The Annals of Statistics
1996, Vol. 24, No. 4, 1511–1520

EXTREMAL PROPERTIES OF PRINCIPAL CURVES IN THE PLANE[1]

BY TOM DUCHAMP AND WERNER STUETZLE

University of Washington

Principal curves were introduced to formalize the notion of "a curve passing through the middle of a dataset." Vaguely speaking, a curve is said to pass through the middle of a dataset if every point on the curve is the average of the observations projecting onto it. This idea can be made precise by defining principal curves for probability densities. In this paper we study principal curves in the plane. Like linear principal components, principal curves are critical points of the expected squared distance from the data. However, the largest and smallest principal components are extrema of the distance, whereas all principal curves are saddle points. This explains why cross-validation does not appear to be a viable method for choosing the complexity of principal curve estimates.

1. Introduction. Principal curves were introduced in Hastie (1984) and Hastie and Stuetzle (1989) to formalize the notion of "a curve passing through the middle of a dataset." Vaguely speaking, a curve Γ is said to pass through the middle of a dataset if every point $\mathbf{x}$ on the curve is the average of the observations projecting onto it.

To make this idea precise, Hastie and Stuetzle (1989) define principal curves for probability densities. Let $\mathbf{X}$ denote a two-dimensional random vector distributed according to a probability density p, and let $\Gamma \subset \mathbf{R}^2$ be a smoothly embedded closed interval (*arc*) or circle (*loop*). For each point $\mathbf{x} \in \mathbf{R}^2$, let $d(\mathbf{x}, \Gamma)$ denote the distance from $\mathbf{x}$ to Γ. Because Γ is compact, for each $\mathbf{x} \in \mathbf{R}^2$ the distance $d(\mathbf{x}, \Gamma)$ is realized by at least one point of Γ. Of course, there may be several such points; a point $\mathbf{x}$ with several closest points on the curve is called an *ambiguity point*. The *projection map*

$$\pi_\Gamma : \mathbf{R}^2 \to \Gamma$$

is the map which assigns to each $\mathbf{x} \in \mathbf{R}^2$ a point $\pi_\Gamma(\mathbf{x}) \in \Gamma$ realizing the distance from $\mathbf{x}$ to Γ, that is,

$$d(\mathbf{x}, \Gamma) = \|x - \pi_\Gamma(\mathbf{x})\|.$$

Received February 1993; revised September 1995.
[1]Research supported by DOE Grant DE-FG06-85-ER25006 and NSF Grants CCR-89-57323, DMS-91-03002 and DMS-94-02734.
AMS 1991 *subject classifications*. Primary 62G07; secondary 62J02, 62H25, 62H30.
Key words and phrases. Principal curves, least squares, curve fitting, nonlinear regression, calculus of variations.

The vague notion that every point on the curve should be the average of the observations projecting onto it can now be formalized:

DEFINITION 1 [Hastie and Stuetzle (1989)]. A curve Γ is called *self-consistent* or a *principal curve* of a density p if $E(\mathbf{X}|\pi_\Gamma(\mathbf{X}) = \mathbf{x}) = \mathbf{x}$ for almost every $\mathbf{x} \in \Gamma$.

The notion of projection also leads to a natural definition of the distance between a random vector $\mathbf{X}$, or its associated density, and a curve Γ:

$$d^2(\mathbf{X}, \Gamma) = \mathrm{E}\left(\|\mathbf{X} - \pi_\Gamma(\mathbf{X})\|^2\right).$$

The main result proved in Hastie and Stuetzle (1989) is that principal curves are critical points of the distance in the variational sense:

THEOREM 1. *Let Γ be a principal curve and let Γ_t be a smooth family of curves with $\Gamma_0 = \Gamma$. Then*

$$\frac{d}{dt} d^2(\mathbf{X}, \Gamma_t)\Big|_{t=0} = 0.$$

Linear principal components share this property if Γ_t is restricted to be a smooth family of straight lines. The largest principal component minimizes the distance to $\mathbf{X}$; the smallest principal component maximizes the distance [among all lines passing through $\mathrm{E}(X)$], and the others are saddle points. In this paper we show that *all* principal curves are saddle points of the distance; there are no local minima. This is in notable contrast to the regression problem, where the conditional expectation $\mathrm{E}(Y|x)$ *minimizes* the expected squared distance $\mathrm{E}(Y - f(X))^2$ among all functions f. In the regression case, we can choose model complexity by minimizing an estimate of $\mathrm{E}(Y - f(X))^2$, for example, the cross-validated residual sum of squares. Because principal curves are not minima of the distance, there is no justification for choosing the complexity that minimizes an estimate of the distance, and indeed cross-validation has been observed to fail in practice.

