Numerical Approximation for Nonlinear Filtering and Finite–Time Observers

Matthew R. James
Department of Mathematics
University of Kentucky
Lexington, KY 40506
USA

François LeGland
INRIA–Sophia Antipolis
2004, Route des Lucioles
F–06565 Valbonne Cedex
FRANCE

I Models of Partially Observed Systems

We consider partially observed systems of the form

$$\left\{ \begin{aligned} dX_t &= b(X_t)\, dt + \sigma(X_t)\, dW_t \\ dY_t &= h(X_t)\, dt + dV_t \end{aligned} \right. \tag{1}$$

where $\{W_t\,,\, t \geq 0\}$ and $\{V_t\,,\, t \geq 0\}$ are independent Wiener processes of appropriate dimensions, with covariance matrices I and R respectively. We are interested in the state estimation problem, under various hypotheses concerning $a = \sigma\, \sigma^*$ and R.

Consider first the two extreme cases : If $a \equiv 0$ and $R \equiv 0$, we are dealing with an *observer* problem for the deterministic system

$$\left\{ \begin{aligned} \dot{X}_t &= b(X_t) \\ z_t &= h(X_t) \end{aligned} \right. \tag{2}$$

At the other extreme, if R is non–singular, we are dealing with a *nonlinear filtering* problem for the diffusion process (1) . We can also easily handle the following intermediate case : If R is non–singular and $a \equiv 0$, we are again dealing with a *nonlinear filtering* problem but the state equation is now an ODE

$$\left\{ \begin{aligned} \dot{X}_t &= b(X_t) \\ dY_t &= h(X_t)\, dt + dV_t \end{aligned} \right.$$

Let us point out that the solution of the state estimation problem is radically different, depending on whether R is non–singular or identically zero. On the other hand, whether a is non–singular, singular or identically zero only affects the algorithms to be used.

Our purpose is to present, for each of the three main cases described above, a solution to the state estimation problem, and to suggest some numerical approximation procedures. The general idea is to study the asymptotics $R \to 0$. As a by–product, we expect to obtain some numerical algorithms for the nonlinear filtering problem, that are robust when the non–singular matrix R is *small.*

II Solutions to State Estimation Problems

We assume for simplicity that $b \in C_b^1(\mathbf{R}^m, \mathbf{R}^m)$ and $h \in C_b^1(\mathbf{R}^m, \mathbf{R}^p)$, unless otherwise stated.

Let us begin with the *nonlinear filtering* (NLF) problem.

□ When R is non-singular, the Bayesian approach to the state estimation problem is to compute the unnormalized conditional probability distribution $\mu_t(dx)$ of the state X_t, given the past observations $\mathcal{Y}_t = \sigma(Y_s\,,\, 0 \leq s \leq t)$. By definition

$$\langle \mu_t, f \rangle = \mathbf{E}^\dagger[f(X_t)\, Z_t \mid \mathcal{Y}_t] \,,$$

for any test function f, where

$$Z_t^s = \exp\left\{ \int_s^t h^*(X_\tau) R^{-1}\, dY_\tau - \tfrac{1}{2} \int_s^t |h(X_\tau)|^2_{R^{-1}}\, d\tau \right\} \quad \text{and} \quad Z_t = Z_t^0 \,,$$

and $P^\dagger$ is the *reference* probability measure. The probability distribution $\mu_t(dx)$ satisfies a stochastic PDE in weak sense. Usually, this p.d.f. has a density w.r.t. the Lebesgue measure, i.e. $\mu_t(dx) = p_t(x)\, dx$. A sufficient condition for this to hold, is that the probability distribution $\mu_0(dx)$ of the initial condition X_0 has already a density w.r.t. the Lebesgue measure, i.e. $\mu_0(dx) = p_0(x)\, dx$. We will assume that $p_0(x) > 0$ for all $x \in \mathbf{R}^m$. The unnormalized conditional density $p_t(x)$ is the unique solution of the Zakai equation

$$dp_t = L^* p_t\, dt + p_t h^* R^{-1}\, dY_t \,, \tag{3}$$

with initial condition $p_0(x)$, where L^* is the formal adjoint of the second-order partial differential operator

$$L = \tfrac{1}{2} \operatorname{tr}[\, a\, \frac{\partial^2}{\partial x^2}\,] + b \cdot \frac{\partial}{\partial x} \,,$$

associated with the SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t \,. \tag{4}$$

□ If in addition $a \equiv 0$, then the Zakai equation ((3)) becomes a first order stochastic PDE, for which a representation result is available in terms of the flow $\Phi_t(x)$ associated with the ODE

$$\dot{X}_t = b(X_t) \,. \tag{5}$$

Actually, define

$$J_{t-s}(x) = \det[\, \frac{\partial \Phi_{t-s}}{\partial x}(x)\,] = \exp\left\{ \int_s^t \operatorname{div}\, b(\Phi_{\tau-s}(x))\, d\tau \right\} \,,$$

$$\Xi_{s,t}(x) = \exp\left\{ \int_s^t h^*(\Phi_{\tau-s}(x)) R^{-1}\, dY_\tau - \tfrac{1}{2} \int_s^t |h(\Phi_{\tau-s}(x))|^2_{R^{-1}}\, d\tau \right\} \,.$$

In this case, the unique solution of the Zakai equation ((3)) satisfies

$$p_t(\Phi_{t-s}(x)) \cdot J_{t-s}(x) = \Xi_{s,t}(x) \cdot p_s(x) \,, \tag{6}$$

or equivalently, introducing the logarithmic transform $W_t(x) = -\log p_t(x)$

$$W_t(\Phi_{t-s}(x)) - \log J_{t-s}(x) = W_s(x) - \log \Xi_{s,t}(x) \,. \tag{7}$$

We turn now to the *observer* problem.

Let $\{x_t^*\,,\, 0 \leq t \leq T\}$ denote the *true* state trajectory producing the available observation trajectory $\{z_t\,,\, 0 \leq t \leq T\}$. The idea is to build an observer by considering the limit of a sequence of nonlinear filtering problems with noise covariances going to zero. Two different cases are possible

- Introduce small noises of similar intensities in both the state equation and the observation, i.e. set $a = \varepsilon I$ and $R = \varepsilon I$,
- Introduce a small noise in the observation only, i.e. set $a \equiv 0$ and $R = \varepsilon I$.

☐ In the first case, it is proved in James [2] that

$$-\varepsilon \log p_t^\varepsilon(x) \xrightarrow{\varepsilon \downarrow 0} m_t'(x)$$

in probability uniformly on compact subsets of $x \in \mathbf{R}^m$, where up to an additive constant independent of x, $m_t'(x)$ is the unique solution of the Hamilton–Jacobi equation

$$\frac{\partial m_t'}{\partial t} + \tfrac{1}{2}\left|\frac{\partial m_t'}{\partial x}\right|^2 + b \cdot \frac{\partial m_t'}{\partial x} - V_t = 0 \ , \tag{8}$$

with initial condition $m_0'(x) = 0$, in the *viscosity* sense, where

$$V_t(x) = \tfrac{1}{2}|z_t - h(x)|^2 \ .$$

In addition, $m_t'(x)$ is the value function associated with the following control problem. Introduce first the action functional

$$I_t(\xi) = \tfrac{1}{2}\int_0^t |\dot{\xi}_s - b(\xi_s)|^2 \, ds$$

if $\xi \in C([0,T]\,;\,\mathbf{R}^m)$ is absolutely continuous, and $I_t(\xi) = +\infty$ otherwise. Define also

$$F_t(\xi) = \int_0^t V_s(\xi_s)\, ds = \tfrac{1}{2}\int_0^t |z_s - h(\xi_s)|^2 \, ds \ .$$

