Chapter 8

Continuous Time: Filtering Algorithms

In this chapter, we describe various algorithms for determination of the filtering distribution μ^t in continuous time. We begin in Section 8.1 with the Kalman–Bucy filter, which provides an exact algorithm for linear problems. Since the filtering distribution is Gaussian in this case, the distribution is entirely characterized by the mean and covariance; the algorithm comprises a system of differential equations for the mean and the covariance. In Section 8.2, we discuss the approximate Gaussian methods introduced in Section 4.2 in the discrete-time setting. Similarly to the case of the Kalman–Bucy filter, we again obtain a differential equation for the mean; for the extended Kalman (ExKF) filter, we also obtain an equation for the covariance, and for the ensemble Kalman filter (EnKF), we have a system of differential equations coupled through their empirical mean and covariance. In Section 8.3, we discuss how the particle filter methodology introduced in Section 4.3 extends to the continuous case, while in Section 8.4, we study the long-time behavior of some of the filtering algorithms discussed in the previous sections. Finally, in Section 8.5, we present some numerical illustrations and conclude with bibliographic notes and exercises in Sections 8.6 and 8.7 respectively.

It is helpful to recall the form of the continuous-time data-assimilation problem. The signal is governed by the SDE from (6.4):

$$\begin{aligned} \frac{dv}{dt} &= f(v) + \sqrt{\Sigma_0}\frac{dB_v}{dt}, \\ v(0) &\sim N(m_0, C_0); \end{aligned}$$

the data is generated by the SDE (6.5):

$$\begin{aligned} \frac{dz}{dt} &= h(v) + \sqrt{\Gamma_0}\frac{dB_z}{dt}, \\ z(0) &= 0. \end{aligned}$$

We let z^t denote $\{z(s)\}_{s\in[0,t]}$, the data accumulated up to time t, and we are interested in the probability measure μ^t governing $v(t)|z^t$, and in particular, in updating this measure sequentially in time. This is the filtering problem. In principle, the Kushner–Stratonovich and Zakai equations provide the solution to the filtering problem, but in general, they do not have closed-form solutions. Thus algorithms are required to approximate their solution. The filtering algorithms that we describe in the remainder of the section attempt to do this.

Electronic supplementary material The online version of this chapter (doi: 10.1007/978-3-319-20325-6_8) contains supplementary material, which is available to authorized users.

K. Law et al., *Data Assimilation*, Texts in Applied Mathematics 62,
DOI 10.1007/978-3-319-20325-6_8

We highlight here the fact that the Kushner–Stratonovich and Zakai equations are stochastic PDEs in spatial dimension n, the size of the state space. Thus their solution poses formidable challenges for high-dimensional problems.

8.1 Linear Gaussian Problems: The Kalman–Bucy Filter

Although the Kushner–Stratonovich and Zakai equations do not, in general, have closed-form solutions, they do have such solutions for linear models; this is entirely analogous to the discrete-time setting and stems from the fact that the desired filtering distribution is Gaussian. The resulting algorithm for the mean and covariance is the Kalman–Bucy filter, which we now describe.

Consider equations (6.4) and (6.5), where $f(v) = Lv$, $h(v) = Hv$, for $L \in \mathbb{R}^{n\times n}$, $H \in \mathbb{R}^{m\times n}$ of full rank m. Then (6.4) and (6.5) become

$$\frac{dv}{dt} = Lv + \sqrt{\Sigma_0}\frac{dB_v}{dt}, \quad v(0) \sim N(m_0, C_0), \tag{8.3a}$$

$$\frac{dz}{dt} = Hv + \sqrt{\Gamma_0}\frac{dB_w}{dt}, \quad z(0) = 0. \tag{8.3b}$$

Theorem 8.1. *The filtering distribution μ^t for $v(t)|z^t$ governed by (8.3) is Gaussian with mean m and covariance C solving the Kalman–Bucy filter*

$$\frac{dm}{dt} = Lm + CH^T\Gamma_0^{-1}\left(\frac{dz}{dt} - Hm\right), \quad m(0) = m_0,$$

$$\frac{dC}{dt} = LC + CL^T + \Sigma_0 - CH^T\Gamma_0^{-1}HC, \quad C(0) = C_0.$$

Sketch Proof We give references to the rigorous derivation of Kalman–Bucy filtering in the bibliographic notes at the end of the chapter. Here we derive the filter by a formal discretization argument, with no proof, since this provides a useful way to understand the structure of the filter. In particular, we consider the discrete-time Kalman filter of Section 4.1, and the prediction and analysis steps given by (4.4), (4.5) and (4.7), (4.8) respectively, with $M = I + \tau L$ and the other scalings detailed in (6.1). The prediction steps (4.4) and (4.5) give, to leading order in τ,

$$\widehat{m}_{j+1} = m_j + \tau L m_j, \tag{8.4}$$

$$\begin{aligned}\widehat{C}_{j+1} &= (I + \tau L)C_j(I + \tau L)^T + \tau\Sigma_0 \\ &= C_j + \tau(LC_j + C_jL^T + \Sigma_0) + \mathcal{O}(\tau^2).\end{aligned} \tag{8.5}$$

Now using the analysis step from Corollary 4.2, again to leading order and substituting (8.4) and (8.5), we obtain

$$\begin{aligned}
d_{j+1} &= (y_{j+1} - Hm_j) + \mathcal{O}(\tau), \\
S_{j+1} &= \frac{1}{\tau}\Gamma_0^{-1} + \mathcal{O}(1), \\
K_{j+1} &= \tau C_j H^T\Gamma_0^{-1} + \mathcal{O}(\tau^2), \\
m_{j+1} &= m_j + \tau L m_j + \tau C_j H^T\Gamma_0^{-1}(y_{j+1} - Hm_j), \\
C_{j+1} &= C_j + \tau(LC_j + C_jL^T + \Sigma_0) - \tau C_j H^T\Gamma_0^{-1}HC_j + \mathcal{O}(\tau^2).
\end{aligned}$$

Recalling that $y_{j+1} = \tau^{-1}(z_{j+1} - z_j)$ and passing to the limit gives

$$\frac{dm}{dt} = Lm + CH^T\Gamma_0\left(\frac{dz}{dt} - Hm\right),$$
$$\frac{dC}{dt} = LC + CL^T + \Sigma_0 - CH^T\Gamma_0^{-1}HC,$$

which concludes the proof sketch. □

8.2 Approximate Gaussian Filters

Here we discuss the family of approximate Gaussian filtersintroduced in Section 4.2, generalizing to continuous time. Our aim is to ascertain the form of continuous-time limits under the scalings detailed in (6.1). Our aim is not to prove theorems about the limiting process, but rather to give an understanding of what the natural continuous-time limiting processes are. We thus use the environment "result" rather than "theorem" to highlight the fact that the forms of the continuous-time limits are derived only formally and not, currently, rigorously proved in the published literature; however, it would not be difficult to make rigorous proofs based on these results.

The starting point of our investigations is equation (4.13), which is based on the assumption that $\mathbb{P}(u_{j+1}|Y_j) = N(\Psi(m_j), \widehat{C}_{j+1})$. In addition to the expression for the update of m_j (4.13a) using Bayes's, rule one sees that the new covariance C_{j+1} satisfies

$$C_{j+1} = (I - K_{j+1}H)\widehat{C}_{j+1}. \tag{8.6}$$

We will now proceed in a similar fashion as in the case of the Kalman–Bucy filter: we derive a differential equation for the time evolution of the mean and covariance of the different approximate Gaussian filters studied in Section 4.2. The resulting Gaussian measures should be viewed as attempts to approximate the true filtering distribution μ^t.