Although principal curves are not local minima, in general, they are local minima of the distance for "low frequency variations." However, the definition of "low-frequency" depends on the principal curve itself, and not only on the underlying density. Thus it is not clear how this fact can be used in an operational way.

Because our main result is essentially negative, we do not attempt to prove it in greatest generality. To avoid various technical problems, we only consider *regular*, closed curves for which the support of the density forms a tubular neighborhood. For such curves, the projection map π is well defined and smooth.

In a companion paper [Duchamp and Stuetzle (1995)] we show that regular principal curves are solutions of a differential equation. By solving this differential equation, we find principal curves for uniform densities on

rectangles and annuli. There are oscillating solutions besides the obvious straight and circular ones, indicating that principal curves, in general, will not be unique. If a density has several principal curves, they have to cross—a property somewhat analogous to the orthogonality of principal components. Finally, we study principal curves for spherical and elliptical distributions.

Notation and conventions. The following notation is used throughout the paper: L_Γ denotes the length of Γ; Λ denotes the circle of circumference L_Γ; $\mathbf{x} = \mathbf{x}(s)$ denotes an arc length parameterization of Γ. The unit tangent and normal vector fields to Γ are written $\mathbf{T}(s)$ and $\mathbf{N}(s)$, respectively, and oriented so that the pair $(\mathbf{T}(s), \mathbf{N}(s))$ is consistent with the standard orientation of $\mathbf{R}^2$. The support Ω of the density p is assumed to be a compact, connected region with smooth boundary. The density is assumed to be strictly positive on the interior of Ω and smooth on all of Ω. The map

$$\lambda: \begin{cases} \Omega \to \Lambda \\ \mathbf{x} \mapsto s = \lambda(\mathbf{x}), \end{cases}$$

defined by the formula $\pi_\Gamma(\mathbf{x}) = \mathbf{x}(\lambda(\mathbf{x}))$, is called the *projection index* [see Hastie and Stuetzle (1989)].

2. Regularity, normal coordinates and transverse moments. Suppose that Γ is a curve for which the support Ω of the density is a tubular neighborhood that does not contain any ambiguity points (see Figure 1). In this case, Ω can be parameterized in terms of *normal coordinates*.

DEFINITION 2. The *normal coordinate map* of Γ is the map $\nu_\Gamma: \Lambda \times \mathbf{R} \to \mathbf{R}^2$ defined by the formula

$$\nu_\Gamma(s, v) = \mathbf{x}(s) + v\mathbf{N}(s)$$

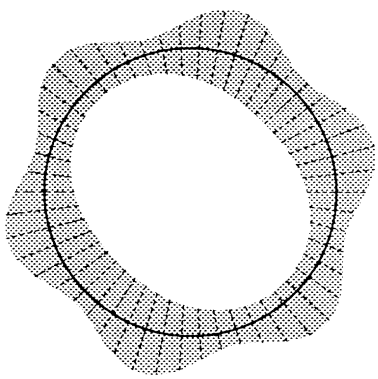

FIG. 1. *A regular loop.*

and the *normal coordinate transformation* is the map $\mu_\Gamma\colon \Omega \to \Lambda \times \mathbf{R}$ defined by the formula

$$\mu_\Gamma(\mathbf{x}) = \big(\lambda(\mathbf{x}), \langle \mathbf{x} - \mathbf{x}(\lambda(\mathbf{x})), \mathbf{N}(\lambda(\mathbf{x}))\rangle\big).$$

The components (s, v) of $\mu_\Gamma(\mathbf{x})$ are called the *normal coordinates* of $\mathbf{x}$.

