

The Geometry of Principal Curves in the Plane

Tom Duchamp

Department of Mathematics

Werner Stuetzle

Department of Statistics

University of Washington

Seattle, WA 98195

March 18, 1994

This research was supported by DOE grant DE-FG06-85-ER25006 and NSF grants CCR-8957323 and DMS-9103002

1991 Mathematics Subject Classification. Primary 62G07; secondary 62J02, 62H25, 62H30

Key words and phrases. Principal curves, least-squares, curve fitting, nonlinear regression, calculus of variations.

Abstract

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. Principal curves can be regarded as a generalization of linear principal components — if a principal curve happens to be a straight line, then it is a principal component. In this paper we study principal curves in the plane. We calculate the principal curves for uniform densities on rectangles and annuli and show that there are oscillating solutions besides the obvious straight respectively circular ones. This indicates 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. Like principal components, principal curves are critical points of the expected squared distance to the data. However, the largest and smallest principal components are extrema of the distance, whereas all principal curves are saddlepoints. This explains why cross-validation does not appear to be a viable method for choosing the complexity of principal curve estimates.

Contents

1	Introduction	4
2	Principal Curves as Solutions of a Differential Equation	12
2.1	Informal Discussion of Admissibility	12
2.2	Transversality Conditions	14
2.3	Formal Definition of Admissibility	15
2.4	Self-consistency as a Curvature Condition	18
2.5	Self-consistency as a Differential Equation	20
3	Principal Curves Cross	23
4	Principal Curves for Uniform Densities	25
4.1	Principal Curves on the Rectangle	26
4.2	Principal Curves on the Annulus	29
5	Extremal Properties of Principal Curves	33
6	Conclusions and Open Problems	39
A	Proof of Theorem 3	41
A.1	Variational Vector Fields	41
A.2	Smoothness of the Projection π_{Γ}	43
A.3	Computation of the Second Derivative	45
B	Proof of Theorem 4	47
B.1	The Sturm-Liouville Problem	48
B.2	The Natural Basis	49
B.3	Eigenvalue Estimates	50
B.4	Extremal Properties	53

1 Introduction

The problem of fitting one or two-dimensional manifolds to point sets in two respectively three dimensions occurs in a variety of contexts, such as modeling of object boundaries in two or three-dimensional images (Banfield and Raftery (1992), Brinkley (1985), Martin et al (1993), Schudy and Ballard (1978, 1979), Sheehan et al (1992)), and reconstruction of objects from range data (Fang and Gossard (1992), Goshtasby (1992), Hoppe et al (1992, 1993), Muraki (1991), Solina and Bajcsy (1990), Vemuri et al (1986)). It is typically not reasonable to assume that the unknown manifold is of a given parametric form, like an ellipsoid. Fitting methods need to be flexible and able to accomodate a variety of shapes.

Manifold fitting is fundamentally different from regression. It is worthwhile to contrast the respective goals for the case of two-dimensional data. In regression, we are given points $(x_1, y_1), \dots, (x_n, y_n)$. The goal is to find a *function* $f(x)$ summarizing the dependence of the response variable Y on the predictor variable X ; the two variables are thus treated asymmetrically. Under the assumption that f is linear, a common choice is the least squares straight line. There has been a large amount of research on nonparametric regression methods that make only very general assumptions about the nature of f .

In manifold fitting we are also given points $\mathbf{x}_1, \dots, \mathbf{x}_n \in R^2$. The goal is to find a *one-dimensional manifold* Γ summarizing the association between the variables X_1 and X_2 . The two variables are treated symmetrically. It is usually clear from the context whether the manifold should be topologically a circle or a closed interval. Under the assumption that Γ is a straight line, a common choice is the largest principal component. Nonparametric methods not relying on the linearity assumption have typically been crafted on an ad-hoc basis. This is not satisfactory, and a theoretical underpinning would be highly desirable.

To statistically analyze the behaviour of fitting methods we need a stochastic model that is thought to give rise to the data. A simple model is to assume that the data points are i.i.d. observations of a two-dimensional random vector distributed according to some (unknown) density p . Then two questions arise: (1) Which characteristic of the density we are trying to estimate, and why? (2) How can we estimate this characteristic?

In the regression context answers to those questions are well understood. We usually estimate the conditional expectation $E(Y | x)$, because it minimizes the expected squared prediction error $E_p(Y - f(X))^2$ among all functions f . There are many approaches to estimating $E(Y | x)$, often based on local averaging. The amount of averaging or, more generally, the complexity of the model is typically chosen to minimize an estimate of expected squared prediction error, like the cross-validated residual sum of squares.

In the context of manifold estimation the situation is not as clear cut — there are no generally accepted answers to questions (1) and (2) above. In this paper we discuss answers based on the concept of *principal curves* (PC's). PC's 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 is said to pass through the middle of a dataset if every point \mathbf{x} on the curve Γ is the average of the observations projecting onto it.

To make this idea precise, Hastie and Stuetzle (1989) define PC's for probability densities.

Definition of principal curves: Let \mathbf{X} denote a two-dimensional random vector distributed according to the 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*

$$\pi_\Gamma : \mathbf{R}^2 \rightarrow \Gamma,$$

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 , i.e.

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

Notice that the map π_Γ is well-defined only on the complement of the set of ambiguity points of Γ . But the set of ambiguity points has Lebesgue measure zero (see Hastie and Stuetzle (1989)) and can be ignored in probability calculations. It is not difficult to show that π_Γ is continuous on the complement of the set of ambiguity points.

The vague concept that every point on the curve 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 $E(\mathbf{X} \mid \pi_\Gamma(\mathbf{X}) = \mathbf{x}) = \mathbf{x}$ 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 :

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

Principal curves as generalizations of linear principal components:

Besides formalizing the notion of “a curve passing through the middle of a dataset”, principal curves share two properties with linear principal components, which make them appear as a natural generalization (Hastie and Stuetzle (1989)):

- If a PC happens to be a straight line, then it is a (linear) principal component.
- PC’s are critical points of the distance function in the variational sense: let Γ be a PC, and let Γ_t be a smooth family of curves with $\Gamma_0 = \Gamma$, then

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

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, and the others are saddlepoints.

A principal curve algorithm: We will first discuss a suggestion for finding PC’s of densities. Given our motivating problem of manifold estimation, we are particularly interested in PC’s with small distance from \mathbf{X} . A strategy is to start with a smooth curve such as the largest linear principal component, and check if it is a principal curve by calculating the conditional expectation $E(\mathbf{X} \mid \pi_\Gamma(\mathbf{X}) = \mathbf{x})$. Either this conditional expectation coincides with Γ , or we get a new curve as a by-product. We then check if the new curve is self-consistent, and so on. If the self-consistency condition is met, we have found a principal curve. It is easy to show that both

the operations of projection and conditional expectation always reduce the expected distance from the points to the curve.

As discussed in Hastie and Stuetzle (1989) there are potential problems with the above algorithm. For example there is no guarantee that the curves produced by the conditional expectation step of the algorithm are differentiable. Still there is some evidence in its favor:

- By definition principal curves are fixed points of the algorithm.
- Assuming that each iteration is well defined and produces a differentiable curve, the distance from the curves to \mathbf{X} converges.
- If the conditional expectation operation in the algorithm is replaced by fitting a least squares straight line, then the procedure converges to the largest principal component of the density p .

So far we have considered principal curves of a density. In reality, however, we always have to work with a finite data set $\mathbf{x}_1, \dots, \mathbf{x}_n$. The idea is to replace the conditional expectation step by scatterplot smoothing. The algorithm alternates between a projection step and a smoothing step. In the projection step we project the data onto the current curve, thereby assigning to each data point \mathbf{x}_i a parameter value λ_i which is the arc-length measured from some (arbitrary) starting point on the curve. In the smoothing step we obtain a new curve by smoothing the two coordinates of the data points against the λ_i . The algorithm is discussed in more detail in Hastie and Stuetzle (1989).

Using a scatterplot smoother (or nonparametric regression procedure) to estimate the coordinate functions of the curve Γ brings up the issue of model complexity: how do we choose the trade-off between fidelity to the data and smoothness? An obvious idea is to minimize the cross-validated squared distance between data points and curve, the natural analog to cross-validation in the regression context. However, as Hastie and Stuetzle (1989) discovered empirically, this does not seem to work. If cross-validation is used for span selection in the projection-smoothing loop, the complexity of the fit continues to increase, till finally the curve interpolates the data points.

Summary of results: The goal of the current paper is to further contribute to the theoretical understanding of PC's. We now present an informal synopsis of our results:

1. Suppose that Γ is a PC for a density p . Under appropriate conditions on Γ and p , the self-consistency condition can be interpreted as a 2nd order ordinary differential equation for Γ . Principal curves that satisfy these conditions and thus are solutions of the ODE are called *admissible*. The results in items (2) ... (5) apply only to admissible PC's.
2. Any two PC's of a density intersect.
3. The PC's for the uniform density on a rectangular strip can be completely characterized. Depending on the aspect ratio of the rectangle, there may be a number of periodically oscillating PC's (besides the obvious straight ones). Their periods can be obtained as the values of certain elliptic integrals.
4. For certain annuli, the uniform density has periodically oscillating PC's.
5. All PC's are saddle points of the distance $d^2(\mathbf{X}, \Gamma)$ — the distance has no local minima. However, PC's are local minima in the variational sense if we restrict ourselves to variations whose “wavelength” is long relative to the variability of \mathbf{X} in directions orthogonal to Γ .

Result (2) — any two PC's intersect — is analogous to the orthogonality property of linear principal components. Results (3) and (4) show that, in general, densities can be expected to have many PC's. Result (5) — PC's are never local minima of the distance — is of particular interest, as it explains the empirical observation that cross-validation is not a viable method for choice of model complexity in the estimation of PC's. PC's are not minima of the distance, hence there is no justification for choosing the complexity that minimizes an estimate of the distance.

Notation and conventions: The following notation is used throughout the paper: L_Γ denotes the length of Γ , Λ denotes the closed interval $[0, L_\Gamma]$, $\mathbf{x} = \mathbf{x}(s)$ 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 angle between the positive x -axis and $\mathbf{T}(s)$ is denoted by $\theta(s)$.

If g is a scalar- or vector-valued function defined on Γ , then the value of g at $\mathbf{x} = \mathbf{x}(s)$ is denoted by both $g(\mathbf{x})$ and $g(s)$; and g' denotes the derivative of g with respect to the arc length parameter. For example, we write the Frenet formula defining the curvature in either of the forms $\mathbf{T}'(\mathbf{x}) = \kappa(\mathbf{x})\mathbf{N}(\mathbf{x})$ or $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$.

The map

$$\lambda : \begin{cases} \mathbf{R}^2 & \rightarrow \Lambda \\ \mathbf{x} & \mapsto s = \lambda(\mathbf{x}) \end{cases}$$

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

Finally, throughout this paper $\Omega \subset \mathbf{R}^2$ denotes a compact, connected region with smooth boundary $\partial\Omega$. (In certain cases, $\partial\Omega$ is allowed to have corners.). The density p is assumed to be supported on Ω , strictly positive on the interior of Ω , and smooth on all of Ω .

2 Principal Curves as Solutions of a Differential Equation

Suppose that Γ is a PC for the density p . The goal of this section is to show that, under appropriate conditions on Γ and p , the function $\mathbf{x} = \mathbf{x}(s)$ is the solution of a 2nd order ordinary differential equation. This provides a way of actually finding PC's for certain densities: We solve the ODE, obtaining a solution $\mathbf{x} = \mathbf{x}(s)$, and then check, *post facto*, whether the conditions are satisfied. If they are, then Γ is a PC.

2.1 Informal Discussion of Admissibility

As mentioned in the introduction, most of our results apply only to a special class of PC's called *admissible* PC's. It is necessary to restrict the class of PC's, for even relatively simple densities support too many PC's, as the next example shows:

Example 1. Consider the uniform density supported on the region $\Omega_{a,b}$ bounded by a rectangle of width a and length b . The lines of symmetry are easily seen to be PC's. But so is any curve Γ intersecting $\Omega_{a,b}$ in a set of n equally spaced, vertical line segments, arranged so that the right- and

left-most segments are at distance $b/(2n)$ from the right and left edges of the rectangle (see Figure 1). The number n of wiggles in Γ can be arbitrarily large. Even worse, Γ can be deformed outside of $\Omega_{a,b}$ in any way we like. Thus, the uniform distribution on the rectangular strip supports infinitely (in fact, uncountably) many principal curves.

We will now argue that in a typical scenario calling for the use of PC's pathological curves like the ones in the example would not be of interest:

Consider a smooth curve $C \subset \mathbf{R}^2$. The process of randomly sampling points of C , with noise, gives rise to a probability density p . Given sufficient information about the sampling process, it is possible to recover C using convolution methods. In the absence of such information, however, we can only hope to find an approximation Γ to C . We will not be able to recover features of C which are small relative to the sampling error, and therefore our approximation Γ should not have such features. It is also reasonable to assume that the support Ω of p is obtained by “fattening” C , and the same should be true for Γ .

