Sequential Bayesian Inference for Dynamical Systems Using the Finite Volume Method

Colin Fox, Richard A. Norton, Malcolm E. K. Morrison and Timothy C. A. Molteno

Abstract Optimal Bayesian sequential inference, or filtering, for the state of a deterministic dynamical system requires simulation of the Frobenius-Perron operator, that can be formulated as the solution of an initial value problem in the continuity equation on filtering distributions. For low-dimensional, smooth systems the finite-volume method is an effective solver that conserves probability and gives estimates that converge to the optimal continuous-time values. A Courant–Friedrichs–Lewy condition assures that intermediate discretized solutions remain positive density functions. We demonstrate this finite-volume filter (FVF) in a simulated example of filtering for the state of a pendulum, including a case where rank-deficient observations lead to multi-modal probability distributions.

1 Introduction

In 2011, one of us (TCAM) offered to improve the speed and accuracy of the Tru-Test scales for 'walk over weighing' (WOW) of cattle, and wagered a beer on the outcome [8]. Tru-Test is a company based in Auckland, New Zealand, that manufactures measurement and productivity tools for the farming industry, particularly for dairy. Tru-Test's XR3000 WOW system was already in the market, though they

Colin Fox
University of Otago, Dunedin, New Zealand, e-mail: fox@physics.otago.ac.nz

Richard A. Norton
University of Otago, Dunedin, New Zealand, e-mail: richard.norton@otago.ac.nz

Malcolm E. K. Morrison
University of Otago, Dunedin, New Zealand, e-mail: malcolmerikking@gmail.com

Timothy C. A. Molteno
University of Otago, Dunedin, New Zealand, e-mail: tim@physics.otago.ac.nz

could see room for improvement in terms of on-farm usage, as well as speed and accuracy. Indeed, advertising material for the XR3000 stated that WOW

> requires that the animals pass the platform regularly and smoothly

which hinted at the existing processing requiring somewhat constrained movement by the cows for it to deliver a weight.

Figure 1 shows a cow walking over the weigh-bridge in the WOW system located on the ground in the path from a milking shed. The weigh bridge consists of a low platform with strain gauges beneath the platform, at each end, that are used to measure a time series of downward force from which weight (more correctly, mass) of the cow is derived.

Fig. 1 A dairy cow walking over a weigh bridge placed near the milking shed. (Photo-credit: Wayne Johnson/Pizzini Productions)

The plan, for improving estimates of cow mass from strain-gauge time series, was to apply Bayesian modeling and computational inference. Bayesian inference allows uncertain measurements to be modeled in terms of probability distributions, and interpreted in terms of physical models that describe how the data is produced. This leads to estimates of parameters in the model, such as the mass of a cow, and meaningful uncertainty quantification on those estimates. At the outset we developed dynamical-systems models for the moving cow, with some models looking like one or more pogo sticks. Operation in real-time would require developing new algorithms for performing the inference sequentially – as the data arrives – and new hardware with sufficient computing speed to implement those algorithms. Figure 2 (right) shows hardware developed for this application, that includes strain-gauge signal conditioning, digitization, and an embedded ARM processor, alongside the XR3000 electronics and display (left).

This paper describes an algorithm for optimal sequential Bayesian inference that we developed in response to this application in cow weighing. We first give a stylized model of WOW, then a method for optimal filtering for tracking the state of a nonlinear dynamical system, then present numerical examples for the stylized model.

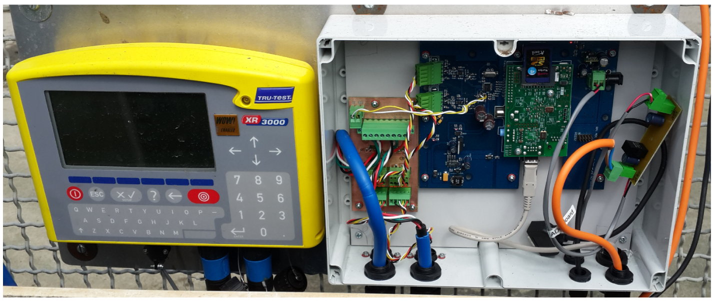

Fig. 2 WOW hardware in 2016: existing commercial unit (left), and prototype with embedded processing (right). (Photocredit and hardware design: Phill Brown)

