Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 00 (2015) 1–5

Journal Logo

www.elsevier.com/locate/procedia

Spatial Statistics 2015: Emerging Patterns – Part 2

State estimation for random closed sets

M.N.M. van Lieshout[a,b]

[a]*CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands*
[b]*University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands*

Abstract

State estimation entails the estimation of an unobserved random closed set from (partial) observation of an associated random set. Examples include edge effect correction, cluster detection, filtering and prediction. We focus on inference for random sets based on points sampled on its boundary. Such data are subject to mis-alignment and noise. First, we ignore mis-alignment and discuss maximum likelihood estimation of the model and noise parameters in the Fourier domain. We estimate the unknown curve by back-transformation and derive the expectation of the integrated squared error. Then, we model mis-alignment by means of a shifted parametric diffeomorphism and minimise a suitable objective function simultaneously over the unknown curve and the mis-alignment parameters.

Keywords: Missing data; random closed set; spectral analysis; state estimation;

1. Introduction

Many geographical or biological objects are observed in image form. The boundaries of such objects are seldom crisp due to measurement error and discretisation, or because the boundaries themselves are intrinsically indeterminate [2]. Moreover, the objects may not be static in the sense that if multiple images are taken, the objects may have been deformed.

One attempt to model natural objects under uncertainty is fuzzy set theory [12]. However, the underlying axioms are too poor to handle topological properties of the shapes to be modelled and cannot deal with correlation. Similarly, the belief functions that lie at the heart of the Dempster–Shafer theory [4, 9] do not necessarily correspond to the containment function of a well-defined random closed set [7].

This paper is organised as follows. We begin by recalling basic facts about planar curves, cyclic Gaussian random processes and spectral analysis. Then we formulate a model for sampling noisy curves, carry out inference in the Fourier domain and quantify the error. We describe how mis-alignment between curves may be dealt with and illustrate the approach on images of a lake in Ethiopia.

2. Noisy curves

2.1. Planar curves

Throughout this paper, we model the boundary of the random object of interest by a smooth (simple) closed curve. Consider the class of functions $\Gamma : I \to \mathbb{R}^2$ from some interval I to the plane. Define an

*Corresponding author. Tel.: +31-20-5924008; E-mail address: M.N.M.van.Lieshout@cwi.nl

equivalence relation $\sim$ on the function class as follows: Two functions Γ and Γ' are equivalent, $\Gamma \sim \Gamma'$, if there exists a strictly increasing function φ from I onto another interval I' such that $\Gamma = \Gamma' \circ \varphi$. Note that φ is a homeomorphism. The relation defines a family of equivalence classes, each of which is called a curve. Its member functions are called parametrisations. Since the images of two parametrisations of the same curve are identical, we shall, with slight abuse of notation, use the symbol Γ for a specific parametrisation, for a curve and for its image. For convenience, we shall often rescale the definition interval to $[-\pi, \pi]$.

In an optimisation context, it is natural to assume a curve to be parametrised by some function Γ that is C^1 and the same degree of smoothness to hold for the functions φ that define the equivalence relation between parametrisations. In effect, φ should be a diffeomorphism. See [11, Chapter 1] for further details.

2.2. Fourier representation

Let $\Gamma = (\Gamma_1, \Gamma_2) : [-\pi, \pi] \to \mathbb{R}^2$ be a C^1 function with $\Gamma(-\pi) = \Gamma(\pi)$. Since the family $\{\cos(j\theta), \sin(j\theta) : j \in \mathbb{N}_0, \theta \in [-\pi, \pi]\}$ forms an orthogonal basis for $L_2([-\pi, \pi])$, the space of all square integrable functions on $[-\pi, \pi]$, Γ can be approximated by a trigonometric polynomial of the form

$$\sum_{j=0}^{J} \Big[\mu_j \cos(j\theta) + \nu_j \sin(j\theta)\Big].$$

The vectors μ_j and ν_j are called the Fourier coefficients of order j and satisfy, for $j \in \mathbb{N}$,

$$\left\{ \begin{array}{lcl} \mu_0 & = & \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma(\theta)\, d\theta; \\ \mu_j & = & \frac{1}{\pi} \int_{-\pi}^{\pi} \Gamma(\theta) \cos(j\theta)\, d\theta; \\ \nu_j & = & \frac{1}{\pi} \int_{-\pi}^{\pi} \Gamma(\theta) \sin(j\theta)\, d\theta. \end{array} \right.$$

