Model Based Smoothing of Linear and Nonlinear Processes

Arthur J. Krener
Department of Mathematics
University of California
Davis, CA, 95616, USA

ajkrener@ucdavis.edu

Dynamic Models

We assume that there are three processes.
The observed "input" process $\boldsymbol{u(t)} \in I\!\!R^m$,
the observed "output" process $\boldsymbol{y(t)} \in I\!\!R^p$
and the hidden "state" process $\boldsymbol{x(t)} \in I\!\!R^n$.
We will restrict to $\boldsymbol{m, \ n, \ p}$ finite.
We will restrict to continuous time.

We also assume that the processes are related by some model, which in the absence of noise is of the form

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f(x, u)} \\ \boldsymbol{y} &= \boldsymbol{h(x, u)} \end{aligned}$$

There may also be additional information such as

$$\boldsymbol{x(0) = x^0} \qquad \text{or} \quad \boldsymbol{b(x(0), x(T)) = 0}$$

In controls community there are two important problems.

Filtering: From the knowledge of the past inputs $\boldsymbol{u}(\mathbf{0}:\boldsymbol{t})$, the past outputs $\boldsymbol{y}(\mathbf{0}:\boldsymbol{t})$, the model $\boldsymbol{f}, \boldsymbol{h}$ and the initial condition $\boldsymbol{x}^{\mathbf{0}}$, estimate the current state $\boldsymbol{x}(\boldsymbol{t})$.

Typically filtering needs to be done in real time as $\boldsymbol{t}$ evolves because the state estimate $\hat{\boldsymbol{x}}(\boldsymbol{t})$ will be used to determine the current control $\boldsymbol{u}(\boldsymbol{t}) = \boldsymbol{\kappa}(\hat{\boldsymbol{x}}(\boldsymbol{t})$ to achieve some goal such as stabilization.

Smoothing: From the knowledge of the inputs $\boldsymbol{u}(\mathbf{0}:\boldsymbol{T})$, the outputs $\boldsymbol{y}(\mathbf{0}:\boldsymbol{T})$, the model $\boldsymbol{f}, \boldsymbol{h}$ and the initial condition or boundary conditions, estimate the state $\boldsymbol{x}(\mathbf{0}:\boldsymbol{T})$.

Typically smoothing is done off-line after the data has been collected. It is done on a fixed interval $\boldsymbol{t} \in [\mathbf{0}, \boldsymbol{T}]$. Since we are using more data, we expect that the smoothed estimate to be more accurate than the filtered estimate.

This talk will focus on smoothing but I will make a few general remarks about filtering.

There are both deterministic and stochastic approaches to filtering.

If the model is linear

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{F}\boldsymbol{x} + \boldsymbol{G}\boldsymbol{u} \\ \boldsymbol{y} &= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{J}\boldsymbol{u} \end{aligned}$$

then the filters are relatively straightforward but if the model is nonlinear significant then there can be significant difficulties.

Observer

This is another dynamical system driven by the input and output

$$\begin{aligned} \dot{z} &= g(z, u, y) \\ \hat{x} &= k(z, u, y) \end{aligned}$$

such that the error $\tilde{x}(t) = x(t) - \hat{x}(t)$ goes to zero as $t \longrightarrow \infty$.
For a linear system the construction of an observer is particularly simple $\hat{x} = z$

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + L(y - \hat{y}) \\ \hat{y} &= H\hat{x} + Ju \end{aligned}$$

Then

$$\dot{\tilde{x}} = (F - LH)\tilde{x}$$

so if

$$\sigma(F - LH) < 0$$

the error goes to zero. When can we find such an L ?

The linear system is observable if the largest $\boldsymbol{F}$ invariant subspace $\boldsymbol{\mathcal{V}}$ in the kernel of $\boldsymbol{H}$ is zero.

In other words the initial state is uniquely determined by the input and the time derivatives of the output. Of course we do not want to differentiate because it greatly accentuates any noise.

The linear system is observable iff the eigenvalues of $\boldsymbol{F} - \boldsymbol{L}\boldsymbol{H}$ can be set arbitraily up to cc by choice of $\boldsymbol{L}$.

If the dimension of $\boldsymbol{\mathcal{V}}$ is $\boldsymbol{r} > \boldsymbol{0}$ then $\boldsymbol{r}$ of the eigenvalues do not depend on $\boldsymbol{L}$. If these $\boldsymbol{r}$ eigenvalues are already in the open left half plane then the linear system is detectable and an asymptotically convergent observer can be constructed.

If not then it is impossible to construct an asymptotically convergent observer.

I have added some references to my slides. They are hardly complete.

A basic reference on linear control and estimation is
Author Kailath, Thomas.
Title Linear systems / Thomas Kailath.
Publisher Englewood Cliffs, N.J. : Prentice-Hall, c1980.
but it is almost too complete.

Simpler introductions are
Author Anderson, Brian D. O. and J. Moore
Title Optimal control–linear quadratic methods / Brian D.O. Anderson, John B. Moore.
Publisher Englewood Cliffs, N.J. : Prentice Hall, c1990.

Author Anderson, Brian D. O.
Title Optimal filtering / Brian D. O. Anderson, John B. Moore.
Publisher Englewood Cliffs, N.J. : Prentice-Hall, c1979.

Author Rugh, Wilson J
Title Linear system theory / Wilson J. Rugh
Edition 2nd ed
Publisher Englewood Cliffs, N.J. : Prentice Hall, 1996

A good reference on nonlinear systems is
Author Khalil, Hassan K., 1950-
Title Nonlinear systems / Hassan K. Khalil
Edition 3rd ed
Publisher Upper Saddle River, N.J. : Prentice Hall, c2002

Observers for linear systems were initiated in
D. G. Luenberger, *Observing the state of a linear system*, IEEE Trans. on Military Electronics, **8** (1964), 74-80.

