A statistical model for signature verification

Ian W. McKeague

May 14, 2004

Abstract

A Bayesian model for off-line signature verification involving the representation of a signature through its curvature is developed. The prior model makes use of a spatial point process for specifying the knots in an approximation restricted to a buffer region close to a template curvature, along with an independent time warping mechanism. In this way, prior shape information about the signature can be built into the analysis. The observation model is based on additive white noise superimposed on the underlying curvature. The approach is implemented using MCMC and applied to a collection of documented instances of Shakespeare's signature.

Key words and phrases: Bayesian nonparametric regression, biometric identification, functional data analysis, shape theory, spatial point processes, time warping.

1 Introduction

People are able to recognize their own handwriting or signature at a glance, and manuscript experts can usually determine genuineness with almost scientific exactitude (or so they claim). The analysis of variations in handwriting style is useful, for example, in forensic investigations, biometric identification, and in the dating of ancient manuscripts. In a study of Shakespeare's handwriting, Charles Hamilton (1985, p. 7) ascribed the ability of recognizing variations in handwriting to the "feel" of a script, to the "sum total of the viewer's knowledge, the infusion of intuition and an immense amount of experience." Manuscript experts often assess the feel of documents very quickly by examining them upside down, so the words themselves become obscured. The "shape" of the handwriting is therefore a key ingredient in such analysis. However, a model-based statistical approach to off-line signature analysis and verification is not yet available. The present paper approaches this problem by formulating a Bayesian shape model for planar curves.

[1]Part of this work was supported by NSF Grants DMS-0204688 and DMS-9971784, and NSA Grant 19984075. The author thanks Regina Liu, Anuj Srivastava, Jim Ramsay, and the referees for helpful comments and encouragement.

Signature verification is a difficult but potentially tractable classification problem, and the discovery of a suitable statistical model would have substantial scientific interest. The key feature of an efficient classification scheme is data condensation, or the reduction of often unmanageably large data sets to a parsimonious form, without the sacrifice of key statistical information. This can lead to the discovery of interesting anomalies, given the availability of adequate data and suitable statistical methods. The instances of the initial letter in Shakespeare's signature in Figure 1 have characteristic statistical structures which could be exploited if the key information can be efficiently condensed into a suitable low-dimensional object (indeed, this may be how the visual cortex processes such information). Shape is the most essential feature, and, while temporal information (e.g., acceleration) is not directly available for off-line analysis, the model should reflect the temporal generation of the observed curves.

Grenander (1970, 1993) introduced the theory of deformable templates to address the problem of recognizing an interesting shape (e.g., hand, galaxy, mitochondrion) in a graylevel image. The shape is described by a flexible template, typically a polygon, with edge lengths and angles governed by a joint prior distribution, typically a Markov chain. The theory has been implemented, e.g., for hand x-rays (Amit, Grenander and Piccioni 1991), and for MRI images of the brain (Miller, Christensen, Amit and Grenander 1993). In representing graylevel images, such as those in Figure 1, one expects to be able to lower the dimensionality considerably because of statistical regularities. This dimension reduction is a crucial first step for any successful high-level image understanding and analysis. Principal components analysis is a useful tool for handwritten digit compression (Hastie, Tibshirani and Friedman 2001, p. 488). Local feature analysis (Penev and Atick 1996) provides compact representations of images (e.g., of human faces) in terms of statistically derived local features and their point locations. There is an extensive statistical theory of shapes based on landmarks (Small 1996; Dryden and Mardia 1998; Lele and Richtsmeier 2000); shape in this approach is treated as the information in an object (specified by planar landmarks) that is invariant under translations, rotations, and isotropic rescalings. Shape representation of planar curves based on curvature has been used extensively in computer vision (Mokhtarian and Mackworth 1992; Mumford 1994; Klassen, Srivastava, Mio and Joshi 2004).

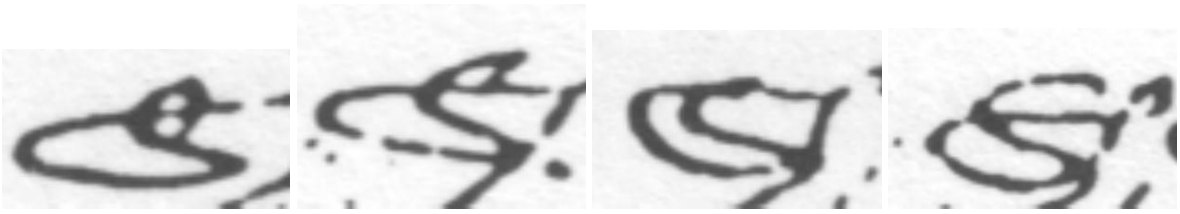

Figure 1: The initial 'S' in Shakespeare's signature from the Welcombe Enclosure Agreement, October 28, 1614, see Hamilton (1985, p. 195).

Despite much progress in signature verification in recent years, a fully-fledged statistical approach has not yet been developed. There is an extensive literature on the computerized recognition of signatures, see Plamondon and Lorette (1989) and Munich and Perona (2003) for references. Much of this work is based on numerical methods designed to extract similarity features (from data provided by on-line or off-line signature digitizers). For example, Sabourin,

Drouhard and Wah (1997) investigate the use of shape matrices as a global shape factor for off-line signature verification. Martens and Claesen (1997) present an on-line signature verification system based on 3D force patterns and pen inclination angles, as recorded during signing. The feature extraction mechanism is based on an elastic matching technique. Lee (1996) develops several neural network based approaches to on-line signature verification. Matsuura and Sakai (1996) consider a stochastic system representation of the handwriting process and its application to on-line signature verification. The stochastic system characterizing the motion in the writing of the signature is described by a random impulse response.

Progress on the problem of signature verification has advanced less rapidly in off-line applications, in part because information such as pen position and velocity is unavailable. Off-line systems using similarity features have been studied experimentally; e.g., Deng, Liao, Ho and Tyan (1999) proposed wavelet analysis of curvature data, and Fang et al. (2003) developed a system to track variations in one-dimensional projection profiles and stroke positions. Some progress has been been made on the recognition problem (i.e., reading) for off-line cursive handwriting, see, e.g., Abuhaiba, Holt and Datta (1994, 1998), who use an algorithm for extracting a polygonal approximation of the handwriting from a graylevel image, and Proctor, Illingworth and Mokhtarian (2000). Machine learning techniques are useful for character recognition from off-line handwritten text based on large training sets of graylevel images (Hastie et al. 2001).

Some statistical modeling approaches to on-line signature verification have been proposed. Hastie, Kishon, Clark and Fan (1991) model the signature trace in the x–y plane as a time-warped and spatially transformed version of a given template signature, and employ cubic spline smoothing techniques to fit the model. When a new signature arrives for verification, it is compared with an optimal affine transformation of the template signature. Kashi, Hu, Nelson and Turin (1997) continued this line of work and demonstrated improved performance by the addition of a signature likelihood feature based on a hidden Markov model. Ramsay and Silverman (1997) use on-line handwriting as a key example in their development of functional data analysis, see also Ramsay (2000).

We investigate the problem of off-line signature verification from a Bayesian perspective, which is a natural approach because of the limited training data and the availability of important prior information about shape. From each graylevel image of a smooth segment of the signature, a spline approximation is extracted (thickness of the handwriting will be ignored). This reduces the problem to finding a suitable stochastic process model to describe the observed smooth planar curves in terms of shape. The key aspects of the approach are an observation model involving white noise superimposed on an underlying curvature, a prior distribution for curvature functions to represent shape information, and a time warping mechanism to register each observed signature. An MCMC algorithm is used to explore the posterior distribution of the curvature and time warping functions, which is feasible because these functions are relatively low dimensional objects.

Various Bayesian nonparametric curve fitting methods are available in the literature. Wahba (1978) shows that smoothing splines result from a type of Bayesian polynomial regression analysis, also see Eubank (1988). The prior on the regression function in this case is a polynomial with independent random coefficients plus an independent multiply-integrated Weiner process. Another approach is given by Silverman (1985), in which the prior for the regression function

is concentrated on the space of natural cubic splines with knots at the design points. For more on Bayesian approaches to nonparametric regression, see Wahba (1983), Barry (1986), Cox (1993), and Denison, Mallick and Smith (1998). These approaches are not directly useful for analyzing signatures, however, because it is not clear how information about the shape of a signature could be expressed via a regression function.

The paper is organized as follows. After some preliminary material on planar curves, the proposed observation model is introduced in Section 2. The prior for the various unknown parts of the observation model is then specified along with an MCMC method for sampling from the posterior. Signature verification is discussed at the end of Section 2. An application to the analysis of Shakespeare's signature is given in Section 3, and some concluding remarks in Section 4. Proofs of the key results can be found in the Appendix.

2 Statistical model

In this section we introduce the proposed Bayesian model. In the observation part of the model, a realized signature will be specified by adding white noise to the curvature. The prior on the space of curvature functions will use a spatial point process to specify the knots in an approximation restricted to a buffer region close to a template curvature, along with a time warping mechanism.

2.1 Preliminaries

We begin by reviewing some background and terminology on the geometry of plane curves (Carmo 1976). Consider a parameterized curve $\alpha : [a, b] \to \mathbb{R}^2$ with coordinate representation $\alpha(t) = (x(t), y(t))^T$ and assume that the curve is smooth in the sense that $x(t)$ and $y(t)$ are twice continuously differentiable. The trace of α is the image set $\alpha([a, b])$, and, at a given $t \in [a, b]$, the velocity (or tangent) vector is $V(t) = \alpha'(t) = (x'(t), y'(t))^T$, and the arc length is $L(t) = \int_a^t |\alpha'(u)|\, du$, where $|\cdot|$ is euclidean norm. If the velocity vector does not vanish, the curve is said to be regular; such curves can be re-parameterized by arc length as $\beta(s) = \alpha(L^{-1}(s))$, $s \in [0, L(b)]$, without changing the trace. The velocity vector of an arc-length parameterized curve has unit norm.

