Nonlinear Estimation

Sanjoy K. Mitter

Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Laboratory for Information and Decision Systems
Cambridge, MA 02139

Nonlinear Estimation

Nonlinear Estimation problems may be divided into two categories:

(1) Static Estimation of Nonlinear Systems

(2) State Estimation for Nonlinear Dynamical Systems.

Static Nonlinear Estimation

A general static non-linear estimation problem can be modelled by considering a state vector $x \in \mathbf{R}^n$ which can be observed through a nonlinear sensor resulting in a signal $y \in \mathbf{R}^p$ which is measured in the presence of additive noise. In many situations the noise can be modelled as a gaussian random variable. The general model is therefore

$$y = h(x) + v \tag{1}$$

where $h : \mathbf{R}^n \to \mathbf{R}^p$ transforms the state into the observed signal and v is an $\mathbf{R}^p$-valued Gaussian random variable. We adopt a probabilistic viewpoint which unifies the static and dynamic situation, namely, we extract the maximum probabilistic information contained in y about x. This means that we are required to compute the conditional probability density (assuming all random variables have a density)

$$p(x|y), \tag{2}$$

which by Bayes Theorem can be computed as

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = L(x,y)p(x). \tag{3}$$

$L(x,y)$ is the so-called likelihood function and $p(x)$ is the prior density of x, representing the prior knowledge of the state x. The estimator is the mapping

$$y \to p(x|y) \tag{4}$$

and specific estimates such as the conditional mean, $\hat{x} = \int_{\mathbf{R}^p} p(x|y)\, dy$ or the maximum likelihood estimate $\tilde{x} = \max_x L(x,y)$ can all be obtained from $p(x|y)$. In general the estimates are non-linear functions of the observed signal y.

The conditional mean estimate $\hat{x}$ can be computed by solving a non-linear least squares problem in the sense that the estimate $\hat{x} = \varphi(y)$ is obtained by minimizing $\min_{\psi} E[(x-\psi(y))^2]$, and in order for this problem to make sense we can assume that all random variables have finite energy. We carry out the same program in a dynamical context in the next sections.

State Estimation for Nonlinear Dynamical Systems

There are several issues to be considered in formulating the state-estimation problem for non-linear dynamical systems. Firstly, there are modelling issues, namely, the modelling of process noise, the modelling of the sensor and modelling of the measurement noise. We use models in the framework of Markov Diffusion processes. The case where the process and observation noise are point-processes can be developed along similar lines.

The Filtering Problem Considered, and the Basic Questions

We consider the signal-observation model:

$$\left.\begin{aligned} dx_t &= b(x_t)dt + \sigma(x_t)dw_t \quad x(0) = x_0 \\ dy_t &= h(x_t)dt + d\eta_t \qquad\qquad 0 \le t \le 1 \end{aligned}\right\} \tag{5}$$

which is a differential notation for the integral equation

$$\begin{aligned} x_t &= x_0 + \textstyle\int_0^t b(x_s)ds + \int_0^t \sigma(x_t)dw_t \\ y_t &= y_0 + \textstyle\int_0^t h(x_s)ds + \eta_t \end{aligned} \tag{5'}$$

In the above, x_t is the state at time t, w_t is Brownian motion whose formal derivative $\dot{w}_t$ is white noise, y_t is the observation at time t and η_t is also Brownian motion which is independent of

w_t. For simplicity all random variables are scalar random variables. $b(\cdot)$ represents the drift of the state process, σ the diffusion coefficient and h models the sensor. Equation (5) is a precise way of describing a non-linear dynamical system with a white noise input. The reader may compare this model with the model for a Kalman filter where the function b, σ and h are linear.

For further details about modelling of stochastic dynamical systems the reader may consult the references cited at the end of the section. Suffice it to say that the calculus involved in the description of stochastic dynamical systems is not ordinary differential calculus but the so-called Ito differential calculus where certain second-order terms have to be included in the differentiation formula since they have first order effects. In turn, this depends on variance properties of Brownian motion in the sense that the variance of w_t behaves like t. Thus the differential of $f(x_t)$ is

$$df = \frac{\partial f}{\partial x} dx_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} dt$$

and not

$$df = \frac{\partial f}{\partial x} dx_t$$

as in ordinary calculus.

A word about the modelling of the observation process y_t is in order. If we write the second equation in (5) as

$$\dot{y}_t = h(x_t) + \dot{\eta}_t,$$

then $\dot{\eta}_t$ should be considered as white noise approximation of wide-band observation noise.

The fundamental description of the filter is obtained by describing the evolution of the conditional density.

In order to describe the evolution of the conditional density, let us first consider the situation where there are no observations present. In this situation, the evolution of the probability density of x_t can be described by the so-called Forward Kolmogoroff equation

$$\begin{cases} \frac{\partial \rho}{\partial t}(t,x) = \mathcal{L}^* \rho(t,x) \\ \rho(0,x) = \rho_0(x) \end{cases} \tag{6}$$

where $\mathcal{L}^*$ is the formal adjoint of the operator

$$\mathcal{L}_\varphi = \frac{1}{2}\sigma(x)\frac{\partial^2 \varphi}{\partial x^2} + b(x)\frac{\partial \varphi}{\partial x}. \tag{7}$$

For example, if x_t were Brownian motion, then the evolution equation for the density of x_t is the familiar heat equation

$$\begin{cases} \frac{\partial \rho}{\partial t}(t,x) = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2}(t,x) \\ \rho(0,x) = \rho_0(x). \end{cases} \tag{8}$$

The description of the evolution of the conditional density equation leads to a non-linear stochastic partial differential equation, the so-called Kushner-Stratanovich equation.

