a), (c) the radius function, (b), (d) $\|\nabla r\|^2$, (e), (g) the unit spoke field $U^{\pm}$, and (f), (h) the medial sheet and reconstructed boundary.

Let $r_i(v)$ for $i = 1 \ldots 5$ be defined analogous to (3.23). We are again free to choose $r_0(v)$, but modifying $r_1(v)$ will affect the first derivative on the right. Therefore, we also modify $r_5(v)$ to compensate. As long as the first and last spline segments have the same size, the coefficients of $r_1(v)$ and $r_5(v)$ in this derivative will be symmetric, e.g., $B^{(2)}_{1,1} + 2B^{(2)}_{2,1} + 3B^{(2)}_{3,1} + B^{(2)}_{1,5} + 2B^{(2)}_{2,5} + 3B^{(2)}_{3,5} = 0$, so we can choose $\tilde{r}_1(v) = r_1(v) + \Delta r_1(v)$ and $\tilde{r}_5(v) = r_5(v) + \Delta r_1(v)$ for some value $\Delta r_1(v)$. In this case, it will also be true that the coefficients of $r_0(v)$, $r_1(v)$, and $r_5(v)$ in the second derivative on the right will be zero, so we will retain C^2 continuity on the right.

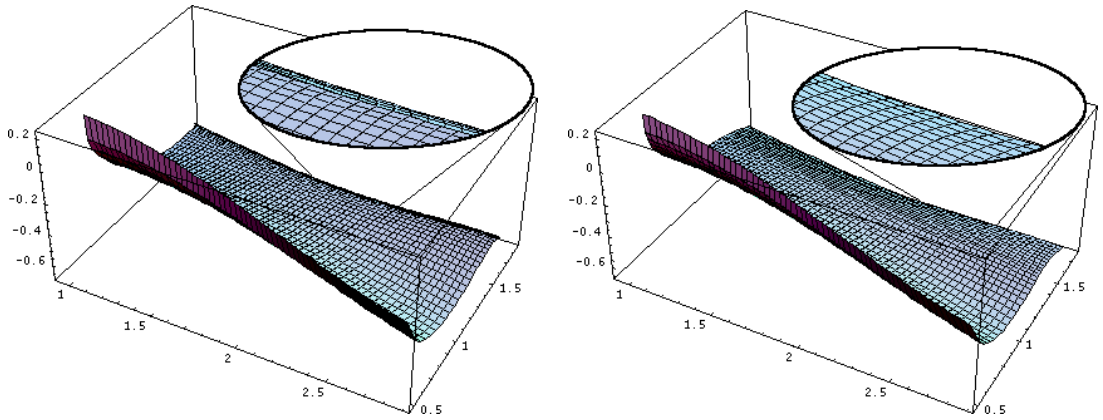
We use $r_0(v)$ to enforce $r_u(0, v) = 0$, producing

$$r_0(v) = -\frac{1}{B^{(0)}_{1,0}} \left(B^{(0)}_{1,1}\tilde{r}_1(v) + B^{(0)}_{1,2}r_2(v) + B^{(0)}_{1,3}r_3(v) + B^{(0)}_{1,4}r_4(v) + B^{(0)}_{1,5}\tilde{r}_5(v) \right) . \quad (3.31)$$

Then, using $\Delta r_1(v)$ to fix $r_{\hat{u}}(0, v)$ gives

$$\Delta r_1(v) = \frac{1}{B_{2,1}^{(0)}} \left(r_{\hat{u}}(0, v) - B_{2,1}^{(0)} r_1(v) - B_{2,2}^{(0)} r_2(v) - B_{2,3}^{(0)} r_3(v) \right) . \tag{3.32}$$

We have assumed $B_{2,0}^{(0)} = B_{2,4}^{(0)} = B_{2,5}^{(0)} = 0$. Without this assumption the above expressions are considerably more complicated.



(a) Bunching in u, $\zeta = 1$

(b) Bunching completely eliminated, $\zeta = \frac{3}{2}$

Figure 3.7: Two close-ups of the right edge of the bottom portion of the boundary pictured in Figure 3.6. At the right edge of (a), where $\zeta = 1$, the iso-parameter lines in u are bunched up. In (b), using $\zeta = \frac{3}{2}$ causes the surface to come to a complete stop on the left edge. Meanwhile, on the right edge, the iso-parameter lines are evenly spaced again.

This approach gives infinite slowdown of m as u approaches zero. As shown in Figure 3.7, without it, iso-parameter lines in u bunch up at the right edge of the patch, eventually causing a kink to appear if the bunching becomes too severe, despite the fact that the original control points in this example were regularly spaced. However, with this approach, one can see that the iso-parameter lines remain evenly spaced, eliminating these kinks.

3.3 Branch Curve Patches

The exact same approach used for edge curves can now be applied to branch curves. A control curve in r on the edge of each of the patches intersecting along the curve provides the means to ensure that the boundaries of each of the medial sheets meet.

The primary difficulty lies at the fin creation point, where the corner on the fin sheet requires both edge curve and branch curve boundary conditions to be enforced on different edges of the same patch. This is done by adding a control curve to m as well as r on one edge of the other two patches, so that a control curve is needed on only one edge of the fin patch. A small number of additional constraints is needed at the fin point itself, but because they are only required at a single point, they can be enforced directly by adjusting control points. The remainder of this section describes how all of this is done.

We gave a succinct expression of the boundary conditions along a branch curve in (2.5), which we repeat for reference:

$$\nabla r^{(k\oplus 2)} - \nabla r^{(k\oplus 1)} = \mathcal{N}^{(k)} \cdot \sqrt{1 - \|\nabla r^{(k)}\|^2} \,, \tag{3.33}$$

However, it is not obvious how these branch conditions should be enforced. We motivate the solution with a geometric approach.

With some algebra, we can decompose ∇r into a component $\nabla r^{(v)}$ in the m_v direction and a component $\nabla r^{(\perp v)}$ in the orthogonal direction $m_u^{(\perp v)} = m_v \times (m_u \times m_v)$:

$$\nabla r = \nabla r^{(v)} \frac{m_v}{\|m_v\|} + \nabla r^{(\perp v)} \frac{m_u^{(\perp v)}}{\left\|m_u^{(\perp v)}\right\|} \,, \tag{3.34}$$

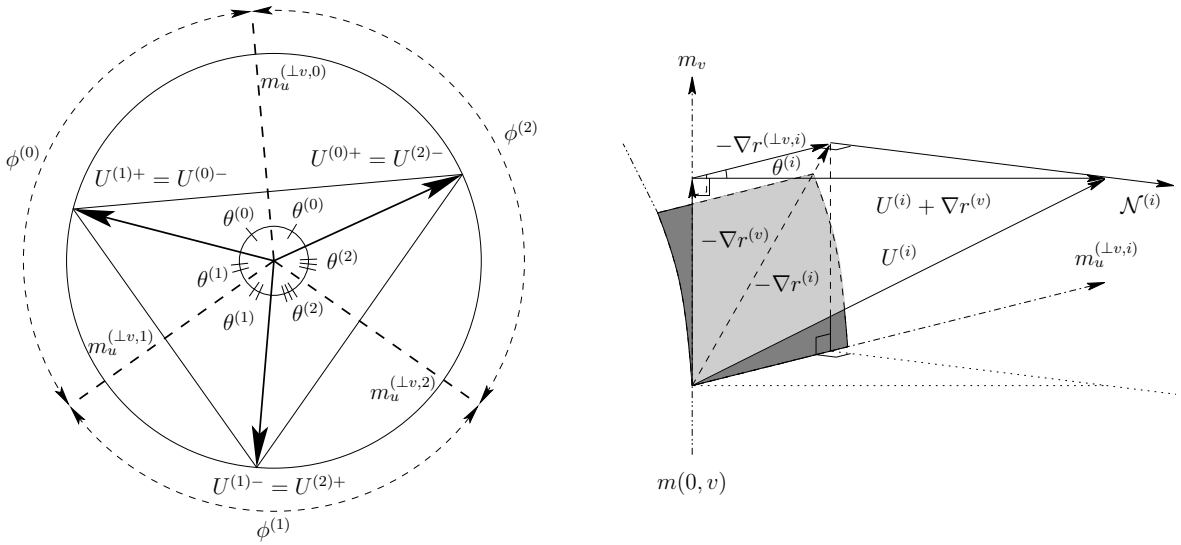
$$\nabla r^{(v)} = \frac{r_v}{\sqrt{G_m}} \,, \qquad\qquad \nabla r^{(\perp v)} = \frac{r_u G_m - r_v F_m}{\sqrt{G_m(E_m G_m - F_m^2)}} \,. \tag{3.35}$$

Without loss of generality, we assume our three patches are oriented so that they meet at the $u = 0$ curve in each. Then r_v and G_m are the same in each patch, and hence so is $\nabla r^{(v)}$.

Now, since all three normal vectors $\mathcal{N}^{(i)}$ are perpendicular to m_v, they lie in the same plane, and hence so do the endpoints of $\nabla r^{(i)}$ and $U^{(i)\pm}$. The *vectors* $\nabla r^{(i)}$ and $U^{(i)\pm}$ do not lie in this plane, just their tips. We project everything into this plane, as illustrated in Figure 3.8.

We use $\theta^{(i)}$ to denote the angle between $\nabla r^{(\perp v,i)}$ and $U^{(i)\pm}$ and use $\phi^{(i)}$ to denote the counter-clockwise angle between $m_u^{(\perp v,i)}$ and $m_u^{(\perp v,i\oplus 1)}$, i.e.

$$\phi^{(i)} = \theta^{(i)} + \theta^{(i\oplus 1)} \,. \tag{3.36}$$



(a) (b)

Figure 3.8: (a) The cross-section of a branch curve, with m_v pointing out of the page. (b) The projection of one of the $U^{(i)}$ vectors into this plane along $\nabla r^{(v)}$.

As one can see, the sums of the angles $\theta^{(i)}$ must be π, and the vectors $m_u^{(\perp v,i)}$ must bisect these angles. This provides a full set of geometric constraints. There are two basic ways these constraints can be enforced, and we have some freedom to choose how we go about it. Either we can adjust $r_u^{(i)}$ to produce the desired angles $\theta^{(i)}$, or we can adjust the orientation of the tangent plane via $m_u^{(i)}$ to produce the desired angles $\phi^{(i)}$.

3.3.1 Satisfying Branch Conditions Away From Fin Creation Points

Some adjustment of $r_u^{(i)}$ is always necessary to ensure that the values of $\theta^{(i)}$ sum to π. Therefore, away from a fin creation point, we hold the tangent vectors $m_u^{(i)}$—and thus the angles $\phi^{(i)}$—fixed and enforce the boundary condition solely via $r_u^{(i)}$. Starting from $\phi^{(i)}$ and solving for $\theta^{(i)}$ in (3.36), we have

$$\theta^{(i)} = \frac{1}{2}(\phi^{(i)} + \phi^{(i\oplus 2)} - \phi^{(i\oplus 1)}) \ . \tag{3.37}$$

This solution is valid (e.g., $\theta^{(i)} \geq 0$) so long as no single $\phi^{(i)}$ is greater than π. A negative $\theta^{(i)}$ indicates that the spokes have pierced through one of the medial sheets.

From the right triangles in Figure (b), we can see that

$$\nabla r^{(\perp v,i)} = -\|U^{(i)} + \nabla r^{(v)}\| \cos\theta^{(i)} = -\sqrt{1 - \frac{r_v^2}{G_m}} \cos\theta^{(i)} \ . \tag{3.38}$$

This constraint can be enforced with a control curve exactly as in Section 3.2. First, we solve (3.38) for $r_u^{(i)}(0,v)$ while holding $m_u^{(i)}(0,v)$, $m_v(0,v)$, and $r_v(0,v)$ fixed:

$$r_u^{(i)} = \frac{1}{G_m}\left(r_v F_m^{(i)} - \cos\theta^{(i)} \sqrt{(G_m - r_v^2)(E_m^{(i)} G_m - {F_m^{(i)}}^2)}\right) \ . \tag{3.39}$$

This solution is defined so long as $\nabla r^{(v)}$ is not already large enough to make $\|\nabla r\| \geq 1$ by itself. Given $r_u(0,v)$, we follow the procedure outlined in Section 3.2.2 to construct a control curve that produces this derivative along the entire edge of the patch.

3.3.2 Satisfying Edge and Branch Conditions at a Fin Creation Point

The fin point is represented as a corner on the fin sheet and a dart vertex on the main medial sheet. Catmull-Clark subdivision does not produce B-spline patches around a dart vertex due to the adjacent crease vertices, but after the first two levels of subdivision, we approximate the medial surface with B-spline patches that interpolate the same branch curve. Since we are about to modify these patches to enforce the branch conditions this deviation from the Catmull-Clark scheme is relatively inconsequential. However, it simplifies the analysis a great deal by avoiding the introduction of a control curve on a patch evaluated with Stam's eigendecomposition method.

Without loss of generality, we will assume $m^{(0)}$ is the fin patch. This patch contains a corner vertex at $(0,0)$ and must satisfy the edge condition $\|\nabla r\| = 1$ on the $v = 0$ edge as well as the branch conditions on the $u = 0$ edge. Since we cannot introduce a control curve for two adjacent edges of the same patch, we choose to enforce the edge constraint with a control curve that modifies $r_v(u,0)$. This fixes $\theta^{(0)}$, so we must enforce the branch conditions by adjusting the remaining angles. This poses two problems. First, the introduction of the edge control curve cannot change $m^{(0)}(0,v)$ or $r^{(0)}(0,v)$, or they will no longer match the other two branch curve patches. Second, at the fin point itself, the other two patches must blend smoothly into the rest of the medial sheet, which limits the adjustments we can make to these patches in order to enforce the branch conditions. However, both of these are much simpler to deal with than

enforcing the general branch conditions along the entire edge, because they can be encapsulated by a finite set of conditions at a single point. Therefore, we can deal with them by adjusting control points before applying the control curve.

The first constraint at a fin creation point is that

$$\theta^{(0)}(0) = 0 \ . \tag{3.40}$$

This is caused by the two sheets $m^{(1)}$ and $m^{(2)}$ flattening out, which implies

$$\phi^{(1)}(0) = \theta^{(1)}(0) + \theta^{(2)}(0) = \pi \ . \tag{3.41}$$

This occurs precisely when $\|\nabla r^{(0)}\| = 1$, which we are enforcing already via the edge constraint, so there is no problem there. However, to keep the control curve from modifying $m^{(0)}(0, v)$ and $r^{(0)}(0, v)$, thus violating our initial assumption that these curves are the same in all three patches, we adjust control points to ensure that $\|\nabla r^{(0)}(0, 0)\| = 1$ *before* application of the control curve. This can be done by modifying $r_u^{(0)}(0, 0)$ while holding $r^{(0)}(0, v)$ fixed. This is a simple matter of applying the mask

$$r(\mathbf{P}^{(0)\prime}) = r(\mathbf{P}^{(0)}) + \begin{bmatrix} 0 & \Delta r' & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\Delta r' & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.42}$$

for a suitable choice of $\Delta r'$. Currently we solve for this value (and the values in the subsequent constraints) numerically. The control points we are modifying have been produced after two levels of Catmull-Clark subdivision, so none of the original control points provided by the modeler are affected. Thus no change is required to how the model is parameterized.

We are assuming the edge constraint is enforced using Catmull-Rom as our interpolating spline. Using our fourth-order spline instead to eliminate sympathetic overfolding is a non-trivial matter, as it requires $r_v^{(0)}(0, 0)$ and $m_v^{(0)}(0, 0)$ to both be zero, which will not in general match the other two branch patches.

The next constraint is that $m_u^{(\perp v, 0)}(0, 0)$ bisect the angle between $U^{(1)+}(0, 0)$ and $U^{(2)-}(0, 0)$. Since these two vectors are equal they must lie in the tangent plane of $m^{(0)}(0, 0)$. From (3.35) it should be clear that rotating m_u around m_v does not change the magnitude of $\nabla r^{(v)}$ or $\nabla r^{(\perp v)}$, merely the $m_u^{(\perp v)}$ direction. The coefficients E_m, F_m,

and G_m remain unchanged. Thus, we rotate $m_u^{(0)}(0,0)$ around $m_v(0,0)$ until $U^{(0)\pm}(0,0)$ coincides with $U^{(1)+}(0,0) = U^{(2)-}(0,0)$. This is done without affecting $m^{(0)}(0,v)$ by applying the mask

$$m(\mathbf{P}^{(0)\prime}) = m(\mathbf{P}^{(0)}) + \begin{bmatrix} 0 & \Delta m' & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\Delta m' & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.43}$$

for a suitable choice of $\Delta m'$.

We will see that one additional constraint is necessary to ensure continuity across the $v = 0$ edge of the two non-fin patches, $m^{(1)}$ and $m^{(2)}$, but to motivate it, we must first describe the control curves used on those patches. After applying the edge constraint to $m^{(0)}$, we are taking $\theta^{(0)}$ as fixed, so we adjust $\theta^{(1)}$ and $\theta^{(2)}$ to ensure their sum is π. We denote the amount of the adjustment by α and β, so that

$$\theta^{(1)} = \phi^{(0)} - \theta^{(0)} - \alpha \ , \tag{3.44}$$

$$\theta^{(2)} = \phi^{(2)} - \theta^{(0)} - \beta \ . \tag{3.45}$$

There are two degrees of freedom, α and β, but the sum of the $\theta^{(i)}$ angles provides only one constraint. As a second constraint, we make the amount of the adjustment proportional to the associated $\phi^{(i)}$ angles:

$$\alpha\phi^{(2)} = \beta\phi^{(0)} \tag{3.46}$$

This ensures that a solution exists whenever $\theta^{(0)} \leq \pi$, which is always the case. Then solving for α and β we have

$$\alpha = \phi^{(0)} \cdot \left(1 - \frac{\pi + \theta^{(0)}}{\phi^{(0)} + \phi^{(2)}}\right) \ , \tag{3.47}$$

$$\beta = \phi^{(2)} \cdot \left(1 - \frac{\pi + \theta^{(0)}}{\phi^{(0)} + \phi^{(2)}}\right) \ . \tag{3.48}$$

The resulting angles $\theta^{(1)}$ and $\theta^{(2)}$ are then produced by adjusting $r_u^{(1)}$ and $r_u^{(2)}$ with a control curve.

We now have choices of $\theta^{(i)}$ that sum to π, but $m_u^{(\perp v,i)}$ may not necessarily bisect the angle between its adjacent $U^{(i)\pm}$ vectors. To accomplish this, we rotate $m_u^{(1)}$ by

α and $m_u^{(2)}$ by $-\beta$ around m_v. We can now apply the procedure in Section 3.2.2 to m_u instead of r_u, producing the desired tangent plane. Because these rotations do not change the magnitude of $\nabla r^{(\perp v,i)}$, just the direction, these two control curves are completely independent of each other. Therefore it does not matter in which order they are applied.

We now return to the final constraint that must be imposed at the fin creation point to ensure continuity across the $v = 0$ edge on the non-fin patches. From (3.41) we have $\phi^{(0)}(0) + \phi^{(2)}(0) = \pi$, which combined with (3.40) implies $\alpha = \beta = 0$. Thus, $\theta^{(1)}(0)$ and $\theta^{(2)}(0)$ remain unchanged and no rotation is applied to $m_u^{(1)}(0,0)$ and $m_u^{(2)}(0,0)$. However, the modifications produced by the remainder of the control curve can affect $r_v^{(1)}(u,0)$, $m_v^{(1)}(u,0)$, $r_v^{(2)}(u,0)$, and $m_v^{(2)}(u,0)$ for any $u > 0$. If these derivatives are modified, these two branch patches will no longer smoothly blend into the rest of the medial sheet with C^1 continuity. These derivatives are held fixed precisely when the two derivatives $\alpha_v(0) = \beta_v(0) = 0$. Both of these latter derivatives are proportional to $\phi_v^{(0)} + \phi_v^{(2)} + \theta_v^{(0)}$. Recall that $\phi^{(0)}$ and $\phi^{(2)}$ are functions of the angles between the patches before any rotation is applied to them, which by construction join the rest of the medial sheet with C^2 continuity at the fin point. Beyond the fin creation point $\phi^{(0)} + \phi^{(2)} = \pi$ everywhere, therefore the derivatives $\phi_v^{(0)} + \phi_v^{(2)}$ must be identically zero. The constraint we want to enforce thus reduces to

$$\theta_v^{(0)}(0) = 0 \ . \tag{3.49}$$

This holds precisely when

$$\frac{\partial \|\nabla r^{(0)}\|^2}{\partial v} = 0 \ . \tag{3.50}$$

This constraint can be enforced by adjusting control points to modify $r_{uu}^{(0)}(0,0)$ while holding $r_u^{(0)}(0,0)$ and $r^{(0)}(0,v)$ fixed by applying the mask

$$r(\mathbf{P}^{(0)\prime\prime}) = r(\mathbf{P}^{(0)\prime}) + \begin{bmatrix} 0 & -2\Delta r'' & 0 & 0 \\ 0 & \Delta r'' & 0 & 0 \\ 0 & -2\Delta r'' & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.51}$$

for a suitable choice of $\Delta r''$. With this final constraint, the formulation for fin points is complete.

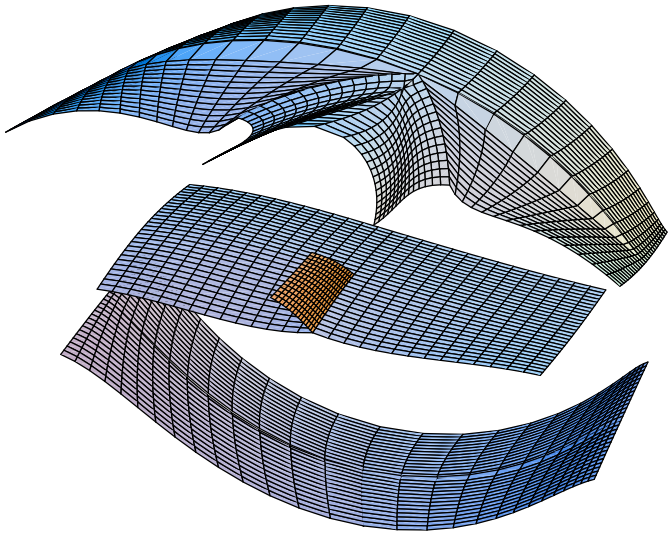


Figure 3.9: An example of the interpolated patches near a fin creation point.

3.3.3 Transition Region

Section 3.3.1 and Section 3.3.2 have described two different procedures for enforcing the branch conditions on different parts of the branch curve. We could have continued the procedure for fin points along the whole curve, but we have no guarantee that the same sheet will be the fin sheet at the other end of the branch curve, so the first procedure, which is agnostic to which sheet is which, is preferred. Therefore we introduce a transition region that moves from the procedure at fin points to that away from them.

Our approach is to blend the two strategies together using a weight function

$$\omega(v) = 1 - 3v^2 + 2v^3 \ , \tag{3.52}$$

chosen specifically so that $\omega(0) = 1$, $\omega(1) = 0$, and $\omega_v(0) = \omega_v(1) = 0$. Then in the

transition region we use

$$\theta^{(i)} = \omega \cdot \theta^{(i)}_{\text{fin}} + (1-\omega) \cdot \theta^{(i)}_{\text{branch}} \ , \qquad \alpha = \omega \cdot \alpha_{\text{fin}} \ , \qquad \beta = \omega \cdot \beta_{\text{fin}} \ , \tag{3.53}$$

where $\theta^{(i)}_{\text{fin}}$, α_{fin}, and β_{fin} are chosen in accordance with the fin point procedure, and $\theta^{(i)}_{\text{branch}}$ is chosen in accordance with the procedure away from fin points.

Because of ω, this will connect the two strategies while retaining C^1 continuity at the patch edges. Other choices of ω are possible, e.g., $\cos(\frac{\pi}{2}v)$, but we have not experimented with them. Another strategy we have not explored is adjusting control points to enforce an explicit set of continuity constraints where the patches meet, like we do at the fin point. The set of constraints is finite, so this should at least in theory be possible.

3.4 Summary and Conclusion

The entire model evaluation process used in the rest of this dissertation is summarized as follows. The model is constructed from a set of control points in m and r arranged in a mesh with ordinary, corner-free boundary. Two levels of Catmull-Clark subdivision are used to construct a collections of patches with creases and edges representing branch and edge curves. The control points around the edge are adjusted via (3.30) to limit sympathetic overfolding, using $\zeta = \frac{11}{8}$. Then the control points around a fin point are adjusted using the masks in (3.42), (3.43), and (3.51). A B-spline like (3.2) is now used to evaluate each interior patch. On patches containing an extraordinary vertex, the actual spline control points are computed using Stam's evaluation method, summarized in (3.7). On those patches containing an edge or branch curve, the evaluation instead uses a two-step asymmetric process. First, the right half of (3.2) is expanded for a specific value of v to produce four control points for a spline in u, where here the v coordinate goes along the edge, and u goes across it. These are converted to the Catmull-Rom basis using (3.10). The first derivatives m_u, m_v, and r_v at $u = 0$ are also evaluated and used to compute the desired r_u, e.g., using with (3.18) for edge curves. From there, (3.25) gives the value of the control curve that replaces the first control point, and the remaining half of the spline is evaluated, as in (3.26). The result is a continuously defined medial axis, $\mathcal{M}$. Applying the same procedure with appropriate modifications to the spline basis functions produces derivatives as well, as summarized in Appendix A. With first derivatives the associated boundary can be reconstructed

using (3.3), and second derivatives allow the computation of S_{rad}.

This chapter has presented a new, generative, continuous 3D medial model. Unlike the problem of computing the medial axis of a given boundary model, the reverse problem of computing the boundary associated with a given medial model is stable. This makes generative models important tools for shape analysis based on the medial axis.

Our model builds on work by (Yushkevich et al., 2003) and removes most of its major limitations. It uses subdivision surfaces, which are closely related to the B-splines used by Yushkevich's original approach. These are highly efficient and well-researched modeling tools, and because of this we can leverage existing work in this field (for example, if we wanted to make the limit surface around extraordinary vertices G^2). Yet we are able to provide a fixed parameterization domain for each model (allowing statistical analysis of model populations) without resorting to a differential equation approach, which is more cumbersome. Finally, ours is the first continuous 3D model that is not restricted to a single medial sheet.

There are, however, still some limitations to our approach. The lack of corners along the edge of the medial sheets, while a desirable feature for most biological objects, could be a hindrance for synthetic ones. Furthermore, we do not explicitly enforce the $\|\nabla r\| \leq 1$ constraint off the edge, nor the Radial Curvature Condition. Instead, these are enforced during model optimization (described in the next chapter) by sampling the medial axis and checking them at the discrete samples. This requires sampling at the finest resolution at which the model will be used (which may not be known a priori) and does not guarantee that a finer sampling will not locate places where these constraints are violated. An analytic method of identifying whether or not a patch satisfies these constraints (without sampling) would be of great utility in creating robust models.

We have provided practical examples of all the concepts in this chapter using the symbolic computation package MATHEMATICA to prove their feasibility. In addition, we have written a C library for computing and displaying single-sheet models, which will be described in more detail in the next chapter. Without any special optimization, this library can compute sampled boundary representations for a medial model at interactive rates.

Much future work remains. First and foremost is the extension of our C library to support branching. Chronologically, the theory behind branching was one of the last pieces developed for this dissertation, and incorporating it into a practical, useful tool is an important next step. Additionally, we would like to explore ways of supporting

both our fourth order spline for eliminating sympathetic overfolding and branching in the same model. This may be as simple as transitioning back to the Catmull-Rom spline at fin creation points, or forcing $r_v(0,0) = m_v(0,0) = 0$ on all three patches at the fin point to allow the branch curve to match. Initial explorations of the latter route have raised the problem that $\theta_v^{(0)}$ becomes independent of $r_{uu}^{(0)}$, meaning a different set of control point adjustments would need to be found to enforce the constraint in (3.49). Finally, the same techniques we use to handle fin creation points should be applicable to six-junctions as well, the last step needed to provide complete support for all of the generic medial topologies in 3D. The challenge there is identifying a way to apply enough control curves to satisfy all of the branch conditions on all six patches when each one has a corner.

Chapter 4

Model Fitting With Explicit Correspondence Optimization

> The problem is closely akin to that of the cartographer who transfers identical data to one projection or another; and whose object is to secure (if it be possible) a complete correspondence, *in each small unit of area*, between the one representation and the other.
>
> (Thompson, 1917)

Having described an object model, this chapter turns to the problem of producing a collection of models that represent a collection of shapes to be used in statistical tests. The basic approach is to start with a template object, described medially, with a fixed branching structure, to align a copy of that template to each target shape, and then to deform it to match. This work only considers fitting a model to an existing, segmented shape, represented as a binary image, a triangulated surface, etc. The approach described is readily adaptable, however, to the job of segmentation itself when coupled with a statistical model of the shape variability and an image match function that determines how well a shape matches the data from a particular image.

The chapter begins by describing a method for sampling a continuous medial axis and then shows how this sampling can be used to approximate medial integrals. Computing volume overlap, an error metric used to evaluate goodness of fit, is used as a motivating example. The complete process for fitting a single model to a single target shape follows. This is involves aligning a template to the target shape and deforming it using a multi-scale constrained optimization. Finally, this is extended to a technique for producing a population of models with a common correspondence. This is

integrated directly into the deformation process by deforming the model in a manner approximately transverse to the fitting process. The models are mapped to a common coordinate system and the control points are adjusted to achieve a parameterization that matches an explicit correspondence given on the boundary, as illustrated in Figure 4.1. This is done by taking advantage of the inherent link between the boundary and the medial axis given by a continuous medial model, moving the problem of formulating correspondence to the boundary, where most of the information lies.