Curvature measures the rate at which a regular curve pulls away from its tangent line, and, for an arc-length parameterized curve, is defined by

$$\kappa(t) = \langle \alpha''(t), \mathbf{n}(t) \rangle \;\; = \;\; x'(t)y''(t) - x''(t)y'(t), \tag{2.1}$$

where $\mathbf{n}(t) = (-y'(t), x'(t))^T$ is a unit normal vector. Curvature has the attractive feature that it is invariant under location shifts and rotations, and gives a unique representation of the curve up to rigid motion. The velocity vector of an arc-length parameterized curve is represented by

$$V(t) \;\; = \;\; \begin{pmatrix} \cos\theta(t) \\ \sin\theta(t) \end{pmatrix}, \qquad \theta(t) = \theta(a) + \int_a^t \kappa(s)\, ds, \tag{2.2}$$

where $\theta(t)$ is the angle of the velocity vector with the x-axis. The curve can be reconstructed (up to a location shift) from its velocity vector by $\alpha(t) = \alpha(a) + \int_a^t V(s)\, ds$.

Another important property of curvature is that it is inversely proportional to scale: the curve $c\alpha(t)$ has curvature $\kappa(t)/c$, for $c > 0$. We therefore need a standardized form for the observed signatures so their curvatures are directly comparable without the interference of scale effects. We do this by re-scaling each observed signature to have a fixed total arc-length $\tau > 0$, so each arc-length parameterized signature has curvature function belonging to $C[0,\tau]$, the space of continuous functions on $[0,\tau]$. Such invariance under length scaling has long been a fundamental principle of signature verification, see, e.g., Lew (1980, p. 29).

2.2 Observation model

In this section we introduce the signature observation model and derive the likelihood. The problem of estimating curvature nonparametrically is shown to be equivalent to estimating the drift in a Gaussian shift experiment.

The model is first formulated for a single observed signature. The graylevel image of the signature is preprocessed to extract a smooth arc-length parameterized approximation $\alpha_{\mathrm{obs}}(t)$, $t \in [0,\tau]$, and its velocity vector $V(t) = \alpha'_{\mathrm{obs}}(t)$ computed numerically (see Section 3 for details). The observed velocity is modeled as

$$V(t) \;=\; \begin{pmatrix} \cos\theta(t) \\ \sin\theta(t) \end{pmatrix}, \qquad \theta(t) = \varphi + \int_0^t \kappa(s)\,ds + \sigma W(t), \tag{2.3}$$

where $\kappa \in C[0,\tau]$ is an underlying curvature function, $W(t)$ is a standard Brownian motion and φ is the angle of $V(0)$ with the x-axis. Here $\theta(t)$ represents the angle resulting from additive white noise with variance $\sigma^2 > 0$ superimposed on the underlying curvature. The noise locally perturbs the curve along the direction of its normal vector.

The likelihood for κ (given that σ is known) can be obtained as the Radon–Nikodym derivative of the distribution of the process V with respect to a certain dominating measure, for which we need to assume that continuous observation of V over $[0,\tau]$ is available (even though the numerical implementation will only involve observation on a fine grid). In the Appendix we show that this leads to the log-likelihood

$$\begin{aligned} \ell(\kappa|V) \;=\; & \frac{1}{\sigma^2}\int_0^\tau \kappa(s)\,\{V_1(s)\,dV_2(s) - V_2(s)\,dV_1(s)\} \\ & -\frac{1}{2\sigma^2}\int_0^\tau \kappa(s)^2\,ds + \frac{1}{8}\sigma^2\tau. \end{aligned} \tag{2.4}$$

For numerical evaluation of the likelihood, the velocity vector $V(t)$, $t \in [0,\tau]$ will be approximated over a regular grid $s_j = j\tau/N$, $j = 0,\ldots,N$. The first term in the log-likelihood is an Itô integral of the form $\int_0^\tau \xi(s)\,dV(s)$ with a predictable integrand $\xi(s)$. This integral is approximated by the finite sum:

$$\sum_{j=1}^{N} \xi(s_{j-1})(V(s_j) - V(s_{j-1})) \;\;\to\;\; \int_0^\tau \xi(s)\,dV(s) \tag{2.5}$$

in quadratic mean as $N \to \infty$, see Liptser and Shiryayev (1977, 4.2). The second integral is approximated by a Riemann sum over the same grid s_j. This results in an approximation to $\ell(\kappa|V)$ which is suitable for use with the data extracted from the graylevel image.

Gaussian shift experiment

The problem of recovering the curvature κ in the case of known noise level σ can be reduced to a Gaussian shift experiment (or white noise with drift). This is already suggested by the form of the log-likelihood in (2.4), which agrees with the white noise model with drift κ. In the Appendix we show that

$$dZ(t) \;=\; \kappa(t)\,dt + \sigma dW(t), \quad t \in [0,\tau], \tag{2.6}$$

where

$$Z(t) = \int_0^t \left\{V_1(s)\,dV_2(s) - V_2(s)\,dV_1(s)\right\},$$

so we have the Gaussian shift experiment with drift κ, noise level σ and observation of Z. Many classical problems have been shown to be equivalent (in terms of asymptotic minimax risk) to the white noise model as $\sigma \to 0$, e.g., nonparametric regression (Brown and Low 1996) and density estimation (Nussbaum 1996). The expression for the log-likelihood (2.4) indicates that the process Z is a sufficient statistic for κ.

A kernel estimator of κ is given by

$$\hat{\kappa}(t) = \frac{1}{b}\int_0^\tau K\left(\frac{t-s}{b}\right)\,dZ(s),$$

where $b > 0$ is the bandwidth and K is a kernel function which integrates to 1. This estimator agrees with the multiscale curvature function representation of Mokhtarian and Mackworth (1992). A wavelet shrinkage estimator is preferable to the kernel estimator when a spiky curvature function is expected.

Under continuous observation of V, the noise level σ is determined because

$$\frac{1}{\tau}\sum_{j=1}^{N}\left\{(V_1(s_j) - V_1(s_{j-1}))^2 + (V_2(s_j) - V_2(s_{j-1}))^2\right\} \;\to\; \sigma^2 \tag{2.7}$$

in probability as $N \to \infty$, which follows from expressions for the optional quadratic variation processes of the components of V, see the Appendix. Moreover, any dominating measure for the likelihood must depend on σ because the distributions μ_σ of the process V corresponding to different values of σ are mutually singular: for any positive noise levels $\sigma_1 \neq \sigma_2$, there exists a subsequence of (2.7) that converges almost surely under both noise levels, so μ_{σ_1} and μ_{σ_2} have disjoint supports. Thus, it is not possible to derive a likelihood based on continuous observation of V unless σ is assumed known.

Although σ is determined from (2.7), the rate of convergence is far too slow for practical use: sharp curves in the signature will cause high *positive* bias in the estimates. Wavelet thresholding comes to mind as a way of controlling this problem, but unfortunately there is a high *negative* bias in this case. The Haar wavelet coefficients of Z are integrals of the form $\int \psi(s)\,dZ(s)$, where ψ is piecewise constant on intervals with dyadic endpoints. These coefficients can be approximated from the discrete data using sums of the form (2.5). Taking the median absolute deviation of the wavelet coefficients at the finest scale of resolution and dividing by 0.6745

gives a crude estimate $\hat{\sigma}$, see Donoho and Johnstone (1995, p. 1218). Although $\hat{\sigma}$ works well for calibrating wavelet shrinkage, simulation experiments show that it grossly underestimates σ when $V(t)$ is extracted from traces that are generated with noise level σ, and it fails to reflect the large variation in shapes seen in real signatures. Indeed, this is to be expected because the approximation procedure used to extract $V(t)$ from the trace is a smoothing operation that removes much of the noise.

A practical way of selecting σ would be through experimental studies, where it could be chosen to optimize performance (in terms of error rates say) of the signature verification procedure. For simplicity, however, in the present paper we treat σ as a fixed tuning parameter.

Time warping

We next extend our model to the case of a sample of n signatures with observed velocity vectors $V^{(i)}$, $i = 1, \ldots, n$. Sharp peaks in the curvature are likely to differ in timing between the different records, so we need a way to give greater flexibility to the model rather than assuming identical curvature functions. This can be done using an increasing time warping function $h_i \colon [0, \tau] \to [0, \tau]$ specific to each observed signature to re-parameterize a given (baseline) curvature κ. The curvature for signature i is then $\kappa_i(t) = \kappa(h_i(t))$. We refer to κ as the *baseline* curvature function to distinguish it from κ_i.

The time warping function h_i is non-identifiable, for note that a constant baseline curvature has the same κ_i irrespective of h_i. We handle this problem by regarding the time warp as a latent variable or random effect (cf. mixed models), which, in our Bayesian framework, is equivalent to specifying a prior distribution for h_i. In the next subsection, this prior will be specified conditionally on κ, and will depend on the data in such a way that the velocity vector corresponding to κ_i favors close agreement with $V^{(i)}$.

Full likelihood

Given a sample of n signatures with velocity vectors $V^{(i)}$, the full log-likelihood for the corresponding vector of time warp functions $\boldsymbol{h} = (h_i, i = 1, \ldots, n)$ and baseline curvature function κ is the sum of the respective log-likelihoods:

$$\ell(\kappa, \boldsymbol{h} | \text{data}) = \sum_{i=1}^{n} \ell(\kappa_i | V^{(i)}),$$

assuming that the signatures are sampled independently from the observation model.

2.3 Prior specification

We design the prior to have the following properties:

- realizations of the trace are consistent with known features of the signature shape;
- flexible enough to represent a wide variety of possible signatures;

- gives a parsimonious representation of the baseline curvature and time warping functions;
- MCMC is feasible for sampling from the posterior distribution.