More formally, Γ should satisfy the following two conditions:

1. The support Ω of p contains no ambiguity points of Γ .
2. For each $\mathbf{x} \in \Omega$ the entire line segment joining \mathbf{x} to $\pi_{\Gamma}(\mathbf{x})$ lies in the support of p . In particular, Γ is a subset of Ω .

Notice that the pathological principal curves of Example 1 satisfy neither of these requirements.

Roughly speaking, a PC Γ is called *admissible* if the two conditions above are satisfied. In the case where Γ has endpoints, the condition that it be a principal curve imposes upon it additional *transversality conditions* (see Section 2.2 below). Here is a further example illustrating the concept of admissibility:

Example 2. Consider the uniform distribution on the annular region Ω_{R_1, R_2} , bounded by two concentric circles of radii $R_1 < R_2$ shown in Figure 2. Symmetry considerations show that any line Γ_1 through the origin is a PC, though it is not admissible because it is not contained in Ω_{R_1, R_2} . In Section 4.2 it is shown that Γ_2 , the concentric circle of radius $2/3 (R_1^2 + R_1 R_2 + R_2^2)(R_1 + R_2)^{-1}$ is an admissible PC. Notice that Ω_{R_1, R_2} is obtained by “fattening” Γ_2 .

2.2 Transversality Conditions

Suppose that $\Gamma \subset \Omega$ is a principal arc of p . We wish to study the endpoints of Γ .

Proposition 1. *If $\Gamma \subset \Omega$ is a principal curve of a probability density p then the identity*

$$\langle (\mathbf{x} - \pi_\Gamma(\mathbf{x})), \mathbf{T}(\pi_\Gamma(\mathbf{x})) \rangle = 0$$

is satisfied for all $\mathbf{x} \in \Omega$.

Proof. Choose a point $\mathbf{x} \in \Omega$. If $\pi_\Gamma(\mathbf{x})$ is an interior point of Γ then the identity follows from the fact that $\pi_\Gamma(\mathbf{x})$ is a point of Γ realizing the distance between \mathbf{x} and Γ .

Suppose that $\mathbf{x}_0 = \pi_\Gamma(\mathbf{x})$ is an endpoint of Γ . Without loss of generality, we may assume that the orientation of Γ has been chosen so that $\mathbf{T}_0 \equiv \mathbf{T}(\mathbf{x}_0)$ is outward-pointing. Note that the inequality $\langle (\mathbf{y} - \mathbf{x}_0), \mathbf{T}_0 \rangle \geq 0$ is satisfied for all $\mathbf{y} \in \Omega$ such that $\pi_\Gamma(\mathbf{y}) = \mathbf{x}_0$. For otherwise, the distance from \mathbf{y} to Γ would be strictly less than $\|\mathbf{y} - \mathbf{x}_0\|$. Consequently, the subset $\pi_\Gamma^{-1}(\mathbf{x}_0) \cap \Omega$ is contained in the half-plane $H = \{\mathbf{x} \in \mathbf{R}^2 : \langle (\mathbf{x} - \mathbf{x}_0), \mathbf{T}_0 \rangle \geq 0\}$.

Suppose that the proposition is false. Then $\langle (\mathbf{y} - \mathbf{x}_0), \mathbf{T}_0 \rangle > 0$ for some point $\mathbf{y} \in \Omega$ with $\pi_\Gamma(\mathbf{y}) = \mathbf{x}_0$. Since p is continuous on Ω and strictly positive on the interior of Ω , there is an open set $Q \subset H$, such that (i) $\pi_\Gamma(Q) = \mathbf{x}_0$, (ii) $p > 0$ on Q and (iii) $\langle (\mathbf{y} - \mathbf{x}_0), \mathbf{T}_0 \rangle > 0$ for all $\mathbf{y} \in Q$. But these conditions together imply that the point $E(\mathbf{X} | \pi_\Gamma(\mathbf{X}))$ is contained in the interior of H , violating the self-consistency condition $E(\mathbf{X} | \pi_\Gamma(\mathbf{X}) = \mathbf{x}_0) = \mathbf{x}_0$ (\mathbf{x}_0 is on the boundary of H). ■

Corollary 1. *If $\Gamma \subset \Omega$ is a principal arc then*

1. *the endpoints of Γ lie on the boundary $\partial\Omega$;*
2. *Γ intersects $\partial\Omega$ orthogonally;*
3. *the endpoints of Γ cannot be concave points of $\partial\Omega$.*

A curve $\Gamma \subset \Omega$ (not necessarily a PC) with these three properties is said to satisfy the *transversality conditions*.

2.3 Formal Definition of Admissibility

To give a formal definition of admissibility, we require the notion of normal coordinates in Ω :

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

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

and the normal coordinate transformation is the map $\mu_\Gamma : \Omega \rightarrow \Gamma \times \mathbf{R}$ defined by the formula

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

In terms of the arc length parameter s and the projection index λ ,

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

and

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

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 admissibility:

Definition 3. *A smooth curve $\Gamma \subset \Omega$ is said to be admissible if the following conditions are satisfied:*

1. Ω contains no ambiguity points of Γ .
2. There are two bounded, continuous functions, $v_+, v_- : \Gamma \rightarrow \mathbf{R}$ with the following properties:
 - (a) they are smooth on the interior of Γ ;
 - (b) $v_-(\mathbf{x}) < 0 < v_+(\mathbf{x})$ for all $\mathbf{x} \in \Gamma$ interior to Γ ; and
 - (c) $\mu_\Gamma(\Omega) = \{(\mathbf{x}, v) \in \Gamma \times \mathbf{R} : v_-(\mathbf{x}) \leq v \leq v_+(\mathbf{x})\}$.

3. The map $\nu_\Gamma : \mu_\Gamma(\Omega) \rightarrow \Omega$ is a diffeomorphism.

The space of all admissible curves of Ω is denoted by the symbol $\mathcal{E}_{Adm,\Omega}$.

Notice that, by virtue of condition (3), admissible curves satisfy the transversality conditions.

Recall that we have required the boundary of Ω to be smooth. This requirement can be relaxed somewhat to allow for corners, provided that they project onto the endpoints of Γ . In this case one or both of the functions v_\pm are non-vanishing at an endpoint of Γ . However, we only wish to allow corners if the induced probability density on the normal lines of the endpoints is non-degenerate:

Definition 4. A boundary point $\mathbf{x} \in \partial\Omega$ is said to be essential if $\partial\Omega$ contains a line segment with \mathbf{x} in its interior and if the probability density function is strictly positive at \mathbf{x} .

Figure 3 illustrates the three types of regions we allow and an admissible curve in each.

2.4 Self-consistency as a Curvature Condition

We will now rewrite the self-consistency condition in normal coordinates.

To do this, we have to compute the Jacobian determinant of the normal coordinate map. Consider an admissible curve Γ . Recall that the *curvature function* $\kappa = \kappa(s)$ is given by the formula

$$\kappa = \frac{d\theta}{ds}.$$

Let \mathbf{i} and \mathbf{j} denote the standard unit vectors in \mathbf{R}^2 . Differentiation with respect to s of the identities

$$\mathbf{T} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j} \quad \text{and} \quad \mathbf{N} = -\sin(\theta)\mathbf{i} + \cos(\theta)\mathbf{j}$$

yields the *Frenet formulas*

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N} \quad \text{and} \quad \frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T}.$$

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 $\frac{\partial(x, y)}{\partial(s, v)}$ of the normal coordinate map is now easily computed:

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

Remark 1. The condition that ν_Γ be a diffeomorphism implies that the Jacobian is nowhere vanishing on the set $\mu_\Gamma(\Omega)$. Since the points $(s, 0)$ are all contained in $\mu_\Gamma(\Omega)$, continuity implies that the inequality $v\kappa(s) < 1$ holds for all (s, v) satisfying the inequality $v_- < v < v_+$.

Recall that Γ is called self-consistent if $E(\mathbf{X} | \pi_\Gamma(\mathbf{X}) = \mathbf{x}) = \mathbf{x}$ for almost all $\mathbf{x} \in \Gamma$. By definition of conditional expectation this is equivalent to the condition that

$$\int_{\pi_\Gamma^{-1}(A)} \mathbf{x} p(\mathbf{x}) \, d\mathbf{x} = \int_{\pi_\Gamma^{-1}(A)} \pi_\Gamma(\mathbf{x}) p(\mathbf{x}) \, d\mathbf{x}$$

or

$$\int_{\pi_\Gamma^{-1}(A)} (\mathbf{x} - \pi_\Gamma(\mathbf{x})) p(\mathbf{x}) \, d\mathbf{x} = 0 \quad (2)$$

for all measurable $A \subset \Gamma$. If the normal map is a diffeomorphism, then equation (2) can be rewritten in normal coordinates:

$$\int_{\{(s, v) \in \mu_\Gamma(\Omega) | s \in A\}} v p(\mathbf{x}(s) + v\mathbf{N}(s)) \frac{\partial(x, y)}{\partial(s, v)} \, dv \, ds = 0,$$

where A now denotes a measurable subset of Λ . This implies that

$$\int_{\{v | (s, v) \in \mu_\Gamma(\Omega)\}} v p(\mathbf{x}(s) + v\mathbf{N}(s)) \frac{\partial(x, y)}{\partial(s, v)} \, dv = 0 \quad s - \text{a.e.}$$

Define

$$\mathcal{V}(s) = \{v | (s, v) \in \mu_\Gamma(\Omega)\} = \{v | v_-(s) \leq v \leq v_+(s)\}.$$

Geometrically, $\mathcal{V}(s)$ can be thought of as the “domain of attraction” of $\mathbf{x}(s)$: all points $\mathbf{x}(s) + v\mathbf{N}(s)$ with $v \in \mathcal{V}(s)$ project onto $\mathbf{x}(s)$. Using equation (1) and recalling that p is smooth, self-consistency implies that

$$\int_{\mathcal{V}(s)} v p(\mathbf{x}(s) + v\mathbf{N}(s)) \, dv - \kappa(s) \int_{\mathcal{V}(s)} v^2 p(\mathbf{x}(s) + v\mathbf{N}(s)) \, dv = 0, \quad (3)$$

for all $s \in \Lambda$.

Equation (3) relates the curvature of Γ to first and second moments of the density induced on the normal line to the curve at s . Figure 4 illustrates the situation for the case of a uniform density. If $\mathbf{x}(s)$ happens to coincide with the mean of the induced density, i.e. the center of the normal line segment, then the curvature $\kappa(s)$ has to vanish. If the mean of the induced density falls below $\mathbf{x}(s)$, then the curvature has to be negative, whereas in the opposite case it has to be positive. This makes intuitive sense: Consider an infinitesimal segment of the curve centered at $\mathbf{x}(s)$. As shown in Figure 5, the set of points projecting onto this segment is wedge shaped. If the center of the normal line segment falls below $\mathbf{x}(s)$, then the part of the wedge below the curve is longer than the part above the curve. In order for the mean of the segment to fall on the curve, the segment thus has to open up as we go upwards, meaning the curvature of Γ has to be negative.

2.5 Self-consistency as a Differential Equation

Equation (3) can be thought of as a differential equation satisfied by principal curves. Unfortunately, it cannot be used to find principal curves because the integration boundaries in the moment integrals themselves depend on the curve. In order to use a differential equation to find principal curves, we proceed differently: We define a differential equation in such a way that its solution curves, *when they are admissible*, are actually principal curves.

Our new differential equation is expressed in terms of *transverse moments* of p .

Definition 5. *The k -th transverse moment of the density p at (\mathbf{x}, θ) is the function $\mu_k : \Omega \times S^1 \rightarrow \mathbf{R}$ defined by the formula*

$$\mu_k(\mathbf{x}, \theta) = \int_{v_-(\mathbf{x}, \theta)}^{v_+(\mathbf{x}, \theta)} v^k p(\mathbf{x} + v\mathbf{N}(\theta)) dv.$$

The integration boundaries $v_+ = v_+(\mathbf{x}, \theta)$ and $v_- = v_-(\mathbf{x}, \theta)$ are determined by the condition that $\mathbf{x} + v_+\mathbf{N}$ and $\mathbf{x} + v_-\mathbf{N}$ be the boundaries of the line segment around \mathbf{x} obtained by intersecting the line $\{\mathbf{x} + v\mathbf{N} : v \in \mathbf{R}\}$ with the support Ω of p . This line segment is called the transverse line segment at (\mathbf{x}, θ) and denoted by $\ell(\mathbf{x}, \theta)$. Similarly $\mathcal{V}(\mathbf{x}, \theta) = \{v : v_-(\mathbf{x}, \theta) \leq v \leq v_+(\mathbf{x}, \theta)\}$ is called the transverse interval of Ω at (\mathbf{x}, θ) .

The moments $\mu_1(\mathbf{x}, \theta)$ and $\mu_2(\mathbf{x}, \theta)$ can be written in terms of the mean and variance of the induced probability density on $\mathcal{V}(\mathbf{x}, \theta)$.