8.2.1. 3DVAR

Result 8.2. *Consider the filtering distribution for the 3DVAR algorithm* (4.14) *arising in the case of deterministic dynamics ($\Sigma = 0$) and linear observations($h(v) = Hv$). Under the scalings detailed in* (6.1)*, and in the continuous-time limit ($\tau \to 0$), the corresponding limiting filtering distribution is Gaussian with mean m and covariance C satisfying*

$$\frac{dm}{dt} = f(m) + CH^T\Gamma_0^{-1}\left(\frac{dz}{dt} - Hm\right), \quad m(0) = m_0, \tag{8.7a}$$
$$\frac{dC}{dt} = 0, \quad C(0) = \widehat{C}. \tag{8.7b}$$

Derivation We begin our derivation by observing that (4.13) implies that

$$m_{j+1} = \Psi(m_j) + \widehat{C}_{j+1}H^T(\Gamma + H\widehat{C}_{j+1}H^T)^{-1}\big(y_{j+1} - H\Psi(m_j)\big).$$

We now apply the scalings appropriate for a continuous-time limit and in particular set

$$\Psi(m) = m + \tau f(m) + \mathcal{O}(\tau^2).$$

In addition, for the case of 3DVAR, we have $\widehat{C}_{j+1} = \widehat{C}$ for all j, which implies that

$$\widehat{C}_{j+1} - \widehat{C}_j = 0.$$

Finally, using the scalings from equation (6.1) and recalling that $y_{j+1} = \tau^{-1}(z_{j+1} - z_j)$, we have that

$$\begin{aligned}
\frac{m_{j+1} - m_j}{\tau} &= f(m_j) + \widehat{C}_{j+1} H^T (\Gamma_0 + \tau H \widehat{C}_{j+1} H^T)^{-1} \left(\frac{z_{j+1} - z_j}{\tau} - H m_j \right) + \mathcal{O}(\tau), \\
\frac{\widehat{C}_{j+1} - \widehat{C}_j}{\tau} &= 0.
\end{aligned}$$

By taking the limit $\tau \to 0$, we obtain

$$\begin{aligned}
\frac{dm}{dt} &= f(m) + C H^T \Gamma_0^{-1} \left(\frac{dz}{dt} - Hm \right), \quad m(0) = m_0, \\
\frac{dC}{dt} &= 0, \quad C(0) = \widehat{C},
\end{aligned}$$

where we have identified C_0 with $\widehat{C}$. This completes our derivation. □

8.2.2. Extended Kalman Filter

The 3DVAR algorithm imposes a fixed covariance on the model in the prediction step of the algorithm, implying also a fixed covariance in the analysis step. The extended Kalman filter attempts to improve on this by propagating the covariance in the prediction step according to the linearized dynamics. In the continuous-time limit, we obtain the following.

Result 8.3. *Consider the filtering distribution for the extended Kalman Filter from Section 4.2.2 in the case of stochastic dynamics and linear observations ($h(v) = Hv$). Under the scalings detailed in* (6.1) *and in the continuous-time limit ($\tau \to 0$), the corresponding limiting filtering distribution is Gaussian with mean m and covariance C satisfying*

$$\begin{aligned}
\frac{dm}{dt} &= f(m) + C H^T \Gamma_0^{-1} \left(\frac{dz}{dt} - Hm \right), \quad m(0) = m_0, \\
\frac{dC}{dt} &= Df(m)C + C(Df(m))^T + \Sigma_0 - C H^T \Gamma_0^{-1} H C, \quad C(0) = C_0.
\end{aligned}$$

Derivation From the formulas given in Section 4.2.2, we have

$$\widehat{C}_{j+1} = D\Psi(m_j) C_j D\Psi(m_j)^T + \Sigma,$$

and using (4.13) and (8.6),

$$m_{j+1} = \Psi(m_j) + \widehat{C}_{j+1}H^T(\Gamma + H\widehat{C}_{j+1}H^T)^{-1}(y_{j+1} - H\Psi(m_j)),$$
$$\widehat{C}_{j+1} = D\Psi(m_j)(I - K_jH)\widehat{C}_jD\Psi(m_j)^T + \Sigma.$$

To deduce the continuous-time limit, we set $\Psi(m) = m + \tau f(m) + \mathcal{O}(\tau^2)$ and impose (6.1). This yields

$$D\Psi(m_j) = I + \tau Df(m_j) + \mathcal{O}(\tau^2),$$
$$(I - K_jH) = I - \tau\widehat{C}_jH^T(\Gamma_0 + \tau H\widehat{C}_jH^T)^{-1}H + \mathcal{O}(\tau^2).$$

Combining the two previous sets of equations and recalling that $y_{n+1} = \tau^{-1}(z_{n+1} - z_n)$, we obtain

$$\frac{m_{j+1} - m_j}{\tau} = f(m_j) + \widehat{C}_{j+1}H^T(\Gamma_0 + \tau H\widehat{C}_{j+1}H^T)^{-1}\left(\frac{z_{j+1} - z_j}{\tau} - Hm_j\right) + \mathcal{O}(\tau),$$
$$\frac{\widehat{C}_{j+1} - \widehat{C}_j}{\tau} = Df(m_j)\widehat{C}_j + \widehat{C}_jDf(m_j)^T - \widehat{C}_jH^T(\Gamma_0 + \tau H\widehat{C}_jH^T)^{-1}H\widehat{C}_j + \Sigma_0 + \mathcal{O}(\tau).$$

By taking the limit $\tau \to 0$, we obtain

$$\frac{dm}{dt} = f(m) + CH^T\Gamma_0^{-1}\left(\frac{dz}{dt} - Hm\right), \quad m(0) = m_0,$$
$$\frac{dC}{dt} = Df(m)C + C(Df(m))^T + \Sigma_0 - CH^T\Gamma_0^{-1}HC, \quad C(0) = C_0,$$

as required. □

8.2.3. Ensemble Kalman Filter

As already discussed in Chapter 4, the ensemble Kalman filter differs from the extended Kalman filter and 3DVAR in that instead of using an appropriate minimization procedure to update a single estimate of the mean, a minimization principle is used to generate an ensemble of particles all of which satisfy the model/data compromise inherent in the minimization; these are coupled through the empirical covariance used to weight the minimization. Thus in studying the EnKF in continuous time, it is natural to derive an SDE for each of the particles, instead of deriving a single equation for the mean and the covariance as in Results 8.2 and 8.3. In deriving the continuous-time limit for each of the particles, it will be useful to rewrite the ensemble Kalman filter from Section 4.2.3 using the family of minimization principles $\mathsf{I}_{\text{filter},n}n$ given by (4.15), with $n = 1, \ldots, N$ indexing the particles. Using such an interpretation leads to the following equation:

$$\widehat{C}_{j+1}^{-1}(I + \widehat{C}_{j+1}H^T\Gamma^{-1}H)v_{j+1}^{(n)} = \widehat{C}_{j+1}^{-1}\widehat{v}_{j+1}^{(n)} + H^T\Gamma^{-1}y_{j+1}^{(n)},$$

with $\widehat{C}_{j+1}$ the predictive covariance found from the properties of the predictive ensemble; in this derivation we simply assume that $\widehat{C}_{j+1}$ is invertible and return to address this issue after the derivation. Applying $\widehat{C}_{j+1}$ allows us to rewrite the analysis step of the ensemble Kalman Filter in the following form:

$$(I + \widehat{C}_{j+1} H^T \Gamma^{-1} H) v_{j+1}^{(n)} = \widehat{v}_{j+1}^{(n)} + \widehat{C}_{j+1} H^T \Gamma^{-1} y_{j+1}^{(n)}, \tag{8.8a}$$

$$y_{j+1}^{(n)} = y_{j+1} + \eta_{j+1}^{(n)}, \quad n = 1, \cdots, N, \tag{8.8b}$$

where $\widehat{v}_{j+1}^{(n)}$, $\widehat{m}_{j+1}$, and $\widehat{C}_{j+1}$ are given by the prediction step detailed in Section 4.2.3. We now have the following result:

Result 8.4. *Consider the ensemble Kalman Filter from Section 4.2.3 in the case of stochastic dynamics and linear observations ($h(v) = Hv$). Under the scalings detailed in* (6.1) *and in the continuous-time limit ($\tau \to 0$), the particles evolve according to the following set of coupled SDEs, for $n = 1, \cdots, N$:*

$$\frac{dv^{(n)}}{dt} = f(v^{(n)}) + C(v) H^T \Gamma_0^{-1} \Big(\frac{dz^{(n)}}{dt} - H v^{(n)} \Big) + \Sigma_0^{1/2} dB_v, \tag{8.9a}$$

$$\frac{dz^{(n)}}{dt} = Hv + \Gamma_0^{1/2} \Big(\frac{dW^{(n)}}{dt} + \frac{dB_z}{dt} \Big), \tag{8.9b}$$

where $W^{(1)}, \cdots, W^{(N)}, B_z, B_u$ are mutually independent standard Wiener processes. The mean $m(v)$ and covariance $C(v)$ are defined empirically from the particles $v = \{v^{(n)}\}_{n=1}^N$ as follows:

$$m(v) = \frac{1}{N} \sum_{n=1}^{N} v^{(n)}, \tag{8.10a}$$

$$C(v) = \frac{1}{N-1} \sum_{n=1}^{N} (v^{(n)} - m)(v^{(n)} - m)^T. \tag{8.10b}$$

Derivation We begin our derivation by noting the formulation of the analysis step given in (8.8) and employing the definition of the particle prediction step to give

$$\begin{aligned} v_{j+1}^{(n)} - v_j^{(n)} &= \widehat{v}_{j+1}^{(n)} - v_j^{(n)} - \widehat{C}_{j+1} H^T \Gamma^{-1} H v_{j+1}^{(n)} + \widehat{C}_{j+1} H^T \Gamma^{-1} y_{j+1}^{(n)} \\ &= \Psi(v_{j+1}^{(n)}) - v_j^{(n)} - \widehat{C}_{j+1} H^T \Gamma^{-1} H v_{j+1}^{(n)} + \widehat{C}_{j+1} H^T \Gamma^{-1} y_{j+1}^{(n)} + \xi_j^{(n)}. \end{aligned}$$