Kalman Filtering

Assume that the linear system is affected by independent standard white Gaussian driving and observation noises and the initial condition is Gaussian ,

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{F}\boldsymbol{x} + \boldsymbol{G}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{w} \\
\boldsymbol{y} &= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{J}\boldsymbol{u} + \boldsymbol{D}\boldsymbol{v} \\
\boldsymbol{x}(0) &= \boldsymbol{x}^0
\end{aligned}
$$

where $\boldsymbol{D}$ is invertible and the mean and covariance of $\boldsymbol{x}^{\mathbf{0}}$ are $\hat{\boldsymbol{x}}^{\mathbf{0}},\ \boldsymbol{P}^{\mathbf{0}}$.

The goal is to compute the conditional expectation $\hat{\boldsymbol{x}}(\boldsymbol{t})$ conditioned on the past inputs and outputs $\boldsymbol{u}(\mathbf{0}:\boldsymbol{t}),\ \boldsymbol{y}(\mathbf{0}:\boldsymbol{t})$

Note the system could be time varying $\boldsymbol{F}=\boldsymbol{F}(t),\ \boldsymbol{G}=\boldsymbol{G}(t),$ $\boldsymbol{B}=\boldsymbol{B}(t), \boldsymbol{H}=\boldsymbol{H}(t),\ \boldsymbol{J}=\boldsymbol{J}(t),\ \boldsymbol{D}=\boldsymbol{D}(t)$.

Kalman Filter

$$
\begin{aligned}
\dot{\hat{x}} &= F\hat{x} + Gu + L(y - \hat{y}) \\
\hat{y} &= H\hat{x} + Ju \\
\hat{x}(0) &= \hat{x}^0 \\
L &= PH'(DD')^{-1} \\
\dot{P} &= FP + PF' + BB' - PH'(DD')^{-1}HP \\
P(0) &= P^0
\end{aligned}
$$

The equation for P is called a Riccati differential equations.

For autonomous systems there is also a asymptotic version (large t) of the Kalman filter which is also autonomous $\dot{P} = 0$.

A readable introduction to linear and nonlinear estimation is

Author Analytic Sciences Corporation. Technical Staff.
Title Applied optimal estimation. Written by: Technical Staff, Analytic Sciences Corporation.
Edited by Arthur Gelb. Principal authors: Arthur Gelb [and others]
Publisher Cambridge, Mass., M.I.T. Press [1974]

Least Squares or Minimum Energy Filtering

For fixed $\boldsymbol{u(0:t)}$ the model describes a mapping

$$(\boldsymbol{x^0},\ \boldsymbol{w(0:t)},\ \boldsymbol{v(0:t)}) \mapsto \boldsymbol{y(0:t)}$$

where $\boldsymbol{w(0:t)},\ \boldsymbol{v(0:t)}$ are $\boldsymbol{L^2[0,t]}$.

Given $\boldsymbol{y(0:t)}$ we seek the noise triple $(\boldsymbol{x^0},\ \boldsymbol{w(0:t)},\ \boldsymbol{v(0:t)})$ of minimum "energy" that generates it. The "energy" is defined to be

$$\frac{1}{2}(\boldsymbol{x^0})'(\boldsymbol{P^0})^{-1}\boldsymbol{x^0} + \frac{1}{2}\int_0^t |\boldsymbol{w(s)}|^2 + |\boldsymbol{v(s)}|^2\, \boldsymbol{ds}$$

The least squares estimate $\boldsymbol{\hat{x}(t)}$ is the endpoint of the trajectory generated by the minimizing $(\boldsymbol{\hat{x}^0},\ \boldsymbol{\hat{w}(0:t)})$. This estimate is the same as that of the Kalman filter.

Minimax Estimation or $\boldsymbol{H^\infty}$ Estimation

For fixed $\boldsymbol{u(0:t)}$ consider the mapping from noise to estimation error

$$\boldsymbol{(x^0,\ w(0:t),\ v(0:t)) \mapsto \tilde{x}(0:t)}$$

One seeks the estimator that minimizes the $\boldsymbol{L^2}$ induced norm of this mapping.

It is difficult to solve this problem directly so usually one sets an upper bound $\boldsymbol{\gamma}$ for induced norm and seeks an estimator that achieves it. As with the Kalman filter this reduces to finding a nonnegative definite solution to a Riccati differential equation. The form of the minimax estimator is similar to that of a Kalman filter.

If such an estimator can be found then $\boldsymbol{\gamma}$ can be lowered, if not $\boldsymbol{\gamma}$ is increased.

There are generalizations of the above to nonlinear systems of the form

$$\begin{aligned} \dot{x} &= f(x, u) \\ y &= h(x, u) \end{aligned}$$

A popular way of finding a nonlinear observer is to seek a local change of state coordinates $z = \theta(x)$ and an input-output injection $\beta(u, y)$ such that the system becomes

$$\dot{z} = Az + \beta(u, y)$$

where $\sigma(A) < 0$ for then the observer

$$\begin{aligned} \dot{\hat{z}} &= A\hat{z} + \beta(u, y) \\ \hat{x} &= \theta^{-1}(\hat{z}) \end{aligned}$$

has linear and exponentially stable error dynamics

$$\dot{\tilde{z}} = A\tilde{z}$$

in the transformed coordinates. This is relatively easy to accomplish if there is no input $m = 0$ but generally impossible if $m > 0$.

Using change of state coordinates and input/output injection was initiated in

A. J. Krener and A. Isidori,
Linearization by output injection and nonlinear observers,
Systems Control Lett., 3 (1983), pp. 47–52.

The method was improved in

A. J. Krener and W. Respondek,
Nonlinear observers with linearizable error dynamics,
SIAM J. Control Optim., 23 (1985), pp. 197–216.

and futher improved in

N. Kazantzis and C. Kravaris, *Nonlinear observer design using Lyapunov's auxiliary theorem*, Systems Control Lett., 34 (1998), pp. 241–247.

Other methods of constructing observers can be found in
E. A. Misawa, J. K. Hedrick, *Nonlinear observers a state of the art survey*, Trans. of ASME, J. of Dynamic Systems, Measurement and Control, **111** (1989), 344-352.