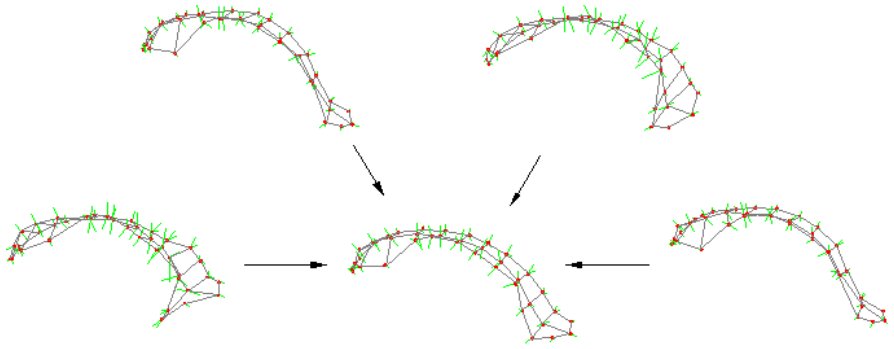


Figure 4.1: The entire population of objects is mapped to a common coordinate system to optimize correspondence.

This chapter considers two specific discretized representations for the target shapes, though of course many others are possible. One is the binary image, $I(x) : \mathbb{R}^3 \rightarrow \{0, 1\}$, defined to be 1 when x is in the shape's interior, and 0 elsewhere. The image is defined on a regular grid—though the spacing along each axis may not be the same—and extended to the whole space via some interpolation method. The second representation is the triangle mesh, T, composed of a set of (oriented) triangles. The fitting process for the latter is not yet implemented, but a complete description of the mathematics involved given here shows how it could easily be incorporated into this optimization framework.

In principle one representation can be converted to the other. Algorithms such as Marching Cubes could construct a mesh from a binary image. Similarly the function $I(x)$ could be implemented as a series of inside-outside tests against the triangular mesh, discretizing the result on a regular grid (though of course this discretization is not always required). However, these conversions can have difficult to predict effects on the model fitting process, making it desirable to work with the original representations when possible.

4.1 Approximating Medial Integrals

Most of the functions used during model fitting are expressed as medial integrals. Medial coordinates provide a natural, object-centric coordinate system for these problems, and being able to operate directly in these coordinates makes object-centric tasks straightforward. The actual computation of these integrals is done by simple numerical approximation.

4.1.1 Sampling the Medial Axis

The first step in computing a medial integral is defining a sampling of the medial axis. The same sampling will be used for display purposes and to check the boundary for illegalities, e.g., using the radial curvature condition defined in Section 2.3.4. One could simply evaluate the splines at a given, fixed resolution, but choosing an appropriate one is difficult. Near the edges of the medial axis, the rate of change of the spoke direction for an infinitesimal step along the medial axis diverges to infinity. Even away from the edge, the unit piece of surface area on the boundary can change substantially relative to a unit piece of area on the medial axis. The real goal is to limit the amount of surface area on the boundary each sample represents to something, say, commensurate with the size of a voxel in a binary image or triangle in a mesh. Recursive subdivision drives this process. It is also possible to to limit the volume element instead of the surface area element, but most of the integrals we are approximating are over the boundary, and using additional subdivision along the spoke direction provides a more uniform sampling of the object interior when required.

The sampling consists of a tiling of the unit square in the (u, v) plane for each patch, starting with a single sample at $(u, v) = (\frac{1}{2}, \frac{1}{2})$, with sample dimensions $\Delta u = \Delta v = 1$. This represents the rectangle in the (u, v) plane with corners $(u - \frac{1}{2}\Delta u, v - \frac{1}{2}\Delta v)$ and $(u + \frac{1}{2}\Delta u, v + \frac{1}{2}\Delta v)$. Subdivision proceeds in the u or v direction (or both) until the corresponding area on the boundary is sufficiently small. For example, to subdivide in the u direction, the original sample at (u, v) is replaced with two samples at $(u - \frac{1}{4}\Delta u, v)$ and $(u + \frac{1}{4}\Delta u, v)$, and Δu is scaled by a factor of $\frac{1}{2}$. The process is then repeated on the new samples.

The stopping condition is based on the differentiable unit of area on the boundary, which is given by

$$d\mathcal{B}^{\pm} = \|\mathcal{B}_u^{\pm} \times \mathcal{B}_v^{\pm}\| \, du \, dv \ . \tag{4.1}$$

By applying (3.3) this can be expanded to

$$d\mathcal{B}^{\pm} = \left\| \left(m_u + r_u U^{\pm} + r U_u^{\pm}\right) \times \left(m_v + r_v U^{\pm} + r U_v^{\pm}\right) \right\| \, du \, dv \tag{4.2}$$

Expressions for computing $U_u^{\pm}$ and $U_v^{\pm}$ are given in Appendix A. The formula in (4.2) is precisely equivalent to the $\det\left(\mathbf{I} - r S_{\text{rad}}\right) dM$ given in Section 2.3.5, but it makes explicit which portion of the area comes from a step in the u direction, and which from a step in v. This allows refinement of the sampling in one variable without affecting the other. This avoids very long, very thin samples near the edge of the medial axis, where a small step in the u direction (towards the edge) produces a large change in $U^{\pm}$, but a step in the v direction (along the edge) does not. Always subdividing in both dimensions would make getting below the requisite area bound require very small steps in both u and v.

Replacing $du\,dv$ above by $\Delta u \Delta v$, the unit steps in u and v at the current sampling level, the sample area falls below a threshold τ^2 when

$$\|\mathcal{B}_u^{\pm} \Delta u \times \mathcal{B}_v^{\pm} \Delta v\| < \tau^2 \ . \tag{4.3}$$

This is true if both

$$\|\mathcal{B}_u^{\pm} \Delta u\| < \frac{\tau}{\sqrt{|\sin\theta|}} \qquad \text{and} \qquad \|\mathcal{B}_v^{\pm} \Delta v\| < \frac{\tau}{\sqrt{|\sin\theta|}} \tag{4.4}$$

are satisfied, where θ is the angle between $\mathcal{B}_u^{\pm}$ and $\mathcal{B}_v^{\pm}$. The $\sqrt{|\sin\theta|}$ term adjusts the threshold to correct for the degree to which $\mathcal{B}_u^{\pm}$ and $\mathcal{B}_v^{\pm}$ fail to be orthogonal. Unless otherwise specified, the sampling resolution τ is the (geometric) average voxel spacing in each direction for binary images or the (geometric) average edge length for triangle meshes.

Noting that

$$|\sin\theta| = \frac{\|\mathcal{B}_u^{\pm} \Delta u \times \mathcal{B}_v^{\pm} \Delta v\|}{\|\mathcal{B}_u^{\pm} \Delta u\| \|\mathcal{B}_v^{\pm} \Delta v\|} \ , \tag{4.5}$$

the check for the first half of (4.4) reduces to

$$\|\mathcal{B}_u^{\pm} \Delta u\|^2 \|\mathcal{B}_u^{\pm} \Delta u \times \mathcal{B}_v^{\pm} \Delta v\|^2 < \|\mathcal{B}_v^{\pm} \Delta v\|^2 \tau^4 \ . \tag{4.6}$$

A similar check can be obtained for the second half by swapping u and v. Thus, once $\mathcal{B}_u^{\pm}$

and $\mathcal{B}_v^{\pm}$ are available, the check for whether or not to stop subdivision can be computed with just one cross product and three dot products. Computing these checks for both the top and bottom sides, subdivision continues if either fails.

At each level of subdivision, samples are split in u, v, or both, depending on which thresholds are exceeded, and replaced with two or four new samples at the next lower subdivision level. As a practical matter, because the sample area is only an approximation, once u or v is left unsplit at a given level, that threshold may later be exceeded again at a deeper level. However, this happens seldom. To simplify the data structures involved, once the threshold is satisfied for one of the variables in a particular sample, its descendants are never subdivided in that direction again. Any further subdivision is done instead with respect to the other variable only.

There is also no inherent requirement that the sampling on the top side be the same as the sampling on the bottom side, and indeed the curvature on the boundary may differ greatly between the two, for example when the medial sheet itself is curved. However, in the current implementation, subdivision always proceeds on both sides simultaneously. This can lead to a larger number of samples than is necessary on the concave side of such a curved medial sheet, but this gives a greater chance of detecting overfolding in this region.

A simplistic recursive scheme yields a polygon mesh for display purposes. A tree traversal connects adjacent points at the same subdivision level with edges and connects adjacent points at different levels with triangle fans. To save memory, the list of faces and edges is determined implicitly directly from the tree structure and never stored.

This scheme avoids sampling at the edge of any of the patches, which avoids all singular points as well as extraordinary vertices. This means that S_{rad} is well defined at each sample location, ensuring that the radial curvature condition can be checked to avoid overfolding and making the point's contribution to a numeric approximation of a medial integral simple. The problems that arise with some patch edges only affect the computation of S_{rad}, either because $\mathcal{M}$ is not C^2 continuous at these points due to the presence of a control curve on the patch, or because they are edge points or extraordinary vertices where S_{rad} itself diverges. The radial vectors U and the associated points on the boundary are still well-defined at these points, and for display purposes it might even be a good idea to compute them, especially along the crest region. However, the size of a local surface patch around these boundary points is no longer well-approximated by derivatives at the patch center. The places that are not C^2 continuous could be handled by using one-sided derivatives to compute multiple values

of S_{rad}, and at edge points we could use S_E instead, if we could find a suitable edge coordinate parameterization. But because the set of such points has measure zero, the simple route is to avoid them altogether.

Taking the simple route produces a *sampling* where the local patch areas are easy to compute, a term whose definition follows. A sampling is a set of sample points $\mathbb{S}$ consisting of their (u, v) coordinates as well as the patch they are located on, along with the area each sample covers in parameter space, $\Delta u \Delta v$. Abusing notation, a tuple of functions evaluated at a single point (u, v) on some patch is treated as if they were part of the sampling itself. Therefore, $(m_0, r, U_0, \Delta M) \in \mathbb{S}$ stands for the values $m(u, v)$, $r(u, v)$, $U^+(u, v)$ or $U^-(u, v)$, and $\Delta M(u, v)$. The latter is defined as

$$\begin{aligned} \Delta M(u,v) &\triangleq \rho(u,v)\,\|m_u(u,v) \times m_v(u,v)\|\,\Delta u \Delta v \\ &= U^+(u,v) \cdot (m_u(u,v) \times m_v(u,v))\,\Delta u \Delta v \\ &= -U^-(u,v) \cdot (m_u(u,v) \times m_v(u,v))\,\Delta u \Delta v\ . \end{aligned} \tag{4.7}$$

The two tuples corresponding to $U^{\pm}$, that is $(m(u,v), r(u,v), U^+(u,v), \Delta M(u,v))$ and $(m(u,v), r(u,v), U^-(u,v), \Delta M(u,v))$, are considered two distinct elements of $\mathbb{S}$. Depending on the precise usage of the sampling in a given context, it will also include various derivatives of m, r, and values constructed from them, such as S_{rad} (again with one copy for each side). This will hopefully become clear as several examples illustrate how a sampling is used to approximate medial integrals.

4.1.2 Moment Integrals

The standard moment integrals up through second order describe the distribution of mass in an object. They can be thought of as approximating the object by an ellipsoid, or equivalently, a multivariate Gaussian distribution. They are interesting primarily as a means of aligning two different shapes into a common coordinate system, though such a simple alignment scheme may fail or be inadequate for bent or twisted objects.

Section 2.3.5 gave formulas for the integrals, using an exact analytic expression for $\tilde{g}$. To evaluate them numerically, one simply sums up the contribution from each sample. For example, to compute the volume of an object (from (2.28)),

$$V(\Omega) \approx \sum_{(r, H_{\text{rad}}, K_{\text{rad}}, \Delta M) \in \mathbb{S}} \left(r - r^2 H_{\text{rad}} + \frac{1}{3} r^3 K_{\text{rad}} \right) \Delta M\ . \tag{4.8}$$

A similar approximation can be constructed for both (2.30) and (2.32).

$$\overline{\Omega} \approx \frac{1}{V(\Omega)} \sum_{(m_0,r,U_0,H_{\text{rad}},K_{\text{rad}},\Delta M)\in\mathbb{S}} \left(m_0 \left(r - r^2 H_{\text{rad}} + \frac{1}{3} r^3 K_{\text{rad}} \right) + U_0 \left(\frac{1}{2} r^2 - \frac{2}{3} r^3 H_{\text{rad}} + \frac{1}{4} r^4 K_{\text{rad}} \right) \right) \Delta M \ , \quad (4.9)$$

$$\mathbf{I}_\Omega \approx \sum_{(m_0,r,U_0,H_{\text{rad}},K_{\text{rad}},\Delta M)\in\mathbb{S}} \left(m_0^T m_0 \left(r - r^2 H_{\text{rad}} + \frac{1}{3} r^3 K_{\text{rad}} \right) + \left(m_0^T U_0 + U_0^T m_0 \right) \left(\frac{1}{2} r^2 - \frac{2}{3} r^3 H_{\text{rad}} + \frac{1}{4} r^4 K_{\text{rad}} \right) + U_0^T U_0 \left(\frac{1}{3} r^3 - \frac{1}{2} r^4 H_{\text{rad}} + \frac{1}{5} r^5 K_{\text{rad}} \right) \right) \Delta M \ - V(\Omega) \overline{\Omega}^T \overline{\Omega} \ . \quad (4.10)$$

4.1.3 Estimating Volume Overlap

Volume overlap is a standard measure of how close the shape given by a model, Ω_m, and the target shape, Ω_t, are to each other. It can be defined in several ways, but herein shall be

$$V_{\text{overlap}}(\Omega_m, \Omega_t) \triangleq \frac{V(\Omega_m \cap \Omega_t)}{V(\Omega_m \cup \Omega_t)} = \frac{V(\Omega_m \cap \Omega_t)}{V(\Omega_m) + V(\Omega_t) - V(\Omega_m \cap \Omega_t)} \ . \quad (4.11)$$

Some authors also call this the Jaccard coefficient (Jaccard, 1901) or sometimes even erroneously the Rogers-Tanimoto coefficient (Rogers and Tanimoto, 1960)[1]. Other authors use $\frac{1}{2}(V(\Omega_m)+V(\Omega_t))$ for the denominator, also called the Czekanowski, Sorenson, or Dice coefficient (Czekanowski, 1932; Sorenson, 1948; Dice, 1945) (or various hyphenated combinations thereof). This results in higher overlap measurements than with our definition. The two are related by $V_{\text{Dice}} = 2\frac{V_{\text{overlap}}}{1+V_{\text{overlap}}}$.

The volume overlap of a medial model with a binary image is again expressed as a medial integral. Letting Ω_t be the shape defined by the image, counting the number of voxels x such that $I(x) = 1$ and scaling by the voxel spacing yields an estimate for $V(\Omega_t)$. Similarly, letting Ω_m be the shape defined by a medial model, via the approximation in (4.8) provides an estimate for $V(\Omega_m)$. Another medial integral is constructed to compute $V(\Omega_m \cap \Omega_t)$.

[1] The real definition of this latter coefficient also includes terms for the points that are in neither set, which is clearly only useful in bounded spaces.

The function being integrated is still $g = 1$, but its domain is restricted to the specific region $\Gamma = \Omega_m \cap \Omega_t$. By Corollary 7 of (Damon, 2005a), if $\Gamma \subseteq \Omega$ is a Borel measurable region, then this can be done simply by replacing g with $\chi_\Gamma \cdot g$ in the definition of a skeletal integral given in (2.13) and (2.14):

$$\tilde{g}_\Gamma(m_0, S_0) = \int_0^1 \chi_\Gamma(m_0 + tS_0) \cdot g(m_0 + tS_0) \cdot \det\left(\mathbf{I} - t \cdot r S_{\text{rad}}\right) \, dt \tag{4.12}$$

$$\int_\Gamma g \, dV = \int_{\tilde{M}} \tilde{g}_\Gamma \cdot r \, dM \ . \tag{4.13}$$

Here $\chi_\Gamma(x)$ is the characteristic function of Γ, defined to be 1 when $x \in \Gamma$, and 0 otherwise. This is precisely the definition of the binary image $I(x)$ (restricted to the interior of Ω_m).

For binary images, instead of deriving an exact formula for $\tilde{g}_\Gamma$, it is approximated numerically given a tuple $(m_0, r, U_0) \in \mathbb{S}$. The number of steps used for the approximation is $n_{\text{steps}} \triangleq \lceil \frac{r}{\tau} \rceil$. An approximation for $m_j(g_\Gamma)$ from (2.22) is given by

$$m_0(g_\Gamma) \approx \sum_{i=1}^{n_{\text{steps}}} I(m_0 + t_i \cdot rU_0) \cdot \Delta t \ , \tag{4.14}$$

$$m_1(g_\Gamma) \approx \sum_{i=1}^{n_{\text{steps}}} I(m_0 + t_i \cdot rU_0) \cdot t_i \Delta t \ , \tag{4.15}$$

$$m_2(g_\Gamma) \approx \sum_{i=1}^{n_{\text{steps}}} I(m_0 + t_i \cdot rU_0) \cdot \left(t_i^2 + \frac{\Delta t^2}{12}\right) \Delta t \ . \tag{4.16}$$

where $t_i = \frac{2i-1}{2n_{\text{steps}}}$ and $\Delta t = \frac{1}{n_{\text{steps}}}$. The terms Δt, $t_i \Delta t$, and $(t_i^2 + \frac{\Delta t^2}{12})\Delta t$ represent the exact analytic integrals of t^j from $(t_i - \frac{1}{2}\Delta t)$ to $(t_i + \frac{1}{2}\Delta t)$ for $j = 0, 1, 2$, respectively. The value of the binary image, $I(x)$, is sampled using nearest neighbor interpolation. Plugging these sums into (2.23) gives an approximation for $\tilde{g}_\Gamma$, and thus the full medial integral.

For triangle meshes, the algorithm given in (Mirtich, 1996) allows direct computation of $V(\Omega_t)$ and the other moment integrals[2] $V(\Omega_m)$ is computed as before, there is a better approximation of $V(\Omega_m \cap \Omega_t)$. Given all the intersection points of the ray $m_0 + t \cdot rU_0$ with the mesh T, there is an exact formula for the contribution from each

[2] The reader is cautioned that the inertia tensor $\mathbf{J}$ computed by Mirtich is not the same as the second moment tensor $\mathbf{I}_\Omega$ used here, even though the former is used to compute the *moment of inertia* about an axis, and even though they are both denoted $\mathbf{I}$ in various texts. However, Mirtich's algorithm produces all the quantities needed to compute either tensor, and both have the same eigenvectors.

segment (t_0, t_1) of the ray that lies inside the target shape:

$$m_j(g_\Gamma) = \frac{t_1^{j+1} - t_0^{j+1}}{j+1} \ . \tag{4.17}$$

The values (t_0, t_1) are clamped to fall in the interval $[0, 1]$. Summing the contributions for each segment and plugging them into (2.23) again completes the integral.

It is possible, of course, to also find such intersection points in a voxelized image using some kind of variation of Bresenham's line drawing algorithm (Bresenham, 1965). However, such a formulation would be complicated by the need to determine which faces of each voxel represent shape boundaries and of those, which are intersected by the ray, and at what time t. The sampling approach used instead is much simpler to implement.

4.2 Single-Subject Model Fitting

Before describing the process for an entire collection, this section first considers fitting a single model to a single image. This assumes that a template that defines the number and arrangement of control points and the branching topology of the medial axis is given as input. Currently, these templates are hand-constructed, though it is certainly reasonable to expect that one could develop automatic template-creation methods similar to those developed for discrete m-reps (Styner et al., 2003).

4.2.1 Template Alignment

The first step is to align the template model with the target shape. This is done using the moment integrals derived in Section 4.1.2 to derive a similarity transform to apply to the template object. From the volumes $V(\Omega_m)$ and $V(\Omega_t)$, the desired scale factor is given by

$$s = \left(\frac{V(\Omega_t)}{V(\Omega_m)} \right)^{\frac{1}{3}} \ . \tag{4.18}$$

The next step is to compute the matrices $\mathbf{U}_{\Omega_m}$ and $\mathbf{U}_{\Omega_t}$, containing an orthonormal set of right eigenvectors of $\mathbf{I}_{\Omega_m}$ and $\mathbf{I}_{\Omega_t}$, respectively, sorted by the descending order of their associated eigenvalues. This gives the desired rotation:

$$\mathbf{R} = \mathbf{U}_{\Omega_t} \mathbf{U}_{\Omega_m}^{-1} \ . \tag{4.19}$$

There is some ambiguity here, since flipping any of the eigenvectors to point in the opposite direction yields a different orthonormal set of eigenvectors.

Using the combination of orientations that produces the minimum total rotation, e.g., as measured by computing the axis of rotation and looking at the magnitude of the angle of rotation around it, and rejecting orientations that would introduce a reflection into $\mathbf{R}$ resolves this ambiguity. This works so long as the two shapes are not too bent or twisted and are expected to be relatively close in orientation. The latter is a reasonable assumption in the case that the original imagery, such as a whole-brain MRI scan, is stored in a standard coordinate system. Another (more computationally expensive) strategy that could handle more general positioning, including reflections, would be to choose the eigenvector orientations that give maximum volume overlap. To handle heavily bent or twisted objects, one could use a small number of easily identifiable landmarks to drive the alignment, such as the tips of the head or tail of the object.

Technically, alignment via moments may also fail when two or more eigenvalues are very close, since when they are equal there are infinitely many valid sets of orthonormal eigenvectors. This occurs when an object is nearly spherically or cylindrically symmetric.

However, the computed $\mathbf{R}$ yields the following similarity transform to apply to each control point:

$$m' = s\mathbf{R}\left(m - \overline{\Omega}_m\right) + \overline{\Omega}_t \ , \tag{4.20}$$

$$r' = sr \ . \tag{4.21}$$

The application of a scale factor means that the new model will have a different sampling than the original template, because the sample size will fall under the sampling resolution at different points in the subdivision process. Due to approximation errors, the moment integrals of the transformed shape might not not match those of the target shape exactly. Assuming that the template is roughly commensurate with the size of the targets, this scale factor will be near 1, and these errors can be ignored. Performing a pre-scaling using just the volumes and then resampling the model before computing the full similarity transform ensures this assumption is true.

4.2.2 Model Deformation

The next step attempts to find a globally optimal choice of control point values that minimize the average distance to the target shape. This is a global nonlinear opti-

mization problem, to which no well-known panacea gives the solution. Our approach is inspired by continuation methods, which have met with some success in, for example, the biomedical engineering problem of determining the configuration of complex protein molecules (Moreé and Wu, 1997).

Continuation methods attempt to find the global minimum of a function with many local minima by applying a diffusion process driven by the heat equation to smooth the objective function. Standard local minimization techniques are then used to find the minimum of the smoothed function, and this minimum is traced back to the original objective function by gradually reducing the amount of diffusion. As the smoothing is reduced, the local minimum produced by one iteration is used as the initial point for a local minimization process in the next iteration. Continuation methods rely on being able to efficiently solve the heat equation, which has an analytic solution for several interesting classes of functions (Gaussians, polynomials, and trigonometric polynomials) (Kostrowicki and Piela, 1991). Without an analytic solution to this partial differential equation, numeric integration is required, which becomes equivalent to sampling the entire (high-dimensional) domain one is optimizing over.

There is no known analytic solution to the heat equation for our objective functions over the domain of possible control point values. However, applying some diffusion in the spatial domain can still reduce the number and magnitude of local extrema in the target shape. That does not imply that this property pulls back into the control point domain at every scale, but it must certainly do so in the limit, as eventually the shape is obliterated and the objective function becomes uniform. Thus the procedure is to compute a coarse scale representation of the target shape, fit a model to that using local nonlinear optimization, and then reduce the scale and refine the model until reaching the desired sampling resolution.

The Objective Function for Images

The first step is to give a precise definition of the objective function for binary images. This formulation is straightforward, but the formulation for triangle meshes that follows is much more interesting. We convert the target image to a scale-space representation

by convolving it with a family of Gaussians, defined by their scale parameter, σ:

$$g_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{2\sigma^2}} \tag{4.22}$$

$$I_\sigma(x) = (g_\sigma * I)(x) = \int_{y\in\mathbb{R}^n} I(y) g_\sigma(x-y)\, dV \ . \tag{4.23}$$

Scale-space theory (Lindeberg, 1997) tells us that such a family of Gaussians is both necessary and sufficient for producing a multi-scale representation that (along with other invariance properties) does not enhance local maxima or minima. The convolution can done by applying a finite filter to the image, which truncates the rapidly decaying Gaussian after several standard deviations. The boundary $\mathcal{B}_t^\sigma$ of the target shape at a scale σ is then given by the level set

$$\mathcal{B}_t^\sigma = \{x \in \mathbb{R}^n : I_\sigma(x) = \ell_0\} \ . \tag{4.24}$$

Here ℓ_0 is the target level of the level set, chosen to be the threshold that gives the maximum volume overlap with the original shape. This choice helps prevent thin structures from being entirely destroyed, possibly breaking the target shape into several disconnected pieces. As the amount of diffusion decreases, it converges to $\ell_0 = \frac{1}{2}$.

It is possible to optimize over the distance to this set by computing the distance transform of the result, either by thresholding it to a binary image (Maurer et al., 2003) or by explicitly solving for the location of $\mathcal{B}_t^\sigma$, using, e.g., Marching Cubes (Mauch, 2000). However, using level set methods directly can avoid the additional computational expense and discretization errors this would introduce. We use a medial integral on the boundary as defined in (2.15) to construct the objective function:

$$F_I^\sigma \triangleq \frac{1}{2}\int_{\tilde{M}} \left(I_\sigma(m_0 + rU_0) - \ell_0\right)^2 \cdot \det\left(\mathbf{I} - rS_{\text{rad}}\right) dM \ . \tag{4.25}$$

This drives the model surface to fit the same shape as the distance function formulation would.

Unfortunately, this objective function has a trivial global minimum: $\mathcal{M} = 0$ everywhere. This forces the total surface area of the object, and thus the objective function, to zero. One solution would be to divide F_I^σ by the area, but it makes computing derivatives much simpler to fix the weights applied to each sample throughout the

optimization. That is, we use the approximation

$$F_I^\sigma \approx \sum_{(m_0,S_0)\in\mathbb{S}} w_i \cdot \left(I_\sigma(m_0+S_0)-\ell_0\right)^2 \tag{4.26}$$

where

$$w_i = \left|\det\left(\mathbf{I} - rS_{\text{rad}}\right)\Delta M\right| . \tag{4.27}$$

The weights w_i are computed once at the beginning of the optimization and periodically updated throughout, but they are assumed to be constant when computing the gradient of F_I^σ. The absolute value handles the case of overfolding at the current sample when the weights are first evaluated. If the $\|\nabla r\| < 1$ constraint is violated, meaning the spoke S_0 cannot be computed, we temporarily set $w_i = 0$.

In particular, the derivative of F_I^σ with respect to a given control point component ν is given by

$$\frac{\partial F_I^\sigma}{\partial \nu} \approx \sum_{(m_0,S_0)\in\mathbb{S}} w_i \cdot \left(I_\sigma(m_0+S_0)-\ell_0\right)\nabla I_\sigma(m_0+S_0)\cdot\left(\frac{\partial m_0}{\partial \nu}+\frac{\partial S_0}{\partial \nu}\right) . \tag{4.28}$$

Details on how to compute the derivative $\frac{\partial S_0}{\partial \nu} = r_\nu U_0 + r\frac{\partial U_0}{\partial \nu}$ can be found in Appendix B. We sample I_σ at non-lattice positions using linear interpolation (for speed) and choose the gradient ∇I_σ to match this interpolation. Not updating w_i at each step avoids the need to differentiate ΔM and S_{rad} with respect to ν, of which the latter, at least, is relatively expensive to compute. This prevents the optimizer from trying to reduce the objective function by shrinking the surface area of samples with large errors.

The Objective Function for Surfaces

The question of how to introduce a scale space on surfaces is far more subtle, and a very recent approach based on work by (Glaunès et al., 2004; Vaillant and Glaunès, 2005; Glaunès and Joshi, 2006) gives a possible solution. The problem is one of putting a distance metric on two unlabeled surfaces. Our two surfaces can, and will, have different triangulations, with no a priori knowledge about which points correspond to each other, other than the canonical choice of the "closest" points, which are difficult to compute and may not be unique. In fact, this correspondence is not fixed precisely to allow it to change as the optimization proceeds. However, without such a fixed

correspondence, the space of surfaces is infinite dimensional. Either one of the surfaces can have arbitrarily many points, and there is no a priori bound on the number of points computed when sampling a medial model.

