PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Finite-Dimensional Filters

Stephen Maybank

Phil. Trans. R. Soc. Lond. A 1996 **354**, 1099-1123
doi: 10.1098/rsta.1996.0042

Finite-dimensional filters

By Stephen Maybank

Department of Computer Science, The University of Reading, PO Box 225, Whiteknights, Reading, Berkshire RG6 6AY, UK

Contents

The class of optimal nonlinear finite-dimensional recursive filters found by Beneš is extended to include cases in which the drift in the state propagation equation is a general linear function plus the gradient of a scalar potential.

It is shown that if the state space is one dimensional, then the deterministic systems underlying the Beneš filter fall into five classes, depending on the asymptotic behaviour of the state at large times. Only two of these classes can be obtained using the Kalman filter. It is shown that an arbitrary deterministic trajectory can be approximated at small times to an accuracy of $O(t^5)$ by a trajectory for which the Beneš filter is appropriate.

The Beneš construction is the starting point for the development of new finite-dimensional recursive approximations to the optimal filter. One of the new filters is applied to a simple tracking problem taken from computer vision, and its performance compared with that of the extended Kalman filter.

Phil. Trans. R. Soc. Lond. A (1996) **354**, 1099–1123

Printed in Great Britain

TEX Paper

1. Introduction

The Kalman–Bucy filter and its many variations are widely used in computer vision for estimating the state of time varying systems (Ayache 1991; Broida & Chellappa 1991; Faugeras 1993). A typical filter contains models of the system dynamics and of the measurement process. It takes as input a sequence of measurements obtained over an extended time. The filter produces from this sequence an estimate of the probability density function for the system state conditional on the measurements. If the filter is optimal, then the estimated density is the exact conditional density. In the case of the Kalman–Bucy filter the estimated density is Gaussian, and the filter outputs the mean and the covariance of the Gaussian. The Kalman–Bucy filter is optimal, but only for the small class of systems satisfying the following constraints: (i) the system evolution and the measurement process are linear, and both are subject to white Gaussian noise; (ii) the prior density for the system state is Gaussian and independent of the noise processes. There are many generalizations of the Kalman filter to nonlinear systems, the best known of which is the extended Kalman filter (EKF). These generalizations are usually suboptimal in that they do not generate the true probability density function for the system state conditional on the measurements. An applications oriented introduction to filtering in general and to the Kalman–Bucy filter in particular can be found in Maybeck (1979*a*, *b*, 1982).

In the nonlinear case the conditional density for the system state is usually very difficult to determine. The underlying reason is that the density depends on an infinite number of parameters or statistics. However, there do exist nonlinear systems which are finite dimensional in that the conditional density for the system state depends on only a finite number of sufficient statistics. Finite-dimensional systems are discussed by Brockett (1980) and Brockett & Clark (1980). The first nonlinear examples were found by Beneš (1981). In the Beneš filters the evolution of the sufficient statistics over time is governed by a finite set of first order, ordinary differential equations which can be solved recursively. In this context 'recursive' means that the conditional density ρ_t at time $t > s \geqslant 0$ can be computed from the single density ρ_s and the measurements obtained in the interval $(s, t]$. It is not necessary to use the measurements obtained in $[0, s]$.

Beneš's original result applies only to systems in which the drift velocity is the gradient of a scalar potential. In this paper the result is extended to include systems in which the drift velocity is an arbitrary linear function plus the gradient of a scalar potential; a similar extension is obtained by Yau (1994). The trajectories of the unperturbed systems to which the Beneš filter is applicable are described for a state space of dimension one. In the case of the Kalman–Bucy filter only two types of unperturbed trajectory are possible: linear or exponential as a function of time. In the case of Beneš systems there are five possible types of unperturbed trajectory: after suitable rescalings of space and time, an unperturbed trajectory $t \mapsto G(t)$, $0 \leqslant t$ with $G(0) = 0$, $(\mathrm{d}G/\mathrm{d}t)(0) > 0$ satisfies one of (i) $\lim_{t\to\infty} G(t) = a$; (ii) $G(t) \sim \sqrt{2t}$; (iii) $G(t) \sim t$; (iv) $G(t) \sim t^2$; (v) $G(t) \sim \exp(t/2)$. These trajectories are referred to as Beneš trajectories. It is shown that at small times an initial segment of a general unperturbed trajectory can be approximated to an accuracy of $O(t^5)$ by a Beneš trajectory, in contrast with the $O(t^3)$ accuracy obtainable with the Kalman–Bucy filter.

The Beneš filter is derived from a path integral representation of the conditional density ρ_t. The representation is the basis of a new class of finite-dimensional suboptimal filters which can be applied to a wide range of nonlinear systems, including those for which the drift velocity is a poynomial function of the system state. A comparison of one of the new filters with the EKF suggests that the new filters give an improved performance when the evolution of the system state has a strong nonlinearity.

The general framework for filtering, as part of the theory of stochastic processes, is given in § 2. The Beneš filter is derived in § 3. The five types of Beneš trajectory are obtained in § 4. A numerical example is given in which the EKF is close to optimal on a Beneš trajectory. An example of the application of the new type of filter to a problem in computer vision is described in § 5. Some concluding remarks are made in § 6.

2. Filtering

The main task in filtering is to obtain an accurate estimate of the probability density function ρ_t for a system state at time $t \geqslant 0$, conditional on a series of measurements obtained in the time interval $[0, t]$. There are many different frameworks for filtering. In the one employed here the system evolution and the measurement processes are modelled by Itô stochastic differential equations (SDEs). More detailed information about SDEs and their applications to filtering is given by Øksendal (1992), Pardoux (1991) and Rogers & Williams (1987).

The SDEs for filtering are stated in § 2 *a*. The Cameron-Martin-Girsanov transformation of an SDE is described in § 2 *b* and used in § 2 *c* to obtain a representation of ρ_t as a path integral.

(*a*) *Formulation*

The system state is modelled by a stochastic process X in $\mathbb{R}^n$ evolving over time according to the SDE,

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t \qquad (0 \leqslant t), \tag{2.1}$$

where the drift $f : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and the diffusion $\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ are measurable, and B is a standard Brownian motion in $\mathbb{R}^n$ independent of X_0. The random variable X_0 has a known law which summarizes the knowledge of the system state available at time $t = 0$, before any measurement. In many cases X_0 is deterministic, $X_0 = x$ a.s. The measurement process Y in $\mathbb{R}^m$ is described by the SDE

$$\mathrm{d}Y_t = g(t, X_t)\,\mathrm{d}t + \mathrm{d}W_t \qquad (0 \leqslant t, Y_0 = 0), \tag{2.2}$$

where $g : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^m$ is measurable and W is a standard Brownian motion in $\mathbb{R}^m$, independent of B and X_0. The equations (2.1), (2.2) are a simplified special case of the more general equations given by Pardoux (1991).

Conditions are imposed on (2.1), (2.2) sufficient to ensure that weak solutions exist. It is assumed that f, g, σ are locally bounded, measurable functions. The diffusion σ is assumed to satisfy the conditions

$$\begin{aligned}
\inf_{0\leqslant s\leqslant t}\{\|\sigma\sigma^{\mathrm{T}}(s,x)\|\} &> 0 && (x \in \mathbb{R}^n)\\
\lim_{y\to x}\sup_{0\leqslant s\leqslant t}\{\|\sigma\sigma^{\mathrm{T}}(s,y) - \sigma\sigma^{\mathrm{T}}(s,x)\|\} &= 0 && (x \in \mathbb{R}^n)\\
\sup_{0\leqslant s\leqslant t}\{\|\sigma\sigma^{\mathrm{T}}(s,x)\|\} &\leqslant c_t(1 + \|x\|^2) && (x \in \mathbb{R}^n).
\end{aligned} \tag{2.3}$$

The functions f, g are assumed to satisfy

$$\sup_{0\leqslant s\leqslant t}\{|x.k(s,x)|\}\leqslant c_t(1+\|x\|^2)\qquad (x\in\mathbb{R}^p),\tag{2.4}$$

where $p=n$, $k=f$ or $p=m$, $k=g$. In (2.3) and (2.4) the variable c_t is independent of x. It is shown by Stroock & Varadhan (1979) that if (2.3) and (2.4) hold with $k=f,g$, then the martingale problem in $\mathbb{R}^{n+m}$ for the drift $(t,x,y)^{\mathrm{T}}\mapsto(f(t,x),g(t,x))^{\mathrm{T}}$ and the diffusion $(t,x,y)^{\mathrm{T}}\mapsto(\sigma(t,x),I)^{\mathrm{T}}$ is well-posed. It follows that (2.1) and (2.2) have a unique weak solution (Øksendal 1992; Rogers & Williams 1987).

In practice, measurements are obtained at discrete times. The model (2.2) for the measurement process is an approximation which is accurate if the measurements are obtained frequently. In the discrete time case the measurement Z is the value of a function g of the system state, subject to a random perturbation '$\mathrm{d}W_t/\mathrm{d}t$',

$$Z_t=g(t,X_t)+\frac{\mathrm{d}W_t}{\mathrm{d}t}.$$

The measurement process Y in (2.2) is an integrated version of Z,

$$Y_t=\int_0^t Z_s\,\mathrm{d}s\qquad(0\leqslant t).\tag{2.5}$$

The use of Y simplifies the analysis, because the combined process $(X,Y)^{\mathrm{T}}$ is an Itô diffusion in $\mathbb{R}^{n+m}$. The relation between Y and Z is discussed by Maybeck (1982).

(b) *Cameron–Martin–Girsanov transformation*

A solution to the Itô SDE,

$$\mathrm{d}X_t=b(t,X_t)\,\mathrm{d}t+\sigma(t,X_t)\,\mathrm{d}B_t\qquad(0\leqslant t, X_0=a)\tag{2.6}$$

consists of a triple (X,P,B), where X, B are stochastic processes with a common sample space Ω and P is a probability measure defined on Ω. The process B is required to be a Brownian motion with respect to P; X is given by the integral version of (2.6),

$$X_t-a=\int_0^t b(s,X_s)\,\mathrm{d}s+\int_0^t\sigma(s,X_s)\,\mathrm{d}B_s\qquad(0\leqslant t).$$

The sample space Ω carries a filtration, that is a collection $\mathcal{F}$ of sigma-algebras $\mathcal{F}_t$, $0\leqslant t$, of subsets of Ω such that $\mathcal{F}_{t_1}\subset\mathcal{F}_{t_2}$ if $t_1\leqslant t_2$. The processes X, B are required to be adapted to $\mathcal{F}$, in that the functions $\omega\mapsto X(t,\omega)$, $\omega\mapsto B(t,\omega)$, $\omega\in\Omega$, are measurable with respect to $\mathcal{F}_t$, $0\leqslant t$. In effect, the sigma-algebra $\mathcal{F}_t$ is large enough to contain all the information about X, B available at time t.