See also
Title New directions in nonlinear observer design
H. Nijmeijer and T.I. Fossen (eds.)
Publisher London ; New York : Springer, c1999

Nonlinear Stochastic Filtering

We add driving and observation noises to the nonlinear model which now must be written as a pair of Ito stochastic differential equations

$$
\begin{aligned}
dx &= f(x,u)dt + Bdw \\
dy &= h(x,u)dt + dv \\
x(0) &= x^0
\end{aligned}
$$

Let $p^0(x)$ be the known density of x^0 .
The unnormalied conditional density $q(x,t)$ of $x(t)$ given the model and past inputs and outputs satisfies the Zakai stochastic partial differential equation (summation convention)

$$
\begin{aligned}
dq &= -\frac{\partial q f_i}{\partial x_i}dt + \frac{\partial^2 q}{\partial x_i \partial x_j}(BB')_{ij}dt + qh_i dy_i \\
q(x,0) &= p^0(x)
\end{aligned}
$$

The Zakai equation is stochastic PDE in the Ito sense driven by the observations. It has to be solved in its Stratonovich form.

It is very difficult to solve if $n > 1$. It can't be solved implicitly and it is parabolic so if the spatial step is small, the temporal step is extremely small.

The function $x \mapsto q(x,t)$ can be thought of as the state at time t of an infinite dimensional observer with inputs $u,\ y$ and output

$$\hat{x}(t) = \frac{\int xq(x,t)\ dx}{\int q(x,t)\ dx}$$

or

$$\hat{x}(t) = \mathbf{argmax}_x\ q(x,t)$$

Many nonlinear estimators are infinite dimensional and one is forced to seek a finite dimensional approximation.

A relatively simple derivation of the Zakai equation can be found in

M. H. A. Davis and S. I. Marcus,
An introduction to nonlinear filtering,
in Stochastic Systems: The Mathematics of Filtering and Identification and Applications,
M. Hazewinkel and J. C. Willems, (eds.),
D. Reidel Publishing, Dordrecht, (1981), 53-76.

Monte Carlo Filtering, Particle Filtering

These are discrete time filters based on the approximation of $\boldsymbol{p(x,t)}$ by point masses.

$$\boldsymbol{p(x,t) \approx \sum \alpha_k(t)\delta(x - x^k(t))}$$

Sample $\boldsymbol{p(x,t)}$, use the noisy system and the Bayes formula to compute $\boldsymbol{\alpha_k(t+1),\ x^k(t+1)}$

There are several different implimentations of this basic philosophy including replacing the point masses with Gaussians.

Nonlinear Minimum Energy Filtering

Some of the techical difficulties associated with stochastic nonlinear models can be avoided by assuming the noises are unknown L^2 functions and the initial condition is also unknown

$$\begin{aligned}\dot{x} &= f(x,u) + Bw \\ y &= h(x,u) + v \\ x(0) &= x^0.\end{aligned}$$

Consistent with $u(0:t),\ y(0:t)$ we seek to minimize

$$\frac{1}{2}|x^0|^2 + \frac{1}{2}\int_0^t |w(s)|^2 + |v(s)|^2\, ds$$

The least squares estimate $\hat{x}(t)$ is the endpoint of the trajectory generated by the minimizing $(\hat{x}^0,\ \hat{w}(0:t))$. This estimate is generally not the same as that of the stochastic nonlinear filter.

To find this estimate we must solve in real time a partial differential equation of Hamilton-Jacobi-Bellman type driven by $u(t),\ y(t)$.

$$\frac{\partial Q}{\partial t} = -\frac{\partial Q}{\partial x_i} f_i - \frac{1}{2}\frac{\partial Q}{\partial x_i}\frac{\partial Q}{\partial x_i} + \frac{1}{2}(y_i - h_i)(y_i - h_i)$$

$$\hat{x}(t) = \operatorname{argmin}_x Q(x, t)$$

This is nearly impossible if $n > 1$.

There is no guarantee that a smooth solution exists so we must allow solutions in the viscosity sense. We have repaced the technical difficulties of stochastic nonlinear filtering with a different set of technical difficulties.

Similar remarks hold for nonlinear minimax filtering except now the PDE is of Hamilton-Jacobi-Isaacs type so it is even harder.

For these reasons most nonlinear filtering is done by approximation. The workhorse is the extended Kalman filter.

Nonlinear Minimum Energy Estimation was initiated in

Mortenson RE (1968) J. Optimization Theory and Applications, 2:386–394

See also

Hijab O (1980) Minimum Energy Estimation PhD Thesis, University of California, Berkeley, California

Hijab O (1984), Annals of Probability, 12:890–902

Extended Kalman Filtering (EKF)

Add WGNs to the nonlinear model

$$\begin{aligned} \dot{x} &= f(x,u) + Bw \\ y &= h(x,u) + Dv \end{aligned}$$

and linearize around the estimated trajectory.

$$F(t) = \frac{\partial f}{\partial x}(\hat{x}(t), u(t)), \quad H(t) = \frac{\partial h}{\partial x}(\hat{x}(t), u(t))$$

Then build a Kalman filter for the linear model and impliment it on the nonlinear model.

$$\begin{aligned} \dot{\hat{x}} &= f(\hat{x},u) - L(t)(y - h(\hat{x},u)) \\ L(t) &= -P(t)H(t)'(DD')^{-1} \\ \dot{P} &= F(t)P(t) + P(t)F(t)' + BB' \\ &\quad -P(t)H(t)'(DD')^{-1}H(t)P(t) \end{aligned}$$

Interpertation

$$\boldsymbol{x}(\boldsymbol{t}) \approx \boldsymbol{N}(\hat{\boldsymbol{x}}(\boldsymbol{t}), \boldsymbol{P}(\boldsymbol{t}))$$

The EKF is the most widely used nonlinear filter.

Like the Kalman filter it can also be derived nonstochastically from the minimum energy filter.

It generally performs well but can diverge.

Recently it has shown that when viewed as an observer it is locally convergent for a broad class of nonlinear systems in both continuous and discrete time.

