Feedback Particle Filter: Application to Continuous State Space Models

Angwenyi David

Institut für Mathematik der Universität Potsdam

October 3, 2019

Contents

Introduction

State space models describing a dynamical system in a continuum can be expressed using the following vector Itô equations.

$$\text{Signal:} \quad dx_t = f(x_t, \theta)dt + G(x_t)d\beta_t; \quad t_0 \leq t, \tag{1a}$$
$$\text{Measurement:} \quad dy_t = h(x_t)dt + d\eta_t, \quad t_0 \leq t, \tag{1b}$$

where

x_t	state vector	$n \times 1$
$f(x_t, \theta)$	drift term	$n \times 1$
θ	vector of parameters	$d \times 1$
$G(x_t)$	matrix function	$n \times m$
$\{\beta_t,\ t > t_0\}$	Brownian motion process	$m \times 1$
y_t	output vector	$r \times 1$
$h(x_t)$	measurement function	$r \times 1$
$\{\eta_t,\ t > t_0\}$	Brownian motion process	$r \times 1$

Statement of the problems

The following problems are of intense interest in many application areas, for example, geosciences, weather forecasting, navigation, financial analysis, *et cetera.*

- **Filtering problem:** What is the current estimate of the state, x_t, given measurements from the past to the present time?
- **Prediction problem:** What is the future estimate of the state, x_τ, given the measurements from the past to the present time?
- **Smoothing problem:** What is the past estimate of the state, x_τ, given the measurements from the past to the present time?
- **Parameter estimation:** What are the estimates of the static parameters, θ, given the measurements from the past to the present time?

Probabilistic/Bayesian framework

In this framework, the probability law for each problem is established by means of conditional probability density functions. Of interest, in this case, are the following:

- The evolution of conditional probability density, $dp(x_t \mid Y_t)$.
- The evolution of the moments—from which we can obtain the mean and the covariances.

These have already been obtained for the system (1).

Probabilistic/Bayesian framework

The system (1) can be rewritten as follows.

$$\begin{aligned}
\text{Signal:} \quad & dx_t/dt = f(x_t, \theta) + G(x_t)v_t; \quad t_0 \leq t, && \text{(2a)} \\
\text{Measurement:} \quad & dy_t/dt = h(x_t) + w_t, \quad t_0 \leq t, && \text{(2b)}
\end{aligned}$$

where

$\{v_t,\ t > t_0\}$	white Gaussian noise process	$m \times 1$
y_t	output vector	$r \times 1$
$\{w_t,\ t > t_0\}$	white Gaussian noise process	$r \times 1$

$v_t \sim \mathcal{N}(0,\ Q(t))$ and $w_t \sim \mathcal{N}(0,\ R(t))$

Theorem 1.1: Evolution of conditional density

For the system (2), the conditional density evolves as follows.

$$dp = \mathcal{L}(p)dt + (h - \hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt)p, \qquad (3)$$

where

$$\mathcal{L}(p) = -\sum_{j=1}^{n} \frac{\partial(pf_j)}{\partial x_i} + \frac{1}{2}\sum_{j,i=1}^{n} \frac{\partial^2[p(GQG^{\mathrm{T}})_{ji}]}{\partial x_j \partial x_i}.$$

Equation (3) is called **Kushner's equation**. In the absence of observations, that is, when $R = 0$, we obtain the so-called **Kolmogorov's forward equation**—which is also known as the **Fokker-Plank equation**

Lemma 1.1: Evolution of moments

For the system (2), the moments, $\psi(x_t \mid Y_t)$, under the conditional density, p, evolve as follows.

$$d\widehat{\psi} = \widehat{\psi_x^{\mathrm{T}} f}\,dt + \frac{1}{2}\mathrm{tr}(GQG^{\mathrm{T}}\psi_{xx})dt + (\widehat{\psi h} - \hat{\psi}\hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt), \tag{4}$$

- From (4), we can obtain the **mean**, μ_t, and the **covariance**, P_t, which correspond to the **first** and **second** moments, respectively.
- The **mean**, μ_t, and the **covariance**, P_t, provide **sufficient statistics** for a Gaussian process.

Theorem 1.2: Evolution of the mean and covariance

For the system (2), the mean **mean**, $\hat{x}_t$, and the **covariance**, P_t, satisfy the following equations.

$$d\hat{x}_t = \hat{f}dt + (\widehat{x_t h^{\mathrm{T}}} - \hat{x}_t \hat{h}^{\mathrm{T}})R^{-1}(t)(dy_t - \hat{h}dt), \tag{5}$$

$$\begin{aligned}(dP_t)_{ij} = {} & (\widehat{x_i f_j} - \hat{x}_i \hat{f}_j)dt + (\widehat{f_i x_j} - \hat{f}_i \hat{x}_j)dt + \\ & (GQG^{\mathrm{T}})_{ij}dt - (\widehat{x_i h} - \hat{x}_i \hat{h})^{\mathrm{T}} R^{-1}(\widehat{h x_j} - \hat{h}\hat{x}_j)dt + \\ & (\widehat{x_i x_j h} - \widehat{x_i x_j}\hat{h} - \hat{x}_i \widehat{x_j h} - \hat{x}_j \widehat{x_i h} + \\ & 2\hat{x}_i \hat{x}_j \hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt),\end{aligned} \tag{6}$$

Mean and Covariance: Continuous-discrete case

Theorem 1.3: Evolution of the mean and covariance

For the system (2) with discrete observations, the mean **mean**, $\hat{x}_t$, and the **covariance**, P_t, satisfy the following equations.