Now using the scalings from equation (6.1), we have that

$$\Psi(v) = v + \tau f(v) + \mathcal{O}(\tau^2), \quad \xi_j^{(n)} = \sqrt{\tau} \Sigma_0^{1/2} \widehat{\xi}_j^{(n)},$$

and thus we obtain

$$v_{j+1}^{(n)} - v_j^{(n)} = \tau (f(v_j^{(n)}) - \widehat{C}_{j+1} H^T \Gamma_0^{-1} H v_{j+1}^{(n)}) + \tau \widehat{C}_{j+1} H^T \Gamma_0^{-1} y_{j+1}^{(n)} + \sqrt{\tau} \Sigma_0^{1/2} \widehat{\xi}_{j+1}^{(n)} + \mathcal{O}(\tau^2),$$

where $\widehat{\xi}_j^{(n)}$ is $N(0, I)$ distributed. Now using the rescaling $y_{j+1}^{(n)} = \tau^{-1}(z_{j+1}^{(n)} - z_j^{(n)})$, we have the coupled difference equations

$$v_{j+1}^{(n)} - v_j^{(n)} = \tau (f(v_j^{(n)}) - \widehat{C}_{j+1} H^T \Gamma_0^{-1} H v_{j+1}^{(n)}) + \widehat{C}_{j+1} H^T \Gamma_0^{-1} (z_{j+1}^{(n)} - z_j^{(n)}) + \sqrt{\tau} \Sigma_0^{1/2} \widehat{\xi}_{j+1}^{(n)}, \tag{8.11}$$

$$z_{j+1}^{(n)} - z_j^{(n)} = \tau H v_{j+1} + \sqrt{\tau} \Gamma_0^{1/2} (\widehat{\eta}_{j+1}^{(n)} + \widehat{\eta}_{j+1}), \tag{8.12}$$

where $\widehat{\xi}_{j+1}^{(n)}$, $\widehat{\eta}_{j+1}^{(n)}$, and $\widehat{\xi}_{j+1}$ are independent $N(0, I)$-distributed random variables. In addition, we have

$$\widehat{m}_{j+1} = \frac{1}{N}\sum_{n=1}^{N} v_{j+1}^{(n)} + \mathcal{O}(\tau),$$

so in the limit of $\tau \to 0$, we have that $(\widehat{m}_{j+1}, \widehat{C}_{j+1}) \to (m, C)$ given by equation (8.10). Furthermore, we see that the coupled difference equations given by (8.11) form a mixed implicit–explicit Euler–Maruyama-type scheme for the system of SDEs (8.9), and thus in the limit of $\tau \to 0$, we obtain the desired equations. □

Recall that in the preceding derivation, we made the assumption that $\widehat{C}_{j+1}$ is invertible. However, this might not be the case; indeed, it cannot be the case if the dimension of the state space exceeds the number of particles. However, one can still obtain the key equations (8.8) in this case, by applying the following lemma to the analysis formula in Section 4.2.3.

Lemma 8.5. *Assume that Γ is invertible. Then*

$$(I - K_{j+1}H)^{-1} = (I + \widehat{C}_{j+1}H^T\Gamma^{-1}H)$$

and

$$(I - K_{j+1}H)^{-1}K_{j+1} = \widehat{C}_{j+1}H^T\Gamma^{-1}.$$

Proof We begin the proof by noting that using the Woodbury matrix identity from Lemma 4.4, we have, for

$$S_{j+1} := (\Gamma + H\widehat{C}_{j+1}H^T),$$

that

$$S_{j+1}^{-1} = \Gamma^{-1} - \Gamma^{-1}H(\widehat{C}_{j+1}^{-1} + H^T\Gamma^{-1}H)^{-1}H^T\Gamma^{-1},$$

where we note that S_{j+1} is invertible because Γ is. Assume that C_{j+1} is invertible; we will relax this assumption below. Thus $Z := I - K_{j+1}H = I - \widehat{C}_{j+1}H^TS_{j+1}^{-1}H$ can be written as

$$\begin{aligned}
Z &= I - \widehat{C}_{j+1}H^T\Gamma^{-1}H - \widehat{C}_{j+1}H^T\Gamma^{-1}H(\widehat{C}_{j+1}^{-1} + H^T\Gamma^{-1}H)^{-1}H^T\Gamma^{-1}H \\
&= I - \widehat{C}_{j+1}B - \widehat{C}_{j+1}B(\widehat{C}_{j+1}^{-1} + B)^{-1}B \\
&= I - \widehat{C}_{j+1}B(I - (\widehat{C}_{j+1}^{-1} + B)^{-1}B) \\
&= I - \widehat{C}_{j+1}B(\widehat{C}_{j+1}^{-1} + B)^{-1}\widehat{C}_{j+1}^{-1},
\end{aligned}$$

where $B = H^T\Gamma^{-1}H$. Manipulating this expression further, we see that

$$Z(I + \widehat{C}_{j+1}B) = I.$$

This identity may be derived even if C_{j+1} is not invertible simply by adding ϵI to C_{j+1} in the preceding derivation and then letting $\epsilon \to 0$. The preceding identity implies that

$$(I - K_{j+1}H)^{-1} = (I + \widehat{C}_{j+1}B) = (I + \widehat{C}_{j+1}H^T\Gamma^{-1}H),$$

which concludes the proof for our first equation. Now using this equation, it is easy to see that

$$K_{j+1} = H^{-1} - (I + \widehat{C}_{j+1}H^T\Gamma^{-1}H)^{-1}H^{-1}$$

and thus

$$(I - K_{j+1}H)K_{j+1} = (I + \widehat{C}_{j+1}H^T\Gamma^{-1}H)H^{-1} - H^{-1} = \widehat{C}_{j+1}H^T\Gamma^{-1},$$

which concludes the proof for our second equation. □

8.3 The Particle Filter

The particle filter in continuous time faces many of the same issues arising in discrete time, as outlined in Section 4.3. In the basic version of the method, analogous to the bootstrap filter of Section 4.3.2, the particles evolve in continuous time according to the SDE (6.4). The particles are weighted according to (6.31), which reflects the change of measure required to take the solution of the SDE into the solution of the SDE conditioned on the observations given by (6.5). As in discrete time, it is helpful to resample from the resulting distribution in order to obtain an approximation with significant weight near the data. References to detailed literature on the subject are given in Section 8.6.

8.4 Large-Time Behavior of Filters

Here we provide some simple examples that illustrate issues relating to the large-time behavior of filters. The discussion from the preamble of Section 4.4 also applies here in continuous time. In particular, the approximate Gaussian filters do not perform well as measured by the Bayesian quality assessment test of Section 2.7 but may perform well as measured by the signal estimation quality assessment test. Also, similarly to the situation in discrete time, the Kalman–Bucy filter for linear problems and the particle filter give accurate approximations of the true posterior distribution, in the latter case in the large-particle limit. The purpose of this section is to illustrate these issues.

8.4.1. The Kalman–Bucy Filter in One Dimension

We consider the case of one-dimensional deterministic linear dynamics (8.3) with $\Sigma_0 = 0$ and

$$f(v) = \ell v, \quad h(v) = v,$$

while we will also assume that

$$\Gamma_0 = \gamma^2, C_0 = c_0.$$

Thus the filter aims to reconstruct signal $v(t)$ solving the equation

$$\frac{dv}{dt} = \ell v$$

from knowledge of $\{z(s)\}_{0\le s\le t}$, where

$$z(s) = \int_0^s v(\tau)d\tau + \gamma B_w(s).$$

Even though the variance of $B_w(t)$ grows linearly with t, we will show that the variance of the Kalman–Bucy filter is asymptotically zero, for $\ell \leq 0$, or of order $\mathcal{O}(\gamma^2)$ otherwise.

With these definitions, the Kalman–Bucy filter of Theorem 8.1 becomes

$$\frac{dm}{dt} = \ell m + \gamma^{-2} c \left(\frac{dz}{dt} - m \right), \quad m(0) = m_0,$$
$$\frac{dc}{dt} = 2\ell c - \gamma^{-2} c^2, \quad c(0) = c_0.$$

Here m denotes the mean of the filter, and c, the variance.

