1

Least Squares Smoothing of Nonlinear Systems

Arthur J. Krener *

Department of Mathematics, University of California, Davis, CA, 95616, USA, ajkrener@ucdavis.edu

Summary. We consider the fixed interval smoothing problem for data from linear or nonlinear models where there is a priori information about the boundary values of the state process. The nonlinearities and boundary values preclude a stochastic approach so instead we use a least squares methodology. The resulting variational equations are a coupled system of ordinary differential equations for the state and costate involving boundary conditions. If the model is linear and the a priori information is only about the initial state then several authors have given methods for solving the resulting equations in two sweeps. If the model is linear but the a priori information is about both the initial and final states then direct methods have been proposed. If the state dimension is large these methods can be very expensive and moreover they don't readily generalize to nonlinear models. Therefore we present an iterative method for solving both linear and nonlinear problems.
Key Words: Nonlinear smoothing, boundary value processes, least squares smoothing

1.1 Introduction

We consider the problem of smoothing a boundary value process, i.e. a nonlinear system of the form

$$\begin{aligned} \dot{x} &= f(t,x,u) \\ y &= h(t,x,u) \\ 0 &= b(x(0),x(T)) \end{aligned} \tag{1.1}$$

where $x(t) \in I\!R^n$ is the unknown state of the process, $u(t) \in I\!R^m$ is the known input or control, $y(t) \in I\!R^p$ is the known output or observation and the boundary condition (1.1) is k dimensional. The notation $x(0:T)$ denotes the curve $t \mapsto x(t)$, for $t \in [0,T]$.

* Research supported in part by NSF DMS-0204390.

The fixed interval smoothing problem [5] is to estimate the state $x(0:T)$ from the knowledge of the functions f, h, b and the inputs and outputs $u(0:T), y(0:T)$. This problem is called fixed interval smoothing because to estimate $x(t)$ at any $t \in [0,T]$, we use all the inputs $u(t)$ and outputs $y(t)$ on the interval $[0,T]$.

But the above formulation is different from the standard fixed interval smoothing problem in two ways. The first is that the model is allowed to be nonlinear. The second is that the a priori partial information that we have available can be about both $x(0)$ and $x(T)$. The usual approach is to assume that only a priori information is known about $x(0)$ so the boundary conditions (1.1) are replaced by initial conditions,

$$0 = b(x(0)) \tag{1.2}$$

Both of these differences add substantial difficulty to the smoothing problem.

There are several closely related problems which we briefly describe. For a more complete discussion of these problems we refer the reader to [5].

The filtering problem is to estimate $x(t)$ for $t > 0$ given the past inputs and outputs $u(0:t), y(0:t)$. In filtering problems it is usually assumed a priori information is only available about $x(0)$.

The fixed point smoothing problem is to estimate $x(t)$ given the inputs $u(0:T)$ and outputs $y(0:T)$ as $T \to \infty$. Again it is usually assumed a priori information is only available about $x(0)$. There are extensions of the above model to the situation where a priori information is available at times interior to $[0,T]$. Then the boundary condition is replaced by a multipoint condition

$$0 = b(x(0), x(t_1), \ldots, x(t_l), x(T)) \tag{1.3}$$

where $0 < t_1 < \ldots < t_l < T$. For simplicity of exposition we shall not discuss such problems but the methods that we will present readily generalize.

In some estimation problems, the input and/or output data may be censored. For example in fixed interval smoothing the output may be unknown on some subset $[t_0, t_1] \subset [0,T]$. The model can be adjusted to handle this by setting $y(t) = 0$ and $h(t,x,u) = 0$ for $t \in [t_0, t_1]$. If the input is unknown on $[t_0, t_1]$ then we can treat it as a driving noise, see below.