The construction has two main parts: a prior baseline curvature process for embedding the shape information, and an independent record-specific time warping mechanism for registering each curve. The proposed approach is by no means the only way of satisfying the above properties, but we believe it provides an intuitively reasonable formulation, well suited to off-line signature verification.

Baseline curvature process

The baseline curvature process is defined to have paths contained in a certain buffer region centered on a template curvature.

We start with a template curvature function κ_0, which in practice is derived from a signature having a shape representative of the form of the observed signatures, or from an estimate (such as $\hat{\kappa}$) of the curvature function of one of the observed signatures. A buffer region is placed around the template as follows:

$$\mathcal{B} = \{(t, y) : \kappa_L(t) \leq y \leq \kappa_U(t),\ t \in [0, \tau]\},$$

where $\kappa_L(t) = \kappa_0(t) - \epsilon$ and $\kappa_U(t) = \kappa_0(t) + \epsilon$, with $\epsilon > 0$ a user-specified constant that controls the size of the buffer region. The bracketting functions κ_L and κ_U represent imposed bounds on perturbations of the template curvature. Hard bounds are used here to ensure that the realized traces remain close in shape to that of the template. In practice, inspection of traces generated by curvature functions inside the buffer region (as under the prior for κ defined below) can be used to guide the choice of ϵ; we found that $\epsilon = 1.5$ works well. The hard bounds are also useful for restricting the scope of proposals in the MCMC implementation.

The baseline curvature process is constrained to go through points (knots) in the buffer region. These knots are specified using a spatial point process X contained in $\mathcal{B}$. For background on spatial point processes, see van Lieshout (2000). Here we take X to be a Strauss process, i.e., X is a random finite set of points in $\mathcal{B}$ having unnormalized density $f(\mathbf{x}) = \beta^{n(\mathbf{x})}\gamma^{d(\mathbf{x})}$ with respect to the distribution of a unit rate Poisson process in $\mathcal{B}$, where $\beta > 0$, $0 < \gamma \leq 1$ are tuning parameters, $d(\mathbf{x})$ is the number of unordered pairs of points in $\mathbf{x}$ within a metric-distance ρ of each other, and $n(\mathbf{x})$ is the number of points in $\mathbf{x}$. The metric is taken as the distance between time coordinates (rather than euclidean distance in $\mathbb{R}^2$). The Strauss process has repulsive pairwise interaction, so the knots tend to be well separated in the time direction, providing an easy way of controlling (through γ and ρ) the prior roughness of the baseline curvature process. List the points in X in order as $(t_j, y_j), j = 1, \ldots, n(X)$, where $0 < t_1 < t_2 < \ldots < t_{n(X)} < \tau$, and set $(t_0, y_0) = (0, \kappa_0(0))$, $(t_{n(X)+1}, y_{n(X)+1}) = (\tau, \kappa_0(\tau))$. Then we define the baseline curvature process by

$$\kappa(t) = \left(\frac{t_{j+1} - t}{t_{j+1} - t_j}\right)\{\kappa_0(t) + y_j - \kappa_0(t_j)\} + \left(\frac{t - t_j}{t_{j+1} - t_j}\right)\{\kappa_0(t) + y_{j+1} - \kappa_0(t_{j+1})\}$$

for $t_j \leq t \leq t_{j+1}$, $j = 0, 1, \ldots, n(X) + 1$. The baseline curvature function represents a smooth departure from κ_0 passing through the knots and contained in the buffer region. The influence

of the knot (t_j, y_j) on the baseline curvature at t is proportional to the distance of t from t_j relative to its distance from t_{j+1}.

Time warping processes

Each time warping process is modeled as

$$h_i(t) = \min\{\tilde{h}_i(t), \tau\}, \;\; t \in [0, \tau],$$

where $\tilde{h}_i$ is an increasing, continuous, piecewise-linear function with $\tilde{h}_i(0) = 0$. We take $\tilde{h}_i$ to be linear between the grid points $u_j = j\tau/p$, $j = 0, \ldots, p$, where $p \geq 1$ is user-specified. The parameters $\theta_{ij} = \tilde{h}_i(u_j) - \tilde{h}_i(u_{j-1})$, $j = 1, \ldots, p$ determine h_i, and without loss of generality may be constrained so that $0 \leq \theta_{ij} \leq \tau$. We then represent $\boldsymbol{h}$ by combining the θ_{ij} into a vector $\boldsymbol{\theta} \in \Theta = [0, \tau]^{np}$. Hastie et al. (1991) and Ramsay (2000) also used piecewise-linear time warping functions.

The functional data analysis method of landmark registration can be adapted to provide a prior on h_i (equivalently on the p components of $\boldsymbol{\theta}$ that represent h_i). For example, the landmarks could be specified as the times t of sharp changes in the direction of the tangent vector $V^{(i)}(t)$, with $h_i(t)$ approximating the corresponding time of peak curvature in the template. Such landmarks are typically located close to the zero crossings of the x-component the velocity vector, see Figure 5. Denote the time points (landmarks) in the template and the ith record to be matched by t_{0r} and t_{ir}, $r = 1, \ldots, m$, respectively. Then the prior density on h_i (in terms of its parameters) is specified by $\pi(h_i) \propto \exp\{-\eta J(h_i)\}$, where

$$J(h_i) = \sum_{r=1}^{m} (h_i(t_{ir}) - t_{0r})^2$$

is the 'cost' function for mis-timing the landmarks and $\eta > 0$ is a user-specified precision parameter.

An alternative to landmark registration is curve registration in which departures of the data from the fitted velocity vectors are penalized: replace the above cost function by

$$J(h_i|\kappa) = \int_0^\tau \text{angle}(V^{(i)}(t), V^{(i)}_{\text{fitted}}(t))\, dt,$$

where $V^{(i)}_{\text{fitted}}(t)$ is the velocity vector corresponding to the curvature function $\kappa(h_i(t))$, with initial value $V^{(i)}_{\text{fitted}}(0) = V^{(i)}(0)$, and $\text{angle}(u, v) = \cos^{-1}(u^T v)$ is the angle in radians between unit vectors u, v. This is essentially equivalent to the Procrustes method of curve registration, although in that case the time warping is applied to the records themselves: a cost function for the differences between the time-warped records and an overall mean record is iteratively minimized, see Ramsay and Silverman (1997, Section 5.4). For the on-line problem, Hastie et al. (1991) used time warping of a template signature in the x–y plane, whereas Ramsay and Silverman (1997) and Ramsay (2000) used time warping of acceleration records; time warping of the baseline curvature is more natural in the present setting.

2.4 Sampling from the posterior

The posterior is the joint conditional distribution (given the data) of the baseline curvature κ (represented by its knots $\mathbf{x}$) and the vector $\boldsymbol{h}$ of time warping functions (represented by $\boldsymbol{\theta}$). By Bayes formula, the posterior density is proportional to

$$g(\mathbf{x}, \boldsymbol{\theta}) = \exp\{\ell(\kappa, \boldsymbol{h}|\text{data})\} f(\mathbf{x}) \prod_{i=1}^{n} \pi(h_i|\kappa, \text{data}),$$

where the dominating measure is the product of the unit-rate Poisson distribution on $\mathcal{B}$ and Lebesgue measure on Θ.

The MCMC scheme we use to explore this posterior is a hybrid sampler having separate Metropolis–Hastings updates for each component of $(\mathbf{x}, \boldsymbol{\theta})$. Such samplers are also known as 'Metropolis-within-Gibbs' or 'variable-at-a-time Metropolis–Hastings'; for background see, e.g., Robert and Casella (1999, p. 322). Geometric ergodicity can be difficult to establish for hybrid samplers (e.g., Fort, Moulines, Roberts and Rosenthal 2003), but it is quite easily shown in our case, essentially due to the compactness of Θ. A similar type of hybrid sampler has recently been studied by Blackwell and Møller (2003, Section 3.3).

We use a deterministic scan (or sweep) to cycle through the components, each proposal having acceptance probability of the form $\min(\alpha, 1)$, where α (the Hastings ratio) depends on current state and the proposed state. Each component θ of $\boldsymbol{\theta}$ is updated (in turn) by a random walk Metropolis move with the proposal θ' normally distributed with mean θ (except that θ' is restricted to the compact interval $[0, \tau]$). The Hastings ratio for a move of this form, $(\mathbf{x}, \boldsymbol{\theta}) \to (\mathbf{x}, \boldsymbol{\theta}')$, is $\alpha_{rw}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\theta}') = g(\mathbf{x}, \boldsymbol{\theta}')/g(\mathbf{x}, \boldsymbol{\theta})$. The $\mathbf{x}$-component is updated using the Metropolis–Hastings algorithm of Geyer and Møller (1994): either a point is born and added to $\mathbf{x}$, or a point in $\mathbf{x}$ is killed (unless $\mathbf{x}$ is empty), each step having equal probability. In the birth step, we generate ξ at a uniform random location in $\mathcal{B}$ and set $(\mathbf{x}, \boldsymbol{\theta}) \to (\mathbf{x} \cup \{\xi\}, \boldsymbol{\theta})$ with the acceptance probability having Hastings ratio given by

$$\alpha_b(\mathbf{x}, \boldsymbol{\theta}, \xi) = \frac{|\mathcal{B}|}{n(\mathbf{x}) + 1} \frac{g(\mathbf{x} \cup \{\xi\}, \boldsymbol{\theta})}{g(\mathbf{x}, \boldsymbol{\theta})},$$

where $|\mathcal{B}| = 2\epsilon\tau$ is the area of the buffer region. In the death step, we select ξ at random from the points in $\mathbf{x}$ and set $(\mathbf{x}, \boldsymbol{\theta}) \to (\mathbf{x} \setminus \{\xi\}, \boldsymbol{\theta})$ with the Hastings ratio given by

$$\alpha_d(\mathbf{x}, \boldsymbol{\theta}, \xi) = \frac{n(\mathbf{x})}{|\mathcal{B}|} \frac{g(\mathbf{x} \setminus \{\xi\}, \boldsymbol{\theta})}{g(\mathbf{x}, \boldsymbol{\theta})}.$$

Theorem A.1 in the Appendix shows that this hybrid sampler is geometrically ergodic.