Prediction step

$$d\hat{x}_t = \hat{f}(x_t)dt, \qquad t_n \leq t \leq t_{n+1}, \tag{7}$$

$$\begin{aligned} dP_t = {} & (E\{x_t f^{\mathrm{T}}\} - \hat{x}_t \hat{f}^{\mathrm{T}})dt + (E\{f x_t^{\mathrm{T}}\} - \hat{f}\hat{x}_t^{\mathrm{T}})dt + \\ & E\{GQG^{\mathrm{T}}\}dt, \end{aligned} \tag{8}$$

Update step

$$\hat{x}_{t_n} = \frac{E\{x_{t_n} p(y_{t_n} \mid x_{t_n})\}}{E\{p(y_{t_n} \mid x_{t_n})\}} \tag{9a}$$

$$P_{t_n} = \frac{E\{x_{t_n} x_{t_n}^{\mathrm{T}} p(y_{t_n} \mid x_{t_n})\}}{E\{p(y_{t_n} \mid x_{t_n})\}} - \hat{x}_{t_n}\hat{x}_{t_n}^{\mathrm{T}}. \tag{9b}$$

Example 1: Nonlinear scalar SDE

Consider the following linear Gaussian Itô state space model.

$$dx_t = ax_t dt + bx_t^2 dt; \qquad t_0 \leq t, \tag{10a}$$

$$y_t = x_t + x_t^3 + R^{1/2} w_t; \qquad t_0 \leq t. \tag{10b}$$

where $\{w_t\}$ is a Gaussian white noise process with $\mathbb{E}\{w_t w_t^T\} = 1$. Let the x_t at time, t_0 be $x_{t_0} \sim \mathcal{N}(0.4, 0.001)$. Let, moreover, x_{t_0} and $\{w_t,\ t \geq t_0\}$ be uncorrelated. The analytic solution for (7a) is:

$$x_t = \frac{ax_{t_0} e^{a(t-t_0)}}{a + bx_{t_0}(1 - e^{a(t-t_0)})} \tag{11}$$

- Take $a = -0.2$, $b = 0.2$ and $R = 0.0001$. Estimate x_t.

Figure: A plot of the analytic solution, numerical approximation, filter estimate and the measurements of the model in Example (1)

Linear Continuous Models

The linear version of the system (1) is obtained by replacing $f(x_t, \theta)$ with $F(t, \theta)x_t$, $G(x_t)$ with $G(t)$ and $h(x_t)$ with $H(t)x_t$—which leads to:

$$\begin{aligned} \text{Signal:} \quad & dx_t = F(t,\theta)x_t dt + G(t)d\beta_t; \quad t_0 \le t, && (12a) \\ \text{Measurement:} \quad & dy_t = H(t)x_t dt + d\eta_t; \quad t_0 \le t, && (12b) \end{aligned}$$

in which

x_t	state vector	$n \times 1$
$F(t, \theta)$	continuous time-function matrix	$n \times n$
θ	vector of parameters	$d \times 1$
$G(t)$	continuous time-function matrix	$n \times m$
$\{\beta_t,\ t > t_0\}$	Brownian motion process	$m \times 1$
y_t	output vector	$r \times 1$
$H(t)$	continuous time-function matrix	$r \times n$
$\{\eta_t,\ t > t_0\}$	Brownian motion process	$r \times 1$

Theorem 2.1: Evolution of the mean and covariance

For the system (7), the **mean**, $\hat{x}_t$, and the **covariance**, P_t, satisfy the following equations.

$$d\hat{x}_t = F\hat{x}_t dt + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - H\hat{x}_t dt), \tag{13a}$$

$$dP_t = FP_t dt + P_t F^{\mathrm{T}} dt + GQG^{\mathrm{T}} dt - P_t H^{\mathrm{T}} R^{-1} HP_t dt. \tag{13b}$$

- The system (8) is the minimum variance (optimal) filter for the time-continuous system (7). It is known as **Kalman-Bucy filter**.

