Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

Estimating the Shape of a Moving Contour

Roger Brockett*
Division of Applied Sciences
Harvard University

Andrew Blake†
Department of Engineering Sciences
University of Oxford

1. Introduction

Most tracking problems in computer vision can be conceptualized as nonlinear estimation problems, with the gray level of each pixel being an observation. The difficulty in making use of this point of view is that i) we lack a plausible model for the underlying stochastic processes, and ii) the resulting problems would be of very high dimension as well as being nonlinear, making them all but intractable. The search for an approach that makes effective use of the existing spatial correlations has led to the study of various methods of simplification involving feature points, contours, blobs, etc. In particular there has been considerable work based on organizing visual evidence around parametrized contours called snakes [1].

It seems that the most geometrically natural evolutionary equations for curves in the plane, such as the smoothing flow for closed curves investigated by Gage and Hamilton [2], are nonlinear. On the other hand, Blake et al. [3] have shown that linear models based on finite dimensional parametrization can also be effective and remain tractable even in a stochastic setting. This paper continues in a similar way. We introduce a coordinate system and represent the curve using its horizontal and vertical components (x_1, x_2), expressed as functions of time and arc length. We introduce continuum models for the evolution of the coordinates and noisy observation models that allow us to formulate and solve a realistic class of estimation problems.

One interesting question brought into focus by this work is that of determining how to optimize the use of spatial correlation along the curve, and temporal correlation in the evolution of the curve, so as to reduce the effects of the observation noise. Our formulas in Section 7 give an answer to this question.

2. Stochastic Models for Contour Evolution

We begin by introducing a family of stochastic processes that evolve in infinite dimensional spaces. These will be spaces of functions, scalar or vector valued, defined on the circle or the real line. In a number of places we ignore technical points whose discussion would lead us too far from the applications of interest here. We will not distinguish between statements that are only true with probability one, and those that are true, etc.

It is a standard idea in control theory that linear differential equation models involving an exogenous input u, e.g.

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$

can serve as a basis for defining stochastic processes. Under suitable hypotheses we may replace u by a Wiener process, a jump process, etc., and generate a new process x whose statistical properties are shaped by A and B to approximate a particular situation of interest. Because we are discussing linear equations here we use a simplified notation, exemplified by $\dot{x} = Ax + B\dot{w}$. For such equations $\bar{x} = \mathcal{E}x$ satisfies

$$\frac{d}{dt}\bar{x}(t) = A(t)\bar{x}(t)$$

and $\Sigma = \mathcal{E}(\mathrm{x} - \bar{\mathrm{x}})(\mathrm{x} - \bar{\mathrm{x}})^{\mathrm{T}}$, satisfies

$$\dot{\Sigma}(t) = \Sigma(t)A^T(t) + A(t)\Sigma(t) + B(t)B^T(t)$$

If A and B are constant then it makes sense to ask about the existence of an equilibrium solution. Assuming that the null solution of the deterministic equation $\dot{x}(t) = Ax(t)$ is asymptotically stable, the steady state value of the mean is zero and the steady state value of the variance is given by

$$\Sigma_\infty = \int_0^\infty e^{-At}BB^Te^{A^Tt}\,dt$$

We will develop some analogous formulae covering cases in which x belongs to a vector space of functions taking on values in a finite dimensional space. Let L be a linear differential operator that generates a semigroup e^{Lt}. The equation

$$\dot{x}(t,s) = Lx(t,s) + \sum_{-\infty}^{\infty}\tilde{b}_i(s)u_i(t)$$

is a direct generalization of the finite dimensional situation. However, if L is an unbounded operator, e.g. if it is a partial differential operator, then some restrictions on the $b's$ may be necessary if we are to claim that there is a solution for a wide class of u's. Introducing a parameter Δ and rewriting this equation as

$$\dot{x}(t,s) = Lx(t,s) + \sum_{-\infty}^{\infty}\Delta b(s,i\Delta)u(t,\Delta)$$

suggests that one may consider

$$\dot{x}(t,s) = Lx(t,s) + \int_{-\infty}^{\infty} b(s,\eta)u(t,\eta)\,d\eta$$

*This work was supported in part by the National Science Foundation under Engineering Research Center Program, NSF D CDR-8803012 and by the US Army Research Office under grant DAAL03-86-K-0171(Center for Intelligent Control Systems) and the Newton Institute.