2.5 Signature verification

Given samples of an individual's signature, how can we recognize whether a new signature comes from the same person? In this section we propose a formal test procedure calibrated via simulation of the fitted observation model.

The new signature is first scaled to have length τ, and, to make its velocity vector V^{new} directly comparable with each sample signature $V^{(i)}$, it is rotated (clockwise) to produce a velocity vector V_i^{new} having initial direction $V^{(i)}(0)$, $i = 1, \ldots, n$. Evidence that the new signature is a forgery is then provided by large values of the 'forgery index'

$$F = \frac{1}{\pi\tau} \min_{i=1,\ldots,n} \int_0^\tau \text{angle}(V_i^{\text{new}}(t), V^{(i)}(t))\, dt,$$

which is a measure of the difference in shape between the new signature and the closest-matching sample signature. We have normalized F to range between 0 and 1, and we expect it to be small for authentic signatures. In the language of hypothesis testing, F is a test statistic for the null hypothesis that the new signature is authentic.

The strength of evidence that the new signature is a forgery can be assessed by the p-value $P(F^* > F)$ that a velocity vector V^* sampled from the fitted observation model has a forgery index F^* greater than the observed F; here F^* is defined by replacing V^{new} by V^* in F. This p-value can be found using bootstrap simulation: V^* is generated from the observation model (2.3) with the curvature function κ replaced by $\hat{\kappa}(\hat{h}_I(\cdot))$, where I is randomly selected from $i = 1, \ldots, n$, and $\hat{\kappa}$, $\hat{h}_i$ are the posterior means of the baseline curvature, and the time warping h_i, respectively. The p-value is then approximated by the proportion of bootstrap draws with forgery index larger than F. We illustrate this method of signature verification in the next section.

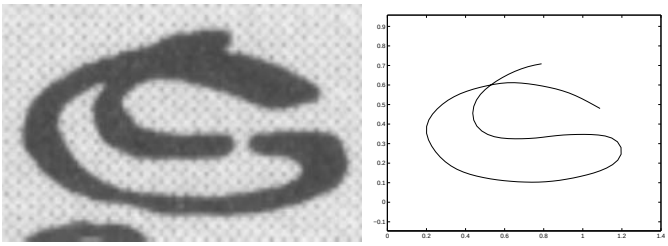

Figure 2: Template trace. Secretary hand 'S' from a book on penmanship published in 1571 when Shakespeare was 7, possibly used by pupils at Stratford Grammar School (see Hamilton 1985, p. 12).

3 Application

The style of handwriting taught in Elizabethan grammar schools is known as Secretary hand. The letter 'S' from a Secretary hand alphabet is illustrated in Figure 2, and provides a possible template to model the instances of the initial 'S' in Shakespeare's signature in Figure 1.

The graylevel images first need to be preprocessed to extract the velocity data. The Matlab command `ginput` records the coordinates of successive cursor locations in response to mouse clicks, and was used to extract the locations of a sequence of regularly spaced points along each curve. About 20 such points were selected for each signature. Then cubic spline interpolation (with the command `interp1`) over a regular grid of $N = 100$ points was applied separately to the x- and y-components. The componentwise derivatives were then found using the `gradient` command, and converted into arc-length parameterization; the results are plotted in Figure

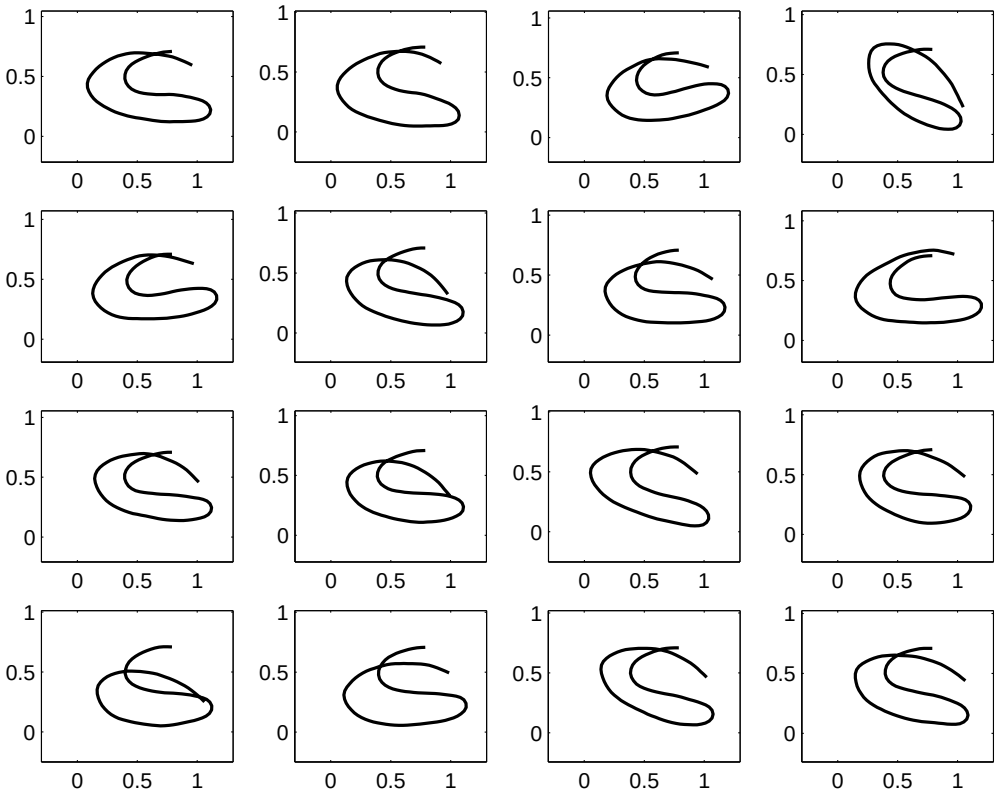

Figure 3: Simulated traces from the observation model with $\sigma = 0.15$ under the template curvature κ_0 and no time warping.

5. This provides a satisfactory approximation to the velocity vector: reconstructions of the curves (see the first row of Figure 7) show good visual agreement with the original signatures in Figure 1. The template was treated in the same way, and its curvature computed using the discrete version of (2.1) applied to its velocity vector (see the last column of Figure 5). The right side of Figure 2 shows the reconstruction of the template trace from its curvature, and is in good visual agreement with the shape of the printed letter. All curves are scaled to have the same arc-length $\tau = 3.56$ as the template trace.

Figure 3 shows 16 traces simulated under the observation model with $\kappa = \kappa_0$ (template curvature) and noise level $\sigma = 0.15$. This noise level appears to provide sufficient variation in shape, and apart from time warping the traces are consistent with the original signatures in Figure 1. In fitting the model later in this section we specify $\sigma = 0.15$; equivalently, $\sigma = 0.28/\sqrt{\tau}$ for other values of τ. A much higher noise level would not be suitable; the traces in Figure 4 were generated using $\sigma = 0.2$ (i.e., $\sigma = 0.37/\sqrt{\tau}$) and appear to be unrealistically distorted.

Figure 5 shows that the velocity vectors in the data appear to be noisy versions of (a time warped version of) the template velocity vector, as required by the model. The zero crossings of the first component of the velocity vector $V_1(t)$ match closely the times of extreme curvature, and could be used to locate landmarks to implement time warping by landmark registration.

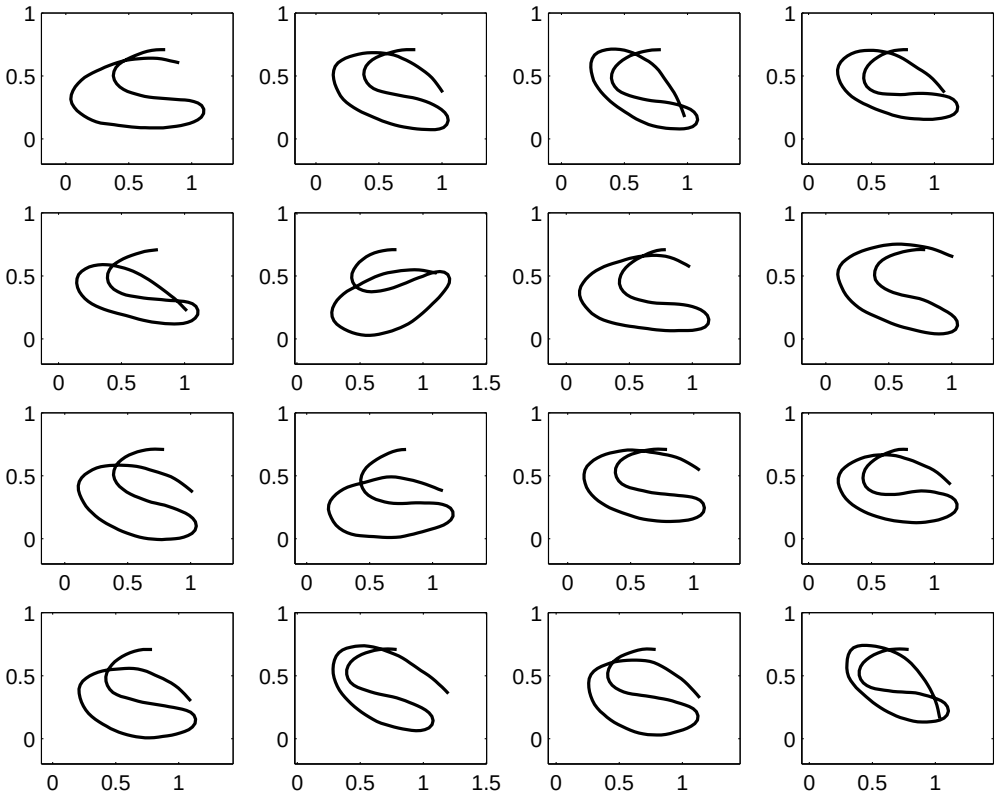

Figure 4: Simulated traces as in Figure 3 but with higher noise level $\sigma = 0.2$.

