



Taylor & Francis
Taylor & Francis Group



Bayesian Multiscale Modeling of Closed Curves in Point Clouds

Author(s): Kelvin Gu, Debdeep Pati and David B. Dunson

Source: *Journal of the American Statistical Association*, December 2014, Vol. 109, No. 508 (December 2014), pp. 1481-1494

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <https://www.jstor.org/stable/24247386>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Taylor & Francis, Ltd. and American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

Bayesian Multiscale Modeling of Closed Curves in Point Clouds

Kelvin GU, Debdeep PATI, and David B. DUNSON

Modeling object boundaries based on image or point cloud data is frequently necessary in medical and scientific applications ranging from detecting tumor contours for targeted radiation therapy, to the classification of organisms based on their structural information. In low-contrast images or sparse and noisy point clouds, there is often insufficient data to recover local segments of the boundary in isolation. Thus, it becomes critical to model the entire boundary in the form of a closed curve. To achieve this, we develop a Bayesian hierarchical model that expresses highly diverse 2D objects in the form of closed curves. The model is based on a novel multiscale deformation process. By relating multiple objects through a hierarchical formulation, we can successfully recover missing boundaries by borrowing structural information from similar objects at the appropriate scale. Furthermore, the model's latent parameters help interpret the population, indicating dimensions of significant structural variability and also specifying a "central curve" that summarizes the collection. Theoretical properties of our prior are studied in specific cases and efficient Markov chain Monte Carlo methods are developed, evaluated through simulation examples and applied to panorex teeth images for modeling teeth contours and also to a brain tumor contour detection problem.

KEY WORDS: Biomedical imaging; Cyclic basis; Deformation; Functional data; Hierarchical modeling.

1. INTRODUCTION

Boundaries of objects are widely studied across many disciplines, including biomedical imaging, cytology, and computer vision. In describing complex boundaries, one can use a parametric curve (2D) or surface (3D), $C(t) : \mathcal{D}_1 \rightarrow \mathbb{R}^2$ or $C(t) : \mathcal{D}_2 \rightarrow \mathbb{R}^3$, respectively, where $\mathcal{D}_1 \subset \mathbb{R}$ and $\mathcal{D}_2 \subset \mathbb{R}^2$. This is different from a typical function estimation problem because the independent variable, t , is unknown. In addition, the curve must be closed to produce a valid boundary.

A collection of introductory works on curve and surface modeling can be found in Su and Liu (1989) and subsequent developments in Müller (2005). Popular representations include Bézier curves, splines, and principal curves (Hastie and Stuetzle 1989), the final one being a nonlinear generalization of principal components involving smooth curves which pass through the middle of a data cloud. Su et al. (2012) dealt with curve modeling based on stochastic processes when the observations are given as a set of time-indexed points on manifolds. Srivastava et al. (2011) developed an elegant theoretical framework for comparing and analyzing shapes of curves once the fitted curves are obtained. Nonparametric representations of parametric curves and surfaces are widely used (Barnhill 1985; Lang and Röschel 1992; Hagen and Santarelli 1992; Aziz et al. 2002), because they provide a flexible model for a broad range of objects.

Although there is a vast literature on estimating curves and surfaces, the majority of this work focuses on estimating unrestricted functions. However, the boundary of a simply connected object must be a closed curve, which is a restriction on the curve representation. Estimating a closed surface or curve involves a

different modeling strategy and there has been little work in this regime, particularly from a Bayesian perspective. To our knowledge, only Pati and Dunson (2014) developed a Bayesian approach for fitting a single 3D closed surface (instead of multiple 2D closed curves) from point cloud data in \mathbb{R}^3 . Fischler and Bolles (1981) developed the random sample consensus (RANSAC) algorithm that is used for finding simpler shapes (lines, circles, etc.) in point clouds. Active contours (Blake and Isard 1998; contours driven by differential equations, such as the Mumford–Shah equation) and Bayesian active contours (Lievin et al. 1999) can be used to find contours in images directly, as opposed to point clouds generated from images. Several articles focus on finding shapes in cluttered point clouds, with shapes defined using simple closed curves, and using explicit models for both noise and clutter. However, none of these methods borrows information to fit multiple objects. In applications featuring low-contrast images or sparse and noisy point clouds, there is often insufficient data to recover the entire boundary in isolation. Thus, it becomes critical to exploit similarities of multiple related objects.

One strategy for modeling multiple complex curves is to refactor them in a multiscale fashion, as done by Fourier and wavelet descriptors (Whitney 1937; Zahn and Roskies 1972; Mortenson 1985; Persoon and Fu 1977). These approaches decompose a curve into components at different scales, so that the coarsest scale components carry the global approximation information, while the finer scale components contain the local detailed information. Mokhtarian and Mackworth (1992) and Désidéri, Abou El Majd, and Janka (2007) also proposed multiscale curves. Such multiscale transforms make it easier to compare objects that share the same coarse outline, but differ on finer details, or vice versa. The finer scale components can also be discarded to yield a finite and low-dimensional representation. However, none of these methods are model-based and do not allow automatic estimation of uncertainty.

Kelvin Gu, Department of Statistics, Stanford University, Stanford, CA 94305 (E-mail: kgu@stanford.edu). Debdeep Pati, Department of Statistics, Florida State University, Gainesville, FL 32611 (E-mail: debdeep@stat.fsu.edu). David B. Dunson, Department of Statistical Science, Duke University, Durham NC 27708-0251 (E-mail: dunson@duke.edu). Kelvin Gu acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-114747. Dr. Pati acknowledges support from the Office of Naval Research (ONR BAA 14-0001). Dr. Dunson acknowledges support from grant number R01 ES017240-01 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH).

© 2014 American Statistical Association
Journal of the American Statistical Association
December 2014, Vol. 109, No. 508, Theory and Methods
DOI: 10.1080/01621459.2014.934825

In this article, we propose a Bayesian hierarchical model for object boundaries, which addresses all of the aforementioned problems: (i) guaranteeing valid boundaries through closed curves, (ii) enabling borrowing of information when fitting multiple similar objects, and (iii) employing a multiscale representation suitable for representing curves at different resolutions.

The key innovation in our model is a closed curve-generating random process which can approximate the whole range of simply connected 2D objects. It is based on applying a sequence of multiscale deformations to a closed curve formed by a cyclic basis (Róth et al. 2009). Because the model is multiscale, it is able to detect and borrow interobject similarities at a particular resolution even if similarities are not present at other resolutions.

En route, we solve several important subproblems that may be generally useful in the study of curve and surface fitting. We develop a model-based approach for parameterizing point cloud data. We show how fully Bayesian joint modeling can be used to incorporate several pieces of auxiliary information in the process of curve-fitting, such as when a surface orientation is reported for each point within a point cloud. The concept of multiscale deformation can be generalized to 3D surfaces in a straightforward manner.

2. CLOSED CURVE-GENERATING RANDOM PROCESS

Our closed curve-generating random process starts with a closed curve and performs a sequence of multiscale deformations to generate a final curve. In Section 2.1, we introduce the Roth curve developed by Róth et al. (2009), which will be used to represent the object boundary and demonstrate how to deform a Roth curve at multiple scales to produce any simply connected object. Using the mechanisms developed in Section 2.1, we present the full random process in Section 2.2. In Section 4, we use this as a prior distribution for fitting multiple curves.

2.1 Roth Curves

A Roth curve is a closed parametric curve, $C : [-\pi, \pi] \rightarrow \mathbb{R}^2$, defined by a set of $J = 2n + 1$ control points in \mathbb{R}^2 , $\{c_j, j = 1, \dots, 2n + 1\}$, with n the positive integer-valued degree of the curve. As a function of t , the curve can be viewed as the trajectory of a particle over time. At every time t , the particle’s location is a convex combination of the control points, with weights varying over time according to a set of basis functions, $\{B_j^n(t), j = 1, \dots, J\}$, where $B_j^n(t) > 0$ and $\sum_{j=1}^J B_j^n(t) = 1$ for all t . In particular, we have

$$C(t) = \sum_{j=1}^J c_j B_j^n(t), \quad t \in [-\pi, \pi], \tag{1}$$

$$B_j^n(t) = \frac{h_n}{2^n} \left\{ 1 + \cos \left(t + \frac{2\pi(j-1)}{2n+1} \right) \right\}^n, \tag{2}$$

$$h_n = \frac{(2^n n!)^2}{(2n+1)!},$$

where $c_j = [c_{j,x} \ c_{j,y}]'$ specifies the location of the j th control point and $B_j^n : [-\pi, \pi] \rightarrow [0, 1]$ is the j th basis function. For simplicity, we omit the superscript n denoting a basis function’s degree. This representation is a type of Bezier curve. The Roth curve has appealing properties:

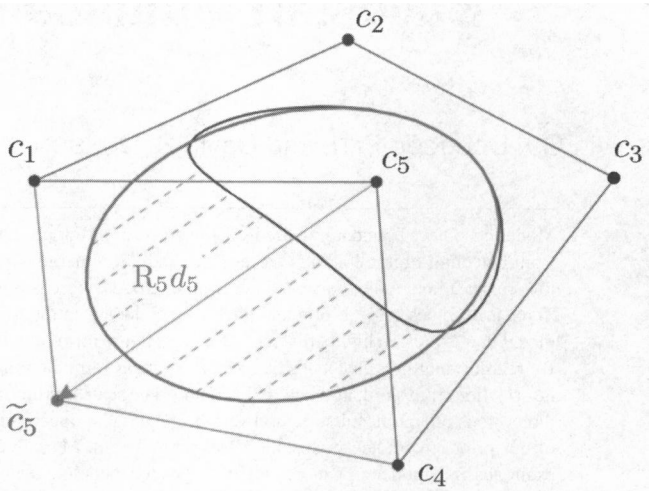


Figure 1. Deformation of a Roth curve.

- 1. $C(-\pi) = C(\pi)$ implies that the curve is always closed. This is necessary to represent the boundary of an object.
- 2. All basis functions are nonlinear translates of each other, and are evenly spaced over the interval $[-\pi, \pi]$. They can be cyclically permuted without altering the curve. This implies that each control point exerts the same “influence” over the curve.
- 3. A degree 1 Roth curve having three control points is always a circle or ellipse.
- 4. Any closed curve can be approximated arbitrarily well by a Roth curve, for some large degree n . This is because the Roth basis, for a given n , spans the vector space of trigonometric polynomials of degree n and as $n \rightarrow \infty$, the basis functions span the vector space of Fourier series.
- 5. Roth curves are infinitely differentiable (C^∞).

Although we could also obtain closed curves using a Fourier basis, the major advantage of the Roth basis is that it has a simple representation for degree elevation and the curve lies in the convex hull of the control points (see Figure 1). Our multiscale model exploits these properties through *deformation* and *degree elevation*.

A Roth curve can be *deformed* by translating some of its control points.

Definition. Suppose we are given two Roth curves,

$$C(t) = \sum_{j=1}^J c_j B_j(t), \quad \tilde{C}(t) = \sum_{j=1}^J \tilde{c}_j B_j(t), \tag{3}$$

where for each j , $\tilde{c}_j = c_j + R_j d_j$, $d_j \in \mathbb{R}^2$ and R_j is a rotation matrix. Then, we say that $C(t)$ is *deformed* into $\tilde{C}(t)$ by the *deformation vectors* $\{d_j, j = 1, \dots, J\}$.