†The support of the Newton Institute, Cambridge, UK, is gratefully acknowledged.

as a limiting case. If b is a difference kernel, i.e. if $b(s,\eta) = b(s-\eta,0)$, then the s dependence exhibits a type of translation invariance. This plays a role in later sections. In particular, if b_α is a one parameter family of difference kernels such that b_α approaches a delta function as α goes to zero, this provides a way to approximate the simpler equation

$$\dot{x}(t,s) = Lx(t,s) + u(t,s)$$

If there is a countable number of controls, we may identify BB^T as

$$BB^T = \sum_{-\infty}^{\infty} b_i(s)b_i(r)$$

and in the integral case the role of BB^T is played by

$$\phi(s,r) = \int_{-\infty}^{\infty} b(s,\eta)b(\eta,r)\, d\eta$$

If $x(t,s)$ is the value of the x-component of a curve then it is important that it evolve in such a way as to maintain continuity in s; otherwise the curve would not hold together! The following observations will be useful.

Remark 1: If $u(t,\cdot)$ is assumed to be bounded but not necessarily continuous with respect to the spatial variable s then $x(t,\cdot)$ will not necessarily be continuous unless the operators L and B are suitably chosen. A full treatment would involve more mathematical questions than we have space to address here, but, roughly speaking, it is necessary for $L^{-1}B$ to provide the required smoothing of u. For example, if α and β are positive, a solution to the equation

$$\dot{x}(t,s) = (\alpha\frac{\partial^2}{\partial s^2} + \gamma)x(t,s) + \frac{1}{\sqrt{2\pi\beta}}\int_{-\infty}^{\infty} e^{\frac{-(s-\eta)^2}{2\beta}} u(t,\eta)\, d\eta$$

will be continuous if u is bounded and integrable. Even the special case obtained by equating α equal to zero evolves in the space of continuous functions if u is bounded and integrable. On the other hand, the equation

$$\dot{x}(t,s) = +a_2 x(t,s) + u(t,s)$$

can not be expected to generate a solution that is continuous with respect to s because $L^{-1}B$ is simply a constant and therefore provides no smoothing of u.

If $x(t,s)$ takes on values in R^n, we let L be an $n\times n$ matrix of operators, $B(s,\eta)$ an $n\times m$ matrix, and $u(t,s)$ an m-vector. This model presents no new difficulties provided that L is the infinitesimal generator of a semigroup and, in the cases where L is unbounded, if the operator B is sufficiently well-behaved so as to assure the existence of a solution for forcing functions u that are not necessarily continuous with respect to time or space.

The cases in which L has constant coefficients and b is translation invariant in the sense that $b(s,r) = b(s-r,0)$ are the most tractable. We will refer to such equations as being *spatially homogeneous.* We are also interested in the case where $x(t,\cdot)$ is periodic. If $(x_1(t,s), x_2(t,s))$ are the coordinates of a closed curve of length l and if $(x_1(t,s), x_2(t,s)) = (x_1(t,s+l), x_2(t,s+l))$, then x will evolve in such a way as to preserve both continuity and the equality of the end points, provided that L is spatially invariant, b is periodic, and $L^{-1}B$ is as above. Thus they provide a model for a stochastic process whose sample paths are one parameter families of closed curves in the plane.