Definition 6. *The density*

$$p_{(\mathbf{x}, \theta)}^\perp(v) = \frac{p(\mathbf{x} + v \mathbf{N}(\theta))}{\mu_0(\mathbf{x}, \theta)}, \quad v_-(\mathbf{x}, \theta) \leq v \leq v_+(\mathbf{x}, \theta).$$

on $\mathcal{V}(\mathbf{x}, \theta)$ is called the transverse density at (\mathbf{x}, θ) . Its mean and variance, written $\hat{v}_\perp = \hat{v}_\perp(\mathbf{x}, \theta)$ and $\sigma_\perp^2 = \sigma_\perp^2(\mathbf{x}, \theta)$, are called, respectively, the transverse mean and the transverse variance at (\mathbf{x}, θ) .

A straightforward calculation then yields the identities

$$\frac{\mu_1(\mathbf{x}, \theta)}{\mu_0(\mathbf{x}, \theta)} = \hat{v}_\perp(\mathbf{x}, \theta) \tag{4}$$

$$\frac{\mu_2(\mathbf{x}, \theta)}{\mu_0(\mathbf{x}, \theta)} = \hat{v}_\perp(\mathbf{x}, \theta)^2 + \sigma_\perp^2(\mathbf{x}, \theta). \tag{5}$$

Consider now the second order differential equation

$$\mu_1(\mathbf{x}(s), \theta(s)) - \kappa(s) \mu_2(\mathbf{x}(s), \theta(s)) = 0 \tag{6}$$

for a curve $\mathbf{x} = \mathbf{x}(s)$. (We do *not* assume here that $\mathbf{x} = \mathbf{x}(s)$ is necessarily admissible.) It is convenient to rewrite this equation as a system of first order differential equations, which we call the *principal curve equations*:

$$\frac{d\mathbf{x}}{ds} = \cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}; \quad \frac{d\theta}{ds} = \frac{\mu_1(\mathbf{x}, \theta)}{\mu_2(\mathbf{x}, \theta)} = \frac{\hat{v}_\perp(\mathbf{x}, \theta)}{\hat{v}_\perp(\mathbf{x}, \theta)^2 + \sigma_\perp^2(\mathbf{x}, \theta)}. \tag{7}$$

The first equation encodes the condition that s is an arc length parameter, the second equation is related to the self-consistency condition defining principal curve.

Suppose that Γ is an admissible curve. It is clear from the definitions of transverse moments and from the calculations of Section 2.4 that the terms of equations (3) and (6) coincide. This simple observation yields the next theorem.

Theorem 1. *An admissible curve Γ is a principal curve of the density p if and only if it is a solution of the system of equations (7).*

Remark 2. It is worth noting that the system (7) may be singular along the boundary of Ω . Suppose that Γ is a principal curve and denote its initial and terminal endpoints by \mathbf{x}_- and \mathbf{x}_+ respectively. Next let $\kappa_{\partial\Omega,\pm}$ denote the curvatures of the boundary $\partial\Omega$ at these two points. By virtue of the transversality conditions, $\kappa_{\partial\Omega,\pm}$ cannot be negative at either endpoint (for otherwise the local concavity of Ω would force a violation of transversality). In the case where the curvature is strictly positive, all transverse moments μ_k vanish; and the self-consistency condition degenerates at the boundary. We do not present an analysis of the boundary behavior of solutions of (7) here.

Remark 3. We have made the *a priori* assumption that principal curves are smooth. In fact, we need only assume principal curves to be of class C^2 . The additional smoothness is a consequence of Theorem 1. To see this, notice that (within the class of admissible curves) the right hand side of equation (7) depends smoothly on \mathbf{x} and θ . We now apply a bootstrapping argument: the derivative $d\theta/ds$ has as many derivatives as $\mathbf{x}(s)$ and $\theta(s)$. Since $\mathbf{x}(s) = \int \mathbf{T}(\theta(s)) ds$, $\mathbf{x}(s)$ has one more derivative than $\theta(s)$. Smoothness of $\theta(s)$ and $\mathbf{x}(s)$ follows by induction.

3 Principal Curves Cross

We show here that any two admissible principal curves of a density must cross. The proof makes use of the following lemma¹, which can be proved by doing computations in the normal coordinate system of one of the curves. We leave the proof to the reader, but illustrate the lemma in Figure 6.

Lemma 1. *Consider two smooth curves, C and C' , crossing the x -axis orthogonally at points $(x, 0)$ and $(x', 0)$ with $x < x'$ and with centers of curvature at the points $(c, 0)$ and $(c', 0)$, respectively. Suppose further that the inequality $c > x$ is satisfied. Then, in order for $|x - x'|$ to be a local minimum for the distance between C and C' it is necessary that c' satisfy the inequality $c \geq c'$.*

Theorem 2. *Let Γ_1 and Γ_2 be two principal curves of the density p . In the case where Γ_1 has endpoints, suppose either that $\partial\Omega$ is strongly convex*

¹We wish to thank Andreas Buja for this observation.

at the endpoints of Γ_1 or that the endpoints of Γ_1 are essential. Then Γ_2 intersects Γ_1 .

Proof. Suppose that Γ_2 does not intersect Γ_1 . Then by compactness, there are points $\mathbf{x}_i \in \Gamma_i$ such that $\text{dist}(\Gamma_1, \Gamma_2) = \text{dist}(\mathbf{x}_1, \mathbf{x}_2)$. In the case where $\partial\Omega$ is strongly convex at the endpoints of Γ_1 , \mathbf{x}_1 and \mathbf{x}_2 must be interior points, and the line segment joining \mathbf{x}_1 and \mathbf{x}_2 intersects both curves orthogonally. In the case where the endpoints are essential, \mathbf{x}_1 and \mathbf{x}_2 are either both endpoints or both interior points, and, again, the line segment joining them intersects both curves orthogonally. The principal curves thus share a common normal line, say L , containing both points. We will use the self-consistency condition, which is satisfied by both curves, to locate their centers of curvature on L . An application of Lemma 1 then leads to a contradiction.

Let $\theta_1 = \theta(\mathbf{x}_1) = \theta(\mathbf{x}_2)$ and let $\ell \subset L$ denote the connected component of the intersection of L with Ω . Recall from Section 2.5 that p induces a density on ℓ which is used to compute transverse moments. Let $\mathbf{x}_0 \in \ell$ be its mean and σ^2 its variance. Parameterize ℓ by $\ell = \{\mathbf{x}_0 + u\mathbf{N}_1 : a < u < b\}$, and set $\mathbf{x}_i = \mathbf{x}_0 + u_i\mathbf{N}_1$.

Writing the self-consistency condition in the form $\kappa_i = -u_i/(u_i^2 + \sigma^2)$ shows that the u -coordinate of the center of curvature of Γ_i is given by the formula

$$c_i = u_i - \frac{u_i^2 + \sigma^2}{u_i} = -\frac{\sigma^2}{u_i}.$$

Consider the possibilities, forced by the condition that the pair $(\mathbf{x}_1, \mathbf{x}_2)$ realize the minimum distance between Γ_1 and Γ_2 (without loss of generality we may assume the inequality $u_1 < u_2$):

- If $0 < u_1$ then the center of Γ_1 is $-\sigma^2/u_1$; and it follows that c_2 must be to the left of c_1 . But the self-consistency condition forces the inequality $c_2 = -\sigma^2/u_2 > -\sigma^2/u_1 = c_1$.
- If $u_1 = 0$ then the inequality $c_2 > 0$ must hold. But self-consistency forces $c_2 < 0$.
- Finally, if the inequality $u_1 < 0$ is satisfied then $c_1 = -\sigma^2/u_1 > 0$ and we must have $c_2 < c_1$. But self-consistency gives $c_2 = -\sigma^2/u_2 > -\sigma^2/u_1 = c_1$

Since all possibilities result in contradiction, Γ_1 and Γ_2 cross. \blacksquare

4 Principal Curves for Uniform Densities

Before continuing with the theoretical development, we study the special case of uniform densities. After a short discussion of uniform densities in general, we consider in detail the uniform densities on a rectangle and an annulus.

Denote by $A(\Omega)$ the area of the support Ω of the uniform density p . Consider a point $\mathbf{x} \in \Omega$ and a direction θ . Let $\bar{\mathbf{x}} = \bar{\mathbf{x}}(\mathbf{x}, \theta)$ and $w = w(\mathbf{x}, \theta)$ be the midpoint and the width, respectively, of the transverse line segment $\ell(\mathbf{x}, \theta) = \{\mathbf{x} + v\mathbf{N}(\theta) \mid v_-(\mathbf{x}, \theta) \leq v \leq v_+(\mathbf{x}, \theta)\}$ (see Definition 5). Because the transverse density on $\ell(\mathbf{x}, \theta)$ is uniform, one calculates the transverse mean and variance to be

$$\hat{v}_\perp(\mathbf{x}, \theta) = \langle \bar{\mathbf{x}} - \mathbf{x}, \mathbf{N}(\theta) \rangle \quad \text{and} \quad \sigma_\perp^2(\mathbf{x}, \theta) = \frac{w^2(\mathbf{x}, \theta)}{12}.$$

An easy calculation then gives for the transverse moments:

$$\mu_0(\mathbf{x}, \theta) = \frac{w}{A(\Omega)}, \quad \mu_1(\mathbf{x}, \theta) = \frac{w\hat{v}_\perp}{A(\Omega)}, \quad \text{and} \quad \mu_2(\mathbf{x}, \theta) = \frac{w^3 + 12w\hat{v}_\perp^2}{12A(\Omega)}.$$

Suppose that Γ was an admissible PC passing through \mathbf{x} with tangential direction θ . According to the differential equation (6), the curvature κ of Γ at \mathbf{x} would then have to be

$$\kappa = \frac{\hat{v}_\perp}{\hat{v}_\perp^2 + \frac{1}{12}w^2}. \quad (8)$$

However, for a PC to be admissible, its centers of curvature have to be disjoint from Ω . The v -coordinate of the center of curvature of Γ at \mathbf{x} is $1/\kappa = \hat{v}_\perp + w^2/(12\hat{v}_\perp)$. The requirement that this number *not* lie between $v_- = \hat{v}_\perp - w/2$ and $v_+ = \hat{v}_\perp + w/2$ implies the inequality (see Figure 7)

$$-\frac{w}{6} < \hat{v}_\perp < \frac{w}{6}.$$

4.1 Principal Curves on the Rectangle

Our first example is the uniform density on the rectangular strip

$$\Omega_{a,b} = \{(x, y) : 0 \leq x \leq b, \quad -a/2 \leq y \leq a/2\}$$

of length b and width a .

For symmetry reasons, the horizontal line $y = 0$ and the vertical line $x = b/2$ are PC's. If the region is square, then the same is true for the two diagonals. We will shortly see, however, that for long strips many other admissible principal curves exist.

Without loss of generality, any admissible PC for the strip can be assumed to be of the form $y = f(x)$, where $f(x)$ is a smooth function which satisfies the transversality conditions $f'(0) = f'(b) = 0$. Curves of a more general type are excluded because they necessarily have ambiguity points in $\Omega_{a,b}$, and curves of the form $x = f(y)$ are dealt with by interchanging the x and y axes. Curves which enter and/or leave are also excluded by the admissibility criterion.

The normal line segment to such a curve has endpoints on the lines $y = \pm a/2$ and midpoint on the x -axis. From this it follows easily that $w = a/\cos \theta$ and $\hat{v}_\perp = y/\cos \theta$, where θ is the angle between the normal \mathbf{N} and the positive x -axis. The differential equation (8) thus becomes

$$\kappa = -\frac{y}{y^2 + a^2/3} \cos \theta.$$

We can explicitly solve this equation, obtaining solutions in terms of elliptic functions. To this end suppose that Γ is a principal curve with initial endpoint $(x, y, \theta) = (0, y_0, 0)$. By symmetry we may as well suppose that y_0 is positive. Integrating the equation

$$\frac{dy}{d\theta} = \frac{dy/ds}{d\theta/ds} = \frac{\sin \theta}{-y \cos(\theta)/(y^2 + a^2/3)} = -\left(\frac{y^2 + a^2/3}{y}\right) \tan \theta$$

by separation of variables results in the equality

$$\cos^2(\theta) = \frac{a^2/3 + y^2}{a^2/3 + y_0^2}.$$

Using the identity $dy/dx = \tan \theta$ to eliminate the variable θ gives rise to the first order differential equation

$$\frac{dy}{dx} = -\sqrt{\frac{y_0^2 - y^2}{a^2/12 + y^2}},$$

where the negative root is taken because the self-consistency conditions force the inequality $y''(x) < 0$ near $x = 0$, but since $y'(0) = 0$ it necessarily follows

that $y'(x)$ is negative for small x . Integration by separation of variables gives

$$x = \int_y^{y_0} \sqrt{\frac{a^2/12 + z^2}{y_0^2 - z^2}} dz = \frac{a}{\sqrt{12}} \int_{y/y_0}^1 \sqrt{\frac{1 + \left(\frac{y_0^2}{a^2/12}\right) q^2}{1 - q^2}} dq.$$

This equation can be rewritten in the form

$$x = \frac{a}{\sqrt{12}} (E(\frac{\pi}{2} \setminus -m) - E(\varphi \setminus -m)), \quad (9)$$

where $\varphi = \sin^{-1}(y/y_0)$, $m = 12y_0^2/a^2$ and $E(\varphi \setminus m) = \int_0^\varphi \sqrt{1 - m \sin^2 \beta} d\beta$ is a standard elliptic integral of the second kind. The period $T = T(a, y_0)$ of the solution is therefore given by the complete elliptic integral of the second kind

$$T = \frac{2a}{\sqrt{3}} E(\frac{\pi}{2} \setminus -m) \quad (10)$$

The transversality condition $\theta = 0$ at $x = b$ imposes an additional condition that the length b must be an integral multiple of half the period. On the other hand, the condition that no ambiguity points of the curve lie in the support of the density places the restriction $|y_0| < a/6$; so we need only consider the periods $T = T(a, y_0)$ for $0 \leq y_0 < a/6$.