If B is given and if X is expressible as a functional of B, i.e. X is adapted to the augmented natural filtration of B, then (X,P,B) is said to be a strong solution of (2.6). If X, B are a given pair of processes such that B is a Brownian motion under the measure P and (2.6) holds, then (X,P,B) is said to be a weak solution of (2.6). Every strong solution of (2.6) is necessarily a weak solution, but there exist SDEs that have weak solutions but no strong solutions. In these cases the construction of X from B is impossible; extra randomness is required, in addition to that provided by B (Karatzas & Shreve 1988; Rogers & Williams 1987). The weak solutions are appropriate for filtering, firstly because the criteria for the existence of a weak solution are more likely to be satisfied; and secondly because the conditional density is the same for the weak solutions and for the strong solutions, when the latter exist.

The Cameron–Martin–Girsanov (CMG) transformation is described. Let $c: [0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$ be a locally bounded, Borel function and let $\tilde{B}$ be the stochastic process in $\mathbb{R}^n$ defined by

$$\mathrm{d}\tilde{B}_t = c(t, X_t)\,\mathrm{d}t + \mathrm{d}B_t \qquad (0 \leqslant t, \tilde{B}_0 = 0). \tag{2.7}$$

It follows from (2.6) and (2.7) that

$$\mathrm{d}X_t = (b(t, X_t) - c(t, X_t))\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}\tilde{B}_t \qquad (0 \leqslant t, X_0 = a). \tag{2.8}$$

Let M be defined by

$$M_t = \exp\left(-\frac{1}{2}\int_0^t \|c(t, X_t)\|^2\,\mathrm{d}s - \int_0^t c(s, X_s)\cdot \mathrm{d}B_s\right) \qquad (0 \leqslant t). \tag{2.9}$$

The Girsanov theorem states firstly that there is a measure $\tilde{P}$ defined on Ω such that $\tilde{B}$ is a Brownian motion with respect to $\tilde{P}$, and secondly that the Radon–Nikodym derivative of $\tilde{P}$ with respect to P exists, and is given by

$$\left.\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}\right|_{\mathcal{F}_t} = M_t. \qquad (0 \leqslant t) \tag{2.10}$$

The conclusion is that $(X, \tilde{P}, \tilde{B})$ is a weak solution of (2.8).

To summarize, the CMG transformation of a given SDE (2.6) yields, firstly a new SDE (2.8) with a different drift term, and secondly, the relation between a solution of the first SDE and a weak solution of the second SDE. The solution to the second SDE is necessarily weak because $\tilde{B}$ is specified by (2.7).

(*c*) *The conditional density*

The CMG transformation is used to obtain a path integral representation of ρ_t, conditional on the measurements Y_s for $0 \leqslant s \leqslant t$ and on the law of X_0. The representation is unnormalized, in that it is correct only up to a scale factor independent of x. In other words, the condition $\int \rho(t,x)\,\mathrm{d}x = 1$ is not enforced. During the calculations terms which contribute only a scale factor to ρ_t may be omitted without comment.

The state process X and the measurement Y are combined to give a process $(X, Y)^{\mathrm{T}}$ in $\mathbb{R}^{n+m}$ with a sample space Ω equal to the product of the sample spaces of X and Y. The CMG transformation is then applied to the SDE describing $(X, Y)^{\mathrm{T}}$. Let P be the probability measure on Ω associated with the Brownian motion $(B, W)^{\mathrm{T}}$. Let $\mathcal{Y}_t$ be the σ-algebra generated by the random variables Y_s for $0 \leqslant s \leqslant t$ and let Z be a random variable in $L^1(\Omega, \mathcal{F}_t, P)$ such that ZM_t^{-1} is in $L^1(\Omega, \mathcal{F}_t, \tilde{P})$. It can be shown that

$$E(Z|\mathcal{Y}_t) = \frac{\tilde{E}(ZM_t^{-1}|\mathcal{Y}_t)}{\tilde{E}(M_t^{-1}|\mathcal{Y}_t)}, \tag{2.11}$$

where E is expectation with respect to P and $\tilde{E}$ is expectation with respect to $\tilde{P}$. Equation (2.11) is known as the Kallianpur–Striebel formula. A proof of it is given by Pardoux (1991). It is observed by Marcus (1984) that (2.11) is a version of Bayes's rule.

Let $(t, x) \mapsto \rho(t, x)$ be the probability density function for the system state X_t

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

conditional on the measurements obtained in the time interval $[0, t]$. It follows from the definition of ρ_t that

$$E(\chi_{\{X_t \in \,\mathrm{d}x\}} | \mathcal{Y}_t) = \rho(t, x)\,\mathrm{d}x, \tag{2.12}$$

where χ is the indicator function. The term $\tilde{E}(M_t^{-1}|\mathcal{Y}_t)$ is independent of x, thus it follows from (2.11) and (2.12) that up to a scale factor

$$\rho(t, x)\,\mathrm{d}x = \tilde{E}(\chi_{\{X_t \in \mathrm{d}x\}} M_t^{-1} | \mathcal{Y}_t). \tag{2.13}$$

It remains to select a CMG transformation such that the right-hand side of (2.13) reduces to a convenient form. This is done in the following two theorems.

Theorem 2.1. *Let X, Y satisfy the filter equations (2.1), (2.2). Let $(X, Y)^{\mathrm{T}}$ be the stochastic process in $\mathbb{R}^{n+m}$ with components X, Y. Let $h : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and $c : [0, \infty)\mathbb{R}^n \to \mathbb{R}^{n+m}$ be locally bounded, Borel functions such that*

$$c(t, x) \equiv \begin{pmatrix} c_1(t, x) \\ c_2(t, x) \end{pmatrix} = \begin{pmatrix} \sigma(t, x)^{-1}(f(t, x) - h(t, x)) \\ g(t, x) \end{pmatrix}. \tag{2.14}$$

It is assumed that h, c satisfy (2.4). Let $\tilde{P}$ be a new probability measure on Ω defined by

$$\left.\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P}\right|_{\mathcal{F}_t} = \exp\left(-\frac{1}{2}\int_0^t \|c(s, X_s)\|^2\,\mathrm{d}s - \int_0^t c(s, X_s)\cdot(\mathrm{d}B_s, \mathrm{d}W_s)\right) \qquad (0 \leqslant t).$$

Then the unnormalized conditional density ρ_t is represented by

$$\rho(t, x)\,\mathrm{d}x = \tilde{E}\left(\chi_{\{\xi_t \in \mathrm{d}x\}} \exp\left(\frac{1}{2}\int_0^t \|c(s, \xi_s)\|^2\,\mathrm{d}s + \int_0^t c(s, \xi_s)\cdot(\mathrm{d}B_s, \mathrm{d}W_s)\right)\right), \tag{2.15}$$

where ξ satisfies the SDE

$$\mathrm{d}\xi_t = h(t, \xi_t)\,\mathrm{d}t + \sigma(t, \xi_t)\,\mathrm{d}\tilde{B}_t \qquad (0 \leqslant t, \xi_0 = X_0).$$

Proof. The result follows from the CMG transformation and the observation that

$$\tilde{E}(\chi_{\{\xi_t \in \,\mathrm{d}x\}} M_t^{-1} \mid \mathcal{Y}_t) = \tilde{E}(\chi_{\{\xi_t \in \,\mathrm{d}x\}} M_t^{-1}).$$

■

The switch of notation in theorem 2.1, from X before the CMG transformation to ξ after the transformation, is made in order to clarify the exposition. The function h does not at this stage play any special role. It can be chosen at will, subject to the condition (2.4). In the applications of theorem 2.1 made below h is a linear function of x, and σ is constant; ξ is then an Ornstein–Uhlenbeck process in $\mathbb{R}^n$ (OU_n process).

In the next theorem the representation (2.15) of ρ_t is rewritten. The proof follows Beneš (1981), except for the way in which the auxiliary function h is used.

Theorem 2.2. *Let X, Y satisfy the filter equations*

$$\left.\begin{aligned} \mathrm{d}X_t &= f(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t \qquad (0 \leqslant t), \\ \mathrm{d}Y_t &= g(t, X_t)\,\mathrm{d}t + \mathrm{d}W_t \qquad (0 \leqslant t), \end{aligned}\right\} \tag{2.16}$$

where $\mathbb{R}^n$ is the state space of X and $\mathbb{R}^m$ is the state space of Y. Let $r : [0, \infty) \times \mathbb{R}^n \to$

$\mathbb{R}^n$ be defined by

$$r(t,x) = (\sigma\sigma^{\mathrm{T}})^{-1}(f(t,x) - h(t,x))$$

and let r also be the gradient of a scalar potential Φ, $r_i = \partial\Phi/\partial x_i$ for $1 \leqslant i \leqslant n$. Let $V : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}$ be defined by

$$V(t,x) = f^{\mathrm{T}}(\sigma\sigma^{\mathrm{T}})^{-1}f - h^{\mathrm{T}}(\sigma\sigma^{\mathrm{T}})^{-1}h + \sum_{i,j=1}^{n}\frac{\partial r_i}{\partial x_j}(\sigma\sigma^{\mathrm{T}})_{ij} + 2\frac{\partial\Phi}{\partial t} + \|g\|^2. \tag{2.17}$$

Let $\tilde{B}$, ξ be defined as in the statement of theorem 2.1. Then the unnormalized density ρ_t is represented by

$$\rho(t,x)\,\mathrm{d}x = \exp(\Phi(t,x)) \times \tilde{E}\left(\chi_{\{\xi_t\in dx\}}\exp\left(-\Phi(0,\xi_0) - \frac{1}{2}\int_0^t V(s,\xi_s)\,\mathrm{d}s + \int_0^t g(s,\xi_s)\cdot\mathrm{d}Y_s\right)\right). \tag{2.18}$$

Proof. Let $c : [0,\infty)\times\mathbb{R}^n \to \mathbb{R}^{n+m}$ be defined as in theorem 2.1, and let $e(t,\xi_t)$ be the exponent on the right-hand side of (2.15),

$$e(t,\xi_t) = \frac{1}{2}\int_0^t \|c(s,\xi_s)\|^2\,\mathrm{d}s + \int_0^t c(s,\xi_s)\cdot(\mathrm{d}B_s,\mathrm{d}W_s). \tag{2.19}$$