Stochastic models can be constructed by replacing the controls by the increments of Wiener processes. If there is a countable number of controls this yields equations of the type

$$\dot{x}(t,s) = Lx(t,s) + \sum_{-\infty}^{\infty} \tilde{b}_i(s)\dot{w}_i(t)$$

A special case of this, corresponding to what one obtains when spatially discretizing the integral form described above, yields

$$\dot{x}(t,s) = Lx(t,s) + \sum_{-\infty}^{\infty} \Delta b(s,i\Delta)\dot{w}_i(t)$$

If we evaluate

$$\phi_1(s,s) = \mathcal{E}\sum_{-\infty}^{\infty} \Delta^2 b(s,i\Delta)b(i\Delta,r)\dot{w}_i(t)\dot{w}_s(t)$$

and compare it to the refined model obtained by dividing Δ is two,

$$\phi_2(s,s) = \mathcal{E}\sum_{-\infty}^{\infty} (\Delta/2)^2 b(s,i\Delta/2)b(i\Delta/2,r)\dot{w}_i(t)\dot{w}_i(t)$$

we see that in order to maintain the effective variance one needs to double the variance of the noise when dividing the spatial discretization by two. Our notation for the limiting form obtained by the repeated application of this process is

$$\dot{x}(t,s) = Lx(t,s) + \int_{-\infty}^{\infty} b(t,s,\eta)\dot{w}_\eta(t,\eta)\, d\eta$$

Any spatial coherence enjoyed by the solution of this equation must come from B and L, not from the driving noise.

3. Scalar Valued Functions

We may define the evolution of functions by means of a linear stochastic model driven by a family of independent white noise terms. Such equations may be formulated in a variety of function spaces. The functions may be supported on a finite interval, with or without boundary conditions, or on an infinite interval, with or without decay conditions, at infinity. If the spatial domain of the functions is an infinite interval, it may be natural to ask that the functions are square integrable along with some of its derivatives with respect to s or one may just ask that they be locally square integrable along with certain derivatives. Equations of the form

$$\dot{x}(t,s) = Lx(t,s) + \int_{-\infty}^{\infty} b(s,\eta)\dot{w}_\eta(t,\eta)\, d\eta$$

are of interest in either case. The simplest examples may be those that are defined on the whole real line and are translation invariant. The state space for such processes can not, of course, be a subspace of the space of square integrable functions but it could be a subspace of the space of locally square integrable functions.

Remark 2: Second order statistical characterizations such as $\mathcal{E}x(t,s)x(t,r)$ will play an important role in the discussion of the estimation problem. In order for the equation displayed above to define a process with finite variance it is necessary that the integral

$$\phi(s) = \int_{-\infty}^{\infty} b(s,\eta)b(\eta,s)\ d\eta$$

should not be too large relative to the size of L^{-1}. Of course $|\mathcal{E}x(t,s)x(t,r)| \leq \mathcal{E}x(t,s)x(t,s)$ so that the spatial correlation exists for all values of s and r if $\mathcal{E}x(t,s)x(t,s)$ is finite.

Remark 3: In order for the variance equation to reach steady state it is necessary that L generate a semigroup with b in its domain of definition.

The Fourier transform, as applied to the spatial variable s, is an effective tool for understanding the translation invariant problems of this type when the domain of definition is the whole line. Fourier series play a corresponding role when the curves are defined on the circle. We use

$$\check{x}(p,s) = \int_{-\infty}^{\infty} e^{-ipt}x(t,s)\ dt$$

to denote the temporal Fourier transform and

$$\tilde{x}(t,k) = \int_{-\infty}^{\infty} e^{-iks}x(t,s)\ ds$$

to denote the spatial Fourier transform.

Example 1: Consider the scalar equation

$$\dot{x}(t,s) = -ax(t,s) + \int_{-\infty}^{\infty} e^{-|s-\eta|}\dot{w}_\eta(t,\eta)\ d\eta$$

to be thought of defining a process on the space of locally square integrable functions on the line. The spatial autocorrelation function of the integral term is

$$\int_{-\infty}^{\infty} e^{-|s-\eta|}e^{-|\eta-r|}\ d\eta = (1+|s-r|)e^{-|s-r|}$$