Each R_j orients the deformation vector d_j relative to the original curve’s surface. As a result, positive values for the y -component of d_j correspond to outward deformation, negative values correspond to inward deformation, and d_j ’s x -component corresponds to deformation parallel to the surface. R_j is also

termed as a *deformation-orienting matrix*. In precise terms,

$$R_j = \begin{bmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{bmatrix}, \quad (4)$$

where θ_j is the angle of the curve's tangent line at $q_j = \frac{-2\pi(j-1)}{2n+1}$, the point where the control point c_j has the strongest influence: $q_j = \arg \max_{t \in [-\pi, \pi]} B_j(t)$. θ_j can be obtained by computing the first-derivative of the Roth curve, also known as its hodograph.

Definition. The hodograph of a Roth curve is given by

$$H(t) = \sum_{j=1}^J c_j \frac{d}{dt} B_j(t), \quad (5)$$

$$\frac{d}{dt} B_j(t) = -\frac{2}{(2n+1)\binom{2n}{n}} \sum_{k=0}^{n-1} \binom{2n}{k} (n-k) \sin\left((n-k)t + \frac{2(n-k)(j-1)\pi}{2n+1}\right), \quad (6)$$

where $t \in [-\pi, \pi]$. If we view $C(t)$ as the trajectory of a particle, $H(t)$ intuitively gives the velocity of the particle at point t .

Writing $H(t) = (H_x(t), H_y(t))'$, observe that

$$\theta_j = \arctan\left(\frac{H_y(q_j)}{H_x(q_j)}\right). \quad (7)$$

Hence R_j is a function of $\{c_j \in \mathbb{R}^2, j = 1, \dots, J\}$.

To alter the scale of deformation, we use an important concept called *degree elevation*.

Definition. Given any Roth curve, we can use degree elevation to reexpress the same curve using a larger number of control points (a higher degree). More precisely, if we are given a curve of degree n , $C(t) = \sum_{j=1}^{2n+1} c_j B_j^n(t)$, we can elevate its degree by any positive integer v , to obtain a new degree elevated curve: $\hat{C}(t) = \sum_{j=1}^{2(n+v)+1} \hat{c}_j B_j^{n+v}(t)$ such that $C(t) = \hat{C}(t)$ for all $t \in [-\pi, \pi]$. In $\hat{C}(t)$, each new degree-elevated control point, \hat{c}_j , can be defined in terms of the original control points, $\{c_i, i = 1, \dots, 2n+1\}$:

$$\hat{c}_j := \frac{1}{2n+1} \sum_{i=1}^{2n+1} c_i + \frac{\binom{2(n+v)}{n+v} h_n}{2^{2n-1}} \sum_{k=0}^{n-1} \frac{\binom{2n}{k}}{\binom{2(n+v)}{v+k}} \sum_{i=1}^{2n+1} \cos\left((n-k)\left(\frac{-2(j-1)\pi}{2(n+v)+1}\right) + \frac{2(n-k)(i-1)\pi}{2n+1}\right) c_i.$$

Although daunting to read, the only crucial points to note about this relationship are that \hat{c}_j is linear in c_i 's, $i = 1, \dots, 2n+1$ and that the "influence" of a single control point shrinks after degree elevation. This is because the curve now depends on a greater total number of control points. This implies that after degree-elevation, the translation of any single control point will cause a smaller, finer-scale deformation to the curve's structure. Thus, degree elevation can be used to adjust the scale of deformation. This is exploited in the construction of the random process proposed in Section 2.2.

To that end, we first rewrite all of the concepts described above in more compact vector notation. The formulas for degree

elevation, deformation, the hodograph and the curve itself all simply involve linear operations on the control points. Rewrite the control points in a "stacked" vector of length $2J$,

$$c = (c_{1,x}, c_{1,y}, c_{2,x}, c_{2,y}, \dots, c_{J,x}, c_{J,y})'. \quad (8)$$

The formula for a Roth curve given in (1) can be rewritten as

$$C(t) = X(t)c \quad (9)$$

$$X(t) = \begin{bmatrix} B_1(t) & 0 & B_2(t) & 0 & \dots & B_J(t) & 0 \\ 0 & B_1(t) & 0 & B_2(t) & \dots & 0 & B_J(t) \end{bmatrix}. \quad (10)$$

The formula for the hodograph given in (5) is rewritten as

$$H(t) = \dot{X}(t)c, \quad \dot{X}(t) = \frac{d}{dt} X(t). \quad (11)$$

Deformation can be written as

$$\begin{aligned} \tilde{c} &= c + T(c)d, \\ d &= (d_{1,x}, d_{1,y}, d_{2,x}, d_{2,y}, \dots, d_{J,x}, d_{J,y})', \\ T(c) &= \text{block}(R_1, R_2, \dots, R_J), \end{aligned} \quad (12)$$

where $\text{block}(A_1, \dots, A_q)$ is a $pq \times pq$ block diagonal matrix using $p \times p$ matrices $A_i, i = 1, \dots, q$. T is also termed as the *stacked deformation-orientating matrix*. It is a function of c , because each R_j depends on c . Degree elevation can be written as the linear operator, E :

$$\hat{c} = Ec, \quad E = (E_{i,j})_{i=1, j=1}^{n+v, n},$$

where

$$E_{i,j} = \frac{1}{2n+1} + \frac{\binom{2(n+v)}{n+v} h_n}{2^{2n-1}} \sum_{k=0}^{n-1} \frac{\binom{2n}{k}}{\binom{2(n+v)}{v+k}} \cos\left[(n-k)\left\{\frac{-2(i-1)\pi}{2(n+v)+1}\right\} + \frac{2(n-k)(j-1)\pi}{2n+1}\right].$$

These properties are crucially used to construct multiscale curves and to incorporate any prior knowledge of the shape of the object through appropriate prior distribution for the deformation vectors. Although we can guarantee closedness using the Fourier basis, the curve produced by Fourier basis does not necessarily lie in the convex hull as the basis functions are not guaranteed to be nonnegative. Basis functions described in Zahn and Roskies (1972) can as well be used for constructing the curves as long as a proper characterization of degree elevation and an interpretation of the coefficients in terms of the shape of the curve is available.

2.2 Hierarchical Multiscale Curve Model

Our multiscale curve-generating process starts by drawing an initial set of control points, $c^{(0)}$, from a prior distribution to induce an initial Roth curve. For convenience in notation, we refer to all curves by their control point vectors c . Exploiting the deformation and degree-elevation operations defined above, we repeatedly apply the following recursion R times:

$$\begin{aligned} \hat{c}^{(r-1)} &= E_r c^{(r-1)}, \quad d^{(r)} \sim N(\mu_r, \Sigma_r), \\ c^{(r)} &= \hat{c}^{(r-1)} + T_r(c^{(r-1)})d^{(r)} \end{aligned} \quad (13)$$

resulting in a final curve $c^{(R)}$. In other words, (i) degree elevate the current curve, (ii) randomly deform it, and repeat a total of R

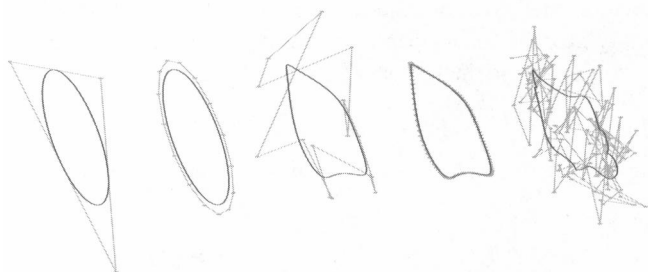


Figure 2. An illustration of the closed curve-generation process. From left to right: (i) initial curve specified by three control points, (ii) the same curve after degree elevation, (iii) deformation, (iv) degree elevation again, and (v) deformation again. Dark lines indicate the curve, pale dots indicate the curve's control points, and pale lines connect the control points in order.

times. This random process induces a prior distribution for the final R -resolution curve $c^{(R)}$. Figure 2 provides an illustration.

We elaborate on the details of this recursive process below. The parameters of the process are

1. R , the number of steps (resolution levels).
2. n_r , the degree of the curve $c^{(r)}$, for each $r = 0, \dots, R$. The sequence $\{n_r\}_0^R$ is strictly monotonically increasing. We let $J_r = 2n_r + 1$.
3. $\mu_r \in \mathbb{R}^{2J_r}$, the average set of deformations applied at step $r = 0, \dots, R$. This vector contains a stack of deformations, not just one.
4. $\Sigma_r \in \mathbb{R}^{2J_r \times 2J_r}$, the covariance in the set of deformations applied at step $r = 0, \dots, R$.

For these parameters, E_r is the degree-elevation matrix going from degree n_{r-1} to n_r , $N(\cdot, \cdot)$ is a $2J_r$ -variate normal distribution and T_r is the stacked deformation orienting matrix. The initial curve $c^{(0)}$ is chosen to be an ellipse, which we generate a priori from $c^{(0)} = T_0 d^{(0)}$, with $d^{(0)} \sim N(\mu_0, \Sigma_0)$.

The random process gives a multiscale representation, because each step produces increasingly fine-scale deformations, through degree elevation. Let S_r denote the class of curves that can be exactly represented by a degree n_r Roth curve. If $\{n_r\}_1^R$ is monotonically increasing, then $S_1 \subset S_2 \subset \dots \subset S_R$. Thus, the deformations $d^{(r)}$ roughly describe the additional details gained going from S_{r-1} to S_r .

Expression (13) provides a multiscale hierarchical model, which can be used not only to generate a single curve but also to borrow information across multiple related curves. Importantly, we allow the level of heterogeneity across curves to vary with resolution level. In some applications, different curves may have similar coarse scale shapes, while varying in fine details. In other applications, there may be more similarities in fine scales.

In Equation (13), μ_r is the mean deformation at level r . Based on $\{\mu_r, r = 0, \dots, R\}$, we define the “central curve” of the random process, c_μ as:

$$c_\mu := c_\mu^{(R)}$$

$$c_\mu^{(r)} = E_r c_\mu^{(r-1)} + T_r (c_\mu^{(r-1)}) \mu_r.$$

As the deformation variance Σ_r decreases, there will be less variance across curves at scale r . If $\Sigma_r \approx 0$ across scales, then

all the curves will closely resemble the central curve. This is illustrated in Figure 3. The central curve provides a convenient summary of a group of curves. The covariance Σ_r also controls correlation between the deformation vectors at the same resolution, allowing incorporation of higher-level assumptions, such as reflected or radial symmetry. For example, if $R = 2$, $n_1 = 1$, and $n_2 = 2$, we can specify perfect correlation in Σ_2 , such that $d_1^{(2)} = d_4^{(2)}$ and $d_2^{(2)} = d_3^{(2)}$. The resulting curves are symmetric along an axis of reflection.

Allard, Chen, and Maggioni (2012) and Xie, Huang, and Willett (2012) introduced multiscale manifold models, which estimate a d -dimensional manifold or more general nonlinear subspace embedded in D -dimensional Euclidean space. In our case, $d = 1$ and $D = 2$. There are several differences between our approach and this literature. Previous multiscale manifold models focus on estimation of a single manifold \mathcal{M} , while we are interested in estimating a population distribution of related manifolds. In addition, instead of point estimation, we estimate a posterior distribution for the population distribution while incorporating the closed curve constraint. In contrast, Allard, Chen, and Maggioni (2012) and Xie, Huang, and Willett (2012) essentially characterize \mathcal{M} on the coarsest scale using a single hyperplane and then introduce additional hyperplanes in going to finer scales. At the coarsest scale, we instead have an ellipse and then degree elevate in the Roth basis in going to finer scales.

3. PROPERTIES OF THE HIERARCHICAL MULTISCALE CURVE MODEL

3.1 Moments

We discuss some basic properties of hierarchical model (13). In particular, we want to study the effect of the parameters at different resolutions on the mean and variance of $c^{(R)}$. The proofs are found in Appendix A.