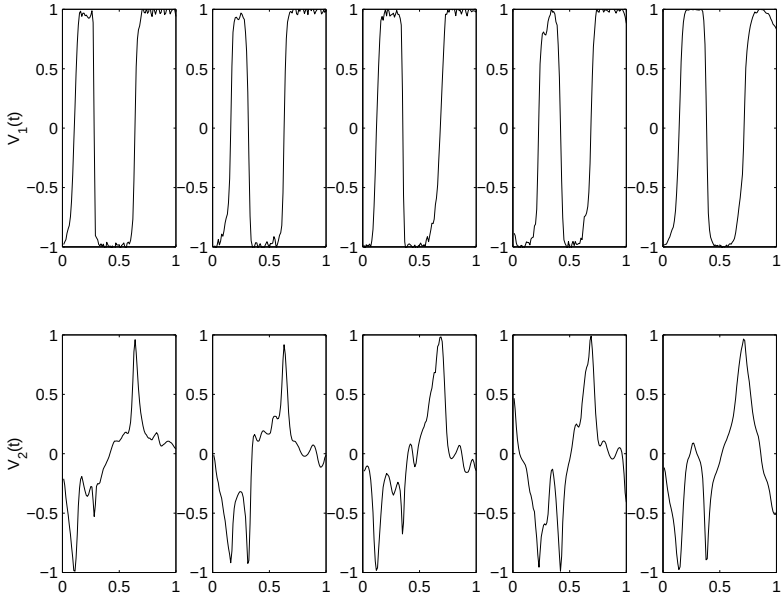

Figure 5: Components of the velocity vector $V^{(i)}(t)$ for the four observed instances of Shakespeare's signature (first 4 plots in each row), and the template (last plot in each row).

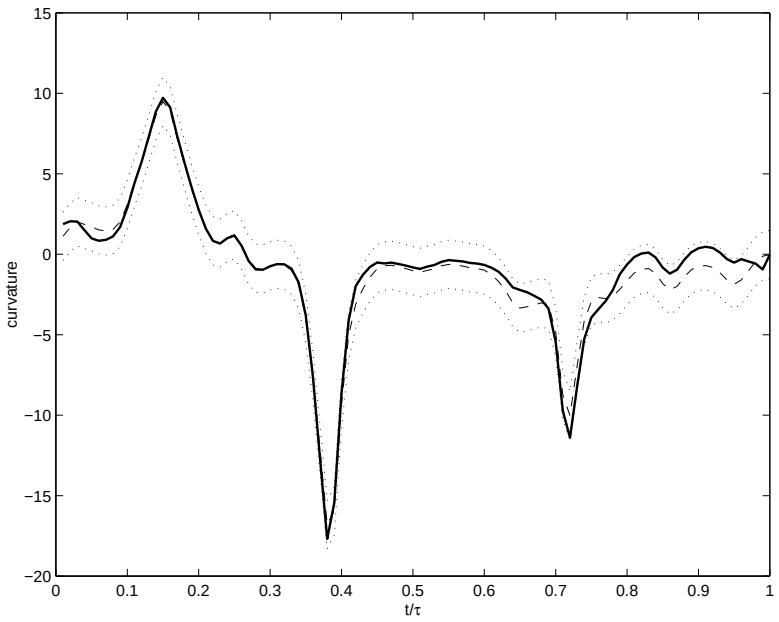

Figure 6: Posterior mean baseline curvature (solid line), template curvature κ_0 (dashed line) and bracketting functions for the buffer region (dotted lines).

We find it preferable, however, to use curve registration and initialize the piecewise-linear time warping h_i in the MCMC by matching the (three) zero crossings of $V_1^{(i)}(t)$ to those of the template.

Figure 6 shows the posterior mean baseline curvature function. The width of the buffer region is specified as 3 ($\epsilon = 1.5$), which was chosen based on the inspection of simulations under the prior. The (prior) Strauss process parameters for the baseline curvature are taken as $\gamma = 0.02$ (a small value that forces the knots to be well separated in the time direction), $\rho = 0.03\tau$, $\beta = 10$. The number of piecewise-linear components in the time warping is set at $p = 20$, and the precision parameter $\eta = 400$. The shape of the generated traces was found to be highly sensitive to perturbations in the time warping, and smaller values of η gave poor results. The MCMC run used a burn-in of 10^4 sweeps, followed by the same number of sweeps in the estimation stage.

The posterior mean baseline curvature (the solid line in Figure 6) is seen to touch the boundary of the buffer region at various points, indicating that the template is not ideal, but it is satisfactory for exploratory purposes. If a more refined analysis is needed, for a subsequent MCMC run the buffer region could be re-centered on the exploratory posterior mean baseline curvature, and its width reduced (replacing ϵ by $\epsilon/2$ say), which would have the effect of increasing the acceptance rate for new knots in κ.

The traces derived from the posterior mean baseline curvature and the posterior mean time warping are displayed in Figure 7. These are comparable in shape (but smoother because the noise has been 'removed') to the corresponding data traces, and also quite similar to the template. The time warping is seen to be substantial in all but the third observed signature. Simulated traces derived from the posterior baseline curvature and time warping for the third

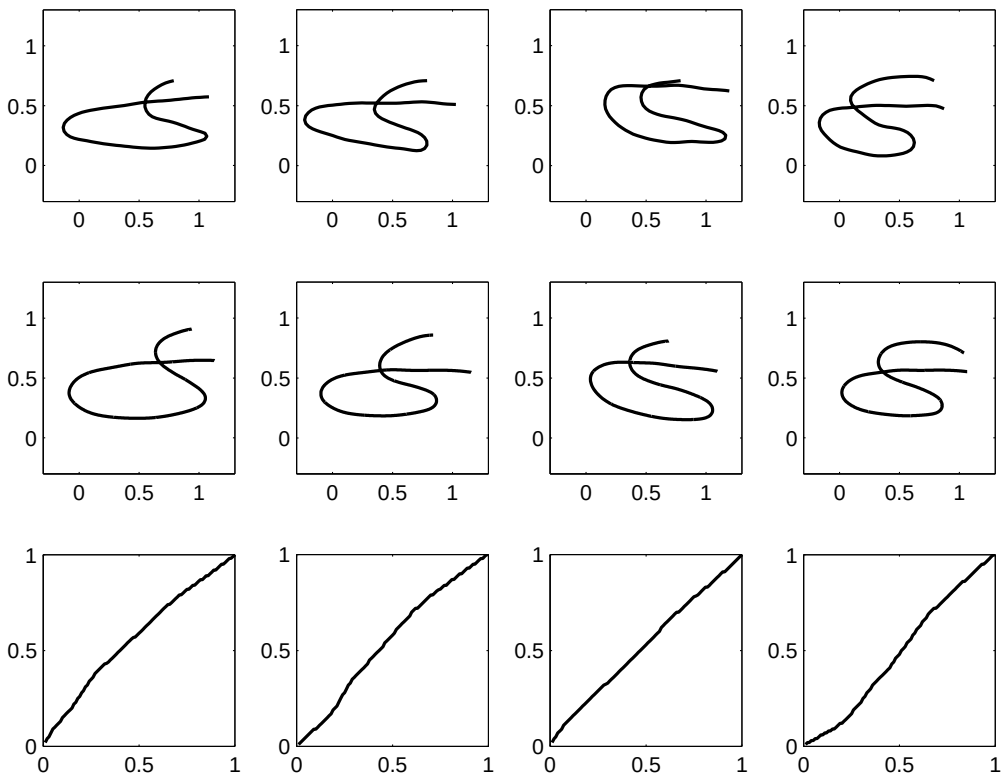

Figure 7: Traces of the data (first row); traces derived from the posterior mean baseline curvature adjusted by each posterior mean time warping (second row); posterior mean time warping functions (last row).

observed signature are given in Figure 8. These are suggestive of the variety of possible realizations of Shakespeare's 'S'.

Toy example

To evaluate the accuracy of the model fitting procedure, we now consider a toy example to see whether the model fit comes close in terms of reproducing pre-specified baseline curvature and time warping functions.

The first row of Figure 9 shows the data, which was generated from the observation model with $\kappa = \kappa_0$ and noise level $\sigma = 0.15$; the time warping functions are plotted in the bottom row (dotted lines). The various tuning parameters are the same as before, except that the time warping has $p = 10$ piecewise-linear segments instead of 20. The MCMC run has a burn-in of $2,000$ sweeps, and the same number of sweeps in the estimation stage. The fit is seen to be excellent: the estimates and the targets (see Figure 10 for the baseline curvature) are so close as to be almost indistinguishable. The same accuracy was found using other target baseline curvature functions κ within the buffer region. Fitting the model with a value of the noise level σ in the range 0.1–0.2 (with $\sigma = 0.15$ in the data) has little effect on these results.