Because the drift term is just $-ax$ we see that in steady state

$$\lim_{t\to\infty} \mathcal{E}x(t,s)x(t,r) = \frac{1+|s-r|}{2a}e^{-|s-r|}$$

The steady state spatio-temporal correlation function involves an additional factor

$$\lim_{t\to\infty} \mathcal{E}x(t,s)x(t+\tau,r) = \left(\frac{1+|s-r|}{2a}\right)e^{-|s-r|}e^{-|\tau|}$$

In this case the power spectrum of the signal, i.e. the temporal Fourier transform of the autocorrelation function, is

$$\int_{-\infty}^{\infty} \mathcal{E}x(t,s)x(\tau,r)e^{ip\tau}d\tau = \frac{1+|s-r|}{2a}e^{-|s-r|}\frac{1}{1+p^2}$$

If we also take the Fourier transform with respect to space, we get the spatio-temporal power spectrum

$$\hat{\phi}(p,k) = \frac{4}{(1+k^2)^2}\frac{1}{(1+p^2)}$$

Example 2: Consider the model

$$\dot{x}(t,s) = (a_1\frac{\partial^2}{\partial s^2} + a_2)x(t,s) + \int_{-\infty}^{\infty} e^{-(s-\eta)^2}\dot{w}_\eta(t,\eta)\ d\eta$$

The spatial correlation of the driving noise is

$$\int_{-\infty}^{\infty} e^{-(s-\eta)^2}e^{-(\eta-r)^2}\ d\eta = \frac{1}{\sqrt{\pi}}e^{-\frac{(s-r)^2}{2}}$$

We think of this as an equation on the subset of the space of locally square integrable functions on the real line consisting of those functions that have at least two locally square integrable derivatives with respect to s. We must assume that a_1 is positive if the drift term is to generate a semigroup. If we assume that a_2 is negative then there will be a steady state value of the variance. Under these assumptions we can look for the steady state value of the expectation of $x(t,s)$ as well as the steady state value of the various correlation functions. In this case we have

$$\begin{aligned}\mathcal{E}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b(s,\eta)\dot{w}(t,\eta)\ b(\chi,r)\dot{w}(t,\chi)\ d\chi\ d\eta \\ = \tfrac{1}{\sqrt{\pi}}e^{-(s-r)^2}\end{aligned}$$

The steady state value of the spatial autocorrelation function is given by

$$\lim_{t\to\infty} \mathcal{E}x(t,s)x(t,r) = \sigma(s-r)$$

with

$$\hat{\sigma}(k) = (2L(k))^{-1}\hat{\psi}(k)$$

where

$$\psi(s-r) = \mathcal{E}\int_{-\infty}^{\infty} b(s,\eta)b(\eta,r)\ d\eta$$

The time dependence of the correlation function can be expressed in terms of the solution of the deterministic initial value problem

$$\dot{x}(t,s) = (a_1\frac{\partial^2}{\partial s^2} + a_2)x(t,s) \quad ; \quad x(0,s) = given$$

This has the solution

$$x(t,s) = e^{a_2 t}\frac{1}{\sqrt{2\pi a_2 t}}\int_{-\infty}^{\infty} e^{-(s-\eta)^2/2a_1 t}x(0,\eta)\ d\eta$$

using this together with the steady state value of the covariance we can get an expression for the spatio-temporal autocorrelation function of the process evolving in the space of curves.

We observe that the larger the magnitude of a_1, the more rapidly the spatial correlation falls off with increasing spatial frequency. The larger a_2 is in magnitude, the faster the temporal correlation falls off in time.

Example 3: This example concerns stochastic processes whose sample values are functions defined on the unit circle. The evolution is again linear and of the same general form as defined above. Let $x(t,s) = x(t,s+2\pi)$. The process defined by

$$\dot{x}(t,s) = (a_1\frac{\partial^2}{\partial s^2} + a_2)x(t,s) + \int_{-\pi}^{\pi} \cos(s-\eta)\dot{w}(t,\eta)\ d\eta$$

has sample paths that are continuous on the circle. If a_2 is negative and a_1 is zero then the steady state value of the spatial covariance is

$$\mathcal{E}x(t,s)x(t,r) = \frac{1}{2a_2}\int_{-\pi}^{\pi}\cos(s-\eta)\cos(\eta-r)\ d\eta$$

or

$$\mathcal{E}x(t,s)x(t,r) = \frac{1}{2a_2}(1+\cos(s-r))$$

Example 4: Again we consider stochastic processes whose sample values are functions defined on the unit circle but now we allow the dynamics to be second order.