Thus we turn to an infinite dimensional vector space, the space of Borel measures on $\mathbb{R}^3$, to represent them. This space has the nice feature that, although infinite dimensional, it can be viewed as a Reproducing Kernel Hilbert Space (Daumé, 2004) with a computationally tractable (for finitely representable elements) inner product and norm. Without delving too deep into the mathematical theory, the measure associated with an oriented surface $T \in \mathbb{R}^3$ is given by the linear functional $[T](\omega) = \int_T \omega$ over 2-forms $\omega(x)$ (see (Vaillant and Glaunès, 2005) for details). This is precisely the measure associated with integrating a vector field over the surface: a collection of unit vectors normal to the surface, $\mathcal{N}_T$, distributed with a density equal to the element of surface area, dT. This is a specific formulation of the more general theory of *currents* from Geometric Measure Theory (Federer and Fleming, 1960), which can generalize the concept to sub-manifolds of any dimension embedded in an ambient space of any larger dimension. In other words, one could apply the concept to point sets, as in (Glaunès et al., 2004), or even hand-drawn contours on a stack of 2-D image slices—a typical method of manual segmentation employed by radiologists. However, this chapter will not go into this generalization further. A suitable matrix-valued kernel $K(x, y) : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ gives the K-norm, defined as

$$\|[T]\|_K^2 = \int_T \int_T \mathcal{N}(y)^T K(x, y) \mathcal{N}(x) dT(x) dT(y) \ . \tag{4.29}$$

For a discretely triangulated surface, the sum over a collection of Dirac delta functions δ_x[3] approximates the integral:

$$\|[T]\|_K^2 \approx \left\| \sum_{i=1}^{n_T} \eta_i \delta_{c_i} \right\|^2 = \sum_{i=1}^{n_T} \sum_{j=1}^{n_T} \eta_i^T K(c_i, c_j) \eta_j \ . \tag{4.30}$$

Here n_T is the number of triangles in T, c_i is the center of mass of the ith triangle, and η_i is the outward pointing (non-unit) normal vector with length equal to the triangle's area. E.g., for a triangle with vertices x_i, y_i, and z_i in counter-clockwise order,

$$c_i = \frac{1}{3}(x_i + y_i + z_i), \qquad \eta_i = \frac{1}{2}(y_i - x_i) \times (z_i - x_i) \ . \tag{4.31}$$

[3] Defined here to be a function $\delta_x(y)$ such that $\delta_x(y) = 0$ when $y \neq x$ and $\int_{y\in\mathbb{R}^3} \delta_x(y)\, dV = x$.

A sampling $\mathbb{S}$ yields a similar measure without needing to triangulate the surface. For a sample $(m_0, r, U_0, \mathcal{B}_u, \Delta u, \mathcal{B}_v, \Delta v) \in \mathbb{S}$ (or equivalently $(m_0, r, U_0, S_{\text{rad}}, \Delta M) \in \mathbb{S}$),

$$c_i' = m_0 + rU_0, \qquad \eta_i' = (\mathcal{B}_u \Delta u) \times (\mathcal{B}_v \Delta v) = U_0 \det(\mathbf{I} - rS_{\text{rad}})\, \Delta M \ . \tag{4.32}$$

All that remains is to choose a suitable kernel, such as a uniform Gaussian $K_\sigma(x, y) = g_\sigma(x - y)\mathbf{I}$. Thus the distance metric naturally has a free scale parameter σ, which again allows a continuation-like method of optimization.

This norm introduces a distance metric between the model boundary $\mathcal{B}$ and the triangulated target surface T via subtraction on measures. Letting $n_{\mathbb{S}}$ denote the number of points in the sampling $\mathbb{S}$, the objective function is

$$\begin{aligned} F_T^\sigma \triangleq \left\| \sum_{i=1}^{n_{\mathbb{S}}} \eta_i' \delta_{c_i'} - \sum_{i=1}^{n_T} \eta_i \delta_{c_i} \right\|^2 &= \sum_{i=1}^{n_{\mathbb{S}}} \sum_{j=1}^{n_{\mathbb{S}}} \eta_i'^T K_\sigma(c_i', c_j') \eta_j' \\ - 2 \sum_{i=1}^{n_{\mathbb{S}}} \sum_{j=1}^{n_T} \eta_i'^T K_\sigma(c_i', c_j) \eta_j &+ \sum_{i=1}^{n_T} \sum_{j=1}^{n_T} \eta_i^T K_\sigma(c_i, c_j) \eta_j \ . \end{aligned} \tag{4.33}$$

The biggest reductions in this distance will come from making the cross terms in the center large, e.g., by making the surfaces "close" to each other. Not only does this require that each point c_i' be near some of the c_j points, but their normal vectors must point in similar directions. It does not have the same trivial global minimum as our objective function for binary images, since even if the surface area of our object goes to zero, there will still be a large contribution from the last term, the norm of the target surface.

The double summations might also at first glance make the method seem expensive to compute. Unlike with binary images, the Gaussian blurring is no longer a simple filter applied on a regular grid. However, one can truncate the Gaussians beyond a certain threshold and take advantage of the fact that the target shape T does not change during the optimization to implement fast query structures to retrieve the nearby points, such as oriented bounding box (OBB) trees (Gottschalk et al., 1996). For large scales or dense point sets one of the Fast Gauss Transforms (Greengard and Strain, 1991; Yang et al., 2003; Lee et al., 2005) provides an $O(n_{\mathbb{S}} + n_T)$ approach, albeit with a large constant factor that depends on the desired accuracy.

The gradient is again given by derivatives with respect to each control point com-

ponent, ν:

$$\begin{aligned}\frac{\partial F_T^\sigma}{\partial \nu} = \sum_{i=1}^{n_{\mathbb{S}}} \Bigg[& \left(\frac{\partial \eta_i'}{\partial \nu}\right)^T \left(\sum_{j=1}^{n_{\mathbb{S}}} K_\sigma(c_i', c_j')\eta_j' - 2\sum_{j=1}^{n_T} K_\sigma(c_i', c_j)\eta_j \right) \\ & + \eta_i'^T \sum_{j=1}^{n_{\mathbb{S}}} \left(\left(\nabla_{c_i'} K_\sigma(c_i', c_j') + \nabla_{c_j'} K_\sigma(c_i', c_j') \right) \eta_j' + K_\sigma(c_i', c_j') \frac{\partial \eta_j'}{\partial \nu} \right) \\ & - 2\eta_i'^T \sum_{j=1}^{n_T} \nabla_{c_i'} K_\sigma(c_i', c_j)\eta_j \Bigg] \ . \end{aligned} \tag{4.34}$$

Here $\nabla_{c_i'} K_\sigma(\cdot,\cdot)$ denotes the directional derivative of the kernel in the $\frac{\partial c_i'}{\partial \nu}$ direction (and similarly for c_j'). That is, for the isotropic Gaussian kernel,

$$\nabla_{c_i'} K_\sigma(c_i', c_j') = -\frac{1}{2\sigma^2} \frac{\partial c_i'}{\partial \nu} \cdot (c_i' - c_j') g_\sigma(c_i' - c_j')\mathbf{I} \ , \tag{4.35}$$

$$\nabla_{c_j'} K_\sigma(c_i', c_j') = \frac{1}{2\sigma^2} \frac{\partial c_j'}{\partial \nu} \cdot (c_i' - c_j') g_\sigma(c_i' - c_j')\mathbf{I} \ . \tag{4.36}$$

4.2.3 Local Nonlinear Optimization

The optimization process itself is skirted over by many researchers, often being summarized in a couple of sentences and seldom described in enough detail to be reproducible by others. In contrast, this section attempts to go into considerable detail, where the strict page limits of conference proceedings and journal publications are not in force. The uninterested reader may skip to the next section after being told that we use a nonlinear conjugate gradient method, resampling the model each time the method is restarted (termed a *macro step* herein). The interested reader is invited to continue.

The model-fitting process is inherently nonlinear, due to the nonlinear relation between the radial vector field U and the radius function r. All nonlinear optimization methods are iterative and depend on starting from a good initial solution. There are two general types of multivariate nonlinear solvers: trust region and line-search methods. Trust region methods estimate a local model of the objective function, as well as the size of the region in which the model is valid, and take steps within that region, updating the region size depending on how closely the actual objective function matched the prediction of the model. Line search methods, on the other hand, choose a single direction at each step, and then perform a one-dimensional optimization along this direction. This can require many more evaluations of the objective function.

Special methods like Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) or Powell's Dog Leg method (Powell, 1970) are available for optimizing sum-of-squares objective functions[4] like F_I^σ, our objective function for binary images. Even though these methods are specifically designed to take advantage of the special structure of these problems and despite the fact that they offer second order convergence near a local minimum, they have two distinct drawbacks. First, they require computation of a Jacobian matrix, with a row for each of the residuals, which typically number in the tens of thousands at small scales. The rows themselves are sparse, with $O(1)$ non-zero elements, and sparse methods can be used to solve the set of Gaussian normal equations that arise, but even so their solution dominates the computational time. Second, both methods are trust-region methods. When a step fails to reduce the objective function in a trust region method, the size of the trust region is reduced and the step is re-computed. However, the objective function fails to be continuous when a step causes the $\|\nabla r\| < 1$ constraint to be violated at one or more samples, because we cannot evaluate the spoke vector S_0 (and simply drop the corresponding terms from the sum). Since continuity is assumed in all trust region methods, a gradient pointing into one of these regions will cause the method to step up very close to its edge, after which the models it constructs will fail to be valid in a region of any size, terminating the optimization. Initial experiments with a Dog Leg optimizer confirm that it does indeed reach points where the objective function does not actually decrease when moving in the gradient descent direction[5], and this prevents the optimizer from producing a reasonable fit. A line search method, in contrast, can select a step that actually moves in the opposite of the search direction, which can help avoid getting stuck in such places.

The line search method used is a nonlinear conjugate gradient method (Shewchuk, 1994), which performs well on objective functions with a steep, narrow valley, such as can be produced by the penalty functions described in the next section. Conjugate gradients have a vastly reduced memory footprint, as only a couple of gradient vectors need be stored, and also avoid expensive linear algebra computations. The bottleneck of the optimizer then becomes the computation of the gradient vector itself, which due

[4] Both methods share similar qualities, and Levenberg-Marquardt is generally more popular, but the Dog Leg method has considerable computational advantages (Lourakis and Argyros, 2005).

[5] The accuracy of the gradient can be confirmed by approximating it numerically using finite differences smaller than the minimum allowed step size. That is, this result is not due to a bug in the gradient calculation.

to locality is roughly a (large) constant factor more expensive to evaluate than the objective function.

The basic method for an objective function $F(x_k)$ proceeds as follows. Given a search direction δ_k at step k, a line search is used to find a step $x_{k+1} \triangleq x_k + \alpha_k \delta_k$. The search direction is initialized to $\delta_0 \triangleq -\nabla F(x_0)$. The next search direction is chosen to be approximately orthogonal to all of the previous ones using the update step

$$\beta_{k+1}^{PR} \triangleq \max\left(\frac{\nabla F(x_{k+1}) \cdot (\nabla F(x_{k+1}) - \nabla F(x_k))}{\|\nabla F(x_k)\|^2}, 0\right) , \tag{4.37}$$

$$\delta_{k+1} \triangleq \beta_{k+1}^{PR} \delta_k - \nabla F(x_{k+1}) . \tag{4.38}$$

This approximation is exact (within numerical precision) in the case of a quadratic function, in which case (for a well conditioned problem) convergence is achieved in exactly n iterations, where n is the number of variables. The optimization continues until the L_∞ norm of $\nabla F(x_k)$ drops below a threshold ϵ_1. It also stops if the chosen step size $\alpha_k \|\delta_k\|$ falls below the threshold $\epsilon_2(\|x_k\| + \epsilon_2)$. We set both ϵ_1 and ϵ_2 to $\epsilon_0^{3/4}$, where ϵ_0 is the floating-point precision of the machine.

The update strategy used here is due to Polak and Ribiere, which generally provides better convergence than the original strategy proposed by Fletcher and Reeves: $\beta_{k+1}^{FR} \triangleq \|\nabla F(x_{k+1})\|^2 / \|\nabla F(x_k)\|^2$. The max operator in the β_{k+1}^{PR} formula restarts the optimization in the gradient descent direction whenever $\beta_{k+1}^{PR} < 0$, a modification to the original Polak-Ribiere update formula that is necessary to avoid potentially getting trapped in an infinite cycle around the minimum. This restart discards information about the local quadratic model from previous points, which may also be helpful if the current point has moved a sufficient distance to make this model invalid. Therefore it is also common practice to restart the method when two successive gradients are not sufficiently orthogonal, e.g., whenever

$$\frac{|\nabla F(x_k) \cdot \nabla F(x_{k+1})|}{\|\nabla F(x_k)\|^2} > c_1 \tag{4.39}$$

for some constant c_1, here fixed at $c_1 = \frac{1}{10}$. One can also choose to restart after a fixed number of iterations (e.g., n), but typically the orthogonality check will fail long before this limit is reached.

A series of steps up until a restart (for any reason) is considered a single *macro step*. The sampling of a model is kept fixed for the duration of a macro step. Whenever the conjugate gradient optimization is restarted, the models are resampled, which of course

means that $F(x_k)$ and $\nabla F(x_k)$ must be recomputed as well. For $F = F_I^\sigma$, each restart also updates the weights, w_i. The total number of macro steps at each scale level is limited to some value $K_{\max}$. All the fits reported in this chapter use $K_{\max} = 200$.

Returning to the line search, the direction vectors δ_k are badly scaled, so using an initial guess for the step size of $\alpha_{k,0} = 1$ will typically land far from the actual minimum. However, there is a strong coherence in the step sizes between iterations, so a good initial guess is

$$\alpha_{k,0} \triangleq \alpha_{k-1} \frac{\|\delta_{k-1}\|}{\|\delta_k\|} \ . \tag{4.40}$$

From there a Golden Section Search similar to the one described in Chapter 10 of (Press et al., 1992) brackets the minimum. That is, three consecutive points on the line are identified where the objective function at the middle one is smaller than that at the outer two. The step size coherence usually allows this to be done with only three or four additional function evaluations.

The bracketing process tries successively larger steps until it finds one where the objective function is larger than the smaller of $F(x_k)$ and $F(x_k + \alpha_{k,0}\delta_k)$. The algorithm described by Press et al. uses steps in the *negative* direction if the $\alpha_{k,0}$ step actually led to an increase of F. However, one expects that the function should decrease in the search direction. Naively reversing the search direction just because the initial step led to an increase runs the danger of the line search failing. To see how, imagine a function that blows up to a very large value outside of some small interval around the base point, x_k (such as the penalty functions introduced below). By reversing the direction, the chosen bracket will surround the base point and will appear to the minimization process to lead to such a large reduction in F that it is immediately chosen as the minimum. The resulting step $\alpha_k = 0$ halts the optimization process.

Therefore if the initial choice of $\alpha_{k,0}$ leads to an increase, it is cut in half repeatedly until a bracket is formed or $\alpha_{k,j}$ falls below a threshold, chosen here to be $\frac{2^{16}\epsilon_2}{\|\delta_k\|}$. This prevents the bracket from moving too close to the same kind of function discontinuities that plague the trust region methods. Only if this process fails is the search direction reversed.

Given a bracket, the actual line search is done using Brent's Method (Brent, 1973), a derivative-free 1D optimization method. Despite the fact that this search may require many function evaluations, it is extremely fast because it does not need to calculate the gradient vector at each step. Doing the line search without derivatives eliminates

the major drawback of line search methods: the computational expense of the extra function evaluations. Typically, the entire bracketing process and 1D minimization are completed in less time than it takes to compute a single gradient, i.e., in fewer than 64 steps, typically much fewer. Since the conjugate gradient method converges better the closer the minimizer is to the true minimum and additional iterations are so cheap, the line search continues all the way to the tolerance used for the step size, ϵ_2.

4.2.4 Constrained Optimization

The preceding optimization techniques are unconstrained; control points can take on any values. However, in reality a model must obey a number of restrictions at each sample:

- The normal to the medial sheet, $m_u \times m_v$, must be non-zero.
- $\|\nabla r\|$ must be less than 1 off the edges.
- For each value of v on an edge patch, $r_v(0, v)$ must not be so large as to cause $\|\nabla r\|$ to be larger than 1 all by itself.
- The radial curvature condition must be satisfied off the edges, for both the top and bottom sides.

A set of control point values that satisfies all of these restrictions is said to lie in the *feasible region*.

Three simple tests can summarize these restrictions. Because one of the conditions they enforce is that the normal is non-zero, they avoid dividing by its magnitude. This makes each test non-commensurate with the others, so division by the sampling threshold τ is used to restore them to approximately the same scale. Converting the equality in (3.16) to an inequality yields the first test:

$$\frac{1}{\tau^4}\left(\|m_u \times m_v\|^2 - \|r_u m_v - r_v m_u\|^2\right) > 0 \tag{4.41}$$

This test actually covers the first two restrictions. The next test simply ensures that the square root term from (3.18) is positive:

$$\frac{1}{\tau^6}\left(\left(\|m_v(0,v)\|^2 - r_v^2(0,v)\right) \cdot \|m_u(0,v) \times m_v(0,v)\|^2\right) > 0 \tag{4.42}$$

The final test is given by the radial curvature condition and is necessary only on edge patches:

$$1 - r\kappa_{r_i}^{\pm} > 0 \tag{4.43}$$

Only the largest eigenvalue on each side needs to be checked, and only if the eigenvalue is real. The latter should always be the case on a Blum medial axis, but a complex value might arise due to numerical imprecision. No constraint need be imposed in this case.

These tests transform an unconstrained method into a constrained method via the standard technique of adding a *penalty function* to the objective function. This increases the value of the objective function when any of these conditions are violated, forcing the optimum to lie in the feasible region. Yet our problem is somewhat unusual in several regards. There are many more constraints than variables, in fact $O(n_{\mathbb{S}})$ of them. Generally only a handful are active at any given time, however. All of the constraints are inequality constraints, while most penalty function methods are first designed to enforce equality constraints and then adapted to inequality constraints through the use of slack variables. This requires adding an extra variable to the problem per constraint, which is clearly impractical. Not only are there tens of thousands of such constraints at small scales, but the number changes every time model is resampled. A similar problem occurs with Lagrange multiplier approaches. Also, each constraint is a strict inequality. No constraints should be active at the final solution, so active set methods are not appropriate. This could be handled by adding a "safety margin" to the right hand side of each of the constraints above, but it raises the question of how large a margin is appropriate. In particular, the left-hand side of (4.41) becomes exactly zero on the edge points, so it can take on arbitrarily small values in the neighborhood of these points.

The following now describes the method actually used, which works reasonably well in practice, yet which could almost certainly be improved by further research. The general approach is to introduce penalty terms of the form

$$\delta_i \triangleq \upsilon(\epsilon\gamma_i - p_i) \ . \tag{4.44}$$

Here p_i is the left-hand side of one of the expressions in (4.41), (4.42), or (4.43). The parameter υ controls the strength of the penalty, and ϵ controls the size of the safety margin. As the optimization progresses, υ is increased, and ϵ is adjusted adaptively.

The weights γ_i act roughly like Lagrange multipliers, but are held constant; they do not require careful tuning. These are fixed at $\gamma_i = 2^{-12}$ for (4.41), $\gamma_i = 2^{-16}$ for (4.42), and $\gamma_i = 2^{-4}$ for (4.43). Collecting all of the terms where $\delta_i > 0$ yields the penalty

$$p \triangleq \sum_{i:\delta_i>0} \delta_i^2 \ . \tag{4.45}$$

This function is C^1 continuous in p_i at the point each constraint is activated, since $\frac{\partial \delta_i}{\partial p_i} = -2\upsilon\delta_i$, and at that point $\delta_i = 0$. A constraint being active ($\delta_i > 0$) is distinct from it being violated ($p_i \leq 0$). A point lies in the infeasible region only if one or more constraints are actually violated.

Making ϵ one of the optimization variables allows the optimization process to adjust it adaptively. However, to avoid introducing yet another constraint to ensure it remains positive, the actual variable used is instead $\log \epsilon$. The derivative for the active δ_i terms

$$p_{\log \epsilon} = \epsilon\upsilon \sum_i \gamma_i \tag{4.46}$$

is always positive whenever there is at least one active constraint, so ϵ will be steadily decreased throughout the optimization. The remaining derivatives with respect to control point parameters ν,

$$p_\nu = -\upsilon p \sum_i \frac{\partial p_i}{\partial \nu} \ , \tag{4.47}$$

may require derivatives of S_{rad}, etc., but they need only be computed for the active constraints. There are typically only a handful of these.

A typical approach to adjusting the strength parameter υ would be to solve the repeated minimization problem

$$\operatorname*{argmin}_{x,\epsilon} F(x) + p(x, \epsilon, \upsilon) \tag{4.48}$$

as $\upsilon \to \infty$. However, solving even one of these minimizations is already very expensive, so to make the optimization practical υ is held fixed just for the duration of each macro step and adjusted upward under certain conditions.

This is done by introducing a factor Υ and setting $\upsilon_{K+1} = \max(\upsilon_K, \frac{\Upsilon}{\epsilon})$ after macro step K. Thus as ϵ decreases, υ increases, keeping the total penalty at the $p_i = 0$ boundary of the feasible region roughly constant. At the start of the coarsest scale, Υ

is chosen according to the scale of the target image, and $\upsilon_0 = \Upsilon$. A good choice for Υ is the square root of the total surface area of the target, or the conservatively low approximation $(6\sqrt{\pi}V(\Omega_t))^{\frac{1}{3}}$, the latter being the choice used here. The parameters ϵ, υ, and Υ are not reset after each scale, but carried over from one to the next, except that Υ is increased by a factor of $\sqrt{2}$ whenever the optimization at a given scale σ finishes with a point in the infeasible region. This doubles the strength of the penalty function because of the squaring therein. Making Υ extremely large makes the problem extremely stiff and could prevent the optimizer from making any progress at all, and so it is only increased a limited number of times, and only when the penalty function is failing to keep the solution in the feasible region.

It is possible that the optimization even at the final scale could finish in the infeasible region. This occurs because we both limit the number of iterations via $K_{\max}$ and do not force $\Upsilon \to \infty$. To make sure it produces a feasible model, every time the model is resampled, if it lies in the feasible region it is copied. After the optimization at the finest scale finishes, the optimization returns the last feasible model as the final result.

The entire process is summarized as follows. The medial model is aligned to the target shape using the procedure outlined in Section 4.2.1. Then the model is fit to this shape over a number of steadily decreasing scales, σ. The sampling of the model is tied directly to the scale, with $\tau = \frac{1}{2}\sigma$. Thus at large scales where the biggest steps are taken, the iteration proceeds very quickly. The factor of $\frac{1}{2}$ ensures the model is still legal well below the sampling resolution of the target image. In the results presented below, the largest scale was computed from geometric average lengths of the edges of the template model and of its spokes after alignment. The smallest scale was taken as the geometric average of the voxel spacing in each direction.

The objective function of the local optimization is $F_I^\sigma + p$, constructed from (4.26) and (4.45). The fitting process itself is the non-linear conjugate gradient method presented in Section 4.2.3 using Brent's method for a derivative-free line search. The variables in this optimization are the components of the control points in the subdivision mesh along with ϵ, which controls the size of the safety margin around the infeasible region in the penalty function. The sample weights w_i from (4.27) and the penalty strength υ are held fixed during this optimization until the conjugate gradient optimization restarts, which we term a macro step. The total number of macro steps at a given scale is limited to a constant $K_{\max}$. After the optimization at that scale is complete, the penalty parameter Υ is increased if the model is infeasible, the scale σ and sampling resolution τ are decreased, and a new local optimization at the finer scale

starts from the result of the previous scale. At the end of the finest scale, the most recent feasible model is returned as the final result of the optimization.

4.2.5 Results

A simple test population was created from an ellipsoid deformed by a set of 20 diffeomorphisms of the form (Han, 2006)

$$\Psi_{\alpha,\beta,\gamma}(x,y,z) \triangleq (x, e^{\gamma x}(y\cos(\beta x) - z\sin(\beta x)), e^{\gamma x}(y\sin(\beta x) + z\cos(\beta x)) + \alpha x^2) \ , \tag{4.49}$$

where α, β, and γ are parameters controlling bending, twisting, and tapering, respectively. These parameters were drawn from normal distributions with standard deviations 1.5, 1.05, and 2.12, respectively, and the resulting deformation was applied to a standard ellipsoid with axis lengths of $(1/2, 1/3, 1/4)$ centered around the origin. Compared to the size of these ellipsoids, these are relatively large deformations, as shown in the figures below. The result was converted to a $128 \times 128 \times 128$ binary image for fitting.

The fitting process used 10 scales and the maximum number of macro steps was $K_{\max} = 200$. The choice of these parameters was conservatively large; during the fitting process itself most of the finer scales changed the model very little beyond eliminating illegalities, which was usually done in the first few iterations. They could probably be reduced significantly without affecting the quality of the fits. The template model was constructed by fitting a regular 4×6 slab with the corners removed (20 control points, 80 parameters) to each model and aligning and averaging the results, using the arithmetic average for the m components and the geometric average for the r components. Then the models were fit again starting from the template model.

Examples of the best, worst, and an average (median) fit obtained are shown in Figures 4.2, 4.3, and 4.4, respectively. The average volume overlap over all 20 subjects was 93.64% ($V_{\text{Dice}} = 96.72\%$). Table 4.1 summarizes the errors between the boundary surfaces for each model and the original binary images, computed using MeshValmet (Xu, 2006). Figure 4.5 shows how these errors are distributed over the three representative models. Except for the worst case, which was an outlier, voxelization effects account for the majority of the error. This outlier merely represents a local minimum. Re-fitting this image starting from the model fit to one of the other bent ellipsoids produced a result with $V_{\text{overlap}} = 93.62\%$ ($V_{\text{Dice}} = 96.71\%$) and dropped the maximum error

Deformed ellipsoids	Average over the surface	Maximum over the surface
Average over the cases	0.318	4.189
Maximum over the cases	0.977	20.300

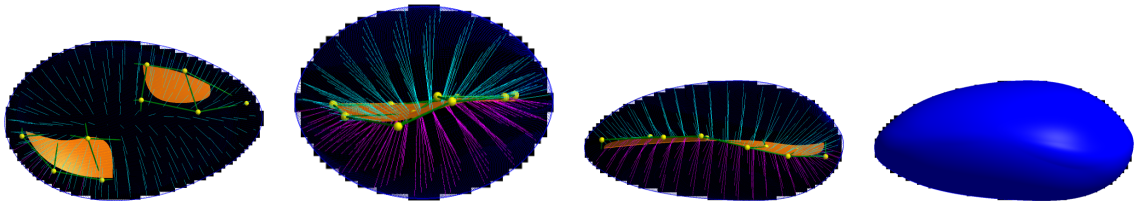
Table 4.1: The absolute error between the reconstructed boundary of the medial model and the original voxelized boundary of the binary image. All measurements given in voxels.

for this case from 20.300 to 2.807 voxels. Using Principal Geodesic Analysis (PGA) (Fletcher et al., 2004) to extract the major modes of variation in the population and performing an initial deformation along just these modes after alignment and before the main optimization could likely eliminate such an outlier. This technique has already been employed with some success in fitting discrete m-reps.

Models were also fit to real-world data from an ongoing clinical longitudinal pediatric autism study (Styner et al., 2006). The study includes two groups: 23 autistic (AUT) and 14 typically developing controls (TYP), with a baseline scan at age two and a follow-up at age four. Four of the control baselines were paired with follow-ups from an unrelated subject, and an additional 4 of the controls had no associated follow-up. This gives a total of 70 images from 46 autistic and 24 typically developing subjects.

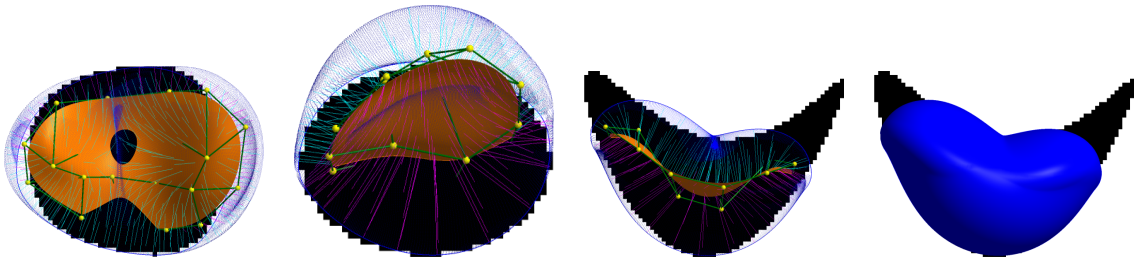
This analysis focuses on the left and right caudate, which were segmented by experts via semi-automatic means into binary images at $0.75 \times 0.75 \times 0.75$ mm^3 resolution. Spherical harmonics (SPHARM) (Brechbühler et al., 1995) were also constructed, using first order harmonics to establish correspondence (Gerig et al., 2001). A template was built by fitting a hand-deformed 5×10 mesh with the corners removed (46 control points, 184 parameters) to a half-dozen subjects, aligning and averaging the results, and then fitting that to the entire population. This matches the number of free parameters in the 3×7 discrete m-reps used in (Styner et al., 2006) (16 end atoms $\times$ $9 + 5$ interior atoms $\times$ $8 = 184$), which the automatic method in (Styner et al., 2003) determined to be the minimum sampling. The same sampling threshold and iteration limits were used as for the deformed ellipsoids.