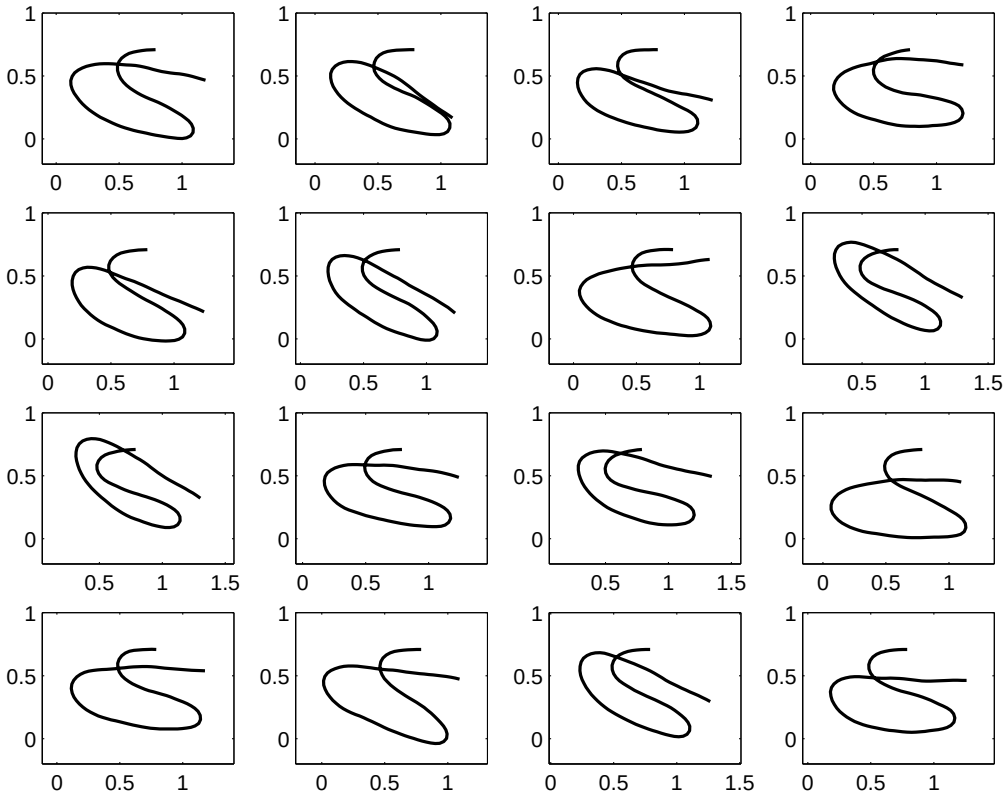

Figure 8: Simulated traces from the fitted observation model ($\sigma = 0.15$) under the posterior mean baseline curvature adjusted by the posterior mean of the time warping function $h_3(t)$.

Signature verification

Hamilton (1985) made a strong case that Shakespeare's (1616) will is holographic (i.e., written by Shakespeare himself). Figure 11 shows an 'S' from the will, along with a modern forgery.

The proposed signature verification method applied to these traces provides strong evidence that the fake 'S' is indeed a forgery: $F = 12.4\%$, p-value $= 0.033$. In contrast, the 'S' from Shakespeare's will does not raise suspicion: $F = 6.5\%$, p-value $= 0.42$. These results are based on 10^4 bootstrap draws from the fitted model with noise level $\sigma = 0.15$; the histogram of F^* is displayed in Figure 12.

The p-values are quite sensitive to the choice of noise level; with $\sigma = 0.2$ the p-value for the fake 'S' exceeds 0.05. In a practical application it would be advisable to calibrate σ from large training samples to control the false-positive and false-negative error rates.

Computational costs

The proposed approach is computationally feasible for larger samples of signatures; a sample of 10 true signatures per individual is a typical requirement for a signature verification system, and this would entail only modest computational cost. The MCMC runtime is approximately proportional to the sample size n and the number of sweeps; increasing n does not appear to

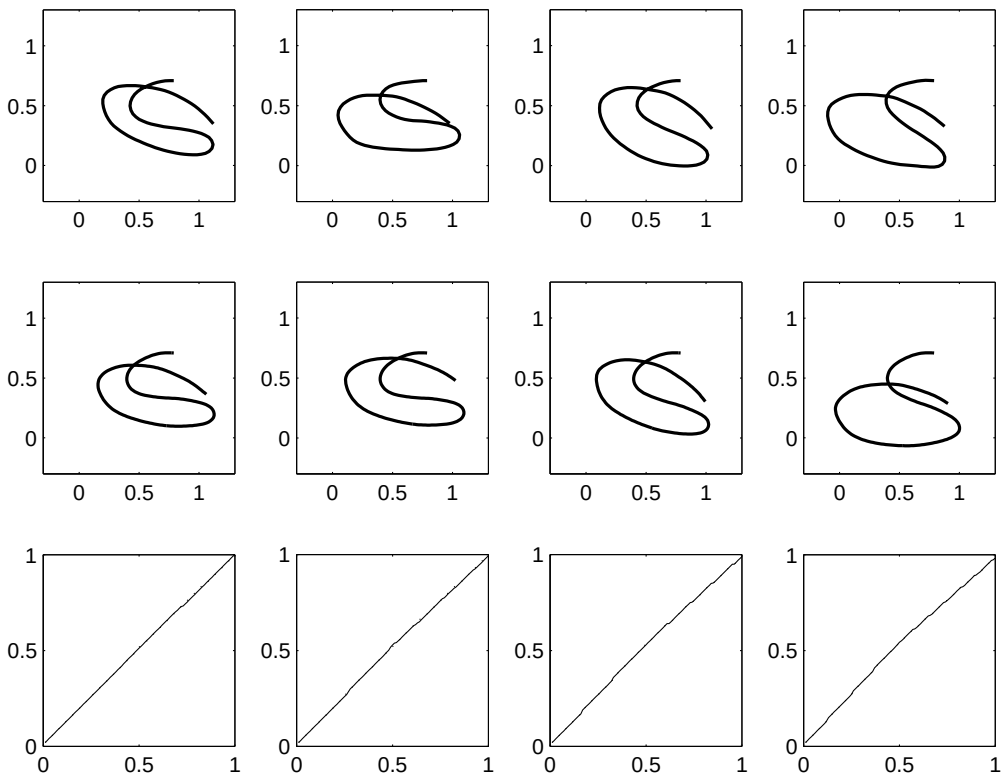

Figure 9: Toy example. Traces of the data (first row); traces derived from the posterior mean baseline curvature adjusted by each posterior mean time warping (second row); time warping (last row), with posterior mean (solid line) and true (dotted line).

slow down the convergence rate of the MCMC.

A time-consuming aspect of handling more signatures, however, is the "manual" preprocessing stage of extracting the velocity vectors from the image data, which can take several minutes per image. Runtime increases dramatically with the number of points N on the grid used to approximate the likelihood, but $N = 100$ was satisfactory and with $n = 4$ signatures and 10^4 sweeps, runtime is less than 10 minutes (on a fast desktop computer). Generation of the bootstrap sample used in the signature verification stage takes only a few seconds.

4 Discussion

In this paper we have utilized an array of modern statistical techniques in an attempt to place the difficult problem of off-line signature verification in a model-based framework. The proposed approach can be applied to any smooth segment of a signature. Cusps amount to infinite curvature, which cannot be handled by the observation model. Nevertheless, in some applications (such as biometric identification and manuscript analysis), the analysis of smooth segments of handwriting is a potentially valuable tool.

The key advantage over methods based on similarity features is that a formal statistical

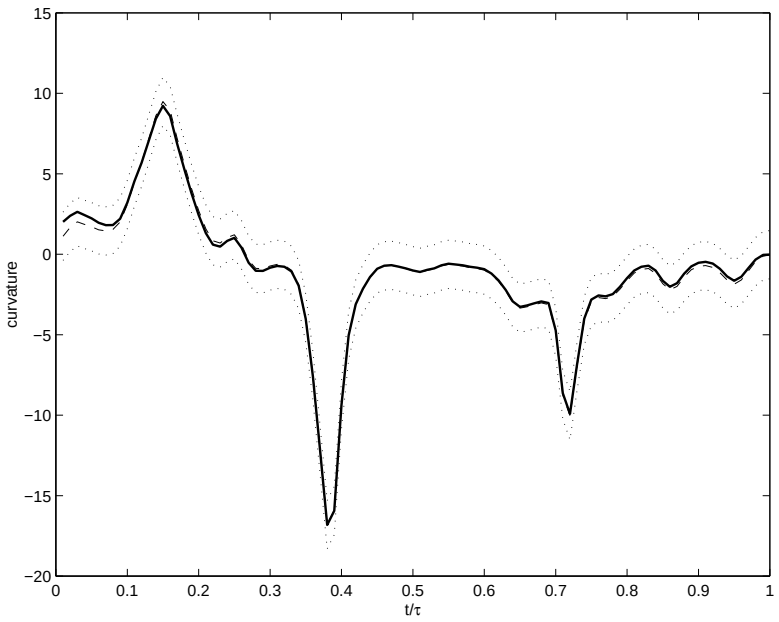

Figure 10: Toy example. Posterior mean baseline curvature (solid line), true baseline curvature (dashed line) and bracketting functions for the buffer region (dotted lines).

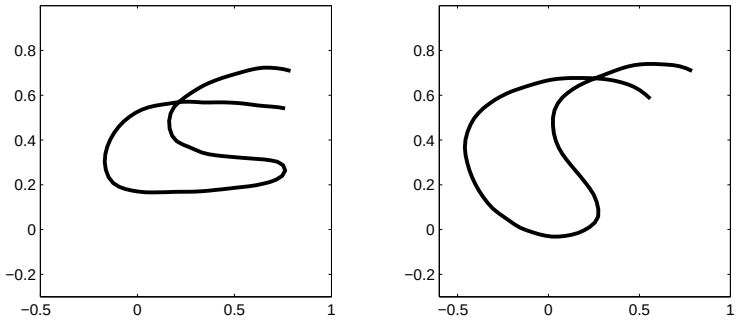

Figure 11: Test signatures for forgery detection. A Secretary hand 'S' from Shakespeare's holographic will, March 25, 1616, and a modern forgery (right).

test for a forgery now becomes available. Much remains to be done, however, to make the approach useful in practice. The method needs to be tested experimentally on large databases of signatures and the noise level calibrated to optimize the procedure in terms of Type I and Type II error rates; see Munich and Perona (2003) for a comprehensive illustration (in the context of biometric identification) of the type of performance testing that is needed. Also, the preprocessing stage needs to be further developed to provide automatic segmentation of the signature into smooth components for analysis.