$$\ddot{x}(t,s) + \alpha\dot{x}(t) = (a_1\frac{\partial^2}{\partial s^2} + a_2)x(t,s) + \int_{-\pi}^{\pi}\cos(s-\eta)\dot{w}(t,\eta)\ d\eta.$$

The process so defined has sample paths that are continuous on the circle and statistics that are rotationally invariant. If a_2 is negative and a_1 is zero then the steady state value of the spatial covariance is

$$\mathcal{E}x(t,s)x(t,r) = \frac{1}{2b}(1+\cos(s-r))$$

$$\mathcal{E}\dot{x}(t,s)x(t,r) = 0$$

$$\mathcal{E}\dot{x}(t,s)\dot{x}(t,r) = \frac{1}{2a_2}(1+\cos(s-r))$$

4. Vector Valued Functions

By considering vector-valued functions, we can obtain models that can be used to study the problem of tracking closed curves in the plane. If we chose a Cartesian coordinate system and parameterize the curve by its arc length then the locus of points $(x_1(s), x_2(s))$ for $0 \le s \le l$ define the curve. Of course $(x_1(s+l), x_2(s+l)) = (x_1(s), x_2(s))$ so we can think of the coordinate functions as being defined on a circle of circumference l. The models given in the previous section tend to a steady state whose expected value is zero, however, by adding a bias term one can get stochastic processes in the space of closed curves whose steady state expectation is any particular closed curve one wants.

Example 5: Let x be a two dimensional vector

$$x(t,s) = (x_1(t,s), x_2(t,s))$$

Suppose that x evolves according to the equation

$$\dot{x}(t,s) = Lx(t,s) + \int B(s,r)\dot{w}(t,r)\ dr$$

with L being a generator of a semigroup and B being a matrix-valued function whose entries are differentiable functions of $s-\eta$, square integrable in the sense used in Example 4. These conditions, and the specification of L, are guided by the idea that the model should yield an x that evolves in the space of smooth curves. We also want the statistical properties of these curves to have limiting values as t goes to infinity. Let $B(s,r)$ be given by diag $((\cos(s-r), \cos(s-r))$. Let L be multiplication by a constant and let (w_1, w_2) be a vector-valued Wiener process. Thus

$$\begin{aligned}\ddot{x}_1(t,s) + a_{11}\dot{x}_1(t,s) + a_{12}\dot{x}_2(t,s) = \\ \textstyle\int_0^{2\pi}\cos(s-r)\dot{w}_1(t,r)\ dr \\ \ddot{x}_2(t,s) + a_{21}\dot{x}_2(t,s) + a_{22}\dot{x}_2(t,s) = \\ \textstyle\int_0^{2\pi}\sin(s-r)\dot{w}_2(t,r)\ dr\end{aligned}$$

In this case the system is simplified in various ways. For example, if we expand x in a Fourier series in the arc length s, then only the first term in the Fourier series participates in the motion.

5. The Observation Model

The main point of this paper is to explore the problem of removing the effects of noise in the process of identifying the location of a moving contour. Because the whole idea of organizing evidence using curves rather than points or areas can be viewed as a way of approximating the optimal estimator, the choice of model is critical. The particular models we introduce here are, in effect, simply ways to exploit spatial correlation in the observation process. The edge detection processes available in computer vision have been refined over many years but the process is intrinsically noisy. A prevalent view is that more effective edge location will require algorithms that take into account more explicitly the coherence of the curve. The models of the previous sections have been constructed with the idea of being able to use the spatial correlation to improve the contour detection and, of course, are designed to treat the more general problems associated with tracking moving contours. To complete our specification of the models, however, it is necessary to give equations for the measurements. These must reflect the inaccuracy of the detection process and any spatial or temporal coherence in the errors made by the contour detection process.