The local covergence of the EKF in discrete time is demonstrated in

Song and Grizzle, Proceedings of the American Control Conference, 1992, pp. 3365-3369.

and in continuous time in Krener AJ (2002) The convergence of the extended Kalman filter.
In: Rantzer A, Byrnes CI (eds)
Directions in Mathematical Systems Theory and Optimization, 173–182.
Springer, Berlin Heidelberg New York,
also at `http://arxiv.org/abs/math.OC/0212255`

Unscented Kalman Filter

Like the particle filters this is a discrete time filter.

At each time it assumes that the conditional density is approximately $N(\hat{x}(t), P(t))$

Loosely speaking one computes an orthonormal frame that is $P^{\frac{1}{2}}(t)$ and propagates the noisy dynamics from $\hat{x}(t)$ and from $\hat{x}(t)$ plus and minus a positive multiple of the vectors of the frame. This deterministic sample of $2n+1$ points is updated by the current observation using Bayes rule and a mean and covariance $\hat{x}(t+1),\ P(t+1)$ is computed.

The computational burden is about the same of the EKF but the UKF is second order accurate rather than first order accurate. It has the same drawback of the EKF, namely that it is a local method in that it approximates the conditional density by a Gaussian.

The UKF was introduced in

Julier et al. Proceedings of the American Control Conference, 1995, pp. 1628-1632

Smoothing of Linear Initial Value Models

$$
\begin{aligned}
\dot{x} &= Fx + Gu + Bw \\
y &= Hx + Ju + Dv \\
x(0) &\approx N(\hat{x}^0, P^0) \\
t &\in [0, T]
\end{aligned}
$$

Estimate $x(0:T)$ from the model and $u(0:t),\ y(0:T)$.

Two Kalman Filters

$$\begin{aligned}
\dot{\hat{x}}_f &= F\hat{x}_f + Gu + L_f(y - \hat{y}_f) \\
\hat{y}_f &= H\hat{x}_f + Ju \\
\hat{x}_f(0) &= \hat{x}^0 \\
L_f &= P_f H'(DD')^{-1} \\
\dot{P}_f &= FP_f + P_f F' + BB' - P_f H'(DD')^{-1} H P_f \\
P_f(0) &= P^0
\end{aligned}$$

$$\begin{aligned}
\dot{\hat{x}}_b &= F\hat{x}_b + Gu - L_b(y - \hat{y}_b) \\
\hat{y}_b &= H\hat{x}_b + Ju \\
\hat{x}_b(T) &= ? \\
L_b &= P_b H'(DD')^{-1} \\
\dot{P}_b &= FP_b + P_b F' - BB' + P_b H'(DD')^{-1} H P_b \\
P_b(T) &= ?
\end{aligned}$$

The two technical problems can be solved by letting

$$Q(t) = P_b^{-1}(t), \qquad z(t) = Q(t)\hat{x}_b(t)$$

then

$$\begin{aligned}
\dot{z} &= (-F' + QBB')z + (QG + H'(DD')^{-1}J)u \\
&\quad -H'(DD')^{-1}y \\
z(T) &= 0 \\
\dot{Q} &= -QF - F'Q + QBB'Q - H'(DD')^{-1}H \\
Q(T) &= 0
\end{aligned}$$

Then the smoothed estimate $\hat{x}(t)$ and its error covariance $P(t)$ are given by combining the independent estimates

$$\begin{aligned}
\hat{x}(t) &= P(t)\left(P_f^{-1}(t)\hat{x}_f(t) + P_b^{-1}(t)\hat{x}_b(t)\right) \\
&= P(t)\left(P_f^{-1}(t)\hat{x}_f(t) + z(t)\right) \\
P^{-1}(t) &= P_f^{-1}(t) + P_b^{-1}(t) \\
&= P_f^{-1}(t) + Q(t)
\end{aligned}$$

The same smoothed estimate can be obtained by a least squares argument. Suppose $w(0:T),\ v(0:t)$ are unknown $L^2[0,T]$ functions and x^0 is an unkown initial condition. We seek to minimize

$$(x^0)'(P^0)^{-1}x^0 + \int_0^T |w|^2 + |v|^2\, dt$$

consistent with $u(0:T),\ y(0:T)$ and the model

$$\begin{aligned} \dot{x} &= Fx + Gu + Bw \\ y &= Hx + Ju + Dv \\ x(0) &= x^0. \end{aligned}$$

Let $\hat{w}(0:T),\ \hat{v}(0:t),\ \hat{x}^0$ be the minimizer then the smoothed estimate $\hat{x}(t)$ is the solution of

$$\begin{aligned} \dot{\hat{x}} &= F\hat{x} + Gu + B\hat{w} \\ \hat{x}(0) &= \hat{x}^0. \end{aligned}$$

For fixed $u(0:T),\ y(0:T)$ define the Hamiltonian

$$\mathcal{H}(\lambda, x, w) = \lambda' \left(Fx + Gu + Bw\right) + \frac{1}{2}\left(|w|^2 + |v|^2\right)$$
$$v = D^{-1}\left(y - Hx - Ju\right)$$

If $\hat{x}(t),\ \hat{w}(0:T),\ \hat{v}(0:t),\ \hat{x}^0$ are minimizing then there exist $\hat{\lambda}(t)$ such that the Pontryagin Minimum Principle holds.

$$\dot{\hat{x}} = \left(\frac{\partial \mathcal{H}}{\partial \lambda}\right)' (\hat{\lambda}, \hat{x}, \hat{w})$$
$$\dot{\hat{\lambda}} = -\left(\frac{\partial \mathcal{H}}{\partial x}\right)' (\hat{\lambda}, \hat{x}, \hat{w})$$
$$\hat{w} = \operatorname{argmin}_w \mathcal{H}(\lambda, x, w)$$
$$\hat{x}(0) = P^0 \hat{\lambda}(0)$$
$$0 = \hat{\lambda}(T)$$