Proposition 1. Let \mathbb{E}_r denote the expectation with respect to the randomness in the r th resolution and \mathbb{E} denote the overall expectation. Then,

$$\mathbb{E}(c^{(R)}) = E_R E_{R-1} \dots E_1 T_0 \mu_0,$$

$$\text{var}(c^{(R)}) = \sum_{r=1}^R E_R \dots E_{r+1} \mathbb{E}_{r-1} (T_r \Sigma_r T_r') E_{r+1}' \dots E_R' \\ + E_R E_{R-1} \dots E_1 T_0 \Sigma_0 T_0' E_1' E_2' \dots E_R'.$$

$\mathbb{E}(c^{(R)}) = c_\mu^{(r)}$ is the mean curve assuming $\mu_r = 0, r = 1, \dots, R$ and is constructed using a sequence of degree elevations on $T_0 \mu_0$. The dispersion matrix at a particular resolution is the sum of low-rank degree-elevated rotated dispersion matrices from all the previous resolutions.

3.2 Prior Support

To justify use of Roth bases, it is important to study their flexibility. Consider for simplicity a single resolution Roth curve with control points $\{c_j, j = 0, \dots, 2n\}$. Assume we have independent Gaussian priors on each of the two coordinates of c_j for $j = 0, \dots, 2n$, $C(t) = \sum_{j=0}^{2n} c_j B_j^n(t)$, $c_j \sim N_2(0, \sigma_j^2 I_2)$, $j = 0, \dots, 2n$. Denote the prior for C by Π_{C^n} . Π_{C^n} defines an independent Gaussian process for each of the

components of C . Intuitively, the support characterizes the variety of prior realizations along with those which are in their limit. We construct a prior distribution to have large support so that the prior realizations are flexible enough to approximate the true underlying target object. As reviewed in van der Vaart and van Zanten (2008), the support of a Gaussian process (in our case Π_{C^n}) is the closure of the corresponding reproducing kernel Hilbert space (RKHS). The following Lemma 1 describes the RKHS of Π_{C^n} , denoted by \mathbb{H}^n . Refer to Appendix A for the proofs.

Let the Hölder class of periodic functions on $[-\pi, \pi]$ of order α be denoted by $C^\alpha([-\pi, \pi])$. Define the class of closed parametric curves $\mathcal{S}_C(\alpha_1, \alpha_2)$ having different smoothness along different coordinates as

$$\begin{aligned} \mathcal{S}_C(\alpha_1, \alpha_2) &:= \{S = (S^1, S^2) : [-\pi, \pi] \rightarrow \mathbb{R}^2, \\ &S^i \in C^{\alpha_i}([-\pi, \pi]), i = 1, 2\}. \end{aligned} \tag{14}$$

Lemma 1. The RKHS \mathbb{H}^n of Π_{C^n} consists of all functions $h : [-\pi, \pi] \rightarrow \mathbb{R}^2$ of the form

$$h(t) = \sum_{j=0}^{2n} c_j B_j^n(t). \tag{15}$$

The following theorem describes how well an arbitrary closed parametric surface $S_0 \in \mathcal{S}_C(\alpha_1, \alpha_2)$ can be approximated by the elements of \mathbb{H}^n for each n .

Theorem 1. For any fixed $S_0 \in \mathcal{S}_C(\alpha_1, \alpha_2)$, there exists $h \in \mathbb{H}^n$ with $\|h\|_{\mathbb{H}^n}^2 \leq K_1 \sum_{j=0}^{2n} 1/\sigma_j^2$ such that

$$\|S_0 - h\|_\infty \leq K_2 n^{-\alpha_1 \wedge \alpha_2} \log n \tag{16}$$

for some constants $K_1, K_2 > 0$ independent of n .

This shows that the Roth basis expansion is sufficiently flexible to approximate any closed curve arbitrarily well. Although we have only shown large support of the prior under independent Gaussian priors on the control points, the multiscale structure should be even more flexible and hence rich enough to characterize any closed curve. We can also expect minimax optimal posterior contraction rates using the prior Π_{C^n} similar to The-

orem 2 in Pati and Dunson (2014) for suitable choices of prior distributions on n .

4. HIERARCHICAL MODEL FOR MULTIPLE CURVES

4.1 Preprocessing to Obtain Point Cloud from a Pixelated Image

We first describe the necessary preprocessing steps to obtain point clouds representing the boundary pixels of closed contours of the objects of interest. We will illustrate the steps on a panorex teeth image (refer to Figure 4). The goal here is to extract the pixels corresponding to the teeth boundaries. We discuss two simple algorithms.

1. Gradient thresholding: The grayscale image can be treated as a function $Z : \mathbb{R}^2 \rightarrow \mathbb{R}$. The gradient of this function, $\nabla Z : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a vector field, where $\nabla Z(x, y)$ is a vector pointing in the direction of steepest ascent at (x, y) . In computer vision, it is well known that the gradient norm of the image, $\|\nabla Z\|_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ approximates a “line-drawing” of all the high-contrast edges in the image. In practice, an image is discretized into pixels $\{z_{a,b} \mid a = 1, \dots, X, b = 1, \dots, Y\}$ but a discrete version of the gradient can still be computed by taking the difference between neighboring pixels, such that one gradient vector, $g_{a,b}$ is computed at each pixel. The image’s gradient norm is then just another image, where each pixel $m_{a,b} = \|\nabla z_{a,b}\|_2$. Finally, we extract a point cloud: $\{(a, b) \mid m_{a,b} > M, a = 1, \dots, X, b = 1, \dots, Y\}$ where M is some user-specified threshold. Each point (a, b) can still be matched to a gradient vector $g_{a,b} = (g_{(a,b),x}, g_{(a,b),y})'$. For convenience, we will reindex them as p_i and g_i . The gradient vector points in the direction of steepest change in contrast, approximating the object’s surface normal. The surface orientation is then just $\omega_i = \arctan(\frac{g_{i,y}}{g_{i,x}})$.
2. Contour operator: In low-resolution images like the one in Figure 4, simply thresholding the gradient might not be sufficient to obtain a reasonable collection of pixels representing the boundary. A contour operator for edge detection (CORF; Azzopardi and Petkov 2012) is another popular tool particularly suited to low contrast images. CORF also gives a bonus estimate for the object’s

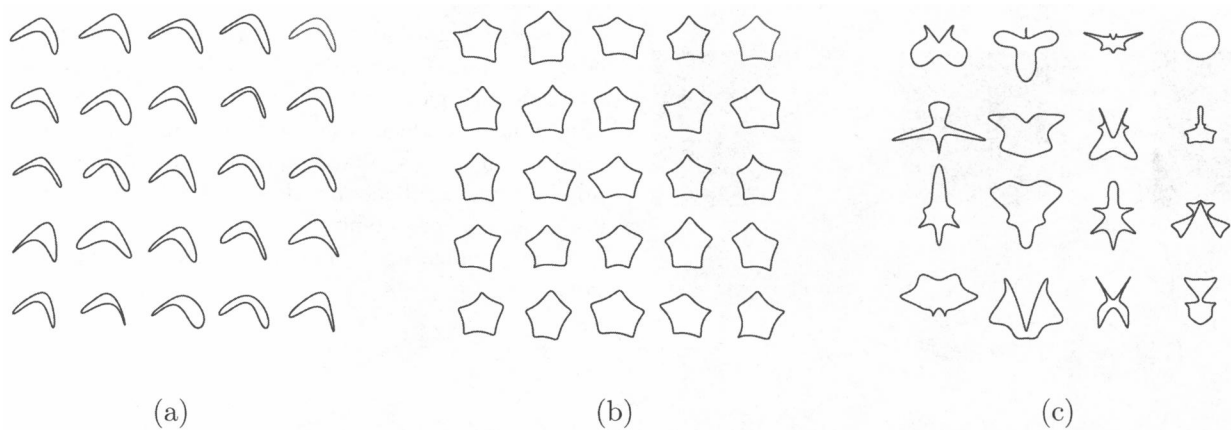


Figure 3. Random samples from the closed curve-generating process (the central curve is in the top right corner of every panel). (a) A moon-shaped collection, (b) star-shaped collection, and (c) high-variance but symmetry-constrained collection.



Figure 4. Panorex image.

surface orientation, ω_i at each point p_i . Due to noise, CORF may fail to detect large portions of a tooth's boundary (false negative), while also identifying false boundaries that do not exist (false positive), typically termed "clutter" (see illustration on three chosen teeth in the top panel of Figure 5). To limit complexity, we manually remove clutter prior to analysis. The resulting boundary pixels are shown in the bottom panel of Figure 5.

4.2 Modeling Multiple Point Cloud Data

Assume that the k th object boundary consists of points $\{p_i^k \in \mathbb{R}^2, i = 1, \dots, N^k\}$ concentrated near the k th 2D closed curve for $k = 1, \dots, K$. Since a Roth curve can be thought of as a function expressing the trajectory of a particle over time, we view each data point, p_i^k from the k th object as a noisy observation of the underlying closed curve C^k evaluated at time t_i^k ,

$$p_i^k = C^k(t_i^k) + \epsilon_i^k, \quad \epsilon_i^k \sim N_2(0, (\sigma^k)^2 I_2). \quad (17)$$

For each k , C^k is constructed using deformations $\{d^{(r),k} \mid r = 0, \dots, R\}$ as in (13),

$$d^{(r),k} \sim N(\mu_r, \Sigma_r),$$

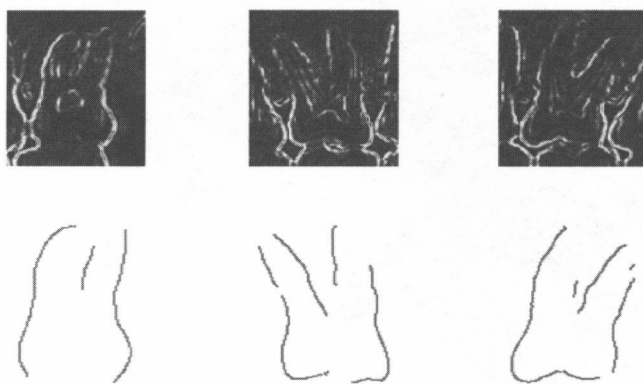


Figure 5. Obtaining point cloud of teeth boundaries from panorex image.

with the following priors placed on the resolution-specific means and covariances:

$$\begin{aligned} \mu_r &\sim N_{2J_r}(\mu_{\mu_r}, \Sigma_{\mu_r}), \\ \Sigma_r^{-1} &= \text{diag} \left(\left[\tau_{1,x}^{(r)}, \tau_{1,y}^{(r)}, \tau_{2,x}^{(r)}, \tau_{2,y}^{(r)}, \dots, \tau_{J_r,x}^{(r)}, \tau_{J_r,y}^{(r)} \right]' \right), \\ \tau_u^{(r)} &\sim \text{Gamma}(\alpha_\tau, \beta_\tau) \quad \text{for } u = \{1, \dots, J_r\} \times \{x, y\}, \end{aligned}$$

where $\text{diag}(v)$ is a diagonal matrix with elements in the vector v along the diagonal. We treat R and $\{n_r \mid r = 0, \dots, R\}$ as elicited hyperparameters. For simplicity we assume a diagonal covariance structure Σ_r , which implies independence between deformations and the x, y components of each deformation. It is straightforward to allow nondiagonal Σ_r using inverse-Wishart priors, but the resulting model becomes heavily parameterized and difficult to compute. This hierarchical structure induces dependence between the K curves, enabling them to borrow information from each other during fitting.