1.1 A Stylized Problem

A (very) stylized model of WOW is the problem of tracking a simple pendulum of length l and mass m when only the force F in the string is measured, as depicted in Figure 3. For this system the kinematic variables are the angular displacement (from

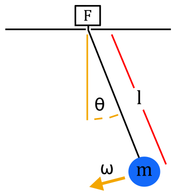

Fig. 3 A pendulum of length l with mass m, undergoing motion with angular displacement θ and angular velocity ω. The force F in the string is measured.

the vertical downwards) θ and the angular velocity ω. The kinematic state (θ, ω) evolves according to

$$\frac{\mathrm{d}}{\mathrm{d}t}(\theta, \omega) = \left(\omega, -\frac{g}{l}\sin\theta\right)$$

where g is the acceleration due to gravity, l the length of the pendulum.

The force F_k is measured at times t_k, $k = 1, 2, 3, \ldots$, with the (noise free) value related to the state variables by

$$F_k = ml\omega^2(t_k) + mg\cos\theta(t_k).$$

Estimation of the parameter m may be performed by considering the augmented system

$$\frac{\mathrm{d}}{\mathrm{d}t}(\theta, \omega, m) = \left(\omega, -\frac{g}{l}\sin\theta, 0\right).$$

See [7] for a computed example of parameter estimation in this system.

2 Sequential Bayesian Inference for Dynamical Systems

Consider now a general dynamical system that evolves according to the (autonomous) differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}x = f(x), \tag{1}$$

where f is a known velocity field and $x(t)$ is the state vector of the system at time t. Given an initial state $x(0) = x_0$ at $t = 0$, Eq. 1 may be solved to determine the future state $x(t)$, $t > 0$, that we also write $x(t;x_0)$ to denote this deterministic solution.

At increasing discrete times t_k, $k = 1,2,3,\ldots$, the system is observed, returning measurement z_k that provides noisy and incomplete information about $x_k = x(t_k)$. We assume that we know the conditional distribution over observed value z_k, given the state x_k,

$$\rho\left(z_k|x_k\right).$$

Let $Z_t = \{z_k : t_k \leq t\}$ denote the set of observations up to time t, and let (the random variable) $x_t = x(t)$ denote the unknown state at time t. The formal Bayesian solution corresponds to determining the time-varying sequence of filtering distributions

$$\rho\left(x_t|Z_t\right) \tag{2}$$

over the state at time t conditioned on all available measurements to time t.

Discrete-time formulation A standard approach [1] is to discretize the system equation 1 and treat the discrete-time system [3]. When uncertainty in f is included via 'process noise' v_k, observation errors via 'observation noise' n_k, the discrete-time problem is written as

$$\begin{aligned} x_k &= f_k\left(x_{k-1}, v_k\right) \\ z_k &= h_k\left(x_k, n_k\right) \end{aligned}$$

with functions f_k and h_k assumed known.

When the random processes v_k and n_k are independently distributed from the current and previous states, the system equation defines a Markov process, as does Eq. 1, while the observation equation defines the conditional probability $\rho\left(z_k|x_k\right)$.

We will treat the continuous-time problem directly, defining a family of numerical approximations that converge in distribution to the desired continuous-time distributions.

Continuous-time Bayesian filtering Sequential Bayesian inference iterates two steps to generate the filtering distributions in Eq. 2 [5].

Prediction Between measurements times t_k and t_{k+1}, Z_t is constant and the continuous-time evolution of the filtering distribution may be derived from the (forward) Chapman-Kolmogorov equation

$$\begin{aligned}\rho\left(x_{t+\Delta t}|Z_{t+\Delta t}\right) = \rho\left(x_{t+\Delta t}|Z_t\right) &= \int \rho\left(x_{t+\Delta t}|x_t, Z_t\right)\rho\left(x_t|Z_t\right)\,\mathrm{d}x_t \\ &= \int \delta(x_{t+\Delta t} - x(\Delta t; x_t))\rho\left(x_t|Z_t\right)\,\mathrm{d}x_t,\end{aligned}$$

which defines a linear operator on the space of probability distributions,

$$S_{\Delta t} : \rho\left(x_t|Z_t\right) \mapsto \rho\left(x_{t+\Delta t}|Z_t\right). \tag{3}$$

$S_{\Delta t}$ is the Frobenius-Perron (or Foias) operator for time increment Δt.