2.3. Stationary cyclic Gaussian random processes

As in [5, Proposition 2.1], let $N = (N_1, N_2)$ be a stationary cyclic Gaussian random process on $[-\pi, \pi]$ with values in $\mathbb{R}^2$ of the form

$$N(\theta) = \sum_{j=0}^{\infty} \Big[A_j \cos(j\theta) + B_j \sin(j\theta)\Big],$$

where the components of A_j and B_j are mutually independent zero-mean Gaussian random variables with variances σ_j^2 that are small enough for the series $\sum_j \sigma_j^2$ to converge. Then N has independent components with zero mean and covariance function $\rho(\theta) = \sum_{j=0}^{\infty} \sigma_j^2 \cos(j\theta)$. For the existence of a continuous version, further conditions are needed. Indeed, Theorem 25.10 in [8] implies that if

$$\sum_{j=1}^{\infty} j^{2k+\epsilon} \sigma_j^2 < \infty \tag{1}$$

for $k \in \mathbb{N} \cup \{0\}$, $\epsilon > 0$, there exists a version of N that is k times continuously differentiable. From now on, we shall assume (1) for $k = 1$.

3. Inference

Suppose that the data consist of multiple observations X_t of an object of interest in discretised form as a list of finitely many points $X_t = (X_t^l)_{l=1,\ldots,n}$ on its boundary. In other words, the lists $(X_t^l)_l$ trace some unknown closed curve Γ affected by noise as described in Section 2.

We set ourselves the goal of estimating Γ and the noise variance parameters σ_j^2. For the moment, assume that the curves are perfectly aligned. (We shall return to the issue of estimating the parametrisations later). Then we obtain the simplified model

$$X_t(\theta) = \Gamma(\theta) + N_t(\theta),$$

which is observed at $\theta_l = -(n+1)\pi/n + 2\pi l/n$, $l = 1, \ldots, n$, to give $X_t^l = X_t(\theta_l)$.

It is natural to carry out inference in the Fourier domain. Let, for $j \in \mathbb{N}$,

$$\left\{ \begin{array}{lcl} F^t_{0,n} & = & \frac{1}{n}\sum_{l=1}^n X^l_t = \frac{1}{n}\sum_{l=1}^n [\Gamma(\theta_l) + N_t(\theta_l)] \\ F^t_{j,n} & = & \frac{2}{n}\sum_{l=1}^n X^l_t \cos(j\theta_l) = \frac{2}{n}\sum_{l=1}^n [\Gamma(\theta_l)\cos(j\theta_l) + N_t(\theta_l)\cos(j\theta_l)] \\ G^t_{j,n} & = & \frac{2}{n}\sum_{l=1}^n X^l_t \sin(j\theta_l) = \frac{2}{n}\sum_{l=1}^n [\Gamma(\theta_l)\sin(j\theta_l) + N_t(\theta_l)\sin(j\theta_l)] \end{array} \right. \tag{2}$$

be the Riemann approximations to the random Fourier coefficients of X_t. We shall write $\mu_{j,n}$ respectively $\nu_{j,n}$ for the deterministic parts of $F^t_{j,n}$ and $G^t_{j,n}$ in (2). Furthermore, let

$$\sigma^2_{j,n} = \frac{2}{n}\sum_{l=1}^n \rho(\theta_l)\cos(j\theta_l)$$

for $j \geq 1$ and $\sigma^2_{0,n} = \sum_l \rho(\theta_l)/n$ be the Riemann approximations of σ^2_j.

From now on, assume that $n \geq 3$ is odd. In this case, the sequence θ_l contains 0 and is symmetric around zero. We estimate $\mu_{j,n}$ and $\nu_{j,n}$ by the empirical means of, respectively, $F^t_{j,n}$ and $G^t_{j,n}$, $t = 0, \ldots, T$, and transform back to the spatial domain to obtain

$$\widehat{\Gamma_n}(\theta) = \hat{\mu}_{0,n} + \sum_{j=1}^J \left[\hat{\mu}_{j,n}\cos(j\theta) + \hat{\nu}_{j,n}\sin(j\theta)\right] = \frac{1}{(T+1)}\sum_{t=0}^T \sum_{l=1}^n X^l_t S_l(\theta) \tag{3}$$

with smoother $S_l(\theta) = 1/n + 2\sum_{j=1}^J \cos(j(\theta-\theta_l))/n$. Finally, the variances $\sigma^2_{j,n}$ are estimated by

$$\hat{\sigma}^2_{j,n} = \frac{1}{4(T+1)}\sum_{t=0}^T \left[\|F^t_{j,n} - \hat{\mu}_{j,n}\|^2 + \|G^t_{j,n} - \hat{\nu}_{j,n}\|^2\right]$$

for $j \in \mathbb{N}$ and $\hat{\sigma}^2_{0,n} = \sum_{t=0}^T \|F^t_{0,n} - \hat{\mu}_{0,n}\|^2/(2(T+1))$ for $j = 0$.