In order for principal curves other than the line $y = 0$ to exist, the ratio b/a must be of the form

$$\frac{b}{a} = \frac{n}{\sqrt{3}} E\left(\frac{\pi}{2} \setminus -\frac{12y_0^2}{a^2}\right)$$

for some integer n . Taking into account that fact that y_0 lies between 0 and $a/6$, it follows that for non-linear principal curves to exist, b/a must lie in an interval of the form

$$\begin{aligned} I_n &= \left(\frac{n}{\sqrt{3}} E\left(\frac{\pi}{2} \setminus 0\right), \frac{n}{\sqrt{3}} E\left(\frac{\pi}{2} \setminus -\frac{1}{3}\right) \right) \\ &= \left(\frac{n\pi}{2\sqrt{3}}, \frac{n}{\sqrt{3}} E\left(\frac{\pi}{2} \setminus -\frac{1}{3}\right) \right) \\ &\approx (0.9069 n, 0.978 n). \end{aligned}$$

For sufficiently large n these intervals overlap, and the number of intervals containing a point b/a increases approximately linearly with b/a . Consequently, rectangles with large aspect ratio have a large number of principal curves.

4.2 Principal Curves on the Annulus

Consider next the annulus

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

where (r, ϕ) are polar coordinates. The principal curve equations for a curve Γ in Ω_{R_1, R_2} are best expressed in terms of the angle $\psi = \theta - \phi$ between the radius vector from the origin and the tangent to Γ (see Figure 8). Provided that endpoints of the transverse line segment to Γ lie on both the inner and outer boundary circles of Ω_{R_1, R_2} , its length is given by the formula:

$$w(r, \psi) = \sqrt{R_2^2 - r^2 \cos^2(\psi)} - \sqrt{R_1^2 - r^2 \cos^2(\psi)}.$$

The signed distance from the point (r, ϕ) to the center of the segment is given by the formula

$$\hat{v}_\perp(r, \psi) = r \sin(\psi) - \frac{1}{2} \left(\sqrt{R_2^2 - r^2 \cos^2(\psi)} + \sqrt{R_1^2 - r^2 \cos^2(\psi)} \right).$$

Substitution of these two identities into the curvature formula (8) for uniform densities gives an expression for κ in terms of r and ψ only; and a straightforward calculation shows that the self-consistency equation assumes the form

$$\frac{dr}{d\phi} = \cot(\psi) r, \quad \frac{d\psi}{d\phi} = \kappa(r, \psi) \left(\cos(\psi) \frac{dr}{d\phi} + r \sin(\psi) \right) - 1,$$

where $\kappa = \hat{v}_\perp / (\hat{v}_\perp^2 + w^2/12)$.

Circular Principal Curves. It is not difficult to use the symmetry of the annulus to find the principal curves which are circles centered at the origin. Observe that on such a circle $\psi = \pi/2$; and, consequently, $w = R_2 - R_1$ and $\hat{v}_\perp = r - (R_1 + R_2)/2$. If the circle $r = r_{circ}$ is to be a principal curve its curvature must be the constant

$$\kappa = \frac{r_{circ} - (R_1 + R_2)/2}{(r_{circ} - (R_1 + R_2)/2)^2 + (R_2 - R_1)^2/12}.$$

Equating this with $1/r_{circ}$ and solving to r_{circ} one finds that the radius of the circle is given by the formula

$$r_{circ} = \frac{2(R_1^2 + R_1 R_2 + R_2^2)}{3(R_1 + R_2)} \quad (11)$$

To facilitate comparison between principal curves of the annulus and those of the rectangle, it is useful to express all data in terms of the width $a = R_2 - R_1$ and the mean circumference $b = \pi(R_1 + R_2)$. The ratio b/a is called the *aspect ratio* of the annulus. Since we are free to choose units in which R_2 is equal to 1, properties of the uniform density are determined by the single parameter b/a .

A little algebra shows that the transverse moments of $r = r_{circ}$ are

$$\mu_0 = \frac{1}{b}, \quad \mu_1 = \frac{\pi a^2}{6b^2}, \quad \mu_2 = \frac{a^2}{12b} + \frac{\pi^2 a^4}{36b^3}, \quad (12)$$

and that the transverse mean and variance are

$$\hat{v}_\perp = \frac{\pi a^2}{6b} = \frac{(R_2 - R_1)^2}{6(R_1 + R_2)} \text{ and } \sigma_\perp^2 = \frac{a^2}{12} = \frac{(R_2 - R_1)^2}{12}, \quad (13)$$

respectively.

Principal Curves of the form $r = f(\phi)$. It is easily seen that *all* admissible curves for the annular distribution can be expressed in the form $r = f(\phi)$. After a possible rotation, we may assume that f attains a local minimum at $\phi = 0$. The admissibility condition requires that no center of curvature of Γ lie in the annulus. A simple calculation at a local minimum and a local maximum of f shows that $f(\phi)$ must lie between the values R_{min} and R_{max} , where

$$R_{min} = \frac{2R_1 + R_2}{3} \text{ and } R_{max} = \frac{R_1 + 2R_2}{3}.$$

Moreover, since all admissible PC's cross (Theorem 2), the minimum of f can be no larger than r_{circ} . Hence, we can find all admissible principal curves on the annulus by solving the initial value problem

$$\begin{aligned} \frac{dr}{d\phi} &= \cot(\psi) r, & \frac{d\psi}{d\phi} &= \kappa(r, \psi) \left(\cos(\psi) \frac{dr}{d\phi} + r \sin(\psi) \right) - 1, \\ r(0) &= r_0, & \psi(0) &= \frac{\pi}{2}, \end{aligned}$$

where $R_{min} < r_0 \leq r_{circ}$.

Because we did not succeed in finding an analytic solution of this problem, we used a variable-step 4th-order Runge-Kutta method (Runge-Kutta-Fehlberg) to compute self-consistent curves for a variety of aspect ratios and

initial values, $f(0)$. Our results indicate that admissible PC's on the annulus have much in common with PC's on the rectangle:

Let $T(R_1, R_2, r_0)$ denote the period of the solution $r = f(\phi)$ of the principal curve equation, satisfying the initial conditions $r(0) = r_0$, $r'(0) = 0$, with $r_{circ} \geq r_0 < R_{max}$. In order for $r = f(\phi)$ to define a principal curve, it must be closed. This forces the condition $T(R_1, R_2, r_0) = 2\pi/n$, $n = 1, 2, 3, \dots$. This condition, in turn, forces r_0 to assume only a discrete set of values (see Figure 9). Numeric calculations indicate that $T(R_1, R_2, r_0)$ assumes a unique minimum at $r_0 = r_{circ}$ and decreases with increasing aspect ratio (see Figure 10).

Our numerical experiments can be summarized as follows:

- annular densities appear to support only a finite number (up to rotations about the origin) of principal loops;
- the number of principal loops of the form $r = f(\phi)$ increases with increasing aspect ratio;
- the ϕ -period of principal loops decreases with increasing aspect ratio.

Consider, for example, the annulus $\Omega_{0.45,1}$. In addition to the circular principal loop $r = r_{circ} \approx 0.760$, there is one other principal loop, given by the initial conditions $r(0) \approx 0.646$, $r'(0) = 0$. This principle loop has period $T = \pi/4$ and oscillates between the values $r = 0.646$ and $r = 0.874$. Initial conditions of the form $r'(0) = 0$ and $r(0) < 0.646$ give curves whose periods are slightly greater than $\pi/2$ and so cannot form closed curves; for $r(0) > 0.646$, the period is slightly smaller than $\pi/2$ and again the curve cannot close (see Figure 9).

5 Extremal Properties of Principal Curves

Let Γ be an admissible curve of the density p . Then the expected squared distance between an observation and Γ is given by

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

We wish to consider I as a functional on the space of all admissible curves:

$$I : \mathcal{E}_{Adm, \Omega} \rightarrow \mathbf{R}.$$

Hastie and Stuetzle (1989) show that principal curves are critical points of I . This means that if Γ is a principal curve then

$$\left. \frac{dI[\Gamma_t]}{dt} \right|_{t=0} = 0$$

for any smooth one-parameter family Γ_t of admissible curves with $\Gamma_0 = \Gamma$. (Such a family is called a *variation* of Γ .)

We will now compute the second derivative

$$\left. \frac{d^2I[\Gamma_t]}{dt^2} \right|_{t=0}.$$

If it were positive for all variations of a PC Γ , then Γ would be a local minimum for the functional I . Unfortunately there are no PC's with this property, as we now show.

Represent the family Γ_t by a function $\Phi : \Gamma \times (-\epsilon, +\epsilon) \rightarrow \mathbf{R}^2$. Differentiating Φ with respect to t at $t = 0$ gives a vector field defined along Γ , of the form $\mathbf{X} : \Gamma \rightarrow \mathbf{R}^2$. We show in the appendix that Φ can be chosen so that \mathbf{X} is normal to Γ . Thus, \mathbf{X} is of the form $\mathbf{X} = f\mathbf{N}$ where f is a real valued function on Γ . Such a vector field is called the *variational vector field* generated by Γ_t . We call f an *infinitesimal variation* of Γ . It can be shown that \mathbf{X} depends only on the family Γ_t (i.e. it does not depend on the choice of Φ).

Theorem 3. *Let Γ be an admissible principal curve of p and let Γ_t be a variation of Γ . Then*

$$\left. \frac{d^2I[\Gamma_t]}{dt^2} \right|_{t=0} = 2 \int_{\Gamma} \{ (\mu_0 - \mu_1 \kappa) f^2 - \mu_2 f'^2 \} ds,$$

where f is the infinitesimal variation generated by Γ_t .

The proof of Theorem 3 will be given in Appendix A.

Remark 4. The formula for second derivative given in Theorem 3 can be written in the form

$$\left. \frac{\partial^2 I[\Gamma_t]}{\partial t^2} \right|_{t=0} = 2 \int_{\Gamma} \left\{ \left(\frac{\sigma_{\perp}^2}{\hat{v}_{\perp}^2 + \sigma_{\perp}^2} \right) f^2 - (\hat{v}_{\perp}^2 + \sigma_{\perp}^2) f'^2 \right\} \mu_0 ds$$

Here we have made use of the identity $\kappa = \mu_1 / \mu_2$ and equation (4).

An immediate consequence of this formula is that the term $(\mu_0 - \mu_1 \kappa)$ is strictly positive (except possibly at the endpoints of Γ).

From Theorem 3 it is easy to see why admissible principal curves can *never* be local minima of I . It suffices to choose a function f whose L^2 -norm relative to the weighted measure $(\mu_0 - \mu_1 \kappa) ds$ is small, but whose derivative f' has large L^2 norm relative to the weighted measure $\mu_2 ds$. One then shows that f can be chosen to be the infinitesimal variation generated by a variation Γ_t . Setting $I(t) = I[\Gamma_t]$, we have $I'(0) = 0$ and $I''(0) < 0$, from which it follows that Γ is not a local minimum of I .

It is worthwhile to consider two simple examples in which a complete calculation of the second derivative of I can be given.