Then

$$m_t'(x) = \inf\left\{ I_t(\xi) + F_t(\xi) \ : \ \xi_t = x \right\} \ .$$

Clearly $m_t'(x) \geq 0$ and $m_t'(x_t^*) = 0$ for the *true* state trajectory, and we define our observer as the set

$$\widehat{x}_t' = \underset{x \in \mathbf{R}^m}{\operatorname{argmin}} \ m_t'(x) = \left\{ x \in \mathbf{R}^m \ : \ m_t'(x) = 0 \right\} \ . \tag{9}$$

Obviously $x_t^* \in \widehat{x}_t'$ for all $t \geq 0$. It is proved in James [3] that, provided the deterministic system ((2)) is *observable* on $[0,T]$ (i.e. the map $x_0 \mapsto \{z_s \, , \, 0 \leq s \leq t\}$ is injective), the set-valued observer ((9)) is actually a *finite-time observer* (FTO) on $[0,T]$ (meaning that $\widehat{x}_t'$ is defined in terms of a recursive system with the property that $\widehat{x}_t' = \{x_t^*\}$ for all $t \geq T$).

☐ In the second case, it follows from equation ((7)) that

$$-\varepsilon \log p_t^\varepsilon(x) = \varepsilon W_t^\varepsilon(x) \xrightarrow{\varepsilon \downarrow 0} m_t(x) \ ,$$

in probability uniformly on compact subsets of $x \in \mathbf{R}^m$, where up to an additive constant independent of x, $m_t(x)$ is given by

$$m_t(\Phi_t(x)) = \int_0^t V_s(\Phi_s(x))\, ds \quad \text{or} \quad m_t(x) = \int_0^t V_s(\Phi_{t-s}^{-1}(x))\, ds \ ,$$

i.e. $m_t(x) = F_t(\xi^{t,x})$, where $\xi^{t,x}$ is the unique solution of the ODE ((5)) ending in x at time t. In addition, $m_t(x)$ is the unique solution of the linear first-order PDE

$$\frac{\partial m_t}{\partial t} + b \cdot \frac{\partial m_t}{\partial x} - V_t = 0 \; ,$$

satisfying the initial condition $m_0 = 0$. Just as above, it is clear that $m_t(x) \geq 0$ and $m_t(x_t^*) = 0$ for the *true* state trajectory, and we define our observer as the set

$$\hat{x}_t = \operatorname*{argmin}_{x \in \mathbf{R}^m} \; m_t(x) = \Big\{ x \in \mathbf{R}^m \; : \; m_t(x) = 0 \Big\} \; . \tag{10}$$

Here again, it is obvious that $x_t^* \in \hat{x}_t$ for all $t \geq 0$, and in addition the set–valued observer defined by ((10)) is actually a FTO on $[0,T]$, provided the deterministic system ((2)) is *observable* on $[0,T]$.

Note that $m_t(x) = F_t(\xi^{t,x})$ where $I_t(\xi^{t,x}) = 0$ (i.e. $\xi^{t,x}$ solves the ODE ((5)) exactly) and $\xi_t^{t,x} = x$, whereas in the definition of $m_t'(x)$, a penalty $I_t(\xi)$ is put on those trajectories ξ that do not solve the ODE ((5)). This is a less severe requirement, and is reflected in the relation $m_t'(x) \leq m_t(x)$. Note however that $\hat{x}_t = \hat{x}_t'$. This is the set of those points that are *indistinguishable* from the true state x_t^*. In conclusion, the observer ((10)) is more *precise* than the observer ((9)), whereas the latter is expected to be more *robust* w.r.t. modeling errors.

III Numerical Approximation

In this section, we restrict ourselves to the situation where the state satisfies an ODE, in which case the solution to the NLF problem is given by ((6)), where R is non–singular, and the corresponding FTO is given by ((10)), where $R \equiv 0$.

Concerning the approximation of the NLF ((6)), we wish to compute an approximate normalized conditional density $p_k^{\Delta,\delta}(x)$ (where Δ and δ denote the time discretization step and the space discretization step respectively) with the following property

($\star$) as $\Delta, \delta \downarrow 0$

$$\mathbf{E} \int_{\mathbf{R}^m} |p_{[t/\Delta]}^{\Delta,\delta}(x) - c_t p_t(x)| \, dx \longrightarrow 0 \quad \text{for all } t \geq 0 \; ,$$

where c_t is a normalization constant.

Concerning the approximation of the FTO ((10)), our approach is to build a family $\hat{x}_k^{\Delta,\delta}$ with the following property

($\star\star$) if the deterministic system ((2)) is observable on $[0,T]$, then as $\Delta, \delta \downarrow 0$

$$\operatorname{dist}(\hat{x}_{[t/\Delta]}^{\Delta,\delta}, \{x_t^*\}) \longrightarrow 0 \quad \text{for all } t \geq T \; .$$

A necessary and sufficient condition for ($\star\star$) to hold is $\operatorname{dist}(\hat{x}_{[t/\Delta]}^{\Delta,\delta}, \hat{x}_t) \to 0$ as $\Delta, \delta \downarrow 0$. The approximate observer $\hat{x}_k^{\Delta,\delta}$ will be defined in terms of an approximate value function $m_k^{\Delta,\delta}(x)$, i.e.

$$\hat{x}_k^{\Delta,\delta} = \Big\{ x \in \mathbf{R}^m \; : \; m_k^{\Delta,\delta}(x) \leq c^{\Delta,\delta} \Big\} \; ,$$

and a sufficient condition for ($\star\star$) to hold is $c^{\Delta,\delta} \downarrow 0$ and $m_{[t/\Delta]}^{\Delta,\delta}(x) \to m_t(x)$ uniformly on compact subsets of $\mathbf{R}^m$, as $\Delta, \delta \downarrow 0$.

Time Discretization

Consider a uniform partition $0 = t_0 < \cdots < t_k < \cdots$ of the time interval $[0, \infty)$, with time step $\Delta = t_k - t_{k-1}$. The first step is to sample the available observation trajectory.

The nonlinear filtering problem. If noisy observations $\{Y_t \, , \, t \geq 0\}$ are available, we first build the following sequence of compressed observations

$$y_k^\Delta = \frac{1}{\Delta}[Y_{t_k} - Y_{t_{k-1}}] = \frac{1}{\Delta}\int_{t_{k-1}}^{t_k} h(X_s)\,ds + \frac{1}{\Delta}[V_{t_k} - V_{t_{k-1}}]$$

and we use the approximate model

$$\left\{ \begin{aligned} \dot{X}_t &= b(X_t) \\ y_k^\Delta &= h(X_{t_k}) + v_k^\Delta \end{aligned} \right. \tag{11}$$

where $\{v_k^\Delta\,,\ k = 1, 2, \cdots\}$ is a Gaussian white noise sequence with covariance matrix R/Δ.