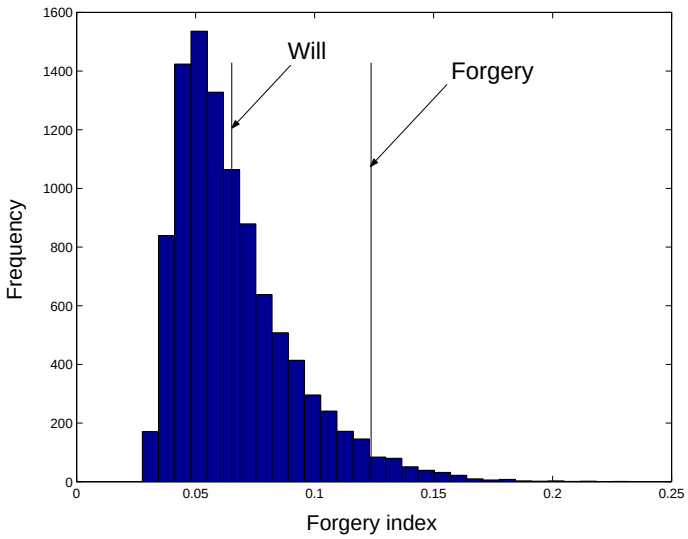

Figure 12: Histogram of bootstrap simulations of F^* under the fitted observation model; the vertical lines show the forgery indices for the 'S' from Shakespeare's will and the fake 'S' (Figure 11).

Appendix: Proofs

Proof of (2.4)

The two components of the vector process $V(t)$ correspond to the real and imaginary parts of $\tilde{V}(t) = F(\theta(t))$, where $F(x) = e^{ix}$. From Itô's formula, using the basic model (2.3) and noting that $F'(\theta(t)) = i\tilde{V}(t)$, $F''(\theta(t)) = -\tilde{V}(t)$, we obtain

$$\tilde{V}(t) = e^{i\varphi} + \int_0^t \tilde{V}(s)\left(i\kappa(s) - \frac{1}{2}\sigma^2\right)\,ds + i\sigma\int_0^t \tilde{V}(s)\,dW(s).$$

Re-expressing this equation in vector form shows that V satisfies the linear stochastic differential equation

$$dV(t) \;=\; A(t)\,dt + B(t)\,dW(t),\ t\in[0,\tau] \tag{A.1}$$

where

$$A(t) \;=\; \begin{pmatrix} -\frac{1}{2}\sigma^2 & -\kappa(t) \\ \kappa(t) & -\frac{1}{2}\sigma^2 \end{pmatrix} V(t), \quad B(t) = \sigma\begin{pmatrix} -V_2(t) \\ V_1(t)\end{pmatrix}, \tag{A.2}$$

with initial condition $V(0) = (\cos\varphi, \sin\varphi)^T$. Referring to a result of Liptser and Shiryayev (1977, p. 279), it follows that the distribution μ of V is absolutely continuous with respect to the distribution $\bar{\mu}$ of the two-dimensional process $\bar{V}$ satisfying the stochastic differential equation

$$d\bar{V}(t) \;=\; \sigma\begin{pmatrix} -\bar{V}_2(t) \\ \bar{V}_1(t)\end{pmatrix} dW(t), \qquad \bar{V}(0) = V(0). \tag{A.3}$$

The corresponding Radon–Nikodym derivative is

$$\frac{d\mu}{d\bar{\mu}}(V) = \exp(\ell(\kappa|V)),$$

where

$$\begin{aligned}\ell(\kappa|V) &= \int_0^\tau A(s)^T (B(s)B(s)^T)^- \, dV(s) \\ &\qquad -\frac{1}{2}\int_0^\tau A(s)^T (B(s)B(s)^T)^- A(s)\, ds,\end{aligned} \tag{A.4}$$

and C^- denotes the Moore–Penrose generalized inverse of a matrix C. This expression can be simplified as follows. The Moore–Penrose inverse above can be found explicitly: $(BB^T)^- = |V|^2/\sigma^2 = 1/\sigma^2$, where the unit norm property of V has been used. The first integral in (A.4) then becomes

$$\begin{aligned}\frac{1}{\sigma^2}\int_0^\tau A(s)^T \, dV(s) &= \frac{1}{\sigma^2}\int_0^\tau \kappa(s)\left\{V_1(s)\, dV_2(s) - V_2(s)\, dV_1(s)\right\} \\ &\qquad -\frac{1}{2}\int_0^\tau \left\{V_1(s)\, dV_1(s) + V_2(s)\, dV_2(s)\right\}.\end{aligned}$$

Integration by parts simplifies the second integral in the above display:

$$\begin{aligned}\int_0^\tau \left\{V_1(s)\, dV_1(s) + V_2(s)\, dV_2(s)\right\} &= \frac{1}{2}\left\{V_1(\tau)^2 - V_1(0)^2 - (\langle V_1\rangle(\tau) - \langle V_1\rangle(0))\right\} \\ &\qquad +\frac{1}{2}\left\{V_2(\tau)^2 - V_2(0)^2 - (\langle V_2\rangle(\tau) - \langle V_2\rangle(0))\right\} \\ &= -\frac{1}{2}\sigma^2\tau\end{aligned}$$

where predictable variation processes $\langle V_1\rangle$ and $\langle V_2\rangle$ coincide with the optional variation processes derived in the proof of (2.7) below. Again using the unit norm property of V, the second integral in the log-likelihood (A.4) reduces to

$$\begin{aligned}\frac{1}{\sigma^2}\int_0^\tau A(s)^T A(s)\, ds &= \frac{1}{\sigma^2}\int_0^\tau \left(\frac{1}{4}\sigma^4 + \kappa(s)^2\right) ds \\ &= \frac{1}{4}\sigma^2\tau + \frac{1}{\sigma^2}\int_0^\tau \kappa(s)^2\, ds.\end{aligned}$$

Combining these simplifications gives

$$\begin{aligned}\ell(\kappa|V) &= \frac{1}{\sigma^2}\int_0^\tau \kappa(s)\left\{V_1(s)\, dV_2(s) - V_2(s)\, dV_1(s)\right\} + \frac{1}{4}\sigma^2\tau \\ &\qquad -\frac{1}{2}\left(\frac{1}{4}\sigma^2\tau + \frac{1}{\sigma^2}\int_0^\tau \kappa(s)^2\, ds\right) \\ &= \frac{1}{\sigma^2}\int_0^\tau \kappa(s)\left\{V_1(s)\, dV_2(s) - V_2(s)\, dV_1(s)\right\} - \frac{1}{2\sigma^2}\int_0^\tau \kappa(s)^2\, ds + \frac{1}{8}\sigma^2\tau.\end{aligned}$$

Proof of (2.6)

From its definition, the differential of the process Z can be expressed, using (A.1) and (A.2), as

$$\begin{aligned} dZ(t) &= (-V_2(t), V_1(t))\, dV(t) \\ &= (-V_2(t), V_1(t))A(t)\, dt + (-V_2(t), V_1(t))B(t)\, dW(t) \\ &= (-V_2(t), V_1(t)) \begin{pmatrix} -\frac{1}{2}\sigma^2 & -\kappa(t) \\ \kappa(t) & -\frac{1}{2}\sigma^2 \end{pmatrix} \begin{pmatrix} V_1(t) \\ V_2(t) \end{pmatrix} dt + \sigma |V(t)|^2 dW(t) \\ &= \kappa(t)\, dt + \sigma dW(t), \end{aligned}$$

where we again used the unit norm property of V.

Proof of (2.7)

The optional quadratic variation process $[X](t)$ of a semimartingale X may be defined in two equivalent ways: as

$$[X](t) = X(t)^2 - X(0)^2 - 2\int_{(0,t]} X(s-)\, dX(s),$$

or as the limit in probability as $N \to \infty$ of the sums

$$\sum_{j=1}^{N} (X(t_j) - X(t_{j-1}))^2$$

where $t_j = jt/N$, see Rogers and Williams (2000, Chapter IV). For continuous semimartingales, $[X]$ coincides with the predictable variation of the martingale part of X. But from (A.1), the martingale parts of (the continuous semimartingales) V_1 and V_2 have differentials $-\sigma V_2(s)\, dW(s)$ and $\sigma V_1(s)\, dW(s)$, respectively, so

$$[V_1](t) = \langle V_1 \rangle(t) = \sigma^2 \int_0^t V_2^2(s)\, ds, \quad [V_2](t) = \langle V_2 \rangle(t) = \sigma^2 \int_0^t V_1^2(s)\, ds.$$

The limit in probability as $N \to \infty$ of the summation in (2.7) is (by definition of optional quadratic variation) therefore

$$[V_1](\tau) + [V_2](\tau) = \sigma^2 \int_0^\tau (V_1(s)^2 + V_2(s)^2)\, ds = \sigma^2 \tau,$$

where we have used $|V(s)| = 1$ in the last step. This completes the proof.