Examples of the best, worst, and an average (median) fit obtained for each side are shown in Figures 4.6, 4.7, and 4.8, respectively. Because the shapes are curved and not well-aligned to the axes of the image, they are plotted against their SPHARM surfaces for better visualization. The average volume overlap over all 70 subjects was 84.89% ($V_{\text{Dice}} = 91.82\%$) on the left, and 85.25% ($V_{\text{Dice}} = 92.04\%$) on the right. A good deal of



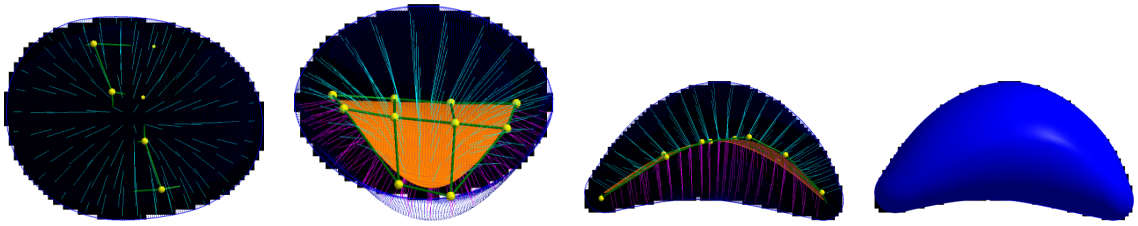
(a) Axial (b) Sagittal (c) Coronal (d) Boundary

Figure 4.2: The best fit obtained for the deformed ellipsoids, $V_{\text{overlap}} = 95.97\%$ ($V_{\text{Dice}} = 97.94\%$).



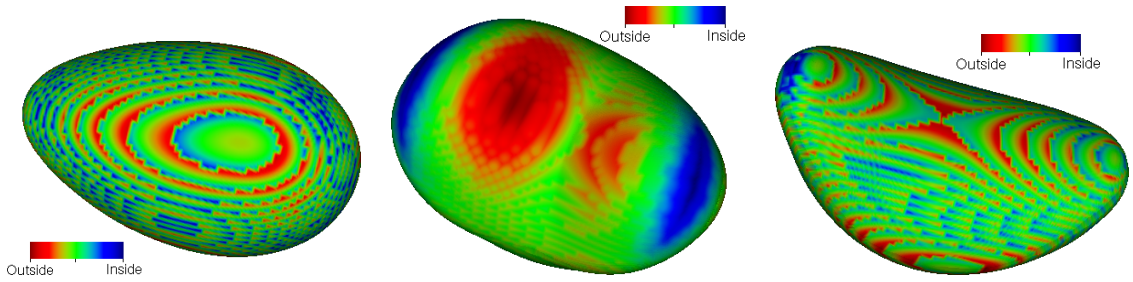
(a) Axial (b) Sagittal (c) Coronal (d) Boundary

Figure 4.3: The worst fit obtained for the deformed ellipsoids, $V_{\text{overlap}} = 79.56\%$ ($V_{\text{Dice}} = 88.62\%$). Here the bending was simply too severe, and the optimization became trapped in a local minimum.



(a) Axial (b) Sagittal (c) Coronal (d) Boundary

Figure 4.4: An example of an average (median) fit obtained for one of the deformed ellipsoids, $V_{\text{overlap}} = 95.74\%$ ($V_{\text{Dice}} = 97.82\%$).



(a) Best (b) Worst (c) Median

Figure 4.5: The signed error between the reconstructed boundary of the medial model and the original voxelized boundary of the binary image. For the best and average cases, small-scale voxel effects dominate the error.

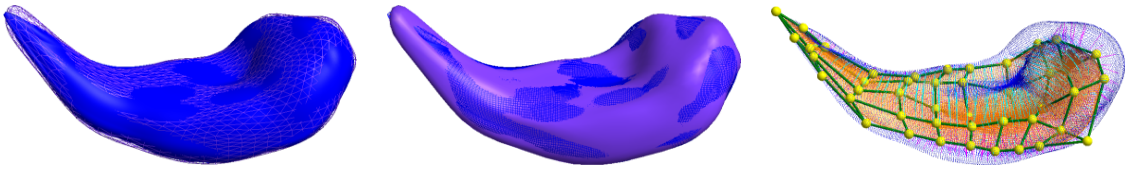
Left caudates	Average over the surface	Maximum over the surface
Average over the cases	0.241	2.812
Maximum over the cases	0.307	4.689
Right caudates	**Average over the surface**	**Maximum over the surface**
Average over the cases	0.243	2.754
Maximum over the cases	0.307	6.502

Table 4.2: The absolute error between the reconstructed boundary of the medial model and the original voxelized boundary of the binary image. All measurements are given in mm.

this error is due to voxelization effects; the errors improve several percentage points on each side—87.97% ($V_{\text{Dice}} = 93.59\%$) and 89.13% ($V_{\text{Dice}} = 94.25\%$), respectively—when volume overlap is computed against the smooth SPHARM representation instead of the voxel images, even though the models were fit to the images. Table 4.2 summarizes the errors between the boundary surfaces for each model and the original binary images, computed using MeshValmet (Xu, 2006). Figure 4.9 shows how these errors are distributed over the median model from each side.

4.3 Population-Based Model Fitting

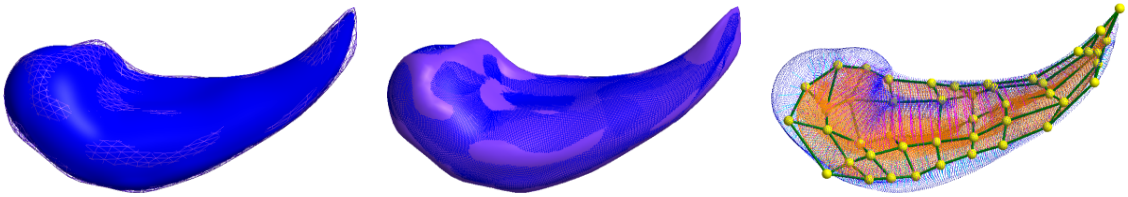
The previous discussion was restricted to fitting a single model against a single target shape. But typically statistics involves an entire population of target shapes. Up to the tolerance afforded by the number of control points, there are often many configurations



(a) The SPHARM boundary over the model

(b) The model over the SPHARM boundary

(c) The medial mesh

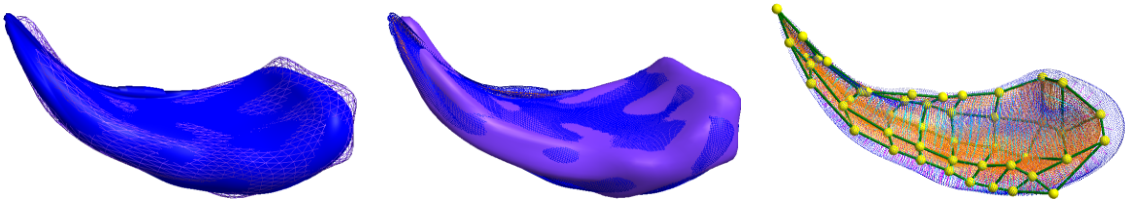


(d) The SPHARM boundary over the model

(e) The model over the SPHARM boundary

(f) The medial mesh

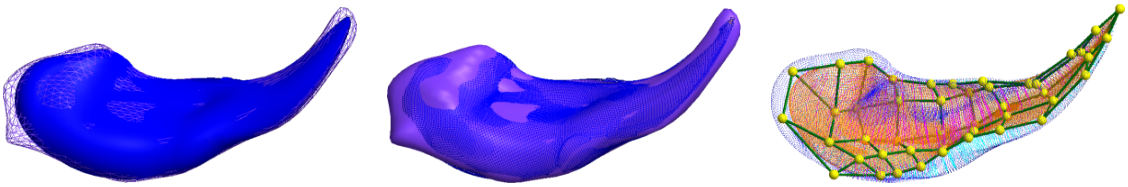
Figure 4.6: The best fit obtained for the caudates on each side. (a)-(c) On the left, $V_{\text{overlap}} = 87.72\%$ ($V_{\text{Dice}} = 93.46\%$). (d)-(f) On the right, $V_{\text{overlap}} = 87.80\%$ ($V_{\text{Dice}} = 93.50\%$).



(a) The SPHARM boundary over the model

(b) The model over the SPHARM boundary

(c) The medial mesh

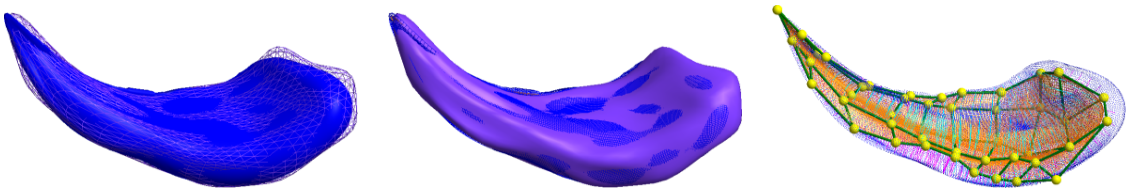


(d) The SPHARM boundary over the model

(e) The model over the SPHARM boundary

(f) The medial mesh

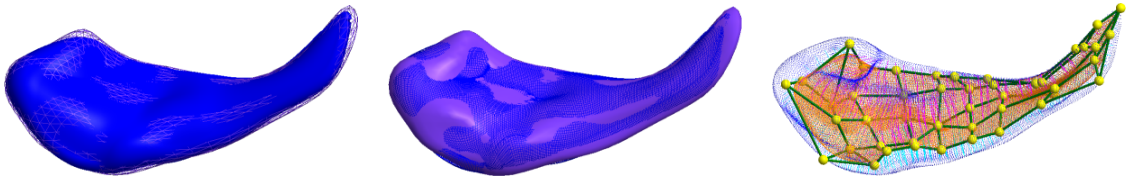
Figure 4.7: The worst fit obtained for the caudates on each side. (a)-(c) On the left, $V_{\text{overlap}} = 82.18\%$ ($V_{\text{Dice}} = 90.22\%$). (d)-(f) On the right, $V_{\text{overlap}} = 82.71\%$ ($V_{\text{Dice}} = 90.54\%$).



(a) The SPHARM boundary over the model

(b) The model over the SPHARM boundary

(c) The medial mesh

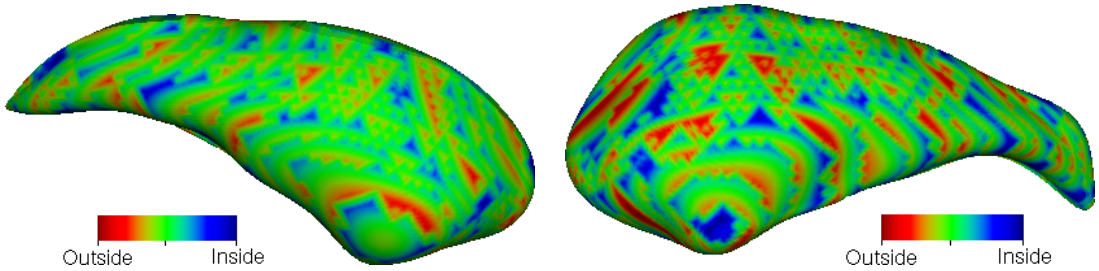


(d) The SPHARM boundary over the model

(e) The model over the SPHARM boundary

(f) The medial mesh

Figure 4.8: An average (median) fit obtained for the caudates on each side. (a)-(c) On the left, $V_{\text{overlap}} = 84.66\%$ ($V_{\text{Dice}} = 91.69\%$). (d)-(f) On the right, $V_{\text{overlap}} = 85.25\%$ ($V_{\text{Dice}} = 92.04\%$).



(a) Left

(b) Right

Figure 4.9: The signed error between the reconstructed boundary of the medial model and the original voxelized boundary of the binary image for the two average cases above.

of those control points that will fit each target equally well. For example, consider a flat slab, with a constant radius function away from the edges. In the interior of the slab, control points can be moved arbitrarily without changing the reconstructed surface.

By performing the optimization individually for every subject, control points are allowed to wander far from their initial configuration, and the resulting pieces of the boundary each represent may be very different from subject to subject. The approach taken when optimizing discrete m-reps has been to add a number of regularization penalties to keep the edges of the mesh roughly of uniform size and to prevent deformation too far from the template itself. This has a good chance of producing the desired effect when the objects themselves are not too different than the template, but seems an unreasonable assumption when inter-subject variation is large.

With a continuous medial model, there is another alternative, which is to reparameterize the medial sheet to produce a better correspondence. This chapter does *not* tackle the problem of defining what the "correct" correspondence is between two shapes. Determining correspondence, while a prerequisite for any kind of statistics, is an extremely difficult problem. For example, in the lateral ventricles (see Chapter 5) the tail sometimes becomes pinched so thin that the end of it is cut off during the segmentation process, meaning that it is not even possible to construct a valid anatomical correspondence with a ventricle whose tail is complete. Careful researchers will thus devote every effort to improving the binary segmentations, but there are no statistical methods that can cope with these errors. Even with other structures that are not subject to this problem, there is an inherent difficulty in separating variability in the observable image features from the variability in shape (the correspondence map).

If one wants to perform statistics on image intensities, for example, and also register those images using the intensities, one has to be careful not to destroy the very statistics being measured. If the correspondence map is allowed to be arbitrary, a perfect registration can be obtained, and there will be no local information left to analyze. The typical response is to impose some smoothness prior on the correspondence map, even if it is not expected that the map should actually be smooth everywhere, or as in DetCov (Kotcheff and Taylor, 1998) or MDL[6] (Davies et al., 2002; Thodberg, 2003; Heimann et al., 2005), to optimize the maps to have a compact statistical model. However, one of the difficulties of MDL techniques in particular is avoiding "data-piling", wherein the optimization process causes landmark points to migrate away from the highly variable

[6] The University of Manchester holds a patent on optimizing the Minimum Description Length (MDL) to establish correspondence.

parts of the shape because it can get more compact models by not sampling them at all. This could damage the statistics of shape change in precisely the places that are most interesting.

Davies et al. try to avoid this by choosing a fixed subject as the "master" that all others are matched against (as do Pizer et. al with their discrete m-reps), but in addition to potentially biasing the results if a poor choice of master is made, (Thodberg, 2003) point out this still fails to prevent the migration when the weight of the majority overwhelms the cost of a single outlier. Thodberg adds an extra term to the optimization to penalize this drift, but this requires a choosing an appropriate weight to balance it against the rest of the cost function (Thodberg, 2003). Heimann et al. optimize the parameterization of each model, instead of the location of the landmarks within that parameterization, and report that no data-piling occurs (Heimann et al., 2005). However, this just transfers the problem to enforcing the smoothness of the parameterization. Recent work by (Cates et al., 2006) seeks to balance a uniform sampling of landmarks within each shape against the variability of the landmark locations across shapes using an information theoretic metric, eliminating the need for a scale factor to relate the two. Yet however mathematically principled, it is unclear how such schemes relate back to the underlying biology.

As explained in Section 1.1.2, any kind of true anatomical correspondence is inherently defined by the problem domain. Population-based methods are a valuable tool for learning these correspondences from a large set of examples, but they are ultimately limited by the class of models they can construct. Different problems may require different types of models to capture the underlying variability of the processes that govern the population. Instead of deriving a new correspondence algorithm and attempting to justify the difficult claim that it is the "correct" correspondence for some class of problems, the method presented here takes as input a fixed correspondence on the boundary (or possibly the entire object interior) and optimizes the implicit medial correspondence to match. This takes advantage of the link between the boundary and the continuous medial axis to leverage any conceivable method of determining correspondence, including population-based methods.

The model is reparameterized by moving control points *along the surface of the medial sheet* to give a better match between the implicit medial correspondence and the explicitly given boundary correspondence. The idea here is that for small steps and within the tolerance afforded by the number of control points in the model, this does not change the model itself, but merely the coordinate system imposed on it. Thus,

this operation is (approximately) transverse to the model fitting process and needs no relative weight to balance against it. Instead the two proceed in lock-step. The optimization takes alternating macro steps first using one of the shape-fit objective functions from the previous section, followed by the correspondence-match objective function described below.

4.3.1 The Objective Function for Correspondence

First consider the situation for two models, $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Correspondence is given by a map $\phi_1^2 : \mathbb{R}^3 \to \mathbb{R}^3$ from the first to the second, and a map $\phi_2^1 : \mathbb{R}^3 \to \mathbb{R}^3$ from the second to the first. In the case where these maps ϕ_j^i are defined only on the boundary of the target shape, $\mathcal{B}_t^{(j)}$, they can be extended to the boundary of the model using projection. If the correspondence is not even available over the whole boundary, but only at a discrete set of points, triangulating the points and linearly interpolating between them can extend the correspondence to the entire boundary. Given a map $\phi_j^{i\prime} : \mathcal{B}_t^{(j)} \to \mathbb{R}^3$ on the boundary $\mathcal{B}_t^{(j)}$, one can construct $\phi_j^i = \phi_j^{i\prime} \circ \pi_j$, where π_j is projection onto the boundary along the medial spokes of $\mathcal{M}^{(j)}$. For a given point $(m_0^{(j)} + S_0^{(j)}) \in \mathcal{B}_m^{(j)}$, this is the intersection of $m_0^{(j)} + tS_0^{(j)}$ closest to $t = 1$, if one exists. If there is no such intersection, the corresponding term is simply dropped from the summations below. It may be worth adding a small constant instead to discourage spokes from being pushed outside the target, but there is little reason to heavily penalize such points. The reparameterization step cannot, by itself, move a piece of the medial axis into the interior of the target, since resampling will simply produce more spokes on the exterior.

Then optimizing the following squared error ensures that the implicit medial correspondence "matches" the explicit correspondence given by these maps in some sense:

$$F_{2C} \triangleq \left(m_0^{(1)} + S_0^{(1)} - \phi_2^1(m_0^{(2)} + S_0^{(2)})\right)^2 + \left(m_0^{(2)} + S_0^{(2)} - \phi_1^2(m_0^{(1)} + S_0^{(1)})\right)^2 , \quad (4.50)$$

for a pair of corresponding samples in $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. This implies that both models must have the same sampling, a fact ensured by continuing to subdivide on each model until the subdivision threshold is satisfied on *all* models in the population. One could certainly imagine other objective functions that would also drive the implicit and explicit correspondences to agreement, such as for example using the geodesic distance along the model boundary $\mathcal{B}_m^{(j)}$ instead of the simple Euclidean distance in the ambient space, but this one has the advantage of being relatively simple to compute and

optimize.

Extending this directly to n models would require $O(n^2)$ maps, and $O(n^2)$ terms in the direct product, which is prohibitive for large populations. Therefore we borrow an idea from recent work on computing unbiased, symmetric averages of shapes (Davis et al., 2004; Škrinjar and Tagare, 2004; Joshi et al., 2004): map all of the points to a common coordinate system and measure the errors there. The number of maps and the number of error terms then become linear in the number of subjects. Without any other prior information, an "average shape" serves as a good choice for the reference coordinate system. All the shapes are Procrustes aligned, and the average is constructed by directly averaging corresponding boundary points on each subject. Then each map ϕ_j simply sends a point to the average of all explicitly corresponding points.

Then, if all the implicitly corresponding points on the models match the explicit correspondence on the targets, they will be mapped to the same point. Therefore one should minimize the intra-cluster distances $\sum_{i \neq j} (\phi_i(m_0^{(i)} + S_0^{(i)}) - \phi_j(m_0^{(j)} + S_0^{(j)}))^2$. However, when holding all the models but $\mathcal{M}^{(j)}$ fixed, the minimizer becomes exactly the cluster mean

$$\mu_{(m_0,S_0)} \triangleq \frac{1}{n} \sum_{i=1}^{n} \phi_i(m_0^{(i)} + S_0^{(i)}) \ . \tag{4.51}$$

Letting $\mu_{(m_0,S_0)}$ be the mean for each sample on the models, the optimization proceeds by taking a macro step for each model in turn, minimizing the objective function

$$F_C^j \triangleq \int_{\tilde{M}} \left(\phi_j(m_0 + S_0) - \mu_{(m_0,S_0)}\right)^2 \cdot \det\left(\mathbf{I} - r S_{\text{rad}}\right) \, dM \ , \tag{4.52}$$

$$\approx \sum_{(m_0,S_0,r,S_{\text{rad}},\Delta M) \in \mathbb{S}^{(j)}} \left(\phi_j(m_0 + S_0) - \mu_{(m_0,S_0)}\right)^2 \cdot \det\left(\mathbf{I} - r S_{\text{rad}}\right) \Delta M \ . \tag{4.53}$$

To preserve the symmetry of the operation, the means $\mu_{(m_0,S_0)}$ are held fixed between macro steps for different models.

Often correspondence is only given on the boundary, but it is possible to extend this approach to the entire volume of an object simply by extending the above objective function from a boundary integral to one over the interior. For neuroanatomy in particular, recent work on symmetric atlas formation can provide a means for obtaining voxel-wise correspondences over the entire brain (Davis et al., 2004; Joshi et al., 2004; Lorenzen et al., 2006). Because these methods map all subjects into a single atlas, they provide the ϕ_j maps directly, without the need to construct them from the all-pairs

maps. They also eliminate the need to project the model boundary onto the target shape.

4.3.2 Parameterizing the Correspondence Optimization

The above mentioned that in order to make the correspondence-match optimization approximately orthogonal to the shape-fit optimization, it moves the control points along the existing medial sheet. In reality, since the medial sheet does not pass through the control points, this is done by moving them so that their corresponding limit points slide along the surface of a *base model*, $\mathcal{M}_0^{(j)}$, the current model produced by the shape-fit optimization. This does not optimize the control point values directly but uses another set of variables that parameterize this sliding process.

This requires constructing a mapping between the control points and their corresponding limit points. This mapping is linear and can be expressed as a sparse mask matrix $\mathbf{K}$ computed from the subdivision process. This mask is fixed for a given model topology and can be computed once in advance, along with its inverse. Although the inverse $\mathbf{K}^{-1}$ is *not* sparse, the LU decomposition $\mathbf{K} = \mathbf{LU}$ is (Halstead et al., 1993), with typically fewer than eight non-zero elements per control point total in the two matrix factors. Analysis of a general subdivision mesh is difficult, but experiments with regular grids confirm Halstead et al.'s empirical observations that the non-zero coefficients grow linearly with the number of control points and the time to compute the decomposition grows quadratically. Also, because the largest element of each row of $\mathbf{K}$ is on the diagonal, one does not even need to use pivoting when computing the decomposition.

Given the LU decomposition, one can solve for a set of control points P that interpolate corresponding limit points Q via $\mathbf{K}P = Q$ using forward and back substitution in linear time. Halstead et al. warn that $\mathbf{K}$ might actually be singular but do not say under what conditions this can occur. A least-squares solution (like the Penrose pseudoinverse) could be used in this case, but all of the mask matrices we have produced have been very well-conditioned.

The next step is the parameterization of the set of limit points, Q. For control points on the edge of the mesh, this is simple, as one can use the $\mathcal{M}_0^{(j)}(0, v)$ curve on patches in the base model around the control point, with v as the parameter. For control points in the interior, there are two degrees of freedom, but around extraordinary vertices it is not obvious how one should map these two parameters into the domain of the

surrounding patches. Given a vertex of valence n, one reasonable choice is the simple bilinear map $\Phi = (\phi^x, \phi^y)$ given by (Loop, 2004), which maps the (u, v) coordinates of a patch to a portion of the unit regular n-gon centered around the origin:

$$\phi^x(u,v) \triangleq \begin{bmatrix} 1 & u \end{bmatrix} \begin{bmatrix} 0 & C_n \\ 1 & -C_n \end{bmatrix} \begin{bmatrix} 1 \\ v \end{bmatrix} , \tag{4.54}$$

$$\phi^y(u,v) \triangleq \begin{bmatrix} 1 & u \end{bmatrix} \begin{bmatrix} 0 & S_n \\ 0 & T_n - S_n \end{bmatrix} \begin{bmatrix} 1 \\ v \end{bmatrix} , \tag{4.55}$$

where $C_n = \cos(\frac{2\pi}{n})$, $S_n = \sin(\frac{2\pi}{n})$, and $T_n = \tan(\frac{\pi}{n})$. The entire unit n-gon is parameterized by using n rotated copies of this map, $\mathbf{R}_n^k \circ \Phi$, where $\mathbf{R}_n$ is a rotation by $\frac{2\pi}{n}$ in the x-y plane. This map is G^2 continuous across patch boundaries, which is more than sufficient.

To go from the unit n-gon back to the patches requires the inverse map, $\Phi^{-1} = (\phi^u, \phi^v)$, given by

$$\phi^u(x,y) = -\frac{C_n(y - T_n x) - S_n + \sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}}{2C_n T_n} , \tag{4.56}$$

$$\phi^v(x,y) = \frac{C_n(y - T_n x) + S_n - \sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}}{2C_n T_n} , \tag{4.57}$$

and with associated derivatives

$$\phi^u_x = \frac{1}{2T_n}\left(\frac{(C_n(y - T_n x) + S_n)T_n}{\sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}} + T_n\right) , \tag{4.58}$$

$$\phi^u_y = \frac{1}{2T_n}\left(\frac{2T_n - (C_n(y - T_n x) + S_n)}{\sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}} - 1\right) , \tag{4.59}$$

$$\phi^v_x = \frac{1}{2T_n}\left(\frac{(C_n(y - T_n x) + S_n)T_n}{\sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}} - T_n\right) , \tag{4.60}$$

$$\phi^v_y = \frac{1}{2T_n}\left(\frac{2T_n - (C_n(y - T_n x) + S_n)}{\sqrt{(C_n(y - T_n x) + S_n)^2 - 4C_n T_n y}} + 1\right) . \tag{4.61}$$

The complete mapping to patch k is then given by $\Phi^{-1} \circ \mathbf{R}_n^{-k}$, and the patch k is identified from $k = \lfloor \frac{n}{2\pi} \arctan(y, x) \rfloor$. Then if $\mathcal{M}_0^{(j,k)}(u,v)$ are the patches surrounding the control point, the limit point to be interpolated is given by $(\mathcal{M}_0^{(j,k)} \circ \Phi^{(-1)} \circ \mathbf{R}_n^{-k})(x,y)$.

The inverse in (4.56) and (4.57) is well-defined on the interior of the n-gon except

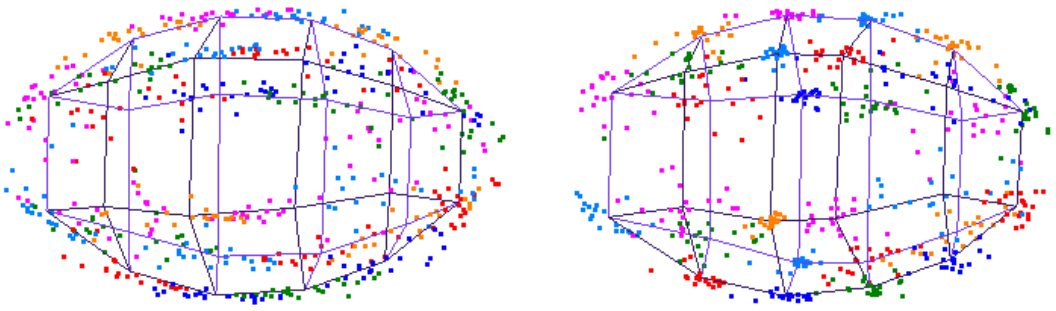
when $n = 4$, since $C_4 = 0$. However, in this case $\Phi(u, v) = (u, v)$, the identity map, and the inverse is trivial. The optimization is constrained to lie in the n-gon interior simply by setting the objective function to ∞ outside of it. This means that the farthest a single control point can move in one macro step is to the edge of an adjacent patch, or approximately one quarter of the distance to a neighboring control point. Edge points are constrained to lie on an adjacent patch in the same way. From these mappings, it is a couple of simple applications of the chain rule to obtain a gradient with respect to the parameters controlling each limit point from the gradient with respect to the control point parameters.

4.3.3 Results

For the ellipse data set, the correspondence is known; it is given by Ψ from (4.49). Therefore ϕ_j is set to $\Psi^{-1}_{\alpha_j,\beta_j,\gamma_j}$. Although this map does describe the deformation applied to obtain the target object, it may not be physically realistic, since one would not, for example, expect two points on the top and bottom of an ellipsoid with the same x coordinate to still have the same x coordinate after bending it. Starting from the same template as in Section 4.2.5 and using the same number of scale levels and the same iteration limit $K_{\max}$, the models were re-fit. However, each iteration comprised one macro step optimizing the binary image match, F^{σ}_{I}, followed by one macro step optimizing the correspondence match for each subject, F^{j}_{C}. The optimization only stopped when one of the stopping criteria was encountered for every model, which in this case meant it proceeded to the iteration limit at every scale. The average volume overlap was extremely similar to the individual fit case, and in fact slightly higher: 93.95% ($V_{\text{Dice}} = 96.88$) vs. the 93.64% ($V_{\text{Dice}} = 96.72\%$) reported in Section 4.2.5. The fact that the correspondence optimization continues to evolve the model even after the image match optimization has converged for a particular subject, giving it the chance to escape a local minimum, likely accounts for this small improvement. However, it did not help the one outlier case.

In order to visually evaluate how well the surface correspondence was maintained by the medial model, Figure 4.10 shows the endpoints of the spokes associated with each control mapped into the reference coordinate system via ϕ_j for the models fit both with and without correspondence optimization enabled. When correspondence optimization is enabled, these points form much tighter clusters, especially towards the center of the ellipsoid, where the radius changes more slowly. This is precisely the place where the

parameterization is most ambiguous, allowing the correspondence match to produce tight clusters without sacrificing fit quality. The clusters on the ends are not as tight, but for the most part they are still well-separated, unlike their counterparts obtained by fitting models individually.