We notice that the equation for the variance evolves independently of that for the mean, and independently of the data. Furthermore, if we define the precision p to be the inverse of c, then straightforward calculation reveals that p solves the linear equation

$$\frac{dp}{dt} = -2\ell p + \gamma^{-2}.$$

For $\ell \neq 0$, this has exact solution

$$p(t) = \exp(-2\ell t)\frac{1}{c_0} + \Big(1 - \exp(-2\ell t)\Big)\frac{1}{2\ell\gamma^2},$$

while for $\ell = 0$, we see that

$$p(t) = \frac{1}{c_0} + \frac{t}{\gamma^2}.$$

Thus for $\ell \leq 0$, we have $p(t) \to \infty$ as $t \to \infty$. and the asymptotic variance is zero, while for $\ell > 0$, the asymptotic variance is $2\ell\gamma^2$. In particular, if the observational variance γ^2 is small, then the asymptotic variance of the Kalman filter is $\mathcal{O}(\gamma^2)$, even when $\ell > 0$, so that the dynamics is unstable. The key point to observe, then, is that asymptotic uncertainty is small, independently of whether the underlying deterministic dynamics is stable. In words, observation can stabilize uncertainty growth in unstable systems. We study the behavior of the error in the mean in the exercises at the end of this chapter.

8.4.2. The 3DVAR Filter

Recall the 3DVAR continuous filtering algorithm (8.7). We will study the behavior of this algorithm with data $z := \{z^\dagger(t)\}_{t\in[0,\infty)}$ constructed as follows. We let $\{v^\dagger(t)\}_{t\in[0,\infty)}$ denote the exact solution of the equations (6.4) in the case $\Sigma_0 = 0$:

$$\frac{dv^\dagger}{dt} = f(v^\dagger), \quad v^\dagger(0) = v_0^\dagger. \tag{8.13}$$

We assume that the data $z^\dagger$ is a single realization of the SDE (6.4) with $v = v^\dagger$ and in the case $h(\cdot) = H\cdot$:

$$\frac{dz^\dagger}{dt} = Hv^\dagger + \sqrt{\Gamma_0}\frac{dB_z}{dt}, \quad z^\dagger(0) = 0. \tag{8.14}$$

In order to study the continuous-time 3DVAR filter, we eliminate $z = z^\dagger$ in equation (8.7) using (8.14) to obtain

$$\frac{\mathrm{d}m}{\mathrm{d}t} = f(m) + CH^T\Gamma_0^{-1}H(v^\dagger - m) + CH^T\Gamma_0^{-\frac{1}{2}}\frac{\mathrm{d}B_z}{\mathrm{d}t}. \tag{8.15}$$

In the following theorem, expectation is with respect to the Brownian motion B_z entering the data equation (8.14).

Theorem 8.6. *Let m solve equation (8.15), let $v^\dagger$ solve equation (8.13), assume that f is globally Lipschitz with constant L and that there exist $\lambda > 0$ and $\epsilon > 0$ such that*

$$\langle CH^T\Gamma_0^{-1}Ha, a\rangle \geq (L + \frac{1}{2}\lambda)|a|^2, \forall a \in \mathbb{R}^n,$$
$$\mathrm{Tr}(\Gamma_0^{-\frac{1}{2}}HC^2H^T\Gamma_0^{-\frac{1}{2}}) \leq \epsilon^2.$$

Then the error in the 3DVAR filter satisfies

$$\mathbb{E}|m(t) - v(t)|^2 \leq e^{-\lambda t}|m(0) - v(0)|^2 + \frac{\epsilon^2}{\lambda}(1 - e^{-\lambda t}). \tag{8.16}$$

Thus

$$\mathrm{limsup}_{t\to\infty}\mathbb{E}|m(t) - v(t)|^2 \leq \frac{\epsilon^2}{\lambda}. \tag{8.17}$$

Proof Define $\delta = m - v^\dagger$ and subtract equation (8.13) from equation(8.15) and apply the Itô formula to $|\delta|^2$ to obtain

$$\frac{1}{2}d|\delta|^2 + \langle CH^T\Gamma_0^{-1}H\delta, \delta\rangle dt \leq \langle f(m) - f(v^\dagger), \delta\rangle + \langle \delta, CH^T\Gamma_0^{-\frac{1}{2}}dB_z\rangle + \frac{1}{2}\mathrm{Tr}\left(\Gamma_0^{-\frac{1}{2}}HC^2H^T\Gamma_0^{-\frac{1}{2}}\right)dt. \tag{8.18}$$

Using the Lipschitz property of f and the definition of λ and ϵ, and taking expectations, gives

$$\frac{d\mathbb{E}|\delta|^2}{dt} \leq -\lambda\mathbb{E}|\delta|^2 + \epsilon^2. \tag{8.19}$$

Use of the Gronwall inequality gives the desired result. □

It is interesting to consider the asymptotic behavior of this 3DVAR filter in the linear Gaussian case from the preceding subsection. We thus assume that

$$f(v) = \ell v, \quad h(v) = v,$$

and that

$$\Gamma_0 = \gamma^2, C = \eta^{-1}\gamma^2 I.$$

We assume that $\gamma^2 \ll 1$ and that $\ell > 0$. Our scaling of C proportional to γ^2 is motivated by the fact that the Kalman–Bucy filter has asymptotic variance on this scale. Theorem 8.6 applies, provided η is chosen to satisfy

$$\eta \leq (\ell + \frac{1}{2}\lambda)^{-1},$$

and then (8.17) shows that the mean-square error in the filter is bounded by $\eta^{-2}\lambda^{-1}\gamma^2$. Thus the asymptotic error of the 3DVAR filter scales in the same way as the error Kalman filter (which we study in the exercises) if the covariance C is tuned appropriately.

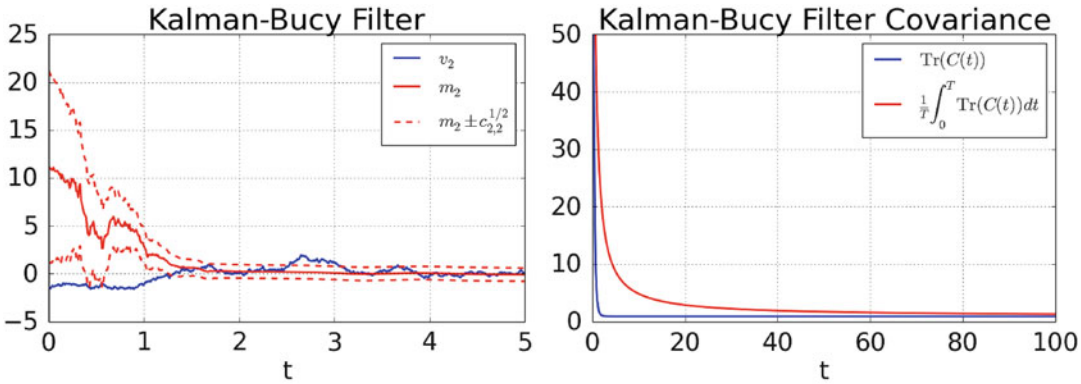
8.5 Illustrations

In this section, the output of the various filtering algorithms will be presented. The text is minimal, since the filters and their respective behaviors relative to one another are analogous to their discrete-time counterparts as detailed in Chapter 4.

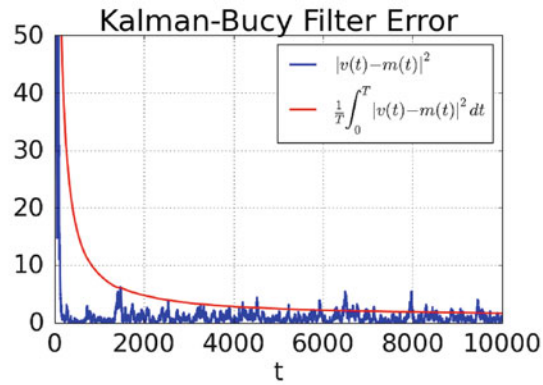
Figure 8.1 shows application of the Kalman Bucy filter to equation (6.21), in dimension $n = 2$, with

$$\Lambda = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}.$$

The observation operator is $H = (1, 0)^T$, so that the second component is unobserved. Figure 8.1a shows that the unobserved component is accurately recovered in the long-time limit, despite the fact that the filter is initialized far from the truth. Figures 8.1b and 8.1c show the asymptotic behavior of the covariance and the square of the Euclidean error between the mean and the true signal underlying the data; both are shown pathwise and in running average form.



(a) Solution. (b) Covariance.



(c) Error.