The approach that we take is an extension of Bryson and Frazier [3]. The above model may not be known exactly or there may be driving noise, observation noise or boundary noise so we don't expect to be able to estimate $x(t)$ exactly. To account for this, we add noises $\beta,\ w(t),\ v(t)$ to the model,

$$\begin{aligned} \dot{x} &= f(t,x,u) + g(t,x)w \\ y &= h(t,x,u) + k(x,t)v \\ \beta &= b(x(0), x(T)) \end{aligned} \tag{1.4}$$

We could take $w(t),\ v(t)$ to be independent white Gaussian noises and β to be an independent random vector but then we would have to deal with the mathematical technicalities of a stochastic ODE with random boundary conditions. Not much is known about this except for linear ODEs with well-posed linear boundary conditions. In this case $x(t)$ is a reciprocal process [8], [9] and there are well-developed linear smoothers [1], [14], [15], [4]. Instead we assume that $w(t) \in L^2([0,T], \mathbb{R}^l)$, $v(t) \in L^2([0,T], \mathbb{R}^q)$ are unknown functions and β is an unknown vector in $\mathbb{R}^k$.

The model (1.4) is well-posed if for every $u(0:T),\ y(0:T),\ \beta,\ w(0:T),\ v(t)$ there exists an unique solution $x(0:T)$ to (1.4). We do not assume that the model is well-posed. If it is not well-posed then for a given $u(0:T),\ y(0:T)$ we restrict our attention to those noise triples $\beta,\ w(0:T),\ v(0:T)$ for which there exists a unique $x(0:T)$ satisfying (1.4). Such noise triples are said to be admissible. To simplify the discussion we shall restrict our attention to the case where $k(x,t) = I$ so that $v(t) = y(t) - h(x(t),t)$.

We define an "energy" associated with the three noises, a simple choice is

$$\frac{1}{2}\left(\beta' P \beta + \int_0^T w'(t)Q(t)w(t) + v'(t)R(t)v(t)\ dt\right) \tag{1.5}$$

where P is positive definite and $Q(t),\ R(t)$ are positive definite for any $t \in [0,T]$. The larger P is relative to $Q(t),\ R(t)$ the more we assume that the boundary condition is satisfied exactly. The larger $Q(t)$ is relative to $P,\ R(t)$ the more we assume that the dynamics is satisfied exactly. The larger $R(t)$ is relative to $P,\ Q(t)$ the more we assume that the observation is exact. One could generalize the definition (1.5) of the energy by allowing $Q,\ R$ to depend on $x,\ u$ or even more general nonlinear functions but for simplicity we do not do so.

We postulate that optimal state estimate $\hat{x}(t),\ t \in [0,T]$ is generated by the admissible noise triple of minimum energy consistent with the known functions $u(0:T),\ y(0:T)$. Therefore for each known $u(0:T),\ y(0:T)$, we must solve the optimal control problem of minimizing the energy (1.5) over all admissible noise triples. If the minimum energy is achieved by the noise triple $\hat{\beta},\ \hat{w}(0:T),\ \hat{v}(0:T)$ then the optimal estimate $\hat{x}(0:T)$ satisfies

$$\begin{aligned}\dot{\hat{x}} &= f(t,\hat{x},u) + g(t,\hat{x})\hat{w} \\ y &= h(t,\hat{x},u) + v \\ \hat{\beta} &= b(\hat{x}(0),\hat{x}(T)).\end{aligned} \tag{1.6}$$

This is an application of the least squares approach to estimation that has been widely used since Gauss. Generally speaking for linear systems the least squares and stochastic approaches give the same estimates when $P,\ Q(t),\ R(t)$ of the least squares approach are the inverses of the covariances of the boundary, driving and observation noises of the stochastic approach. The advantage of the least squares approach is that it readily generalizes to nonlinear systems with a priori boundary information while the stochastic approach can lead to considerable technical difficulties.

1.2 The Variational Approach

We apply the Pontryagin minimum principle to the above problem. This yields the first order necessary that must be satisfied by an optimal solution. For fixed, known $u(0:T), y(0:T)$, the noise triple $\hat{\beta}$, $\hat{w}(0:T)$, $\hat{v}(0:T)$ that minimizes the energy (1.5) generates the state $\hat{x}(0:T)$ and adjoint $\hat{\lambda}(0:T)$ trajectories that satisfy the following conditions. Define the Hamiltonian

$$\begin{aligned} H(t,\lambda,x,w) &= \lambda'\left(f(t,x,u(t)) + g(t,x)w\right) + \frac{1}{2}w'(t)Q(t)w(t) \\ &\quad + \frac{1}{2}(y(t)-h(t,x,u(t)))'R(t)(y(t)-h(t,x,u(t))). \end{aligned} \tag{1.7}$$