The exponent $e(t,\xi_t)$ is reduced to a convenient form. It follows from (2.16) that

$$\begin{pmatrix}\mathrm{d}B_t\\ \mathrm{d}W_t\end{pmatrix} = \begin{pmatrix}\sigma(t,X_t)^{-1}(\mathrm{d}X_t - f(t,X_t)\,\mathrm{d}t)\\ \mathrm{d}Y_t - g(t,X_t)\,\mathrm{d}t\end{pmatrix},$$

thus

$$\begin{aligned}\int_0^t c(s,\xi_s)\cdot(\mathrm{d}B_s,\mathrm{d}W_s) &= \int_0^t c_1(s,\xi_s)\cdot\mathrm{d}B_s + \int_0^t c_2(s,\xi_s)\cdot\mathrm{d}W_s\\ &= \int_0^t r(s,\xi_s)\cdot\mathrm{d}\xi_s - \int_0^t r(s,\xi_s)f(s,\xi_s)\,\mathrm{d}s\\ &\quad + \int_0^t g(s,\xi_s)\cdot\mathrm{d}Y_s - \int_0^t \|g(s,\xi_s)\|^2\,\mathrm{d}s.\end{aligned} \tag{2.20}$$

It follows from (2.17), (2.19) and (2.20) that

$$\begin{aligned}e(t,\xi_t) = &-\frac{1}{2}\int_0^t\left[V(s,\xi_s) - 2\frac{\partial\Phi}{\partial s}(s,\xi_s) - \sum_{i,j=1}^{n}\frac{\partial r_i}{\partial x_j}(\sigma\sigma^{\mathrm{T}})_{ij}\right]\mathrm{d}s\\ &+ \int_0^t r(s,\xi_s)\cdot\mathrm{d}\xi_s + \int_0^t g(s,\xi_s)\cdot\mathrm{d}Y_s.\end{aligned} \tag{2.21}$$

An application of the Itô formula to $\Phi(t,\xi_t)$ yields

$$\begin{aligned}\mathrm{d}(\Phi(t,\xi_t)) &= \frac{\partial\Phi}{\partial t}(t,\xi_t)\,\mathrm{d}t + \sum_{i=1}^{n}\frac{\partial\Phi}{\partial x_i}(t,\xi_t)\,\mathrm{d}\xi_i(t) + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2\Phi}{\partial x_i\partial x_j}(t,\xi_t)\,\mathrm{d}\xi_i(t)\,\mathrm{d}\xi_j(t)\\ &= \frac{\partial\Phi}{\partial t}(t,\xi_t)\,\mathrm{d}t + r(t,\xi_t)\cdot\mathrm{d}\xi_t + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial r_i}{\partial x_j}(t,\xi_t)(\sigma\sigma^{\mathrm{T}})_{ij}\,\mathrm{d}t,\end{aligned}$$

thus

$$\int_0^t r(s,\xi_s)\cdot \mathrm{d}\xi_s = \Phi(t,\xi_t) - \Phi(0,\xi_0) - \int_0^t \frac{\partial \Phi}{\partial s}(s,\xi_s)\,\mathrm{d}s - \frac{1}{2}\sum_{i,j=1}^{n}\int_0^t \frac{\partial r_i}{\partial x_j}(s,\xi_s)(\sigma\sigma^{\mathrm{T}})_{ij}\,\mathrm{d}s. \quad (2.22)$$

It follows from (2.21) and (2.22) that

$$e(t,\xi_t) = -\frac{1}{2}\int_0^t V(s,\xi_s)\,\mathrm{d}s + \Phi(t,\xi_t) - \Phi(0,\xi_0) + \int_0^t g(s,\xi_s)\cdot \mathrm{d}Y_s$$

as required. ■

The path integral representation (2.18) of ρ_t is similar to those obtained by Bensoussan (1992). If ξ_0 is known, $\xi_0 = a$, then the term $\Phi(0,\xi_0)$ on the right-hand side of (2.18) is constant, independent of x. It can then be omitted.

3. Finite-dimensional filters

A finite-dimensional filter is one in which the conditional density for the system state depends on only a finite number of statistics, and in addition the evolution of the statistics over time is governed by a set of ordinary differential equations. The best known example is the Kalman–Bucy filter. The finite-dimensional filters are important because they are computationally tractable. In § 3 *b* it is shown that the representation (2.18) of the unnormalized conditional density leads to class of finite-dimensional optimal filters, the Beneš filters. The results extend those of Beneš (1981) to cases in which the drift velocity of the system state differs by a linear function from the gradient of a scalar potential. The construction of the Beneš filter depends in part on the properties of bridge processes, as summarized in § 3 *a*. Some examples of Beneš filters are described in § 3 *c*.

The characterization of those systems for which the optimal conditional density has only a finite number of sufficient statistics appears to be difficult. Brockett (1980) and Brockett & Clark (1980) suggest a theoretical framework for this problem based on Lie algebras and the Duncan–Mortensen–Zakaï equation (Zakaï 1969). Among the successes of this theory are rigorous proofs that certain simple SDEs define filters which are truly infinite dimensional (Marcus 1984; Pardoux 1991). Chiou & Yau (1994) use Lie algebra methods to show that if the state space is of dimension two, then the optimal filter is finite dimensional only if the drift f differs by a general linear function from the gradient of a scalar potential.

(*a*) *Bridge processes*

The path integral representation (2.18) of ρ_t contains the indicator function χ. This function can be removed by using the bridge process associated with ξ, provided the bridge process exists. Let X be a stochastic process in $\mathbb{R}^n$, let $X \mapsto F(X_u, 0 \leqslant u \leqslant t)$ be a general, measurable, bounded functional of X, and let $Z = X_0$. A process $\eta_t^{Z\to b}$, $0 < t$, in $\mathbb{R}^n$ is a bridge process for X if

$$E(F(X_s, 0 \leqslant s \leqslant t) \mid X_0 = Z, X_0 = b) = E(F(\eta_{s,t}^{Z\to b}, 0 \leqslant s \leqslant t)) \qquad (0 < t). \quad (3.1)$$

It follows from (3.1) that $\eta_{t,t}^{Z\to b} = b$. If $X_0 = a$ a.s., then $\eta_{0,t}^{a\to b} = a$.

The bridge process is easily constructed if X is Gaussian. Let Λ be defined by

$$\Lambda(s,t) = -\mathrm{Cov}(X_s, X_t)\mathrm{Cov}(X_t, X_t)^{-1} \qquad (0 \leqslant s \leqslant t, t \neq 0). \tag{3.2}$$

The random variable X_s has the following orthogonal decomposition with respect to X_t,

$$X_s = (X_s + \Lambda(s,t)X_t) - \Lambda(s,t)X_t \qquad (0 \leqslant s \leqslant t, 0 \neq t)$$

in that X_t is independent of $X_s + \Lambda(s,t)X_t$. The bridge process $\eta_t^{Z\to b}$ for X is defined by

$$\eta_{s,t}^{Z\to b} = (X_s + \Lambda(s,t)X_t) - \Lambda(s,t)b \qquad (0 \leqslant s \leqslant t, 0 \neq t). \tag{3.3}$$

The expectation and the covariance functions of $\eta_t^{Z\to b}$ are readily calculated from (3.3) and the expectation and covariance functions of X. It is noted for future reference that for a Gaussian process X, $E(\eta_{s,t}^{Z\to b})$ is linear in b, and $\mathrm{Cov}(\eta_{s_1,t}^{Z\to b}, \eta_{s_2,t}^{Z\to b})$ is independent of b.

The bridge process is also defined if X is a Bessel process or the square of a Bessel process (Revuz & Yor 1991).

(*b*) *Beneš filters*

The Beneš filters are obtained by constraining the filter equations such that the representation (2.18) of ρ_t takes a particularly simple form. It is assumed that the auxiliary process ξ has a bridge process.

Theorem 3.1. *With the notation of theorem 2.2, let $p_\xi(t,x)$ be the probability density function for ξ_t, let $\xi_0 = Z \sim \mathcal{N}(a, \Sigma)$ and let $\eta_t^{Z\to b}$ be the bridge process for ξ from Z to b in time $t > 0$. If the process $L(t, \eta_t^{Z\to x})$ defined by*

$$L(t, \eta_t^{Z\to x}) = -\Phi(0, \xi_0) - \frac{1}{2}\int_0^t V(s, \eta_{s,t}^{Z\to x})\,\mathrm{d}s + \int_0^t g(s, \eta_{s,t}^{Z\to x})\,\mathrm{d}Y_s \tag{3.4}$$

is Gaussian with expectation $\mu(t,x)$ and variance $S(t,x)$, then the conditional density ρ_t for the system state is given by

$$\rho(t,x) = \exp(\Phi(t,x))\,p_\xi(t,x)\exp(\mu(t,x) + \tfrac{1}{2}S(t,x)). \tag{3.5}$$

Proof. It follows from (2.18) and the definitions of $\eta_t^{Z\to b}$, $L(t, \eta_t^{Z\to x})$, that

$$\rho(t,x)\,\mathrm{d}x = \exp(\Phi(t,x))p_\xi(t,x)\tilde{E}(\exp(L(t, \eta_t^{Z\to x})))\,\mathrm{d}x,$$

thus

$$\begin{aligned}\rho(t,x) &= \exp(\Phi(t,x))p_\xi(t,x)(2\pi S(t,x))^{-1/2} \\ &\quad\times \int_{-\infty}^{\infty} \exp(r - \tfrac{1}{2}S(t,x)^{-1}(r - \mu(t,x))^2)\,\mathrm{d}r \\ &= \exp(\Phi(t,x))p_\xi(t,x)\exp(\mu(t,x) + \tfrac{1}{2}S(t,x)).\end{aligned}$$

■

Conditions sufficient to ensure that $L(t, \eta_t^{Z\to x})$, as defined by (3.4), is Gaussian are given in the next theorem.