The solution of the NLF problem for the approximate model ((11)) is given in terms of the *a priori* and *a posteriori* conditional probability densities defined by

$$p_{k-\frac{1}{2}}^\Delta(x)\,dx = P(X_{t_k} \in dx \mid \mathcal{Y}_{k-1}^\Delta) \quad \text{and} \quad p_k^\Delta(x)\,dx = P(X_{t_k} \in dx \mid \mathcal{Y}_k^\Delta)\ ,$$

respectively, where $\mathcal{Y}_k^\Delta = \sigma(y_1^\Delta, \cdots, y_k^\Delta)$. The transition from $p_{k-1}^\Delta(x)$ to $p_k^\Delta(x)$ is divided into two steps

- *prediction step* : Transport by the flow gives $p_{k-\frac{1}{2}}^\Delta(x) = T_\Delta\ p_{k-1}^\Delta(x)$ where $\{T_t\,,\ t \geq 0\}$ is the semigroup associated with the linear first-order PDE

  $$\frac{\partial p_t}{\partial t} = L^* p_t\ . \tag{12}$$

  An explicit solution is available for this equation

  $$p_{k-\frac{1}{2}}^\Delta(\Phi_\Delta(x)) \cdot J_\Delta(x) = p_{k-1}^\Delta(x)\ , \tag{13}$$

  or equivalently

  $$\int_A p_{k-\frac{1}{2}}^\Delta(x)\,dx = \int_{\Phi_\Delta^{-1}(A)} p_{k-1}^\Delta(x)\,dx\ , \tag{14}$$

  for all Borel set $A \subset \mathbf{R}^m$.

- *correction step* : According to the Bayes formula

  $$p_k^\Delta(x) = c_k \cdot \Psi_k^\Delta(x) \cdot p_{k-\frac{1}{2}}^\Delta(x)\ , \tag{15}$$

  where

  $$\Psi_k^\Delta(x) = \exp\left\{-\tfrac{1}{2}\Delta\ |y_k^\Delta - h(x)|_{R^{-1}}^2\right\}\ ,$$

  is the *likelihood function* for the estimation of X_{t_k} in the approximate model ((11)), based on the observation y_k^Δ alone, and c_k is a normalization constant.

Introducing the logarithmic transform $W_k^\Delta(x) = -\log p_k^\Delta(x)$, it follows from from ((13)) and ((15)) that

$$W_k^\Delta(x) - \log J_\Delta(\Phi_\Delta^{-1}(x)) = -\log c_k + W_{k-1}^\Delta(\Phi_\Delta^{-1}(x)) + \tfrac{1}{2}\Delta\ |y_k^\Delta - h(x)|_{R^{-1}}^2\ . \tag{16}$$

The observer problem. If perfect observations $\{z_t\,,\ t \geq 0\}$ are available, i.e. $R \equiv 0$, we can simply use $z_k = z_{t_k}$, and our model becomes

$$\left\{ \begin{aligned} \dot{X}_t &= b(X_t) \\ z_k &= h(X_{t_k}) \end{aligned} \right. \tag{17}$$

Introducing the asymptotics $R = \varepsilon I$ in the NLF problem and sending ε to zero, it follows from equation ((16)) that

$$-\varepsilon \log p_k^{\Delta,\varepsilon}(x) = \varepsilon W_k^{\Delta,\varepsilon}(x) \xrightarrow{\varepsilon\downarrow 0} m_k^\Delta(x)\ ,$$

in probability uniformly on compact subsets of $x \in \mathbf{R}^m$, where $m_k^\Delta(x)$ satisfies the following relation

$$m_k^\Delta(x) = m_{k-1}^\Delta(\Phi_\Delta^{-1}(x)) + \Delta\, V_k^\Delta(x) \;,$$

where

$$V_k^\Delta(x) = \tfrac{1}{2}|z_k^\Delta - h(x)|^2 \quad \text{and} \quad z_k^\Delta = \frac{1}{\Delta}\int_{t_{k-1}}^{t_k} z_s\, ds = \frac{1}{\Delta}\int_{t_{k-1}}^{t_k} h(X_s)\, ds \neq z_k \;.$$

It is clear that $m_k^\Delta(x) \geq 0$. However, because the averaged observation z_k^Δ used in the definition of $m_k^\Delta(x)$ is different from the actual observation z_k, we have $V_k^\Delta(x_{t_k}^*) \neq 0$ in general for the *true* state trajectory. Therefore, we decide to use the actual observation z_k in the definition of $m_k^\Delta(x)$, instead of the averaged observation z_k^Δ, i.e.

$$m_k^\Delta(x) = m_{k-1}^\Delta(\Phi_\Delta^{-1}(x)) + \Delta\, V_k(x) \;, \tag{18}$$

where

$$V_k(x) = \tfrac{1}{2}|z_k - h(x)|^2 \;.$$

This relation can be divided into two steps

- *prediction step* : Transport by the flow gives $m_{k-\frac{1}{2}}^\Delta(x) = S_\Delta\; m_{k-1}^\Delta(x)$ where $\{S_t\,,\, t \geq 0\}$ is the semigroup associated with the linear first-order PDE

  $$\frac{\partial m_t}{\partial t} + b \cdot \frac{\partial m_t}{\partial x} = 0 \;. \tag{19}$$

  An explicit solution is available for this equation

  $$m_{k-\frac{1}{2}}^\Delta(\Phi_\Delta(x)) = m_{k-1}^\Delta(x) \;. \tag{20}$$

- *correction step* : The contribution of the new observation z_k to the approximate value function is given by

  $$m_k^\Delta(x) = m_{k-\frac{1}{2}}^\Delta(x) + \Delta\, V_k(x) \;.$$

We note that

$$m_{k-\frac{1}{2}}^\Delta(x) = F_{k-1}^\Delta(\xi^{t_k,x}) \quad \text{and} \quad m_k^\Delta(x) = F_k^\Delta(\xi^{t_k,x}) \;,$$

where $\xi^{t,x}$ is the unique solution of the ODE ((5)) ending in x at time t, and the functional $F_k^\Delta(\xi)$ satisfies for all $\xi \in C([0,T]\,;\,\mathbf{R}^m)$

$$F_k^\Delta(\xi) = F_{k-1}^\Delta(\xi) + \Delta\, V_k(\xi_{t_k}) = \Delta\, \left\{V_1(\xi_{t_1}) + \cdots + V_k(\xi_{t_k})\right\} \;.$$

Now it is clear that $m_k^\Delta(x) \geq 0$ and $m_k^\Delta(x_{t_k}^*) = 0$ for the *true* state trajectory, and we define our observer as the set

$$\widehat{x}_k^\Delta = \underset{x \in \mathbf{R}^m}{\operatorname{argmin}}\; m_k^\Delta(x) = \left\{x \in \mathbf{R}^m \,:\, m_k^\Delta(x) = 0\right\} \;. \tag{21}$$

Obviously $x_{t_k}^* \in \widehat{x}_k^\Delta$ for all k, and one can verify using the explicit formulas that $m_{[t/\Delta]}^\Delta(x) \to m_t(x)$ uniformly on compact subsets as $\Delta \downarrow 0$, with the consequence that property $(\star\star)$ holds for this discrete-time approximation.

Model Approximation and PDE Discretization

To obtain computable algorithms, it is necessary to discretize the linear first-order PDE (12) or (19) involved in the prediction step. Generally speaking, two classes of methods can be used : in the *finite difference* approximation (FD) a fixed bounded grid is used, and partial differential operators are approximated by finite differences on grid points, whereas in the *flow-based* approximation (FLOW) the explicit representation (13) or (20) is used to move grid points (or alternatively cells) along the flow of the ODE (5) .