Then

$$\dot{\hat{x}}(t) = \left(\frac{\partial H}{\partial \lambda}(t,\hat{\lambda}(t),\hat{x}(t),\hat{w}(t))\right)' \tag{1.8}$$

$$\dot{\hat{\lambda}}(t) = -\left(\frac{\partial H}{\partial x}(t,\hat{\lambda}(t),\hat{x}(t),\hat{w}(t))\right)' \tag{1.9}$$

$$\hat{w}(t) = \mathrm{argmin}_w H(t,\hat{\lambda}(t),\hat{x}(t),w) \tag{1.10}$$

$$\hat{\lambda}(0) = \left(\frac{\partial b}{\partial x^0}(\hat{x}(0),\hat{x}(T))\right)' Pb(\hat{x}(0),\hat{x}(T)) \tag{1.11}$$

$$\hat{\lambda}(T) = \left(\frac{\partial b}{\partial x^T}(\hat{x}(0),\hat{x}(T))\right)' Pb(\hat{x}(0),\hat{x}(T)) \tag{1.12}$$

Because of the form of the Hamiltonian we have that

$$\hat{w}(t) = -Q^{-1}(t)g'(t,\hat{x}(t))\hat{\lambda}(t) \tag{1.13}$$

and so

$$\begin{aligned} \dot{\hat{x}}(t) &= f(t,\hat{x}(t),u(t)) - g(t,\hat{x}(t))Q^{-1}(t)g'(t,\hat{x}(t))\hat{\lambda}(t) \\ \dot{\hat{\lambda}}(t) &= -\left(\frac{\partial f}{\partial x}(t,\hat{x}(t),u(t)) - \frac{\partial g}{\partial x}(t,\hat{x}(t))Q^{-1}(t)g'(t,\hat{x}(t))\hat{\lambda}(t)\right)'\hat{\lambda}(t) \\ &\quad + \left(\frac{\partial h}{\partial x}(t,\hat{x}(t),u(t))\right)' R(t)\left(y(t)-h(t,\hat{x}(t),u(t))\right) \\ \hat{\lambda}(0) &= \left(\frac{\partial b}{\partial x^0}(\hat{x}(0),\hat{x}(T))\right)' Pb(\hat{x}(0),\hat{x}(T)) \\ \hat{\lambda}(T) &= \left(\frac{\partial b}{\partial x^T}(\hat{x}(0),\hat{x}(T))\right)' Pb(\hat{x}(0),\hat{x}(T)). \end{aligned} \tag{1.14}$$

This is a nonlinear two point boundary value problem in $2n$ variables that can be solved by direct methods, shooting methods or iterative methods [6], [7], [12]. Below we shall introduce iterative methods that takes advantage of the variational nature of the problem.

1.3 Smoothing of Linear Boundary Value Processes

Let's look at the linear case where the noisy system takes the form

$$\begin{aligned} \dot{x} &= F(t)x + B(t)u + G(t)w \\ y &= H(t)x + D(t)u + v \\ \beta &= V^0 x(0) + V^T x(T) \end{aligned} \tag{1.15}$$

then the two point boundary value problem becomes

$$\begin{aligned} \dot{\hat{x}}(t) &= F(t)\hat{x}(t) + B(t)u(t) - G(t)Q^{-1}(t)G'(t)\hat{\lambda}(t) \\ \dot{\hat{\lambda}}(t) &= -F'(t)\hat{\lambda}(t) + H'(t)R(t)\left(y(t) - H(t)\hat{x}(t) - D(t)u(t)\right) \\ \hat{\lambda}(0) &= (V^0)'PV^0\hat{x}(0) + (V^0)'PV^T\hat{x}(T) \\ \hat{\lambda}(T) &= (V^T)'PV^0\hat{x}(0) + (V^T)'PV^T\hat{x}(T). \end{aligned} \tag{1.16}$$