This yields the two point boundary value problem

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{\lambda}} \end{bmatrix} = \begin{bmatrix} F & -BB' \\ -H'(DD')^{-1}H & -F' \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{\lambda} \end{bmatrix} + \begin{bmatrix} G & 0 \\ -H'(DD')^{-1}J & H'(DD')^{-1} \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix}$$

$$\hat{x}(0) = P^0\hat{\lambda}(0)$$

$$0 = \hat{\lambda}(T)$$

There are numerous ways of solving this problem, perhaps the simplest is to define

$$\hat{\mu}(t) = \hat{\lambda}(t) - M(t)\hat{x}(t)$$

where

$$\begin{aligned} \dot{M} &= -MF - F'M - H'(DD')^{-1}H + MBB'M \\ M(T) &= 0 \end{aligned}$$

This transformation triangularizes the dynamics

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{\mu}} \end{bmatrix} = \begin{bmatrix} F - MBB' & -BB' \\ 0 & -F' + BB'M \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{\mu} \end{bmatrix} + \begin{bmatrix} G & 0 \\ -MG - H'(DD')^{-1}J & H'(DD')^{-1} \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix}$$

$$\begin{aligned} \left(I + P^0 M(0)\right) \hat{x}(0) &= P^0 \hat{\mu}(0) \\ 0 &= \hat{\mu}(T) \end{aligned}$$

The triangularized dynamics can be solved by integrating $\hat{\mu}(t)$ backward from $\hat{\mu}(T)$ and then integrating $\hat{x}(t)$ forward from $\hat{x}(0)$

Stochastic Interpertation of the Variational Equations

(For simplicity $\boldsymbol{G} = \mathbf{0},\ \boldsymbol{J} = \mathbf{0}$)

The model

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{F}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{w} \\ \boldsymbol{y} &= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{v} \\ \boldsymbol{x}(\mathbf{0}) &= \boldsymbol{x}^{\mathbf{0}}. \end{aligned}$$

defines a linear map

$$\boldsymbol{\mathcal{T}}_{\mathbf{1}} : \left(\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{w}(\mathbf{0}:\boldsymbol{T}), \boldsymbol{v}(\mathbf{0}:\boldsymbol{T})\right) \mapsto \boldsymbol{y}(\mathbf{0}:\boldsymbol{T})$$

The complementary model

$$
\begin{aligned}
\dot{\lambda} &= -F'\lambda + H'(D')^{-1}v \\
\psi &= -B'\lambda + w \\
\lambda(0) &= 0 \\
\xi &= x^0 - P^0\lambda(0)
\end{aligned}
$$

defines a map

$$\mathcal{T}_2 : \left(x^0, w(0:T), v(0:T)\right) \mapsto (\xi, \psi(0:T))$$

such that the combined map

$$\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2 : \left(x^0, w(0:T), v(0:T)\right) \mapsto (y(0:T), \xi, \psi(0:T))$$

is invertible and the ranges of $\mathcal{T}_1$, $\mathcal{T}_2$ are independent

$$E(y(t)\xi') = 0, \qquad E(y(t)\psi'(s)) = 0$$

Hence

$$\left(\hat{x}^0, \hat{w}(0:T), \hat{v}(0:T)\right) = \mathcal{T}^{-1}(y(0:T), 0, 0)$$

For more on linear smoothing, boudary value models and complementary models see

Weinert H L (2001) Fixed interval smoothing for state space models.
Kluwer Academic Publishers, Norwell MA

and its extensive references.

In filtering it is reasonable to assume that there is apriori information about the initial condition but no apriori information about the terminal condition. But this is not reasonable in smoothing.

A better model for the smoothing problem is the two point boundary value problem

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{w} \\
\boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{v} \\
\boldsymbol{b} &= \boldsymbol{V}^{\boldsymbol{0}}\boldsymbol{x}(0) + \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{x}(\boldsymbol{T})
\end{aligned}
$$

where as before $\boldsymbol{w}$, $\boldsymbol{v}$ are standard white Gaussian noises and $\boldsymbol{b}$ is an independent Gaussian vector.

What is the meaning of such a model? What kind of process is $\boldsymbol{x}(\boldsymbol{t})$?

More generally we might consider multipoint value processes,

$$
\boldsymbol{b} = \boldsymbol{V}^{\boldsymbol{0}}\boldsymbol{x}(\boldsymbol{t}_0) + \boldsymbol{V}^{\boldsymbol{1}}\boldsymbol{x}(\boldsymbol{t}_1) + \cdots + \boldsymbol{V}^{\boldsymbol{k}}\boldsymbol{x}(\boldsymbol{t}_{\boldsymbol{k}})
$$

The two point boundary process is well-posed if

$$\boldsymbol{W} = \boldsymbol{V^0} + \boldsymbol{V^T}\boldsymbol{\Phi}(\boldsymbol{T},\mathbf{0})$$

is nonsingular where

$$\begin{aligned} \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{t}}(\boldsymbol{t},\boldsymbol{s}) &= \boldsymbol{A}(\boldsymbol{t})\boldsymbol{\Phi}(\boldsymbol{t},\boldsymbol{s}) \\ \boldsymbol{\Phi}(\boldsymbol{s},\boldsymbol{s}) &= \boldsymbol{I} \end{aligned}$$

Then there is a Green's matrix

$$\boldsymbol{\Gamma}(\boldsymbol{t},\boldsymbol{s}) = \begin{cases} \boldsymbol{\Phi}(\boldsymbol{t},\mathbf{0})\boldsymbol{W}^{-1}\boldsymbol{\Phi}(\mathbf{0},\boldsymbol{s}) & \boldsymbol{t} > \boldsymbol{s} \\ -\boldsymbol{\Phi}(\boldsymbol{t},\mathbf{0})\boldsymbol{W}^{-1}\boldsymbol{\Phi}(\boldsymbol{T},\boldsymbol{s}) & \boldsymbol{t} < \boldsymbol{s} \end{cases}$$

and

$$\boldsymbol{x}(\boldsymbol{t}) = \boldsymbol{\Phi}(\boldsymbol{t},\boldsymbol{s})\boldsymbol{W}^{-1}\boldsymbol{b} + \int_{\mathbf{0}}^{\boldsymbol{T}} \boldsymbol{\Gamma}(\boldsymbol{t},\boldsymbol{s})\boldsymbol{B}(\boldsymbol{s})\boldsymbol{w}(\boldsymbol{s})\, \boldsymbol{ds}$$

where the integral is in the Weiner sense (integration by parts).