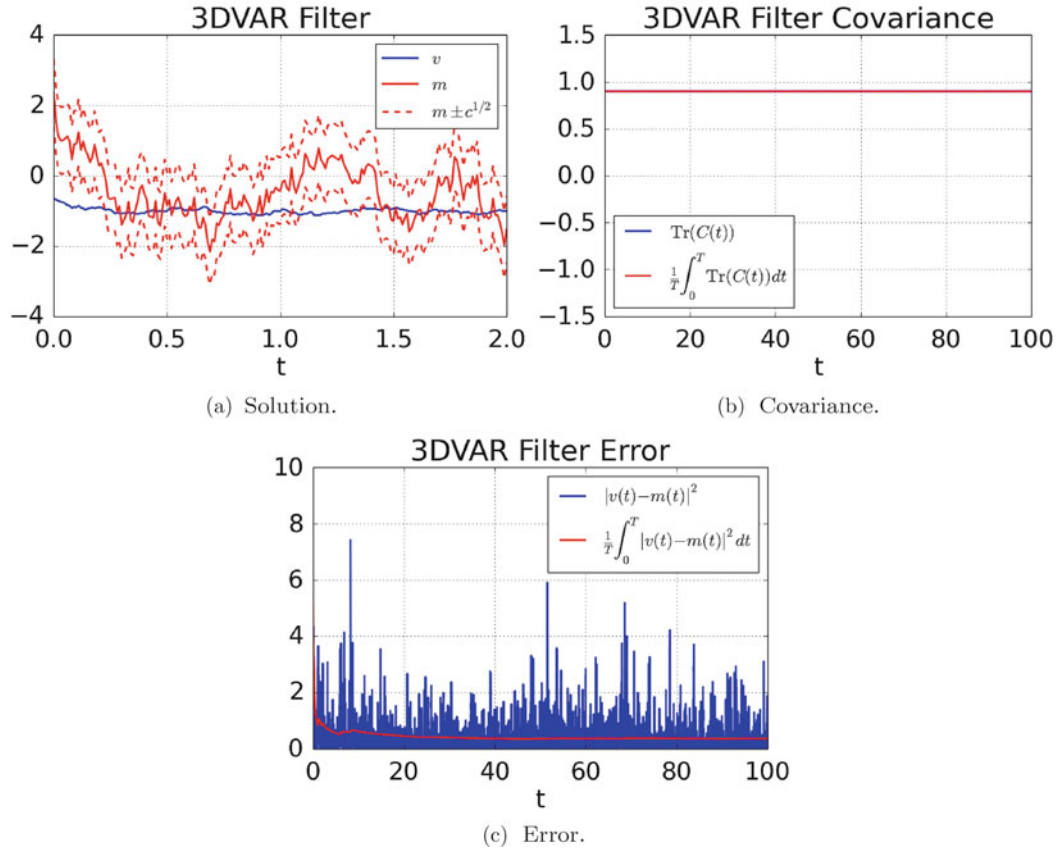
Fig. 8.1: Kalman–Bucy filter for Example 6.4 with Λ given by (8.20) with $\gamma = \sigma = 1$, as given in Section 8.20.

The next figures all concern the application of approximate Gaussian filters, together with the particle filter, to Example 6.6. Figures 8.2 and 8.3 concern application of the 3DVAR filter for $(\gamma, \sigma, \eta) = (0.3, 0.3, 0.1)$ and $(0.1, 0.3, 0.1)$ respectively, where γ^2 and σ^2 are the variance of the observations and the model dynamics respectively, and $C = \eta^{-1}\gamma^2$ is the fixed variance defining the filter. Notice the decrease the error from Figure 8.2 to 8.3 resulting from the decrease in γ. This is consistent with Theorem 8.6. Similar behavior is observed for the other filters, as shown in Figures 8.4–8.8, where the extended Kalman filter, two forms of the ensemble Kalman filter, and two forms of the particle filter are displayed. The extended Kalman filter has, arguably, the best performance; but it is a method that does not scale well to high-dimensional problems.

Remark 8.7. *It is important to remark that stability can be a significant issue, especially for filters in continuous time when complex nonlinear models are considered. As examples of this, consider the continuous-time extended Kalman filter applied to the chaotic dynamical systems Lorenz '63* (6.23) *and Lorenz '96* (6.24)*, as presented in Section 6.2. In both cases, whether instability is observed depends on the observation operator, for example. Figures 8.9 and 8.10 show the second component (left) and mean-square error (right) of the Lorenz '63 model* (6.23) *with* $\sigma = 2$ *and* $\gamma = 0.2$*. In both cases, we make a scalar linear observation* $h(v) = Hv$*. The difference is that in Figure 8.9, the observation operator is given by* $H = (1, 0, 0)$*, while in Figure 8.10, the observation operator is given by* $H = (0, 0, 1)$*. Figures 8.11 and 8.12 show the second component (left) and mean-square error (right) of the Lorenz '96 model* (6.24) *with* $\sigma = 1$ *and* $\gamma = 0.1$*. The difference is that in Figure 8.11, we observe two out of every three degrees of freedom, again linearly, while in Figure 8.12, we observe one out of every three degrees of freedom, also linearly. Note that for both Lorenz models, for observations that are insufficient to keep the filter close to the truth, large excursions in the error can occur. These large excursions can easily induce numerical instabilities and destabilize the algorithms unless care is taken in the choice of integrator and time step.* ♠

8.6 Bibliographic Notes

- In Section 8.1, we consider the continuous-time limit of the Kalman filter. This is the celebrated Kalman–Bucy filter, which was published in [80], the year after Kalman's original paper in the discrete-time setting [79], as described in Section 4.1. The Kalman–Bucy smoother concerns the related continuous-time smoothing problem; it may be solved by a continuous-time analogue of LU factorization, in which the first triangular sweep corresponds to application of the Kalman–Bucy filter—see [64] and the discussion at the end of the previous chapter. Theorem 8.1 can also be derived from the Kushner–Stratonovich or Zakai equation equation of Theorem 6.16 by computing moments.
- Section 8.2 concerns approximate Gaussian filters and, more specifically, filters that are derived as continuous-time limits of the discrete-time approximate Gaussian filters of Section 4.2. The idea of deriving continuous-time limits of the 3DVAR, extended, and ensemble Kalman filters was developed systemically in the papers [21, 82]. Furthermore, those papers, along with the paper [88], contain analyses of the large-time stability and accuracy of the filters, with results similar in spirit to Theorem 4.10.
- Section 8.3 concerns the continuous-time particle filter. This methodology for solving continuous-time filtering problems arising in SDEs can be viewed as constituting a particle method for solution of the underlying stochastic PDE (Zakai or Kushner–Stratonovich)



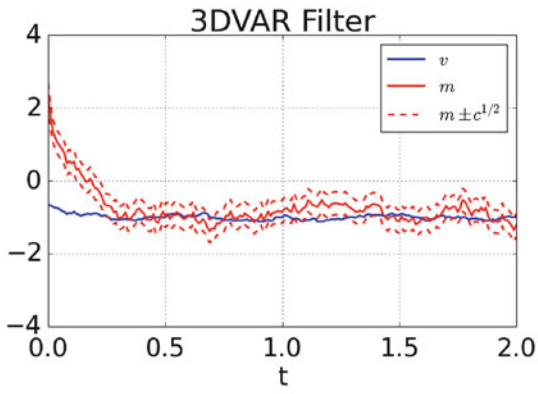
(a) Solution. (b) Covariance.

(c) Error.

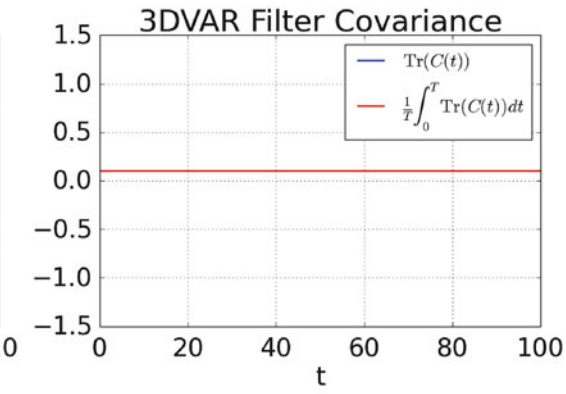
Fig. 8.2: Continuous 3DVAR filter for double-well Example 6.5 with $\gamma = \sigma = 0.3$, as given in Section 9.3.2.

introduced in Section 6.4 and governing the probability density of $v|z^t$. The method is analyzed in detail in [9] and in [45]; these two books also provide copious references to the literature on this subject.