We define

$$\begin{aligned} \xi &= \begin{bmatrix} \hat{x} \\ \hat{\lambda} \end{bmatrix} \\ \mu &= \begin{bmatrix} u \\ y \end{bmatrix} \\ \mathcal{A}(t) &= \begin{bmatrix} F(t) & -G(t)Q^{-1}(t)G'(t) \\ -H'(t)R(t)H(t) & -F'(t) \end{bmatrix} \\ \mathcal{B}(t) &= \begin{bmatrix} B(t) & 0 \\ -H'(t)R(t)D(t) & H'(t)R(t) \end{bmatrix} \\ \mathcal{C}(t) &= \begin{bmatrix} I & 0 \end{bmatrix} \\ \mathcal{V}^0 &= \begin{bmatrix} (V^0)'PV^0 & -I \\ (V^T)'PV^0 & 0 \end{bmatrix} \\ \mathcal{V}^T &= \begin{bmatrix} (V^T)'PV^0 & 0 \\ (V^T)'PV^T & -I \end{bmatrix} \end{aligned}$$

then the two point boundary value problem becomes

$$\begin{aligned} \dot{\xi} &= \mathcal{A}(t)\xi + \mathcal{B}(t)\mu \\ 0 &= \mathcal{V}^0\xi(0) + \mathcal{V}^T\xi(T) \end{aligned} \tag{1.17}$$

This problem is well-posed if

$$\mathcal{F} = \mathcal{V}^0 + \mathcal{V}^T\Phi(T,0)$$

is invertible where $\Phi(t,s)$ is the $2n \times 2n$ matrix solution to

$$\begin{aligned} \frac{\partial \Phi}{\partial t}(t,s) &= \mathcal{A}(t)\Phi(t,s) \\ \Phi(s,s) &= I. \end{aligned} \tag{1.18}$$

Then the Green's matrix is

$$\mathcal{G}(t,s) = \begin{cases} \Phi(t,0)\mathcal{F}^{-1}\mathcal{V}^0\Phi(0,s) & t > s \\ -\Phi(t,0)\mathcal{F}^{-1}\mathcal{V}^T\Phi(T,s) & t < s \end{cases} \tag{1.19}$$

and

$$\xi(t) = \int_0^T \mathcal{G}(t,s)\mathcal{B}(s)\mu(s)\ ds$$

so the least squares smoothed estimate is

$$\hat{x}(t) = \int_0^T \mathcal{C}(t)\mathcal{G}(t,s)\mathcal{B}(s)\mu(s)\ ds.$$

We say that $F(t),\ G(t)$ is controllable on $[0,T]$ if the gramian

$$\int_0^T \Phi(T,s)G(s)G'(s)\Phi'(T,s)\ ds$$

is positive definite. The pair $H(t),\ F(t)$ is observable on $[0,T]$ if the gramian

$$\int_0^T \Phi'(t,0)H'(t)H(t)\Phi(t,0)\ ds$$

is positive definite It can be shown that if $F(t),\ G(t)$ is controllable on $[0,T]$, $H(t),\ F(t)$ is observable on $[0,T]$ and the eigenvalues of $Q(t), R(t)$ are bounded away from zero then there exists a unique solution to the optimization problem and this solution must satisfy (1.16) so it is well-posed.

1.4 Solution of the Linear Two Point Boundary Value Problem by Direct Methods

The most direct way of solving the two point boundary value problem is to compute the Green's matrix (1.19). This requires finding the fundamental matrix which satisfies an $2n\times 2n$ linear differential equation (1.18). If n is large then this is a nontrivial task. Moreover since the dynamics is Hamiltonian (1.16) we expect it to have modes that are unstable in forward time and also modes that are unstable in backward time. This is a difficulty for any numerical scheme

Several authors [13], [1], [11], [2], [14] have proposed other direct methods based on diagonalization or triangularization of the dynamics (1.16). We shall describe the triangularization method due to Weinert [14]. Our use of notation is different from his. He assumes that the input $u(t) = 0$. The first step is to make a change of variables in x, λ space so as to triangularize the dynamics (1.16). Let $N(t)$ be the $n \times n$ matrix that satisfies the backward Riccati equation

$$\begin{aligned}\dot{N} &= -NF - F'N + NGQ^{-1}G'N - H'RH \\ N(T) &= 0\end{aligned}$$