To appropriately borrow information, objects need to be aligned to remove differences in position and orientation. We choose the initial curve $c^{(0),k}$ to be degree $n_0 = 1$, which guarantees that it is an ellipse. For $j = 1, 2, 3$, we define each control point as

$$c_j^{(0),k} = (0, 0)' + R_{\theta_j} d_j^{(0),k}, \quad (18)$$

$$R_{\theta_j} = \text{rotation matrix where } \theta_j = \frac{2\pi j}{3}, \quad (19)$$

and where each $d_j^{(0),k} \in \mathbb{R}^2$ is a random deformation vector. We start with a curve that is just a point at the origin, $C^k(t) \equiv (0, 0)$, and apply three random deformations which are rotated by a radially symmetric amount: 0° , 120° , and 240° (the final deformations are not radially symmetric, since each d_j is randomly drawn). We can use the following vector notation:

$$d^{(0),k} \sim N(\mu_0, \Sigma_0), \quad c^{(0),k} = \mathbf{0} + T_0 d^{(0),k}.$$

According to (18), our random process generates closed curves centered at $(0, 0)$ and rotated to a fixed angle. However, in an actual collection of curves, each object is rotated to a different angle and centered at a different location. We can modify (18) to account for this simply by adding latent variables for the position, $m^k \in \mathbb{R}^2$, and orientation, $\phi^k \in [-\pi, \pi]$, of each object k :

$$c_j^{(0),k} = m^k + R_{\phi^k} R_{\theta_j} d_j^{(0),k},$$

where R_{ϕ^k} is a rotation matrix. We place a uniform prior on ϕ^k and a normal prior on m^k . It can be desirable to put a more sophisticated spatial prior on m^k , but our focus here is on modeling object boundaries, not their location. The orientation of level $r = 0$ orients all subsequent levels. This is sufficient to align the entire collection and make the deformation vectors of each curve directly comparable.

To summarize the hierarchical model, we introduce a few notations:

$$\begin{aligned} p^k &= (p_{1,x}^k, p_{1,y}^k, \dots, p_{N^k,x}^k, p_{N^k,y}^k)', \\ \epsilon^k &= (\epsilon_{1,x}^k, \epsilon_{1,y}^k, \dots, \epsilon_{N^k,x}^k, \epsilon_{N^k,y}^k)', \\ t^k &= (t_1^k, t_2^k, \dots, t_{N^k}^k)', \\ \mathbb{X}^k(t^k)' &= [X^k(t_1^k)' X^k(t_2^k)' \dots X^k(t_{N^k}^k)'], \end{aligned}$$

where $X^k(t_i^k)$ is as defined in (11). Thus, we have

$$p^k = \mathbb{X}^k(t)c^{(r),k} + \epsilon^k, \quad r = 0, \dots, R \quad (20)$$

$$\begin{aligned} \epsilon &\sim N_{2N^k}(0, (\sigma^k)^2 I_{2N^k}) \\ c_j^{(0),k} &= m^k + R_{\phi^k} R_{\theta_j} d_j^{(0),k} \end{aligned} \quad (21)$$

$$\begin{aligned} \tilde{c}^{(r-1)} &= E_r c^{(r-1)}, \quad d^{(r),k} \sim N(\mu_r, \Sigma_r), \\ c^{(r)} &= \tilde{c}^{(r-1)} + T_r(c^{(r-1)})d^{(r)} \end{aligned} \quad (22)$$

$$\mu_r \sim N_{2J_r}(\mu_{\mu_r}, \Sigma_{\mu_r}), \quad (23)$$

$$\Sigma_r^{-1} = \text{diag} \left(\left[\tau_{1,x}^{(r)}, \tau_{1,y}^{(r)}, \tau_{2,x}^{(r)}, \tau_{2,y}^{(r)}, \dots, \tau_{J_r,x}^{(r)}, \tau_{J_r,y}^{(r)} \right]' \right), \quad (24)$$

$$\tau_u^{(r)} \sim \text{Gamma}(\alpha_\tau, \beta_\tau) \quad \text{for } u = \{1, \dots, J_r\} \times \{x, y\} \quad (25)$$

(20)–(25) complete a Bayesian formulation for inferring $\{c^{(r),k}, r = 0, \dots, R\}, k = 1, \dots, K$, given $p^k, k = 1, \dots, K$ and $t^k, k = 1, \dots, K$. Typically, t_i^k is not known and can be treated as a latent covariate corresponding to point p_i^k . There is no ordering constraint on the t_i^k s, and we let $t_i^k \sim U(-\pi, \pi)$. We discuss posterior computation using our model in Appendix B.

A natural question is whether the posterior distribution is invariant to rotating and translating the point cloud. Suppose the data points p_i are transformed to $p_i^* = Op_i + T$, where O is a rotation matrix and $T \in \mathbb{R}^2$. Then, the likelihood obtained by fitting the curve $C(t)$ to the points p_i^* is same as the likelihood obtained by fitting $O'C(t) + T$ to the points p_i , as the error variance is invariant to rotation. Clearly, if $C(t) = \sum_{j=1}^J c_j B_j(t)$, the curve $O'C(t) + T = \sum_{j=1}^J (O'c_j + T)B_j(t)$ is obtained using rotating and translating the original control points. Assuming the same prior for the control points, the posterior distribution will be translated and rotated using T and O' , respectively.

4.3 Modeling Surface Orientation

In the following, we show how to include data on surface orientation ω_i^k into (20)–(25). We incorporate this extra information into our model to improve fitting, with essentially no sacrifice in computational efficiency. Denote by $v_i^k = (H_x(t_i^k), H_y(t_i^k)) \in \mathbb{R}^2$ the velocity vector of the curve $C^k(t)$ at the parameterization location $t_i^k, i = 1, \dots, N^k$. v_i^k is always tangent to the curve. Since each ω_i^k points roughly normal to the curve, we can rotate all of them by 90° , $\theta_i^k = \omega_i^k + \frac{\pi}{2}$, and treat each θ_i^k as a noisy estimate of v_i^k 's orientation. We cannot rotate the vector g_i^k by 90° and directly treat it as a noisy observation of v_i^k . In particular, g_i^k 's magnitude bears no relationship to the magnitude of v_i^k : $\|g_i^k\|$ is the rate of change in image brightness when crossing the edge of the object, while $\|v_i^k\|$ describes the speed at which the curve passes through p_i^k .

Suppose we did have some noisy observation of v_i^k , denoted u_i^k . Then, we could have specified the following linear model relating the curve C^k to the u_i^k 's:

$$u_i^k = v_i^k + \delta_i^k \quad (26)$$

$$= \frac{d}{dt} C^k(t)|_{t=t_i^k} + \delta_i^k \quad (27)$$

for $i = 1, \dots, N^k$ where $\delta_i \sim N_2(0, (\tau^k)^2 I_2)$. Instead, we only know the angle of u_i^k, θ_i^k . In Appendix B, we show that using this model, we can still write the likelihood for θ_i^k , by marginalizing out the unknown magnitude of u_i^k . The resulting likelihood still results in conditional conjugacy of the control points.

5. SIMULATION STUDY

To demonstrate borrowing of information at multiple scales, we define several curves which share the same features at a fine-scale resolution, but differ in their coarse-scale structure. Each curve is sparsely sampled with noise to produce a point cloud, and our goal is to recover the original curves. Due to gaps in the point cloud, the original curves are very hard to recover without detecting and leveraging structural similarities at the fine scale. This scenario is similar to real-world applications where a single object has been observed in multiple poses, or a collection of similar objects have been observed. We compare our model (HBM: hierarchical Bayesian model) against a simpler version of our model that fits the different curves independently (IM: independence model), and against principal curves (PC; Hastie and Stuetzle 1989).

Because the curves differ in structure at the coarse scale, there is no trivial way to align the point clouds and identify shared fine-scale features. Thus, it was not apparent how to achieve borrowing of information using principal curves prompting us to fit PC independently to each of the point clouds. For HBM, it is only necessary to initialize the parameterization of each point cloud via polar-coordinate parameterization (see Appendix B.4). The method is robust to small errors in initial parameterization.

We define several concepts to help interpret our results. Given some curve c , let $B(c) = \{\mathbb{X}(t)c \mid t \in [-\pi, \pi]\}$ denote the set of all points along the curve, and let $A(c)$ denote the interior region enclosed by the curve. For a given distribution over curves, $P(c)$, we define its boundary heatmap, $M_{P(c)}^B : \mathbb{R}^2 \rightarrow \mathbb{R}$, and its region heatmap, $M_{P(c)}^A : \mathbb{R}^2 \rightarrow \mathbb{R}$, as:

$$M_{P(c)}^A(x, y) = P((x, y) \in A(c)),$$

$$M^B(x, y) dx dy = P((x + dx, y + dy) \cap B(c)).$$

Given a set of samples from the distribution $P(c)$, $\{c_s \mid s = 1, \dots, S\}$, we can discretely approximate $M_{P(c)}^B$ and $M_{P(c)}^A$ as:

$$M_{P(c)}^B(x, y) \approx \frac{1}{S} \sum_{s=1}^S \mathbf{1}[W(x, y) \cap B(c_s)],$$

$$M_{P(c)}^A(x, y) \approx \frac{1}{S} \sum_{s=1}^S \mathbf{1}[(x, y) \in A(c_s)],$$

where $W(x, y) = \{(x', y') \mid (x', y') \in [x \text{ div } \Delta_x, x \text{ div } \Delta_x + \Delta_x] \times [y \text{ div } \Delta_y, y \text{ div } \Delta_y + \Delta_y]\}$ (div denotes integer division). The function W simply maps values in \mathbb{R}^2 to a regular grid of bins with width Δ_x and height Δ_y . Finally, the mean can be approximated by $\hat{c} = \frac{1}{S} \sum_{s=1}^S c_s$.

For the example given in Figure 6, we have used two resolutions. The hyperparameters were set to: $r_1 = 1, r_2 = 8$, $\Sigma_1 = 1000 I_{2J_{r_1}}, \Sigma_2 = 0.001 I_{2J_{r_2}}, \mu_{\mu_1} = [1 \ 1 \ 1]' \otimes [0 \ 40]'$, $\mu_{\mu_2} = \mathbf{0}_{2J_{r_2}}, \Sigma_{\mu_1} = 1000 I_{2J_{r_1}}, \Sigma_{\mu_2} = 1000 I_{2J_{r_2}}, \alpha = 1, \beta = 1, \sigma_p = 0.1$.