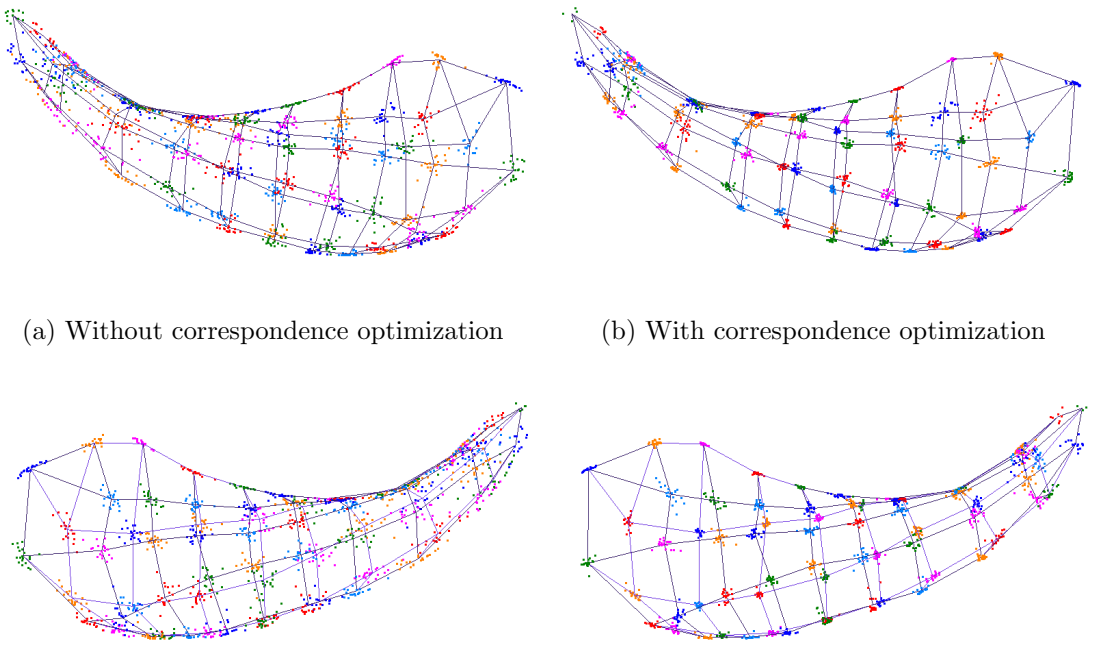
(a) Without correspondence optimization (b) With correspondence optimization

Figure 4.10: Clusters of corresponding points in the reference coordinate system, both with and without correspondence optimization enabled. Points with the same color belong to the same cluster. The wireframe model connects the cluster centers. Enabling correspondence optimization gives much tighter clusters, especially near the center of the ellipsoid, where the parameterization is most ambiguous.

For the caudates, ten subjects each from the AUT and TYP groups were chosen and their scans from age four used to test the correspondence optimization. The SPHARM models gave correspondence, with ϕ_j chosen to map to their average after Procrustes alignment. Ray-triangle intersections computed with the algorithm in (Möller and Trumbore, 1997) were used to project $m_0 + S_0$ onto the SPHARM surface, and OBB trees (Gottschalk et al., 1996) were used to reduced the number of triangles that needed to be tested. To speed up these tests even further, the most recent triangle intersecting each ray was cached, and the OBB tree was searched only if the intersection test against the cached triangle failed. The results in Figure 4.11 are even more striking than for the ellipse data, especially in the tail region.

4.4 Conclusion

This chapter has described a fitting process for the medial model developed in Chapter 3. As presented, it is not a complete, robust system for segmentation, but it



(a) Without correspondence optimization

(b) With correspondence optimization

(c) Without correspondence optimization

(d) With correspondence optimization

Figure 4.11: Clusters of corresponding points in the reference coordinate system, both with and without correspondence optimization enabled. Points with the same color belong to the same cluster. The wireframe model connects the cluster centers for both (a)-(b) the left and (c)-(d) the right caudates.

demonstrates that the same kinds of optimizations used with discrete m-reps can also be performed with their continuous counterparts. The same kind of one-sample statistics used to drive a Bayesian segmentation algorithm in the discrete case can be applied here, and future work in this direction is certainly planned.

In addition, it shows how specific knowledge of correspondence can be incorporated into the model fitting process. This eliminates excess variability in the parameterization of the objects which could mask real statistical effects of the shape change. The correspondence match optimization introduced does not sacrifice the quality of the fit. Instead, it operates transversely to the fitting process up to the tolerance of the model, requiring no tuning parameter to trade off between the two.

The only downside is the additional time required, since the optimization generally does not converge within the iteration limit. However, the limit was set conservatively high here, and the thresholds for convergence were set very low. Both could likely be aggressively relaxed without sacrificing much fit quality. The optimization is also

easily parallelizable, since each subject is fit to its target shape and to the cluster centers independently. Coordination is only needed to resample the models and compute the cluster centers, $\mu_{(m_0,S_0)}$. These operations require only a trivial amount of the optimization time, the majority of which is spent in the gradient calculations.

Chapter 5

Nonlinear Hypothesis Testing[1]

> If no chain hangs in a perfect catenary and no raindrop is a perfect sphere, this is for the reason that forces and resistances other than the main one are inevitably at work. The same is true of organic form, it is for the mathematician to unravel the conflicting forces which are at work together. And this process of investigation may lead us on step by step to new phenomena...
>
> (Thompson, 1917)

5.1 Introduction

Statistical shape analysis is a powerful tool for understanding the structure of anatomical objects in ways that simpler method like volume analysis cannot. This chapter focuses this analysis on discrete m-reps, via their medial atoms $\underline{m} = (m_0, r, n_0, n_1) \in \mathbb{M}$, where $\mathbb{M} = \mathbb{R}^3 \times \mathbb{R}_+ \times S^2 \times S^2$ and S^2 is the unit sphere in $\mathbb{R}^3$. An example of such a model is given in Figure 5.1. A continuous m-rep model, like the one described in Chapter 3, can easily be converted to this representation. However, because this work was done chronologically before this model was developed, it focuses on the discrete case.

Except for the first term, the spaces involved are nonlinear, although one can easily map the radius to a linear space. This is not so simple for the spoke directions, however. One could argue that since the spoke directions can be recovered directly via

[1] This chapter is based on work done in collaboration with Sarang C. Joshi and Guido Gerig and previously published in (Terriberry et al., 2005).

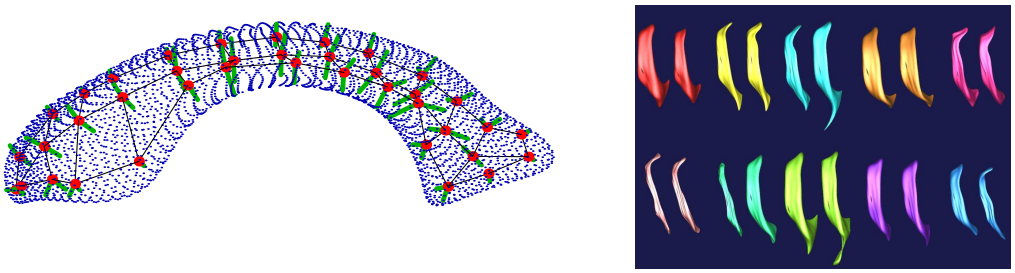


Figure 5.1: Left: An example m-rep of a left lateral ventricle. The mesh vertices and off-shooting spokes make up the medial atoms. The shape the m-rep was fit to is shown as a point cloud surrounding it. Right: Ventricle pairs from five monozygotic twin pairs (top) and five dizygotic twin pairs (bottom).

derivatives of m and r, they do not contain any additional information. However, the belief is that they are valuable in capturing the types of deformations for which medial representations are designed, such as tapering.

A simple motivating example illustrates how the nonlinear combination of a set of linear variates can lead to improvements in the ability to separate two populations. Consider the problem of trying to find a separating plane between two groups of points in the x-y plane, where one group is contained inside the unit circle, and the other outside of it. If one considers only linear combinations of x and y to define this plane, then there is no good solution that provides separation between the groups. However, if one instead parameterizes the points by the 4-tuple (x, y, x^2, y^2), the plane equation $x^2 + y^2 - 1 = 0$ exactly separates them. The latter two coordinates of the tuple do not provide any new information but are still essential to determining the type of class boundaries one can represent.

However, the situation for the spoke directions is not as simple as for polynomials; we cannot simply fall back to the ambient space. The average of two squares in $\mathbb{R}$ is still a square, but the average of two unit vectors in $\mathbb{R}^3$ in general no longer lies in S^2. Therefore this (and many other useful shape spaces) require an alternative to the traditional analysis tools designed for Euclidean spaces.

This chapter formalizes the notion of hypothesis testing against data that lies in the direct product of a large number of nonlinear spaces as a tool for understanding growth and disease. Unlike previous analysis, which has focused on a single (linear) component of the medial atom at a time, we incorporate the entire atom in a single test. We build on the recent methods of Fletcher et al. for one-sample statistical shape

analysis based on m-reps (Fletcher et al., 2003; Fletcher et al., 2004; Fletcher, 2004) to develop two-sample statistics designed to answer the following question: given two samples from two different populations, do they have the same statistical distribution? This is the classic problem of testing the null hypothesis, H_0, that the populations are identical, against its complement, H_1.

The main difficulty arises from the fact that m-reps lie on high-dimensional nonlinear manifolds where assumptions of Gaussianity are unreasonable, making traditional parametric or linear methods inapplicable. We present a multivariate permutation test approach (more accurately described as a multivariate combination of univariate tests) that is equivalent to traditional nonparametric permutation tests in the univariate case and converges to the same result as Hotelling's well-known T^2 test in the linear, normally-distributed case in the limit as the number of observations in each sample goes to infinity. The only tool required on the underlying space the shapes live in is the existence of a metric.

The mechanics of the method are similar to those used in correction for multiple tests (Pantazis et al., 2004), discussed later in more detail. Unlike methods of direct combination, which sum up various test statistics (Chung and Fraser, 1958; Blair et al., 1994), our method is invariant to the scale of each term. This is critical when different shape parameters have different physical units and the choice of weighting between them can be arbitrary. The test also accounts for the dependencies between model parameters.

5.1.1 A Metric Space for M-reps

Fletcher et al. treat the space $\mathbb{M}$ of medial atoms as a Riemannian symmetric space (Fletcher et al., 2004). Such a space is a differentiable manifold and has a Riemannian metric that is invariant to certain transformations of the space. $\mathbb{R}^3$, invariant to translation, uses the normal Euclidean metric, while the positive reals, $\mathbb{R}_+$, use the metric $d(r_1, r_2) = |\log(r_1) - \log(r_2)|$, which is invariant to scale. The unit sphere, S^2, uses distance measured along the surface of the sphere, which is invariant to rotation. We now proceed to describe some of the additional structure of Riemannian symmetric spaces that allows for fast computation of distances and of the mean (defined below), but we emphasize that this structure is not essential for the results in the remainder of this chapter.

Every point $\underline{m}_0$ on the manifold has a tangent plane $T_{\underline{m}_0}\mathbb{M}$, which is a vector

space, and an *exponential map* $\mathrm{Exp}_{\underline{m}_0} : T_{\underline{m}_0}\mathbb{M} \to \mathbb{M}$ that projects points from the tangent plane onto the manifold. Shortest paths, or *geodesics*, on the manifold are all of the form $\mathrm{Exp}_{\underline{m}_0}(tv)$ for a vector $v \in T_{\underline{m}_0}\mathbb{M}$ and $t \in \mathbb{R}$. The set of points where these geodesics cross is called the *cut locus*. Inside the cut locus, the exponential map preserves distances to the base point $\underline{m}_0$ between the vector space and the manifold and has a well-defined inverse, the *log map*, $\mathrm{Log}_{\underline{m}_0} : \mathbb{M} \to T_{\underline{m}_0}\mathbb{M}$. This provides a simple computational framework for computing distances on the manifold: map to the tangent plane at one of the points and compute the length of the corresponding tangent vector.

In $\mathbb{R}^n$, the exponential and log maps are the identity, since it is already a vector space. For $r_0, r \in \mathbb{R}_+$, the maps are given by the ordinary exponential and natural logarithm: $\mathrm{Exp}_{r_0}(r) = e^{r-r_0}$ and $\mathrm{Log}_{r_0}(r) = \log(r) - \log(r_0)$. Geodesics never cross in both of these spaces, so the log map is well-defined everywhere. For S^2, we define the maps with respect to the base point $p = (0, 0, 1)$, the z-axis. The tangent plane is parameterized by points $v = (v_1, v_2) \in \mathbb{R}^2$.

$$\mathrm{Exp}_p(v_1, v_2) = \left(v_1 \frac{\sin\|v\|}{\|v\|}, v_2 \frac{\sin\|v\|}{\|v\|}, \cos\|v\| \right) , \tag{5.1}$$

$$\mathrm{Log}_p(x, y, z) = \left(x \frac{\theta}{\sin\theta}, y \frac{\theta}{\sin\theta} \right) , \tag{5.2}$$

where $\theta = \arccos(z)$. Notice that the log map is not well-defined when $z = 1$ or $z = -1$. In the first case, it can be defined by the limit, which is $(0, 0)$. However in the second case no such limit exists; every geodesic through p crosses at $-p$, which is the only point in the cut locus. These maps can be generalized to other base points by first applying a rotation to align the base point with the z-axis, but there is an inherent ambiguity in the resulting parameterization of the tangent plane. Any rotation in the x-y plane does not move the z-axis, but yields different tangent vectors.

This is merely an outline of the mathematics behind these spaces. For a more complete treatment, we refer the reader to Fletcher's Ph.D. thesis (Fletcher, 2004).

5.1.2 One-sample Statistics in Nonlinear Spaces

In linear spaces the most important property of a probability distribution is often its first moment, the mean. Fréchet generalized the notion of an arithmetic mean of a sample of n points x_i drawn from a distribution in a general metric space $\mathbb{M}$ as the

point that minimizes the sum-of-squared distances (Fréchet, 1948):

$$\hat{\mu} = \underset{x \in \mathbb{M}}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^{n} d(x, x_i)^2 \ . \tag{5.3}$$

This is sometimes referred to as the Fréchet mean or the *intrinsic mean*, but hereafter will just be called the mean.

In general, this mean may not exist, or may not be unique, and without additional structure on the metric space, the minimization may be difficult to perform. However, for Riemannian manifolds, it is possible to compute the gradient of this functional (Karcher, 1977), making a gradient descent algorithm possible (Pennec, 1999). The resulting algorithm is a simple iterative computation

$$\hat{\mu}_{k+1} = \operatorname{Exp}_{\hat{\mu}_k} \left(\frac{1}{n} \sum_{i=1}^{n} \operatorname{Log}_{\hat{\mu}_k}(x_i) \right) \ . \tag{5.4}$$

A reasonable choice of starting point $\hat{\mu}_0$ is one of the data points, x_i. Kendall showed that existence and uniqueness is guaranteed if the data is well-localized (Kendall, 1990). The additional structure provided by a differentiable manifold that allows this gradient calculation is helpful from a computational standpoint but is not required for the ideas presented here.

Fletcher et al. extend this notion to describe the variance of one sample around the mean through the notion of Principal Geodesic Analysis (PGA) (Fletcher et al., 2004). They define this as the set of one-dimensional subspaces (geodesics) that maximizes the (remaining) variance after projection onto the geodesic. However, these geodesics are not directly computable, so they resort to approximating them with principal component analysis (PCA) in the tangent plane at the mean.

There are a number of theoretical issues with this. First, maximizing the variance of the projection is not the only way to define PCA. Minimizing the sum of squared distances to the subspace provides an equivalent definition in the linear case, but this equivalence breaks down in the nonlinear case, even for simple spaces such as the sphere. Furthermore, the geodesic that maximizes the variance of the projection does not of necessity even pass through the mean. Consider a set of points evenly distributed on the equator of a sphere, with one point offset slightly into the northern hemisphere. The mean of these points will be (approximately) the north pole, but the geodesic that maximizes the projected variance of the points is (approximately) the equator. Nev-

ertheless, for well-localized data the PCA-in-the-tangent-plane approximation works well.

5.1.3 Two-sample Statistics

If we assume that both of our distributions are identical around the mean and that they can be characterized entirely by the distance from the mean (e.g., they are isotropic), then a single global distance value is sufficient to construct a univariate permutation test for equality of the two means. Permutation tests are appealing because they make no other distributional assumptions, requiring only that the data in each group be exchangeable under the null hypothesis that they do in fact come from the same distribution. The interested reader is referred to Bradley (Bradley, 1968) or Nichols and Holmes (Nichols and Holmes, 2002) for details.

However, geometric models contain parameters in nonlinear spaces, like the sphere. Some parameters may have a large variance, masking the effects of other variables with a smaller variance that might provide greater discrimination. Some may be highly correlated, unduly increasing their contribution to the distance over that of parameters with less correlation. Some will have completely different scales, and appropriate scale factors need to be determined to combine them in a single metric. These factors make the assumption that the distance from the mean entirely characterizes the distribution hard to justify.

For example, scaling the model will change the distance between medial atom centers, $\mathbf{x}$, without affecting the distance between radii or spoke directions. To combat this, Fletcher et al. propose scaling the latter by the average radius across corresponding medial atoms (Fletcher et al., 2004), but this choice is somewhat arbitrary. It does restore invariance to scale, but it does nothing to handle differing degrees of variability or correlation. Different choices of scale factors will produce tests with differing amounts of power.

In $\mathbb{R}^n$ when relaxing the assumption that the distribution is characterized by the distance from the mean and instead assuming only a common covariance, the classic Hotelling's T^2 test provides a test invariant to coordinate transformations. For normally distributed data, it is uniformly the most powerful (see a standard text, such as Anderson's (Anderson, 1958), for a derivation). The test is based on the statistic $T^2 \propto D^2 = (\hat{\mu}_1 - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$, where $\hat{\mu}_1$ and $\hat{\mu}^2$ are the sample means and $\hat{\Sigma}$ the pooled sample covariance. Any linear change of coordinates yields a corresponding

change in metric, but this is absorbed by the $\hat{\Sigma}^{-1}$ term, leaving the statistic unchanged.

5.2 Multivariate Permutation Tests

The hypothesis test proposed is an attempt to generalize the desirable properties of Hotelling's T^2 test to a nonparametric, nonlinear setting. We cannot take advantage of the vector space structure of the tangent plane, as Fletcher et al. do, to apply Hotelling's test directly, because there is a different tangent space around each sample's mean, and there may be no unique map between them. For example, on the sphere the ambiguity in the coordinate axes used to parameterize the tangent planes gives such a map one degree of freedom. Instead, we take a more general approach, only requiring that our objects lie in a metric space.

Our approach is based upon a general framework for nonparametric combination introduced by (Pesarin, 2001). The general idea is to perform a set of partial tests, each on a different aspect of the data, and then combine them into a single summary statistic, taking into account the dependence between the variables and the true multivariate nature of the data. These results of these partial tests produce a new set of features, the p-values, that unlike the original features are insensitive to scale and have a known marginal distribution. The latter can be used to construct a mapping to a linear space and develop a test based on a standard Gaussian distribution to determine if the means are equal under the assumption of two distributions with the same structure around the mean. We now describe the details.

5.2.1 The Univariate Case

We begin by introducing notation and describing the procedure for a single, univariate permutation test. Suppose we have two data sets of size n_1 and n_2, $x_1 = \{x_{1,i}, i \in 1 \dots n_1\}$ and $x_2 = \{x_{2,i}, i \in 1 \dots n_2\}$, and a test statistic, $T(x_1, x_2)$. To test for a difference in the means, a natural test statistic is

$$T(x_1, x_2) = d(\hat{\mu}_1, \hat{\mu}_2) \ , \tag{5.5}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the sample (Frechét) means of the two data sets computed via the optimization in (5.3) and d is the metric on the nonlinear manifold containing the data points. For other tests, other statistics are possible.

Under the null hypothesis, both samples are drawn from the same distribution, so we may randomly permute the data between the two groups without affecting the distribution of $T(x_1, x_2)$. We pool the data together and then generate $N = \binom{n_1+n_2}{n_1}$ random partitions into two new groups, still of size n_1 and n_2. We label these $x^k_{1,i}$ and $x^k_{2,i}$, with $k \in 1 \dots N$, and compute the value of the test statistic, T^k, for all of them. We always include the actual observed groupings among this list and denote its test statistic T^o. This forms an empirical distribution of the statistic, from which we can calculate the probability of observing T^o under the null hypothesis:

$$p(T^o) = \frac{1}{N} \sum_{k=1}^{N} H(T^k, T^o) \ , \qquad H(T^k, T^o) = \begin{cases} 1, & T^k \geq T^o \\ 0, & T^k < T^o \end{cases} . \tag{5.6}$$

5.2.2 Partial Tests

If our data can be adequately summarized by a single test statistic, then this is the end of the story. We now turn to the case where we have M test statistics, one for each of the parameters in our shape model. Let $\mu_{1,j}$ and $\mu_{2,j}$ be the means of the jth model parameter for each population. Then we wish to test whether any hypothesis $H_{1,j} : \{\mu_{1,j} \neq \mu_{2,j}\}$ is true against the alternative, that each null hypothesis $H_{0,j} : \{\mu_{1,j} = \mu_{2,j}\}$ is true. The partial test statistics $T_j(x_1, x_2), j \in 1 \dots M$ are defined analogously to (5.5), and the values for permutations of this data are denoted T^k_j, with $j \in 1 \dots M$, $k \in 1 \dots N$.

Given that each $T_j(x_1, x_2)$ is significant for large values, consistent, and marginally unbiased, Pesarin shows that a suitable combining function (described in the next section) will produce an unbiased test for the global hypothesis H_0 against H_1 (Pesarin, 2001). The meaning of each of these criteria is as follows:

1. **Significant for large values:** Given a significance level α and the critical value of $T_j(x_1, x_2)$ at α—T^α_j—the probability that $T^o_j \geq T^\alpha_j$ is at least α. For a two-sided test $T_j(x_1, x_2)$ must be significant for both large and small values.

2. **Consistent:** As the sample size $n = n_1 + n_2$ goes to infinity, the probability that $T^o_j \geq T^\alpha_j$ must converge to 1.

3. **Marginally unbiased:** For any threshold z, the probability that $T^o_j \leq z$ given $H_{0,j}$ must be greater than the probability that $T^o_j \leq z$ given $H_{1,j}$, irrespective of

the results of any other partial test. This implies that T_j^o is positively dependent in $H_{1,j}$ regardless of any dependencies between variables.

Since each of our tests are restricted to the data from a single component of the direct product and we have assumed that the distributions around the means are identical, they are marginally unbiased. However, consider adding a separate test for equality of the distributions around the means. Our test statistic for the means is only unbiased if the distributions around them are in fact equal, so this unbiasedness would be conditioned on the outcome of the additional test. Hence no such test can be added without also constructing a different test against the means.

To illustrate these ideas, we present a simple example, which we will follow through the next few sections. We take two samples of $n_1 = n_2 = 10$ data points from the two-dimensional space $\mathbb{R} \times \mathbb{R}_+$, corresponding to a position and a scale parameter. The samples are taken from a multivariate normal distribution by exponentiating the second coordinate and then scaling both coordinates by a factor of ten. They are plotted together in Figure 5.2a. They have the common covariance (before the exponentiation) of $\frac{1}{2}\left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right)$, and the two means are slightly offset in the second coordinate. That is, $\mu_{1,1} = \mu_{2,1}$, but $\mu_{1,2} < \mu_{2,2}$.

We construct $M = 2$ partial test statistics using (5.5) for each coordinate and evaluate them using Monte Carlo simulation. To avoid an exponential complexity, we use a fixed $N = 10,000$ permutations, which still provides an unbiased test. The results are shown in Figure 5.2b. The first partial test value, T_1^o, lies in the middle of the abscissa distribution, while the second, T_2^o, lies near the edge of the ordinate distribution. However, the scale of the first test is much larger, because no logarithm is involved in its metric.

5.2.3 Multivariate Combination

Given the partial tests from the previous section, we wish to combine them into a single test while preserving the underlying dependence relations between the tests. This is done in the following manner. We apply the same N permutations to the data when computing each of partial tests and then compute a p-value using the empirical distribution for that test over all of the other permutations:

$$p(T_j^k) = \frac{1}{N} \sum_{l=1}^{N} H(T_j^l, T_j^k) \ . \tag{5.7}$$

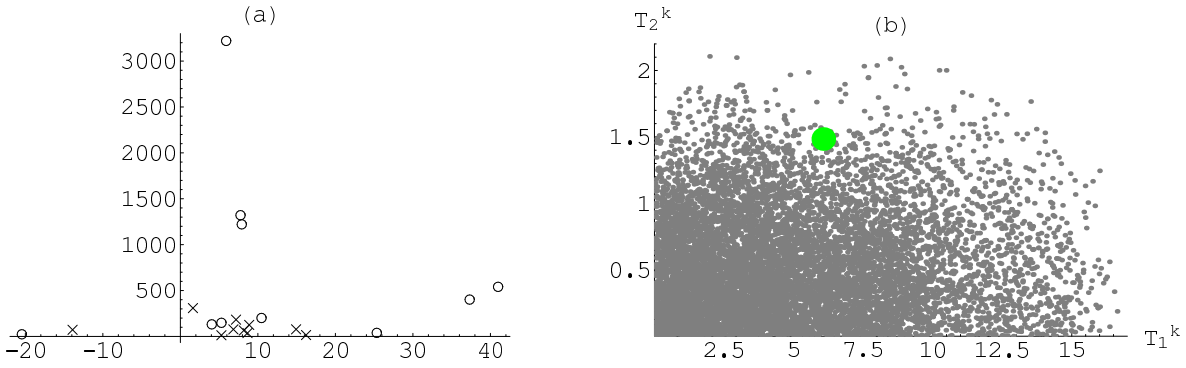


Figure 5.2: The observed data and test statistics for our simple example. (a) shows the distribution of our two samples, with ×'s for the first and ∘'s for the second. (b) shows the distribution of the partial test statistics under permutation. The large dot indicates the location of the observed data point.

Thus, for every permutation k we have a column vector of p-values,

$$p(T^k) = (p(T_1^k), \ldots, p(T_j^k), \ldots, p(T_M^k))^T \ . \tag{5.8}$$

It is critical to use the same permutations for each partial test, as this is what captures the nature of the joint distribution. Because the p-value vector is needed for every permutation instead of just the observed test statistic, computing them is equivalent to sorting the empirical distribution for each feature, and thus is $O(MN \log N)$.

We now wish to design a combining function to produce a single summary statistic, $T'(p(T^k))$, from each p-value vector. For one-sided tests, this statistic must be monotonically non-increasing in each argument, must obtain its (possibly infinite) supremum when any p-value is zero, and the critical value T'^{α} must be finite and strictly smaller than the supremum. If these conditions are satisfied, along with those on the partial tests from the previous section, then $T'(p(T^k))$ will be an unbiased test for the global hypothesis H_0 against H_1 (Pesarin, 2001).

Our combining function is motivated by the two-sided case, where we can use the Mahalanobis distance. First, we compute a U^k vector for each permutation, where

$$U_j^k = \Phi^{-1}(p(T_j^k) - \frac{1}{2N}) \tag{5.9}$$

and $j \in 1 \ldots M$. Here Φ is the cumulative distribution function for the standard normal distribution. The extra $\frac{1}{2N}$ term keeps the values finite when the p-value is 1 and is

negligible as N goes to infinity. The resulting vector lies in a linear space, making a linear test appropriate.

Because the distribution of p-values for each partial test is uniform by construction, the marginal distribution of the U_j^k values over k for a single j is standard normal. Arranging these vectors into a single $N \times M$ matrix U, we can estimate the covariance matrix $\hat{\Sigma}_U = \frac{1}{N} U^T U$ and use the Mahalanobis statistic: $T'^k = (U^k)^T \hat{\Sigma}_U^{-1} U^k$. In the event that the data really is linear and normally distributed, $\hat{\Sigma}_U$ matrix converges to the true covariance as the sample size goes to infinity (Pallini and Pesarin, 1992), making it asymptotically equivalent to Hotelling's T^2 test. Even if the sample size is small, the matrix Σ_U is well-conditioned regardless of the number of variables, since it is the covariance over the N permutations.

Typically, our distances are not signed, so we are interested in a one-sided test. In this case, we use the positive half of the standard normal c.d.f., $U_j^k = \Phi^{-1}(1 - \frac{1}{2}(p(T_j^k) - \frac{1}{2N}))$ and assume the U^k distribution is symmetric about the origin. This assumption, however, implies that the covariance between $U_{j_1}^k$ and $U_{j_2}^k$ when $j_1 \neq j_2$ is exactly zero. The diagonal entries of $\hat{\Sigma}_U$ are 1 by construction, so $\hat{\Sigma}_U = I$, the identity matrix. The fact that the p-values of the partial tests are invariant to scale obviates the need for arbitrary scaling factors. Thus, our one-sided combining function is

$$T'^k = (U^k)^T \cdot U^k \ . \tag{5.10}$$

Normality of the partial test statistics is not required, and even though the marginal distributions of the U^k vectors are normal, the joint distribution may not be. Therefore, we must use a nonparametric approach to estimating the distribution of the T'^o statistic under the null hypothesis. Just as in the univariate case, this produces a single p-value:

$$p(T'^o) = \frac{1}{N} \sum_{k=1}^{N} H(T'^k, T'^o) \ . \tag{5.11}$$

It is this nonparametric approach that corrects for correlation among the tests, even without explicit diagonal entries in the covariance matrix.