The abstract linear model that underlies the Kalman-Bucy filter, postulates the specification of a model for the observation process as well as a linear differential equation model for the signal itself. In one popular notational scheme the observations are written with an over dot so that the model appears as

$$\dot{x}(t) = Ax(t) + B\dot{w}(t)$$

$$\dot{y}(t) = Cx(t) + D\dot{\nu}(t)$$

with the understanding that ν is a, possibly vector-valued, Wiener process which we take to be independent of w.

Having discussed a set of choices for the differential equation appropriate for generating stochastic process in the space of contours, it remains to be discussed the choice of C and D. When the cardinality of the observation set is infinite, i.e. when $y(t,\cdot)$ belongs to an infinite dimensional space, to know y one requires an infinite number of measurements. If $x(t,\cdot)$ is continuous with respect to its second argument and if the observations are too numerous, in a sense that we will illustrate shortly, the estimation problem may be meaningless because the observations determine $x(t,s)$ with vanishing mean-square-error. A similar problem arises when discussing observation processes that are continuous in time. We simply adapt the standard treatment of that situation to the spatial domain.

Let $\dot{\nu}_i(t)$ be a set of unit strength white noise processes. Consider a countable set of observations uniformly distributed in space and separated by Δ units.

$$y_i(t) = x(t, i\Delta) + \dot{\nu}_i(t) \; ; i = 0, \pm 1, \pm 2, \ldots$$

If we now make observations that are separated in space by $\Delta/2$ but with additive noise having twice the variance, i.e. if we switch to

$$y_i(t) = x(t, i\Delta/2) + 2\dot{\nu}_i(t) \; ; i = 0, \pm 1, \pm 2, \ldots$$

we would determine a constant $x(t, \cdot)$ with the same error variance. Repeating this process of reducing the distance between measurements by one-half and doubling the variance of the additive noise, we approach a limit which we write as

$$y(t, s) = x(t, s) + \dot{\nu}_s(t, s)$$

This corresponds to very dense spatial sampling with large, independent noise variances. The effective noise variance associated with such a model can be associated with the ratio of the strength of the noise variance divided by Δ, the distance between the initial observations. By letting the variance of the noise go to infinity as the spacing goes to zero, we can obtain a meaningful limiting form. A class of observation models that are spatially invariant, and therefore more tractable, are

$$\dot{y}(t, s) = \int_{-\infty}^{\infty} c(t, s-\eta) x(t, \eta) \, d\eta + \int_{-\infty}^{\infty} d(t, s-\eta) \dot{\nu}_\nu(t, \eta) \, d\eta$$

These incorporate some spatial smearing of the value of $x(t, \eta)$.

If the process to be estimated is vector valued, we can take the observation process to be a two component vector of the form

$$\dot{y}_i(t, s) = \int h_i(t, s - r) x(t, r) \, dr \, dt + \dot{w}_i(t, s)$$

It is also useful to consider models of the above type in a more general setting involving an unstructured, finite dimensional part that is coupled to the infinite dimensional part. Such supplements can be used, for example, to model the classical dynamics of a rigid body whose motion determines aspects of the moving contour.

6. Recursive Estimation

Given a Gauss-Markov model for a stochastic process and a linear observation model of the type discussed above, the Kalman-Bucy filter generates the conditional mean of the state via an equation of the form

$$\frac{d}{dt}\hat{x}(t) = A\hat{x}(t) + \Sigma C^T (y(t) - C\hat{x}(t))$$

where Σ is the error covariance. In the spatially homogeneous case this takes the more concrete form

$$\frac{d}{dt}\hat{x}(t, s) = L\hat{x}(t, s) + \int_{-\infty}^{\infty} f(s - \eta)(y(t, \eta) - C\hat{x}(t, \cdot)) d\eta$$

The problem of finding the optimal filter is the problem of finding f. The remainder of this section and all of the next is devoted to showing how to do this.