Example 3 (The Rectangle). Recall that the line segment $\Gamma = \{(x, 0) : 0 \leq x \leq b\}$ is a principal curve for the uniform density on the rectangle $\Omega_{a,b} = \{(x, y) : 0 \leq x \leq b, -a/2 \leq y \leq a/2\}$. Consider a variation of Γ of the form

$$\Gamma_t = \{(x, tf(x)) : 0 \leq x \leq b\},$$

where $f(x)$ is a smooth function on $[0, b]$ with $f'(0) = f'(b) = 0$. In this case, $\mu_0 = 1/b$, $\mu_1 = 0$, and $\mu_2 = a^2/12b$; hence, setting $I(t) = I[\Gamma_t]$, we have

$$I''(0) = \frac{1}{b} \int_0^b f(s)^2 ds - \frac{a^2}{12} \int_0^b f'(s)^2 ds.$$

Substitution of the Fourier expansion

$$f(x) = \sum_{n=0}^{\infty} c_n \cos\left(\frac{n\pi}{b} x\right)$$

into this formula gives

$$I''(0) = \sum_{n=0}^{\infty} \left(1 - \frac{n^2 a^2 \pi^2}{12b^2}\right) c_n^2 = \frac{1}{b} \sum_{n=0}^{\infty} \left(1 - \left(\frac{2\pi\sigma_{\perp}}{T_n}\right)^2\right) c_n^2, \quad (14)$$

where $\sigma_{\perp} = a/\sqrt{12}$ is the transverse standard deviation and $T_n = 2b/n$ is the period of $\cos(n\pi x/b)$.

It is easy to find variations for which $I''(0)$ is negative: just choose $f(x) = \cos(n\pi x/b)$, with n large. Thus, the line $y = 0$ is not a local minimum of I . Notice, though, that if the Fourier expansion of f only has terms of period longer than $2\pi\sigma_{\perp}$, then the second derivative $I''(0)$ is positive. In more colorful language, *the principal curve Γ is a local minimum of I within the class of perturbations of Γ of period longer than $2\pi\sigma_{\perp}$*

Example 4 (The Annulus). Next consider the circular principal curve $r = r_{circ}$ on the annulus $\Omega_{R_1, R_2} = \{(r, \phi) : R_1 \leq r \leq R_2\}$. Let $f(s)$ be the infinitesimal variation generated by the variation Γ_t . (In this case, \mathbf{N} is the unit vector field pointing towards the origin.) As before, set $I(t) = \mathbb{I}[\Gamma_t]$. We again wish to compute $I''(0)$ by Fourier expanding $f(s)$.

Since the circumference of Γ is $L_\Gamma = 2\pi r_{circ}$, we can expand as follows:

$$f(s) = a_0 + \sum_{n=1}^{\infty} a_n \cos\left(2n\pi \frac{s}{L_\Gamma}\right) + b_n \sin\left(2n\pi \frac{s}{L_\Gamma}\right).$$

Setting $I(t) = \mathbb{I}[\Gamma_t]$, we get

$$I''(0) = 2 \int_{\Gamma} (\mu_0 - \mu_1 \kappa) f^2 - \mu_2 (f')^2 ds.$$

Expanding f in Fourier series and using the fact that the moments μ_i are all constant along Γ gives

$$I''(0) = L_\Gamma(\mu_0 - \mu_1 \kappa) a_0^2 + \sum_{n=1}^{\infty} \left(L_\Gamma(\mu_0 - \mu_1 \kappa) - 4n^2 \pi^2 \frac{\mu_2}{L_\Gamma} \right) (a_n^2 + b_n^2) \quad (15)$$

Using formulas (12), one finds that along Γ

$$L_\Gamma(\mu_0 - \mu_1 \kappa) = 1 \quad \text{and} \quad 4n^2 \pi^2 \frac{\mu_2}{L_\Gamma} = \left(\frac{2n\pi \sigma_\perp}{b} \right)^2,$$

where $b = (R_1 + R_2)/2$. One can also show that $b = L_\Gamma + (2\pi \sigma_\perp)^2/b$. Substituting these formulas into equation (15) yields the identity

$$I''(0) = a_0^2 + \sum_{n=1}^{\infty} \left(1 - \left(\frac{2\pi \sigma_\perp}{T_n + \frac{4\pi \sigma_\perp^2}{nb}} \right)^2 \right) (a_n^2 + b_n^2), \quad (16)$$

where $T_n = L_\Gamma/n$ is the period of $\cos(2\pi s/L_\Gamma)$ and $\sin(2\pi s/L_\Gamma)$.

Notice that the equations (14) and (16) are quite similar. Again we see that, if the Fourier expansion of f only has terms of period longer than $T_n > 2\pi \sigma_\perp$, then the second derivative $I''(0)$ is positive. Thus, the curve Γ is a local minimum of \mathbb{I} , provided that we consider only variations of Γ of period longer than $2\pi \sigma_\perp$.

In the two previous examples, the transverse moments μ_k , $k = 0, 1, 2$ were constant along the principal curve Γ . For this reason, we were able to Fourier expand the infinitesimal variation f and compute the second derivative $I''(0)$ in terms of this expansion.

It turns out that similar computations can be carried out if the transverse moments are only approximately constant. When this is the case, the transverse moments are said to be *slowly varying*. In this case, $I''(0)$ can be computed in terms of a certain (generalized) Fourier expansion of the infinitesimal variation f with respect to a basis $\{f_n\}_{n=1}^{\infty}$ of functions on Γ which is naturally associated with the transverse moments of p . The details of how this basis is constructed are discussed in Appendix B.

There are three important features of this basis:

1. 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 non-decreasing sequence of positive real numbers.

2. The functions f_n are approximately periodic functions whose periods decrease with increasing n .
3. $\lambda_n > 1$ when the period of f_n is smaller than $2\pi\sigma_{\perp}(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2)$ and $\lambda_n < 1$ when the period of f_{n+1} is greater than $2\pi\sigma_{\perp}(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2)$.

Thus, $I''(0) > 0$ if the Fourier expansion of f only contains terms of long period relative to $2\pi\sigma_{\perp}(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2)$. Within the class of such variations, Γ is thus a local minimum of I . We formulate this as a theorem, whose proof we defer to Appendix B:

Theorem 4. *Suppose that Γ is a principal curve of a density along which the transverse moments μ_0 , μ_1 and μ_2 are slowly varying, and suppose that Γ is either a principal loop or has essential endpoints. Let Γ_t be a variation of Γ whose infinitesimal variation f has a Fourier expansion $f = \sum_n a_n f_n$ with $a_n = 0$ for period of $f_n < 2\pi\sigma_{\perp}(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2)$. Then $I[\Gamma]$ is a local minimum of $I[\Gamma_t]$.*

6 Conclusions and Open Problems

Principal curves were invented to formalize the concept of “a curve passing through the middle of a data set”. In view of the results presented in this paper, however, it appears that the current definition is only a qualified success.

The fact that even simple densities can have many PC’s is unfortunate, but hardly surprising — after all, linear principal components are not unique either.

More serious is the fact that PC’s are never local minima of the distance functional. This is where nonparametric manifold estimation differs fundamentally from nonparametric regression. In the regression context, the conditional expectation minimizes distance (i.e. expected squared prediction error). This justifies choosing the complexity of the model to minimize an estimate of distance, such as the cross-validated residual sum of squares. Principal curves, on the other hand, do not minimize distance. Thus there is no justification for using cross-validation for model selection, and indeed cross-validation has been observed to fail in practice.

PC’s are local minima of the distance functional if we restrict ourselves to “low frequency” variations, but the definition of “low frequency” in turn depends on the PC. Therefore it is not clear how this extremal property could be used in estimating PC’s.

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, and this remains an important open problem.

A Proof of Theorem 3

Suppose that Γ is an admissible curve and that $\Gamma_t \in \mathcal{E}_{Adm, \Omega}$, $-a < t < a$ is a smooth family of admissible curves with $\Gamma_0 = \Gamma$. Such a family is called a *variation* of Γ . To prove Theorem 3, we need only compute the second derivative of $I(t) = \mathbb{I}[\Gamma_t]$ at $t = 0$. Our computation relies on a careful analysis of the dependence of the projection map $\pi_{\Gamma_t} : \Omega \rightarrow \Gamma_t$ on t .

A.1 Variational Vector Fields

We begin with a short exposition of variational vector fields. The material here is fairly standard and can be found in a number of places (see, for example, Griffiths (1983)).

The family $\{\Gamma_t\}$ can be expressed as a map $\Phi : \Gamma \times (-a, a) \rightarrow \Omega$, with $\Phi(\Gamma, t) = \Gamma_t$. The map Φ is uniquely determined by the condition

$$\left\langle \frac{\partial \Phi(s, t)}{\partial t}, \frac{\partial \Phi(s, t)}{\partial s} \right\rangle = 0, \quad (17)$$

where Φ is expressed in terms of the arclength parameter s of Γ .² It will be convenient to adopt the notation $\mathbf{x}_t(s) = \Phi(s, t)$.

To prove that equation (17) uniquely determines Φ , consider the smooth surface in $\mathbf{R}^2 \times \mathbf{R}$ of the form

$$\Sigma = \{(\mathbf{x}, t) : \mathbf{x} \in \Gamma_t\} \subset \Omega \times \mathbf{R}.$$

It is not difficult to see that there is a unique vector field on Σ , of the form

$$\frac{\partial}{\partial t} + f_t(\mathbf{x}) \mathbf{N}_t(\mathbf{x}),$$

and characterized by the condition that it be tangent to Σ . (Subscripted quantities refer to the curve Γ_t .) Set $\mathbf{X}_t = f_t(\mathbf{x}) \mathbf{N}_t(\mathbf{x})$ and let $\mathbf{X} = \mathbf{X}_0$. The vector field \mathbf{X} is called an *infinitesimal variation* of Γ . For t sufficiently small, integration of $\partial/\partial t + \mathbf{X}_t$ gives a smooth 1-parameter family of diffeomorphisms $\Phi_t : \Gamma \rightarrow \Gamma_t$, which can be assembled into a single map Φ . By construction, the equation

$$\frac{\partial \Phi(s, t)}{\partial t} = f_t(s) \mathbf{N}_t(s)$$

²Note that for $t \neq 0$, s is *not* an arc length parameter on Γ_t .

holds for all s and t . In particular, Φ satisfies the identity (17).

In the case where Γ has endpoints, so do each of the curves Γ_t , and the transversality conditions—i.e. the requirement that Γ_t intersect the boundary $\partial\Omega$ orthogonally for all t —impose certain boundary conditions on f , called the *infinitesimal transversality conditions*, specified in next lemma.

Lemma 2. *Let Γ be an admissible arc of Ω with initial and terminal endpoints \mathbf{x}_- and \mathbf{x}_+ , respectively. Let $\kappa_{\partial\Omega}$ denote the curvature of $\partial\Omega$, oriented so that its normal points into Ω . Then every infinitesimal variation $\mathbf{X} = f\mathbf{N}$ of Γ satisfies the pair of boundary conditions*

$$f' = \pm\kappa_{\partial\Omega} f \quad \text{at } \mathbf{x}_{\pm}.$$

Proof. It suffices to consider only the initial endpoint \mathbf{x}_- (the computation at \mathbf{x}_+ is identical, except for sign). Let Γ_t and X_t be as above and set $\mathbf{x} = \Phi(s, t)$. Differentiate the identity $\left\langle \mathbf{N}_t(0), \frac{\partial\mathbf{x}}{\partial t} \right\rangle = 0$ with respect to t at the endpoint $\Phi(0, 0) = \mathbf{x}_-$:

$$\begin{aligned} \frac{\partial}{\partial t} \left\langle \mathbf{N}_t, \frac{\partial\mathbf{x}}{\partial s} \right\rangle &= \left\langle \frac{\partial\mathbf{N}}{\partial t}, \frac{\partial\mathbf{x}}{\partial s} \right\rangle + \left\langle \mathbf{N}, \frac{\partial^2\mathbf{x}}{\partial t\partial s} \right\rangle \\ &= \kappa_{\partial\Omega} f \langle \mathbf{T}, \mathbf{T} \rangle + \langle \mathbf{N}, (f'\mathbf{N} - kf\mathbf{T}) \rangle \quad (\text{at } t = 0). \end{aligned}$$

Thus $\kappa_{\partial\Omega} f + f' = 0$. \blacksquare

Remark 5. It can be shown that any function f satisfying the conditions of Lemma 2 arises from a variation of Γ through admissible curves. For this reason we denote the set of all variational vector fields along Γ by the symbol $T_{\Gamma}\mathcal{E}_{Adm,\Omega}$, and we will at times refer to the *tangent space*³ of $\mathcal{E}_{Adm,\Omega}$ at Γ . Because the infinitesimal transversality conditions are linear, the space $T_{\Gamma}\mathcal{E}_{Adm,\Omega}$ is a vector space.

³For those conversant with the theory of infinite dimensional manifolds, we remark that $\mathcal{E}_{Adm,\Omega}$ can be given the structure of a Fréchet manifold modeled on the Fréchet space of C^{∞} -functions on the unit interval with either vanishing derivative at the endpoints or periodic boundary conditions.