Theorem 3.2. *With the notation of theorems 2.2 and 3.1, it is assumed that h,*

g, V are linear in x, with

$$\left.\begin{aligned} h(t,x) &= -M(t)x && (0 \leqslant t),\\ g(t,x) &= \Pi(t)x && (0 \leqslant t),\\ V(t,x) &= v_1(t)\cdot x + v_0(t) && (0 \leqslant t). \end{aligned}\right\} \tag{3.6}$$

Let H, Λ, μ, c, D be defined by

$$H(t) = \int_0^t M(s)\,\mathrm{d}s \qquad (0 \leqslant t) \tag{3.7}$$

$$\Lambda(s,t) = -\operatorname{Cov}(\xi_s,\xi_t)\operatorname{Cov}(\xi_t,\xi_t)^{-1} \qquad (0 \leqslant s \leqslant t, 0 \neq t)$$

$$\mu(t,x) = \frac{1}{2}\int_0^t v_1(s)^{\mathrm{T}}\Lambda(s,t)x\,\mathrm{d}s - x^{\mathrm{T}}\int_0^t \Lambda(s,t)^{\mathrm{T}}\Pi(s)\,\mathrm{d}Y_s \qquad (0 \leqslant t)$$

$$c(t) = \exp(-H(t))a \qquad (0 \leqslant t)$$

$$D(t) = \exp(-H(t))\int_0^t \exp(H(u))\exp(H(u)^{\mathrm{T}})\,\mathrm{d}u\,\exp(-H(t)^{\mathrm{T}}) \qquad (0 \leqslant t). \tag{3.8}$$

It is assumed that a scalar potential $\Phi : [0,\infty)\times\mathbb{R}^n \to \mathbb{R}$ exists such that

$$\frac{\partial \Phi}{\partial x_i}(t,x) = f_i(t,x) - h_i(t,x) \qquad (1 \leqslant i \leqslant n).$$

Then the unnormalized conditional density for the system state is

$$\rho(t,x) = \exp(\Phi(t,x))\exp(-\tfrac{1}{2}(x-c(t))^{\mathrm{T}}D(t)^{-1}(x-c(t)))\exp(\mu(t,x)). \tag{3.9}$$

Proof. The process ξ is an OU_n process,

$$\mathrm{d}\xi_t = -M(t)\xi_t\,\mathrm{d}t + \mathrm{d}\tilde{B}_t \qquad (0 \leqslant t, \xi_0 = a). \tag{3.10}$$

The solution of (3.10) is

$$\xi_t = \exp(-H(t))\left(\xi_0 + \int_0^t \exp(H(s))\,\mathrm{d}\tilde{B}_s\right) \qquad (0 \leqslant t)$$

from which it follows that $E(\xi_t) = c(t)$, $\operatorname{Cov}(\xi_t,\xi_t) = D(t)$, $0 \leqslant t$. Next, it is shown that

$$\tilde{E}(\exp(L(t,\eta_t^{a\to x}))) = \exp(\mu(t,x)) \tag{3.11}$$

up to a scale factor independent of x. The term $v_0(t)$ in (3.6) is set equal to zero without loss of generality, because it contributes only a scale factor to ρ_t. The hypotheses of the theorem ensure that $L(t,\eta_t^{a\to x})$ is a linear functional of $\eta_t^{a\to x}$. It follows from the observations made in § 3 *a* that $E(L(t,\eta_t^{a\to x}))$ is linear in x and that the covariance function of $\eta_t^{a\to x}$ is independent of x. Equation (3.11) and hence (3.9) follow on calculating $E(L(t,\eta_t^{a\to x}))$. ∎

(*c*) *Special cases*

The original definition of the Beneš filter (Beneš 1981) excludes the Kalman–Bucy filters in which the drift is not a symmetric function of x. The more general methods developed in § 3 *b* allow the extension of the class of Beneš filters to include all the

Kalman–Bucy filters. In detail, let the filter equations be

$$\left.\begin{aligned} \mathrm{d}X_t &= F(t)X_t\,\mathrm{d}t + \mathrm{d}B_t \qquad (0 \leqslant t, X_0 = a), \\ \mathrm{d}Y_t &= \Pi(t)X_t\,\mathrm{d}t + \mathrm{d}W_t \qquad (0 \leqslant t), \end{aligned}\right\} \tag{3.12}$$

where $F : [0, \infty) \to \mathbb{R}^{n\times n}$, $\Pi : [0, \infty) \to \mathbb{R}^{m\times n}$. Let $h : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ be linear in x, $h(t, x) = -M(t)x$. It follows from (2.17), (3.12) and the linearity of h as a function of x that

$$V(t,x) = x^{\mathrm{T}}(F(t)^{\mathrm{T}}F(t) - M(t)^{\mathrm{T}}M(t) + \Pi(t)^{\mathrm{T}}\Pi(t))x + 2\frac{\partial \Phi}{\partial t} \quad (0 \leqslant t), \tag{3.13}$$

$$\Phi(t,x) = \tfrac{1}{2}x^{\mathrm{T}}(F(t) + M(t))x \quad (0 \leqslant t). \tag{3.14}$$

It remains to choose the matrix valued function $t \mapsto M(t)$, $0 \leqslant t$, such that the right-hand side of (3.13) is zero, and such that $F(t)+M(t)$ is symmetric for all $0 \leqslant t$. It is more convenient to first choose the symmetric matrix $Q(t) = F(t) + M(t)$ and then to obtain $M(t)$ from the equation $M(t) = Q(t) - F(t)$. On substituting $Q(t) - F(t)$ for $M(t)$ in (3.13) it follows that

$$V(t,x) = x^{\mathrm{T}}\left(-Q(t)^2 + Q(t)F(t) + F(t)^{\mathrm{T}}Q(t) + \Pi(t)^{\mathrm{T}}\Pi(t) + \frac{\mathrm{d}Q}{\mathrm{d}t}\right)x.$$

It suffices to choose $t \mapsto Q(t)$ such that Q satisfies the matrix Riccati equation (Choi & Laub 1990; Kenney & Leipnik 1985)

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = Q(t)^2 - Q(t)F(t) - F(t)^{\mathrm{T}}Q(t) - \Pi(t)^{\mathrm{T}}\Pi(t) \qquad (0 \leqslant t).$$

The next proposition gives an example of constraints on f, g sufficient to ensure that the hypotheses of theorem 3.2 are satisfied for a nonlinear drift f.

Proposition 3.1. *With the notation of theorems 2.2–3.2, let σ be the identity matrix, and let $g(t,x) = \Pi x$ where Π is a constant $m \times n$ matrix. Let $V(t,x) = v_1 x + v_0$ where v_1, v_0 are constant. If the drift velocity f is independent of t and if $h(x) = -Mx$ where M is a constant $n \times n$ matrix, then f satisfies the Riccati equation*

$$\|f\|^2 + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} = Q(x), \tag{3.15}$$

where Q is a quadratic form with constant coefficients, $Q(x) = x^{\mathrm{T}}q_2 x + q_1 x + q_0$, such that $q_2 + \Pi^{\mathrm{T}}\Pi$ is a positive matrix.

Proof. Equation (2.17) and the hypotheses of the proposition yield

$$\begin{aligned} v_1 x + v_0 &= V(x) \\ &= \|f\|^2 - \|h\|^2 + \sum_{i=1}^{n}\left(\frac{\partial f_i}{\partial x_i} - \frac{\partial h_i}{\partial x_i}\right) + \|\Pi x\|^2. \end{aligned} \tag{3.16}$$

The result follows on observing that (3.16) reduces to

$$\|f\|^2 + \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i} = v_1 x + v_0 + x^{\mathrm{T}}(M^{\mathrm{T}}M - \Pi^{\mathrm{T}}\Pi)x - \mathrm{tr}(M). \tag{3.17}$$

■

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

With the notation and the assumptions of proposition 3.1, let $f-h$ be the gradient of a scalar potential Φ,

$$f_i(x) + \sum_{j=1}^{n} M_{ij} x_j = \frac{\partial \Phi}{\partial x_i} \qquad (1 \leqslant i \leqslant n). \tag{3.18}$$

On using (3.18) to eliminate f from (3.17) it follows that

$$\|\nabla \Phi\|^2 - 2x^{\mathrm{T}} M^{\mathrm{T}} (\nabla \Phi) + \nabla^2 \Phi = -x^{\mathrm{T}} \Pi^{\mathrm{T}} \Pi x + v_1 x + v_0.$$

If (3.18) holds, then f itself is the gradient of a scalar potential only if M is symmetric. In the one-dimensional case M is just a scalar, $M = \beta$. A potential Φ can always be defined by

$$\Phi(x) = \int^x f(u)\,\mathrm{d}u + \tfrac{1}{2}\beta x^2 \tag{3.19}$$

and (3.15) reduces to a Riccati equation (Bellman 1969),

$$f^2 + \frac{\mathrm{d}f}{\mathrm{d}x} = q_2 x^2 + q_1 x + q_0 \qquad (q_2 \geqslant -\Pi^2). \tag{3.20}$$

4. Beneš trajectories in one dimension

The advantage of the Beneš filter over the Kalman–Bucy filter is that it can be applied to a wider range of dynamical systems. As indicated in § 1, a Beneš trajectory in one dimension is defined to be a trajectory $t \mapsto x(t)$ in $\mathbb{R}$ obtained by solving the ODE $\mathrm{d}x/\mathrm{d}t = f(x)$, where f satisfies the Riccati equation (3.20). In this section it is shown that there are five possible types of behaviour for the Beneš trajectories at large times. It is also shown that at small times an arbitrary smooth trajectory can be approximated to $O(t^5)$ by a Beneš trajectory.

It is assumed throughout this section that in the filter equations (2.16), $\sigma(t,x) \equiv 1$, $g(t,x) = x$ and that f satisfies the Riccati equation (3.20). It is also assumed that $x(0) = 0$ and that f is normalized such that $f(0) > 0$. If $f(0) < 0$, then the normalization is effected by applying the transformation $x \mapsto -x$.

Some general properties of Beneš trajectories are obtained in § 4 *a*. In particular it is shown that the normalized Beneš trajectories are strictly monotonic increasing with time. The trajectories are obtained explicitly in § 4 *b* for the case in which $q_1 = q_2 = 0$ on the right-hand side of (3.20). The trajectories for $q_1 \neq 0$ and $q_2 \neq 0$ are discussed in § 4 *c*. It is shown in § 4 *d* that an arbitrary trajectory can be approximated to an accuracy of $O(t^5)$ by a Beneš trajectory. Finally, in § 4 *e*, a numerical example is given in which the EKF is applied to a Beneš system and the resulting density compared with the optimal probability density function for the system state.

(*a*) *General results*

The Beneš trajectories are obtained by solving the ODE $\mathrm{d}x/\mathrm{d}t = f(x)$. A single integration yields

$$\int^x f(u)^{-1}\,\mathrm{d}u = t \tag{4.1}$$

from which x can in principle be obtained as a function of t. To avoid overuse of the symbol 'x', $G(t)$ is used for the system state $x(t)$,

$$\mathrm{d}G/\mathrm{d}t = f(G(t)). \tag{4.2}$$

A differential equation for G is obtained. It follows from (4.2) that

$$\begin{aligned}\frac{\mathrm{d}^2G}{\mathrm{d}t^2} &= \frac{\mathrm{d}f}{\mathrm{d}x}(G)\frac{\mathrm{d}G}{\mathrm{d}t} \\ &= \left(-\left(\frac{\mathrm{d}G}{\mathrm{d}t}\right)^2 + Q(G)\right)\frac{\mathrm{d}G}{\mathrm{d}t},\end{aligned} \tag{4.3}$$

where $Q(G) = q_0 + q_1 G + q_2 G^2$, in the notation of (3.20). A rearrangement of (4.3) yields

$$\frac{\mathrm{d}^2G}{\mathrm{d}t^2} + \left(\frac{\mathrm{d}G}{\mathrm{d}t}\right)^3 = \frac{\mathrm{d}G}{\mathrm{d}t}Q(G). \tag{4.4}$$

In the Kalman–Bucy filter, $f(x) = a_0 + a_1 x$, where a_0, a_1 are constants. The integral in (4.1) is evaluated to yield

$$\left.\begin{aligned} G(t) &= a_1^{-1}a_0(\exp(a_1 t) - 1) \qquad (G(0) = 0, a_1 \neq 0), \\ G(t) &= a_0 t \qquad (G(0) = 0, a_1 = 0). \end{aligned}\right\} \tag{4.5}$$

There are only two possible behaviours of G at large times: either G increases linearly with time or G increases exponentially with time.