This in general is not a Markov process. It is a reciprocal process in the sense of S. Bernstein.

A process on $[0,T]$ is reciprocal if conditioned on $x(t_1), x(t_2)$ where $0 \leq t_1 < t_2 \leq T$, what happens on (t_1, t_2) is independent of what happens on $[0, t_1) \cup (t_2, T]$.

In other words a reciprocal process is a Markov random field on the interval $[0,T]$.

Every Markov process is reciprocal but not vice versa.

The covariance $R(t,s)$ of a reciprocal process with full noise

$$Q(t) = \frac{\partial R}{\partial t}(t, t^+) - \frac{\partial R}{\partial t}(t^+, t) > 0$$

satisfies a second order, self-adjoint differential equation

$$\frac{\partial^2 R}{\partial t^2}(t,s) = F(t)R(t,s) + G(t)\frac{\partial R}{\partial t}(t,s) - Q(t)\delta(t-s)$$

If $Q(t)$ is not positive definite then $R(t,s)$ satisfies a higher order, self-adjoint differential equation.

For $\boldsymbol{x}$ defined by the linear boundary value model above

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{B}\boldsymbol{B}' \\ \boldsymbol{G}\boldsymbol{Q} &= \boldsymbol{A}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{A}' + \frac{d\boldsymbol{Q}}{dt} \\ \boldsymbol{F} &= \frac{d\boldsymbol{A}}{dt} + \boldsymbol{A}^2 - \boldsymbol{G}\boldsymbol{A} \end{aligned}$$

Any reciprocal process with full noise is the solution of a second order, stochastic boundary value problem

$$-d^2\boldsymbol{x}(t) + \boldsymbol{F}(t)\boldsymbol{x}(t)\;dt^2 + \boldsymbol{G}(t)d\boldsymbol{x}(t)\;dt = \boldsymbol{Q}(t)\boldsymbol{\xi}(t)dt^2$$

$$\boldsymbol{x}(0) = \boldsymbol{x}^0, \quad \boldsymbol{x}(T) = \boldsymbol{x}^T$$

To leading order $\boldsymbol{\xi}(t)dt^2 = d^2\boldsymbol{w}(t) + \ldots$ where $\boldsymbol{w}(t)$ is a standard Wiener process.

This is not an Ito stochastic differential equation!

The quick and dirty way to derive this equation is to note that if

$$\boldsymbol{d}^{+}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}\,\boldsymbol{dt} + \boldsymbol{B}\,\boldsymbol{d}^{+}\boldsymbol{w}$$

where $\boldsymbol{d}^{+}$ denotes the forward differential in the Ito sense then

$$\boldsymbol{B}^{-1}\left(\boldsymbol{d}^{+}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{x}\,\boldsymbol{dt}\right) = \boldsymbol{d}^{+}\boldsymbol{w}$$

We apply the adjoint of the operator on the left to obtain

$$\begin{aligned}\left(-\boldsymbol{d}^{-} + \boldsymbol{A}'\,\boldsymbol{dt}\right)(\boldsymbol{B}')^{-1}\boldsymbol{B}^{-1}\left(\boldsymbol{d}^{+}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{x}\,\boldsymbol{dt}\right)\\ = \left(-\boldsymbol{d}^{-} + \boldsymbol{A}'\,\boldsymbol{dt}\right)(\boldsymbol{B}')^{-1}\boldsymbol{d}^{+}\boldsymbol{w}\end{aligned}$$

which yields

$$-\boldsymbol{d}^{2}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{x}\,\boldsymbol{dt}^{2} + \boldsymbol{G}\boldsymbol{dx}\,\boldsymbol{dt} = \boldsymbol{Q}\boldsymbol{d}^{2}\boldsymbol{w} + \ldots$$

For more on reciprocal processes see

Krener A J, Frezza R, Levy B C (1991)
Gaussian reciprocal processes and self-adjoint stochastic differential equations of second order.
Stochastics 34:29-56.

Krener, A. J., Reciprocal diffusions in flat space,
Probability Theory and Related Fields (1997) pp. 243-281

Suppose that we have observations of the reciprocal process

$$\boldsymbol{dy = Cxdt + Ddv}$$

where $\boldsymbol{v}$ is a standard Wiener process and $\boldsymbol{D}$ is invertible.

If the boundary conditions are known to be zero $\boldsymbol{x^0 = x^T = 0}$ then the optimal smoothed estimate $\boldsymbol{\hat{x}(t)}$ satisfies

$$\boldsymbol{-d^2\hat{x}(t) + F(t)\hat{x}(t)\ dt^2 + G(t)d\hat{x}(t)\ dt}$$

$$\boldsymbol{= C'(t)(DD')^{-1}(t)\left(dy\ dt - C(t)\hat{x}(t)\ dt^2\right)}$$

$$\boldsymbol{\hat{x}(0) = 0, \quad \hat{x}(T) = 0}$$

The formula is a little more complicated if $\boldsymbol{x^0 \neq 0,\ x^T \neq 0}$

The error process $\boldsymbol{\tilde{x}(t) = x(t) - \hat{x}(t)}$ is also reciprocal.

The reciprocal smoother can be found in

Frezza R (1990)
Models of higher order and mixed order Gaussian reciprocal processes with applications to the smoothing problem.
PhD Thesis, University of California, Davis, CA

Deterministic Smoothing of Linear Boundary Processes

$$\begin{aligned} \dot{x} &= Fx + Gu + Bw \\ y &= Hx + Ju + Dv \\ b &= V^0 x(0) + V^T x(T) \end{aligned}$$

As before we view $w(0:T), v(:T)$ as unknown L^2 functions and $b \in \mathbb{R}^k$ as an unknown boundary condition.