We return to our example from the previous section. The U^k vectors are plotted in Figure 5.3a, along with the $\alpha = 0.95$ decision boundary, and our sample is shown to lie outside of it. Equal power is assigned to alternatives lying at the same distance from the origin in this space. Figure 5.3b shows this boundary mapped back into the space of the original p-values. The p-values of the individual partial tests are 0.36 and 0.022,

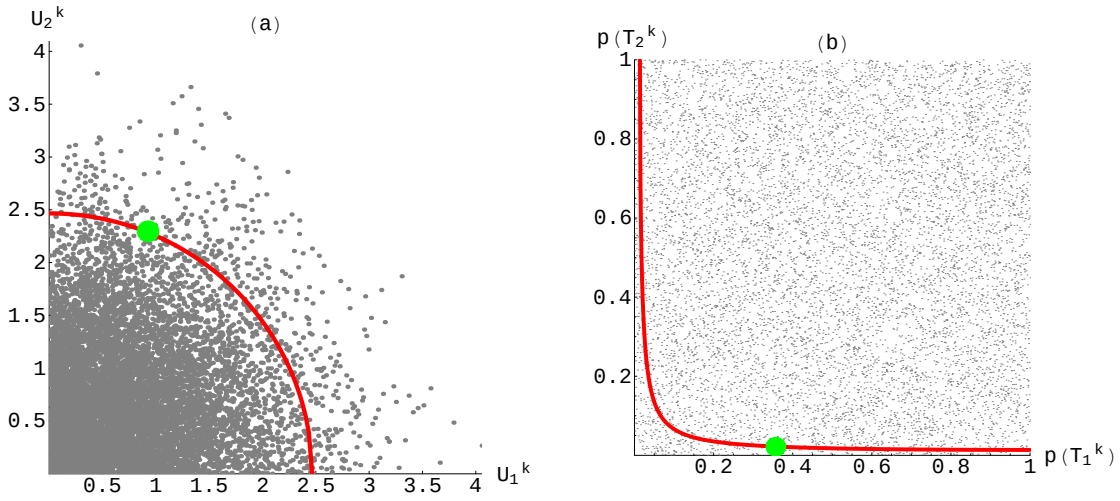


Figure 5.3: The empirical distribution of our example plotted against the decision boundary at $\alpha = 0.95$. (a) The distribution of the U^k vectors, where the cutoff is a circle centered around the origin. (b) The distribution of the original p-values with the decision boundary pulled back into this space.

and the combined result is 0.049.

The entire procedure can be summarized as follows. Compute N permutations and compute the partial test statistics T_j^k for each feature j and permutation k. Then map these to p-values with (5.7) using the empirical distribution for each feature. These p-values have a uniform marginal distribution and are invariant to scale, or indeed any monotonic transformation. Next, stack them into a vector $p(T^k)$ for each permutation and map them via (5.9) to $\mathbb{R}^M$. Here they have a Gaussian (or half-Gaussian, if unsigned) marginal distribution. Finally, combine these vectors into a single test statistic T'^k by (5.10) and compute the overall p-value from its empirical distribution with (5.11).

5.2.4 Relation to Other Testing Procedures

The method presented here is very similar to procedures used in correction for multiple tests, such as that proposed by Pantazis et al. (Pantazis et al., 2004). In fact, another alternative for a combining function is Tippet's $T'^k = \max_{j=1}^{M}(1 - p(T_j^k))$, which results in a Bonferroni-style correction (Pesarin, 2001). Some authors have suggested methods of direct combination applied to the T_j^k statistics themselves (Chung and Fraser, 1958;

Blair et al., 1994). They are more appealing computationally, being $O(MNn)$ instead of our method's $O(MN(n+\log(N)))$, but they do not avoid problems of differing scale or strong correlation.

Consider what happens when $T'^k = \sqrt{(T_1^k)^2 + (T_2^k)^2}$. Now, the first test dominates the results, and the overall p-value becomes 0.34. With $n_1 = n_2 = 100$ samples, our test becomes much more significant ($p = 0.0008$), while the direct combination test becomes even worse ($p = 0.44$).

5.3 Experimental Data and Results

The data for our experiments comes from a twin pair schizophrenia study conducted by Weinberger et al. (Weinberger et al., 2001). High resolution ($0.9375 \times 0.9375 \times 1.5$ mm^3) Magnetic Resonance Imaging (MRI) scans were acquired from three different subject groups: 9 healthy monozygotic twin pairs (MZ), 10 healthy dizygotic twin pairs (DZ), and 9 monozygotic twin pairs with one twin discordant for schizophrenia and the co-twin unaffected. See Figure 5.1 for some examples. A fourth group of 10 healthy non-related subject pairs (NR) was constructed by matching unrelated members of the two healthy groups. All four groups were matched for age, gender, and handedness. A tenth healthy, monozygotic twin pair was discarded due to segmentation problems attributed to head trauma suffered by one of the twins in a car accident at age seven. A tenth twin pair discordant for schizophrenia was discarded due to hydrocephaly in the unaffected twin.

The left and right lateral ventricles were segmented using supervised classification and 3-D connectivity (van Leemput et al., 1999). An automatic morphological closing operation was applied to ensure a spherical topology. An area-preserving map was used to map them to a sphere, after which they were converted to a spherical harmonics representation (SPHARM) (Brechbühler et al., 1995). Correspondence on the boundary was established using the first order harmonics (Gerig et al., 2001). Point Distribution Models (PDMs) were constructed by uniformly sampling the boundary at corresponding points. The m-rep models were constructed using a robust method that ensures a common medial topology (Styner et al., 2003). For our data, this consists of a single medial sheet with a 3×13 grid of medial atoms, which provides 98% volume overlap with the original segmentations.

From this data set, we wish to determine if the twin pairs that were more closely related had smaller variations in shape. We also wish to see if the shape variations

between the discordant and the unaffected twins in the schizophrenic pairs is similar to the normal variation between healthy monozygotic twins. For this purpose, we use the partial test statistics

$$T_j(x_1, y_1, x_2, y_2) = \frac{1}{n_2}\sum_{i=1}^{n_2} d(x_{2,i,j}, y_{2,i,j}) - \frac{1}{n_1}\sum_{i=1}^{n_1} d(x_{1,i,j}, y_{1,i,j}) \ . \tag{5.12}$$

Here (x_1, y_1) form the twin pairs for one group, while (x_2, y_2) form the twin pairs for the other. The partial tests are applied separately to all three components of the medial atom location, m_0, as well as the radius and two spoke directions. This gives six partial tests per medial atom, for a total of $M = 3 \times 13 \times 6 = 234$, much larger than the sample size. Each is a one-sided test that the variability in group 2 is larger than that in group 1.

For consistency with previous studies (Styner et al., 2005), all shapes were volume normalized. After normalization, we also applied m-rep alignment, as described by Fletcher et al. (Fletcher et al., 2004), to minimize the sum of squared geodesic distances between models in a medial analog of Procrustes alignment. First, the members of each twin pair were aligned with each other, and then the pairs were aligned together as a group, applying the same transformation to each member of a single pair.

In order to ensure invariance to rotations, we had to choose data-dependent coordinate axes for the m_0 component of each medial atom. Our choice was the axes that diagonalized the sample covariance of the displacement vectors from one twin's atom position to the other at each site. While this had some small influence on the results, the general trend was the same irrespective of the axes used.

For each pair of twin groups, we generated $N = 50,000$ permutations and computed their p-value vectors using (5.7). Following Sect. 5.2.3, these were mapped into U^k vectors, from which the empirical distribution of the combined test statistic T'^k from (5.10) was estimated, producing a single global p-value via (5.11).

The results are summarized in Table 5.1. For comparison, we list the results of a previous study that used a univariate test on the average distance between corresponding points on the PDMs (Styner et al., 2005). While the significance of a p-value on an experimental data set is not a useful metric for comparing different methods, it is interesting to see the differences between the two. Our tests give a consistent ranking: MZ $\approx$ DS $<$ DZ $\approx$ NR, which is fully transitive. The boundary study, however, finds a significant difference between DZ and NR, but fails to identify the difference between DS and DZ.

	Our Study		Boundary Study (Styner et al., 2005)	
	Left	Right	Left	Right
MZ vs. DS	0.12	0.38	0.28	0.68
MZ vs. DZ	**0.00006**	**0.0033**	**0.0082**	**0.0399**
MZ vs. NR	**0.00002**	**0.00020**	**0.0018**	**0.0006**
DS vs. DZ	**0.020**	**0.0076**	0.25	0.24
DS vs. NR	**0.0031**	**0.00026**	**0.018**	**0.0026**
DZ vs. NR	0.16	0.055	**0.05**	**0.016**

Table 5.1: p-values for paired tests for the difference in the amount of shape variability in groups with different degrees of genetic similarity. Results from our method are in the first two columns, while results from a previous study (Styner et al., 2005) are in the last two for comparison. Groups are: monozygotic (MZ), monozygotic twins with one twin discordant for schizophrenia (DS), dizygotic (DZ), and non-related (NR). Results significant at the $\alpha = 0.95$ level are shown in **bold**.

We also performed local tests, to identify specific medial atoms with with strong differences. A multivariate test was conducted using our procedure on the 6 components of each atom, and the results were corrected for multiple tests using the minimum p-value distribution across the shape, as described by Pantazis et al. (Pantazis et al., 2004). The results are shown in Figure 5.4.

However, we get a very different result when we try to test for differences between the schizophrenic groups and the controls. We pooled the healthy MZ and DZ subjects and created a control group by randomly selecting one twin from each pair. A template healthy ventricle was then constructed from the mean of the other twins in each of the pairs. We then tested the distance to this mean for the control group, the group of affected twins from each DS pair, and the group of unaffected twins. *Unlike* the (Styner et al., 2005) study on the boundary, *none* of our results were significant.

What went wrong? If one examines the ventricles pictured in Figure 5.1, it is immediately apparent that the shape variation between two unrelated subjects is much larger than the variation between two twins. In the face of this large variation, the implicit correspondence of the samples on the medial axis breaks down.

Consider Figure 5.5, which shows plots the atom locations of 18 m-reps after alignment. If the alignment had succeeded, small clusters in a regular grid pattern corresponding to the 3×13 grid in each individual m-rep should be readily apparent. Instead, there is no visible structure at all because atoms were allowed to move arbitrarily during the model fitting process, with no regard paid to the correspondence

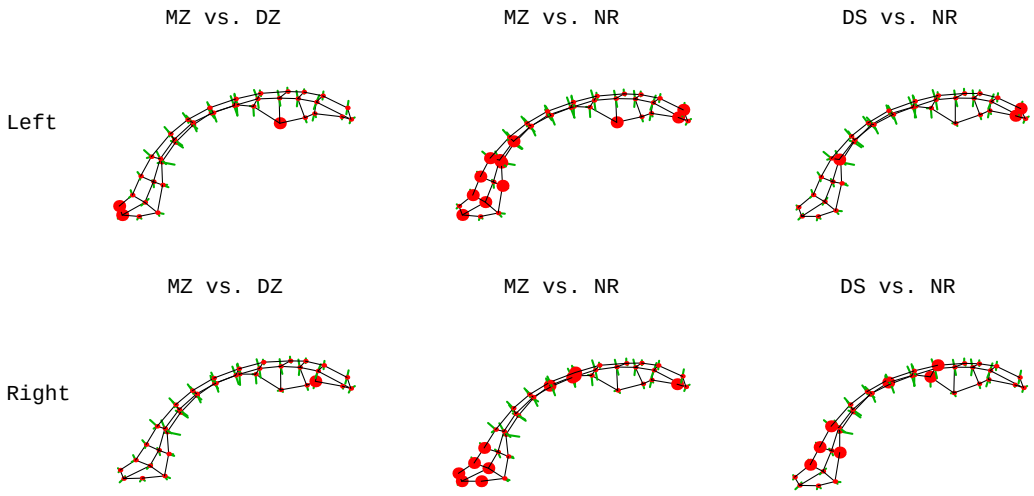


Figure 5.4: Results for local tests for the difference in shape variability in groups with different degrees of genetic similarity. Atoms with differences significant at the $\alpha = 0.95$ level are shown in a larger size. Tests not shown had no significant local differences.

across subjects of the associated piece of boundary each atom represents. As a result, the total variation within the group dwarfs any variation between groups. This was precisely the motivation for the correspondence optimization method introduced in Chapter 4.

5.4 Conclusion

We have presented a multivariate permutation test approach for hypothesis testing in direct products of metric spaces. The resulting test does not require a priori scaling factors to be chosen and captures the true multivariate nature of the data. It is well-defined even in the high-dimensional, low-sample size case. The method has been applied to shape discrimination using m-reps, though it is suitable for any type of metric data, including potentially categorical data.

An important area for future research is the design of suitable partial tests to use in each space. Because they cannot be broken into smaller pieces than a single component of the direct product, the distance to the mean and similar tests are limited in the types of distributions they can describe. For example, the distance from the mean can only characterize an isotropic distribution on the sphere. An interesting candidate is the

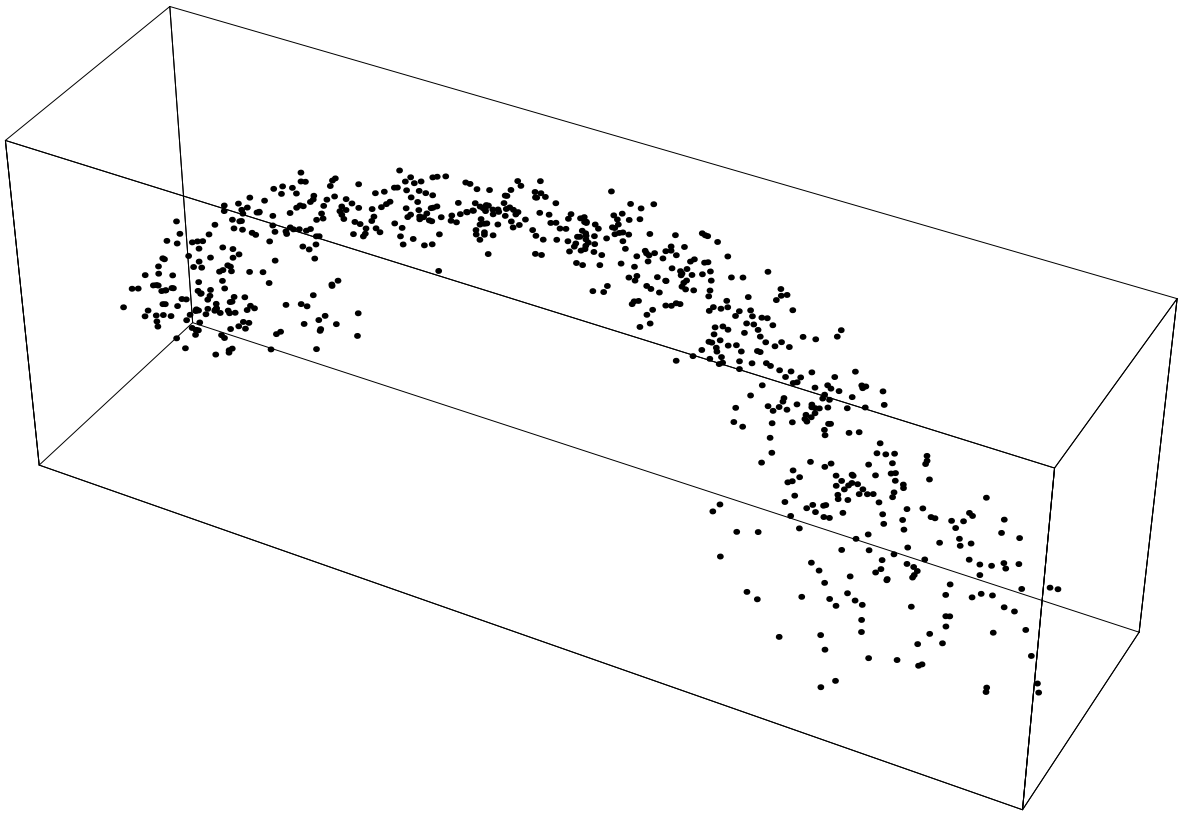


Figure 5.5: The atom centers (m_0) for the m-reps of all 18 of the left lateral ventricles in the MZ group, after Procrustes alignment.

test designed by Hall and Tajvidi, which can test for equality of entire distributions in a single metric space with commensurate coordinates (Hall and Tajvidi, 2002). This would allow relaxation of the assumption of identical distributions about the mean.

For manifolds, another possibility is the use of tests based on Distance Weighted Discrimination (DWD) (Marron and Todd, 2002). Styner et al. have recently explored using DWD in the tangent space of the pooled mean and achieved promising results (Styner et al., 2006). However, this must still contend with the fact that the distance between two points on the manifold is *not* preserved in the tangent plane in general. Rather, only the distance of each point to the mean (within the cut locus) is preserved. As long as the portion of the manifold containing the population is nearly flat, this will not have a large influence, but it is an open question how gross of an approximation this is.

It is also possible to extend our approach to different shape models, such as PDMs with surface normals or deformable templates (Grenander and Keenan, 1993; Miller and

Younes, 2001; Fletcher and Whitaker, 2006). These latter in particular describe shapes as points on an infinite dimensional Riemannian manifold representing the deformations. In such spaces, the degree of curvature and structure of the cut locus are much less well-understood than on the sphere, and even simple linearization techniques must contend with the fact that the tangent space is also infinite dimensional. Nevertheless, steps have been taken towards computationally tractable methods of approximating the log and exponential maps of these spaces (Arsigny et al., 2006).

Chapter 6

Conclusion

> But we come in time to forms which, though both may still be simple, yet stand so far apart that direct comparison is no longer legitimate.
>
> (Thompson, 1917)

At the end of this work, this chapter takes a moment to revisit the claims laid out in Chapter 1 and to discuss possible future research directions. These claims were:

1. *A novel "control curve" formulation capable of enforcing first- and second-order boundary conditions at the edges of subdivision surfaces.*

   This was the primary mechanism for enforcing boundary conditions laid out in Chapter 3. Although it was put to a specific application there, it could be used to enforce any boundary condition that can be expressed in terms of the derivative of the subdivision surface across the edge. The fourth-order interpolating spline from Section 3.2.3 gives a second free parameter which can be used to enforce a second-order boundary condition as well.

   The main drawback to this method is its inability to handle corners on the edge due to the asymmetry of the interpolation. Though eliminating these corners has long been desired for anatomical objects, this limitation of the control curve method was the real impetus for moving to a subdivision surface formulation. It might be possible to include corners given boundary conditions that do not interfere with each other. For example, it is very easy to see how a constraint on the tangent vector to the medial sheet m_u on the $u = 0$ edge could co-exist with a constraint on $\|\nabla r\|$ on the $v = 0$ edge. Enforcing the latter via the derivative r_v would not interfere with the former. Such an approach may eventually lead to a method for handling six-junctions.

2. *A new continuous medial model based on this formulation, the first such model capable of representing branching in three dimensions.*

   This is the featured result of this dissertation, the main focus of Chapter 3. It is a complicated construction, with each of the pieces—subdivision surfaces with ordinary, corner-free boundary, control curves, and interpolating splines—included to solve very specific parts of the problem. Even with all these elements in place, the method for handling fin-points was non-obvious. However, it has only marginally more computational complexity than subdivision surfaces by themselves. The computation of the control curves affects evaluation in only one direction on only some of the patches. The majority of the points, even on such patches, are evaluated with a cubic B-spline.

3. *A method of approximating medial integrals driven by a uniform sampling of these continuous models.*

   These medial integrals are the fundamental tool developed in Chapter 4 for taking advantage of the link between the medial axis and the boundary. It allows properties of the boundary to be expressed on the boundary but be evaluated in medial coordinates. This gives continuous medial models advantages of both representations. Although Damon introduced the concept, this work is the first application to demonstrate that they are computable and to apply them to computational problems. However, it does not derive error bounds for the surface area element approximation of (4.2). Especially near the crest region, where $U_u^{\pm}$ diverges, this approximation error may be non-trivial. A higher-order approximation could be helpful here.

4. *Applications of these medial integrals. These include computing volume overlap to evaluate goodness-of-fit and computing second-order moments to align models to a common position, orientation and scale.*

   Chapter 4 gave a number of examples of the way medial integrals could be used to compute basic object properties. Although the applications were simple, they paved the way for understanding the more complex functions to follow. Their most important feature was demonstrating that these properties could be computed without first converting to a boundary representation, avoiding the additional complexity and approximation error such a conversion would involve.

5. *A multi-scale model fitting framework utilizing a highly constrained multivariate nonlinear optimization to fit these models to a target shape.*

   The optimization framework developed in Chapter 4 serves as a validation that the medial model presented in Chapter 3 can be integrated into a complete shape modeling system. To that end, the methods used were fairly basic. More robust methods of alignment and initialization would produce better and more reliable results, and the automatic learning of population statistics could provide a more reliable means of avoiding local minima.

   A number of problems could be avoided simply by moving from the F_I^σ objective function for binary images to the F_T^σ objective function for triangle meshes. For example, if the initialization places a part of the boundary of the model on the wrong side of the true medial axis of the target, F_I^σ will actually drive the boundary farther away, in the wrong direction. In contrast, the surface normal agreement required by F_T^σ would flip the sign of the gradient in these regions, pulling the surface in the proper direction. Similarly, the model sometimes fails to match thin tips well because the boundary in these regions is nearly parallel to the medial axis. This means that the gradient of F_I^σ points nearly orthogonal to the direction needed to push the model in or out of this tip. While F_T^σ as presented has a similar problem, it could be extended to also require curvature agreement instead of just normal agreement, which would help point the gradient in the right direction and drive the end curves into the crests.

6. *A new correspondence optimization method that works in tandem with the model fitting process to produce a group of models with a common parameterization.*

   This feature serves as a good example of the reasons to use a continuous medial representation over a discrete one, and it is another non-trivial example of how the link between the medial axis and the boundary can be exploited to give a medial model the advantages of the latter without sacrificing those of the former. The sliding process used to remove ambiguities in the parameterization of the axis would not be possible with a discrete representation. Although the amount of variability this process eliminates is visually remarkable, little has been done to evaluate the quantitative effect this has on the power of statistical tests, and more work needs to be done in this area. Moreover, the effects of conflicts between the explicit correspondence given on the boundary and the implicit correspondences achievable while respecting the medial geometry need

further consideration. While it is believed that shape changes due to growth and deformation should be well-represented by the latter, if this is not the case for a given population then medial methods may not be an appropriate representation for it.

7. *A novel nonlinear hypothesis test for m-reps, which simultaneously considers all of the parameters of the shape model.*

   This last accomplishment returns to the driving problem of the dissertation: being able to compare multiple populations of shapes. It was applied to real data and shown to provide more consistent results than boundary methods. Yet it was not as powerful in the presence of very large shape variation, discovering no significant effects where boundary analysis found some. It remains to be seen whether this is due to the nature of the shape representation, or the test itself. The obvious experiment to perform would be to see if the correspondence optimization of Chapter 4 improves the situation, but the ventricles are a very challenging data set, and the aforementioned improvements in model-fitting robustness are needed before this can be done. A formal power analysis on the test itself would also be instructive as to the types of changes it is good at detecting.

One other fundamental question this dissertation has not addressed is what the "right" distance metric to use for comparing medial structures is. With a discrete model, there are a limited number of options, and the metric of Fletcher et al. made use of here is a logical one, in the sense that it is directly computable from and intrinsic to the model parameters. However, when a continuous representation is available, one is not limited to the information available at discrete samples. For example, the intrinsic medial coordinates of objects with the same medial topology could allow the expression of shape change as a diffeomorphism on the object interior, although there is some difficulty in representing this computationally.

Restricting these coordinates to the boundary produces a similar diffeomorphism. Either space of diffeomorphisms has its own set of intrinsic metrics, though the fact that the domain of the transformation is no longer all of $\mathbb{R}^n$ complicates the issue. An approach along these lines can be viewed as a natural extension of the discrete case in the following sense: whereas the rotation, translation, and elongation of a spoke can be likened to the transport of the small, differential piece of surface attached to its endpoint, the medial measure theory used throughout this dissertation would seem to provide exactly the means to weight the relative geometric significance of these changes.

However, differing degrees of variability and correlation may make the statistical significance quite different. Unfortunately, such spaces are also infinite dimensional, so the direct product decomposition on which the tests presented here rely does not apply. Perhaps it could be brought to bear on an appropriate numerical approximation, and in some sense this is precisely what the method presented in Chapter 5 does, but in all likelihood fundamentally different statistical tools for infinite dimensional spaces will need to be developed. Given the difficulty of the problem already in high dimensions, it is unlikely that simple extensions of existing techniques will engender much success.

Appendix A

Computing $U^{\pm}$ and $S_{\mathbf{rad}}$ in Three Dimensions

This appendix describes how to compute all of the values necessary to reconstruct the boundary, check for overfolding, and perform medial integration at an arbitrary smooth point of a parameterized 3D medial axis. These include the radial vectors $U^{\pm}$ and its derivatives, the radial shape operator S_{rad}, and the medial measure dM. At non-edge singular points, several patches will meet, and each one contributes a valid pair of values for U and S_{rad}. At edge singular points, and on the edge patches of edge-closure points, U may still be computed, but its derivatives, and thus S_{rad}, may not be. The same holds true for places where the medial axis is only C^1, though if second derivatives may be evaluated when approaching from one or more sides, multiple valid values of S_{rad} may still be computed. The radial curvature condition must then be enforced for all of them.

The computations given here are mostly straightforward, but they are provided for completeness. They require derivatives of m and r up to second order, which are easily computable on a normal B-spline patch and everywhere except the extraordinary vertex on an interior patch containing one. On edge patches and branch patches, their computation is more involved. They can be approximated numerically, but Section A.1 derives exact expressions for them on edge patches.

First, two quantities will appear repeatedly:

$$a \triangleq r_u m_v - r_v m_u \qquad\qquad b \triangleq m_u \times m_v \tag{A.1}$$

We will also write $a^2 = a \cdot a$ in place of $\|a\|^2$ for simplicity, and similarly with b^2. As an aside, we note that $b^2 = E_m G_m - F_m^2$, the determinant of the metric tensor on m. With this notation and some tedious algebra, we can write

$$\nabla r = \frac{a \times b}{b^2} \tag{A.2}$$

Substituting this into (2.3) gives us an expression for $U^{\pm}$:

$$U^{\pm} = \frac{1}{b^2} \cdot \left(b \times a \pm \sqrt{b^2 - a^2} \cdot b \right) \ . \tag{A.3}$$

In order to compute S_{rad}, the derivatives of $U^{\pm}$—and thus also of a and b—are required:

$$a_u = r_u m_{uv} + r_{uu} m_v - r_v m_{uu} - r_{uv} m_u \ , \tag{A.4}$$
$$a_v = r_u m_{vv} + r_{uv} m_v - r_v m_{uv} - r_{vv} m_u \ , \tag{A.5}$$
$$b_u = m_u \times m_{uv} + m_{uu} \times m_v \ , \tag{A.6}$$
$$b_v = m_u \times m_{vv} + m_{uv} \times m_v \ . \tag{A.7}$$

The following are also needed:

$$(a^2)_u = 2a \cdot a_u \ , \qquad (a^2)_v = 2a \cdot a_v \ , \tag{A.8}$$
$$(b^2)_u = 2b \cdot b_u \ , \qquad (b^2)_v = 2b \cdot b_v \ , \tag{A.9}$$
$$(b \times a)_u = b \times a_u + b_u \times a \ , \qquad (b \times a)_v = b \times a_v + b_v \times a \ , \tag{A.10}$$
$$(\sqrt{b^2 - a^2})_u = \frac{(b^2)_u - (a^2)_u}{2\sqrt{b^2 - a^2}} \ , \qquad (\sqrt{b^2 - a^2})_v = \frac{(b^2)_v - (a^2)_v}{2\sqrt{b^2 - a^2}} \ . \tag{A.11}$$

The derivatives of $U^{\pm}$ may now be expressed as

$$U^{\pm}_u = \frac{1}{b^2} \left((b \times a)_u \pm \sqrt{b^2 - a^2} \cdot b_u \pm (\sqrt{b^2 - a^2})_u \cdot b - U^{\pm} \cdot (b^2)_u \right) \ , \tag{A.12}$$
$$U^{\pm}_v = \frac{1}{b^2} \left((b \times a)_v \pm \sqrt{b^2 - a^2} \cdot b_v \pm (\sqrt{b^2 - a^2})_v \cdot b - U^{\pm} \cdot (b^2)_v \right) \ . \tag{A.13}$$

Along the edge of the medial axis, the expression $\sqrt{b^2 - a^2}$ goes to zero, so these derivatives do not exist. This is the reason for the introduction of S_E, the edge shape operator.