Convergence was monitored using the Raftery and Lewis diagnostic test as well as trace plots of the deviance parameters. Also, we get essentially identical posterior summaries with different MCMC starting points and moderate changes to hyperparameters. We additionally conduct sensitivity analyses in

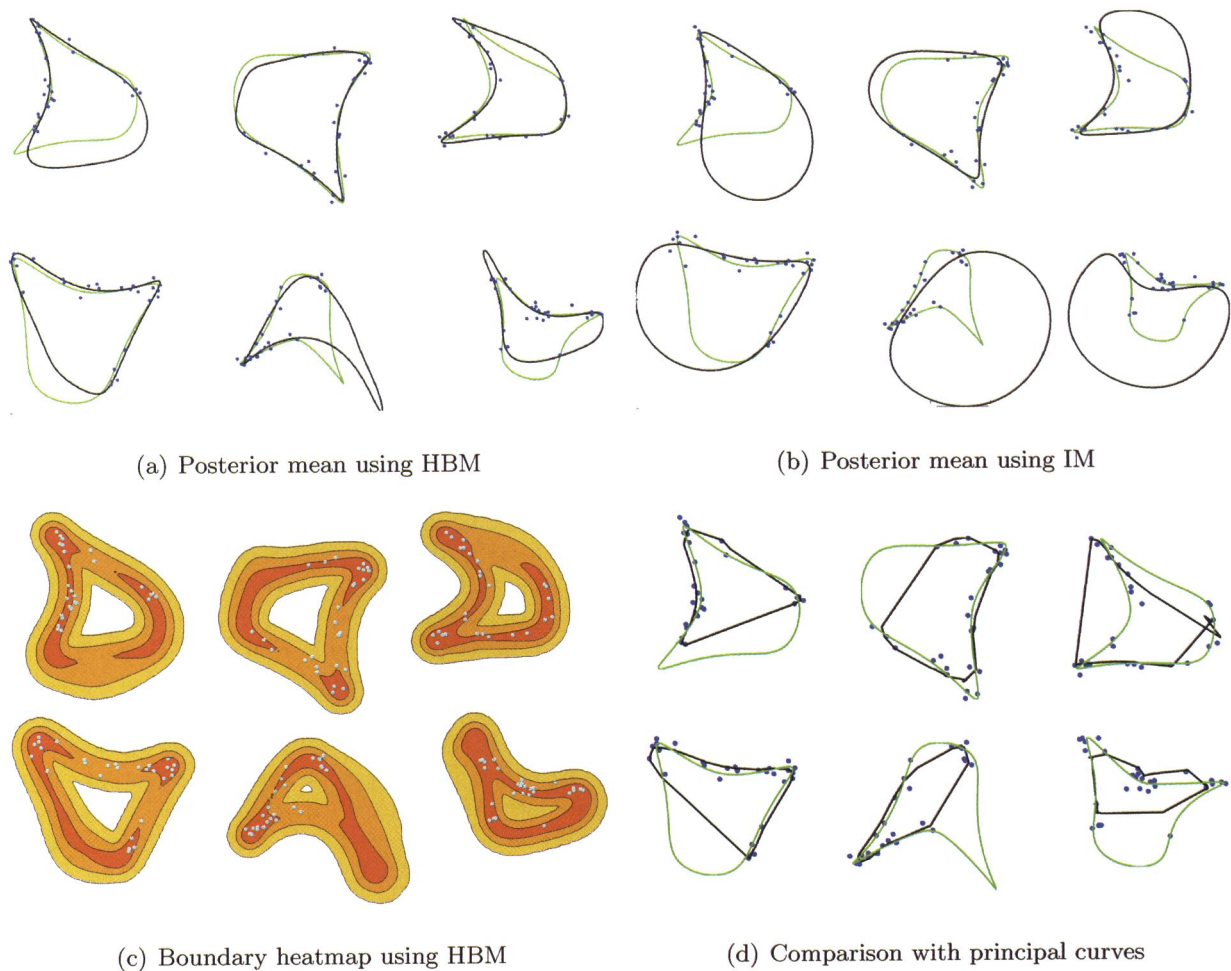


Figure 6. Simulation study (blue/cyan dots: data points, green line: true curve, black line: posterior mean, M^B discretized into three colored regions—red: > 0.95 , orange: > 0.5 , yellow: > 0.01).

which we vary hyperparameters in the priors for μ_0 and Σ_0 , demonstrating robustness to the central curve initialization.

It is evident from Figures 6(a) and 6(b) that HBM leads to more accurate estimates by borrowing information across different point clouds. In most of the cases, IM fails to recover the true boundary from sparse data. The heatmaps in Figures 6(c) suggest that HBM is successful in properly characterizing uncertainty near sparse regions. PC, not being able to borrow information across the different point clouds, fails to successfully recover the missing chunks of the boundary (refer to Figure 6(d)).

The posterior distribution for c_μ successfully captures the artificially injected variability in the structure of the curve. Again, we interpret $P(c_\mu \mid \{p^k\})$ using its boundary heatmap $M^B_{P(c_\mu \mid \{p^k\})}$ (just M^B for notational convenience). Figure 7 shows that the posterior estimate of c_μ using HBM provides a good estimate of the true curve.

6. REAL-DATA APPLICATIONS

We consider two real-data applications to demonstrate the efficacy of our closed curve-generating model. The first one shows that the model can borrow information across multiple closed curves to fill in missing boundaries of teeth from low

resolution panorex images. The second application illustrates how to intelligently combine expert traces of a brain tumor using our model to come up with a heatmap of the tumor contour, a useful graphical tool for excising tumors.

6.1 Filling Missing Teeth Boundaries in Panorex Images

The panorex X-ray (refer to Figure 4) is a useful technique to diagnose dental problems related to cavities and jaw. The X-ray provides an image of the teeth and surrounding bones. Sometimes referred to as a single FMX, or full mouth X-ray,

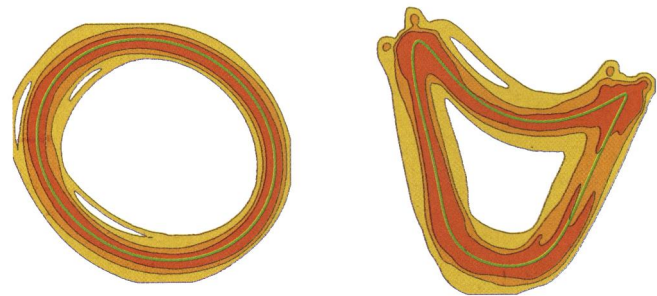


Figure 7. Central curve heatmap for tw resolutions (the left one with 3 control points and the right with 33 control points).

the panorex provides a two-dimensional panoramic view of the mouth. The image quality is highly dependent on patient positioning and features of the laser technology used. Hence, teeth outlines are significantly hard to determine using simple edge detection algorithms which fail to take into account the noise associated with each pixel or allow borrowing of information from similar teeth. Clearly, in these low contrast panorex images, there is insufficient data to recover local segments of the boundary of a single tooth in isolation. Our goal is to use the multiscale random closed curve-generating process to borrow information from similar teeth to fill in missing boundaries.

As mentioned before, the image in Figure 4 is processed with CORF to extract the edge pixels, shown in the top panel of Figure 8(a). Then, the clutter is manually removed to obtain the boundary pixels (middle panel of Figure 8(a)). Observe that the removal of clutter does not provide any extra information about missing boundaries. We implement HBM on the boundary pixels of 6 chosen teeth using 5000 MCMC iterations after discarding a burn-in of 2000. The hyperparameters of our model were set to: $r_1 = 1, r_2 = 5, r_3 = 15, \mu_{\mu_1} = \mathbf{J}'_{2J_{r_1}} \otimes [100\ 0]', \mu_{\mu_2} = \mathbf{0}_{2J_{r_2}}, \mu_{\mu_3} = \mathbf{0}_{2J_{r_3}}, \Sigma_{\mu_1} = 2000 \mathbf{I}_{2J_{r_1}}, \Sigma_{\mu_2} = 1000 \mathbf{I}_{2J_{r_2}}, \Sigma_{\mu_3} = 2000 \mathbf{I}_{2J_{r_3}}, \alpha = 1/30, \beta = 1/30, \sigma_p = 2$.

In the bottom panel of Figure 8(a), we display the results of HBM. The teeth enclosed by the posterior mean contours are shown in gray. Clearly, HBM successfully recovered most of the missing boundaries. The central curves for each of the resolutions are shown in Figure 8(b). The radii of the circles centered at the control points represent the posterior uncertainty associated with each control point and provide useful information on the uncertainty of the teeth contours.

We compared our method with two state-of-the-art techniques (i) active contours Chan and Vese (2001), suited to curve estimation from images (shown in the bottom panel of Figure 9) and (ii) principal curves, suited to curve estimation from point clouds (shown in the top panel of Figure 9). As there are no obvious ways we can allow the methods to borrow information across related teeth, we implement them independently on each tooth. To put active contours and principal curves on even footing with our model, we seeded both methods with an initial curve formed by taking all of the boundary pixels for a given tooth (after removing clutter) and manually “connecting the dots” in the correct order. This favorable initialization is not typically provided when running either method. In addition to existing preprocessing steps, the active contours require delineating part of the boundary which is shared by the neighboring teeth to prevent the contours from spilling into the neighboring pixels. This is a highly undesirable phenomenon as it requires manual input of extra information on the shape of the boundary. The results of the principal curves are highly sensitive to the choice of the smoothing parameters. A large value of the smoothing parameter fails to capture finer details while smaller values lead to deranged fits involving self-intersecting curves. We search over different smoothing parameters to find a single value that performed reasonably for all the teeth. As evident from Figures 8(a) and 9, the hierarchical Bayesian model performs much better relative to the competitors.

6.2 Brain Tumor Tracing

In brain tumor diagnosis and therapy, it is very important to account for uncertainty in the tumor’s outline. This information is crucial for assessing whether a tumor has grown/regressed over time, and even more important if a surgeon must target the tumor for excision or radiation therapy. In that situation, there is a critical tradeoff between false positives (targeting healthy tissue) and extremely undesirable false negatives (missing the tumor). Furthermore, tumor outlines are notoriously hard to determine. Error stems from the poor contrast between tumor and healthy tissue in magnetic resonance imaging (MRI), the prevalent modality for diagnosis. Even seasoned experts differ greatly when tracing an estimate.

In the current application of combining information across multiple radiologist segmentations of a brain tumor, a primary advantage is having a posterior distribution over segmentations which we used to create heatmaps demonstrating uncertainty. Such uncertainty regions are quite important in making treatment decisions, but approaches for consensus segmentation in the computer vision literature produce a point estimate.

We use our model to intelligently combine the input traces of multiple experts obtained from a brain tumor segmentation study in Archip, Jolesz, and Warfield (2007), by treating each trace as a different realization from the same random process. We can then interpret $P(c_\mu | \{p^k\})$ as the posterior distribution of the tumor, fully describing the variability and uncertainty among the experts. One might also run additional tumor segmentation algorithms, and combine their outputs using the same approach. In this setting, the region heatmap of the posterior, $M^A_{P(c_\mu|\{p^k\})}$ (shortened to M^A), is especially informative. For every point x , $M^A(x)$ gives the probability that it is part of the tumor. This enables a neurosurgeon to manage the tradeoff between false positives/negatives in a principled manner. Let the true tumor region be $X_{\text{tumor}} \subset \mathbb{R}^2$ and X_{tumor}^c its complement. Then, define the loss function for targeting a region X to be

$$L(X) = \lambda_+ \text{Area}(X \cap X_{\text{tumor}}^c) + \lambda_- \text{Area}(X^c \cap X_{\text{tumor}}).$$

Depending on the ratio of the penalties λ_+ and λ_- , the surgeon can minimize L simply by cutting along a level set of M^A .

Figure 10(a) shows the raw MRI image and Figure 10(b) shows the region heatmap corresponding to the highest resolution. The hyperparameters were set to: $r_1 = 1, r_2 = 5, r_3 = 17, \mu_{\mu_1} = [1\ 1\ 1]' \otimes [0\ 300]', \mu_{\mu_2} = \mathbf{0}_{2J_{r_2}}, \mu_{\mu_3} = \mathbf{0}_{2J_{r_3}}, \Sigma_{\mu_1} = 2200 \mathbf{I}_{2J_{r_1}}, \Sigma_{\mu_2} = 2200 \mathbf{I}_{2J_{r_2}}, \Sigma_{\mu_3} = 2200 \mathbf{I}_{2J_{r_3}}, \alpha = 1, \beta = 0.01, \sigma_p = 0.05$.

The lower middle region in the resolution 3 region heatmap shows more uncertainty compared to other regions demonstrating the experts tend to differ more near the inward indentation at the lower middle region of the tumor. Thus, one can cut along the desired level set of the heatmap to obtain a better surgery than to just cut along any one of the four traces.