We assume $J < n/2$, so that the number of Fourier parameters to estimate does not exceed the number of observed boundary points. Then $\hat{\mu}_{j,n}$ and $\hat{\nu}_{j,n}$ are unbiased independent Gaussian maximum likelihood estimators with diagonal covariance $\sigma^2_{j,n}/(T+1)$, whereas $\hat{\sigma}^2_{j,n}$ is proportional to a χ^2. Moreover, $\hat{\Gamma}_n$ is a cyclic Gaussian random process whose expected mean integrated squared error reads [6]

$$\sum_{j=J+1}^\infty \left[\|\mu_j\|^2 + \|\nu_j\|^2\right] + \frac{4}{T+1}\sum_{j=0}^J \sigma^2_{j,n} + 2c_{0,n} + \sum_{j=1}^J c_{j,n}, \tag{4}$$

where

$$c_{j,n} = \|\mu_{j,n} - \mu_j\|^2 + \|\nu_{j,n} - \nu_j\|^2$$

for $j = 1, \ldots, J$ with $c_{0,n}$ equal to $\|\mu_{0,n} - \mu_0\|^2$. The first term in (4) is the bias caused by taking into account only a finite number of Fourier coefficients. The second term corresponds to the variance, and the last two terms describe the discretisation error in the Fourier coefficients. Note that the smoothness assumptions on Γ and the covariance function ρ imply that both $\sum_j(\sigma^2_{j,n} - \sigma^2_j)$ and $\sum_j c_{j,n}$ are of the order J^3/n^2.

4. Parametrisation

As we saw previously, given a root, any parametrisation Γ of a (simple) closed C^1 curve can be written as a composition $\Gamma' \circ \varphi$ of a fixed parametrisation Γ' with a diffeomorphism φ. Thus, given two curves parametrised by, say, Γ and Γ_1, alignment of Γ_1 to Γ amounts to finding a shift α and a diffeomorphism φ such that $\Gamma_1(\theta) \approx \Gamma(\varphi(\theta-\alpha))$ interpreted cyclically. Without loss of generality, we consider diffeomorphisms φ from $[-\pi, \pi]$ onto itself.

Parametric diffeomorphisms can be constructed as the flow of differential equations [11, Chapter 8]. In our context, it is convenient to consider the differential equation

$$x'(t) = f_w(x(t)), \qquad t \in \mathbb{R}, \tag{5}$$

with initial condition $x(0) = \theta \in [-\pi, \pi]$. Heuristically, consider a particle whose position at time 0 is θ. If the particle travels with speed governed by the function f_w, then $x(t)$ is its position at time t. To emphasise the dependence on the initial state we shall also write $x_\theta(t)$.

We take f_w to be a trigonometric polynomial, that is, a linear combination of Fourier basis functions with pre-specified values w_i at equidistant $x_i \in [-\pi, \pi]$ under the constraint that $f_w(-\pi) = f_w(\pi) = 0$. More precisely, let $-\pi = x_0 < x_1 < \cdots < x_{2m} < \pi$, $w_0 = 0$, and define $f_w(x) = \sum_{j=0}^{2m} w_j t_j(x)$ where

$$t_j(x) = \prod_{j \neq k=0}^{2m} \sin\Big(\frac{x - x_k}{2}\Big) \Big/ \prod_{j \neq k=0}^{2m} \sin\Big(\frac{x_j - x_k}{2}\Big)$$

for arbitrary $w_1, \ldots, w_{2m}$ and $m \geq 1$. Note that f_w vanishes at $-\pi$. By inspection of the derivative, one finds that f_w is a C^1 function on $(-\pi, \pi)$ whose derivative is Lipschitz. Therefore, the function $\theta \mapsto x_\theta(1)$, the solution of (5) at time 1, is a diffeomorphism. This function is known as the flow of the differential equation and denoted by $\varphi(\theta) = x_\theta(1)$. As it depends on the weights, we shall also write $\varphi_w(\theta)$ to emphasise this fact. Note that in total, there are $2m + 1$ alignment parameters, $2m$ for the diffeomorphism and one for the shift.

Returning to our model

$$X_t(\theta) = \Gamma(\varphi_{w_t}(\theta - \alpha_t)) + N_t(\varphi_{w_t}(\theta - \alpha_t))$$

observed at $\theta_l = -(n+1)\pi/n + 2\pi l/n$, $l = 1, \ldots, n$, and extended to $[-\pi, \pi]$ by trigonometric interpolation. The latter is valid under our assumption that n is odd. For each choice of w_t and α_t, the theory developed so far may be applied to the transformed curves $Y_t(\theta) = X_t(\varphi_{-w_t}(\theta) + \alpha_t)$. Indeed, by (3), $\widehat{\Gamma_n}(\theta) = \sum_{t=0}^{T} \hat{\Gamma}_t(\theta)/(T+1)$ where

$$\hat{\Gamma}_t(\theta) = \sum_{l=1}^{n} Y_t(\theta_l) S_l(\theta) = \sum_{l=1}^{n} X_t(\varphi_{-w_t}(\theta_l) + \alpha_t) S_l(\theta).$$

We pick the 'best' $\hat{\alpha}_t$, $\hat{w}_t$ by minimising

$$\sum_{t=0}^{T} \sum_{l=1}^{n} \|\hat{\Gamma}_t(\theta_l) - \widehat{\Gamma_n}(\theta_l)\|^2$$

over a compact cube containing the origin under the constraint $\alpha_0 = w_0 = 0$. Such a criterion is well known in the shape recognition literature. For an overview, see for example [3]; asymptotic properties can be found in [1].