By virtue of our assumption that Ω does not contain ambiguity points of Γ, the normal map is a left inverse of the normal coordinate transformation μ_Γ:

$$\nu_\Gamma \circ \mu_\Gamma = id_\Omega.$$

We can now state a formal definition of regularity:

DEFINITION 3. A smooth curve $\Gamma \subset \Omega$ is called *regular* if the following conditions are satisfied:

(i) Ω contains no ambiguity points of Γ.
(ii) The map $\mu_\Gamma\colon \Omega \to \Lambda \times \mathbf{R}$ is a diffeomorphism onto its image.
(iii) The image $\mu_\Gamma(\Omega)$ is of the form

$$\mu_\Gamma(\Omega) = \big\{(s, v) \in \Lambda \times \mathbf{R} \,|\, v_-(s) \le v \le v_+(s)\big\},$$

where v_- and v_+ are smooth and $v_-(s) < 0 < v_+(s)$ on the interior of Λ.

Regularity implies that for any $s \in \Lambda$ the set $\mathscr{V}(s) = \{v \,|\, (s, v) \in \mu_\Gamma(\Omega)\}$ of points in Ω projecting onto $\mathbf{x}(s)$ is an interval.

Define

$$\mu_k(s) = \int_{\mathscr{V}(s)} v^k p(\mathbf{x}(s) + v\mathbf{N}(s))\, dv.$$

Then $\mu_\perp(s) = \mu_1(s)/\mu_0(s)$ is the mean of the *transverse density* on $\mathscr{V}(s)$ induced by p, and $\sigma_\perp^2(s) = \mu_2(s)/\mu_0(s) - \mu_\perp^2$ is its variance.

For later reference, we now calculate the Jacobian determinant of the normal map. Recall the *Frenet formulas*

$$\frac{d\mathbf{T}(s)}{ds} = \kappa(s)\mathbf{N}(s) \quad \text{and} \quad \frac{d\mathbf{N}(s)}{ds} = -\kappa(s)\mathbf{T}(s),$$

where $\kappa(s)$ is the curvature of Γ. Then

$$\frac{\partial \nu_\Gamma(s, v)}{\partial s} = \mathbf{x}'(s) + v\mathbf{N}'(s) = (1 - v\kappa(s))\mathbf{T}(s)$$

and

$$\frac{\partial \nu_\Gamma(s, v)}{\partial v} = \mathbf{N}(s).$$

The Jacobian determinant $(\partial(x, y))/(\partial(s, v))$ of the normal coordinate map is now easily computed:

$$\frac{\partial(x, y)}{\partial(s, v)} = \left| \frac{\partial \nu_\Gamma(s, v)}{\partial s} \times \frac{\partial \nu_\Gamma(s, v)}{\partial v} \right| = 1 - v\kappa(s). \tag{1}$$

3. The first variation. The goal of this section is to compute the first (variational) derivative of the distance functional

$$\mathrm{I}[\Gamma] = d^2(\mathbf{X}, \Gamma) = \int_{\mathbf{x} \in \Omega} \|\mathbf{x} - \pi_\Gamma(\mathbf{x})\|^2 p(\mathbf{x})\, d\mathbf{x}$$

at a regular curve Γ. For this purpose consider a smooth *variation* of Γ, that is, a smooth function

$$\Phi\colon \Lambda \times (-\varepsilon, +\varepsilon) \to \mathbf{R}^2$$

for which $\Phi(s, 0)$ is the arc-length parameterization of Γ. Let Γ_t be the curve parameterized by $\Phi(s, t)$. Differentiating Φ with respect to t at $t = 0$ gives a vector field $\mathbf{Y}$ defined along Γ. The fact that $\mathrm{I}[\Gamma_t]$ is independent of the parameterization of Γ_t implies that Φ can be chosen so that $\mathbf{Y}$ is normal to Γ. Thus, $\mathbf{Y}$ can be assumed to be of the form $\mathbf{Y} = f\mathbf{N}$, where f is a real valued function on Λ. We call f an *infinitesimal variation* of Γ.