Define

$$\rho(t) = \lambda(t) - N(t)x(t) \tag{1.20}$$

then the dynamics is triangular

$$\begin{bmatrix}\dot{x} \\ \dot{\rho}\end{bmatrix} = \begin{bmatrix}F - GQ^{-1}G'N & -GQ^{-1}G' \\ 0 & -F' + GQ^{-1}G'N\end{bmatrix}\begin{bmatrix}x \\ \rho\end{bmatrix} + \begin{bmatrix}0 \\ H'R\end{bmatrix} y. \tag{1.21}$$

The boundary conditions become

$$\mathcal{V}^0 \begin{bmatrix}I & 0 \\ N(0) & I\end{bmatrix}\begin{bmatrix}x(0) \\ \rho(0)\end{bmatrix} + \mathcal{V}^T \begin{bmatrix}x(T) \\ \rho(T)\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix}. \tag{1.22}$$

Moreover Weinert asserts that the dynamics of x is stable in the forward direction when $\rho = 0$ and the dynamics of ρ is stable in the backward direction.

If there is only a priori data about the initial condition ($V^0 = I,\ V^T = 0$) then

$$\begin{aligned}\mathcal{V}^0 &= \begin{bmatrix}P & -I \\ 0 & 0\end{bmatrix} \\ \mathcal{V}^T &= \begin{bmatrix}0 & 0 \\ 0 & -I\end{bmatrix}\end{aligned}$$

and the boundary conditions also triangularize

$$\begin{aligned}(P - N(0))x(0) - \rho(0) &= 0 \\ \rho(T) &= 0.\end{aligned} \tag{1.23}$$

We can find the solution of (1.21) and (1.23) by first integrating the differential equation for ρ backward from the final condition $\rho(T) = 0$ and then integrating the differential equation for x forward from the initial condition $(P - N(0))x(0) - \rho(0) = 0$.

But if there is a priori data about both the initial and final conditions then the boundary conditions do not triangularize. We can still find a particular solution by first integrating the lower half of (1.21) backward from the final condition $\rho(T) = 0$ and then integrating the upper half forward from the initial condition $x(0) = 0$. Denote this solution by $\bar{x}(t),\ \bar{\rho}(t)$, it will probably not satisfy the correct boundary conditions (1.22).

Therefore we need to compute the $2n \times 2n$ matrix solution of

$$\begin{aligned}\frac{d}{dt}\Sigma(t) &= \begin{bmatrix}F - GQ^{-1}G'N & -GQ^{-1}G' \\ 0 & -F' + GQ^{-1}G'N\end{bmatrix}\Sigma(t) \\ \Sigma(0) &= I.\end{aligned} \tag{1.24}$$

This can be reduced to the solution of two $n \times n$ matrix differential equations. Let $\Psi(t,s)$ be the solution of

$$\begin{aligned} \frac{\partial}{\partial t}\Psi(t,s) &= \left(F - GQ^{-1}G'N\right)\Psi(t,s) \\ \Psi(s,s) &= I \end{aligned} \tag{1.25}$$

and let $M(t)$ be the solution of the Lyapunov equation

$$\begin{aligned} \dot{M} &= \left(F - GQ^{-1}G'N\right)M + M\left(F - GQ^{-1}G'N\right)' - GQ^{-1}G' \\ M(0) &= 0 \end{aligned} \tag{1.26}$$

then

$$\Sigma(t) = \begin{bmatrix} \Psi(t,0) & M(t)\Psi'(T,t) \\ 0 & \Psi'(T,t) \end{bmatrix}. \tag{1.27}$$

The solution to (1.21) that we seek is of the form

$$\begin{bmatrix} x(t) \\ \rho(t) \end{bmatrix} = \begin{bmatrix} \bar{x}(t) \\ \bar{\rho}(t) \end{bmatrix} + \Sigma(t)\xi \tag{1.28}$$

where ξ is determined by the boundary conditions (1.22).

Weinert's method is more efficient than the direct computation of the Green's matrix (1.19). The latter requires finding the fundamental solution of a $2n \times 2n$ linear differential equation. Weinert's method requires finding the fundamental solution of a $n \times n$ linear differential equation and the $n \times n$ solutions of a Riccati differential equation and a Lyapunov differential equation. But neither method is practical if n is large.