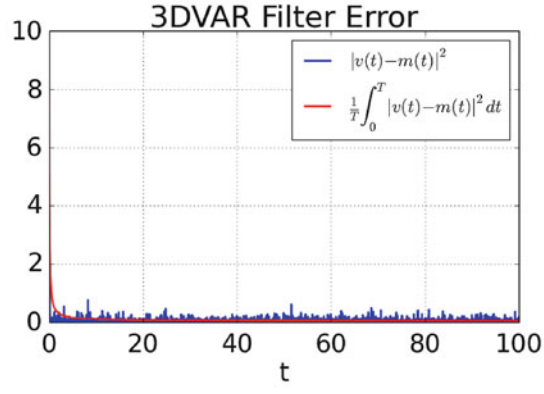
- Section 8.4 concerns stability of filters. In Section 8.4.1, we study the one-dimensional Kalman filter, while Section 8.4.2 concerns the 3DVAR filter. The example and theorem, respectively, covered in these two subsections are entirely analogous to those in discrete time in Section 4.4.
  The example from Section 8.4.1 concerning the Kalman–Bucy filter is very specific to one-dimensional deterministic dynamics. However, the general setting is thoroughly studied, as in discrete time, and the reader is directed to the book [86], concerning Riccati equations, for details. Theorem 8.6 is a simplified version of a result first proved in the context of the Navier–Stokes equation in [21], and then for the Lorenz '63 model in [88]. The first stability analysis in continuous time concerned noise-free data and a synchronization filter in which the observed variables are simply inserted into the governing equations for the unobserved variables, giving rise to a nonautonomous dynamical system for the unobserved variables [116]; this analysis forms the backbone of the analyses of the 3DVAR filter for the Navier–Stokes and Lorenz '63 models. The analysis was recently extended to the Lorenz '96 model in [87]. The synchronization filter is discussed in discrete time in



(a) Solution.



(b) Covariance.

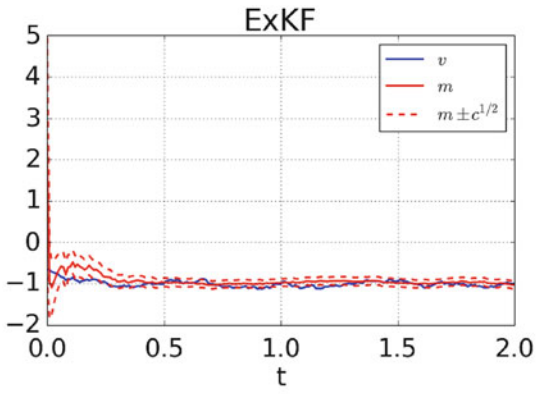


(c) Error.

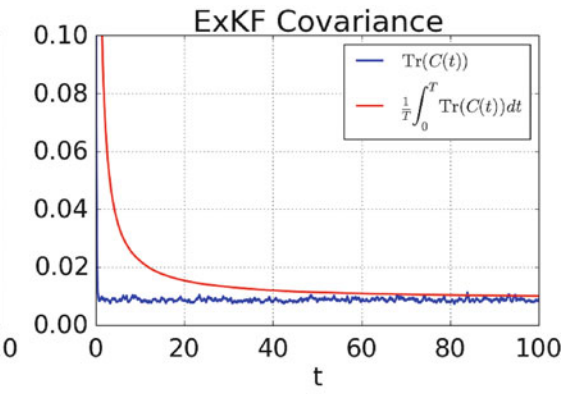
Fig. 8.3: Continuous 3DVAR filter for double well Example 6.5 with $\gamma = 0.1, \sigma = 0.3$, as given in Section 9.3.2.

Section 4.4.3. The continuous-time 3DVAR filter acts as a control system, forcing the filter toward the data. In the paper [21], the data for the Navier–Stokes equation comprised low-frequency Fourier information, and this control perspective was generalized in [8] to cover the technically demanding case of data based on pointwise observations. The large-time behavior of the EnKF is studied, in both discrete and continuous time, in [82]. Finally, we note that as discussed in Section 4.6 in the discrete-time setting, the 3DVAR filter may be used to bound the error in the mean of the Kalman filter for linear problems, because of the optimality of the Kalman filter; this latter optimality property follows, as in discrete time, from a Galerkin orthogonality interpretation of the error resulting from taking conditional expectation. The paper [126] implements this idea in the discrete setting.

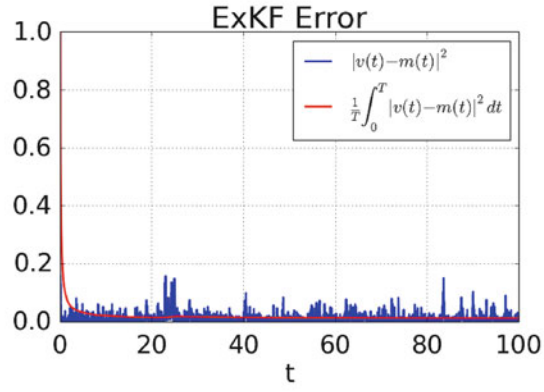
- Section 8.5 concerns various numerical illustrations. Remark 8.7 highlights the fact that implementing filters in a stable fashion, especially for complex models, can be nontrivial, and the reader is cautioned that blind transfer of the programs in the next chapter to other models may well lead to numerical instabilities. These can be caused by an interaction of the numerical integration method with noisy data. See [59] for a discussion of this.



(a) Solution.



(b) Covariance.



(c) Error.

Fig. 8.4: Extended Kalman filter for double-well Example 6.5 with $\gamma = 0.1, \sigma = 0.3$, as given in Section 9.3.3.

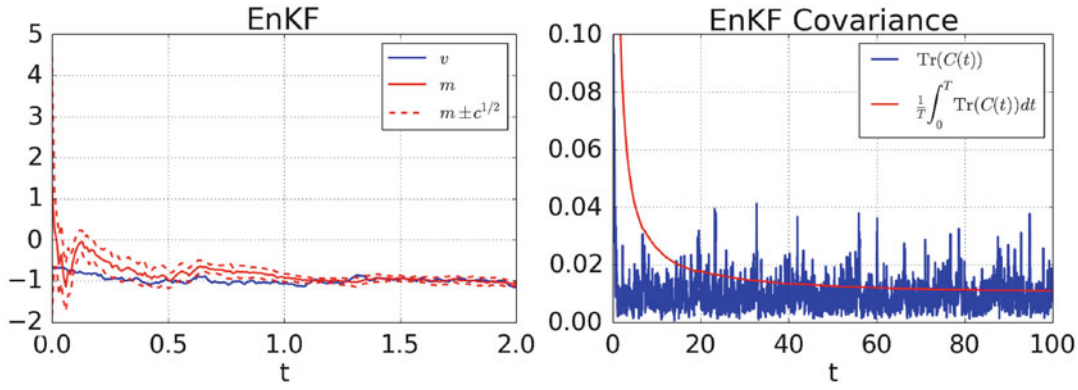
8.7 Exercises

1. Consider the linear example of Section 8.4.1. Implement the Kalman–Bucy filter for this problem by modifying program p8c.m. Verify that the code reproduces the large-time asymptotic behavior of the variance as proved in Section 8.4.1. Carefully distinguish between $\ell < 0, \ell = 0$, and $\ell > 0$. Now extend your code to include the case $\Sigma_0 = \sigma^2 > 0$ and study the large-time behavior of the covariance. What can you prove about the large-time behavior of the covariance in this case?
2. In this exercise, we study the properties of the mean for the one-dimensional linear dynamics example considered in Section 8.4.1, in the large-time asymptotic. More specifically, we study the error between the filter and the truth $v^\dagger$. Assume that the truth satisfies the equation

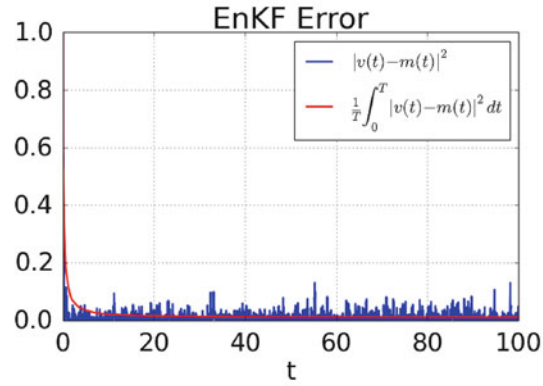
$$\frac{dv^\dagger}{dt} = \ell v^\dagger,$$

while the data z is given by

$$\frac{dz}{dt} = v^\dagger + \gamma \frac{dB}{dt},$$



(a) Solution. (b) Covariance.



(c) Error.