Let $\lambda\colon \Omega \times (-\varepsilon, \varepsilon) \to \Lambda$ be the projection index associated with the curve $\Phi(s, t)$. It is defined by the condition

$$\Phi(\lambda(\mathbf{x}, t), t) = \pi_{\Gamma_t}(\mathbf{x}),$$

where $\pi_{\Gamma_t}\colon \Omega \to \mathbf{\Gamma}_t \subset \Omega$ is the projection map. One can show that, for sufficiently small t, Γ_t is regular and thus $\lambda(\mathbf{x}, t)$ is a smooth function of $\mathbf{x}$ and t.

THEOREM 2. *Let Γ be a regular curve of p and let Γ_t be a smooth variation of Γ. Then*

$$\left.\frac{d\mathrm{I}[\Gamma_t]}{dt}\right|_{t=0} = -2\int_\Lambda f(s)\{\mu_\perp(s) - \kappa(s)(\sigma_\perp^2(s) + \mu_\perp^2(s))\}\mu_0(s)\, ds,$$

where f is the infinitesimal variation generated by Γ_t.

PROOF. The proof is a direct calculation. Set $I(t) = \mathrm{I}[\Gamma_t]$. Then

$$\begin{aligned} I'(t) &= \frac{d}{dt}\int\int_\Omega \|\mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t)\|^2 p(\mathbf{x})\, d\mathbf{x} \\ &= -2\int\int_\Omega \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \right. \\ &\qquad \left. \left(\frac{\partial \Phi(\lambda(\mathbf{X}, t), t)}{\partial s} \frac{\partial \lambda(\mathbf{X}, t)}{\partial t} + \frac{\partial \Phi(\lambda(\mathbf{X}, t), t)}{\partial t} \right) \right\rangle p(\mathbf{x})\, d\mathbf{x}. \end{aligned}$$

Note that, for $\mathbf{x} \in \Omega$, the vector $\partial\Phi(\lambda(\mathbf{x}, t), t)/\partial s$ is tangential to Γ_t at $\Phi(\lambda(\mathbf{x}, t), t)$. Since $\Phi(\lambda(\mathbf{x}, t), t)$ is the point on Γ_t nearest $\mathbf{x}$ and since Γ_t is

regular, it follows that $\mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t)$ is orthogonal to $\partial\Phi(\lambda(\mathbf{x}, t), t)/\partial s$. Thus,

$$I'(t) = -2\int\int_{\Omega} \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x}. \tag{2}$$

By switching to normal coordinates, we obtain

$$\begin{aligned} I'(0) &= -2\int\int_{\Omega} \langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), f(\pi_\Gamma(\mathbf{x}))\mathbf{N}(\pi_\Gamma(x))\rangle p(\mathbf{x})\, d\mathbf{x} \\ &= -2\int_{\Lambda} \left\{ \int_{\mathscr{V}(s)} vf(s)\, p(\mathbf{x}(s) + v\mathbf{N}(s))(1 - v\kappa(s))\, dv \right\} ds \\ &= -2\int_{\Lambda} f(s)\{\mu_\perp(s) - \kappa(s)(\sigma_\perp^2(s) + \mu_\perp^2(s))\}\mu_0(s)\, ds. \qquad \square \end{aligned}$$

For Γ to be a critical point of the distance functional, $I'(0)$ has to vanish for all choices of f. This is equivalent to the condition that

$$\kappa(s) = \frac{\mu_\perp(s)}{\mu_\perp^2(s) + \sigma_\perp^2(s)}. \tag{3}$$

4. The second variation. We will now compute the second derivative of $\mathrm{I}[\Gamma_t]$.

THEOREM 3. *Let Γ be a regular principal curve of p and let Γ_t be a smooth variation of Γ. Then*

$$\left.\frac{d^2\mathrm{I}[\Gamma_t]}{dt^2}\right|_{t=0} = 2\int_{\Gamma} \left\{ \left(\frac{\sigma_\perp^2}{\mu_\perp^2 + \sigma_\perp^2} \right) f^2 - (\mu_\perp^2 + \sigma_\perp^2) f'^2 \right\} \mu_0\, ds,$$

where f is the infinitesimal variation generated by Γ_t.