We can also allow for experts to express varying confidence in different portions of their trace. This is desirable, because certain boundaries of the tumor will have high contrast with the surrounding tissue while other parts would not, and the expert

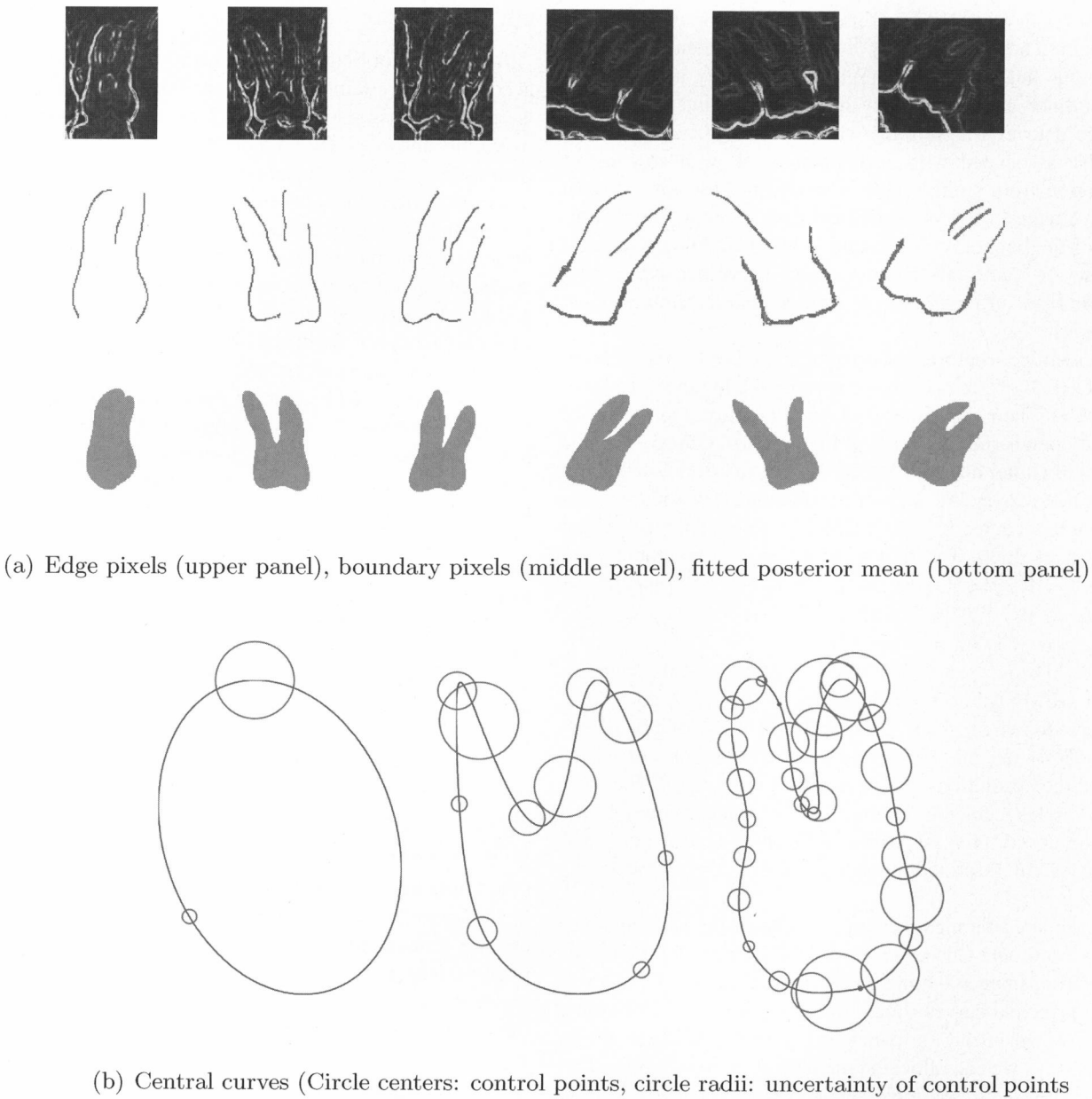


Figure 8. HBM on panorex images.

should not be forced to make an equal opinion on both. We can achieve this by slightly modifying the point cloud model given in (17). There, we assumed that each point p_i was generated with fixed variance σ_p^2 . Instead, we can let $\sigma_{p_i}^2 = \sigma_p^2/\kappa_i$, where κ_i is the expert’s confidence in that point. Furthermore, if the expert has no confidence at all, they can simply leave a gap in their trace. The model automatically closes the gap, as shown in simulation examples. Finally, it is also easy to compute the posterior distribution for quantities such as the size of the tumor, simply by computing the size of each sample.

7. DISCUSSION

We have developed a fully Bayesian hierarchical model based on multiscale deformations for modeling a collection of 2D

closed curves. In using a parametric curve, a fundamental issue is to render the analysis of curves invariant to the choice of parameterization. Although we have characterized a collection of curves using our fully Bayesian approach, it is not clear whether our approach leads to a parameterization invariant analysis. There is an elegant literature on parameterization invariant shape analysis (Kurtek et al. 2010; Kurtek et al. 2011; 2012). However their approaches are not fully model based. Comparing the shapes of objects in a rigorous Riemannian framework will involve further work, such as defining a loss function involving an appropriate metric between shapes.

In defining the multiscale process, we would like to have a more automatic way of choosing the different resolutions. It is clear that the highest resolution is obtained by maximizing the fit subject to minimizing the Bayesian penalty for model

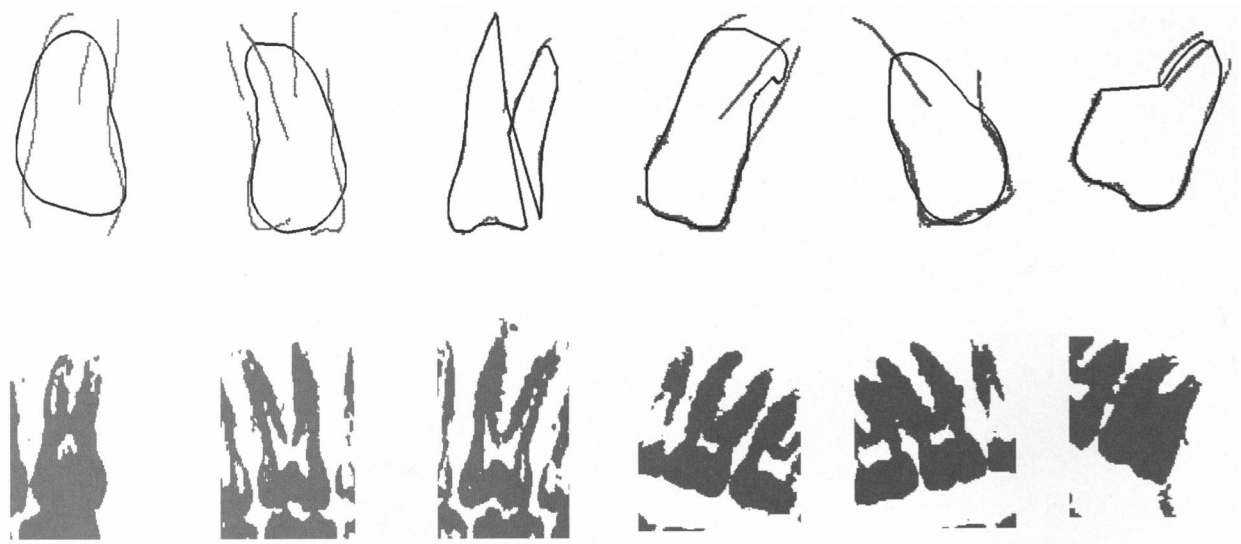


Figure 9. Comparison with principal curves (top panel) and active contours (bottom panel).

complexity. In our future research, we would like to have a more informed way of selecting the lower resolutions.

Our multiscale model differs in its purpose from other multiscale methods such as the wavelet transform. With wavelet methods, the goal is often to compress the data, whereas our goal is to define levels that isolate dimensions of similarity or variability within a collection of objects. However, one can use any wavelet basis instead of the Roth basis to define analogously a multiscale random curve-generating process. In certain applications where it is desirable to accurately estimate the fine scale local features, a compactly supported basis might be more suitable.

Typically real images are contaminated with noise from different sources (Srivastava and Jermyn 2009) (i) shape variations within classes, (ii) variability in sampling continuous curves, (iii) pose and scale variability, (iv) observation noise, and (v) points introduced by clutter. Taking into account all these will be a challenge in a model based framework. However, clutter can be taken care of by introducing a mixture model involving two components—the first one for pixels coming from the target image contour and the second one for the clutter pixels.

For simplicity of exposition in the current article, we focus our attention only on modeling the boundary pixels of the target object.

APPENDICES

A.1. PROOFS OF MAIN RESULTS

Proof of Proposition 1. Assuming $\mu_r = 1, \dots, R$,

$$\begin{aligned} \mathbb{E}(c^{(R)}) &= E_R \mathbb{E}(c^{(R-1)}) = \dots = E_R E_{R-1} \dots E_1 \mathbb{E}(c^{(0)}) \\ &= E_R E_{R-1} \dots E_1 T_0 \mu_0. \end{aligned}$$

Observe that

$$\begin{aligned} \text{var}(c^{(R)}) &= \mathbb{E}_{R-1}(\text{var}(c^{(R)} \mid c^{(R-1)}) + \text{var}(\mathbb{E}_{R-1}(c^{(R)} \mid c^{(R-1)})) \\ &= \mathbb{E}_{R-1}(T_R \Sigma_R T_R') + \text{var}(E_R c^{(R-1)}) \\ &= \mathbb{E}_{R-1}(T_R \Sigma_R T_R') + E_R \text{var}(c^{(R-1)}) E_R'. \end{aligned}$$

Iterating $R + 1$ times, we obtain the required result. □

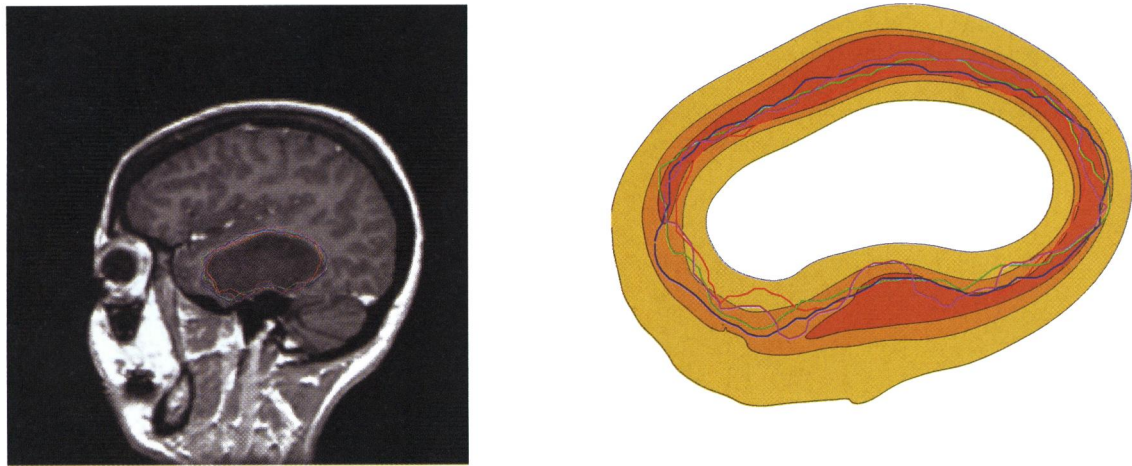


Figure 10. (a) Four expert traces overlaid on MRI. (b) Resolution 3 region heatmap.