A.2 Smoothness of the Projection π_{Γ_t}

The projection index $\lambda = \lambda(\mathbf{x}, t)$ is defined by the formula

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

where $\pi_{\Gamma_t} : \Omega \rightarrow \Gamma_t \subset \Omega$ is the projection map. We require a formula for $\partial\lambda(\mathbf{x}, t)/\partial t$ at $t = 0$. But to justify such a formula, we need to know that λ is smooth for t sufficiently near 0:

Lemma 3 . *There exists a number $0 < a' \leq a$ such that λ is smooth at all points of $\Omega \times (-a', a')$.*

Proof. Let $\mathbf{x}_0 \in \Omega$ and let $s_0 = \lambda(\mathbf{x}_0, 0)$. Since $\Phi(s_0, 0)$ is the unique point of Γ_0 nearest to \mathbf{x}_0 , the line joining \mathbf{x}_0 to $\Phi(s_0, 0)$ intersects Γ orthogonally at $\Phi(s_0, 0)$; i.e.

$$\left\langle \mathbf{x}_0 - \Phi(\lambda(\mathbf{x}_0, 0), 0), \frac{\partial\Phi(\lambda(\mathbf{x}_0, 0), 0)}{\partial s} \right\rangle = 0.$$

We wish to apply the implicit function theorem to prove that there are real numbers $\epsilon > 0$ and $\delta' > 0$ such that the following condition is satisfied:

For all $(\mathbf{x}, t) \in \Omega \times (-a, a)$ with $\|\mathbf{x} - \mathbf{x}_0\| < \delta'$, $|t| < \delta$, there is a unique solution $s = S(\mathbf{x}, t)$, $|s - s_0| < \epsilon$, of the equation

$$\left\langle \mathbf{x} - \Phi(s, t), \frac{\partial\Phi(s, t)}{\partial s} \right\rangle = 0. \quad (18)$$

Moreover, $S(\mathbf{x}, t)$ depends smoothly on (\mathbf{x}, t) .

We only need to verify that the derivative of the left hand side of (18) with respect to s is non-zero for $(\mathbf{x}, t, s) = (\mathbf{x}_0, 0, s_0)$. But

$$\begin{aligned} \frac{\partial}{\partial s} \left\langle \mathbf{x}_0 - \Phi(s, 0), \frac{\partial\Phi(s, 0)}{\partial s} \right\rangle &= - \left\langle \frac{\partial\Phi(s, 0)}{\partial s}, \frac{\partial\Phi(s_0, 0)}{\partial s} \right\rangle + \\ &\quad \left\langle \mathbf{x}_0 - \Phi(s, 0), \frac{\partial^2\Phi(s, 0)}{\partial s^2} \right\rangle \\ &= -1 + \langle \mathbf{x}_0 - \Phi(s, 0), \kappa(s)\mathbf{N}(s) \rangle ; \end{aligned}$$

hence, setting $s = s_0$ gives

$$\frac{\partial}{\partial s} \left\langle \mathbf{x}_0 - \Phi(s, 0), \frac{\partial \Phi(s_0, 0)}{\partial s} \right\rangle = -1 + \langle \mathbf{x}_0 - \pi_\Gamma(\mathbf{x}), \kappa(s_0) \mathbf{N}(s_0) \rangle < 1.$$

On the other hand, the fact that Ω contains no ambiguity points of Γ , together with continuity of Φ and compactness of Γ , implies the existence of a real number $\delta'' > 0$ such the inequality

$$\|\mathbf{x} - \Phi(s, t)\| \geq \|\mathbf{x} - \Phi(s_0, 0)\|$$

is satisfied for all $(\mathbf{x}, t) \in \Omega \times (-a, a)$ with $\|\mathbf{x} - \mathbf{x}_0\| < \delta''$, $|t| < \delta'$ and s with $|s - s_0| \geq \epsilon$.

Consequently, setting $\delta = \min(\delta', \delta'')$, we find that for all for $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ and $|t| < \delta$, the inequality $|\lambda(\mathbf{x}, t) - s_0| < \epsilon$ holds. Thus, $\lambda(\mathbf{x}, t) = S(\mathbf{x}, t)$, showing that λ is smoothly dependent on (\mathbf{x}, t) in a neighborhood of $(\mathbf{x}_0, 0)$. The result now follows from compactness of Ω . ■

Lemma 4 .

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

Proof. Differentiate the identity $\left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial \Phi(\lambda(\mathbf{x}, t), t)}{\partial s} \right\rangle = 0$ with respect to t at $t = 0$ and expand as follows:

$$\begin{aligned} \frac{\partial}{\partial t} & \left(\left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial \Phi(\lambda(\mathbf{x}, t), t)}{\partial s} \right\rangle \right) \Big|_{t=0} \\ &= - \left\langle \left(\frac{\partial \Phi(\lambda(\mathbf{x}, 0), 0)}{\partial s} \frac{\partial \lambda(\mathbf{x}, 0)}{\partial t} + \frac{\partial \Phi(\lambda(\mathbf{x}, 0), 0)}{\partial t} \right), \frac{\partial \Phi(\lambda(\mathbf{x}, 0), 0)}{\partial s} \right\rangle \\ & \quad + \left\langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \left(\frac{\partial^2 \Phi(\lambda(\mathbf{x}, 0), 0)}{\partial s^2} \frac{\partial \lambda(\mathbf{x}, 0)}{\partial t} + \frac{\partial^2 \Phi(\lambda(\mathbf{x}, 0), 0)}{\partial s \partial t} \right) \right\rangle \\ &= - \frac{\partial \lambda(\mathbf{x}, 0)}{\partial t} + \langle \mathbf{x} - \pi_\Gamma(\mathbf{x}), \mathbf{N}(\lambda(\mathbf{x}, 0)) \rangle \left(\kappa(\lambda(\mathbf{x}, 0)) \frac{\partial \lambda(\mathbf{x}, 0)}{\partial t} + f'(\lambda(\mathbf{x}, 0)) \right) \end{aligned}$$

Solving for $\partial \lambda / \partial t$ gives the identity we seek. ■

A.3 Computation of the Second Derivative

Set $I(t) = \mathbb{I}[\Gamma_t]$. Then

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

Note that, for $\mathbf{x} \in \Omega$, the vector $\partial\Phi(\lambda(\mathbf{x}, t), t)/\partial s$ is tangent to Γ_t at $\mathbf{x} = \Phi(\lambda(\mathbf{x}, t), t)$. Since $\Phi(\lambda(\mathbf{x}, t), t)$ is the point on Γ_t nearest \mathbf{x} and since Γ_t is admissible, it follows that $\mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t)$ is orthogonal to $\partial\Phi(\lambda(\mathbf{x}, t), t)/\partial s$. For this reason, the formula for $I'(t)$ reduces to the equation

$$I'(t) = -2 \int \int_{\Omega} \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x}) d\mathbf{x}. \quad (19)$$

By equation (19),

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

Differentiation inside the integral and an application of the chain rule gives:

$$\begin{aligned} I''(t) &= 2 \int \int_{\Omega} - \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial^2\Phi(\lambda(\mathbf{x}, t), t)}{\partial s \partial t} \frac{\partial\lambda(\mathbf{x}, t)}{\partial t_2} \right\rangle \\ &\quad - \left\langle \mathbf{x} - \Phi(\lambda(\mathbf{x}, t), t), \frac{\partial^2\Phi(\lambda(\mathbf{x}, t), t)}{\partial t^2} \right\rangle \\ &\quad + \left\langle \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial s} \frac{\partial\lambda(\mathbf{x}, t)}{\partial t}, \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle \\ &\quad + \left\langle \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t}, \frac{\partial\Phi(\lambda(\mathbf{x}, t), t)}{\partial t} \right\rangle p(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Setting $t = 0$ and noting that $\mathbf{T} = \partial\Phi/\partial s$ and $X = f\mathbf{N} = \partial\Phi/\partial t$, yields the formula

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

Next employ the Frenet formulas $\mathbf{T}' = \kappa\mathbf{N}$ and $\mathbf{N}' = -\kappa\mathbf{T}$ and recall that $\mathbf{x} - \pi_{\Gamma}(\mathbf{x})$ is orthogonal to \mathbf{T} to obtain the formula

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

where $h(s)$ is the function on Γ defined by the formula

$$h(s) = \left\langle \mathbf{N}(s), \frac{\partial^2 \Phi(s, 0)}{\partial t^2} \right\rangle.$$

By Lemma 4 the equality

$$I''(0) = \int \int_{\Omega} \left\{ f(\pi_{\Gamma}(\mathbf{x}))g(\pi_{\Gamma}(\mathbf{x})) - \frac{\langle \mathbf{x} - \pi_{\Gamma}(\mathbf{x}), \mathbf{N}(\pi_{\Gamma}(\mathbf{x})) \rangle^2 f'(\pi_{\Gamma}(\mathbf{x}))g'(\pi_{\Gamma}(\mathbf{x}))}{(1 - \langle \mathbf{x} - \pi_{\Gamma}(\mathbf{x}), \mathbf{N}(\pi_{\Gamma}(\mathbf{x})) \rangle \kappa(\lambda(\mathbf{x}, 0)))} - \langle (\mathbf{x} - \pi_{\Gamma}(\mathbf{x})), h(\pi_{\Gamma}(\mathbf{x}))\mathbf{N}(\pi_{\Gamma}(\mathbf{x})) \rangle \right\} p(\mathbf{x}) d\mathbf{x}$$

is satisfied.

Reverting to normal coordinates and carrying out the integration with respect to v gives the identity

$$I''(0) = \int_{\Gamma} \left\{ (\mu_0 - \kappa\mu_1)f^2 - \mu_2 f'^2 \right\} ds - \int_{\Gamma} h(\mu_1 - \kappa\mu_2) ds.$$

Since Γ is a principal curve, the last integral vanishes, yielding the desired formula.

B Proof of Theorem 4

We present here the technical details needed to justify the discussion in Section 6. We assume the reader is somewhat familiar with the theory of Sturm-Liouville problems as presented, for example, in Courant and Hilbert (1953).

B.1 The Sturm-Liouville Problem

Set $P(s) = \mu_2(s)$ and $Q(s) = \mu_0(s) - \mu_1^2(s)/\mu_2(s)$, and consider the Sturm-Liouville problem consisting of the differential equation

$$(P(s)f'(s))' + \lambda Q(s)f(s) = 0, \quad 0 \leq s \leq L_{\Gamma} \quad (20)$$

together with one of the boundary conditions:

$$f(0) = f(L_{\Gamma}) \text{ and } f'(0) = f'(L_{\Gamma}) \text{ (when } \Gamma \text{ is a PL)}$$

or

$f'(0) = 0$ and $f'(L_{\Gamma}) = 0$ (when Γ is a principal arc with essential endpoints).

(When Γ is a principal loop, periodic boundary conditions are required in order for f to be a C^1 -function on Γ , and when Γ is a principal arc the condition that admissible curves intersect the boundary of Ω orthogonally imposes the boundary condition $f'(0) = f'(L_\Gamma) = 0$ (see Lemma 2).)

Notice that in terms of P and Q , the second derivative $I''(0)$ assumes the form

$$I''(0) = \int_0^{L_\Gamma} \{Q(s)f(s)^2 - P(s)f'(s)^2\} ds. \quad (21)$$

The fact that Γ is an admissible PC implies that both $P(s)$ and $Q(s)$ are strictly positive, except possibly at the end points of Γ (see Remark 4).

If P and Q are strictly positive everywhere (endpoints included), then the Sturm-Liouville problem (20) is non-singular and the general theory of such problems applies. This is the case if Γ is a principal loop or if the endpoints of Γ are essential (see Definition 4). From now on, we only consider principal loops and principal arcs with essential endpoints.

B.2 The Natural Basis

The formula

$$\langle f, g \rangle = \int_0^{L_\Gamma} Q(s) f(s) g(s) ds$$

defines an inner product on the space of C^1 -functions on $[0, L_\Gamma]$ satisfying the boundary conditions. We identify this space with the space of infinitesimal variations of Γ . Let \mathcal{H}_Γ be its L^2 -completion with respect to the inner product. Sturm-Liouville theory gives a complete, orthonormal basis for \mathcal{H}_Γ , consisting of smooth functions f_n , $n = 1, 2, 3, \dots$. More specifically, there is a discrete set of positive numbers $\lambda_1 \leq \lambda_2 \leq \dots$, such that

$$(Pf'_n)' + \lambda_n Q f_n = 0 \quad \text{and} \quad \int_0^{L_\Gamma} f_n(s) f_m(s) Q(s) ds = \delta_{n,m}.$$

Now suppose that Γ_t is a variation of Γ with variational vector field $f\mathbf{N}$. Let $f = \sum_{n=1}^{\infty} a_n f_n$ be the Fourier expansion of f . Substituting into formula (21) for $I''(0)$ gives

$$\left. \frac{d^2 I}{dt^2} \right|_{t=0} = 2 \int_0^{L_\Gamma} Q f^2 - P f'^2 ds = \sum_{n=1}^{\infty} a_n^2 (1 - \lambda_n). \quad (22)$$

The second derivative $I''(0)$ is positive if $a_n = 0$ for all n such that $\lambda_n \geq 1$. Thus, $I''(0)$ is positive if f lies in the finite dimensional space

$$\mathcal{H}_+ = \text{Span}\{f_n : \lambda_n < 1\} \subset \mathcal{H}. \quad (23)$$

This result can be reformulated as a theorem.