Now using $\mathbf{v} = \{m_u, m_v\}$ as a basis for $T_{m_0} M$ allows the computation of a matrix representation of S_{rad}. We also denote the ith vector of the basis as v_i and the matrix representation of S_{rad} as $S^{\pm}_{\mathbf{v}}$. The superscript emphasizes that it is multivalued. This choice of basis means that $\frac{\partial U^{\pm}}{\partial v_1} = U^{\pm}_u$ and $\frac{\partial U^{\pm}}{\partial v_2} = U^{\pm}_v$. For each v_i, the coefficients of

$S_{\mathbf{v}}^{\pm}$ must satisfy the following equation (Damon, 2003):

$$\frac{\partial U^{\pm}}{\partial v_i} = t_i^{\pm} \cdot U^{\pm} - \sum_{j=1}^{n} s_{ji}^{\pm} v_j \ . \tag{A.14}$$

The $t_i^{\pm}$ coefficients measure the amount $\frac{\partial U^{\pm}}{\partial v_i}$ is displaced in the $U^{\pm}$ direction by the proj_U operator in (2.6), while $s_{ji}^{\pm}$ are the coefficients of $S_{\mathbf{v}}^{\pm}$. In three dimensions this gives rise to the following linear system:

$$\begin{bmatrix} U^{\pm} & m_u & m_v \end{bmatrix} \cdot \begin{bmatrix} t_i^{\pm} & -s_{1i}^{\pm} & -s_{2i}^{\pm} \end{bmatrix}^T = \frac{\partial U^{\pm}}{\partial v_i} \ , \tag{A.15}$$

whose solutions are

$$\begin{bmatrix} t_i^{\pm} & -s_{1i}^{\pm} & -s_{2i}^{\pm} \end{bmatrix}^T = \begin{bmatrix} U^{\pm} & m_u & m_v \end{bmatrix}^{-1} \cdot \frac{\partial U^{\pm}}{\partial v_i} \ , \tag{A.16}$$

However, there is no need to explicitly compute the inverse, especially as the $t_i^{\pm}$ coefficients are not needed. Instead, a simple solution is given by applying Cramer's rule, which allows us to write $S_{\mathbf{v}}^{\pm}$ as

$$S_{\mathbf{v}}^{\pm} = \frac{-1}{\det \begin{pmatrix} U^{\pm} & m_u & m_v \end{pmatrix}} \begin{bmatrix} \det \begin{pmatrix} U^{\pm} & U_u^{\pm} & m_v \end{pmatrix} & \det \begin{pmatrix} U^{\pm} & U_v^{\pm} & m_v \end{pmatrix} \\ \det \begin{pmatrix} U^{\pm} & m_u & U_u^{\pm} \end{pmatrix} & \det \begin{pmatrix} U^{\pm} & m_u & U_v^{\pm} \end{pmatrix} \end{bmatrix} \ . \tag{A.17}$$

The determinants can be evaluated via the scalar triple product: $\det(x \quad y \quad z) = x \cdot (y \times z)$. In particular, note $\det(U^{\pm} \quad m_u \quad m_v) = U^{\pm} \cdot b = \pm\sqrt{b^2 - a^2}$, an expression already used to compute $U^{\pm}$ itself. The two expressions $U^{\pm} \times m_v$ and $m_u \times U^{\pm}$ can be factored out of the remaining triple products to save computation:

$$S_{\mathbf{v}}^{\pm} = \frac{\pm 1}{\sqrt{b^2 - a^2}} \begin{bmatrix} U_u^{\pm} \cdot (U^{\pm} \times m_v) & U_v^{\pm} \cdot (U^{\pm} \times m_v) \\ U_u^{\pm} \cdot (m_u \times U^{\pm}) & U_v^{\pm} \cdot (m_u \times U^{\pm}) \end{bmatrix} \ . \tag{A.18}$$

One might question whether or not it is worth computing S_{rad}, given that we already compute $\mathcal{B}_u^{\pm}$ and $\mathcal{B}_v^{\pm}$ in order to apply the test in (4.6) to decide when to stop refining our sampling. The differential unit of area on the boundary is readily obtainable from their cross product, and as pointed out in (Yushkevich et al., 2003), one can test for overfolding by computing the principal curvatures of the boundary directly. These can be expressed as the eigenvalues of $\mathbf{II}_{\mathcal{B}^{\pm}} \mathbf{I}_{\mathcal{B}^{\pm}}^{-1}$, where $\mathbf{II}_{\mathcal{B}^{\pm}}$ is the *second fundamental form*.

The metric tensor on the boundary $\mathbf{I}_{\mathcal{B}^\pm}$ is defined as in (3.14), with $\mathcal{B}_u^\pm$ and $\mathcal{B}_v^\pm$ in place of m_u and m_v. A matrix representation of the second fundamental form, the traditional shape operator, is given by

$$\mathbf{II}_{\mathcal{B}^\pm} \triangleq \begin{bmatrix} \mathcal{B}_{uu}^\pm \cdot U^\pm & \mathcal{B}_{uv}^\pm \cdot U^\pm \\ \mathcal{B}_{vu}^\pm \cdot U^\pm & \mathcal{B}_{vv}^\pm \cdot U^\pm \end{bmatrix} = - \begin{bmatrix} \mathcal{B}_u^\pm \cdot U_u^\pm & \mathcal{B}_v^\pm \cdot U_u^\pm \\ \mathcal{B}_u^\pm \cdot U_v^\pm & \mathcal{B}_v^\pm \cdot U_v^\pm \end{bmatrix} . \tag{A.19}$$

The simplification on the right is given by the relation

$$\frac{\partial}{\partial u}\left(\mathcal{B}_u^\pm \cdot U^\pm\right) = \mathcal{B}_{uu}^\pm \cdot U^\pm + \mathcal{B}_u^\pm \cdot U_u^\pm = 0 . \tag{A.20}$$

This holds because $\mathcal{B}_u^\pm \cdot U^\pm$ is identically zero everywhere, since $\mathcal{B}_u^\pm$ is orthogonal to $U^\pm$, and hence its derivative is also identically zero. Similar relations hold when v is substituted for one or more occurrences of u.

Careful reuse of expressions makes this approach almost exactly as computationally expensive as computing S_{rad}, however it is not as versatile. By storing just two numbers derived from S_{rad}, namely the H_{rad} and K_{rad} defined in (2.8), one can compute its eigenvalues to check the radial curvature condition, compute the Jacobian of the radial vector flow, and conveniently separate out the contributions from different powers of r in medial integrals.

A.1 Derivatives of the Edge Control Curve

On edge patches, differentiating r requires differentiating the control curve, $r_0(v)$. The surface $m(u, v)$ and the other $r_i(v)$ curves defined in (3.23) are simple polynomials, so their derivatives are easily obtained. In the following, we assume all expressions are evaluated at $(0, v)$—or simply at (v) for the $r_i(v)$ curves—and drop the explicit argument lists to improve readability.

First, derivatives of the metric tensor coefficients are needed:

$$\frac{\partial E_m}{\partial v} = 2m_u \cdot m_{uv} , \qquad \frac{\partial^2 E_m}{\partial v^2} = 2\left(m_u \cdot m_{uvv} + m_{uv} \cdot m_{uv}\right) , \tag{A.21}$$

$$\frac{\partial F_m}{\partial v} = m_u \cdot m_{vv} + m_{uv} \cdot m_v , \qquad \frac{\partial^2 F_m}{\partial v^2} = m_u \cdot m_{vvv} + 2m_{uv} \cdot m_{vv} + m_{uvv} \cdot m_v , \tag{A.22}$$

$$\frac{\partial G_m}{\partial v} = 2m_v \cdot m_{vv} , \qquad \frac{\partial^2 G_m}{\partial v^2} = 2\left(m_v \cdot m_{vvv} + m_{vv} \cdot m_{vv}\right) . \tag{A.23}$$

Three new common sub-expressions follow:

$$c \triangleq F_m \frac{\partial r_1}{\partial v} \ , \qquad d \triangleq G_m - \left(\frac{\partial r_1}{\partial v}\right)^2 \ , \qquad e \triangleq d\left(E_m G_m - F_m^2\right) = d \cdot b^2 \tag{A.24}$$

Using these to rewrite (3.18) produces

$$r_u = \frac{1}{G_m}\left(c + \sqrt{e}\right) \ . \tag{A.25}$$

The fact that the spline for r is interpolative in u allows the replacement of $r(0, v)$ with $r_1(v)$ here. Differentiating the various sub-expressions yields

$$(b^2)_v = E_m \frac{\partial G_m}{\partial v} + \frac{\partial E_m}{\partial v} G_m - 2F_m \frac{\partial F_m}{\partial v} \ , \tag{A.26}$$

$$c_v = F_m \frac{\partial^2 r_1}{\partial v^2} + \frac{\partial F_m}{\partial v} \cdot \frac{\partial r_1}{\partial v} \ , \tag{A.27}$$

$$d_v = \frac{\partial G_m}{\partial v} - 2\frac{\partial r_1}{\partial v} \cdot \frac{\partial^2 r_1}{\partial v^2} \ , \tag{A.28}$$

$$e_v = d \cdot (b^2)_v + d_v \cdot b^2 \ , \qquad (\sqrt{e})_v = \frac{e_v}{2\sqrt{e}} \ . \tag{A.29}$$

Re-deriving $(b^2)_v$ instead of using the expression given in (A.9) allows the reuse of the derivatives of the metric tensor coefficients, saving computation. Putting these together, one obtains

$$r_{uv} = \frac{1}{G_m}\left(c_v + (\sqrt{e})_v - \frac{\partial G_m}{\partial v} r_u\right) \ . \tag{A.30}$$

A second round of differentiation proceeds in much the same way.

$$(b^2)_{vv} = E_m \frac{\partial^2 G_m}{\partial v^2} + \frac{\partial^2 E_m}{\partial v^2} G_m + 2\left(\frac{\partial E_m}{\partial v} \cdot \frac{\partial G_m}{\partial v} - F_m \frac{\partial^2 F_m}{\partial v^2} - \left(\frac{\partial F_m}{\partial v}\right)^2\right) \ , \tag{A.31}$$

$$c_{vv} = F_m \frac{\partial^3 r_1}{\partial v^3} + 2\frac{\partial F_m}{\partial v} \cdot \frac{\partial^2 r_1}{\partial v^2} + \frac{\partial^2 F_m}{\partial v^2} \cdot \frac{\partial r_1}{\partial v} \ , \tag{A.32}$$

$$d_{vv} = \frac{\partial^2 G_m}{\partial v^2} - 2\left(\frac{\partial r_1}{\partial v} \cdot \frac{\partial^3 r_1}{\partial v^3} + \frac{\partial^2 r_1}{\partial v^2} \cdot \frac{\partial^2 r_1}{\partial v^2}\right) \ , \tag{A.33}$$

$$e_{vv} = d \cdot (b^2)_{vv} + 2d_v \cdot (b^2)_v + d_{vv} b^2 \ , \qquad (\sqrt{e})_{vv} = \frac{e_{vv} - 2\left((\sqrt{e})_v\right)^2}{2\sqrt{e}} \ , \tag{A.34}$$

$$r_{uvv} = \frac{1}{G_m}\left(c_{vv} + (\sqrt{e})_{vv} - \frac{\partial^2 G_m}{\partial v^2} r_u - 2\frac{\partial G_m}{\partial v} r_{uv}\right) \ . \tag{A.35}$$

The only expressions that involve third derivatives in v are $\frac{\partial^2 F_m}{\partial v^2}$, $\frac{\partial^2 G_m}{\partial v^2}$, c_{vv}, and d_{vv}.

Finally, plugging these into (3.25) gives the complete derivatives for r_0:

$$\frac{\partial r_0}{\partial v} = \frac{1}{C_{1,0}} \left(r_{uv} - C_{1,1} \frac{\partial r_1}{\partial v} - C_{1,2} \frac{\partial r_2}{\partial v} - C_{1,3} \frac{\partial r_3}{\partial v} \right) , \tag{A.36}$$

$$\frac{\partial^2 r_0}{\partial v^2} = \frac{1}{C_{1,0}} \left(r_{uvv} - C_{1,1} \frac{\partial^2 r_1}{\partial v^2} - C_{1,2} \frac{\partial^2 r_2}{\partial v^2} - C_{1,3} \frac{\partial^2 r_3}{\partial v^2} \right) . \tag{A.37}$$

Replacing r_i in (3.26) with $\frac{\partial r_i}{\partial v}$ and $\frac{\partial^2 r_i}{\partial v^2}$ then yields the corresponding derivatives of r itself. Since r is a polynomial in u, derivatives in that direction are simple to obtain.

Appendix B

Derivatives for Optimizing $U^{\pm}$ and $S_{\mathbf{rad}}$

This appendix describes the derivatives needed by the model fitting process. It assumes that ν is the variable of optimization and that at a given sample location the following derivatives are available: m_ν, r_ν, $m_{u\nu}$, $r_{u\nu}$, $m_{v\nu}$, $r_{v\nu}$, $m_{uu\nu}$, $r_{uu\nu}$, $m_{uv\nu}$, $r_{uv\nu}$, $m_{vv\nu}$, $r_{vv\nu}$. The last six are only necessary if the radial curvature constraint is violated or the objective function for triangle meshes is used. Away from the singular points, these are all linear in the control point derivatives and can be computed in exactly the same manner as m and r, substituting the derivatives $\mathbf{P}_\nu$ for the control point matrix $\mathbf{P}$ in (3.1). Near singular points, one could resort to numeric derivatives, but like before Section B.1 gives exact expressions for edge patches. The remainder of this appendix gives the expressions for the derivatives of the quantities computed in the previous appendix.

The first derivatives computed are for the two common sub-expressions from (A.1):

$$a_\nu = r_u m_{v\nu} + r_{u\nu} m_v - r_v m_{u\nu} - r_{v\nu} m_u \tag{B.1}$$

$$b_\nu = m_u \times m_{v\nu} + m_{u\nu} \times m_v \tag{B.2}$$

These lead to the related derivatives

$$(a^2)_\nu = 2a \cdot a_\nu \ , \qquad (b^2)_\nu = 2b \cdot b_\nu \ , \tag{B.3}$$

$$(b \times a)_\nu = b \times a_\nu + b_\nu \times a \ , \qquad (\sqrt{b^2 - a^2})_\nu = \frac{(b^2)_\nu - (a^2)_\nu}{2\sqrt{b^2 - a^2}} \ . \tag{B.4}$$

These are sufficient to compute the derivatives of the unit spoke vectors. The first two also allow the computation of the derivative of the constraint in (4.41). If this condition

is violated, computation stops here. Otherwise, it continues with

$$U_\nu^\pm = \frac{1}{b^2}\left((b \times a)_\nu \pm \sqrt{b^2 - a^2} \cdot b_\nu \pm (\sqrt{b^2 - a^2})_\nu \cdot b - U^\pm \cdot (b^2)_\nu\right) \ . \tag{B.5}$$

When differentiating the objective function for binary images, as long as the radial curvature constraint is satisfied these are all the derivatives necessary. The derivatives of the associated boundary points are given by

$$\mathcal{B}_\nu^\pm = m_\nu + r_\nu U^\pm + r U_\nu^\pm \ . \tag{B.6}$$

The objective function for triangle meshes requires the derivative of the medial measure ΔM, given by

$$\Delta M_\nu = (\sqrt{b^2 - a^2})_\nu \Delta u \Delta v \tag{B.7}$$

If the radial curvature constraint in (4.43) is violated or triangle meshes are used, derivatives of S_{rad} and its eigenvalues are required. Starting with the second derivatives of the common sub-expressions,

$$\begin{aligned} a_{u\nu} &= r_u m_{uv\nu} + r_{u\nu} m_{uv} + r_{uu} m_{v\nu} + r_{uu\nu} m_v \\ &\quad - r_v m_{uu\nu} - r_{v\nu} m_{uu} - r_{uv} m_{u\nu} - r_{uv\nu} m_u \ , \end{aligned} \tag{B.8}$$

$$\begin{aligned} a_{v\nu} &= r_u m_{vv\nu} + r_{u\nu} m_{vv} + r_{uv} m_{v\nu} + r_{uv\nu} m_v \\ &\quad - r_v m_{uv\nu} - r_{v\nu} m_{uv} - r_{vv} m_{u\nu} - r_{vv\nu} m_u \ , \end{aligned} \tag{B.9}$$

$$(a^2)_{u\nu} = 2(a \cdot a_{u\nu} + a_\nu \cdot a_u) \ , \tag{B.10}$$

$$(a^2)_{v\nu} = 2(a \cdot a_{v\nu} + a_\nu \cdot a_v) \ , \tag{B.11}$$

$$b_{u\nu} = m_u \times m_{uv\nu} + m_{u\nu} \times m_{uv} + m_{uu} \times m_{v\nu} + m_{uu\nu} \times m_v \ , \tag{B.12}$$

$$b_{v\nu} = m_u \times m_{vv\nu} + m_{u\nu} \times m_{vv} + m_{uv} \times m_{v\nu} + m_{uv\nu} \times m_v \ , \tag{B.13}$$

$$(b^2)_{u\nu} = 2(b \cdot b_{u\nu} + b_\nu \cdot b_u) \ , \tag{B.14}$$

$$(b^2)_{v\nu} = 2(b \cdot b_{v\nu} + b_\nu \cdot b_v) \ , \tag{B.15}$$

$$(b \times a)_{u\nu} = b_u \times a_\nu + b_{u\nu} \times a + b \times a_{u\nu} + b_\nu \times a_u \ , \tag{B.16}$$

$$(b \times a)_{v\nu} = b_v \times a_\nu + b_{v\nu} \times a + b \times a_{v\nu} + b_\nu \times a_v \ , \tag{B.17}$$

$$(\sqrt{b^2-a^2})_{u\nu} = \frac{(b^2)_{u\nu} - (a^2)_{u\nu} - 2(\sqrt{b^2-a^2})_u(\sqrt{b^2-a^2})_\nu}{2\sqrt{b^2-a^2}} \ , \tag{B.18}$$

$$(\sqrt{b^2-a^2})_{v\nu} = \frac{(b^2)_{v\nu} - (a^2)_{v\nu} - 2(\sqrt{b^2-a^2})_v(\sqrt{b^2-a^2})_\nu}{2\sqrt{b^2-a^2}} \ . \tag{B.19}$$

This allows the computation of the derivative of the change in the spokes in each direction:

$$\begin{aligned} U^\pm_{u\nu} = \frac{1}{b^2}\Big(&(b\times a)_{u\nu} \pm \sqrt{b^2-a^2}\cdot b_{u\nu} \pm (\sqrt{b^2-a^2})_\nu \cdot b_u \pm (\sqrt{b^2-a^2})_u \cdot b_\nu \\ &\pm(\sqrt{b^2-a^2})_{u\nu}\cdot b - U^\pm\cdot(b^2)_{u\nu} - U^\pm_\nu\cdot(b^2)_u - U^\pm_u\cdot(b^2)_\nu\Big) \ , \end{aligned} \tag{B.20}$$

$$\begin{aligned} U^\pm_{v\nu} = \frac{1}{b^2}\Big(&(b\times a)_{v\nu} \pm \sqrt{b^2-a^2}\cdot b_{v\nu} \pm (\sqrt{b^2-a^2})_\nu \cdot b_v \pm (\sqrt{b^2-a^2})_v \cdot b_\nu \\ &\pm(\sqrt{b^2-a^2})_{v\nu}\cdot b - U^\pm\cdot(b^2)_{v\nu} - U^\pm_\nu\cdot(b^2)_v - U^\pm_v\cdot(b^2)_\nu\Big) \ . \end{aligned} \tag{B.21}$$

These allow the computation of the derivative of the matrix $S^\pm_{\mathbf{v}}$ from those of its coefficients, $s^\pm_{ji}$:

$$\begin{aligned} \frac{\partial s^\pm_{11}}{\partial\nu} = \frac{\pm1}{\sqrt{b^2-a^2}}\Big(&U^\pm_u\cdot(U^\pm\times m_{v\nu} + U^\pm_\nu\times m_v) + \\ &U^\pm_{u\nu}\cdot(U^\pm\times m_v) - s^\pm_{11}(\sqrt{b^2-a^2})_\nu\Big) \ , \end{aligned} \tag{B.22}$$

$$\begin{aligned} \frac{\partial s^\pm_{12}}{\partial\nu} = \frac{\pm1}{\sqrt{b^2-a^2}}\Big(&U^\pm_v\cdot(U^\pm\times m_{v\nu} + U^\pm_\nu\times m_v) + \\ &U^\pm_{v\nu}\cdot(U^\pm\times m_v) - s^\pm_{12}(\sqrt{b^2-a^2})_\nu\Big) \ , \end{aligned} \tag{B.23}$$

$$\begin{aligned} \frac{\partial s^\pm_{21}}{\partial\nu} = \frac{\pm1}{\sqrt{b^2-a^2}}\Big(&U^\pm_u\cdot(m_u\times U^\pm_\nu + m_{u\nu}\times U^\pm) + \\ &U^\pm_{u\nu}\cdot(m_u\times U^\pm) - s^\pm_{21}(\sqrt{b^2-a^2})_\nu\Big) \ , \end{aligned} \tag{B.24}$$

$$\begin{aligned} \frac{\partial s^\pm_{22}}{\partial\nu} = \frac{\pm1}{\sqrt{b^2-a^2}}\Big(&U^\pm_v\cdot(m_u\times U^\pm_\nu + m_{u\nu}\times U^\pm) + \\ &U^\pm_{v\nu}\cdot(m_u\times U^\pm) - s^\pm_{22}(\sqrt{b^2-a^2})_\nu\Big) \ , \end{aligned} \tag{B.25}$$

$$\frac{\partial S^\pm_{\mathbf{v}}}{\partial\nu} = \begin{bmatrix} \frac{\partial s^\pm_{11}}{\partial\nu} & \frac{\partial s^\pm_{12}}{\partial\nu} \\ \frac{\partial s^\pm_{21}}{\partial\nu} & \frac{\partial s^\pm_{22}}{\partial\nu} \end{bmatrix} \ . \tag{B.26}$$

These coefficients give $\frac{\partial \kappa_{ri}}{\partial \nu}$, the derivative of the radial curvatures:

$$\frac{\partial H_{\text{rad}}}{\partial \nu} = \frac{1}{2}\left(\frac{\partial s_{11}^{\pm}}{\partial \nu} + \frac{\partial s_{22}^{\pm}}{\partial \nu}\right) , \tag{B.27}$$

$$\frac{\partial K_{\text{rad}}}{\partial \nu} = s_{11}^{\pm}\frac{\partial s_{22}^{\pm}}{\partial \nu} + \frac{\partial s_{11}^{\pm}}{\partial \nu}s_{22}^{\pm} - s_{12}^{\pm}\frac{\partial s_{21}^{\pm}}{\partial \nu} - \frac{\partial s_{12}^{\pm}}{\partial \nu}s_{21}^{\pm} , \tag{B.28}$$

$$\frac{\partial \kappa_{ri}}{\partial \nu} = \frac{\partial H_{\text{rad}}}{\partial \nu} \pm \frac{1}{\sqrt{{H_{\text{rad}}}^2 - K_{\text{rad}}}}\left(H_{\text{rad}}\frac{\partial H_{\text{rad}}}{\partial \nu} - \frac{1}{2}\cdot\frac{\partial K_{\text{rad}}}{\partial \nu}\right) . \tag{B.29}$$

For triangle meshes, assembling the above expressions yields the derivatives of the weighted normal vectors, η_i':

$$\begin{aligned}\frac{\partial \eta_i}{\partial \nu} &= U_{\nu}^{\pm}\left(1 - 2rH_{\text{rad}} + r^2K_{\text{rad}}\right)\Delta M \\ &\quad + U^{\pm}\left(1 - 2\left(r_{\nu}H_{\text{rad}} + r\frac{\partial H_{\text{rad}}}{\partial \nu}\right) + 2rr_{\nu}K_{\text{rad}} + r^2\frac{\partial K_{\text{rad}}}{\partial \nu}\right)\Delta M \\ &\quad + U^{\pm}\left(1 - 2rH_{\text{rad}} + r^2K_{\text{rad}}\right)\Delta M_{\nu} .\end{aligned} \tag{B.30}$$

This completes the set of derivatives needed to evaluate the objective functions and the associated penalty functions from Chapter 4.

B.1 Derivatives of the Edge Control Curve

Once again, evaluating the derivatives in the preceding section on edge patches requires first differentiating the control curve, $r_0(v)$. This section essentially applies another round of differentiation to the expressions in Section A.1. It assumes that the other derivatives on the patch with respect to ν are already available, namely those of the medial surface m and of the three spline r_i curves, $i \in 1\ldots3$. In the following, all expressions are evaluated at $(0, v)$—or simply at (v) for the $r_i(v)$ curves—allowing the explicit argument lists to be dropped to improve readability.

First, derivatives of the metric tensor coefficients are needed:

$$\frac{\partial E_m}{\partial \nu} = 2m_u \cdot m_{u\nu} , \tag{B.31}$$

$$\frac{\partial F_m}{\partial \nu} = m_u \cdot m_{v\nu} + m_{u\nu} \cdot m_v , \tag{B.32}$$

$$\frac{\partial G_m}{\partial \nu} = 2m_v \cdot m_{v\nu} . \tag{B.33}$$

Then, differentiating the common sub-expressions from Section A.1,

$$(b^2)_\nu = E_m \frac{\partial G_m}{\partial \nu} + \frac{\partial E_m}{\partial \nu} G_m - 2F_m \frac{\partial F_m}{\partial \nu} \ , \tag{B.34}$$

$$c_\nu = F_m \frac{\partial^2 r_1}{\partial v \partial \nu} + \frac{\partial F_m}{\partial \nu} \cdot \frac{\partial r_1}{\partial v} \ , \qquad d_\nu = \frac{\partial G_m}{\partial \nu} - 2\frac{\partial r_1}{\partial v} \cdot \frac{\partial^2 r_1}{\partial v \partial \nu} \ , \tag{B.35}$$

$$e_\nu = d \cdot (b^2)_\nu + d_\nu \cdot b^2 \ , \qquad (\sqrt{e})_\nu = \frac{e_\nu}{2\sqrt{e}} \tag{B.36}$$

gives

$$r_{u\nu} = \frac{1}{G_m} \left(c_\nu + (\sqrt{e})_\nu - \frac{\partial G_m}{\partial \nu} r_u \right) \ , \tag{B.37}$$

$$\frac{\partial r_0}{\partial \nu} = \frac{1}{C_{1,0}} \left(r_{u\nu} - C_{1,1} \frac{\partial r_1}{\partial \nu} - C_{1,2} \frac{\partial r_2}{\partial \nu} - C_{1,3} \frac{\partial r_3}{\partial \nu} \right) \ . \tag{B.38}$$

Repeating the process on the first-order derivatives with respect to v,

$$\frac{\partial^2 E_m}{\partial v \partial \nu} = 2 \left(m_u \cdot m_{uv\nu} + m_{u\nu} \cdot m_{uv} \right) \ , \tag{B.39}$$

$$\frac{\partial^2 F_m}{\partial v \partial \nu} = m_u \cdot m_{vv\nu} + m_{u\nu} \cdot m_{vv} + m_{uv} \cdot m_{v\nu} + m_{uv\nu} \cdot m_v \ , \tag{B.40}$$

$$\frac{\partial^2 G_m}{\partial v \partial \nu} = 2 \left(m_v \cdot m_{vv\nu} + m_{v\nu} \cdot m_{vv} \right) \ , \tag{B.41}$$

$$\begin{aligned} (b^2)_{v\nu} = {} & E_m \frac{\partial^2 G_m}{\partial v \partial \nu} + \frac{\partial E_m}{\partial \nu} \cdot \frac{\partial G_m}{\partial v} + \frac{\partial E_m}{\partial v} \cdot \frac{\partial G_m}{\partial \nu} + \frac{\partial^2 E_m}{\partial v \partial \nu} G_m \\ & - 2F_m \frac{\partial^2 F_m}{\partial v \partial \nu} - 2 \frac{\partial F_m}{\partial \nu} \cdot \frac{\partial F_m}{\partial v} \ , \end{aligned} \tag{B.42}$$

$$c_{v\nu} = F_m \frac{\partial^3 r_1}{\partial v^2 \partial \nu} + \frac{\partial F_m}{\partial \nu} \cdot \frac{\partial^2 r_1}{\partial v^2} + \frac{\partial F_m}{\partial v} \cdot \frac{\partial^2 r_1}{\partial v \partial \nu} + \frac{\partial^2 F_m}{\partial v \partial \nu} \cdot \frac{\partial r_1}{\partial v} \ , \tag{B.43}$$

$$d_{v\nu} = \frac{\partial^2 G_m}{\partial v \partial \nu} - 2 \left(\frac{\partial r_1}{\partial v} \cdot \frac{\partial^3 r_1}{\partial v^2 \partial \nu} + \frac{\partial^2 r_1}{\partial v \partial \nu} \cdot \frac{\partial^2 r_1}{\partial v^2} \right) \ , \tag{B.44}$$

$$e_{v\nu} = d \cdot (b^2)_{v\nu} + d_\nu \cdot (b^2)_v + d_v \cdot (b^2)_\nu + d_{v\nu} \cdot b^2 \ , \tag{B.45}$$

$$(\sqrt{e})_{v\nu} = \frac{1}{\sqrt{e}} \left(e_{v\nu} - 2(\sqrt{e})_v \cdot (\sqrt{e})_\nu \right) \ , \tag{B.46}$$

$$r_{uv\nu} = \frac{1}{G_m} \left(c_{v\nu} + (\sqrt{e})_{v\nu} - \frac{\partial^2 G_m}{\partial v \partial \nu} r_u - \frac{\partial G_m}{\partial v} r_{u\nu} - \frac{\partial G_m}{\partial \nu} r_{uv} \right) \ , \tag{B.47}$$

$$\frac{\partial^2 r_0}{\partial v \partial \nu} = \frac{1}{C_{1,0}} \left(r_{uv\nu} - C_{1,1} \frac{\partial^2 r_1}{\partial v \partial \nu} - C_{1,2} \frac{\partial^2 r_2}{\partial v \partial \nu} - C_{1,3} \frac{\partial^2 r_3}{\partial v \partial \nu} \right) \ . \tag{B.48}$$