1.5 Solution of the Linear Smoothing Problem by an Iterative Method

We prsent an iterative method to solve the linear smoothing problem and generalize it to the nonlinear problem in the next section. We are given the input $u(0:T)$ and observation $y(0:T)$ and we wish to find the solution of (1.15) that minimizes the least square criterion (1.5). We do not assume that the boundary value problem (1.15) is well-posed so for given β and $w(0:T)$ there may not be a solution and/or it may not be unique. Therefore to parametrize the solutions of (1.15) we must choose a well-posed boundary condition like

$$x(0) = x^0.$$

We fix the observations $y(0:t)$, then given x^0 and $w(0:T)$ there is a unique solution $x(0:T),\ v(0:T)$ to

$$
\begin{aligned}
\dot{x} &= F(t)x + B(t)u + G(t)w \\
y &= H(t)x + D(t)u + v \\
x(0) &= x^0
\end{aligned}
\tag{1.29}
$$

to which we associate the cost

$$
\pi(x^0, w) = \frac{1}{2}\left(\beta' P \beta + \int_0^T w'(t)Q(t)w(t) + v'(t)R(t)v(t)\ dt\right) \tag{1.30}
$$

We compute the first variation in π due to variations in $x^0, w(0:T)$,

$$
\begin{aligned}
\delta\pi &= \pi(x^0 + \delta x^0, w + \delta w) - \pi(x^0, w) \\
&= \beta' P \left(V^0 + V^T \Phi(T, 0)\right) \delta x^0 \\
&\quad - \int_0^T v'(s)R(s)H(s)\Phi(s, 0)\ ds\ \delta x^0 \\
&\quad + \beta' P V^T \int_0^T \Phi(T, s)G(s)\delta w(s)\ ds \\
&\quad + \int_0^T w'(s)Q(s)\delta w(s)\ ds \\
&\quad - \int_0^T \int_s^T v'(t)R(t)H(t)\Phi(t, s) dt G(s)\delta w(s)\ ds \\
&\quad + O(\delta x^0, \delta w)^2
\end{aligned}
$$

where $\Phi(t, s)$ is the $n \times n$ matrix solution of

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

If n is large it is expensive to compute $\Phi(t, s)$ so instead we define

$$
\mu'(s) = \beta' P V^T \Phi(T, s) - \textstyle\int_s^T v'(t)R(t)H(t)\Phi(t, 0)\ dt.
$$

Then $\mu(s)$ satisfies the final value problem

$$
\begin{aligned}
\frac{d}{ds}\mu(s) &= -F(s)\mu(s) + H'(s)R(s)v(s) \\
\mu(T) &= (V^T)' P \beta
\end{aligned}
\tag{1.31}
$$

and the first variation is

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

We solve the problem by gradient descent. Given $u(0:T)$ and $y(0:T)$, choose any x^0 and $w(0:T)$, solve the initial value problem (1.29) and compute the cost (1.30). Then compute the gradient of the cost with respect to $x^0,\ w(0:T)$ by solving the final value problem (1.31). Choose a step size ϵ and define

$$\begin{aligned} \delta x^0 &= -\epsilon\left((V^0)'P\beta + \mu(0)\right) \\ \delta w(s) &= -\epsilon\left(Q(s)w(s) + G'(s)\mu(s)\right). \end{aligned} \tag{1.32}$$

Replace $x^0,\ w(0:T)$ by $x^0 + \delta x^0,\ w(0:T) + \delta w(0:T)$, solve the new initial value problem (1.29) and compute the new cost (1.30). If the new cost is not sufficiently less than the previous cost then we change the step size and recompute (1.29) and (1.30). We repeat until we find a variation that reduces the cost sufficiently. Standard step size rules such as Armijo's can be used [10].

Once we find a variation that sufficiently lowers the cost, we accept it. Then we again compute the gradient of the cost with respect to $x^0,\ w(0:T)$ by solving the final value problem (1.31) and repeat until it is too difficult to find a variation that reduces the cost any further. The last solution to the initial value problem (1.29) is the smoothed estimate $\hat{x}(0:T)$.