A Finite Difference

A finite difference nonlinear filter. To derive a finite difference algorithm, we must first constrain the nonlinear filtering problem to a bounded domain. Let $D \subset \mathrm{R}^m$ be a m–dimensional open cube. After Dupuis–Ishii [1], we constrain the ODE (5) to the convex set $\bar{D}$ as follows. For $x \in \partial D$, let $\nu(x) = \{ \nu \in \mathrm{R}^m : |\nu| = 1, \langle \nu, x - z\rangle \leq 0$ for all $z \in \bar{D} \}$ denote the set of inward unit normals. For $x \in \bar{D}$, $v \in \mathrm{R}^m$, the projection $\pi(x, v)$ of the velocity vector v at x is given by v if $x \in D$, or $v + [\langle v, -\nu^*(x,v)\rangle \vee 0]\, \nu^*(x,v)$ if $x \in \partial D$, where $\nu^*(x,v)$ is an element of $\nu(x)$ which maximizes $\langle v, -\nu\rangle$, $\nu \in \nu(x)$. Define then $\tilde{b}(x) = \pi(x, b(x))$, $x \in \bar{D}$. By Theorem 5.1 of Dupuis–Ishii [1], there exists a unique absolutely continuous solution of the constrained ODE

$$\dot{\xi}_s = \tilde{b}(\xi_s) \quad \text{a.e. } 0 \leq s \leq t \tag{22}$$

satisfying $\xi_0 = x \in \bar{D}$.

A finite difference algorithm is obtained using a Markov chain scheme similar to those described in Kushner [5]. Let R^m_δ denote a coordinate grid of size $\delta > 0$. We define a system of neighborhoods $N_\delta(x) = \{ z \in \mathrm{R}^m_\delta : z = x$ or $z = x \pm \delta e_i$ for some $i = 1, \ldots, m \}$ for $x \in \mathrm{R}^m_\delta$, where $e_i \in \mathrm{R}^m$ denotes the i-th unit vector. We define next $\bar{D}^\delta = \bar{D} \cap \mathrm{R}^m_\delta$, $D^\delta = \{ x \in \bar{D}^\delta : N_\delta(x) \subset \bar{D}^\delta \}$, and $\partial D^\delta = \bar{D}^\delta \setminus D^\delta$. We define the jump intensity matrix $L_\delta(x, z)$ of a pure jump Markov process $\{\overline{X}^\delta_t, t \geq 0\}$ taking values in $\bar{D}^\delta$ by

$$L_\delta(x,z) = \begin{cases} -|\tilde{b}(x)|_1/\delta & \text{if } z = x, \\ \tilde{b}^\pm_i(x)/\delta & \text{if } z = x \pm \delta e_i \text{ and } i = 1, \ldots, m \\ 0 & \text{if } z \notin N_\delta(x), \end{cases} \tag{23}$$

with the notation $|u|_1 = |u_1| + \cdots + |u_m|$ for any $u = (u_1, \cdots, u_m)$. If we use an *implicit* time discretization scheme, we obtain the finite difference equation

$$p^{\Delta,\delta}_k(x) - \Delta \sum_{z \in N_\delta(x)} L^*_\delta(x,z)\, p^{\Delta,\delta}_k(z) = c_k \cdot \Psi^\Delta_k(x) \cdot p^{\Delta,\delta}_{k-1}(x) , \tag{24}$$

for $x \in \bar{D}^\delta$ and $k = 1, \cdots$, where c_k is a normalization constant, and the initial condition $p^{\Delta,\delta}_0(x)$ is a suitable approximation of the density $p_0(x)$. This relation can be divided into two steps

- *prediction step* : Transport by the flow gives

  $$[I - \Delta\, L^*_\delta]\, p^{\Delta,\delta}_{k-\frac{1}{2}}(x) = p^{\Delta,\delta}_{k-1}(x) .$$

- *correction step* : According to the Bayes formula

  $$p^{\Delta,\delta}_k(x) = c_k \cdot \Psi^\Delta_k(x) \cdot p^{\Delta,\delta}_{k-\frac{1}{2}}(x) ,$$

  where c_k is a normalization constant.

The following result is proved in Kushner [5] using weak convergence $\overline{X}^\delta_\cdot \Longrightarrow X_\cdot$ as $\delta \downarrow 0$.

Theorem 1 *Property* $(\star)$ *holds for the finite difference nonlinear filtering algorithm.*

A finite difference observer. To derive a finite difference algorithm, we still need to constrain the observer problem to a bounded domain. However, because we are going to approximate (18) , we must consider the ODE (5) as running backward in time, before we constrain it to the convex set $\bar{D}$. We use the same definition as above for the set $\nu(x)$ of inward unit normals. For $x \in \bar{D}$, $v \in \mathbf{R}^m$, the projection $\pi(x, v)$ of the velocity vector v at x is now given by v if $x \in D$, or $v + [\langle v, \nu^*(x, v)\rangle \vee 0]\, \nu^*(x, v)$ if $x \in \partial D$, where $\nu^*(x, v)$ is an element of $\nu(x)$ which maximizes $\langle v, \nu\rangle$, $\nu \in \nu(x)$. Define then $\tilde{b}(x) = -\pi(x, -b(x))$, $x \in \bar{D}$. By Theorem 5.1 of Dupuis–Ishii [1] again, there exists a unique absolutely continuous solution $\xi = \xi^{x,t}$ of the constrained ODE

$$\dot{\xi}_s = \tilde{b}(\xi_s) \quad \text{a.e. } 0 \le s \le t\,, \tag{25}$$

satisfying $\xi_t = x \in \bar{D}$.

Select $\beta \in C(\mathbf{R}^m)$ non-negative, $\beta \equiv 0$ on $D' \subset D$, with $D' \cap \partial D = \emptyset$, and $\beta > 0$ on ∂D. Now define the value function for $x \in \bar{D}$, $t \ge 0$ by

$$m_t(x) = \beta(\xi_0) + \int_0^t [V_s(\xi_s) + \beta(\xi_s)]\, ds\,, \tag{26}$$

where ξ is the solution of the constrained ODE (25). Then $m_t(x)$ is the unique viscosity solution of the Hamilton–Jacobi equation, see Lions [6]

$$\left\{ \begin{aligned} \frac{\partial m_t}{\partial t} + b \cdot \frac{\partial m_t}{\partial x} - V_t - \beta &= 0 \quad \text{in } D \times (0, S]\,, \\ -\nu \cdot \frac{\partial m_t}{\partial x} &= 0 \quad \text{on } \partial D \times (0, S]\,, \end{aligned} \right. \tag{27}$$

satisfying the initial condition $m_0(x) = \beta(x)$ for $x \in D$. In addition, m satisfies in the viscosity sense

$$\frac{\partial m_t}{\partial t} + \tilde{b} \cdot \frac{\partial m_t}{\partial x} - V_t - \beta = 0 \quad \text{on } \partial D \times (0, S]\,. \tag{28}$$

Define the corresponding observer as the set

$$\hat{x}_t = \underset{x \in \mathbf{R}^m}{\operatorname{argmin}}\; m_t(x) = \left\{ x \in \mathbf{R}^m \,:\, m_t(x) = 0 \right\}\,. \tag{29}$$

Let $\mathcal{I} = \{x_0 \in D' \,:\, \Phi_s(x_0) \in D',\ 0 \le s \le S\}$. If $x_0^* \in \mathcal{I}$, then $x_t^* \in \hat{x}_t$ for all $0 \le t \le S$, and the observer (29) defines a FTO provided the deterministic system ((2)) is *observable* on $[0, T]$, see James [4].