For given $u(0:T),\ y(0:T)$ we seek to minimize

$$\frac{1}{2}\left(|b|^2 + \int_0^T |w|^2 + |v|^2\, dt\right)$$

Define the Hamiltonian

$$\mathcal{H} = \lambda' \left(Fx + Gu + Bw\right) + \frac{1}{2}\left(|w|^2 + |y - (Hx + Ju + Dv)|^2\right)$$

If $\hat{\boldsymbol{x}}(\boldsymbol{t}),\ \hat{\boldsymbol{w}}(\boldsymbol{0}:\boldsymbol{T}),\ \hat{\boldsymbol{v}}(\boldsymbol{0}:\boldsymbol{t}),\ \hat{\boldsymbol{b}}$ are minimizing then there exist $\hat{\boldsymbol{\lambda}}(\boldsymbol{t})$ such that the Pontryagin Minimum Principle holds.

$$\dot{\hat{x}} = \left(\frac{\partial\mathcal{H}}{\partial\lambda}\right)'(\hat{\lambda}, \hat{x}, u, y, \hat{w})$$

$$\dot{\hat{\lambda}} = -\left(\frac{\partial\mathcal{H}}{\partial x}\right)'(\hat{\lambda}, \hat{x}, u, y, \hat{w})$$

$$\hat{w} = \mathrm{argmin}_w \mathcal{H}(\hat{\lambda}, \hat{x}, u, y, w)$$

$$\hat{b} = V^0 \hat{x}(0) + V^T \hat{x}(T)$$

$$\hat{\lambda}(0) = (V^0)'\hat{b}$$

$$\hat{\lambda}(T) = (V^T)'\hat{b}$$

This yields the two point boundary value problem

$$\begin{bmatrix} \dot{\hat{\boldsymbol{x}}} \\ \dot{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F} & -\boldsymbol{B}\boldsymbol{B}' \\ -\boldsymbol{H}'(\boldsymbol{D}\boldsymbol{D}')^{-1}\boldsymbol{H} & -\boldsymbol{F}' \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ -\boldsymbol{H}'(\boldsymbol{D}\boldsymbol{D}')^{-1}\boldsymbol{J} & \boldsymbol{H}'(\boldsymbol{D}\boldsymbol{D}')^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{bmatrix}$$

$$\hat{\boldsymbol{\lambda}}(0) = (\boldsymbol{V}^0)' \left(\boldsymbol{V}^0 \hat{\boldsymbol{x}}(0) + \boldsymbol{V}^T \hat{\boldsymbol{x}}(\boldsymbol{T})\right)$$

$$\hat{\boldsymbol{\lambda}}(\boldsymbol{T}) = (\boldsymbol{V}^T)' \left(\boldsymbol{V}^0 \hat{\boldsymbol{x}}(0) + \boldsymbol{V}^T \hat{\boldsymbol{x}}(\boldsymbol{T})\right)$$

In contrast to the stochastic approach, we do not have to assume that the model is well-posed to pursue the deterministic approach.

Under reasonable assumptions (controllability and observability) the variational equations will be well-posed even if the model is not.

In the initial value formulation, one of the boundary conditions of the variational equations was particularly simple

$$\boldsymbol{\lambda}(\boldsymbol{T}) = \mathbf{0}$$

This allowed us to triangularize the dynamics and solve by backward sweep of $\hat{\boldsymbol{\lambda}}(\boldsymbol{t})$ and a forward sweep of $\hat{\boldsymbol{x}}(\boldsymbol{t})$.

In the general boundary value formulation, the boundary conditions of the variational equations are fully coupled so we cannot solve by two simple sweeps. This is a big problem if the dimension $\boldsymbol{n}$ of $\boldsymbol{x}$ is large.

Another advantage of the deterministic approach is that it readily generalizes to nonlinear boundary value models of the form

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{g}(\boldsymbol{t}, \boldsymbol{x})\boldsymbol{w} \\
\boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{v} \\
\boldsymbol{b} &= \boldsymbol{\beta}(\boldsymbol{x}(\mathbf{0}), \boldsymbol{x}(\boldsymbol{T}))
\end{aligned}
$$

If we try to formulate this stochastically, e.g. $\boldsymbol{w}, \ \boldsymbol{v}$ independent white Gaussian noises and $\boldsymbol{b}$ an independent random $\boldsymbol{k}$ vector, we cannot give meaning to the equations even in an Ito sense.

But for fixed $\boldsymbol{u}(\mathbf{0}:\boldsymbol{T}), \ \boldsymbol{y}(\mathbf{0}:\boldsymbol{T})$ we can seek to minimize

$$
\frac{\mathbf{1}}{\mathbf{2}}\left(|\boldsymbol{b}|^{\mathbf{2}} + \int_{\mathbf{0}}^{\boldsymbol{T}} |\boldsymbol{w}|^{\mathbf{2}} + |\boldsymbol{v}|^{\mathbf{2}} \, \boldsymbol{dt}\right)
$$

The Pontryagin minimum principle yields the first order necessary that must be satisfied by an optimal solution. The Hamiltonian is

$$
\begin{aligned}
\boldsymbol{\mathcal{H}} &= \boldsymbol{\lambda}' \left(\boldsymbol{f}(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{u}(\boldsymbol{t})) + \boldsymbol{g}(\boldsymbol{t}, \boldsymbol{x})\boldsymbol{w}\right) + \frac{1}{2}|\boldsymbol{w}|^2(\boldsymbol{t}) \\
&\quad + \frac{1}{2}|\boldsymbol{y}(\boldsymbol{t}) - \boldsymbol{h}(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{u}(\boldsymbol{t})|^2.
\end{aligned}
$$