And once more for the second-order v derivatives:

$$\frac{\partial^3 E_m}{\partial v^2 \partial \nu} = 2\left(m_u \cdot m_{uvv\nu} + m_{u\nu} \cdot m_{uvv} + 2m_{uv} \cdot m_{uv\nu}\right) \ , \tag{B.49}$$

$$\begin{aligned}\frac{\partial^3 F_m}{\partial v^2 \partial \nu} &= m_u \cdot m_{vvv\nu} + m_{u\nu} \cdot m_{vvv} + 2(m_{uv} \cdot m_{vv\nu} + m_{uv\nu} \cdot m_{vv}) \\ &\quad + m_{uvv} \cdot m_{v\nu} + m_{uvv\nu} \cdot m_v \ ,\end{aligned} \tag{B.50}$$

$$\frac{\partial^3 G_m}{\partial v^2 \partial \nu} = 2\left(m_v \cdot m_{vvv\nu} + m_{v\nu} \cdot m_{vvv} + 2m_{vv} \cdot m_{vv\nu}\right) \ , \tag{B.51}$$

$$\begin{aligned}(b^2)_{vv\nu} &= E_m \frac{\partial^3 G_m}{\partial v^2 \partial \nu} + \frac{\partial E_m}{\partial \nu} \cdot \frac{\partial^2 G_m}{\partial v^2} + \frac{\partial^2 E_m}{\partial v^2} \cdot \frac{\partial G_m}{\partial \nu} + \frac{\partial^3 E_m}{\partial v^2 \partial \nu} G_m \\ &\quad + 2\left(\frac{\partial E_m}{\partial v} \cdot \frac{\partial^2 G_m}{\partial v \partial \nu} + \frac{\partial^2 E_m}{\partial v \partial \nu} \cdot \frac{\partial G_m}{\partial v}\right. \\ &\qquad \left. - F_m \frac{\partial^3 F_m}{\partial v^2 \partial \nu} - \frac{\partial F_m}{\partial \nu} \cdot \frac{\partial^2 F_m}{\partial v^2} - 2\frac{\partial F_m}{\partial v} \cdot \frac{\partial^2 F_m}{\partial v \partial \nu}\right) \ ,\end{aligned} \tag{B.52}$$

$$\begin{aligned}c_{vv\nu} &= F_m \frac{\partial^4 r_1}{\partial v^3 \partial \nu} + \frac{\partial F_m}{\partial \nu} \cdot \frac{\partial^3 r_1}{\partial v^3} + 2\left(\frac{\partial F_m}{\partial v} \cdot \frac{\partial^3 r_1}{\partial v^2 \partial \nu} + \frac{\partial^2 F_m}{\partial v \partial \nu} \cdot \frac{\partial^2 r_1}{\partial v^2}\right) \\ &\quad + \frac{\partial^2 F_m}{\partial v^2} \cdot \frac{\partial^2 r_1}{\partial v \partial \nu} + \frac{\partial^3 F_m}{\partial v^2 \partial \nu} \cdot \frac{\partial r_1}{\partial v} \ ,\end{aligned} \tag{B.53}$$

$$\begin{aligned}d_{vv\nu} &= \frac{\partial^3 G_m}{\partial v^2 \partial \nu} - 2\left(\frac{\partial r_1}{\partial v} \cdot \frac{\partial^4 r_1}{\partial v^3 \partial \nu} + \frac{\partial^2 r_1}{\partial v \partial \nu} \cdot \frac{\partial^3 r_1}{\partial v^3}\right. \\ &\quad \left. + \frac{\partial^2 r_1}{\partial v^2} \cdot \frac{\partial^3 r_1}{\partial v^2 \partial \nu} + \frac{\partial^3 r_1}{\partial v^2 \partial \nu} \cdot \frac{\partial^2 r_1}{\partial v^2}\right) \ ,\end{aligned} \tag{B.54}$$

$$\begin{aligned}e_{vv\nu} &= d \cdot (b^2)_{vv\nu} + d_\nu \cdot (b^2)_{vv} + 2\left(d_v \cdot (b^2)_{v\nu} + d_{v\nu} \cdot (b^2)_v\right) \\ &\quad + d_{vv} \cdot (b^2)_\nu + d_{vv\nu} b^2 \ ,\end{aligned} \tag{B.55}$$

$$(\sqrt{e})_{vv\nu} = \frac{1}{2\sqrt{e}}\left(e_{vv\nu} - 2\left((\sqrt{e})_{vv}(\sqrt{e})_\nu + 2(\sqrt{e})_v(\sqrt{e})_{v\nu}\right)\right) \ , \tag{B.56}$$

$$\begin{aligned}r_{uvv\nu} &= \frac{1}{G_m}\left(c_{vv\nu} + (\sqrt{e})_{vv\nu} - \frac{\partial^2 G_m}{\partial v^2} r_{u\nu} - \frac{\partial^3 G_m}{\partial v^2 \partial \nu} r_u\right. \\ &\quad \left. - 2\left(\frac{\partial G_m}{\partial v} r_{uv\nu} + \frac{\partial^2 G_m}{\partial v \partial \nu} r_{uv}\right) - \frac{\partial G_m}{\partial \nu} r_{uvv}\right) \ ,\end{aligned} \tag{B.57}$$

$$\frac{\partial^3 r_0}{\partial v^2 \partial \nu} = \frac{1}{C_{1,0}}\left(r_{uvv\nu} - C_{1,1}\frac{\partial^3 r_1}{\partial v^2 \partial \nu} - C_{1,2}\frac{\partial^3 r_2}{\partial v^2 \partial \nu} - C_{1,3}\frac{\partial^3 r_3}{\partial v^2 \partial \nu}\right) \ . \tag{B.58}$$

Although these expressions appear expensive to evaluate, this only needs to be done once for each unique value of v on a patch. Their values are cached and re-used for all subsequent samples at the same v coordinate on the same patch. Also, at least for the binary image objective function, the second-order v derivatives are only required if one of the samples violates the radial curvature constraint.

Bibliography

Amenta, A. B., Choi, S., and Kolluri, R. K. (2001). The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2–3):127–153.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc., New York.

Arsigny, V., Commonwick, O., Pennec, X., and Ayache, N. (2006). A log-Euclidean framework for statistics on diffeomorphisms. In *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)*, volume 4190 of *Lecture Notes in Computer Science*, pages 924–931, Montreal, Canada. Springer-Verlag.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 95(2):115–147.

Biermann, H., Levin, A., and Zorin, D. (2000). Peicewise smooth subdivision surfaces with normal control. In *Proceedings of the 27th Annual SIGGRAPH Confererence on Computer Graphics and Interactive Techniques*, pages 113–120.

Blair, R. C., Higgins, J. J., Karniski, W., and Kromrey, J. D. (1994). A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. *Multivariate Behavioral Research*, 29:141–164.

Blu, T., Thévenaz, P., and Unser, M. (2003). Complete parameterization of piecewise-polynomial interpolation kernels. *IEEE Transactions on Image Processing*, 12(11):1297–1309.

Blum, H. (1967). A transformation for extracting new descriptors of shape. In Wathen-Dunn, W., editor, *Models for the Perception of Speech and Visual Form*, pages 362–380, Cambridge, MA. MIT Press.

Blum, H. (1973). Biological shape and visual science (part I). *Journal of Theoretical Biology*, 32(2):205–287.

Blum, H. and Nagle, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, 10(3):167–180.

Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis (MEDIA)*, 1(3):225–243.

Borgefors, G., Nyström, I., and di Baja, G. S. (1999). Computing skeletons in three dimensions. *Pattern Recognition*, 32(7):1225–1236.

Bouix, S., Pruessner, J. C., Collins, D. L., and Siddiqi, K. (2005). Hippocampal shape analysis using medial surfaces. *NeuroImage*, 25(4):1077–1089.

Bradley, J. V. (1968). *Distribution-Free Statistical Tests*. Prentice Hall, Englewood Cliffs, New Jersey.

Brady, M. and Asada, H. (1984). Smoothed local symmetries and their implementation. Technical Report A. I. Memo 757, MIT Artificial Intelligence Laboratory.

Brechbühler, C., Gerig, G., and Kübler, O. (1995). Parameterization of closed surfaces for 3-D shape description. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 61(2):195–170.

Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30.

Burbeck, C. A., Pizer, S. M., Morse, B. S., Ariely, D., Zauberman, G. S., and Rolland, J. P. (1996). Linking object boundaries at scale: a common mechanism for size and shape judgements. *Vision Research*, 36(3):361–372.

Calabi, L. and Hartnett, W. E. (1968). Shape recognition, prairie fires, convex deficiencies and skeletons. *American Mathematical Monthly*, 75(4):335–342.

Cates, J., Meyer, M., Fletcher, P. T., and Whitaker, R. T. (2006). Entropy-based particle systems for shape correspondence. In *Proceedings of the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA'06)*, pages 90–99, Copenhagen, Denmark.

Catmull, E. E. and Clark, J. H. (1978). Recursively generated B-spline surfaces on arbitrary topological meshes. *Computer Aided Design*, 10(6):350–355.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1993). A deformable neuroanatomy textbook based on viscous fluid mechanics. In *Proceedings of the 27th Conference on Information Sciences and Systems (CISS'93)*, pages 211–216, Baltimore, Maryland.

Chung, J. H. and Fraser, D. A. S. (1958). Randomization tests for a multivariate two-sample problem. *Journal of the American Statistical Association*, 53(283):729–735.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1992). Training models of shape from sets of examples. In *Proceedings of the 3rd British Machine Vision Conference*, pages 9–18, Leeds, United Kingdom.

Czekanowski, J. (1932). "Coefficient of racial likeness" und "durchschnittliche difference". *Anthropologischer Anzeiger*, 9:227–249.

Damon, J. N. (2003). Smoothness and geometry of boundaries associated to skeletal structures I: Sufficient conditions for smoothness. *Annales de l'Institut Fourier*, 53(6):1941–1985.

Damon, J. N. (2004). Smoothness and geometry of boundaries associated to skeletal structures II: Geometry in the Blum case. *Compositio Mathematica*, 140(6):1657–1674.

Damon, J. N. (2005a). Global geometry of regions and boundaries via skeletal and medial integrals. Preprint: http://www.math.unc.edu/Faculty/jndamon/Skel.Str.IV.v5.pdf.

Damon, J. N. (2005b). Tree structure for contractible regions in $\mathbb{R}^3$. To appear in the International Journal of Computer Vision: http://www.math.unc.edu/Faculty/jndamon/TreeStr.Contr.Regn.r.pdf.

Damon, J. N. (2006). Swept regions and surfaces: Modeling and volumetric properties. Preprint: http://www.math.unc.edu/Faculty/jndamon/Swept.Surf.v1.pdf.

Daumé, III, H. (2004). From zero to reproducing kernel Hilbert spaces in twelve pages or less. Online: http://pub.hal3.name/daume04rkhs.ps.

Davies, R. H., Twining, C. J., Cootes, T. F., Waterton, J. C., and Taylor, C. J. (2002). A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537.

Davis, B., Lorenzen, P., and Joshi, S. C. (2004). Large deformation minimum mean squared error template estimation for computational anatomy. In *Proceedings of the International Symposium on Biomedical Imaging (ISBI'04)*, pages 173–176, Arlington, Virginia.

DeRose, T., Kass, M., and Truong, T. (1998). Subdivision surfaces in character animation. In *Proceedings of the 25th Annual SIGGRAPH Confererence on Computer Graphics and Interactive Techniques*, pages 85–94.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Federer, H. and Fleming, W. H. (1960). Normal and integral currents. *The Annals of Mathematics, 2nd Series*, 72(3):458–520.

Fletcher, P. T. (2004). *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. PhD thesis, University of North Carolina at Chapel Hill.

Fletcher, P. T., Lu, C., and Joshi, S. C. (2003). Statistics of shape via principal component analysis on Lie groups. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 95–101, Madison, Wisconsin.

Fletcher, P. T., Lu, C., Pizer, S. M., and Joshi, S. C. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23:995–1005.

Fletcher, P. T. and Whitaker, R. T. (2006). Riemannian metrics on the space of solid shape. In *Proceedings of the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA'06)*, pages 47–57, Copenhagen, Denmark.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de L'Institut Henri Poincaré*, 10:215–310.

Gelston, S. M. and Dutta, D. (1995). Boundary surface recovery from skeleton curves and surfaces. *Computer Aided Geometric Design*, 12(1):27–51.

Gerig, G., Styner, M. A., Jones, D. W., Weinberger, D. R., and Lieberman, J. A. (2001). Shape analysis of brain ventricles using SPHARM. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01)*, pages 171–178.

Giblin, P. J. and Brassett, S. A. (1985). Local symmetry of plane curves. *American Mathematical Monthly*, 92(10):689–707.

Giblin, P. J. and Kimia, B. B. (2004). A formal classification of 3D medial axis points and their local geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):238–251.

Glaunès, J. and Joshi, S. C. (2006). Template estimation from unlabeled point set data and surfaces for computational anatomy. In *Proceedings of the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA'06)*, pages 29–39, Copenhagen, Denmark.

Glaunès, J., Trouvé, A., and Younes, L. (2004). Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 2, pages 712–718, Villetaneuse, France.

Golland, P. and Grimson, W. E. L. (2000). Fixed topology skeletons. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, volume 1, pages 10–17, Head Island, South Carolina.

Golland, P., Grimson, W. E. L., and Kikinis, R. (1999). Statistical shape analysis using fixed topology skeletons: Corpus callosum study. In *Proceedings of the 16th International Conference on Information Processing in Medical Imaging (IPMI'99)*, volume 1613 of *Lecture Notes in Computer Science*, pages 382–387, Visegrád, Hungary. Springer-Verlag.

Golland, P., Grimson, W. E. L., Shenton, M. E., and Kikinis, R. (2002). Small sample size learning for shape analysis of anatomical structures. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted*

Intervention (MICCAI'00), volume 1935 of *Lecture Notes in Computer Science*, pages 72–82, Pittsburgh, Pennsylvania. Springer-Verlag.

Gottschalk, S., Lin, M. C., and Manocha, D. (1996). OBB-tree: A hierarchical structure for rapid interference detection. In *Proceedings of the 23rd Annual SIGGRAPH Confererence on Computer Graphics and Interactive Techniques*, pages 171–180.

Greengard, L. F. and Strain, J. (1991). The fast Gauss transform. *SIAM Journal on Scientific and Statistical Computing*, 12(1):79–94.

Grenander, U. and Keenan, D. M. (1993). Towards automated image understanding. *Advances in Applied Statistics: Statistics and Images*, 1:89–103.

Gunnells, P. E. (2003). Stratified spaces TWIGS. Online lecture notes: http://www.math.umass.edu/~gunnells/talks/singspc.ps.

Hall, P. and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, 89(2):359–374.

Halstead, M., Kass, M., and DeRose, T. (1993). Efficient, fair interpolation using Catmull-Clark surfaces. In *Proceedings of the 20th Annual SIGGRAPH Conference on Computer Graphics and Interactive Techniques*, pages 35–44, Anaheim, California.

Han, Q. (2006). Personal communication.

Han, Q., Lu, C., Liu, G., Pizer, S. M., and Thall, A. (2004). Representing multi-figure anatomical objects. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'04)*, pages 1251–1254, Arlington, Virginia.

Han, Q., Pizer, S. M., and Damon, J. N. (2006). Interpolation in discrete single figure medial objects. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'06)*, New York City, New York.

Heimann, T., Wolf, I., Williams, T., and Meinzer, H.-P. (2005). 3D active shape models using gradient descent optimization of description length. In *Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 566–577, Glenwood Springs, Colorado. Springer-Verlag.

Hoffman, C. M. and Vermeer, P. J. (1994). Validity determination for MAT surface representation. In *Proceedings of the 6th IMA Conference on the Mathematics of Surfaces*, pages 249–265, London, United Kingdom.

Hoppe, H., DeRose, T., Duchamp, T., Halstead, M., Jin, H., McDonald, J., Schweitzer, J., and Stuetzle, W. (1994). Piecewise smooth surface reconstruction. Technical Report TR 94-01-01, University of Washington Department of Computer Science and Engineering.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Societé Vandoise des Sciences Naturelles*, 37:547–579.

Joshi, S. C., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage; Supplemental issue on Mathematics in Brain Imaging*, 23(Supplement 1):S151–S160.

Joshi, S. C., Pizer, S. M., Fletcher, P. T., Thall, A., and Tracton, G. S. (2001). Multi-scale 3-D deformable model segmentation based on medial description. In *Proceedings of the 17th International Conference on Information Processing in Medical Imaging (IPMI'01)*, volume 2082 of *Lecture Notes in Computer Science*, pages 64–77, Davis, California. Springer-Verlag.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Math*, 30:509–541.

Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. In *Proceedings of the London Mathematical Society*, volume 61, pages 371–406.

Keys, R. G. (1981). Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160.

Kimia, B. B., Tannenbaum, A., and Zucker, S. W. (1990). Toward a computational theory of shape: an overview. In Faugeras, O. D., editor, *Proceedings of the First European Conference on Computer Vision (ECCV'90)*, volume 427 of *Lecture Notes in Computer Science*, pages 402–407. Springer-Verlag.

Kostrowicki, J. and Piela, L. (1991). Diffusion equation method of global minimization: Performance for standard test functions. *Journal of Optimization Theory and Applications*, 69(2):269–284.

Kotcheff, A. C. W. and Taylor, C. J. (1998). Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis (MEDIA)*, 2(4):303–314.

Lee, D., Gray, A., and Moore, A. (2005). Dual-tree fast Gauss transforms. In Weiss, Y., Schölkopf, B., and Platt, J., editors, *Advances in Neural Information Processing Systems (NIPS'05)*, volume 18, Vancouver, Canada.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

Leventon, M. E., Grimson, W. E. L., and Faugeras, O. (2000). Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, volume 1, pages 316–323, Head Island, South Carolina.

Leymarie, F. F. and Kimia, B. B. (2003). Computation of the shock scaffold for unorganized point clouds in 3D. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, volume 1, pages 821–827, Madison, Wisconsin.

Leyton, M. (1987). Symmetry-curvature duality. *Computer Vision, Graphics, and Image Processing*, 38(3):327–341.

Lindeberg, T. (1997). *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision*, chapter On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance. Kluwer Academic Publishers, Dordrecht. (Copenhagen, Denmark, May 1996).

Loop, C. (2004). Second order smoothness over extraordinary vertices. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pages 165–174, Nice, France.

Lorenzen, P. J., Prastawa, M., Davis, B. C., Gerig, G., Bullitt, E., and Joshi, S. C. (2006). Multi-modal image set registration and atlas formation. *Medical Image Analysis (MEDIA)*, 10(3):440–451.

Lourakis, M. I. A. and Argyros, A. A. (2005). Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In *Proceedings of the 10th International Conference on Computer Vision (ICCV'05)*, volume 2, pages 1526–1531, Beijing, China.

Machado, A. M. C. and Gee, J. C. (1998). Atlas warping for brain morphometry. In *Medical Imaging 1998: Image Processing*, volume 3338 of *Proceedings of the SPIE*, pages 642–651.

Malandain, G., Ayache, N., and Bertrand, G. (1991). Topological segmentation of discrete surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 444–449, Maui, Hawaii.

Manzanera, A., Bernard, T. M., Prêteux, F., and Longuet, B. (1999). Medial faces from a concise 3D thinning algorithm. In *Proceedings of the 7th International Conference on Computer Vision (ICCV'99)*, pages 337–343, Kerkyra, Greece.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Marr, D. C. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B, Biological Sciences*, 200(1140):269–294.

Marron, J. S. and Todd, M. J. (2002). Distance weighted discrimination. Technical Report 1339, School of Operations Research and Industrial Engineering, Cornell University.

Martin, J., Pentland, A., and Kikinis, R. (1994). Shape analysis of brain structures using physical and experimental modes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 752–755, Seattle, Washington.

Mather, J. (1970). Notes on topological stability. Online lecture notes: http://www.math.princeton.edu/facultypapers/mather/notes_on_topological_stability.pdf.

Mauch, S. (2000). A fast algorithm for computing the closest point and distance transform. Technical Report ASCI/2000.077, California Institute of Technology.

Maurer, Jr., C. R., Qi, R., and Raghavan, V. (2003). A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270.

Miller, M. I. and Younes, L. (2001). Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1–2):61–84.

Mirtich, B. (1996). Fast and accurate computation of polyhedral mass properties. *Journal of Graphics Tools*, 1(2):31–50.

Möller, T. and Trumbore, B. (1997). Fast, minimum storage ray/triangle intersection. *Journal of Graphics Tools*, 2(1):21–28.

Moreé, J. J. and Wu, Z. (1997). Global continuation for distance geometry problems. *SIAM Journal of Optimization*, 7(3):814–836.

Nackman, L. R. (1982). *Three-Dimensional Shape Description Using the Symmetric Axis Transform*. PhD thesis, University of North Carolina at Chapel Hill Department of Computer Science.

Näf, M., Kübler, O., Kikinis, R., Shenton, M. E., and Székely, G. (1996). Shape characterization and recognition of 3D organ shape in medical image analysis using skeletonization. In *Proceedings of the IEEE/SIAM Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'96)*, pages 139–150, San Francisco, California.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15:1–25.

Pallini, A. and Pesarin, F. (1992). A class of combinations of dependent tests by a resampling procedure. In Jöckel, K. H., Rothe, G., and Sendler, W., editors, *Bootstrapping and Related Techniques*, volume 376 of *Lecture Notes in Economics and Mathematical Systems*, pages 93–97, Berlin. Springer-Verlag.

Pantazis, D., Leahy, R. M., Nichols, T. E., and Styner, M. A. (2004). Statistical surface-based morphometry using a non-parametric approach. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'04)*, pages 1283–1286, Arlington, Virginia.

Pennec, X. (1999). Probabilities and statistics on Riemmanian manifolds: Basic tools for geometric measurements. In Cetin, A., Akarun, L., Ertuzun, A., Gurcan, M. N., and Yardimci, Y., editors, *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99)*, pages 194–198.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. John Wiley & Sons, Ltd., Chirchester.

Pizer, S. M., Eberly, D., Morse, B. S., and Fritsch, D. S. (1998). Zoom-invariant vision of figural shape: The mathematics of cores. *Computer Vision and Image Understanding*, 69(1):55–71.

Pizer, S. M., Fritsch, D. S., Yushkevich, P. A., Johnson, V. E., and Chaney, E. L. (1999). Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18:851–865.

Pizer, S. M., Siddiqi, K., Székely, G., Damon, J. N., and Zucker, S. W. (2003). Multiscale medial loci and their properties. *International Journal of Computer Vision*, 55(2–3):155–179.

Powell, M. J. D. (1970). *Numerical Methods for Nonlinear Algebraic Equations*, chapter A hybrid method for nonlinear equations, pages 87–161. Gordon and Breach Science, London.

Prautzsch, H. and Umlauf, G. (1998). A G^2-subdivision algorithm. In Farin, G. E., Bieri, H., Brunnett, G., and DeRose, T., editors, *Proceedings of the 1996 Dagstuhl Workshop on Geometric Modelling*, volume 13 of *Computing Supplement*, pages 217–224, Wadern, Germany. Springer-Verlag.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York City, New York, second edition.

Psotka, J. (1978). Perceptual processes that may create tick figures and balance. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1):101–111.

Reif, U. and Schröder, P. (2001). Curvature integrability of subdivision surfaces. *Advances in Computational Mathematics*, 14(2):157–174.

Rogers, D. J. and Tanimoto, T. T. (1960). A computer program for classifying plants. *Science*, 132(3434):1115–1118.

Sheehy, D. J., Armstrong, C. G., and Robinson, D. J. (1996). Shape description by medial surface construction. *IEEE Transactions on Visualization & Computer Graphics*, 2(1):62–72.

Shenton, M. E., Gerig, G., McCarley, R. W., Székely, G., and Kikinis, R. (2002). Amygdala-hippocampal shape differences in schizophrenia: the application of 3D shape models to volumetric MR data. *Psychiatry Research: Neuroimaging*, 115(1–2):15–35.

Sherbrooke, E. C., Patrikalakis, N. M., and Brisson, E. (1996). An algorithm for the medial axis transform of 3D polyhedral solids. *IEEE Transactions on Visualization & Computer Graphics*, 2(1):46–61.

Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Technical Report CS-94-125, School of Computer Science, Carnegie Mellon University.

Shokoufandeh, A., Macrini, D., Dickinson, S., Siddiqi, K., and Zucker, S. W. (2005). Indexing hierarchical structures using graph spectra. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7):1125–1140.

Siddiqi, K., Bouix, S., Tannenbaum, A., and Zucker, S. W. (2002). Hamilton-Jacobi skeletons. *International Journal of Computer Vision*, 48(3):215–231.

Škrinjar, O. and Tagare, H. (2004). Symmetric, transitive, geometric deformation and intensity variation invariant nonrigid image registration. In *Proceedings of the International Symposium on Biomedical Imaging (ISBI'04)*, pages 920–923, Arlington, Virginia.

Sorenson, T. A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Det Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter*, 5(4):1–34.

Stam, J. (1998). Exact evaluation of Catmull-Clark subdivision surfaces at arbitrary parameter values. *Computer Graphics*, 32:395–404.

Stam, J. (1999). Exact evaluation of Catmull-Clark subdivision surfaces at arbitrary parameter values. In Zorin, D. and Schröder, P., editors, *Lecture Notes for the SIGGRAPH'99 Course on Subdivision for Modeling and Animation*, pages 89–110.

Styner, M. A., Gerig, G., Joshi, S. C., and Pizer, S. M. (2003). Automatic and robust computation of 3D medial models incorporating object variability. *International Journal of Computer Vision*, 55(2–3):107–122.

Styner, M. A., Gorczowski, K., Fletcher, P. T., Jeong, J. Y., Pizer, S. M., and Gerig, G. (2006). Statistics of pose and shape in multi-object complexes using prinipal geodesic analysis. In *Proceedings of the Third International Workshop on Medical Imaging and Augmented Reality (MIAR'06)*, volume 4091 of *Lecture Notes in Computer Science*, pages 1–8, Shanghai, China. Springer-Verlag.

Styner, M. A., Lieberman, J. A., McClure, R. K., Weinberger, D. R., Jones, D. W., and Gerig, G. (2005). Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. *Proceedings of the National Academy of Science*, 102(12):4872–4877.

Székely, G., Näf, M., Brechbuhler, C., and Kübler, O. (1994). Calculating 3D Voronoi diagrams of large unrestricted point sets for skeleton generation of complex 3D shapes. In *Proceedings of the 2nd International Workshop on Visual Form*, pages 532–541.

Teixeira, R. C. (1998). *Curvature motions, medial axes and distance transforms*. PhD thesis, Harvard University Department of Mathematics.

Terriberry, T. B. and Gerig, G. (2006). A continuous 3-D medial shape model with branching. In *Proceedings of the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA'06)*, pages 80–89, Copenhagen, Denmark.

Terriberry, T. B., Joshi, S. C., and Gerig, G. (2005). Hypothesis testing with nonlinear shape models. In Christensen, G. E. and Sonka, M., editors, *Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 15–26, Glenwood Springs, Colorado. Springer-Verlag.

Thall, A. (2004). *Deformable Solid Modeling via Medial Sampling and Displacement Subdivision*. PhD thesis, University of North Carolina at Chapel Hill.

Thodberg, H. H. (2003). Minimum description length shape and appearance models. In *Proceedings of the 18th International Conference on Information Processing in Medical Imaging (IPMI'03)*, volume 2732 of *Lecture Notes in Computer Science*, pages 51–62, Ambleside, United Kingdom. Springer-Verlag.

Thompson, D. W. (1917). *On Growth and Form*, chapter XVII. On the Theory of Transformations, or the Comparison of Related Forms, pages 1026–1095. Dover Publications, complete revised (June 1992) edition.

Torsello, A. and Hancock, E. R. (2006). Correcting curvature-density effects in the Hamilton-Jacobi skeleton. *IEEE Transactions on Image Processing*, 15(4):877–891.

Vaillant, M. and Glaunès, J. (2005). Surface matching via currents. In Christensen, G. E. and Sonka, M., editors, *Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 381–392, Glenwood Springs, Colorado. Springer-Verlag.

van Leemput, K., Maes, F., Vandermeulen, D., and Seutens, P. (1999). Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908.

Vermeer, P. J. (1994). *Medial Axis Transform to Boundary Representation Conversion*. PhD thesis, Purdue University Department of Computer Science.

Warren, J. and Schaefer, S. (2004). A factored approach to subdivision surfaces. *IEEE Computer Graphics and Applications*, 24(3):74–81.

Weinberger, D. R., Egan, M. F., Bertolino, A., Callicott, J. H., Mattay, V. S., Lipska, B. K., Berman, K. F., and Goldberg, T. E. (2001). Prefrontal neurons and the genetics of schizophrenia. *Biological Psychiatry*, 50:825–844.

Xu, S. (2006). MeshValmet. Available from http://www.cs.unc.edu/~xushun/research/MeshValmet.html.

Yang, C., Duraiswami, R., Gumerov, N. A., and Davis, L. (2003). Improved fast Gauss transform and efficient kernel density estimation. In *Proceedings of the 9th International Conference on Computer Vision (ICCV'03)*, pages 464–471, Nice, France.

Yushkevich, P. A. (2003). *Statistical shape characterization using the medial representation*. PhD thesis, University of North Carolina at Chapel Hill Department of Computer Science.

Yushkevich, P. A., Fletcher, P. T., Joshi, S. C., Thall, A., and Pizer, S. M. (2003). Continuous medial representations for geometric object modeling in 2D and 3D. *Image and Vision Computing*, 21(1):17–28.

Yushkevich, P. A., Zhang, H., and Gee, J. C. (2005). Parametric medial shape representation in 3-D via the Poisson partial differential equation with non-linear boundary conditions. In Christensen, G. E. and Sonka, M., editors, *Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 162–173, Glenwood Springs, Colorado. Springer-Verlag.

Zhang, J., Siddiqi, K., Macrini, D., Shokoufandeh, A., and Dickinson, S. (2005). Retrieving articulated 3-D models using medial surfaces and their graph spectra. In *Proceedings of the 5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'05)*, St. Augustine, Florida.