Theorem A.1 *The Markov chain given by the hybrid sampler for the posterior distribution of* $(\mathbf{x}, \boldsymbol{\theta})$ *is geometrically ergodic.*

Proof. A Markov chain is geometrically ergodic if its transition kernel P is ϕ-irreducible, aperiodic and satisfies a Foster–Lyapunov drift condition (see Meyn and Tweedie 1993, Theorems 15.0.1 and 16.0.1). To check these conditions, we use of the following easily verified bounds on the Hastings ratios: $\alpha_{rw}(\cdot)$ is bounded away from zero, and there exist constants $0 < b < 1$, $c > 0$ such that

$$\frac{b^{n(\mathbf{x})}}{n(\mathbf{x})+1} \le \alpha_b(\mathbf{x}, \boldsymbol{\theta}, \xi) \le \frac{c}{n(\mathbf{x})+1}; \qquad \alpha_d(\mathbf{x}, \boldsymbol{\theta}, \xi) \ge \frac{n(\mathbf{x})}{c}. \tag{A.5}$$

Let ϕ denote the dominating measure for the posterior density. To establish ϕ-irreducibility we need to show that for any initial position $(\mathbf{x}, \boldsymbol{\theta})$ and measurable set A such that $\phi(A) > 0$, the m-step transition probability $P^m((\mathbf{x}, \boldsymbol{\theta}), A) > 0$ for some $m \ge 1$. First note that a transition $(\mathbf{x}, \boldsymbol{\theta}) \to (\mathbf{x}', \boldsymbol{\theta}')$ can be made by $n(\mathbf{x})$ death steps immediately followed by $n(\mathbf{x}')$ birth steps. From (A.5), the Hastings ratios involved in such a sequence of updates are bounded away from zero. Thus, since the random walk moves (on the components of $\boldsymbol{\theta}$) have proposal densities that are bounded away from zero on compact intervals, for any $m \ge n(\mathbf{x})$ there is a non-trivial absolutely continuous part of P^m and the corresponding density $p^m((\mathbf{x}, \boldsymbol{\theta}), \cdot)$ (with respect to ϕ) is bounded away from zero on $B_m = \{(\mathbf{x}', \boldsymbol{\theta}') : n(\mathbf{x}') = m - n(\mathbf{x})\}$. But

$$\phi(A) = \sum_{m=n(\mathbf{x})}^{\infty} \phi(A \cap B_m) > 0$$

so there exists an $m \ge n(\mathbf{x})$ such that $\phi(A \cap B_m) > 0$, and we have

$$P^m((\mathbf{x}, \boldsymbol{\theta}), A) \ge P^m((\mathbf{x}, \boldsymbol{\theta}), A \cap B_m) = \int_{A \cap B_m} p^m((\mathbf{x}, \boldsymbol{\theta}), \cdot)\, d\phi > 0,$$

as required

Next we show that $C = \{(\mathbf{x}, \boldsymbol{\theta}) : n(\mathbf{x}) < K\}$ is a small set for any constant K. Fix $m \ge K$, and note that, arguing as above, there is a non-trivial absolutely continuous part of P^m and $p^m((\mathbf{x}, \boldsymbol{\theta}), \cdot)$ is bounded away from zero on $\{\emptyset\} \times \Theta$ uniformly in $(\mathbf{x}, \boldsymbol{\theta}) \in C$, where $\emptyset$ is the empty configuration of points in $\mathcal{B}$. Thus there exists a constant $\varrho > 0$ such that for all $(\mathbf{x}, \boldsymbol{\theta}) \in C$ and all measurable sets A,

$$P^m((\mathbf{x}, \boldsymbol{\theta}), A) \ge \int_{A \cap [\{\emptyset\} \times \Theta]} p^m((\mathbf{x}, \boldsymbol{\theta}), \cdot)\, d\phi \ge \varrho \nu(A),$$

where ν is the uniform distribution on $\{\emptyset\} \times \Theta$. This shows that C is a small set.

The Foster–Lyapunov drift condition can be shown to hold for the small set C provided K is chosen large enough. This involves using the upper bound (A.5) on the Hastings ratio for birth proposals to find constants $0 < \lambda < 1$, $\eta < \infty$, such that $PV \le \lambda V + \eta 1_C$, where $V(\mathbf{x}, \boldsymbol{\theta}) = \max(1, c^{n(\mathbf{x})})$. This can be done by simple modification of an argument of Geyer and Møller (1994, p. 365). Aperiodicity is easily checked, and we conclude that the Markov chain is geometrically ergodic.

References

Abuhaiba, I. S. I., Holt, M. J. J., and Datta, S. (1994), "Processing of off-line handwritten text: polygonal approximation and enforcement of temporal information," *Graphical Models and Image Processing*, 56, 324–335.

Abuhaiba, I. S. I., Holt, M. J. J., and Datta, S. (1998), "Recognition of off-line cursive handwriting," *Computer Vision and Image Understanding*, 71, 19–38.

Amit, Y., Grenander, U., and Piccioni, M. (1991), "Structural image restoration through deformable templates," *Journal of the American Statistical Association*, 86, 376–387.

Barry, D. (1986), "Nonparametric Bayesian regression," *Annals of Statistics*, 14, 934–953.

Blackwell, P. G., and Møller, J. (2003), "Bayesian analysis of deformed tessellation models," *Advances in Applied Probability*, 35, 4–26.

Brown, L. D., and Low, M. G. (1996), "Asymptotic equivalence of nonparametric regression and white noise," *Annals of Statistics*, 24, 2384–2398.

Carmo, M. P. do (1976), *Differential Geometry of Curves and Surfaces*, Prentice–Hall.

Cox, D. D. (1993), "An analysis of Bayesian inference for nonparametric regression," *Annals of Statistics*, 21, 903–923.

Deng, P. S., Liao, H.-Y. M., Ho, C. W., and Tyan, H.-R. (1999), "Wavelet-based off-line signature verification," *Computer Vision and Image Understanding*, 76, 173–190.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), "Automatic Bayesian curve fitting," *Journal of the Royal Statistical Society*, Ser. B, 60, 333–350.

Donoho, D. L., and Johnstone, I. M. (1995), "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, 90, 1200–1224.

Dryden, I. L., and Mardia, K.V. (1998), *Statistical Shape Analysis*, Wiley, New York.

Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, Dekker, New York.

Fang, B., Leung, C. H., Tang, Y. Y., Tse, K. W., Kwok, P. C. K., and Wong, Y. K. (2003), "Off-line signature verification by the tracking of feature and stroke positions," *Pattern Recognition*, 36, 91–101.

Fort, G., Moulines, E., Roberts, G. O., and Rosenthal, J. S. (2003), "On the geometric ergodicity of hybrid samplers," *Journal of Applied Probability*, 40, 123–146.

Geyer, C. J., and Møller, J. (1994), "Simulation and likelihood inference for spatial point processes," *Scandinavian Journal of Statistics*, 21, 359–373.

Grenander, U. (1970), "A unified approach to pattern analysis," *Advances in Computers*, 10, 175–216.

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

Hamilton, C. (1985), *In Search of Shakespeare: A Reconnaissance into the Poet's Life and Handwriting*, Harcourt Brace Jovanovich, New York.

Hastie, T. J., Kishon, E., Clark, M., and Fan, J. (1991), "A model for signature verification," in *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, 1, pp. 191–196. Technical report (Feb. 1992): `http://stat.stanford.edu/~trevor/Papers/signature.ps`.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001), *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York.

Kashi, R. S., Hu, J., Nelson, W. L., and Turin, W. (1997), "On-line handwritten signature verification using hidden Markov model features," in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, 253–257.

Klassen, E., Srivastava, A., Mio, W., and Joshi, S. (2004), "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 372–383.

Lee, L. L. (1996), "Neural approaches for human signature verification," in *Third International Conference on Signal Processing Proceedings*, pp. 1346–1349.

Lele, S., and Richtsmeier, J. T. (2000), *An Invariant Approach to Statistical Analysis of Shapes*, Chapman and Hall/CRC Press, London.

Lew, J. S. (1980), "Kinematic theory of signature verification measurements," *Mathematical Biosciences*, 48, 25–51.

Liptser, R. S., and Shiryayev, A. N. (1977), *Statistics of Random Processes I*, Springer, New York.

Martens, R., and Claesen, L. (1997), "On-line signature verification: discrimination emphasised," in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, pp. 657–660.

Matsuura, T., and Sakai, H. (1996), "On stochastic system representation of handwriting process and its application to signature verification," in *Third International Conference on Signal Processing Proceedings*, pp. 1330–1333.

Meyn, S. P., and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, Springer, London.

Miller, M., Christensen, G., Amit, Y., and Grenander, U. (1993), "A mathematical textbook of deformable neuro-anatomies," *Proceedings of the National Academy of Sciences*, R90, 11944–11948.

Mokhtarian, F., and Mackworth, A. K. (1992), "A theory of multi-scale, curvature-based shape representation for planar curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 789–805.

Mumford, D. (1994), "Elastica and computer vision," in *Algebraic Geometry and Its Applications*, C. Bajaj (ed.), Springer, New York, pp. 491–506.

Munich, M. E., and Perona, P. (2003), "Visual identification by signature tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 200–217.

Nussbaum, M. (1996), "Asymptotic equivalence of density estimation and Gaussian white noise," *Annals of Statistics*, 24, 2399–2430.

Penev, P. S., and Atick, J. J. (1996), "Local feature analysis: a general statistical theory for object representation," *Network: Computation in Neural Systems*, 7, 477–500.

Plamondon, R., and Lorette, G. (1989), "Automatic signature verification and writer identification—the state of the art," *Pattern Recognition*, 22, 107–131.

Proctor, S., Illingworth, J., and Mokhtarian, F. (2000), "Cursive handwriting recognition using hidden Markov models and a lexicon-driven level building algorithm," *IEE Proceedings on Vision, Image and Signal Processing*, 147, 332–339.

Ramsay, J. O. (2000), "Functional components of variation in handwriting," *Journal of the American Statistical Association*, 95, 9–15.

Ramsay, J. O., and Silverman, B. W. (1997), *Functional Data Analysis*, Springer, New York.

Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer, New York.

Rogers, L. C. G., and Williams, D. (2000), *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*, Cambridge University Press.

Sabourin, R., Drouhard, J.-P., and Wah, E. S. (1997), "Shape matrices as a mixed shape factor for off-line signature verification," in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, pp. 661–665.

Shakespeare, W. (1616), "Last Will and Testament," `http://www.documentsonline.pro.gov.uk/shakespeare.pdf`.

Silverman, B. W. (1985), "Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion)," *Journal of the Royal Statistical Society*, Ser. B, 47, 1–52.

Small, C. G. (1996), *The Statistical Theory of Shape*, Springer, New York.

Van Lieshout, M. N. M. (2000), *Markov Point Processes and their Applications*, Imperial College Press, London.

Wahba, G. (1978), "Improper priors, spline smoothing, and the problem of guarding against model errors in regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 364–372.