We again use a Markov chain finite difference scheme. However, we discretize the boundary equation (28) rather than the boundary condition in (27). We use the same definition as above for the grid $\mathbf{R}^m_\delta$, the system of neighborhoods $N_\delta(x)$, and the subsets $\bar{D}^\delta$, D^δ and $\partial D^\delta = \bar{D}^\delta \setminus D^\delta$ of the grid $\mathbf{R}^m_\delta$. Assume that $v = \delta/\Delta$ is a fixed real number, indicating the "speed" of the algorithm, satisfying

$$v \ge \max_{x \in D} |\tilde{b}(x)|_1\,. \tag{30}$$

We define the transition probabilities $\pi^{\Delta,\delta}(x, z) = P(\xi^{\Delta,\delta}_{k-1} = z \mid \xi^{\Delta,\delta}_k = x)$ for a *backward* Markov chain $\{\xi^{\Delta,\delta}_k,\ k = [S/\Delta], \ldots, 0\}$ by

$$\pi^{\Delta,\delta}(x, z) \;=\; \left\{ \begin{array}{ll} 1 - |\tilde{b}(x)|_1 / v & \text{if } z = x, \\[2ex] \tilde{b}_i^{\mp}(x)/v & \text{if } z = x \pm \delta e_i \text{ and } i = 1, \ldots, m \\[2ex] 0 & \text{if } z \notin N_\delta(x)\,. \end{array} \right. \tag{31}$$

Note that $\mathrm{E}[\xi^{\Delta,\delta}_{k-1} - \xi^{\Delta,\delta}_k \mid \xi^{\Delta,\delta}_k = x] = -\Delta\, \tilde{b}(x)$.

If we replace $\Phi_\Delta^{-1}(x)$ in (18) by the state $\xi^{\Delta,\delta}_{k-1}$ of the backward Markov chain starting at $\xi^{\Delta,\delta}_k = x$ and take expectations, we obtain the finite difference equation

$$m_k^{\Delta,\delta}(x) = \sum_{z\in N_\delta(x)} \pi^{\Delta,\delta}(x,z)\, m_{k-1}^{\Delta,\delta}(z) + \Delta\,[V_k(x)+\beta(x)]\ , \tag{32}$$

for $x \in \bar{D}^\delta$ and $k = 1,\dots,[S/\Delta]$, with initial condition $m_0^{\Delta,\delta}(x) = \beta(x)$. This relation can be divided into two steps

- *prediction step* : Transport by the flow gives
$$m_{k-\frac{1}{2}}^{\Delta,\delta}(x) = \pi^{\Delta,\delta}\, m_{k-1}^{\Delta,\delta}(x)\ .$$
- *correction step* : The contribution of the new observation z_k to the approximate value function is given by
$$m_k^{\Delta,\delta}(x) = m_{k-\frac{1}{2}}^{\Delta,\delta}(x) + \Delta\,[V_k(x)+\beta(x)]\ .$$

The finite difference observer set is defined by

$$\hat{x}_k^{\Delta,\delta} = \operatorname*{argmin}_{x\in D^\delta}\ m_k^{\Delta,\delta}(x)\ . \tag{33}$$

Obviously, there is no reason for this approximate observer to satisfy the non–asymptotic consistency property : in general we can not guarantee that $x^*_{t_k} \in \hat{x}_k^{\Delta,\delta}$.

The following result is proved in James [4].

Theorem 2 *If $x_0^* \in \mathcal{I}$, then property $(\star\star)$ holds for the finite difference observer algorithm.*

Remark 3 It is also shown in [4] that under additional regularity assumptions $\mathrm{dist}(\hat{x}^{\Delta,\delta}_{[t/\Delta]}, x_t^*) = O(\sqrt{\delta})$ as $\delta \downarrow 0$.

Remark 4 The speed constraint (30) which appears in the finite difference observer algorithm is actually a stability condition for the *explicit* time–discretization scheme used in (32) . From the probabilistic point of view, it ensures that (31) defines transition probabilities. We do not need a similar constraint for the finite difference NLF algorithm, because we are using there an *implicit* time–discretization scheme.

B Flow–Based Approximation

Let us first describe the approximate model we are going to use.

We assume that at each time t_k, a partition $\{B_k^i\,,\ i\in I_k^\delta\}$ of the state space $\mathbb{R}^m$ is given, and we define the discrete I_k^δ–valued state n_k^δ by the relation

$$\left\{n_k^\delta = i\right\} = \left\{X_{t_k} \in B_k^i\right\}\ . \tag{34}$$

The idea behind our approximation is to suppose that, at each step of the algorithm, any information (e.g. probability distribution, likelihood function, value function) about the continuous state X_{t_k} is immediately compressed into some information about the discrete state n_k^δ. We can think of memory constraint as a justification for this compression mechanism. As a consequence, whenever information is needed about the continuous state X_{t_k}, it has to be deduced from the corresponding information about the discrete state n_k^δ, resulting in *compression error.*

Making explicit use of the flow associated with ((5)), we have

$$\left\{\xi_{t_k}\in B_k^i\right\} = \left\{\xi_{t_{k-1}} \in \Phi_\Delta^{-1}(B_k^i)\right\} = \bigcup_{j\in I^\delta_{k-1}[i]} \left\{\xi_{t_{k-1}} \in B^j_{k-1}\cap \Phi_\Delta^{-1}(B_k^i)\right\}\ , \tag{35}$$

where

$$I^\delta_{k-1}[i] = \left\{ j \in I^\delta_{k-1} \ : \ B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k) \neq \emptyset \right\} ,$$

provided ξ solves the ODE ((5)). Notice that in general the set $I^\delta_{k-1}[i]$ has finite cardinality.

Various possible choices are available for the partitions, e.g.

- $I^\delta_k = I^\delta_{k-1} \equiv I^\delta$ and $B^i_k = \Phi_\Delta(B^i_{k-1})$ for all $i \in I^\delta$. In this case $n^\delta_k = n^\delta_{k-1}$, i.e. the discrete state process is constant over time, but the sets B^i_k can become very complicated after some steps.
- $I^\delta_k = I^\delta_{k-1} \equiv I^\delta$ and $B^i_k = B^i_{k-1}$ for all $i \in I^\delta$. In this case, the partition is constant over time, but updating the discrete state can be cumbersome.

Between these two extreme cases, a trade–off has to be found in order to reduce the computational burden of updating both the partition and the discrete probability distribution : B^i_k should both be "close" to $\Phi_\Delta(B^i_{k-1})$ and have a simple geometry.

A flow–based nonlinear filter. According to our approximation approach, we introduce the discrete *a priori* and *a posteriori* conditional probability distributions

$$\tilde{\mu}^i_{k-\frac{1}{2}} = P(X_{t_k} \in B^i_k \mid \mathcal{Y}^\Delta_{k-1}) \quad \text{and} \quad \tilde{\mu}^i_k = P(X_{t_k} \in B^i_k \mid \mathcal{Y}^\Delta_k) ,$$

respectively, where again $\mathcal{Y}^\Delta_k = \sigma(y^\Delta_1, \cdots, y^\Delta_k)$. Making use of (35) transport by the flow gives

$$\begin{aligned} \tilde{\mu}^i_{k-\frac{1}{2}} &= \sum_{j \in I^\delta_{k-1}[i]} P(X_{t_{k-1}} \in B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k) \mid \mathcal{Y}^\Delta_{k-1}) \\ &= \sum_{j \in I^\delta_{k-1}[i]} \lambda(B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k)) \cdot \frac{1}{\lambda(B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k))} \int_{B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k)} p^\Delta_{k-1}(x)\, dx \\ &\simeq \sum_{j \in I^\delta_{k-1}[i]} \lambda(B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k)) \cdot \frac{1}{\lambda(B^j_{k-1})} \int_{B^j_{k-1}} p^\Delta_{k-1}(x)\, dx \\ &\simeq \sum_{j \in I^\delta_{k-1}[i]} \tilde{\mu}^j_{k-1} \, \frac{\lambda(B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k))}{\lambda(B^j_{k-1})} . \end{aligned}$$

Next, according to the Bayes formula

$$\begin{aligned} \tilde{\mu}^i_k &= c_k \int_{B^i_k} \Psi^\Delta_k(x)\, p^\Delta_{k-\frac{1}{2}}(x)\, dx \\ &= c_k \int_{B^i_k} p^\Delta_{k-\frac{1}{2}}(x)\, dx \cdot \frac{\displaystyle\int_{B^i_k} \Psi^\Delta_k(x)\, p^\Delta_{k-\frac{1}{2}}(x)\, dx}{\displaystyle\int_{B^i_k} p^\Delta_{k-\frac{1}{2}}(x)\, dx} \\ &\simeq c_k \cdot \tilde{\mu}^i_{k-\frac{1}{2}} \cdot \max_{x \in B^i_k} \Psi^\Delta_k(x) , \end{aligned}$$

where c_k is a normalization constant. This approximation can be justified in the small noise case, using the Laplace asymptotic formula.