Bellman (1969) shows that any solution to the Riccati equation (3.20) is of the form

$$f = \phi^{-1}\frac{\mathrm{d}\phi}{\mathrm{d}x}, \tag{4.6}$$

where ϕ is a solution of the linear ODE

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} - Q(x)\phi = 0. \tag{4.7}$$

It follows from (4.6) and (4.7) that f is real analytic in any region not including the zeros of ϕ or the poles of $\mathrm{d}\phi/\mathrm{d}x$. If $q_2 \neq 0$, then ϕ is a parabolic cylinder function. If $q_2 = 0$, $q_1 \neq 0$, then ϕ is an Airy function. If $q_2 = q_1 = 0$ and $q_0 \neq 0$, then ϕ is an exponential function, possibly with a complex exponent. If $Q \equiv 0$, then ϕ is linear. Further information about the various solutions to (4.7) is given by Abramovitz & Stegun (1965).

Proposition 4.1. *Let f have a zero in $[0, \infty)$, and let $a > 0$ be the least such zero. Let f be bounded on $[0, a]$. Then G is strictly monotonically increasing with t, $G(t) < a$ for all $t \geqslant 0$ and $\lim_{t\to\infty} G(t) = a$.*

Proof. Let $\epsilon > 0$ be a small positive number. Then f is bounded away from zero on $[0, a-\epsilon]$ thus there exists a time $t_0 < \infty$ such that $G(t_0) = a - \epsilon$, and G is strictly monotonically increasing for $0 \leqslant t \leqslant t_0$.

Suppose, if possible, that $f'(a) = 0$. Then it follows from (4.6) and (4.7) that $\phi'(a) = \phi''(a) = 0$, thus ϕ is identically zero. The case $f'(a) = 0$ is thus excluded because f is not properly defined. In addition, $f'(a) > -\infty$, because f satisfies the Riccati equation (3.20) at a. It follows that $0 > f'(a) > -\infty$. After rescaling f it can be and is assumed that $f'(a) = -1$. It follows that

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + O((x-a)^2) \\ &= -(x-a) + O((x-a)^2). \end{aligned}$$

Let $\alpha_1 < 1 < \alpha_2$ be positive numbers such that

$$\alpha_1(a-x) \leqslant f(x) \leqslant \alpha_2(a-x) \qquad (a-\epsilon \leqslant x \leqslant a). \tag{4.8}$$

Equations (4.1) and (4.8) yield

$$\alpha_1^{-1} \int_{a-\epsilon}^{x} (a-u)^{-1}\,\mathrm{d}u \geqslant t - t_0 \geqslant \alpha_2 \int_{a-\epsilon}^{x} (a-u)^{-1}\,\mathrm{d}u$$

from which it follows that

$$\epsilon \exp(-\alpha_2(t-t_0)) \leqslant a - x \leqslant \epsilon \exp(-\alpha_1(t-t_0)).$$

It follows that $\lim_{t\to\infty} G(t) = a$ and that G does not reach a in a finite time. Finally, G is strictly monotonically increasing for all t because f is strictly positive on $[0, a)$. ■

Proposition 4.2. *Let f be bounded on each finite interval $[0, N]$ for $N > 0$. Each trajectory $t \mapsto G(t)$, $t \geqslant 0$ is a monotonic function of t. If f has no zero in $[0, \infty)$, then $\lim_{t\to\infty} G(t) = \infty$.*

Proof. The result is proved in proposition 4.1 for the case in which f has a zero in $[0, \infty)$. If f has no zero in $[0, \infty)$, then it is strictly bounded away from zero on any finite interval $[0, N]$. Thus the trajectory G reaches N in a finite time t_0 and G is strictly monotonically increasing for $0 \leqslant t \leqslant t_0$. The result follows. ■

Propositions 4.1 and 4.2 are special cases of more general results for the Riccati equation (Bellman 1969).

(*b*) *A special case*

The case $Q(x) \equiv q_0$ can be resolved completely by elementary integration. After the appropriate rescalings of space and time, it is sufficient to consider the three cases $q_0 = 0, \pm 1$.

Proposition 4.3. *If $Q(x) \equiv 1$, then $\lim_{t\to\infty} t^{-1}G(t) = 1$. If $Q(x) \equiv -1$, then $\lim_{t\to\infty} G(t)$ is finite.*

Proof. If $q_2 = 0, q_1 = 0$ and $q_0 = 1$, then (4.4) reduces to

$$\frac{\mathrm{d}^2 G}{\mathrm{d}t^2} + \left(\frac{\mathrm{d}G}{\mathrm{d}t}\right)^3 = \frac{\mathrm{d}G}{\mathrm{d}t}. \tag{4.9}$$

If $\mathrm{d}^2G/\mathrm{d}t^2 \equiv 0$, then (4.9) has the solution $G(t) = t$. Otherwise, define the function $t \mapsto u(t)$ by $u = \mathrm{d}G/\mathrm{d}t$. Then (4.9) yields

$$\frac{\mathrm{d}u}{\mathrm{d}t} = u - u^3$$

from which it follows that

$$u(t) = \frac{A\exp(t)}{\sqrt{1 + A^2\exp(2t)}}, \tag{4.10}$$

where $0 < A < 1$ is a disposable constant. On integrating (4.10) it follows that

$$G(t) = \ln\left(\frac{A\exp(t) + \sqrt{1 + A^2\exp(2t)}}{A + \sqrt{1 + A^2}}\right). \tag{4.11}$$

The result $\lim_{t\to\infty} t^{-1}G(t) = 1$ follows.

If $q_2 = 0, q_1 = 0$ and $q_0 = -1$, then a similar calculation yields

$$G(t) = \arcsin(A) - \arcsin(A\exp(-t)), \tag{4.12}$$

where $A > 0$ is a disposable constant. The trajectory (4.12) is bounded such that $\lim_{t\to\infty} G(t) = \arcsin(A)$. ■

Proposition 4.4. *Let $Q(x) \equiv 0$. Then $f(x) = (x+a)^{-1}$ for some constant a. If $a > 0$, then $\lim_{t\to\infty} t^{-1/2}G(t) = \sqrt{2}$.*

Proof. The function $f(x) = (x+a)^{-1}$ is obtained by solving (3.15) with $Q \equiv 0$. If $a > 0$, then $\mathrm{d}x/\mathrm{d}t = f(x)$ has a solution in the interval $t \geqslant 0$. The solution satisfies $x^2/2 + ax = t$ from which it follows that

$$G(t) = -a + \sqrt{a^2 + 2t}.$$

■

(*c*) *Large times*

If $f(x) > 0$ for $x \geqslant 0$, then a Beneš trajectory $t \mapsto G(t)$ such that $G(0) = 0$ attains large values at large times, $\lim_{t\to\infty} G(t) = \infty$. In this subsection the rate of growth of a Beneš trajectory at large times is investigated. The case $q_2 = q_1 = 0$ has already been described in § 4 *b*. It remains to consider the case $q_2 = 0, q_1 \neq 0$ and the case $q_2 \neq 0$.

If $q_2 = 0$, $q_1 \neq 0$, then

$$f(x)^2 + \frac{\mathrm{d}f}{\mathrm{d}x} = q_1 x + q_0.$$

After rescaling space and time, and then applying a translation in space, it suffices to examine the case $q_1 = 1$, $q_0 = 0$. Then f is a ratio, $f = \phi^{-1}\mathrm{d}\phi/\mathrm{d}x$ where ϕ is a solution of

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} - x\phi = 0. \tag{4.13}$$

The solutions of (4.13) are known as Airy functions (Abramovitz & Stegun 1965). There are two linearly independent Airy functions Ai, Bi with contrasting behaviours at large $x > 0$.

$$\left.\begin{aligned} \mathrm{Ai}(x) &\sim \tfrac{1}{2}\pi^{-1/2}x^{-1/4}\exp(-\tfrac{2}{3}x^{3/2})(1 + O(x^{-3/2})), \\ \mathrm{Bi}(x) &\sim \pi^{-1/2}x^{-1/4}\exp(\tfrac{2}{3}x^{3/2})(1 + O(x^{-3/2})). \end{aligned}\right\} \tag{4.14}$$

In any linear combination $r_1\,\mathrm{Ai} + r_2\,\mathrm{Bi}$ with $r_2 \neq 0$ the function Bi predominates over Ai for $x > 0$ and large. It is thus sufficient to consider two cases, $f(x) = \mathrm{Ai}(x)^{-1}\mathrm{d}\,\mathrm{Ai}/\mathrm{d}x$ and $f(x) = \mathrm{Bi}(x)^{-1}\mathrm{d}\,\mathrm{Bi}/\mathrm{d}x$. The first case is ruled out by the condition $f(0) > 0$ because $\mathrm{Ai}(x) > 0$, $\mathrm{d}\,\mathrm{Ai}/\mathrm{d}x < 0$ for $x \geqslant 0$. In the second case $\mathrm{Bi}(x) > 0$, $\mathrm{d}\,\mathrm{Bi}/\mathrm{d}x > 0$ for $x \geqslant 0$. The hypothesis $f(x) > 0$ for $x \geqslant 0$ is satisfied and the growth of f is given at large x by

$$f(x) = \mathrm{Bi}(x)^{-1}\frac{\mathrm{d}\,\mathrm{Bi}}{\mathrm{d}x} \sim x^{1/2}$$

from which it follows that $G(t) \sim t^2$.

If $q_2 \neq 0$, then after rescaling x, t and applying a constant translation to the x coordinate it suffices to consider the two cases:

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} - (\tfrac{1}{4}x^2 + a)\phi = 0, \tag{4.15}$$

$$\frac{d^2\phi}{dx^2} + (\tfrac{1}{4}x^2 - a)\phi = 0. \tag{4.16}$$

The solutions ϕ to (4.15) or (4.16) are known as parabolic cylinder functions (Abramovitz & Stegun 1965). The solutions to (4.16) are oscillatory if $a < 0$. If $a \geqslant 0$, then the solutions to (4.16) are oscillatory for $|x| > 2\sqrt{a}$. Thus (4.16) does not yield velocities f that are non-zero in $[0, \infty)$. The equation (4.15) has two linearly independent solutions U, V with contrasting behaviours for large $x > 0$,

$$\left.\begin{aligned} U(a,x) &= \exp(-\tfrac{1}{4}x^2)x^{-a-1/2}(1 + O(x^{-2})), \\ V(a,x) &= \surd(2/\pi)\exp(\tfrac{1}{4}x^2)x^{a-1/2}(1 + O(x^{-2})) \end{aligned}\right\} \tag{4.17}$$

If $a < 0$, then U, V are oscillatory for $|x| < 2\sqrt{|a|}$. Thus it suffices to examine the case $a \geqslant 0$. In the range $x \geqslant 0$ the function V dominates any linear combination $r_1U + r_2V$ at large x provided $r_2 \neq 0$. The case $f(x) = U(x)^{-1}dU/dx$ is ruled out because $f(x) \sim -x/2 < 0$ if x is large. There remains only the case

$$f(x) = V(a,x)^{-1}\frac{dV}{dx} \sim \tfrac{1}{2}x + O(x^{-1})$$

from which it follows that $G(t) \sim \exp(t/2)$.