Update At measurement times t_k, Z_t changes, from Z_{k-1} to Z_k, and the filtering distribution changes, typically discontinuously, as

$$\rho\left(x_k|Z_k\right) = \frac{\rho\left(z_k|x_k\right)\rho\left(x_k|Z_{k-1}\right)}{\rho\left(z_k|Z_{k-1}\right)}, \tag{4}$$

which is simply Bayes' rule written at observation time t_k. We have written $x_k = x_{t_k}$ and $Z_k = Z_{t_k}$, and used conditional independence of z_k and Z_{k-1} given x_k.

2.1 The Frobenius-Perron Operator is a PDE

The Frobenius-Perron operator in Eq. 3, that evolves the filtering density forward in time, may be written as the solution of an initial value problem (IVP) in a partial differential equation (PDE) for the probability density function (pdf).

For pdf $\rho(x;t)$ over state x and depending on time t, the velocity field $f(x)$ implies a flux of probability equal to $\rho(x;t)f(x)$. Fig. 4 shows a schematic of the pdf and probability flux in region $(x, x+\mathrm{d}x)$, and for the time interval $(t, t+\mathrm{d}t)$. Equating the

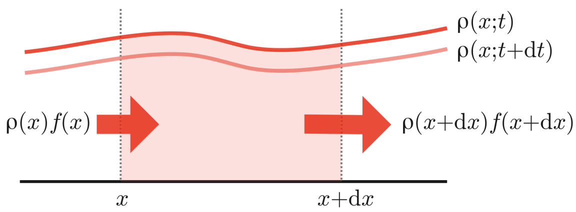

Fig. 4 A schematic of probability flux in region $(x, x+\mathrm{d}x)$, and for time $(t, t+\mathrm{d}t)$. The schematic shows greater flux exiting the region than entering, correspondingly the pdf at $t+\mathrm{d}t$ is decreased with respect to the pdf at t.

rate of change in the pdf with the rate at which probability mass enters the region, and taking $\mathrm{d}x, \mathrm{d}t \to 0$, gives the continuity equation

$$\frac{\partial}{\partial t}\rho = -\nabla\cdot(\rho f). \tag{5}$$

The Frobenius-Perron operator $S_{\Delta t}$, for time interval Δt, may be simulated by solving the PDE 5 with initial condition $\rho(x;0) = \rho(x_t|Z_t)$ to evaluate $\rho(x;\Delta t) = \rho(x_{t+\Delta t}|Z_t)$.

Eq. 5 is a linear advection equation. When the state equation has additive stochastic forcing, as is often used to represent model error, evolution of the filtering pdf is governed by a linear advection-diffusion (Fokker-Planck) equation.

3 Finite Volume Solver

The finite volume method (FVM) discretizes the continuity equation in its integral form, for each 'cell' K in a mesh,

$$\frac{\partial}{\partial t}\int_K \rho \,\mathrm{d}x + \oint_{\partial K} \rho\,(f\cdot\hat{n})\,\mathrm{d}S = 0.$$

Write $L \sim K$ if cells L and K share a common interface, denoted E_{KL}, and denote by $\hat{n}_{KL}$ the unit normal on E_{KL} directed from K to L. Define the initial vector of cell values by $P_K^0 = \frac{1}{|K|}\int_K \rho(x;0)\,\mathrm{d}x$ then for $m = 0, 1, \cdots, r$ compute P^{m+1} as

$$\frac{P_K^{m+1} - P_K^m}{\Delta t} + \frac{1}{|K|}\sum_{L\sim K} f_{KL}P_{KL}^m = 0,$$

where

$$f_{KL} = \int_{E_{KL}} f\cdot\hat{n}_{KL}\,\mathrm{d}S \qquad \text{and} \qquad P_{KL}^m = \begin{cases} P_K^m & \text{if } f_{KL} \geq 0 \\ P_L^m & \text{if } f_{KL} < 0 \end{cases}$$

is the normal velocity on E_{KL} and first-order *upwinding* scheme, respectively.