To obtain a computable algorithm, we introduce new discrete probability distributions $\mu^i_{k-\frac{1}{2}}$ and μ^i_k, and the corresponding densities

$$p^{\Delta,\delta}_{k-\frac{1}{2}}(x) = \mu^i_{k-\frac{1}{2}} / \lambda(B^i_k) \qquad \text{and} \qquad p^{\Delta,\delta}_k(x) = \mu^i_k / \lambda(B^i_k) \qquad \text{iff } x \in B^i_k .$$

We then *define* the transition from $\{\mu^i_{k-1}\,,\, i \in I^\delta_{k-1}\}$ to $\{\mu^i_k\,,\, i \in I^\delta_k\}$, by the following two steps

- *prediction step* : Transport by the flow gives

$$\mu^i_{k-\frac{1}{2}} = \sum_{j \in I^\delta_{k-1}[i]} \mu^j_{k-1} \, \frac{\lambda(B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k))}{\lambda(B^j_{k-1})} \, .$$

- *correction step* : The contribution of the new observation y^Δ_k is given by

$$\mu^i_k = c_k \cdot R^i_k \cdot \mu^i_{k-\frac{1}{2}} \, , \tag{36}$$

where c_k is a normalization constant, and

$$R^i_k = \max_{x \in B^i_k} \Psi^\Delta_k(x) \, ,$$

is the *generalized likelihood function* for the estimation of n^δ_k based on the observation y^Δ_k alone.

Theorem 5 *In the case* $I^\delta_k = I^\delta_{k-1} \equiv I^\delta$, *let* $\{B^i_0\,,\, i \in I^\delta\}$ *denote a finite partition of a bounded domain* D *with* $\mathrm{diam}(B^i_0) \leq \delta$. *Then property* $(\star)$ *holds for this flow-based nonlinear filtering algorithm.*

A flow-based observer. According to our approximation approach, we introduce the *a priori* and *a posteriori* discrete value functions

$$\widetilde{m}^i_{k-\frac{1}{2}} = \inf\left\{F^\Delta_{k-1}(\xi^{t_k,x}) \,:\, x \in B^i_k\right\} \quad \text{and} \quad \widetilde{m}^i_k = \inf\left\{F^\Delta_k(\xi^{t_k,x}) \,:\, x \in B^i_k\right\} \, , \tag{37}$$

respectively. Making use of ((35)) transport by the flow gives

$$\widetilde{m}^i_{k-\frac{1}{2}} = \inf_{j \in I^\delta_{k-1}[i]} \inf\left\{F^\Delta_{k-1}(\xi^{t_{k-1},x}) \,:\, x \in B^j_{k-1} \cap \Phi^{-1}_\Delta(B^i_k)\right\} \geq \inf_{j \in I^\delta_{k-1}[i]} \widetilde{m}^j_{k-1} \, .$$

Next, by definition of the functional $F^\Delta_k(\xi)$,

$$\widetilde{m}^i_k = \inf\left\{F^\Delta_{k-1}(\xi^{t_k,x}) + \Delta\, V_k(x) \,:\, x \in B^i_k\right\} \geq \widetilde{m}^i_{k-\frac{1}{2}} + \Delta \inf_{x \in B^i_k} V_k(x) \, .$$

Thus the discrete value functions satisfy difference *inequalities*. Unfortunately, this does not give a recursive mechanism for computation. Instead, we introduce new discrete value functions $m^i_{k-\frac{1}{2}}$ and m^i_k, and the corresponding value functions

$$m^{\Delta,\delta}_{k-\frac{1}{2}}(x) = m^i_{k-\frac{1}{2}} \qquad \text{and} \qquad m^{\Delta,\delta}_k(x) = m^i_k \qquad \text{iff } x \in B^i_k \, .$$

We then *define* the transition from $\{m^i_{k-1}\,,\, i \in I^\delta_{k-1}\}$ to $\{m^i_k\,,\, i \in I^\delta_k\}$ by the following two steps

- *prediction step* : Transport by the flow gives

$$m^i_{k-\frac{1}{2}} = \inf_{j \in I^\delta_{k-1}[i]} m^j_{k-1} \, .$$

- *correction step :* The contribution of the new observation z_k is given by

$$m^i_k = m^i_{k-\frac{1}{2}} + \Delta \inf_{x \in B^i_k} V_k(x) \, .$$

By construction, it is clear that $m_k^{\Delta,\delta}(x) \geq 0$ and $m_k^{\Delta,\delta}(x^*_{t_k}) = 0$ for the *true* state trajectory, and we define our observer as the set

$$\hat{x}_k^{\Delta,\delta} = \operatorname*{argmin}_{x \in \mathbf{R}^m} m_k^{\Delta,\delta}(x) = \left\{ x \in \mathbf{R}^m \, : \, m_k^{\Delta,\delta}(x) = 0 \right\} , \tag{38}$$

or equivalently

$$\hat{x}_k^{\Delta,\delta} = \bigcup_{i \in \hat{I}_k^\delta} B_k^i \quad \text{with} \quad \hat{I}_k^\delta = \left\{ i \in I_k^\delta \, : \, m_k^i = 0 \right\} .$$

By an inductive comparison argument, it is easy to show that $m_k^{\Delta,\delta}(x) \leq m_k^\Delta(x)$, with the consequence that $\hat{x}_k^\Delta \subset \hat{x}_k^{\Delta,\delta}$. Therefore, $x^*_{t_k} \in \hat{x}_k^{\Delta,\delta}$.

Theorem 6 *In the case* $I_k^\delta = I_{k-1}^\delta \equiv I^\delta$ *and* $B_k^i = \Phi_\Delta(B_{k-1}^i)$ *for all* $i \in I^\delta$*, let* $\{B_0^i \, , \, i \in I^\delta\}$ *denote a finite partition of a bounded domain* D *with* $\operatorname{diam}(B_0^i) \leq \delta$*. If* $x_0^* \in D$*, then property* $(\star\star)$ *will hold for this flow–based observer algorithm.*

As noticed in James [3], the only thing that matters is the argmin set, not the value function itself. This remark can be used to design a simplified algorithm for the construction of the set–valued observer ((38)). We introduce the piecewise–constant logical value functions $\overline{m}_k^{\Delta,\delta}(x)$ taking values TRUE or FALSE, and defined iteratively by the following relations

$$\overline{m}_{k-\frac{1}{2}}^i = \bigvee_{j \in I_{k-1}^\delta[i]} \overline{m}_{k-1}^j ,$$

$$\overline{m}_k^i = \overline{m}_{k-\frac{1}{2}}^i \wedge \overline{V}_k^i ,$$

where

$$\overline{V}_k^i = \begin{cases} \text{TRUE} & \text{if } \inf\limits_{x \in B_k^i} V_k(x) = 0 \\ \\ \text{FALSE} & \text{otherwise} \end{cases}$$