Theorem 5. *Let Γ be a principal loop or a principal arc with essential endpoints and suppose that Γ_t is a variation of Γ whose variational vector field is of the form $\mathbf{X} = f\mathbf{N}$ with $f \in \mathcal{H}_+$. Then there is a number $\epsilon > 0$ such that $I[\Gamma] < I[\Gamma_t]$ for all $0 < |t| < \epsilon$.*

B.3 Eigenvalue Estimates

When the length of Γ is large relative to the transverse standard deviation σ_\perp and the transverse moments $\mu_k(s)$, $k = 0, 1, 2$ are slowly varying, i.e. have small first and second derivatives with respect to s , the situation is much like that of Examples 3 and 4 above.

This is shown by obtaining estimates of the eigenvalues λ_n in terms of the transverse moments of p along Γ . The calculations in the case where Γ is a principal loop are quite similar to those in which Γ is a principal arc. Therefore, we present only the case where Γ is a principal arc and leave to the reader the (easier) case where Γ is a principal loop.

The argument proceeds by applying the asymptotic estimates given in Courant and Hilbert (1953) to the Sturm-Liouville problem (20). It will prove useful to work with the new variables

$$z(t) = \sqrt{\frac{\pi}{L_\Gamma}} \sqrt[4]{P(s)Q(s)} f(s), \text{ and } t = T(s) = \int_0^s \sqrt{\frac{Q(q)}{P(q)}} dq$$

(see page 292 of Courant and Hilbert (1953)). Using the identities (4) and (5), the change of variables can be rewritten in a form which is more useful for our purposes:

$$z(t) = \sqrt{\frac{\pi}{L_\Gamma}} \sqrt{\mu_0(s) \sigma_\perp(s)} f(s) \text{ where } t = \int_0^s \frac{\sigma_\perp(q)}{\hat{v}_\perp^2(q) + \sigma_\perp^2(q)} dq. \quad (24)$$

With respect to these new variables the Sturm-Liouville problem (20) assumes the form

$$\frac{d^2 z}{dt^2} - r \frac{dz}{dt} + \lambda z = 0, \quad 0 \leq t \leq \ell = T(L_\Gamma),$$

where

$$r = r(t) = \frac{1}{\sqrt{\mu_0(s) \sigma_\perp(s)}} \frac{d^2}{dt^2} \left(\sqrt{\mu_0(s) \sigma_\perp(s)} \right), \quad s = T^{-1}(t).$$

In these coordinates the boundary conditions assume the form

$$z'(0) - a(0)z(0) = 0 \text{ and } z'(\ell) - a(\ell)z(\ell) = 0, \quad (25)$$

where

$$a(t) = \frac{1}{2\mu_0(s)} \frac{d}{ds} (\hat{v}_\perp(s) \mu_0(s)) \left(1 + \frac{\sigma_\perp^2(s)}{\hat{v}_\perp^2(s)} \right), \quad s = T^{-1}(t).$$

Assumptions: At this point, we must make some *a priori* assumptions about the moments and the length of Γ :

A1 $|r(t)| < r_M$ for all $0 \leq t \leq \ell$, where r_M is a small constant such that $0 < r_M < \ell^{-2}$.

A2 $|a(0)| < b$ and $|a(\ell)| < b$ where $b < \pi/2\ell$ is a small constant.

A3 $L_\Gamma \gg \max_{0 \leq s \leq L_\Gamma} 2\pi \sigma_\perp(s) (1 + \hat{v}_\perp^2(s)/\sigma_\perp^2(s))$

Conditions (A1) and (A2) hold in the case where the moments $\mu_k(s)$, $k = 0, 1, 2$ have sufficiently small first and second derivatives with respect to s . Condition (A3) is the condition that L_Γ be very long relative to the transverse standard deviation σ_\perp . Since

$$\ell = \int_0^{L_\Gamma} \frac{ds}{\sigma_\perp (1 + \hat{v}_\perp^2(s)/\sigma_\perp^2(s))},$$

condition (A3) implies the inequality $\ell \gg 2\pi$.

Notice that all three of these assumptions will be satisfied if the first and second derivatives of the moments $\mu_k(s)$, $k = 0, 1, 2$ are sufficiently small. When this is the case, the transverse moments of p along Γ are said to be *slowly varying*. In this case, we may replace $\sigma_\perp(s)$ and $\hat{v}_\perp(s)$ by their average values and obtain the approximate formula

$$\ell \approx \frac{L_\Gamma}{\sigma_\perp (1 + \hat{v}_\perp^2/\sigma_\perp^2)}. \quad (26)$$

Under these three assumptions, the above Sturm-Liouville problem is well-approximated by the Sturm-Liouville problem

$$\frac{d^2 z}{dt^2} + \bar{\lambda} z = 0, \quad 0 \leq t \leq \ell, \quad (27)$$

with boundary conditions $z'(0) - a(0)z(0) = 0$ and $z'(\ell) - a(\ell)z(\ell) = 0$. Let $\bar{\lambda}_n$ denote the n -th positive eigenvalue of (27). An elementary argument shows that $\bar{\lambda}_n$ is the n -th positive solution of the equation

$$\tan(\sqrt{\bar{\lambda}}\ell) = (a(\ell) - a(0)) \frac{\sqrt{\bar{\lambda}}}{\bar{\lambda} + a(0)a(\ell)} \quad (28)$$

with associated eigenfunction

$$\bar{z}_n(t) = A_n \cos\left(\sqrt{\bar{\lambda}_n}t + \phi_n\right),$$

where $\tan(\phi_n) = a(0)/\sqrt{\bar{\lambda}_n}$ and A_n is a positive constant.

The argument on pages 414–415 of Courant and Hilbert (1953) yields the formula $\lambda_n = \bar{\lambda}_n + e_n$, where $|e_n| \leq r_M$ for all n . Now using (28) and the inequalities $b < \pi/2\ell$ and $\ell \gg 2\pi$, it is not difficult (though somewhat tedious) to obtain the estimate

$$\left| \bar{\lambda}_n - \left(\frac{n\pi}{\ell}\right)^2 \right| < w_n \left(\frac{2n\pi}{\ell} + w_n\right) \ll \frac{2\pi}{\ell^2(1 - 1/n)} + \frac{2}{\ell^2(n - 1)^2},$$

where

$$w_n = \frac{2b}{(n - 1)\pi \left(1 - \left(\frac{b\ell}{(n-1)\pi}\right)^2\right)}.$$

Thus,

$$\lambda_n = \left(\frac{n\pi}{\ell}\right)^2 + E_n,$$

where

$$|E_n| < r_M + \frac{2\pi}{\ell^2(1 - 1/n)} + \frac{2}{\ell^2(n - 1)^2}.$$

It follows that $\lambda_n < 1$ if $(n + 1)\pi < \ell$ and $\lambda_n > 1$ if $n\pi > \ell$.

Using the approximate formula (26), we obtain the condition that $\lambda_n > 1$ when

$$2\pi\sigma_{\perp}(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2) > \frac{L_{\Gamma}}{n}. \quad (29)$$

B.4 Extremal Properties

We now show that the eigenfunctions associated with eigenvalues larger than 1 correspond to variational vector fields whose periods are small compared to the transverse standard deviation σ_{\perp} .

Calculations similar to those on pages 336–337 of Courant and Hilbert (1953) give the asymptotic estimate

$$z_n(t) = A_n \cos \left(\sqrt{\lambda_n} t + \phi_n \right) + O \left(\frac{1}{n} \right).$$

In terms of the original variables,

$$f_n(s) = A_n(s) \cos \left(\sqrt{\lambda_n} T(s) \right) + O \left(\frac{1}{n} \right),$$

where $A_n(s)$ now denotes a non-negative function independent of n . In the case where $\hat{v}_{\perp}(s)$ and $\sigma_{\perp}^2(s)$ are slowly varying, the eigenfunctions $f_n(s)$ oscillate with period dependent on s and given by the approximate formula

$$\frac{2\pi}{\left(\sqrt{\lambda_n} T'(s) \right)} \approx \frac{2\ell}{n} \sigma_{\perp} \left(1 + \frac{\hat{v}_{\perp}^2(s)}{\sigma_{\perp}^2(s)} \right).$$

Replacing $\hat{v}_{\perp}(s)$ and $\sigma_{\perp}^2(s)$ by their average values on Γ and ℓ by $(L_{\Gamma}/\sigma_{\perp})(1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2)^{-1}$ gives the approximation

$$\text{period of } f_n \approx \frac{2L_{\Gamma}}{n}.$$

Combining this with equation (29), we find that $\lambda_n > 1$ if $2\pi \sigma_{\perp} (1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2) > \text{period of } f_n$ while $\lambda_n < 1$ if $2\pi \sigma_{\perp} (1 + \hat{v}_{\perp}^2/\sigma_{\perp}^2) > \text{period of } f_{n+1}$.

References

- [1] J. D. Banfield and A. E. Raftery. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *JASA*, 87:7–16, 1992.
- [2] J.F. Brinkley. Knowledge-driven ultrasonic three-dimensional organ modeling. *IEEE Trans. Pat. Anal. Mach. Intell.*, 7(4):431–441, July 1985.
- [3] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience Publishers, Inc., New York, 1953.
- [4] L. Fang and D.C. Gossard. Reconstruction of smooth parametric surfaces from unorganized data points. In J. Warren, editor, *Curves and Surfaces in Computer Vision and Graphics 3*, volume 1830, pages 226–236. SPIE, 1992.
- [5] A. Goshtasby. Surface reconstruction from scattered measurements. In J. Warren, editor, *Curves and Surfaces in Computer Vision and Graphics 3*, volume 1830, pages 247–256. SPIE, 1992.
- [6] P.A. Griffiths. *Exterior Differential Systems and the Calculus of Variations*, volume 25 of *Progress in Mathematics*. Birkhauser, Boston-Basel-Stuttgart, 1983.
- [7] T.J. Hastie. *Principal Curves and Surfaces*. PhD thesis, Stanford University, 1984.
- [8] T.J. Hastie and W. Stuetzle. Principal curves. *JASA*, 84:502–516, 1989.
- [9] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface reconstruction from unorganized points. *Computer Graphics*, 26(2):71–78, 1992.
- [10] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Mesh optimization. *Computer Graphics*, 27:19–26, 1993.
- [11] R.W. Martin, G. Bashein, M.L. Nessly, and F. Sheehan. Methodology for three-dimensional reconstruction of the left ventricle from transesophageal echocardiograms. *Ultrasound in Med. and Biol.*, 19(1):27–38, 1993.

- [12] S. Muraki. Volumetric shape description of range data using “blobby model”. *Computer Graphics*, 25(4):227–235, July 1991.
- [13] R.B. Schudy and D.H. Ballard. Model detection of cardiac chambers in ultrasound images. Technical Report 12, Computer Science Department, University of Rochester, 1978.
- [14] R.B. Schudy and D.H. Ballard. Towards an anatomical model of heart motion as seen in 4-d cardiac ultrasound data. In *Proceedings of the 6th Conference on Computer Applications in Radiology and Computer-Aided Analysis of Radiological Images*, 1979.
- [15] F.H. Sheehan, M.P. Feneley, N.P. DeBruijn, J.S. Rankin, J.W. Davis, E.L. Bolson, P.S. Glass, and F.M. Clements. Quantitative analysis of regional wall thickening by transesophageal echocardiography. *Journal of Thoracic and Cardiovascular Surgery*, 103(2):347–354, 1992.
- [16] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Trans. Pat. Anal. Mach. Intell.*, 12(2):131–147, February 1990.
- [17] B.C. Vemuri, A. Mitiche, and J.K. Aggarwal. Curvature-based representation of objects from range data. *Image and Vision Computing*, 4(2):107–114, 1986.

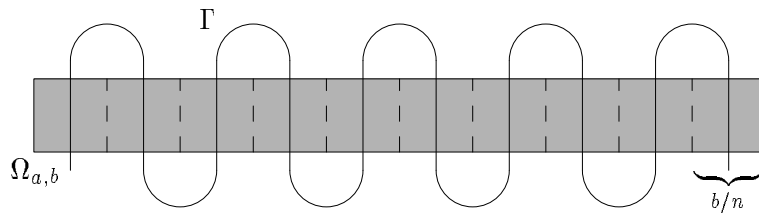


Figure 1: A principal curve for the uniform distribution on a rectangular strip. Points between dotted lines project horizontally onto the portion of Γ lying between them.