PROOF. By (2),

$$I''(t) = -2\frac{d}{dt}\int\int_{\Omega} \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x}.$$

Differentiation inside the integral and an application of the chain rule give $I''(t)$ as the sum of the following four terms:

$$2\int\int_{\Omega} \left\langle \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial s} \frac{\partial\lambda(\mathbf{x}, t)}{\partial t}, \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x}, \tag{4}$$

$$2\int\int_{\Omega} \left\langle \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t}, \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x}, \tag{5}$$

$$-2\int\int_{\Omega} \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial^2\Phi(\lambda(\mathbf{x}, t), t)}{\partial s\, \partial t} \frac{\partial\lambda(\mathbf{x}, t)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x}, \tag{6}$$

$$-2\int\int_{\Omega} \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial^2\Phi(\lambda(\mathbf{x}, t), t)}{\partial t^2} \right\rangle p(\mathbf{x})\, d\mathbf{x}. \tag{7}$$

We will now evaluate the four terms at $t = 0$. By construction, $\partial\Phi(s,0)/\partial s = \mathbf{T}(s)$ and $\partial\Phi(s,0)/\partial t = f(s)\mathbf{N}(s)$. Hence the first term (4) vanishes.

Decomposing the vector field $\partial^2\Phi(s,0)/\partial t^2$ into a tangential component $h_{\|}(s)\mathbf{T}(s)$ and a normal component $h_{\perp}(s)\mathbf{N}(s)$ allows us to rewrite the fourth term (7) as:

$$-2\int\int_{\Omega}\langle \mathbf{x} - \pi_{\Gamma}(\mathbf{x}), h_{\perp}(\lambda(\mathbf{x},0))\mathbf{N}(\lambda(\mathbf{x},0))\rangle p(\mathbf{x})\,d\mathbf{x}.$$

This is a first variation [see (2)] and vanishes because Γ is assumed to be a principal curve.

We can rewrite the second term (5) by changing to normal coordinates and using the fact that Γ is a principal curve:

$$\begin{aligned}
&2\int\int_{\Omega}\left\langle \frac{\partial\Phi(\lambda(\mathbf{x},t),t)}{\partial t}, \frac{\partial\Phi(\lambda(\mathbf{x},t),t)}{\partial t}\right\rangle p(\mathbf{x})\,d\mathbf{x}\\
&\quad = 2\int\int_{\Omega} f^2(\lambda(\mathbf{x},0))\,p(\mathbf{x})\,d\mathbf{x}\\
&\quad = 2\int_{\Lambda} f^2(s)\left\{\int_{\mathscr{V}(s)}(1 - v\kappa(s))\,p(\mathbf{x}(s) + v\mathbf{N}(s))\,dv\right\}ds\\
&\quad = 2\int_{\Lambda} f^2(s)(\mu_0(s) - \mu_1(s)\kappa(s))\,ds\\
&\quad = 2\int_{\Lambda} f^2(s)\left(\frac{\sigma_{\perp}^2(s)}{\mu_{\perp}^2(s) + \sigma_{\perp}^2(s)}\right)\mu_0(s)\,ds.
\end{aligned}$$

It remains to examine the third term (6). We will first rewrite the mixed partial:

$$\begin{aligned}
\frac{\partial^2\Phi(\lambda(\mathbf{x},0),0)}{\partial s\,\partial t} &= \frac{\partial}{\partial s}(f(\lambda(\mathbf{x},0))\mathbf{N}(\lambda(\mathbf{x},0)))\\
&= f'(\lambda(\mathbf{x},0))\mathbf{N}(\lambda(\mathbf{x},0))\\
&\quad - f(\lambda(\mathbf{x},0))\,\kappa(\lambda(\mathbf{x},0))\mathbf{T}(\lambda(\mathbf{x},0)).
\end{aligned}$$

Hence,

$$\begin{aligned}
(8)\quad &-2\int\int_{\Omega}\left\langle \mathbf{x} - \pi_{\Gamma}(\mathbf{x}), \frac{\partial^2\Phi(\lambda(\mathbf{x},0),0)}{\partial s\,\partial t}\,\frac{\partial\lambda(\mathbf{x},0)}{\partial t}\right\rangle p(\mathbf{x})\,d\mathbf{x}\\
&\quad = -2\int\int_{\Omega}\left\langle \mathbf{x} - \pi_{\Gamma}(\mathbf{x}), f'(\lambda(\mathbf{x},0))\mathbf{N}(\lambda(\mathbf{x},0))\,\frac{\partial\lambda(\mathbf{x},0)}{\partial t}\right\rangle p(\mathbf{x})\,d\mathbf{x}.
\end{aligned}$$