and

$$
\begin{aligned}
\dot{\hat{\boldsymbol{x}}}(\boldsymbol{t}) &= \left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\lambda}}(\boldsymbol{t}, \hat{\boldsymbol{\lambda}}(\boldsymbol{t}), \hat{\boldsymbol{x}}(\boldsymbol{t}), \boldsymbol{u}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}), \hat{\boldsymbol{w}}(\boldsymbol{t}))\right)' \\
\dot{\hat{\boldsymbol{\lambda}}}(\boldsymbol{t}) &= -\left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{x}}(\boldsymbol{t}, \hat{\boldsymbol{\lambda}}(\boldsymbol{t}), \hat{\boldsymbol{x}}(\boldsymbol{t}), \boldsymbol{u}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}), \hat{\boldsymbol{w}}(\boldsymbol{t}))\right)' \\
\hat{\boldsymbol{w}}(\boldsymbol{t}) &= \mathbf{argmin}_{\boldsymbol{w}} \boldsymbol{H}(\boldsymbol{t}, \hat{\boldsymbol{\lambda}}(\boldsymbol{t}), \hat{\boldsymbol{x}}(\boldsymbol{t}), \boldsymbol{u}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}), \boldsymbol{w}) \\
\hat{\boldsymbol{\lambda}}(\mathbf{0}) &= \left(\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{x}^{\mathbf{0}}}(\hat{\boldsymbol{x}}(\mathbf{0}), \hat{\boldsymbol{x}}(\boldsymbol{T}))\right)' \boldsymbol{b}(\hat{\boldsymbol{x}}(\mathbf{0}), \hat{\boldsymbol{x}}(\boldsymbol{T})) \\
\hat{\boldsymbol{\lambda}}(\boldsymbol{T}) &= \left(\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{x}^{\boldsymbol{T}}}(\hat{\boldsymbol{x}}(\mathbf{0}), \hat{\boldsymbol{x}}(\boldsymbol{T}))\right)' \boldsymbol{b}(\hat{\boldsymbol{x}}(\mathbf{0}), \hat{\boldsymbol{x}}(\boldsymbol{T}))
\end{aligned}
$$

Then

$$\hat{w}(t) = -g'(t, \hat{x}(t))\hat{\lambda}(t)$$

and so

$$\begin{aligned}
\dot{\hat{x}}(t) &= f(t, \hat{x}(t), u(t)) - g(t, \hat{x}(t))g'(t, \hat{x}(t))\hat{\lambda}(t) \\
\dot{\hat{\lambda}}(t) &= -\left(\frac{\partial f}{\partial x}(t, \hat{x}(t), u(t)) - \frac{\partial g}{\partial x}(t, \hat{x}(t))g'(t, \hat{x}(t))\hat{\lambda}(t)\right)'\hat{\lambda}(t) \\
&\quad + \left(\frac{\partial h}{\partial x}(t, \hat{x}(t), u(t))\right)'(y(t) - h(t, \hat{x}(t), u(t))) \\
\hat{\lambda}(0) &= \left(\frac{\partial b}{\partial x^0}(\hat{x}(0), \hat{x}(T))\right)' b(\hat{x}(0), \hat{x}(T)) \\
\hat{\lambda}(T) &= \left(\frac{\partial b}{\partial x^T}(\hat{x}(0), \hat{x}(T))\right)' b(\hat{x}(0), \hat{x}(T)).
\end{aligned}$$

This is a nonlinear two point boundary value problem in $2n$ variables that we can try solve by direct methods, shooting methods or iterative methods. We shall use an iterative method that takes advantage of the variational nature of the problem.

We use gradient descent. First solve the initial value problem,

$$\begin{aligned} \dot{x} &= f(t,x,u) + g(t,x(t))w \\ x(0) &= x^0 \end{aligned}$$

and compute the cost

$$\pi(x^0, w(0:T)) = \frac{1}{2}\left(|b|^2 + \int_0^T |w|^2 + |v|^2 \, dt\right)$$

Then solve the final value problem for $\mu(s)$,

$$\begin{aligned} \frac{d}{ds}\mu(s) &= -F(s)\mu(s) + H'(s)\left(y(s) - H(s)x(s)\right) \\ \mu(T) &= \left(\frac{\partial b}{\partial x^T}(x(0),x(T))\right)' b(x(0),x(T)) \end{aligned}$$

$$\begin{aligned} F(t) &= \frac{\partial f}{\partial x}(t,x(t),u(t)) \\ G(t) &= g(t,x(t)) \\ H(t) &= \frac{\partial h}{\partial x}(t,x(t),u(t)) \end{aligned}$$

The first variation of the cost due to the changes $\delta x^0,\ \delta w(0:T)$ is

$$\begin{aligned}\delta\pi &= \pi(x^0, w) + \left(\beta' \frac{\partial b}{\partial x^0}(x(0), x(T)) + \mu'(0)\right)\delta x^0 \\ &\quad + \int_0^T \left(w'(s) + \mu'(s)G(s)\right)\delta w(s)\, ds \\ &\quad + O(\delta x^0, \delta w)^2.\end{aligned}$$

Choose a step size ϵ and define

$$\begin{aligned}\delta x^0 &= -\epsilon\left(\left(\frac{\partial b}{\partial x^0}(x(0), x(T))\right)'\beta + \mu(0)\right) \\ \delta w(s) &= -\epsilon\left(w(s) + G'(s)\mu(s)\right).\end{aligned}$$

Replace $x^0,\ w(0:T)$ by $x^0 + \delta x^0,\ w(0:T) + \delta w(0:T)$ and repeat.

The nonlinear smoother can be found in

Krener, A. J., Least Squares Smoothing of Nonlinear Systems, preprint, ajkrener@ucdavis.edu

Conclusions

We have briefly surveyed methods for filtering and smoothing data from dynamic models. For many applications the relevant problem is smoothing a nonlnear dynamic model that involves two point boundary data (or multipoint data).

Well-posed boundary value linear models driven by white Gussian noise lead to reciprocal processes.

A combination of boundary information and nonlinearities rule out a stochastic formulation of the smoothing problem. A deterministic (least squares) formulation leads to a two point boundary problem that is well-posed for controllable and observable linear models.

We have presented a gradient descent algorithm for solving linear and nonlinear problems. It will converge for linear models. Its convergence for nonlinear models is an open question.