The possible behaviours of the Beneš trajectories at large times are summarized.

Theorem 4.1. *Let $t \mapsto G(t)$ be a Beneš trajectory such that $G(0) = 0$,*

$$\frac{dG}{dt}(0) > 0.$$

Then after the appropriate rescalings of space and time the trajectory exhibits one of the following behaviours as $t \to \infty$:
(i) $\lim_{t\to\infty} G(t) = a$;
(ii) $G(t) \sim \sqrt{2t}$;
(iii) $G(t) \sim t$;
(iv) $G(t) \sim t^2$;
(v) $G(t) \sim \exp(t/2)$.
In all cases G is strictly monotonic increasing for $t \geqslant 0$.

Examples of the five types of trajectory are shown in figure 1, drawn with the aid of Mathematica (Wolfram 1991). The trajectories are chosen such that $(dG/dt)(0) = 1/2$. At time $t = 5$ the trajectories in figure 1 are in order (i) to (v) as shown.

(*d*) *Small times*

Let H be a general trajectory such that

$$H(t) = h_1t + h_2t^2 + h_3t^3 + h_4t^4 + O(t^5) \qquad (h_1 > 0). \tag{4.18}$$

At small times H can be approximated to an accuracy of $O(t^3)$ by a trajectory for which f is linear. It suffices to choose a_0, a_1 in (4.5) such that

$$a_0 = h_1,$$
$$a_1 = 2h_1^{-1}h_2.$$

Let $f = \phi^{-1}d\phi/dx$ where ϕ is as in (4.6). The inverse drift velocity f^{-1} has an expansion as a power series in x away from the zeros of $d\phi/dx$ and the poles of ϕ. It

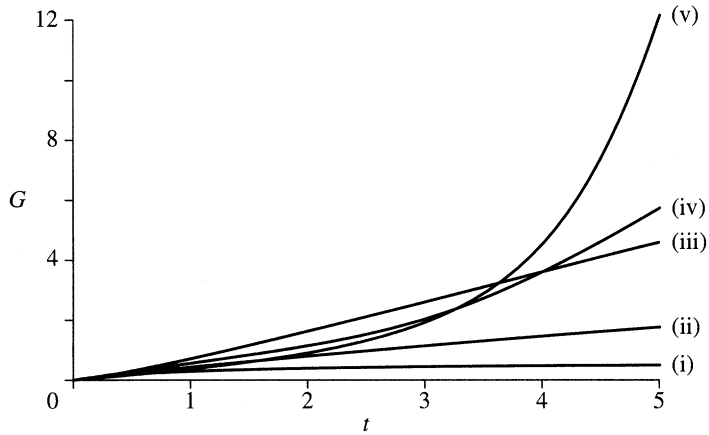

Figure 1. The five types of trajectory.

follows that G also has an expansion as a power series in t. Let this expansion be

$$G(t) = \sum_{i=1}^{\infty} g_n t^n \qquad (g_1 > 0).$$

On substituting the power series for G into (4.4) it follows that

$$2g_2 + 6g_3 t + 12 g_4 t^2 + (g_1 + 2g_2 t + 3g_3 t^2)^3 \\ = q_2 g_1^3 t^2 + q_1(g_1^2 t + 3g_1 g_2 t^2) + q_0(g_1 + 2g_2 t + 3g_3 t^2) + O(t^3). \qquad (4.19)$$

The equation (4.19) holds for $t \geqslant 0$ if and only if

$$\left. \begin{aligned} q_0 &= g_1^2 + 2g_1^{-1} g_2, \\ q_1 &= 6g_1^{-2} g_3 + 4g_2 - 4g_1^{-3} g_2^2, \\ q_2 &= 12 g_1^{-5} g_2^3 - 24 g_1^{-4} g_2 g_3 + 12 g_1^{-3} g_4 + 6 g_1^{-1} g_3. \end{aligned} \right\} \qquad (4.20)$$

It follows that the g_i for $1 \leqslant i \leqslant 4$ can be assigned arbitrary values, subject only to the conditions $g_1 > 0$ and $q_2 \geqslant -\Pi^2$. The g_i for $i \geqslant 5$ are uniquely determined by a set of equations of the form

$$n(n-1) g_n t^{n-2} = c_n(g_i, 1 \leqslant i \leqslant n-1) t^{n-2} \qquad (n \geqslant 5),$$

where c_n is a complicated polynomial in the g_i for $1 \leqslant i \leqslant n-1$. Thus the coefficients g_i, $1 \leqslant i \leqslant 4$ once chosen, determine G uniquely. These observations have the following consequence.

Proposition 4.5. *Let the coefficients $h = (h_1, h_2, h_3, h_4)^{\mathrm{T}}$ of H, where H is as defined in (4.18), satisfy $q_2(h) > -\Pi^2$. Then there exists a Beneš filter for which the deterministic trajectory G approximates H to an accuracy of $O(t^5)$.*

Proof. It suffices to set $g_i = h_i$ for $1 \leqslant i \leqslant 4$. The remaining coefficients g_i of G are then determined uniquely by the condition that G is a solution to (4.4). ■

(*e*) *Comparison with the EKF*

The Beneš systems can be used to test the performance of filters used in practice for nonlinear estimation. The advantage of the Beneš system in such a test is that the filter output can be compared directly with the optimal probability density function for the system state. The filter equations are assumed to be

$$\left.\begin{aligned} \mathrm{d}X_t &= f(X_t)\,\mathrm{d}t + \mathrm{d}B_t \qquad (0 \leqslant t, X_0 = a), \\ \mathrm{d}Y_t &= \lambda X_t\,\mathrm{d}t + \mathrm{d}W_t \qquad (0 \leqslant t). \end{aligned}\right\} \tag{4.21}$$

The extended Kalman filter (EKF) is a good example of a widely used filter. The implementation of the EKF for continuous time measurements is adapted from the book by Maybeck (1979*b*, § 9.5, eqns (9-77a) and (9-77b)). In the EKF the optimal density for the system state conditional on the measurements is approximated at time t by a Gaussian density $\sim \mathcal{N}(\hat{x}_{\mathrm{e}}(t), \hat{v}_{\mathrm{e}}(t))$. Let F be the function defined by

$$F(x) = \frac{\mathrm{d}f}{\mathrm{d}x}.$$

Then the time evolution of $\hat{x}_{\mathrm{e}}$, $\hat{v}_{\mathrm{e}}$ is given by the ODEs

$$\left.\begin{aligned} \mathrm{d}\hat{x}_{\mathrm{e}} &= f(\hat{x}_{\mathrm{e}}(t))\,\mathrm{d}t + k_t(\mathrm{d}Y_t - \lambda\hat{x}_{\mathrm{e}}(t)\,\mathrm{d}t) \qquad (0 \leqslant t), \\ \frac{\mathrm{d}\hat{v}_{\mathrm{e}}}{\mathrm{d}t} &= 2F(\hat{x}_{\mathrm{e}}(t))\hat{v}_{\mathrm{e}}(t) + 1 - \lambda^2\hat{v}_{\mathrm{e}}(t)^2 \qquad (0 \leqslant t), \end{aligned}\right\} \tag{4.22}$$

where k_t is the gain, $k_t = \lambda\hat{v}_{\mathrm{e}}(t)$.

The EKF is compared with the Beneš filter for the nonlinear system with drift $f(x) = \mathrm{Bi}(x)^{-1}\mathrm{d}\,\mathrm{Bi}/\mathrm{d}x$, where Bi is the Airy function with an asymptotic expansion as given in (4.14). The initial state is deterministic, $X_0 = 0$ a.s. The drift f is singular at $x = -1.17367\ldots$, because Bi has a zero at this point. It is the right-most zero of Bi. The function h in theorem 2.2 is defined by $h(x) = -\lambda x$. It follows from (3.17) that

$$\begin{aligned} v_1 x + v_0 &= f(x)^2 - h(x)^2 + \frac{\mathrm{d}f}{\mathrm{d}x} - \frac{\mathrm{d}h}{\mathrm{d}x} + \lambda^2 x^2 \\ &= x + \lambda, \end{aligned}$$

thus $v_1 = 1$ and $v_0 = \lambda$.

The functions Λ, μ, c, D, in the notation of theorem 3.2, are evaluated using (3.8),

$$\begin{aligned} \Lambda(s,t) &= -\frac{\sinh(\lambda s)}{\sinh(\lambda t)} \qquad (0 \leqslant s \leqslant t, 0 \neq t), \\ \mu(t,x) &= x\left(\frac{1}{2}\int_0^t \Lambda(s,t)\,\mathrm{d}s - \lambda\int_0^t \Lambda(s,t)\,\mathrm{d}Y_s\right) \\ &= x\left(\frac{1}{2}\int_0^t \Lambda(s,t)\,\mathrm{d}s + \lambda\int_0^t Y_s\frac{\partial\Lambda}{\partial s}(s,t)\,\mathrm{d}s + \lambda Y_t\right) \qquad (0 \leqslant t), \qquad (4.23) \\ c(t) &= 0, \\ D(t) &= \lambda^{-1}\exp(-\lambda t)\sinh(\lambda t) \qquad (0 \leqslant t). \qquad (4.24) \end{aligned}$$

In the implementation the form (4.23) of μ is used to avoid numerical difficulties arising from the estimation of small increments ΔY in the measurement process. The potential Φ is

$$\Phi(x) = \int_0^x f(u)\,\mathrm{d}u + \tfrac{1}{2}\lambda x^2 \qquad (u_0 < x),$$

where u_0 is the right-most zero of Bi.

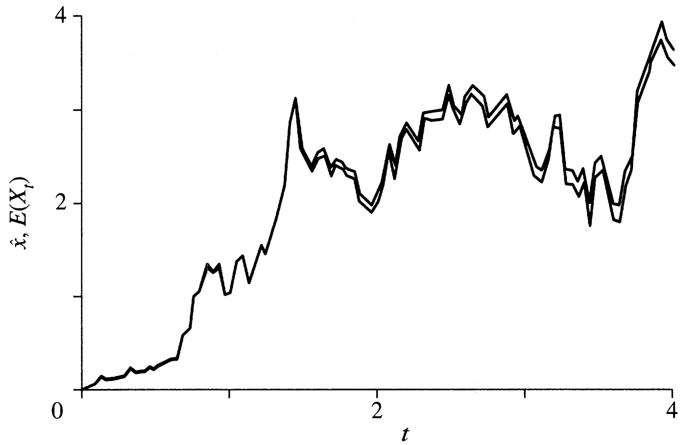

Figure 2. Comparison of expected values.