It is clear that $\overline{m}_k^i = \text{TRUE}$ iff $m_k^i = 0$, so that an equivalent expression for the set–valued observer ((38)) is given by

$$\hat{x}_k^{\Delta,\delta} = \bigcup_{i \in \hat{I}_k^\delta} B_k^i \quad \text{with} \quad \hat{I}_k^\delta = \left\{ i \in I_k^\delta \, : \, \overline{m}_k^i = \text{TRUE} \right\} .$$

Corollary 7 *Under the assumptions of Theorem 6, property* $(\star\star)$ *will hold for the simplified algorithm.*

Remark 8 In the particular case where $I_k^\delta = I_{k-1}^\delta \equiv I^\delta$ and $B_k^i = \Phi_\Delta(B_{k-1}^i)$ for all $i \in I^\delta$, the algorithms exhibit a parallel structure explicitly. On the other hand, these algorithms assume that certain calculations can be made exactly. This is not always possible, in which case one would have e.g. to discretize the ODE (5) or use the following approximations

$$\frac{1}{\lambda(B_k^i)} \int_{B_k^i} p(x) \, dx \simeq p(x_k^i) \qquad \max_{x \in B_k^i} \Psi_k^\Delta(x) \simeq \Psi_k^\Delta(x_k^i) ,$$

$$\inf_{x \in B_k^i} m(x) \simeq m(x_k^i) \qquad \inf_{x \in B_k^i} V_k(x) \simeq V_k(x_k^i) ,$$

where x_k^i is some point in B_k^i.

IV Numerical Experiments

A A One Dimensional Example

We consider a one dimensional model with

$$b(x,t) = -0.2x + 0.8\cos(2.5t) \qquad h(x) = \operatorname{sgn}(x)\ .$$

Even though the observation function is *discontinuous*, the convergence results are still valid, see James [4]. The location of the trajectory is determined at the first time t^* it crosses the origin, so the system is observable.

Figure 1 (below) shows results for the simplified (logical) flow-based observer algorithm, with the choice $I_k^\delta = I_{k-1}^\delta \equiv I^\delta$ and $B_k^i = \Phi_\Delta(B_{k-1}^i)$ for all $i \in I^\delta$, $\Delta = 0.05$, $\delta = 0.02$, and noise-free observations. The estimate $\hat{x}_t$ is a one-dimensional set for times t before t^*, and zero-dimensional after this time.

Figure 2 illustrates the numerical results obtained from the finite difference nonlinear filter algorithm. Here, $\Delta = 0.05$, $\delta = 0.005$, $R = 10^{-4}$, and the observation path was noise-free. Notice the jumps in the conditional mean trajectory and the peaking of the conditional density function each time the origin is crossed. Numerical viscosity causes the density to spread between these times.

Figure 3 shows results for the finite difference observer algorithm, with $\delta = 0.02$, $\Delta = 0.0198$, $v = 1.01$, and noise-free observations. The plot of the value function clearly shows the valley containing the state trajectory.

Figure 4 shows results for the flow-based nonlinear filtering algorithm, with the choice $I_k^\delta = I_{k-1}^\delta \equiv I^\delta$ and $B_k^i = \Phi_\Delta(B_{k-1}^i)$ for all $i \in I^\delta$, $\Delta = 0.05$, $\delta = 0.02$, $R = 10^{-4}$, and noise-free observations. Marginals for the conditional density are shown for times before and after time t^*.

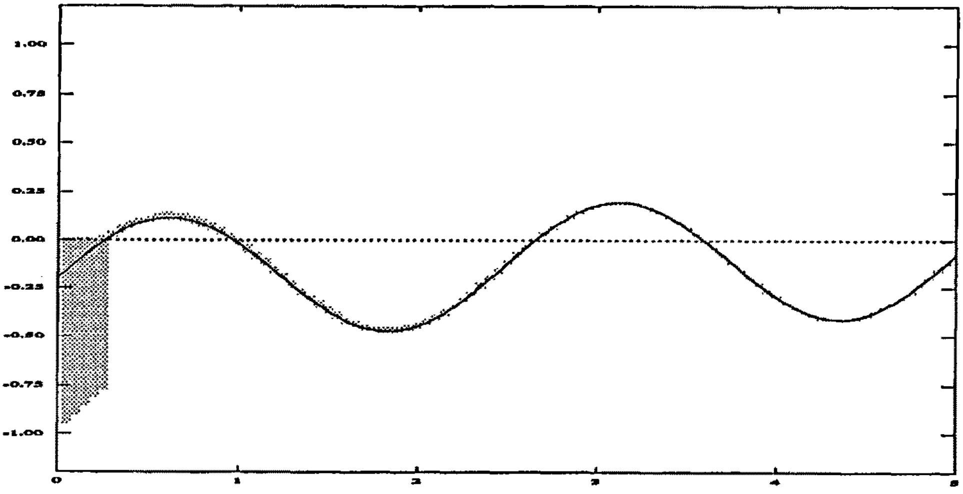

Figure 1. Flow-based observer, simplified algorithm.
State x_t and estimate $\hat{x}_t$ trajectories.

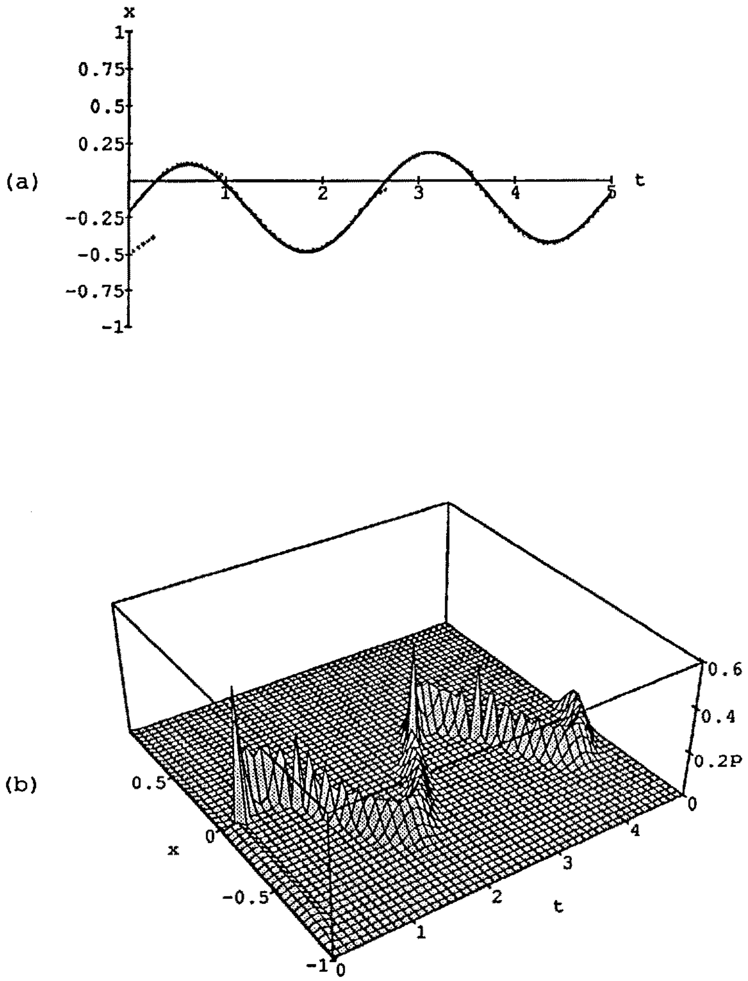

Figure 2. Finite difference nonlinear filter.
(a) State x_t and conditional mean $\mathbf{E}[x_t|\mathcal{Y}_t]$ trajectories; (b) Conditional density function.