To conclude the calculation, we need a formula for $\partial\lambda(\mathbf{x},0)/\partial t$. Observe that

$$\left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x},t),t), \frac{\partial\Phi(\lambda(\mathbf{x},t),t)}{\partial s} \right\rangle = 0$$

identically in t, because $(\mathbf{x} - \Phi(\lambda(\mathbf{x},t),t))$ is normal to Γ_t and $\partial\Phi(\lambda(\mathbf{x},t),t)/\partial s$ is tangential to Γ_t. Differentiating with respect to t at $t=0$ and solving for $\partial\lambda(\mathbf{x},t)/\partial t$ gives

$$(9)\quad \left.\frac{\partial\lambda(\mathbf{x},t)}{\partial t}\right|_{t=0} = f'(\lambda(\mathbf{x},0)) \frac{\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \mathbf{N}(\lambda(\mathbf{x},0))\rangle}{1 - \kappa(\lambda(\mathbf{x},0))\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \mathbf{N}(\lambda(\mathbf{x},0))\rangle}.$$

(The calculation is somewhat lengthy, but proceeds along the same lines as the calculations above.) Substituting (9) into (8) and reverting to normal coordinates, we obtain

$$\begin{aligned} &-2\int\int_\Omega \left\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), f'(\lambda(\mathbf{x},0))\mathbf{N}(\lambda(\mathbf{x},0)) \frac{\partial\lambda(\mathbf{x},0)}{\partial t} \right\rangle p(\mathbf{x})\, d\mathbf{x} \\ &\quad = -2\int\int_\Omega f'^2(\lambda(\mathbf{x},0)) \frac{\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \mathbf{N}(\lambda(\mathbf{x},0))\rangle^2}{1 - \kappa(\lambda(\mathbf{x},0))\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \mathbf{N}(\lambda(\mathbf{x},0))\rangle} p(\mathbf{x})\, d\mathbf{x} \\ &\quad = -2\int_\Lambda f'^2(s)\{\sigma_\perp^2(s) + \mu_\perp^2(s)\}\mu_0(s)\, ds. \end{aligned}$$

Collecting terms gives the final formula for the second variation. □

From Theorem 3 it is easy to see why all regular principal curves are saddle points of I: Choosing a variation with constant f leads to a positive second derivative, whereas choosing a variation with large f'^2 and small f (in the L^2 sense) leads to a negative second derivative. In particular, regular principal curves are never local minima.

5. Low frequency variations. In this section we show that principal curves are local minima of the distance if we restrict ourselves to low frequency variations. We fist examine the case of the circular principal curve for the uniform distribution on an annulus, and then discuss the general case.

5.1. *The annulus.* Consider the uniform distribution on the annulus

$$\Omega_{R_1,R_2} = \{(r,\phi): R_1 \le r \le R_2\},$$

where (r,ϕ) are polar coordinates.

For symmetry reasons, an annulus has a circular principle curve. Using (3) it is not hard to show that

$$r_{\text{circ}} = \frac{2(R_1^2 + R_1R_2 + R_2^2)}{3(R_1 + R_2)}.$$

Let $f(s)$ be the infinitesimal variation generated by a smooth variation Γ_t of the principal circle.

Observing that the moments μ_0, $\mu_\perp$ and $\sigma_\perp^2$ are constant, and using Theorem 3, we obtain

$$I''(0) = \frac{2\mu_0\sigma_\perp^2}{\mu_\perp^2 + \sigma_\perp^2}\int_\Gamma f^2(s)\,ds - 2\mu_0\left(\mu_\perp^2 + \sigma_\perp^2\right)\int_\Gamma f'^2(s)\,ds. \tag{10}$$