5. An application

Figure 1 shows three images of Lake Tana, the largest lake in Ethiopia and the source of the Blue Nile. It is located near the centre of the high Ethiopian plateau and covers some 1,400 square miles. Clearly visible is Dek island in the south-central portion of the lake, which we shall use as the centre of our coordinate system. The three images were taken on February 17th, 2001, at one second intervals by astronauts on the STS098 mission from a space craft altitude of 383 km and were downloaded from NASA's 'The Gateway to Astronaut Photography of Earth' website.

Note that the lake's border is rather fuzzy, resulting in a low image gradient. The output of edge detection algorithms is degraded even further by the substantial cloud cover. Therefore, the border was traced manually. The result is shown in the left-most panel in Figure 2. There are 73 points along each border curve.

We start by considering shift parameters only. In other words, $w_t = 0$ for $t = 0, 1, 2$. Using $J = 20$ Fourier coefficients and $\alpha_0 = 0$, the optimal parameters are $\hat{\alpha}_1 = -0.44$ and $\hat{\alpha}_2 = -2.33$ radians. The value of the optimisation function is 1195.048 corresponding to an average error of 2.34 pixels. The result can be improved by including diffeomorphic changes in speed. Optimising w_1, w_2 for vectors w_t, $t = 1, 2$, in $\mathbb{R}^{2m}$ with $m = 5$, we obtain a value of 568.0997 corresponding to an average error of 1.61 pixels. The estimated curve is plotted in the right-most panel in Figure 2.

Fig. 1. Images courtesy of the Image Science & Analysis Laboratory, NASA Johnson Space Center.

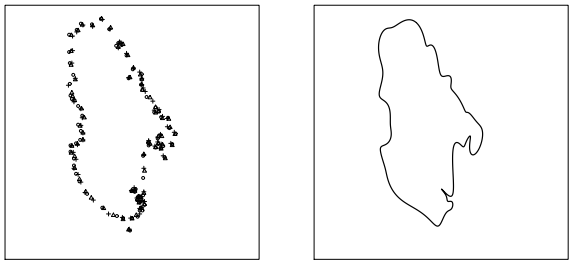

Fig. 2. Left panel: Sampled boundary curves corresponding to Figure 1. Circles trace the boundary of Lake Tana in the left-most panel, triangles correspond to the middle panel, and crosses trace the lake boundary in the right-most panel of Figure 1. Right panel: Estimated border.

Acknowledgements

This research was supported by The Netherlands Organisation for Scientific Research NWO (613.000.809). Calculations were carried out in `R` using the package `deSolve` (Soetaert *et al.*, 2010).

References

[1] Bigot J, Gendre X. Minimax properties of Fréchet means of discretely sampled curves. *Ann Statist* 2013; **41**:923–956.
[2] Burrough P, Frank A. *Geographic objects with indeterminate boundaries.* London: Taylor & Francis; 1996.
[3] Davies R, Twining C, Taylor C. *Statistical models of shape: Optimisation and evaluation.* London: Springer; 2008.
[4] Dempster AP. Upper and lower probabilities induced by a multivalued mapping. *Ann Statist* 1967; **38**:325–329.
[5] Jónsdóttir KY, Jensen EBV. Gaussian radial growth. *Image Anal Stereol* 2005; **24**:117–126.
[6] Lieshout, MNM. A spectral mean for point sampled closed curves. ArXiv 1310.7838.
[7] Molchanov IS. *Theory of random sets.* London: Springer; 2005.
[8] Rogers LCG, Williams D. *Diffusions, Markov processes, and martingales. Volume One: Foundations*. 2nd ed. Chichester: Wiley; 1994.
[9] Shafer G. *Mathematical theory of evidence.* Princeton: Princeton University Press; 1976.
[10] Soetaert K, Petzoldt T, Woodrow Setzer R. Solving differential equations in R: Package deSolve. *J Statist Softw* 2010; **33**:1–25.
[11] Younes L. *Shapes and diffeomorphisms.* Berlin: Springer; 2010.
[12] Zimmermann HJ. *Fuzzy set theory and its applications.* 4th ed. Dordrecht: Kluwer; 2001.