In matrix form, the FVM step for time increment Δt is

$$P^{m+1} = (I - \Delta t A)\,P^m,$$

where I is the identity matrix and A is a sparse matrix defined above. This formula is essentially Euler's famous formula for the (matrix) exponential.

Since $f_{KL} = -f_{LK}$, the FVM conserves probability at each step, i.e., $\sum_K |K|P_K^{m+1} = \sum_K |K|P_K^m$. The FVM also preserves positivity of the pdf when the time step Δt is small enough that the matrix $I - \Delta t A$ has all non-negative entries. It is straightforward to show that positive entries of the matrix A can occur on the diagonal, only. Hence, the Courant–Friedrichs–Lewy (CFL) type condition, that assures that the FVM iteration is positivity preserving, may be written

$$\Delta t \leq \frac{1}{\max_i A_{ii}}. \tag{6}$$

With this condition, the FVM both conserves probability and is positivity preserving, hence is a (discrete) *Markov operator*. In contrast, the numerical method for the matrix exponential in MATLAB, for example, does not preserve positivity for the class of matrices considered here.

4 Continuous-Time Frobenius-Perron Operator and Convergence of the FVM Approximation

In this section we summarize results from [7], to develop some analytic properties of the continuous-time solution, and establish properties and distributional convergence of the numerical approximations produced by the FVM solver.

Let $X(\cdot,t) : \mathbb{R}^d \to \mathbb{R}^d$ denote the map from initial condition to the solution of Eq. 1 at time $t \geq 0$, and $Y(\cdot,t) = X(\cdot,t)^{-1}$. If $X(\cdot,t)$ is *non-singular* ($|Y(E,t)| = 0$ if $|E| = 0$ $\forall$ Borel subsets $E \subset \mathbb{R}^d$), then $\forall t \geq 0$, the associated Frobenius-Perron operator [6] $S_t : L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ is defined by

$$\int_E S_t \rho \,\mathrm{d}x = \int_{Y(E,t)} \rho \,\mathrm{d}x \qquad \forall \text{ Borel subsets } E \subset \mathbb{R}^d.$$

Given an initial pdf p_0, the pdf $p(\cdot;t)$ at some future time, $t > 0$, may be computed by solving (see, e.g., [6, Def. 3.2.3 and §7.6])

$$\begin{aligned} \tfrac{\partial}{\partial t} p + \operatorname{div}(f p) &= 0 \qquad && \forall x \in \mathbb{R}^d, t > 0 \\ p(x;0) &= p_0(x) \qquad && \forall x \in \mathbb{R}^d \end{aligned} . \tag{7}$$

Then, $\forall t \geq 0$, the Frobenius-Perron operator $S_t : L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ is defined such that for any $\rho \in L^1(\mathbb{R}^d)$,

$$S_t \rho := p(\cdot;t),$$

where p is a solution to the IVP 7 with $p_0 = \rho$. Existence of a Frobenius-Perron operator and (weak) solutions to the IVP depends on the regularity of f.

Definition 1. (Definition 3.1.1. in [6]) A linear operator $S : L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ is a *Markov operator* (or satisfies the *Markov property*) if for any $f \in L^1(\mathbb{R}^d)$ such that $f \geq 0$,

$$Sf \geq 0 \qquad \text{and} \qquad \|Sf\|_{L^1(\mathbb{R}^d)} = \|f\|_{L^1(\mathbb{R}^d)}.$$

If f has continuous first order derivatives and solutions to Eq. 1 exist for all initial points $x_0 \in \mathbb{R}^d$ and all $t \geq 0$ then the Frobenius-Perron operator is well-defined, satisfies the *Markov property*, and $\{S_t : t \geq 0\}$ defines a continuous semigroup of Frobenius-Perron operators.

FVM Approximation For computational purposes it is necessary to numerically approximate the Frobenius-Perron operators. We use piece-wise constant function approximations on a mesh and the FVM.