Proof of Lemma 1. The Gaussian process prior Π_{C^n} given $\{\sigma_j, j = 0, \dots, 2n\}$ has the following representation.

$$C^n(t) = \sum_{j=0}^{2n} c_j B_j^n(t), \quad c_j \sim N_2(0, \sigma_j^2 I_2), \quad (t) \in [-\pi, \pi]. \quad (\text{A.1})$$

Since $c_j \sim N_2(0, \sigma_j^2 I_2)$, Π_{C^n} can be written as

$$C^n(t) = \sum_{j=0}^{2n} c_j^* \sigma_j B_j^n(t). \quad (\text{A.2})$$

where $c_j^* \sim N_2(0, I_2)$. Hence from Proposition 1 in Pati and Dunson (2014), \mathbb{H}^n consists of $h : [-\pi, \pi] \rightarrow \mathbb{R}^2$ such that

$$h(t) = \sum_{j=0}^{2n} c_j B_j^n(t), \quad (\text{A.3})$$

where $c_j \in \mathbb{R}^2$. The RKHS norm of h in (A.3) is given by $\|h\|_{\mathbb{H}^n}^2 = \sum_{j=0}^{2n} \|c_j\|^2 / \sigma_j^2$. \square

Proof of Theorem 1. From Stepanets (1974) and observing that the basis functions $\{B_j^n, j = 0, \dots, 2n\}$ span the vector space of trigonometric polynomials of degree at most n , it follows that given any $S_0^i \in C^{\alpha_i}([-\pi, \pi])$, there exists $h^i(u) = \sum_{j=0}^{2n} c_j^i B_j^n(u)$, $h^i : [-\pi, \pi] \rightarrow \mathbb{R}$ with $|c_j^i| \leq M_i$, such that $\|h^i - S_0^i\|_{\infty} \leq K_i n^{-\alpha_i} \log n$ for some constants $M_i, K_i > 0, i = 1, 2$. Setting $h(u) = \sum_{j=0}^{2n} (c_j^1, c_j^2)' B_j^n(u)$, we have

$$\|h - S_0\|_{\infty} \leq M n^{-\alpha(1)} \log n$$

with $\|h\|_{\mathbb{H}}^2 \leq K \sum_{j=0}^{2n} \phi_j$ where $M = M_{(2)}, K = K_{(2)}$. \square

B. POSTERIOR COMPUTATION

In Section 4, we presented a model for characterizing and fitting a collection of K closed curves, with unknown underlying parameterization. Below we describe an MCMC algorithm for sampling from the joint posterior of this model. This involves deriving the conditional posteriors of $m^k, d^{(r),k}, \mu_r, \Sigma_r$ and t^k for $r = 0, \dots, R$ and $k = 1, \dots, K$. We obtain the posterior samples from $\sum_{j=1}^J c_j B_j(t)$ for a grid of points $t_1, \dots, t_k \in [-\pi, \pi]$ and obtain the average across MCMC iterations for each gridpoint.

B.1 Conditional Posteriors for m^k and $d^{(r),k}$

The conditional posteriors for m^k and $d^{(r),k}$ are the most challenging to sample from, because our model's likelihood function is nonlinear in these terms, preventing conditional conjugacy. To overcome this, we derive a linear approximation to the true likelihood function, which does yield conditional conjugacy, and then use samples from the approximate conditional posterior as proposals in a Metropolis-Hastings step.

We first discuss the source of nonlinearity in the likelihood function. Assuming $\sigma^k = \sigma$ for all k for simplicity of exposition, recall from Section 4.2 that

$$P(p^k | c^{(R),k}) = N_{2N}(p^k; \mathbb{X}(t^k) c^{(R),k}, \sigma^2 I_{2N}).$$

From Section 2.2, $c^{(R),k}$ is the result of combining the deformations $\{d^{(r),k} | r = 0, \dots, R\}$ through the following recursive relation, with m^k appearing in the base case:

$$c^{(r),k} = E_r c^{(r-1),k} + T_r (c^{(r-1),k}) d^{(r),k} \quad (\text{B.1})$$

$$c^{(0),k} = m^k + T_{\phi^k} T_0 d^{(0),k}. \quad (\text{B.2})$$

At each step r of the recursive process, the deformation-orienting matrix T_r is a nonlinear function of the previous $c^{(r-1),k}$. As a result, $c^{(R),k}$ is nonlinear in $d^{(r),k}$ for $r = 0, \dots, R-1$. For any given step

of the process, we can replace the true recursive relation with a linear approximation. In particular, we will substitute $T_r(c^{(r-1),k}) d^{(r),k}$ with $\hat{T}_{r,k} c^{(r-1),k}$, where $\hat{T}_{r,k}$ will be derived shortly. The new approximate step is then

$$c^{(r),k} \approx (E_r + \hat{T}_{r,k}) c^{(r-1),k}. \quad (\text{B.3})$$

If we wish to write $c^{(R),k}$ linearly in terms of $c^{(r),k}$ for any $r = 0, \dots, R$, we can replace every recursive step from r to R with the approximate step given in (B.3). We emphasize that steps $0, \dots, r-1$ follow the original recursive relation. This yields the approximation

$$c^{(R),k} \approx \Omega_{r+1}^R c^{(r),k}, \quad \Omega_a^b = \prod_{\rho=a}^b (E_\rho + \hat{T}_{\rho,k}). \quad (\text{B.4})$$

Combining (B.1) and (B.4), we have that

$$c^{(R),k} \approx \begin{cases} \Omega_{r+1}^R [E_r c^{(r-1),k} + T_r (c^{(r-1),k}) d^{(r),k}] & r > 0 \\ \Omega_1^R (m^k + T_{\phi^k} T_0 d^{(0),k}) & r = 0 \end{cases}.$$

Thus, the approximation of $c^{(R),k}$ can be written linearly in terms of any $d^{(r),k}$ or m^k . It is still nonlinear in $d^{(\rho),k}$ for any $\rho \neq r$. However, for MCMC sampling, we only need one $d^{(r),k}$ to be linear at a time, holding all others fixed. The approximation becomes increasingly good as r approaches R , because the number of approximate steps (contained in Ω_a^b) decrease.

We are now in a position to derive the approximate conditional posteriors for m^k and $d^{(r),k}$. We claim that these posteriors can all be written in the following form for generic 'x', 'y', and 'z'.

$$P(x | -) \propto N(y; Qx, \Sigma_y) N(x; z, \Sigma_x) \quad (\text{B.5})$$

$$P(x | -) = N(\hat{\mu}, \hat{\Sigma}), \quad \hat{\Sigma}^{-1} = \Sigma_x^{-1} + \sum_k Q' \Sigma_y^{-1} Q,$$

$$\hat{\mu} = \hat{\Sigma} \left(\Sigma_x^{-1} z + \sum_k Q' \Sigma_y^{-1} y \right). \quad (\text{B.6})$$

Each approximate conditional posterior is simply a multivariate normal. We show that each approximate posterior can be rearranged to match the form of (B.5) and (B.6).

$$\begin{aligned} P_{\text{approx}}(m^k | -) &\propto N(p^k; \mathbb{X}(t^k) c^{(R),k}, \sigma^2 I_{2N^k}) N(m^k; \mu_m, \Sigma_m) \\ &\propto N(p^k; \mathbb{X}(t^k) \Omega_1^{R,k} (m^k + T_0 d^{(0),k}), \sigma^2 I_{2N^k}) \\ &\times N(m^k; \mu_m, \Sigma_m) \\ &\propto N(p^k - \mathbb{X}(t^k) \Omega_1^{R,k} T_0 d^{(0),k}; \\ &\quad \mathbb{X}(t^k) \Omega_1^{R,k} m^k, \sigma^2 I_{2N^k}) \\ &\times N(m^k; \mu_m, \Sigma_m) \end{aligned}$$

$$\begin{aligned} P_{\text{approx}}(d^{(r),k} | -) &\propto N(p^k; \mathbb{X}(t^k) c^{(R),k}, \sigma^2 I_{2N^k}) N(d^{(r),k}; \mu_r, \Sigma_r) \\ &\propto N(p^k; \mathbb{X}(t^k) \Omega_{r+1}^R [E_r c^{(r-1),k} \\ &\quad + T_r (c^{(r-1),k}) d^{(r),k}], \sigma^2 I_{2N^k}) \\ &\times N(d^{(r),k}; \mu_r, \Sigma_r) \\ &\propto N(p^k - \mathbb{X}(t^k) \Omega_{r+1}^R E_r c^{(r-1),k}; \\ &\quad \mathbb{X}(t^k) \Omega_{r+1}^R T_r (c^{(r-1),k}) d^{(r),k}, \sigma^2 I_{2N^k}) \\ &\times N(d^{(r),k}; \mu_r, \Sigma_r). \end{aligned}$$

We use $P_{\text{approx}}(m^k | -)$ and $P_{\text{approx}}(d^{(r),k} | -)$ as M-H proposal distributions to sample from their true counterparts. Both are multivariate normals and if necessary, their variance parameters may be tuned to improve sampling efficiency.

B.2 Derivation of the Approximate Deformation-Orienting Matrix, \hat{T}

For visual clarity of the derivation, we will temporarily drop superscripts denoting the resolution r and object index k of each variable. First, we recall from Section 2 that $T(c)$ is a block diagonal matrix with blocks consisting of the rotation matrices R_j , for $j = 1, \dots, J$ (where J is the total number of control points at resolution r). Each R_j rotates its corresponding deformation, d_j , by $\theta_j = \arctan(\frac{H_y(q_j)}{H_x(q_j)})$. Using the identities

$$\cos(\arctan(x/y)) = \frac{y}{\sqrt{x^2 + y^2}}, \quad \sin(\arctan(x/y)) = \frac{x}{\sqrt{x^2 + y^2}},$$

we can write R_j as

$$R_j = \frac{1}{s_j(c)} \begin{bmatrix} H_x(q_j) & -H_y(q_j) \\ H_y(q_j) & H_x(q_j) \end{bmatrix},$$

where $s_j(c) = \sqrt{H_x(q_j)^2 + H_y(q_j)^2}$. This term intuitively represents the “speed” of the curve at parametric position q_j . We approximate this term with the fixed constant $S_j = s_j(c_{\text{prev}})$, where c_{prev} is just the curve sampled in the previous iteration of the M-H sampler. The hodograph, $H(t) = \dot{X}(t)c$, is a linear function of c . So, we can now approximate R_j as a linear function of c :

$$R_j \approx \frac{1}{S_j} \begin{bmatrix} \dot{X}_x(q_j)c & -\dot{X}_y(q_j)c \\ \dot{X}_y(q_j)c & \dot{X}_x(q_j)c \end{bmatrix}.$$

Then, we can write $R_j d_j$ as

$$R_j d_j \approx \hat{R}_j c, \quad \hat{R}_j = \frac{1}{S_j} \begin{bmatrix} (\dot{X}_x(q_j)d_{j,x} - \dot{X}_y(q_j)d_{j,y}) \\ (\dot{X}_y(q_j)d_{j,x} + \dot{X}_x(q_j)d_{j,y}) \end{bmatrix},$$

and finally we define $\hat{T} = \text{block}(\hat{R}_1, \dots, \hat{R}_J)$.