There is a nonlinear differential equation of the Riccati type that Σ satisfies. In all but the simplest finite dimensional problems, Σ is usually found numerically. The most tractable situations correspond to solving for its steady state value, in which case one can also use Wiener's spectral factorization method. In the spatially homogeneous situation the error variance is characterized by a difference kernel. We will show how the Fourier transform with respect to space can be used to "diagonalize" the process of finding a solution, thereby giving us an expression for the Fourier transform of f. The spatially homogeneous case is of interest in its own right and also gives insight about more general situations.

Consider a scalar, spatially invariant model of the form

$$\dot{x}(t, s) = Lx(t, s) + \int_{-\infty}^{\infty} b(s - \eta) \dot{w}(t, \eta) d\eta$$

$$y(t, s) = x(t, s) + \dot{\nu}(t, s)$$

To find the optimal filter we must solve a Riccati equation in the error variance. Denoting this quantity by σ, the equation is

$$\dot{\sigma}(t, r, s) = 2L\sigma(t, r, s) + R(t, r, s) - \sigma^2(t, r, s)$$

where R is a double integral formed from b. If the large t limit of the solution exists then it satisfies

$$\sigma^2(s, r) - 2L\sigma(s, u) - R(s, r) = 0$$

In terms of the solution of error variance, the conditional mean is generated by

$$d\tilde{x}(t, s)/dt = L\tilde{x}(t, s) + \sigma(t, s, r)(y(t, s) - \tilde{x}(t, s))$$

We require a few standard formulae from estimation theory. A standard problem in Wiener theory involves separating a signal having power spectrum $\phi_{yy}(p)$, from additive noise having power spectrum $\phi_{nn}(p)$. In terms of differential equations, the problems can be posed using

$$\dot{x}(t) = -ax(t) + b\dot{w}(t)$$

$$\dot{y}(t) = cx(t) + d\dot{\nu}(t)$$

In this case the power spectrum of the y-process is

$$\phi_{yy}(p) = \frac{b^2 c^2}{a^2 + p^2}$$

The power spectrum of the noise is

$$\phi_{nn}(p) = d^2$$

If we define $y_1 = y/d$ then the estimation problem takes on the standard form. The corresponding Riccati equation is

$$\dot{\sigma} = -2a\sigma(t) + b^2 - \sigma^2 \frac{c^2}{d^2}$$

This equation has an equilibrium solution

$$\sigma_\infty = \frac{-ad^2}{c^2} - \sqrt{\frac{a^2 d^4}{c^4} + \frac{b^2 d^2}{c^2}}$$

The best filter, expressed in terms of y_1, is first order and takes the form

$$\frac{d}{dt}\hat{x}(t) = -a\hat{x}(t) + \frac{\sigma c}{d}(\dot{y}_1(t) - \frac{c}{d}\hat{x}(t))$$

with $-a - \sigma c/d$ being given by

$$a - \frac{\sigma c^2}{d^2} = -\frac{ad}{c} + \frac{ad}{c}\sqrt{1 + \left(\frac{ca^{-1}b}{d}\right)^2}$$

Expressing matters in terms of the original notation, i.e. in terms of y, we get

$$\frac{d}{dt}\hat{x}(t) = -a\hat{x}(t) + \beta(\dot{y}(t) - c\hat{x}(t))$$

with

$$\beta = -\frac{a}{c}(1 + \sqrt{1 + \left(\frac{ca^{-1}b}{d}\right)^2}$$

7. Fourier Diagonalization

We now turn to the question of finding the optimal steady state filter in the spatially homogeneous situation. By working with the Fourier transformed version of the evolution equation, we reduce the Riccati equation to a family of scalar equations parameterized by the spatial Fourier transform variable k. If we limit ourselves to reasonably simple time-evolution equations then the steady state values of the Riccati equation can be determined and the choice of branches for the various algebraic functions can be described explicitly.