1.6 Solution of the Nonlinear Smoothing Problem by an Iterative Method

We return to the nonlinear smoothing problem and seek to minimize (1.5) subject to (1.4). Again we use gradient descent. Given $u(0:T)$ and $y(0:T)$ choose any x^0 and $w(0:T)$, solve the initial value problem,

$$\begin{aligned} \dot{x} &= f(t,x,u) + g(t,x(t))w \\ x(0) &= x^0 \end{aligned} \tag{1.33}$$

and compute the cost (1.30). Then compute the gradient of the cost with respect to $x^0,\ w(0:T)$ by solving the final value problem (1.31) where

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

We continue as in the linear case. Choose a step size ϵ and define

$$\begin{aligned} \delta x^0 &= -\epsilon\left((V^0)'P\beta + \mu(0)\right) \\ \delta w(s) &= -\epsilon\left(Q(s)w(s) + G'(s)\mu(s)\right). \end{aligned} \tag{1.34}$$

Replace $x^0,\ w(0:T)$ by $x^0 + \delta x^0,\ w(0:T) + \delta w(0:T)$ and solve the new initial value problem (1.33) and compute the new cost (1.30). If the new cost is not sufficiently less than the previous cost then we change the step size and recompute (1.33) and (1.30). We repeat until we find a variation that reduces the cost sufficiently.

Once we find a variation that sufficiently lowers the cost, we accept it. Then we again compute the gradient of the cost with respect to $x^0,\ w(0:T)$ by solving the final value problem (1.31) and repeat until it is too difficult to find a variation that reduces the cost any further. The last solution to the initial value problem (1.33) is the smoothed estimate $\hat{x}(0:T)$.

1.7 Conclusion

We have discussed the problem of obtaining smoothed estimates of a nonlinear process from continuous observations and boundary information which we call boundary value processes. Because of technical difficulties associated with the stochastic approach, we have opted for a least squares approach. The latter is applicable even when the description of the boundary value process is not well-posed. We reviewed an existing direct method for smoothing linear boundary value processes and presented new iterative methods for smoothing both linear and nonlinear boundary value process. For low dimensional linear processes direct methods can be used but for high dimensional and/or nonlinear processes the iterative method is preferable.

Acknowledgement. It is a pleasure and honor to dedicate this paper to my esteemed colleague, Michael Zeitz on the occasion of his sixty fifth birthday.

References

1. Adams M B ,Willsky A S, Levy B C (1984) Linear estimation of boundary value stochastic processes, Part I and II. IEEE Transactions on Automatic Control 29:803–821
2. Bagchi, A, Westdijk H (1989) Smoothing and likelihood ratio for Gaussian boundary value processes. IEEE Transactions on Automatic Control 34:954–962
3. Bryson A E, Frazier M (1963) Smoothing for linear and nonlinear dynamic systems. In: Proc. Optimum Systems Synthesis Conference, Reprinted in: Kailath T (ed) Linear least-squares estimation. Dowden, Hutchinson and Ross, Stroudsberg, PA
4. 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
5. Gelb A (1975) Applied Optimal Estimation. MIT Press, Cambridge, MA
6. Keller H B (1968) Numerical Methods for two-point boundary-value problems. Blaisdell, Waltham, MA

7. Keller H B (1976) Numerical Solution of two point boundary value problems. SIAM, Philadelphia
8. Krener A J (1986) Reciprocal processes and the stochastic realization problem for acausal systems. In: Byrnes C B, Lindquist A (eds) Modeling, identification and robust control. Elsevier Science.
9. 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.
10. McCormick G P (1983) Nonlinear programming, theory, algorithms and applications. Wiley Interscience, New York
11. Nikoukhah R, Adams M B ,Willsky A S, Levy B C (1989) Estimation for boundary value descriptor systems. Circuits, Systems and Signal Processing 8:25–48
12. Roberts S M, Shipman J S (1972) Two-point boundary value problems: shooting methods. Blaisdell, Waltham, MA.
13. Weinert H L, Desai U (1981) On complementary models and fixed interval smoothing. IEEE Transactions on Automatic Control 26:863–867
14. Weinert H L (1991) A note on efficient smoothing for boundary value models. Int. J. Control, 53:503–507
15. Weinert H L (2001) Fixed interval smoothing for state space models. Kluwer Academic Publishers, Norwell MA