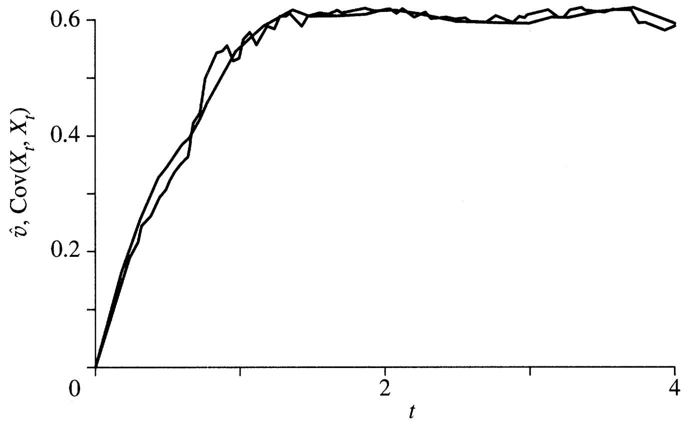

Figure 3. Comparison of variances.

The expression (3.9) for ρ_t reduces to

$$\rho(t, x) = \exp(\Phi(x) - \tfrac{1}{2}D(t)^{-1}(x - c(t))^2 + \mu(t, x))$$

The Beneš filter and the EKF are run in the time interval $[0, 4.008]$, with $\lambda = 2.0$. The continuous time system was approximated by a discrete time system with a time step of $t/500$ where $t = 4.008$. The optimal expected value $\hat{x}_o$ and the optimal variance $\hat{v}_o$ were obtained at each time step by numerical integration of the optimal density. The EKF generated estimates $\hat{x}_e$, $\hat{v}_e$ of $\hat{x}_o$, $\hat{v}_o$ according to (4.22). The graphs of $\hat{x}_o$, $\hat{x}_e$ are compared in figure 2. The estimated expected value $\hat{x}_e$ tends to be very slightly higher that the true expected value $\hat{x}_o$. It is not known if the discrepancy is significant.

The graphs of $\hat{v}_o$ and $\hat{v}_e$ are compared in figure 3. The graph of $\hat{v}_e$ is the more irregular of the two, but otherwise the two graphs are similar. To a first approximation, the covariance increases linearly until it reaches a value of approximately 0.6. It then levels off, with only small variations over time.

The graphs are plotted using 100 samples at times drawn from $[0, 4.008]$. It is apparent that the EKF is a good approximation to the optimal filter for this nonlinear system. The optimal density and the density generated by the EKF are shown

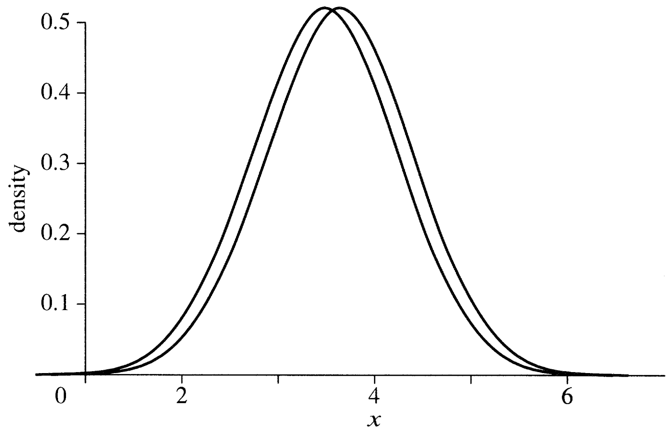

Figure 4. Comparison of the optimal density and the EKF density.

Table 1. *Comparison of expected values and variances for the densities in figure 4*

	expected value	variance
Beneš	3.48	0.585
EKF	3.65	0.589

together in figure 4 at $t = 4.008$. The expected values and the variances are given in table 1.

5. An application to computer vision

The derivation in § 3 of the Beneš filters is the starting point of a method for constructing finite-dimensional, recursive approximations to the optimal filter in more general, nonlinear cases. The new method has two advantages over previous ones, for example those described by Maybeck (1979*a*, *b*, 1982). Firstly, a wide variety of functions can appear as approximations to ρ_t; the method is not restricted to Gaussian approximations. Secondly, the new method is more accurate because it uses a direct approximation to the path integral representation of ρ_t. It is thus easier to control the errors arising in the approximation. A full description is given by Maybank (1996).

In this section a numerical example is given in which the new method is applied to a simple nonlinear estimation problem taken from computer vision. The results are compared with those obtained using the extended Kalman filter (EKF). The problem is stated in § 5 *a*, and the approximating filter described in § 5 *b*. In § 5 *c* the results obtained using the new filter are compared with the results obtained using the EKF. Background material on the use of filtering in computer vision is given by Faugeras (1993).

(*a*) *Tracking*

The problem concerns an object with a time varying position $t \mapsto p(t)$, $0 \leqslant t$, approaching a camera as shown in figure 5. Initially the object is far from the camera, the image θ of the object is near to the centre of the field of view, and the image

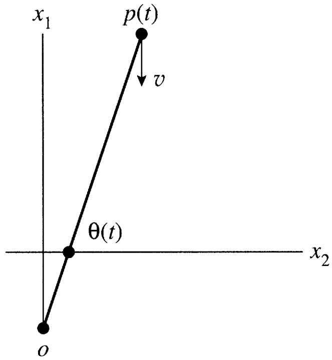

Figure 5. Projection to a one-dimensional image.

motion is small. As the object approaches the camera its image moves away from the centre of the field of view and the image motion becomes larger. The task is to track the image for a fixed time t and then to compute the probability density function summarizing the available information concerning the true value of the point $\theta(t)$. In order to simplify the calculations it is assumed that the motion takes place in a fixed plane which projects down to a line in the camera image.

The image is obtained by polar projection. Let coordinates x_1, x_2 be chosen in space with origin at the optical centre o of the camera. The camera is oriented in space such that the plane in which the object moves projects down to the line $x_2 = 1$. A point $(x_1, x_2)^{\mathrm{T}}$ in space with $x_2 \neq 0$ projects to x_1/x_2 in the image. It is assumed that in the absence of noise the object has a constant velocity $(0, -v)^{\mathrm{T}}$. Let the image of $p(t)$ be $\theta(t) = p_1(t)/p_2(t)$. On rescaling as necessary,

$$\theta(t) = \theta(0)(1 - vt)^{-1}$$

in the noise free case. It follows that

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = v\theta(0)(1 - vt)^{-2} = v\theta(0)^{-1}\theta(t)^2.$$

Let b be defined by $b = v\theta(0)^{-1}$. The projection of the object into the image is modelled as a random variable Θ evolving over time according to the SDE,

$$\mathrm{d}\Theta_t = b\Theta_t^2\,\mathrm{d}t + \mathrm{d}B_t \qquad (0 \leqslant t). \tag{5.1}$$

The model (5.1) captures the qualitative properties of the image motion discussed above, i.e. initial random wandering close to the origin, followed by rapid motion rightwards, away from the origin. The model is, however, unsuitable for deducing detailed information about the motion of the object in space. In these experiments it is assumed that b is known. The drift velocity f is defined by

$$f(\theta) = b\theta^2. \tag{5.2}$$

The random variable Θ_0 is assumed to have a Gaussian distribution $\Theta_0 \sim \mathcal{N}(a, \alpha)$. The measurement Y_t of $\Theta(t)$ evolves according to the SDE,

$$\mathrm{d}Y_t = \lambda\Theta_t\,\mathrm{d}t + \mathrm{d}W_t \qquad (0 \leqslant t), \tag{5.3}$$

where $\lambda > 0$ is a constant. If λ is large, then the measurements are accurate.

(*b*) *Application of the new filter*

It follows from (5.2) that

$$f(\theta)^2 + \frac{\mathrm{d}f}{\mathrm{d}\theta} = b^2\theta^4 + 2b\theta. \tag{5.4}$$

Let $h(x) = -\beta\theta$. The function V, as defined by (2.17), reduces in this example to

$$V(\theta) = b^2\theta^4 + 2b\theta - \beta^2\theta^2 + \beta + \lambda^2\theta^2.$$

It is convenient to set $\beta = \lambda$ and to omit the constant terms from V, giving

$$V(\theta) = b^2\theta^4 + 2b\theta.$$

The potential Φ is obtained from (3.19),

$$\Phi(\theta) = \tfrac{1}{3}b\theta^3 + \tfrac{1}{2}\lambda\theta^2. \tag{5.5}$$

The functions Λ, c, D, in the notation of theorem 3.2, are evaluated using (3.8),

$$\begin{aligned}\Lambda(s,t) &= -\exp(\lambda(t-s))\left(\frac{\sinh(\lambda s)\exp(\lambda s)}{\sinh(\lambda t)\exp(\lambda t)}\right) \qquad (0 \leqslant s \leqslant, 0 \neq t),\\ c(t) &= \exp(-\lambda t)a \qquad (0 \leqslant t),\\ D(t) &= \lambda^{-1}\exp(-\lambda t)\sinh(\lambda t) \qquad (0 \leqslant t).\end{aligned} \tag{5.6}$$

The initial state ξ_0 is deterministic, $\xi_0 = a$ a.s. The random variable $L(t, \eta_t^{a\to\theta})$ of (3.4) reduces in this example to

$$\begin{aligned}L(t,\eta_t^{a\to\theta}) &= -\frac{1}{2}\int_0^t V(\eta_{s,t}^{a\to\theta})\,\mathrm{d}s + \lambda\int_0^t \eta_{s,t}^{a\to\theta}\,\mathrm{d}Y_s\\ &= -\frac{1}{2}\int_0^t [b^2(\eta_{s,t}^{a\to\theta})^4 + 2b\eta_{s,t}^{a\to\theta}]\,\mathrm{d}s + \lambda\theta Y_t - \lambda\int_0^t Y_s\,\mathrm{d}\eta_{s,t}^{a\to\theta}.\end{aligned}$$

Let $m(s)$ be the expected value of $\eta_{s,t}^{a\to\theta}$ and let $C(s_1,s_2)$ be the covariance of $\eta_{s_1,t}^{a\to\theta}$, $\eta_{s_2,t}^{a\to\theta}$. The function μ defined in theorem 3.2 is given in this example by

$$\begin{aligned}\mu(t,\theta) &= E(L(t,\eta_t^{a\to\theta}))\\ &= -\frac{1}{2}\int_0^t [b^2E((\eta_{s,t}^{a\to\theta})^4) + 2bE(\eta_{s,t}^{a\to\theta})]\,\mathrm{d}s + \lambda\theta Y_t - \lambda\int_0^t Y_s\,\mathrm{d}(E(\eta_{s,t}^{a\to\theta}))\\ &= -\frac{1}{2}\int_0^t [b^2(6C(s,s)m(s)^2 + m(s)^4) + 2bm(s)]\,\mathrm{d}s + \lambda\theta Y_t - \lambda\int_0^t Y_s\,\mathrm{d}m(s).\end{aligned}$$

The approximation $\hat{\rho}_t$ to ρ_t is suggested by (3.9),

$$\hat{\rho}(t,x) = \exp(\Phi(x) - \tfrac{1}{2}D(t)^{-1}(x-c(t))^2 + \mu(t,x)). \tag{5.7}$$

The function $\hat{\rho}_t$ differs from ρ_t because the variance of $L(t,\eta_t^{a\to b})$ is discarded.