Define a mesh $\mathscr{T}$ on $\mathbb{R}^d$ as a family of bounded, open, connected, polygonal, disjoint subsets of $\mathbb{R}^d$ such that $\mathbb{R}^d = \cup_{K\in\mathscr{T}}\overline{K}$. We assume that the common interface between two cells is a subset of a hyperplane of $\mathbb{R}^d$, and the mesh is *admissible*, i.e.,

$$\exists\alpha > 0 : \begin{cases} \alpha h^d \leq |K| \\ |\partial K| \leq \frac{1}{\alpha} h^{d-1} \end{cases} \qquad \forall K \in \mathscr{T}$$

where $h = \sup\{\mathrm{diam}(K) : K \in \mathscr{T}\}$, $|K|$ is the d-dimensional Lebesgue measure of K, and $|\partial K|$ is the $(d-1)$-dimensional Lebesgue measure of ∂K.

We will use superscript h to denote numerical approximations (though, strictly, we should use $\mathscr{T}$ as h does not uniquely define the mesh).

The following gives the CFL condition for the (unstructured) mesh $\mathscr{T}$. Suppose that for some $\xi \in [0,1)$ and $c_0 \geq 0$, we say that Δt satisfies the CFL condition if

$$\Delta t \sum_{L\sim K} \max\{0, f_{KL}\} \leq (1-\xi)|K| \qquad \forall K \in \mathscr{T} \text{ and } \Delta t \leq c_0 h. \tag{8}$$

Lemma 1. *If Δt satisfies the CFL condition in Eq. 8 and $p_0 \geq 0$ then*

$$p^h(x;t) \geq 0 \qquad \forall x \in \mathbb{R}^d, t > 0,$$

and S_t^h is a Markov operator.

The following theorems establish convergence of solutions of the FVM, and convergence of expectations with respect to the filtering distributions.

Theorem 1. *Suppose* $\mathrm{div}\, f = 0$*, $\rho \in BV(\mathbb{R}^d)$, and Δt satisfies the CFL condition for some $\xi \in (0,1)$. Then $\forall t \geq 0$,*

$$\|S_t\rho - S_t^h\rho\|_{L^1(\mathbb{R}^d)} \leq C\xi^{-1}\|\rho\|_{TV}(t^{1/2}h^{1/2} + \xi^{1/2}th).$$

Convergence of expectations is a consequence of convergence of our FVM.

Theorem 2. *Suppose $H, T < \infty$. Under the same assumptions as previous Theorem, if:*

1. *$g \in L^\infty(\mathbb{R}^d)$, or*
2. *$g \in L^\infty_{\mathrm{loc}}(\mathbb{R}^d)$ and ρ has compact support,*

then there exists a constant C independent of h and t such that

$$\left|\mathrm{E}_{S_t^h\rho}[g] - \mathrm{E}_{S_t\rho}[g]\right| \leq Ch^{1/2} \qquad \forall t \in [0,T], h \in (0,H].$$

This guarantees convergence in distribution of the discrete approximation to the continuous-time filtering pdfs in the limit $h \to 0$.

In numerical tests ([7]) we found convergence to be $\mathscr{O}(h)$, which is twice the order predicted by Theorem 2. Since the CFL condition requires the time step is also $\mathscr{O}(h)$, the method is $\mathscr{O}(\Delta t)$ accurate. Thus the FVM method we use achieves the highest order permitted by the meta theorem of Bolley and Crouzeix [2], that positivity-preserving Runge-Kutta methods can be first order accurate, at most.

5 Computed Examples

We now present a computed example of the finite volume filter (FVF) that uses the FVM for implementing the Frobenius-Perron operator during the prediction phase, and the update rule, Eq. 4 evaluated using mid-point quadrature.

Further details of these numerical experiments can be found in [4], including comparison to filtering produced by the unscented Kalman filter (UKF).

5.1 FVF Tracking of a Pendulum From Measured Force

Fig. 5 shows four snapshots of the filtering pdf for the stylized model of Sec. 1.1. The 'true' pendulum was simulated with initial condition $(\theta_0, \omega_0) = (0.2\pi, 0)$. Eight measured values of the force were recorded, per 2π time, over time period of 3π, with added Gaussian noise having $\sigma = 0.2$. The FVF was initialized with $N(0, 0.8^2 I)$.