We express the second variation $I''(0)$ in terms of the Fourier expansion of $f(s)$:

$$f(s) = a_0 + \sum_{n=1}^{\infty} a_n \cos(\omega_n s) + b_n \sin(\omega_n s),$$

where $\omega_n = 2\pi n/L_\Gamma$ are the Fourier frequencies. Substituting

$$\int_\Gamma f^2(s)\,ds = \frac{L_\Gamma}{2}\left(2a_0^2 + \sum_{n=1}^{\infty}\left(a_n^2 + b_n^2\right)\right),$$

$$\int_\Gamma f'^2(s)\,ds = \frac{L_\Gamma}{2}\sum_{n=1}^{\infty}\omega_n^2\left(a_n^2 + b_n^2\right)$$

into (10) and simplifying gives

$$I''(0) = \mu_0 L_\Gamma\left\{\frac{2\sigma_\perp^2}{\mu_\perp^2 + \sigma_\perp^2}a_0^2 + \sum_{n=1}^{\infty}\left(\frac{\sigma_\perp^2}{\left(\mu_\perp^2 + \sigma_\perp^2\right)} - \omega_n^2\left(\mu_\perp^2 + \sigma_\perp^2\right)\right)\left(a_n^2 + b_n^2\right)\right\}.$$

Thus, $I''(0)$ will be positive, and the principal circle will be a local minimum of the distance functional, within the space of variations with frequency

$$\omega_n < \frac{\sigma_\perp}{\mu_\perp^2 + \sigma_\perp^2}.$$

5.2. *The general case*. In the previous example the transverse moments were constant along the principal curve Γ. This allowed us to obtain a simple expression for the second variation $I''(0)$ in terms of the Fourier coefficients of f. Similar computations can be carried out if the transverse moments are approximately constant. In this case, $I''(0)$ can be expressed in terms of a (generalized) Fourier expansion of the infinitesimal variation f with respect to a basis $\{f_n\}_{n=1}^{\infty}$ of functions on Γ that is naturally associated with the transverse moments of p. The details of how this basis is constructed can be found in Duchamp and Stuetzle (1993).

There are three important features of this basis:

1. The functions f_n are approximately periodic functions whose frequencies increase with increasing n.
2. If $f = \sum_{n=1}^{\infty} a_n f_n$, then

$$I''(0) = \sum_{n=1}^{\infty}(1 - \lambda_n)|a_n|^2,$$

where λ_n is a monotone nondecreasing sequence of positive real numbers.

3. $\lambda_n > 1$ when the frequency is greater than $\sigma_\perp(s)/(\mu_\perp^2(s) + \sigma_\perp^2(s))$ and $\lambda_n < 1$ when the frequency is smaller than $\sigma_\perp(s)/(\mu_\perp^2(s) + \sigma_\perp^2(s))$.

Thus, $I''(0) > 0$ if the Fourier expansion of f only contains terms of frequency lower than $\sigma_\perp(s)/(\mu_\perp^2(s) + \sigma_\perp^2(s))$. Therefore, Γ is a local minimum of I within the class of such variations. A more precise statement and a proof are presented in Duchamp and Stuetzle (1993).

6. Conclusion. Principal curves were invented to formalize the concept of "a curve passing through the middle of a data set". In this paper we show that, while principal curves are critical points of the distance function, they are never local minima. Thus there is no justification for the use of cross-validation to determine the complexity of principal curve estimates, and indeed cross-validation has been observed to fail in practice.

To our knowledge, nobody has as yet suggested a reasonably motivated, automatic method for choice of model complexity in the context of manifold estimation or nonparametric orthogonal distance regression. This remains an important open problem.

REFERENCES

DUCHAMP, T. and STUETZLE, W. (1993). The geometry of principal curves in the plane. Report TR 250, Dept. Statistics, Univ. Washington. Revised August 1995. Available through http://www.stat.washington.edu:80/tech.reports/

DUCHAMP, T. AND STUETZLE, W. (1996). Geometric properties of principal curves in the plane. In *Robust Statistics*, *Data Analysis*, *and Computer Intensive Methods* (Helmut Rieder, ed.) Springer Lecture Notes in Statistics **109** 135–152.

HASTIE, T. J. (1984). Principal curves and surfaces. Ph.D. dissertation, Stanford Univ.

HASTIE, T. J. and STUETZLE, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84** 502–516.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
E-MAIL: duchamp@math.washington.edu

DEPARTMENT OF STATISTICS
B313 PADELFORD HALL
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
E-MAIL: wxs@stat.washington.edu