It turns out that the conditional density $\tilde{p}(t,x)$ (which depends on the observation y; the dependency has been suppressed in the notation) can be written as

$$\tilde{p}(t,x) = \frac{p(t,x)}{\int_{\mathbf{R}} p(t,x)dx}. \tag{9}$$

In view of this it is natural to call $p(t,x)$ the unnormalized conditional density. The advantage of working with $p(t,x)$ is that is satisfies a linear stochastic partial differential equation, the so-called Zakai equation

$$\begin{cases} dp(t,x) = \mathcal{L}^* p(t,x)dt + h(x)dy_t \\ p(0,x) = p_0(x), \end{cases} \tag{10}$$

where $\mathcal{L}^*$ is as described by equation (7) and the term $h(x)dy_t$ represents the interaction of the state process x_t with the observation process y_t.

As an example, suppose that x_t is Brownian motion and let the observation process be described by

$$dy_t = x_t dt + d\eta_t. \tag{11}$$

In this case equation (10) becomes

$$dp(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}p(t,x)dt + xdy_t. \tag{12}$$

The reader should recognize that we are now in the Kalman filter situation and hence $p(t,x)$ must be conditionally gaussian. Equation (12) can be solved explicitly by wiring

$$p(t,x) = K \exp -\frac{1}{2}P(t)(x - m(t))^2 \tag{13}$$

where K is a constant and obtaining equations for $m(t)$ and $P(t)$. The equation for $m(t)$ and $P(t)$ are nothing else but the Kalman filter with $m(t)$ representing the evolution of the conditional mean.

In order to describe the evolution of the normalized conditional density we introduce some notation. Denote by

$$\pi_t(h) = \int_{\mathbf{R}} h(x)\tilde{p}(t,x)dx \tag{14}$$

$\pi_t(h)$ denotes the estimate (conditional) of h. For example, the conditional mean estimate is given by

$$\pi_t(x) = \int_{\mathbf{R}} x\tilde{p}(t,x)dx \tag{15}$$

The normalized conditional density $\tilde{p}(t,x)$ satisfies the so-called Kushner-Stratanovich equation

$$d\tilde{p}(t,x) = \mathcal{L}^*\tilde{p}(t,x)dt + \tilde{p}(t,x)[h(x) - \pi_t(h)]d\nu_t \tag{16}$$

In the above, ν_t, the innovation process representing the new information contained in the observation is given by

$$d\nu_t = dy_t - \pi_t(h)dt \tag{17}$$

It can be shown that ν_t is Brownian motion which is adapted to the observation process y_t. The fact that ν_t is Brownian motion has the interpretation that the optimal filter has extracted all relevant information from the observation and has left us with white noise ν_t which contains no information. This question is however subtle since in a certain sense ν_t contains the same information as y_t, that is, one can pass from y_t to ν_t via a causal and causally-invertible transformation.

A stochastic differential equation for the conditional mean (or other conditional moments) can be written down, but the computation of the conditional mean depends on the conditional variance which in turn depends on the third conditional moment. Thus, in general, the non-linear filtering problem does not lead to finite-dimensional recursive filters (excepting in the linear and certain special cases).

Since the conditional mean filter is not finite-dimensionally recursively computable one has to resort to approximations. The most popular of these approximate filters is the Extended Kalman filter which is obtained through a process of linearization around the current estimate. Assuming σ in equation (5) is the identity leads to the filter

$$d\hat{x}_t = b(\hat{x}_t)dt + K(t)[dy_t - h(\hat{x}_t)dt]$$

where the gain $K(t)$ is obtained as follows:

$$\dot{P}(t) = B(\hat{x}_t, t)P(t) + P(t)B^T(\hat{x}_t, t) + I$$
$$-P(t)H^T(\hat{x}_t, t)R^{-1}(t)H(\hat{x}_t, t)P(t); P(0) = P_0$$
$$B(\hat{x}_t, t) = \frac{\partial b(x_t, t)}{\partial x_t}|_{x_t = \hat{x}_t}$$
$$H(\hat{x}_t, t) = \frac{\partial h(x_t, t)}{\partial x_t}|_{x_t = \hat{x}_t}$$

In the above note that $\hat{x}_t$ is not the conditional mean but may provide a good approximation to it.

In many situations the Extended Kalman Filter gives good performance but if local observability fails the filter may become unstable.

Applications

Ideas of non-linear filtering, especially the Extended Kalman filter has been widely applied in guidance of missiles, tracking and in the design of receivers. In general, linear filtering suffices when the signal to noise ratio is high. To improve performance in regimes where the signal to noise ratio is low some nonlinear effects need to be introduced. Some success in doing this has been achieved in the design of phase-locked loops.

References

[1] A. Gelb (Editor): *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, Mass. 1974.

[2] A.H. Jazwinski: *Stochastic Processes and Filtering Theory*, Academic Press, N.Y. 1970.

[3] E. Wong: *Stochastic Processes in Information and Dynamical Systems*, McGraw Hill, New York, 1971.