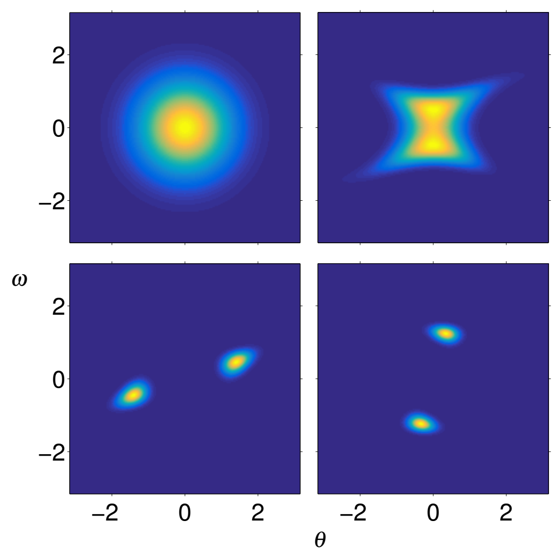

Fig. 5 Initial ($t = 0$) and filtered pdfs in phase-space after measurements at times $t = \pi/4$, π, and 3π (left to right, top to bottom).

Since the initial and filtering pdfs are symmetric about the origin, the means of angular displacement and velocity are always identically zero. Hence, filtering methods that such as the UKF, or any extension of the Kalman filter that assumes Gaussian pdfs, or that focus on the mean as a 'best' estimate, will estimate the state as identically zero, for all time. Clearly, this is uninformative.

In contrast, the FVF has localized the true state after 3π time (about 1.5 periods), albeit with ambiguity in sign. Properties of the system that do not depend on the sign of the state, such as the period, the length of the pendulum, or the mass of the pendulum, can then be accurately estimated. The computed example in [7] shows that the length of the pendulum is correctly determined after just 1.5, and that the accuracy of the estimate improves with further measurements (and time). The same feature holds for estimating mass. Hence, the FVM is successful in accurately estimating the parameters in the stylized model, even when the measurements leave ambiguity in the kinematic state.

6 Conclusions

Bayes-optimal filtering for a dynamical system requires solving a PDE. It is interesting to view filters in terms of the properties of the implicit, or explicit, PDE solver, such as the density function representation and technology used in the PDE solver. This paper develops a FVM solver using the simplest-possible discretization to implement a Bayes-optimal filter, that turned out to be computationally feasible for low-dimensional smooth systems.

The reader may be interested to know that TCAM won his wager, and beer, with new sequential inference algorithms now producing useful results on the farm. In the interests of full disclosure we should also report that the original notion of utilizing a simple dynamical-systems model for a walking cow did not perform well, as the model 'cow' would eventually walk upside down, just as the pendulum prefers to hang downwards. In response, we developed models for cow locomotion based on energy conservation, that are beyond the scope of this paper. However, the FVF has found immediate application in other dynamic estimation problems where a dynamical model that evolves a state vector works well, such as estimating the equilibrium temperature of milk during steaming as a tool for training coffee baristas.

References

1. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Proces. **50**(2), 174–188 (2002)
2. Bolley, C., Crouzeix, M.: Conservation de la positivité lors de la discrétisation des problémes d'évolution paraboliques. R.A.I.R.O. Analyse Numérique **12**, 237–245 (1978)
3. Cappé, O., Moulines, E., Ryden, T.: Inference in Hidden Markov Models. Springer Series in Statistics. Springer, New York (2005). DOI 10.1007/0-387-28982-8
4. Fox, C., Morrison, M.E.K., Norton, R.A., Molteno, T.C.A.: Optimal non-linear filtering using the finite volume method. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. (2017). In the press
5. Jazwinski, A.H.: Stochastic processes and filtering theory. Acad. Press, New York (1970)
6. Lasota, A., Mackey, M.C.: Chaos, fractals, and noise, *Applied Mathematical Sciences*, vol. 97, second edn. Springer-Verlag, New York (1994). DOI 10.1007/978-1-4612-4286-4
7. Norton, R.A., Fox, C., Morrison, M.E.K.: Numerical approximation of the Frobenius-Perron operator using the finite volume method. SIAM J. Numer. Anal. (2017). In the press
8. Treblicock, K.: Building better scales – for a beer. N. Z. Dairy Exporter **November 2011**, 101 (2011)