Example 6: In order to illustrate these ideas we postulate a model that allows us to investigate how spatial correlation can be best used in estimating the function. Consider

$$\dot{x}(t,s) = -ax(t,s) + \hat{b}(t,s)$$
$$\dot{y}(t,s) = x(t,s) + \hat{d}(t,s)$$

with $\hat{b}$ and $\hat{d}$ being white in time and with spatial power spectra of $1/(1+k^2)$ and $1/(\mu^2+k^2)$, respectively.

In a typical application, the spatial scale of the noise would be much shorter than the spatial scale of the contour and so k would be significantly larger than one.

The optimal filter is given by

$$\frac{d}{dt}\hat{x}(t,s) = -ax(t,s) - \int_{-\infty}^{\infty} \beta(s-r)(\hat{x}(t,r) - y(t,r))\,dr$$

With β being $-\sqrt{a^2+1}\delta(s-r)$ plus the inverse Fourier transform of the square integrable function

$$\tilde{\beta}(k) = -\sqrt{a^2 + \frac{\mu^2+k^2}{1+k^2}} + \sqrt{a^2+1}$$

This function takes on the value $-\sqrt{a^2+\mu^2} + \sqrt{a^2+1}$ at $k=0$ and the value 0 at $k=\infty$. Approximating this by a rational function of the form

$$\tilde{\beta}_1(k) = \frac{-\sqrt{a^2+\mu^2} + \sqrt{a^2+1}}{1+\alpha_1^2 k^2}$$

we see that

$$\beta_1(s-n) \approx -\sqrt{a^2+1}\delta(s-r) -$$
$$(\sqrt{a^2+\mu^2} - \sqrt{a^2+1})e^{-\alpha_1|s-r|}$$

will be an approximation to the kernel of the optimal filter. Written out in full, the approximate filter is

$$\tfrac{d}{dt}\hat{x}(t,s) = (-a-\gamma)x(t,s) +$$
$$\int_{-\infty}^{\infty} \gamma e^{-\alpha_1|s-r|}(y(t,r) - x(t,r))\,dr + \gamma y(t,s)$$

with

$$\gamma = -\sqrt{\alpha^2+\mu^2} + \sqrt{a^2+1}$$

More generally, in the case where a is a constant, one obtains

$$\beta(k) = -a\sqrt{1 + \left(\frac{\phi_{yy}}{a\phi_{nn}}\right)^2}$$

which, when reexpressed as

$$\beta(k) = \beta_1(k) + f$$

with f a constant and $\beta_1(k)$ an $L_2(-\infty,\infty)$ function, allows one to find the optimal kernel.

If a is a constant coefficient differential operator this last formula still defines the optimal kernel provided we replace a^2 by $\tilde{a}(k)\tilde{a}(-k)$, with $\tilde{a}$ being the Fourier transform of the operator.

8. References

1. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour models," *Proc. 1st Int. Conf. on Computer Vision*, pp. 259-268, 1987.
2. M. Gage and R. S. Hamilton, "The Heat Equation shrinking convex plane curves," *J. Differential Geometry*, Vol. 23, (1986), pp. 69-96.
3. Andrew Blake, Rupert Curwen, and Andrew Zisserman, "A framework for spatio-temporal control in the tracking of visual contours," *Int. Journal of Computer Vision*, Vol. 11, pp. 127-145., 1993.
4. R. W. Brockett and J. L. Willems, "Discretized Partial Differential Equations: Examples of Control Systems Defined on Modules," *Automatica*, Vol. 10 (1974) pp. 507-515.
5. Andrew Blake, Michael Isard and David Reynard. "Learning to track curves in motion," *Proc. IEEE Conf. CDC*, (1994).
6. R. W. Brockett, "Gramians, Generalized Inverses, and the Least Squares Approximation of Optical Flow," *Journal of Visual Communications and Image Representation*, Vol. 1, No 1, (1990) pp.3-12.
7. R. W. Brockett, "Least Squares Matching Problems," *Linear Algebra and Its Applications*, Vols. 122/123/124, pp. 761-777, 1989.
8. Robert J. Adler, *The Geometry of Random Fields*, New York, Wiley,1981.