B.3 Conditional Posteriors for μ^r and Σ_r

$$\begin{aligned} P(\mu_r | -) &= N(\hat{\mu}_r, \hat{\Sigma}_{\mu_r}) \\ \hat{\Sigma}_{\mu_r}^{-1} &= \Sigma_{\mu_r}^{-1} + K \Sigma_r^{-1} \\ \hat{\mu}_r &= \hat{\Sigma}_{\mu_r} \left(\Sigma_{\mu_r}^{-1} \mu_{\mu_r} + \Sigma_r^{-1} \sum_{k=1}^K d^{(r),k} \right) \\ P\left(\Sigma_r = \text{diag}(\tau_{1,x}^{(r)}, \tau_{1,y}^{(r)}, \tau_{2,x}^{(r)}, \tau_{2,y}^{(r)}, \dots, \tau_{J,x}^{(r)}, \tau_{J,y}^{(r)})^{-1} | -\right) \\ &= \prod_{j=1}^{J_r} P(\tau_{j,x}^{(r)} | -) P(\tau_{j,y}^{(r)} | -) \\ P(\tau_{j,x}^{(r)} | -) &= Ga(\hat{\alpha}_{j,x}^{(r)}, \hat{\beta}_{j,x}^{(r)}) \\ \hat{\alpha}_{j,x}^{(r)} &= \alpha_\tau + K \quad \hat{\beta}_{j,x}^{(r)} = \beta_\tau + \frac{\sum_{k=1}^K (d_{j,x}^{(r),k} - \mu_{r,j,x})^2}{2}. \end{aligned}$$

B.4 Griddy Gibbs Updates for the Parameterizations t_i^k and Orientation ϕ^k

We discretize the possible values of $t_i^k \in [-\pi, \pi]$ to obtain a discrete approximation of its conditional posterior:

$$P(t_i^k | -) \sim \frac{N(p_i^k; \mathbb{X}(t_i)c^{(R),k}, \sigma^2 I_2)}{\sum_{\tau \in [-\pi, \pi]} N(p_i^k; \mathbb{X}(\tau)c^{(R),k}, \sigma^2 I_2)}.$$

We can make this arbitrarily accurate, by making a finer summation over τ . To achieve quick burn-in, we initialize t_i^k using polar-coordinate

parameterization, where

$$(p_i^k)^k = p_i^k - \bar{p}, \quad t_i^k = \left\lfloor \frac{\tan^{-1}\left(\frac{(p_i^k)^k}{(p_i^k)^k}\right) + \phi^k}{2\pi} \right\rfloor - \pi.$$

The point \bar{p}^k is the average of $\{p_i^k, i = 1, \dots, N\}$ and ϕ^k is the orientation variable defined in Section 4.

We discretize the possible value of $\phi^k \in [-\pi, \pi]$ in a similar manner to obtain the full conditionals.

B.5 Likelihood Contribution from Surface-Normals

Define

$$\dot{X}_x(t_i) = \left[\frac{dB_0^{n_1}(t_i)}{dt}, 0, \frac{dB_1^{n_1}(t_i)}{dt}, 0, \dots, \frac{dB_{2n_1}^{n_1}(t_i)}{dt}, 0 \right] \quad (\text{B.7})$$

$$\dot{X}_y(t_i) = \left[0, \frac{dB_0^{n_1}(t_i)}{dt}, 0, \frac{dB_1^{n_1}(t_i)}{dt}, \dots, 0, \frac{dB_{2n_1}^{n_1}(t_i)}{dt} \right]. \quad (\text{B.8})$$

Proposition B.1. The likelihood contribution of the tangent directions $\theta_i^k, i = 1, \dots, N^k$ ensures conjugate updates of the control points for a multivariate normal prior.

Proof. Recall the noisy tangent direction vectors u_i^k 's and v_i^k 's in (26). Using a simple reparameterization

$$u_i^k = (e_i^k, e_i^k \tan \theta_i^k),$$

where only θ_i 's are observed and e_i 's are not. Observe that

$$v_i^k = (H_x(t_i), H_y(t_i)) = (\dot{X}_x(t_i^k)c^{(3),k}, \dot{X}_y(t_i^k)c^{(3),k}). \quad (\text{B.9})$$

Assuming a uniform prior for the e_i^k 's on \mathbb{R} , the marginal likelihood of the tangent direction θ_i^k given τ^2 and the parameterization t_i^k is given by

$$\begin{aligned} l(\theta_i^k) &= \frac{1}{2\pi\tau^2} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\tau^2} \left\{ (e_i^k - \dot{X}_x(t_i^k)c^{(3),k})^2 \right. \right. \\ &\quad \left. \left. + (e_i^k \tan(\theta_i) - \dot{X}_y(t_i^k)c^{(3),k})^2 \right\} \right] de_i^k, \end{aligned}$$

It turns out the above expression has a closed form given by

$$\begin{aligned} l(\theta_i^k) &= \frac{1}{2\pi\tau^2} \frac{\sqrt{2\pi\tau^2}}{\sqrt{1 + \tan^2(\theta_i^k)}} \exp \left[-\frac{1}{2\tau^2} \left\{ (\dot{X}_x(t_i^k)c^{(3),k})^2 \right. \right. \\ &\quad \left. \left. + (\dot{X}_y(t_i^k)c^{(3),k})^2 - \frac{(\dot{X}_x(t_i^k)c^{(3),k} + \dot{X}_y(t_i^k)c^{(3),k} \tan(\theta_i))^2}{1 + \tan^2(\theta_i^k)} \right\} \right]. \end{aligned}$$

The likelihood for the $\{\theta_i^k, i = 1, \dots, N^k\}$ is given by

$$\begin{aligned} L(\theta_1^k, \dots, \theta_{N^k}^k) &\propto \frac{1}{\tau^{N^k} \prod_{i=1}^{N^k} \sqrt{1 + \tan^2(\theta_i^k)}} \\ &\times \exp \left[-\frac{1}{2\tau^2} (c^{(3),k})' \left\{ \sum_{i=1}^N (S_i^k)' \Gamma_i^k S_i^k \right\} c^{(3),k} \right], \end{aligned}$$

where

$$\Gamma_i^k = \begin{pmatrix} \frac{\tan^2(\theta_i^k)}{1 + \tan^2(\theta_i^k)} & -\frac{\tan(\theta_i^k)}{1 + \tan^2(\theta_i^k)} \\ -\frac{\tan(\theta_i^k)}{1 + \tan^2(\theta_i^k)} & \frac{1}{1 + \tan^2(\theta_i^k)} \end{pmatrix}$$

and $S_i^k = [(\dot{X}_x(t_i^k))', (\dot{X}_y(t_i^k))']$ is a $2(2n_3 + 1) \times 2$ matrix. Clearly, an inverse-Gamma for τ^2 and a multivariate normal prior for the control points are conjugate choices. \square

[Received Jun 2012. Revised May 2014.]

REFERENCES

- Allard, W., Chen, G., and Maggioni, M. (2012), "Multi-Scale Geometric Methods for Data Sets II: Geometric Multi-Resolution Analysis," *Applied and Computational Harmonic Analysis*, 32, 435–462. [1484]
- Archip, N., Jolesz, F., and Warfield, S. (2007), "A Validation Framework for Brain Tumor Segmentation," *Academic Radiology*, 14, 1242–1251. [1489]
- Aziz, N., Bata, R., and Bhat, S. (2002), "Bézier Surface/Surface Intersection," *Computer Graphics and Applications, IEEE*, 10, 50–58. [1481]
- Azzopardi, G., and Petkov, N. (2012), "Contour Detection by CORF Operator," in *Artificial Neural Networks and Machine Learning—ICANN 2012* New York: Springer, pp. 395–402. [1485]
- Barnhill, R. (1985), "Surfaces in Computer Aided Geometric Design: A Survey With New Results," *Computer Aided Geometric Design*, 2, 1–17. [1481]
- Blake, A., and Isard, M. (1998), *Active Shape Models*, New York: Springer. [1481]
- Chan, T. F., and Vese, L. A. (2001), "Active Contours Without Edges," *Image Processing, IEEE Transactions*, 10, 266–277. [1489]
- Désidéri, J., Abou El Majd, B., and Janka, A. (2007), "Nested and Self-Adaptive Bézier Parameterizations for Shape Optimization," *Journal of Computational Physics*, 224, 117–131. [1481]
- Fischler, M. A., and Bolles, R. C. (1981), "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, 24, 381–395. [1481]
- Hagen, H., and Santarelli, P. (1992), "Variational Design of Smooth B-spline Surfaces," *Topics in Surface Modeling*, Philadelphia: Society for Industrial and Applied Mathematics, pp. 85–92. [1481]
- Hastie, T., and Stuetzle, W. (1989), "Principal Curves," *Journal of the American Statistical Association*, 84, 502–516. [1481, 1487]
- Kurtek, S., Klassen, E., Ding, Z., Jacobson, S., Jacobson, J., Avison, M., and Srivastava, A. (2011), "Parameterization-Invariant Shape Comparisons of Anatomical Surfaces," *Medical Imaging, IEEE Transactions*, 30, 849–858. [1490]
- Kurtek, S., Klassen, E., Ding, Z., and Srivastava, A. (2010), "A Novel Riemannian Framework for Shape Analysis of 3D Objects," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference*, IEEE, pp. 1625–1632. [1490]
- Kurtek, S., Srivastava, A., Klassen, E., and Ding, Z. (2012), "Statistical Modeling of Curves Using Shapes and Related Features," *Journal of American Statistical Association*, 107, 1152–1165. [1490]
- Lang, J., and Röschel, O. (1992), "Developable $(1, n)$ -Bézier Surfaces," *Computer Aided Geometric Design*, 9, 291–298. [1481]
- Lievin, M., Delmas, P., Coulon, P.-Y., Luthon, F., and Fristol, V. (1999), "Automatic Lip Tracking: Bayesian Segmentation and Active Contours in a Cooperative Scheme," in *Multimedia Computing and Systems, 1999. IEEE International Conference on*, Vol. 1, IEEE, pp. 691–696. [1481]
- Mokhtarian, F., and Mackworth, A. (1992), "A Theory of Multiscale, Curvature-based Shape Representation for Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 789–805. [1481]
- Mortenson, M. (1985), *Geometric Modeling*, New York: Wiley. [1481]
- Müller, H. (2005), "Surface Reconstruction—An Introduction," in *Scientific Visualization Conference, 1997, IEEE*, p. 239. [1481]
- Pati, D., and Dunson, D. (2014), "Bayesian Modeling of Closed Surfaces Through Tensor Products," *Journal of Machine Learning Research*, accepted manuscript. [1481, 1485, 1492]
- Persoon, E., and Fu, K. (1977), "Shape Discrimination Using Fourier Descriptors," *Systems, Man and Cybernetics, IEEE Transactions*, 7, 170–179. [1481]
- Róth, Á., Juhász, I., Schicho, J., and Hoffmann, M. (2009), "A Cyclic Basis for Closed Curve and Surface Modeling," *Computer Aided Geometric Design*, 26, 528–546. [1482]
- Srivastava, A., and Jermyn, I. (2009), "Looking for Shapes in Two-dimensional Cluttered Point Clouds," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 31, 1616–1629. [1491]
- Srivastava, A., Klassen, E., Joshi, S., and Jermyn, I. (2011), "Shape Analysis of Elastic Curves in Euclidean Spaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 99, 1–1. [1481]
- Stepanets, A. (1974), "The Approximation of Certain Classes of Differentiable Periodic Functions of Two Variables by Fourier Sums," *Ukrainian Mathematical Journal*, 25, 498–506. [1492]
- Su, B., and Liu, D. (1989), *Computational Geometry: Curve and Surface Modeling*, San Diego, CA: Academic Press Professional, Inc. [1481]
- Su, J., Dryden, I., Klassen, E., Le, H., and Srivastava, A. (2012), "Fitting Optimal Curves to Time-Indexed, Noisy Observations of Stochastic Processes on Nonlinear Manifolds," *Image and Vision Computing*, 30, 428–442. [1481]
- van der Vaart, A., and van Zanten, J. (2008), "Reproducing Kernel Hilbert Spaces of Gaussian Priors," *IMS Collections*, 3, 200–222. [1485]
- Whitney, H. (1937), "On Regular Closed Curves in the Plane," *Compositio Math*, 4, 276–284. [1481]
- Xie, Y., Huang, J., and Willett, R. (2012), "Multiscale Online Tracking of Manifolds," *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 620–623. [1484]
- Zahn, C., and Roskies, R. (1972), "Fourier Descriptors for Plane Closed Curves," *Computers, IEEE Transactions*, 100, 269–281. [1481, 1483]