Fig. 8.5: Ensemble Kalman filter for double-well Example 6.5 with $\gamma = 0.1, \sigma = 0.3$, as given in Section 9.3.4.

and B is a realization of the standard unit Brownian motion. Define $e = m - v^\dagger$ and show that

$$\frac{de}{dt} + fe = \gamma^{-1} c \frac{dB}{dt},$$

where $f = F'$ and

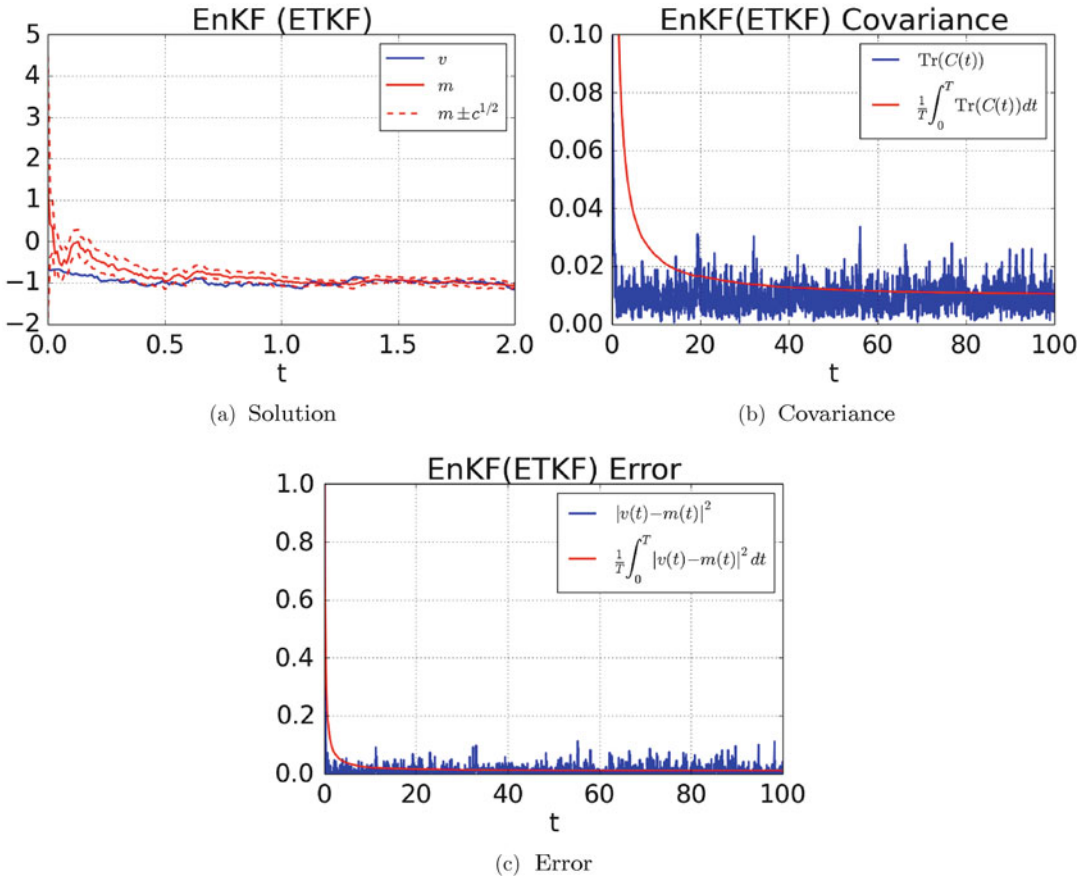
$$F'(t) = \gamma^{-2} c(t) - \ell, \quad F(0) = 0.$$

Apply the Itô formula of Lemma 6.3 to a judiciously chosen function to show that

$$e(t) = \exp\big(-F(t)\big)e(0) + \mathsf{SI}(t), \tag{8.20}$$

where

$$\mathsf{SI}(t) = \int_0^t \gamma^{-1} \exp\big(F(s) - F(t)\big) c(s) dB(s).$$



(a) Solution (b) Covariance

(c) Error

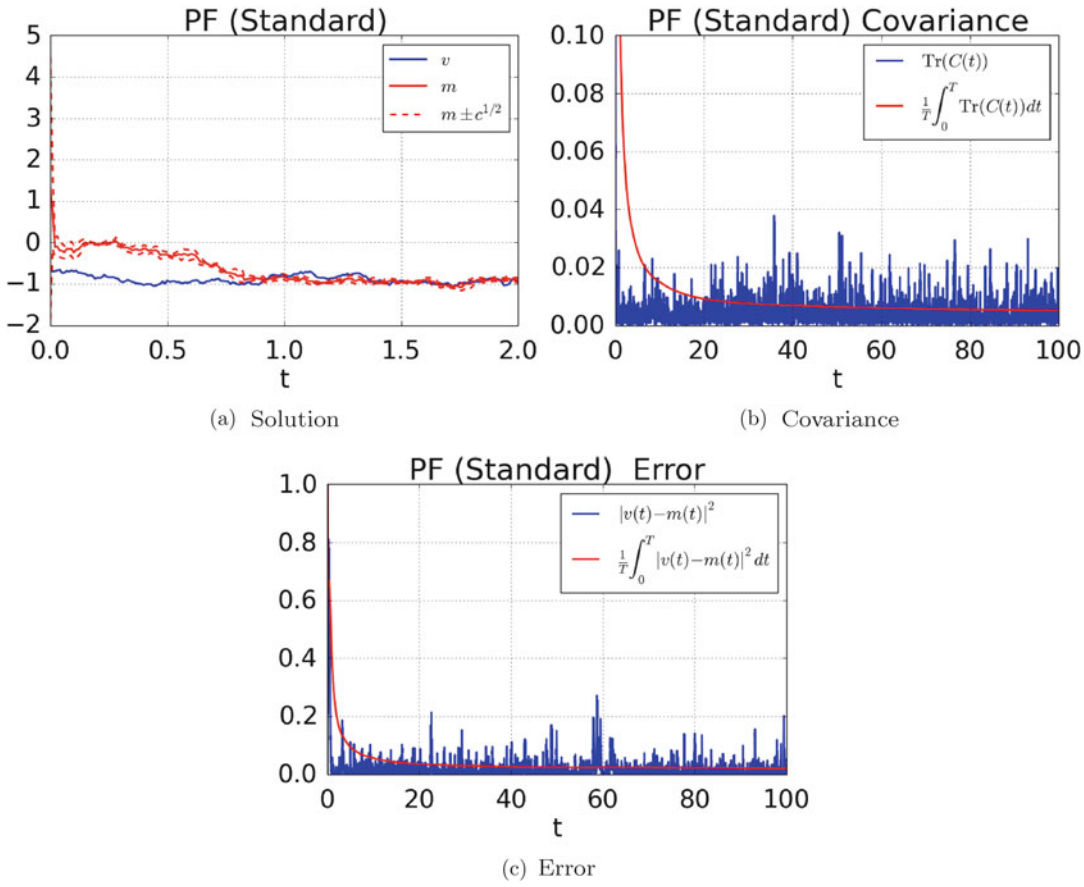
Fig. 8.6: Ensemble transform Kalman filter for double-well Example 6.5 with $\gamma = 0.1, \sigma = 0.3$, as given in Section 9.3.5.

Use the Itô isometry of Lemma 6.1(iii) to show that

$$\mathbb{E}|\mathsf{SI}(t)|^2 = \int_0^t \gamma^{-2} \exp\bigl(2F(s) - 2F(t)\bigr) c^2(s) ds. \tag{8.21}$$

Using the properties of the variance established in Section 8.4.1, show that the asymptotic error in the filter mean is bounded by $\mathcal{O}(\gamma^2)$.

3. Extend the 3DVAR code of program `p10c.m` so that it may be applied to the Lorenz '63 example of Example 6.7. Consider both the fully observed case, in which $H = I$, and the partially observed case, with $H = (1, 0, 0,)^T$. Compare the output of the filter with the truth underlying the data, using different observational noise levels.
4. Extend the ExKF code of program `p11c.m` so that it may be applied to the Lorenz '63 example of Example 6.7. Consider the fully observed case, in which $H = I$. Compare the output of the filter with the truth underlying the data using different observational noise levels.