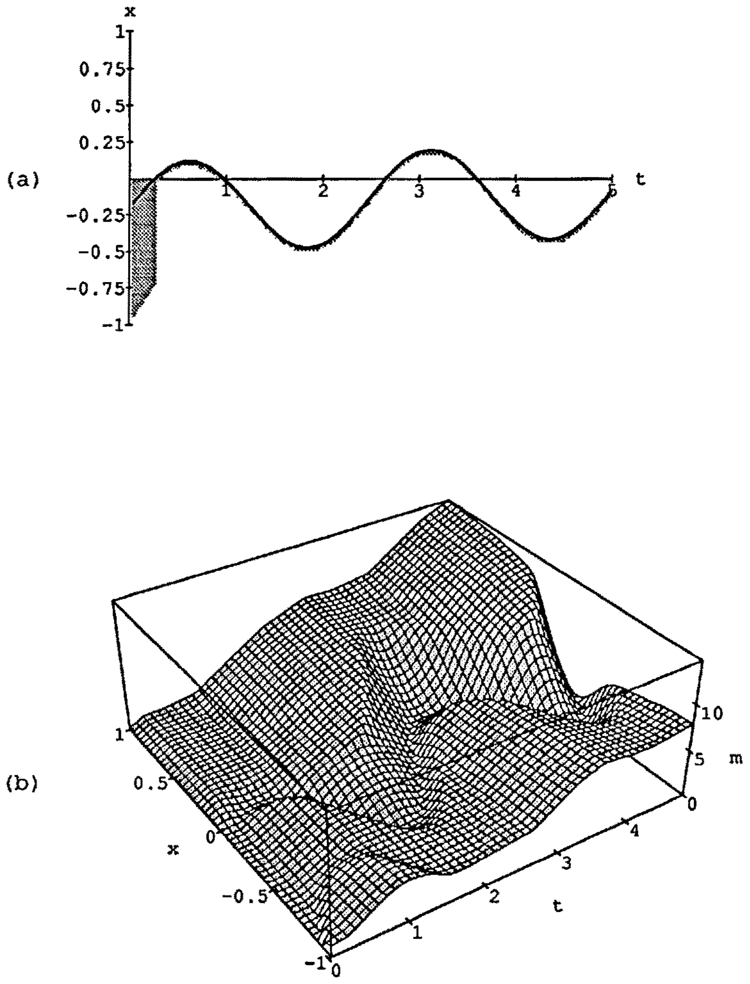

Figure 3. Finite difference observer.
(a) State x_t and estimate $\hat{x}_t$ trajectories; (b) Value function.

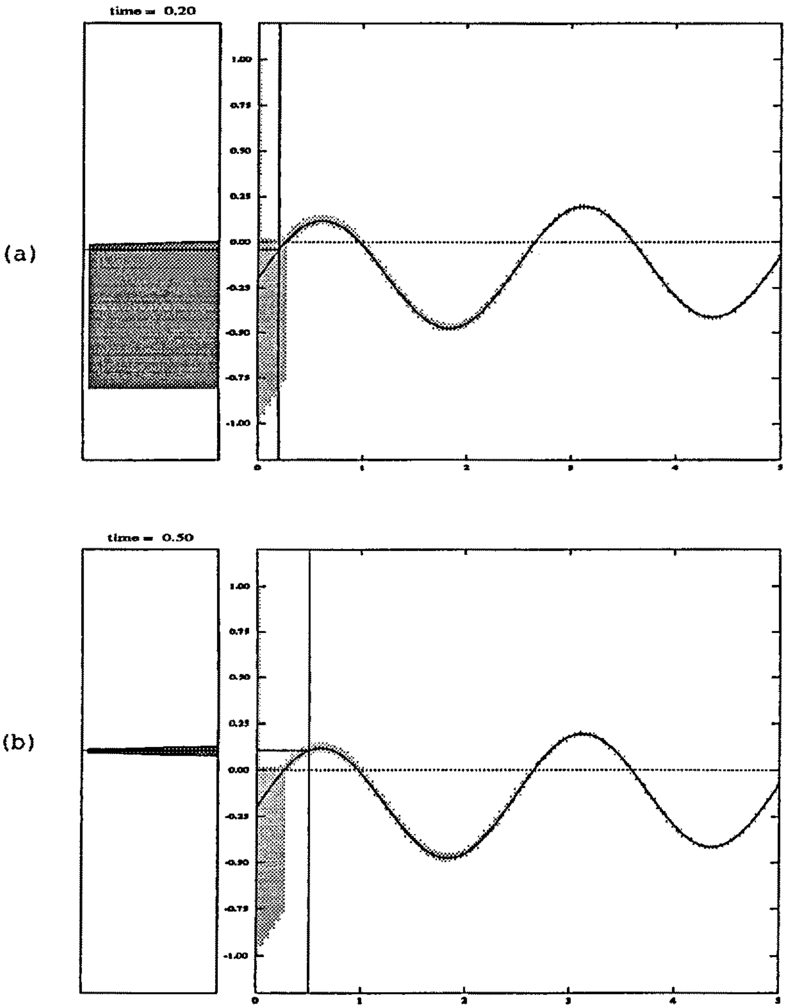

Figure 4. Flow-based nonlinear filter.
(a) State x_t trajectory, 90% confidence region, density at $t = 0.2$;
(b) State x_t trajectory, 90% confidence region, density at $t = 0.5$.

B A Four Dimensional Example

We consider here the problem of target motion analysis, which is to estimate the trajectory (position and velocity) of a target moving at constant speed along a straight line at the surface of the sea. We suppose that bearings-only measurements are available in discrete time, taken from a moving observation platform. If the observation platform itself moves at constant speed along a straight line, the problem is *non-observable*. However, as soon as the observation platform changes its course, the problem becomes *observable*. Assuming that the direction of motion of the target is known, which is true in the case of perfect observations, we can reduce the problem to three dimensions. The state vector is $X = (x, y, v)$ and the state equation

$$\dot{x}_t = v_t \qquad \dot{y}_t = 0 \qquad \dot{v}_t = 0 \,.$$

The observation function is

$$h(x,y,v,t) = \arctan[\, \frac{x - x_t^P}{y - y_t^P} \,] \,,$$

where (x_t^P, y_t^P) is the (known) position of the observation platform at time t.

For this problem, the flow is known explicitly, and the flow-based algorithms (for both the nonlinear filtering and the observer case) are explicitly parallelizable. A variant of the flow-based NLF algorithm has been implemented at INRIA on a 16K Connection Machine from Thinking Machines Corporation. Numerical experiments have been carried out, using noisy observations with standard deviations ranging between one and five degrees. The goal is to find better maneuvers, and to investigate them off-line. The filter is not intended to be run in real-time on the ship.

Acknowledgment: Research supported by NSF Grant "USA-France (INRIA) Collaborative Research in Stochastic Control" NSF-INT-89-06965.

References

[1] P. DUPUIS and H. ISHII, On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications, *Stochastics and Stochastics Reports* 35 (1+2) 31-62 (1991).

[2] M.R. JAMES, *Asymptotic Nonlinear Filtering and Large Deviations with Application to observer Design*, Ph.D. Dissertation, University of Maryland (SRC Technical Report Ph.D. 88-1) (1988).

[3] M.R. JAMES, Finite time observer design by probabilistic-variational methods, to appear, *SIAM Journal on Control and Optimization* 29 (4) (1991).

[4] M.R. JAMES, A numerical method for finite time observers, preprint (1990).

[5] H.J. KUSHNER, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977.

[6] P.L. LIONS, Neumann type boundary conditions for Hamilton-Jacobi equations, *Duke Math. J.* 52 (3) 793-820 (1985).