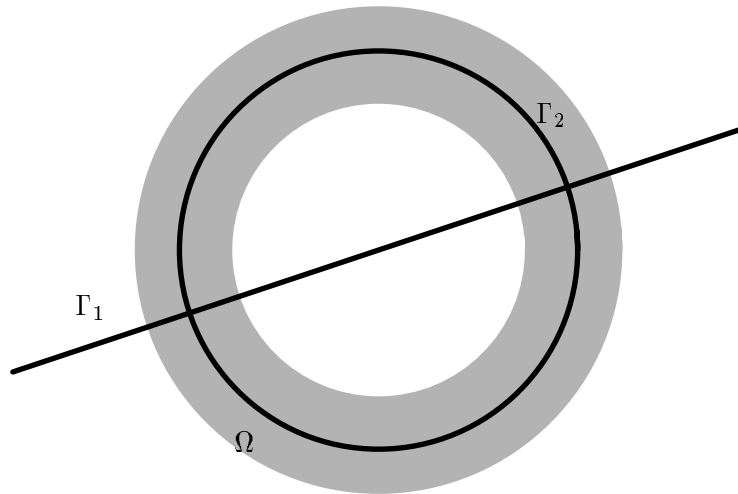


Figure 2: Γ_1 is a PC, but it is not admissible because it is not contained in Ω . Γ_2 is an admissible PC.

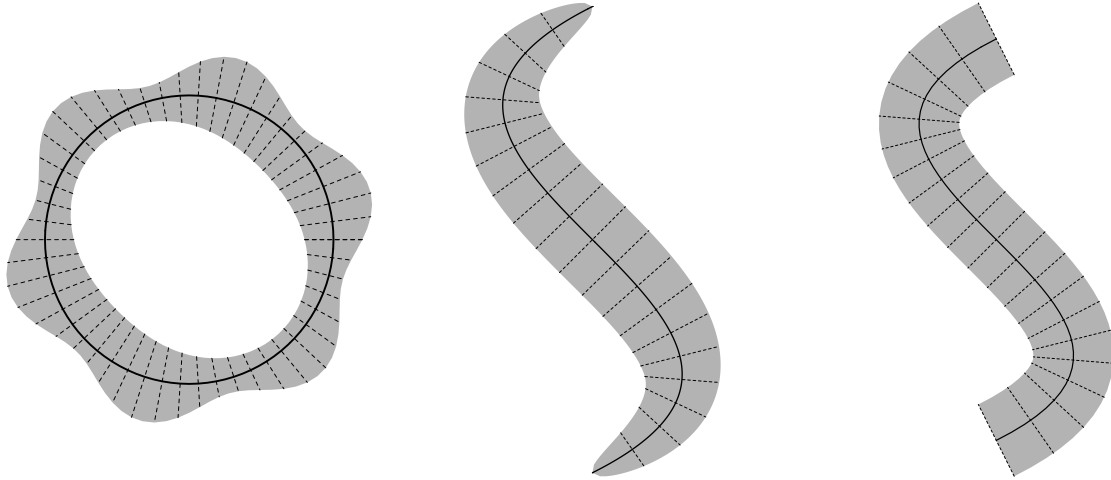


Figure 3: From left to right are: (a) an admissible loop, (b) an admissible curve with $\partial\Omega$ smooth, and (c) an admissible curve with essential endpoints. The dashed lines are the normal lines to the admissible curves.

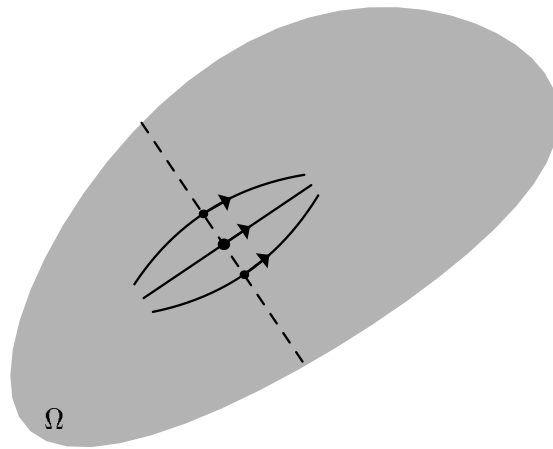


Figure 4: The curvature of a PC depends on the location of its intersection with the normal line

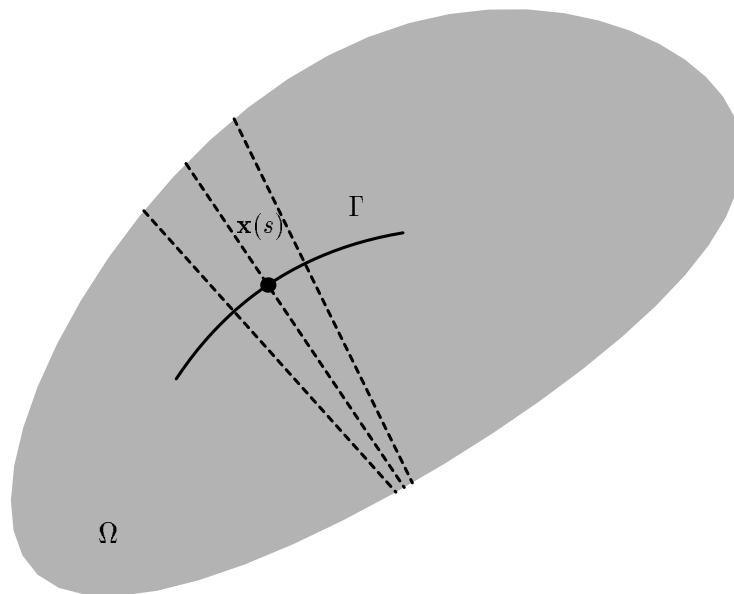


Figure 5: If the center of the normal line segment falls below the curve, the curvature has to be negative.

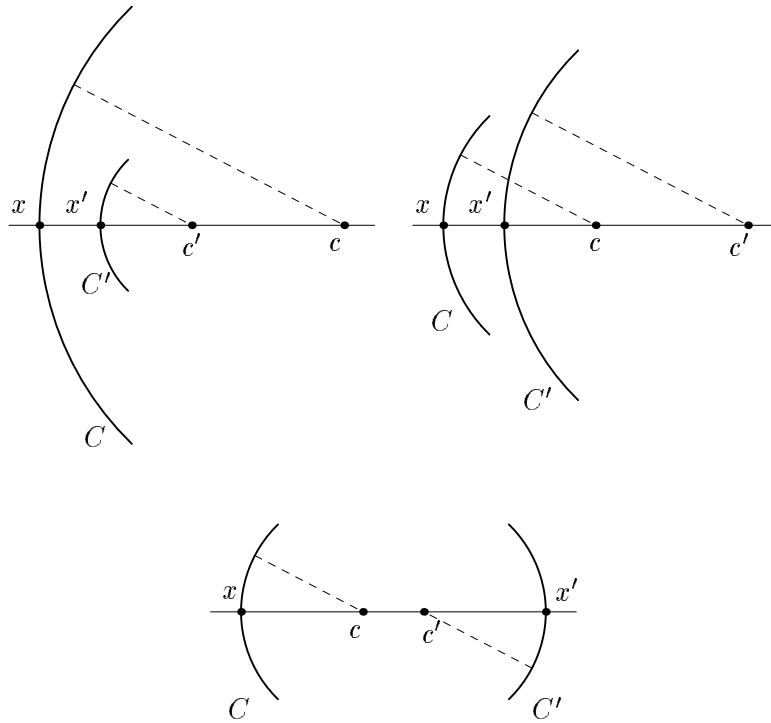


Figure 6: The three configurations above are the only ones satisfying the hypotheses of Lemma 1; but only the configuration at the left satisfies the condition that $|x - x'|$ equal the distance between C and C' .

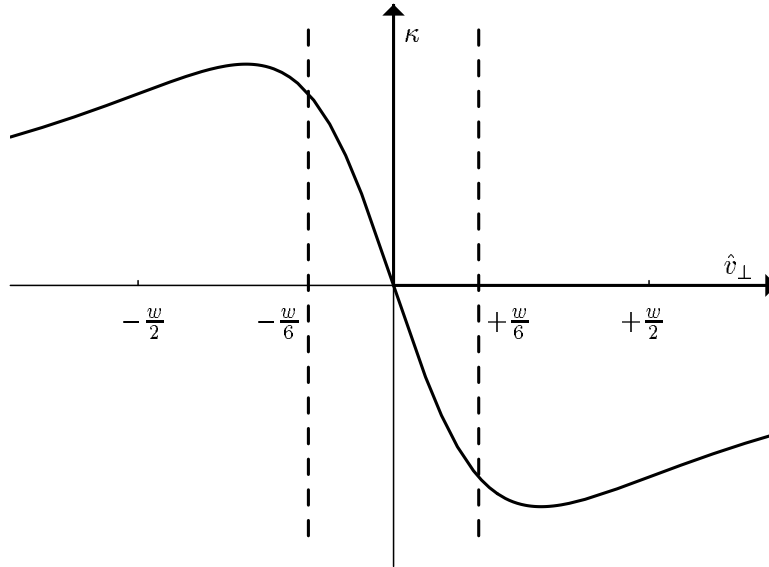


Figure 7: Curvature as a function of \hat{v}_\perp .

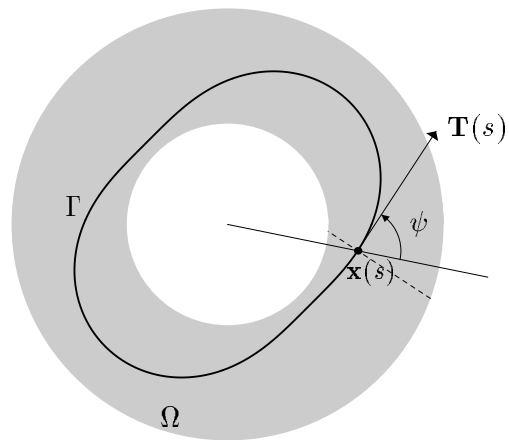


Figure 8: An admissible curve in the annulus $R_1 \leq r \leq R_2$ and one of its normal lines. The angle ψ is the angle between the radius vector and the tangent vector.

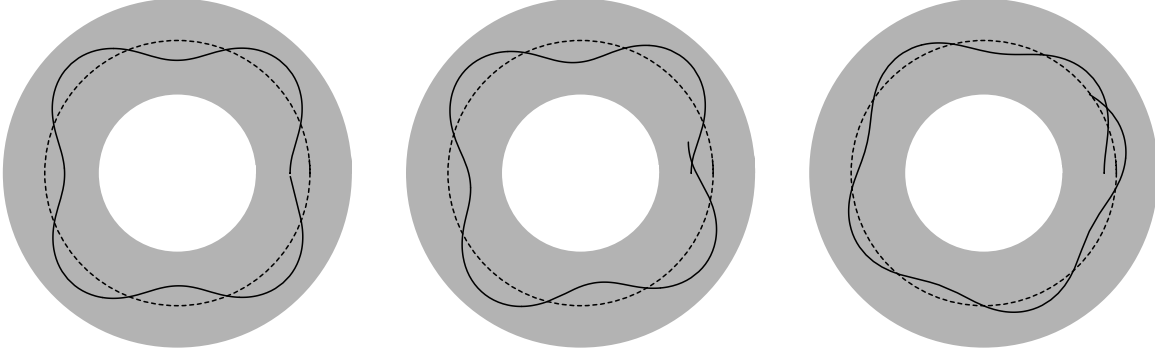


Figure 9: Three solutions of the principal curve equations (7) for the annulus $\Omega_{0.45,1}$. From left to right the initial conditions are: $r(0) = 0.646$, $r(0) = 0.634$ and $r(0) = 0.690$. Only the first is a closed curve, and thus a principal loop.

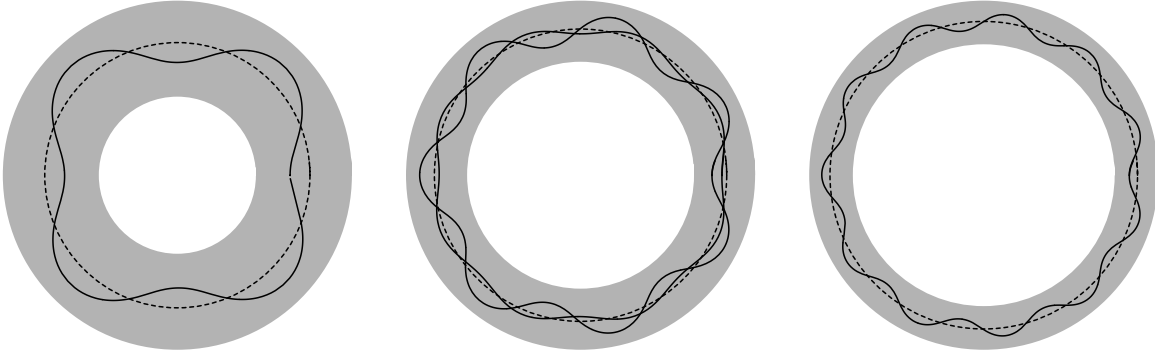


Figure 10: As the aspect ratio increases, the period of principal curves decreases. From left to right are the three annuli $\Omega_{0.45,1}$, $\Omega_{0.65,1}$ and $\Omega_{0.75,1}$ with some principal curves on each. Two PC's are shown for the middle annulus. The right annulus also supports several PC's (not shown).