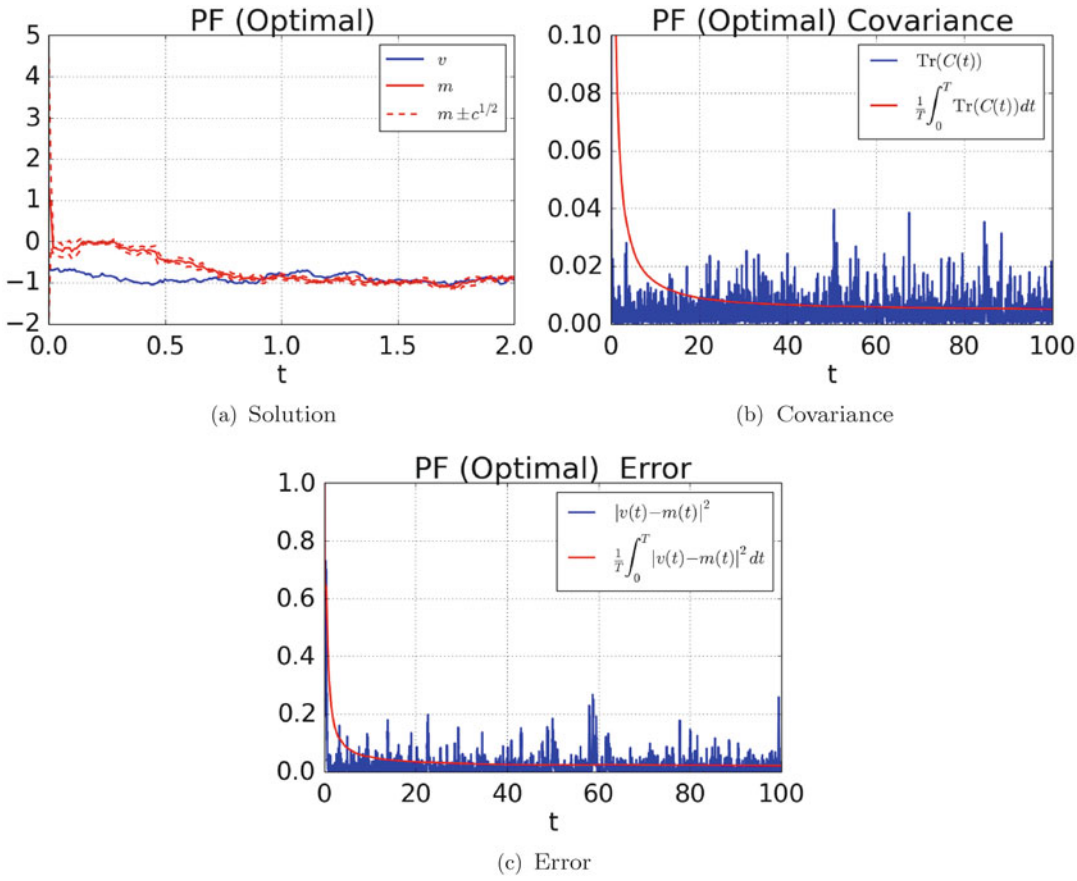


(a) Solution (b) Covariance

(c) Error

Fig. 8.7: Continuous particle filter (standard) for double-well Example 6.5 with $\gamma = 0.1$, $\sigma = 0.3$, as given in Section 9.3.6.

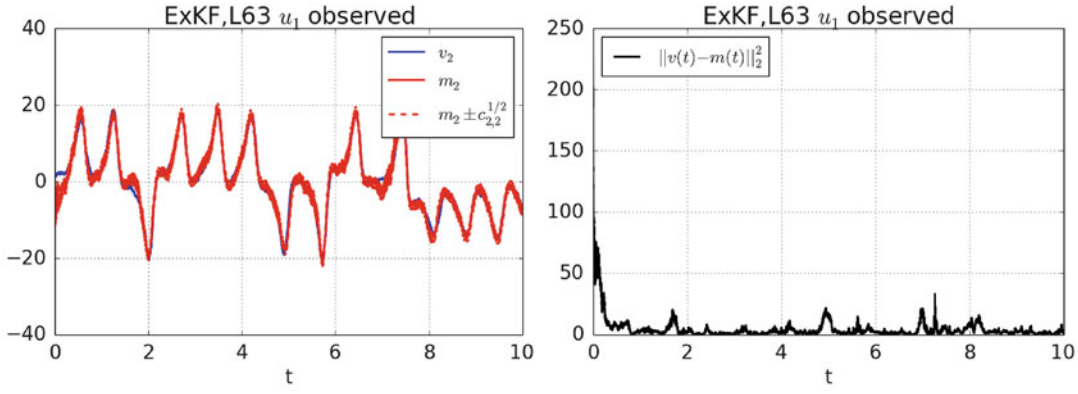
5. Extend the EnKF code of program `p12c.m` so that it may be applied to the Lorenz '96 example of Example 6.8. Consider the case in which two out of every three points are observed. Compare the output of the filter with the truth underlying the data, using different observational noise levels.



(a) Solution (b) Covariance

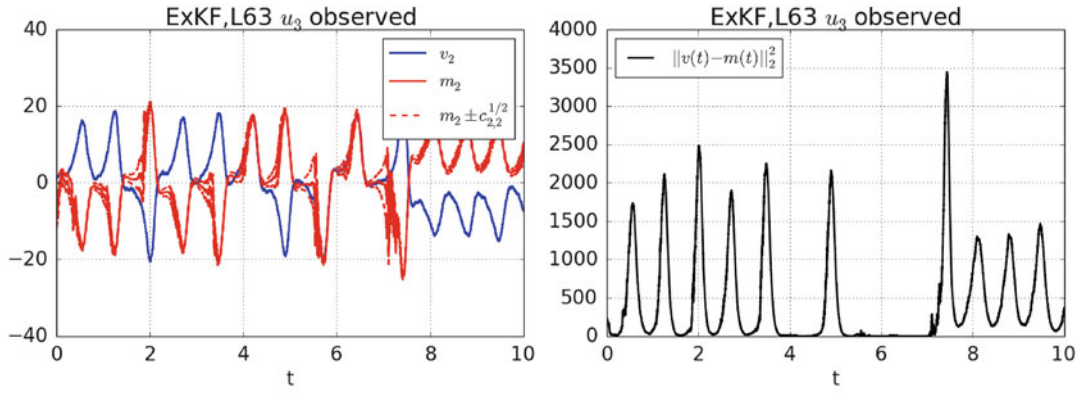
(c) Error

Fig. 8.8: Continuous particle filter (optimal) for double-well Example 6.5 with $\gamma = 0.1, \sigma = 0.3$, as given in Section 9.3.7.



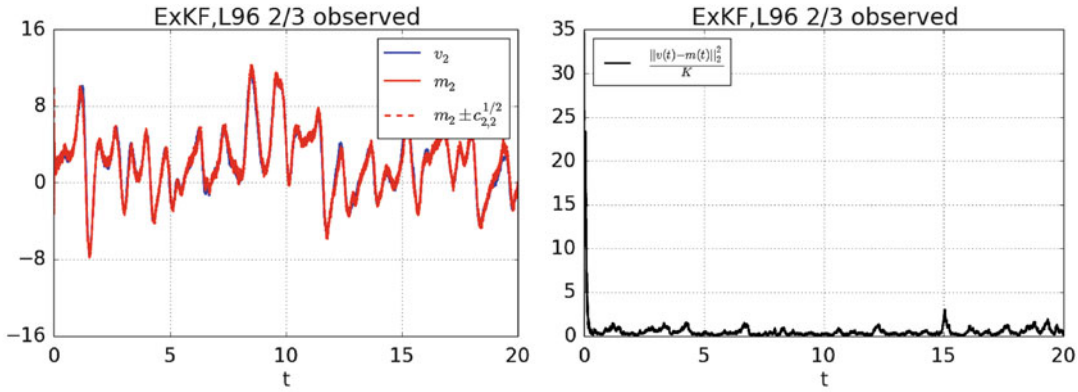
(a) Solution (b) Error

Fig. 8.9: Extended Kalman filter for Lorenz '63 Example 6.7 with observation operator $H = (1, 0, 0)$.



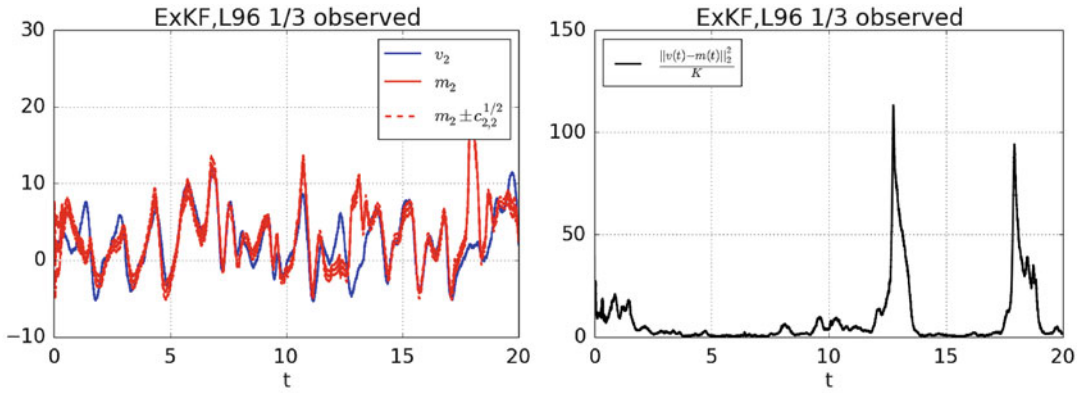
(a) Solution (b) Error

Fig. 8.10: Extended Kalman filter for Lorenz '63 example 6.7 with observation orator $H = (0, 0, 1)$.



(a) Solution (b) Error

Fig. 8.11: Extended Kalman filter for Lorenz '96 example 6.8 with 2/3 of components observed.



(a) Solution (b) Error

Fig. 8.12: Extended Kalman filter for Lorenz '96 example 6.8 with 1/3 of components observed.