Theorem 2.2: Continuous Linear Smoother

For the system (7), the **smoothed estimate**, $\hat{x}_t^s$, and the **smoothed covariance**, P_t^s, satisfy the following equations.

$$d\hat{x}_t^s = F\hat{x}_t^s dt + GQG^{\mathrm{T}} P_t^{-1}(\hat{x}_t^s dt - \hat{x}_t dt), \tag{14a}$$

$$dP_t^s/dt = (F + GQG^{\mathrm{T}} P_t^{-1})P_t^s + P_t^s(F^{\mathrm{T}} + P_t^{-1}GQG^{\mathrm{T}}) - GQG^{\mathrm{T}} \tag{14b}$$

- The system (9) is the **Rauch-Tung-Striebel Smoother**.
- Beginning with the filter estimates of the **mean**, $\hat{x}_t$, and the **covariance**, P_t, at time t the equations (9a-9b) are integrated backwards in time to obtain smoothed estimates of mean, $\hat{x}_t^s$, and covariance, P_t^s.

Ensemble Kalman-Bucy Filter

Suppose, instead of having a single hypothesis of the state, x_t, we formulate say N hypotheses, $X_t := \{x_t^i\}_{i=1}^N$. The EnKB filter for the linear system (7) is of the following form.

- Deterministic: (Bergemann and Reich, 2012; de Wiljes, Reich and Stannart 2016)

$$dx_t^i = Fx_t^i dt + G(t)d\beta_t^i + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - 0.5H(x_t^i + \hat{X}_t)dt). \tag{15}$$

- Stochastic: (Law, Stuart and Zygalakis, 2015; Reich, 2011)

$$dx_t^i = Fx_t^i dt + G(t)d\beta_t^i + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - Hx_t^i dt + dw_t). \tag{16}$$

The **mean**, $\hat{x}_t$, and the **covariance**, P_t, at time t are obtained empirically as shown.

$$\hat{x}_t = \frac{1}{N}\sum_{i=1}^{N} x_t^i, \qquad P_t = \frac{1}{N-1}\sum_{i=1}^{N}(x_t^i - \hat{X}_t)(x_t^i - \hat{X}_t)^{\mathrm{T}}. \tag{17}$$

Much like in the deterministic variant of the EnKB filter, the FPF—for the nonlinear system (1)— has a batch of hypotheses of the state, $X_t := \{x_t^i\}_{i=1}^N$, which are propagated using the **Statonovich SDE** below.

$$dx_t^i = f(x_t^i, \theta)dt + G(x_t^i)d\beta_t^i + K(x_t^i)(dy_t - 0.5H(x_t^i + \hat{X}_t)dt). \quad (18)$$

where K, for the linear SDE, is the **Kalman gain**.

Example 1: Scalar SDE

Consider the following linear Gaussian Itô state space model.

$$dx_t = (ax_t + b)dt + Q^{1/2}dv_t; \qquad t_0 \leq t, \tag{19a}$$

$$dy_t = cx_t dt + R^{1/2}dw_t; \qquad t_0 \leq t. \tag{19b}$$

where $\{v_t\}$ and $\{w_t\}$ are Brownian motion processes with, respectively, $\mathbb{E}\{dv_t dv_t^T\} = dt$ and $\mathbb{E}\{dw_t dw_t^T\} = dt$. Let the x_t at time, t_0 be $x_{t_0} \sim \mathcal{N}(0, 0.001)$. Let, moreover, x_{t_0}, $\{v_t,\ t \geq t_0\}$ and $\{v_t,\ t \geq t_0\}$ be uncorrelated.

- Take $a = -0.2$, $b = 0.2$, $c = 1.01$, $Q = 0.001$, $R = 0.0001$. Estimate x_t.
- Pretend that you do not know a and b. Estimate a and b.

Figure: Filtering, Prediction and smoothing estimates for example 1.

Figure: Top: the boxplot for the estimates of parameters a and b using combined EnKB filter and the FPF. Bottom: left—estimates for a and, right—estimates for b over time.

Conclusion

- The (linear and non-linear) filtering has been introduced.
- An overview of the Feedback Particle filter has been presented.
- An application to a scalar SDE has been made both for state and parameter estimation.
- The FPF gives—in the scalar SDE considered—a converging estimate for the parameters.

References

1. Amirhossein T., de Wiljes J., Prashant G., Reich S., (2016) *Kalman Filter and its Modern Extensions for the Continuous-time Nonlinear Filtering Problem*, arXiv:1702.07241v2[math.OC].
2. Bergemann., Reich S., (2012), *An Ensemble Kalman-Bucy Filter for Continuous Data Assimilation*, Meteorolog. Zeitschrift, vol. 21, pp. 213-219.
3. Jazwinski A., (2007), Stochastic Processes and Filtering Theory, Academic Press, Inc., New York, Dover edition.
4. Kalman R., Bucy R., (1961), *New Result in Linear Filtering and Prediction Theory*, Journal of Basic Engineering.
5. Reich S. and Cotter C., (2015), Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press.
6. Wiljes J., Reich S., Stannat W., (2016), *Long-time Stability and Accuracy of the Ensemble Kalman-Bucy Filter for Fully Observed Processes and Small Measurement Noise*, arXiv:1612.06065v1.