The term $\mu(t,x)$ in the exponent on the right-hand side of (5.7) contains a term proportional to θ^4, with a negative coefficient. This term ensures that $\hat{\rho}_t$ tends rapidly to zero as $|\theta|$ becomes large.

(*c*) *Comparison with the EKF*

Let $\hat{\theta}_\mathrm{e}(t)$, $\hat{v}_\mathrm{e}(t)$ be the expected value and the variance of the Gaussian approximation to the optimal density produced by the EKF at time t. Then $\hat{\theta}_\mathrm{e}$, $\hat{v}_\mathrm{e}$ satisfy

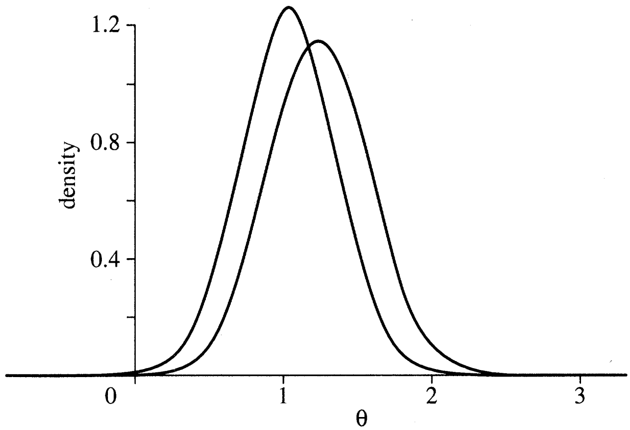

Figure 6. Comparison of estimated densities for $\lambda = 20.0$.

Table 2. *Comparison of expected values and standard deviations for* $\lambda = 20.0$

	expected value	variance
new filter	2.256	0.102
EKF	2.038	0.122

the ODEs

$$\left.\begin{aligned} \mathrm{d}\hat{\theta}_{\mathrm{e}} &= b\hat{\theta}_{\mathrm{e}}(t)^2\,\mathrm{d}t + \lambda\hat{v}_t(\mathrm{d}Y_t - \lambda\hat{\theta}_{\mathrm{e}}(t)\,\mathrm{d}t) \qquad (0 \leqslant t, \hat{\theta}_{\mathrm{e}}(0) = a), \\ \frac{\mathrm{d}\hat{v}_{\mathrm{e}}}{\mathrm{d}t} &= 4b\hat{\theta}_{\mathrm{e}}(t)\hat{v}_{\mathrm{e}}(t) + 1 - \lambda^2\hat{v}_{\mathrm{e}}(t)^2 \qquad (0 \leqslant t). \end{aligned}\right\} \tag{5.8}$$

The equations (5.8) are obtained from (4.22) on setting $f(\theta) = b\theta^2$. The initial values $\hat{\theta}_0 = a$, $\hat{v}_0 = \alpha$ are as for the prior Θ_0 defined in § 5 *a*.

The equations (5.6), (5.8) are integrated numerically using Mathematica. A typical set of results is displayed in figure 6. The leftmost peak is the estimated density of Θ_t at time $t = 0.8032$ obtained using the EKF and the rightmost peak is the estimated density of Θ_t at $t = 0.8032$ obtained using the new filter. In the computer implementation the continuous time system was approximated by a discrete time system with a time step 0.8032/500. The values of the parameters in the two filters are

$$\Theta_0 = 0.01 \text{ a.s.}, \qquad b = 4.0, \qquad \lambda = 20.00. \tag{5.9}$$

The expected values and variances of the densities shown in figure 6 are listed in table 2.

The true value of the system state is 2.765. This is well within the range predicted by the new filter but at the edge of the range predicted by the EKF.

The general behaviour of the random trajectory Θ is as follows. Initially it remains in the region of $\theta = 0$, moving in a random fashion driven by the noise term in the SDE (5.1). In this part of the motion the new filter and the EKF have a similar performance. Eventually the noise process in (5.1) moves Θ away from the origin and into a region $\theta > 0$ where the nonlinear term $b\Theta_t^2\,\mathrm{d}t$ in (5.1) becomes significant.

Then Θ begins to move rapidly rightward. In the example the filter is stopped shortly after Θ begins the rapid rightward motion.

6. Conclusion

A class of nonlinear systems is described for which there exists an efficient optimal filter for computing the probability density function of the system state conditional on the measurements. Such systems were first described by Beneš (1981). The class of Beneš systems is extended to include certain nonlinear systems in which the drift velocity is a linear function plus the gradient of a scalar potential.

The Beneš filter is closely related to the Kalman–Bucy filter. Marcus (1984) shows in § 5 *b* that the state space can be reparametrized such that the Beneš filter is reduced to the Kalman–Bucy filter. The advantage of the development of the Beneš filter given in this paper over such a reduction is that the different terms appearing in the expression for the conditional density are directly related to physical quantities. This facilitates the implementation of the filter and also suggests ways in which the filter can be adapted to systems outside the class for which it generates the exact unnormalized conditional density.

The Beneš filter is the basis of a new class of suboptimal, finite-dimensional filters which are applicable to a wide range of nonlinear systems. One of the new filters is applied to a typical tracking problem drawn from computer vision and its performance compared with that of the EKF. The new filter appears to perform better when the evolution of the system state becomes strongly nonlinear. The full development of this new approach to nonlinear filtering is given by Maybank (1996).

Thanks are due to Roger Brockett for information about finite-dimensional filters. Part of this work was carried out during the Computer Vision Programme at the Isaac Newton Institute, Cambridge, UK (July–December 1993). The author thanks the Institute for its hospitality. The work was funded by the Royal Society through an Industrial Fellowship held at the Department of Engineering Science, University of Oxford, UK, for the period 1993–1995. During this time the author was an employee of the Hirst Division, GEC-Marconi Materials Technology Ltd, Elstree Way, Borehamwood, Hertfordshire, WD6 1RX. The author is grateful to the referees for their comments.

References

Abramovitz, M. & Stegun, I. A. 1965 *Handbook of mathematical functions.* New York: Dover.

Ayache, N. 1991 *Artificial vision for mobile robots: stereo vision and multisensory perception.* Cambridge, MA: MIT Press.

Bellman, R. 1969 *Stability theory of differential equations.* New York: Dover.

Beneš, V. E. 1981 Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5**, 65–92.

Bensoussan, A. 1992 *Stochastic control of partially observable systems.* Cambridge University Press.

Brockett, R. W. 1980 Remarks on finite dimensional nonlinear estimation. In *Analyse des systèmes* (Bordeaux, 11–16 September 1978). *Astérisque* **75–76**. Société Mathématique de France.

Brockett, R. W. & Clark, J. M. C. 1980 The geometry of the conditional density equation. In *Analysis and optimisation of stochastic systems* (ed. O. L. R. Jacobs, M. H. A. Davis, M. A. H. Dempster, C. J. Harris & P. C. Parks). (The Institute of Mathematics and Its Applications Conference Series.) Academic Press.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Broida, T. J. & Chellappa, R. 1991 Estimating the kinematics and structure of a rigid object from a sequence of monocular images. *PAMI* **13**, 497–513.

Chiou, W.-L. & Yau, S. S.-T. 1994 Finite dimensional filters with nonlinear drift. II Brockett's problem on classification of finite-dimensional estimation algebras. *SIAM J. Control Optimization* **32**, 297–310.

Choi, C. H. & Laub, A. J. 1990 Efficient matrix-valued algorithms for solving stiff Riccati differential equations. *IEEE Trans. Automatic Control* **35**, 770–776.

Faugeras, O. D. 1993 *Three dimensional computer vision: a geometric approach.* Cambridge, MA: MIT Press.

Karatzas, I. & Shreve, S. E. 1988 *Brownian motion and stochastic calculus.* (Graduate Texts in Mathematics vol. 113.) New York: Springer.

Kenney, C. S. & Leipnik, R. B. 1985 Numerical integration of the differential matrix Riccati equation. *IEEE Trans. Automatic Control* **30**, 962–970.

Marcus, S. I. 1984 Algebraic and geometric methods in nonlinear filtering. *SIAM J. Control Optimization* **22**, 817–844.

Maybank, S. J. 1996 Recursive approximation to the optimal filter. (Submitted.)

Maybeck, P. S. 1979*a* *Stochastic models, estimation and vontrol*, vol. 1. Mathematics in Science and Engineering Series, vol. 141-1. San Diego, CA: Academic Press.

Maybeck, P. S. 1979*b* *Stochastic models, estimation and control*, vol. 2. Mathematics in Science and Engineering Series, vol. 141-2. New York: Academic Press.

Maybeck, P. S. 1982 *Stochastic models, estimation and control*, vol. 3. Mathematics in Science and Engineering Series, vol. 141-3. London: Academic Press.

Øksendal, B. 1992 *Stochastic differential equations: an introduction with applications*, 3rd edn. Berlin: Springer.

Pardoux, E. 1991 Filtrage nonlinéaire et équations aux dérivées partielles stochastiques associeés. In *Ecole d'Eté de Probabilités de Saint-Flour XIX* (Lecture Notes in Mathematics, vol. 1464.) Berlin: Springer.

Revuz, D. & Yor, M. 1991 *Continuous martingales and Brownian motion.* (Grundlehren der mathematischen Wissenschaften, vol. 293.) Berlin: Springer.

Rogers, L. C. G. & Williams, D. W. 1987 *Diffusions, Markov processes and martingales*, vol. 2. *Itô calculus.* Chichester: Wiley.

Stroock, D. W. & Varadhan, S. R. S 1979 *Multidimensional diffusion processes.* (Grundlehren der mathematischen Wissenschaften vol. 233.) Berlin: Springer.

Wolfram, S. 1991 *Mathematica: a system for doing mathematics by computer*, 2nd edn. Redwood City, CA: Addison Wesley.

Yau, S. S.-T. 1994 Finite dimensional filters with nonlinear drift. I. A class of filters including both Kalman–Bucy filters and Beneš filters. *J. Math. Systems Estimation Control* **4**, 181–203.

Zakaï, M. 1969 On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **11**, 230–243.

Received 11 March 1994; revised 21 February 1995; accepted 2 August 1995

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES