Simultaneous State and Parameter Estimation

Angwenyi David

(The University Of Choice)

Department of Mathematics
Data Science and Computational Intelligence

5th Strathmore International Mathematics Conference

August 12th – August 16th, 2019

Contents

Motivation

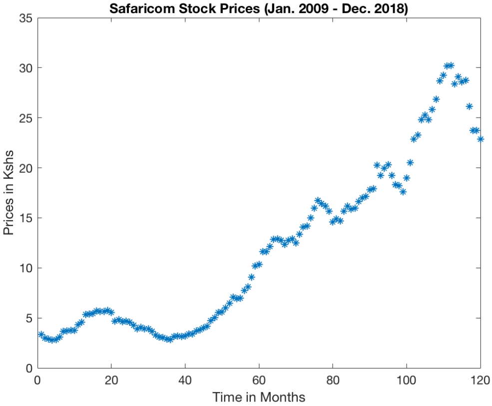


Figure: Safaricom stock prices. Source: https://www.safaricom.co.ke/investor-relation/stocks/share-price/share-price-performance

Problem: How can we predict future prices?

Geometric Brownian motion is given by

$$dx_t = \mu x_t dt + \sigma x_t d\beta_t, \quad t \geq 0, \tag{1a}$$

in which case μ is the interest rate of an asset and σ is its volatility. The Itô equation has an analytical solution,

$$x_t = x_{t_0} \exp\left((\mu - \sigma^2/2)t + \sigma\beta_t\right), \tag{1b}$$

where $\beta(t)$ is Brownian motion process. The Stratonovich equation

$$dx_t = \mu x_t dt + \sigma x_t \circ d\beta_t, \quad t \geq 0, \tag{1c}$$

has an equivalent Itô representation

$$dx_t = \left(\mu + \frac{1}{2}\sigma^2\right) x_t dt + \sigma x_t d\beta_t, \tag{1d}$$

and its analytical solution is

$$x_t = x_{t_0} \exp\left(\mu t + \sigma\beta_t\right). \tag{1e}$$

With data in hand, the problem reduces to estimating two parameters: the interest rate, μ, and the volatility, σ.

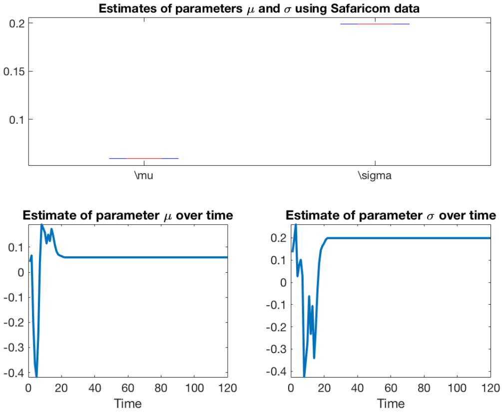


Figure: Parameter estimates are: $\mu = 0.06$ and $\sigma = 0.18$.

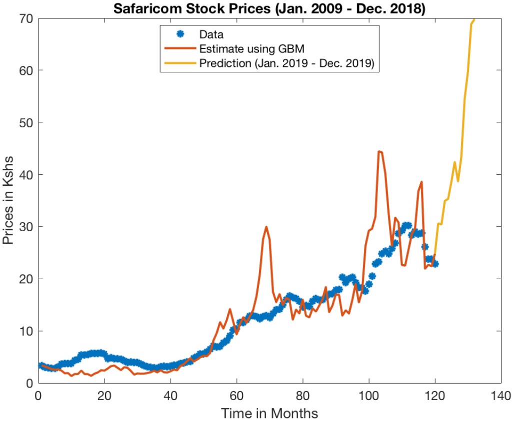


Figure: Prediction for 1 year ahead.

Introduction

- The question of state—and parameter—estimation is, in many fields, very important and has a long history [Lewis et al. (2006)].
- Data assimilation is concerned with the incorporation of measurements of the state—which, mostly, come with measurement errors—into a model to improve the model's approximation of a given state [Reich and Cotter (2015)].
- This can be accomplished by a number of ways: least squares approach, statistical approach and by Bayesian inference. In this study, we take up the Bayesian inference approach to data assimilation.
- Some of the areas which employ data assimilation include surveillance—detection, identification and tracking of, say, astrological objects; navigation which comprises dead reckoning, radio and celestial navigation; numerical weather prediction; robotics and in other control systems [Mohinder and Angus (2001)].

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

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

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	$r \times 1$

Definition 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 $Y_t := y_{[t_0,t]}$? Solution: $\pi_t(x \mid Y_t)$.
- **Prediction problem:** What is the future estimate of the state, x_t, given the measurements from the past to the present time $Y_\tau := y_{[t_0,\tau<t]}$? Solution: $\pi_t(x \mid Y_\tau)$, $\tau < t$.
- **Smoothing problem:** What is the past estimate of the state, x_t, given the measurements from the past to the present time $Y_\tau := y_{[t_0,\tau>t]}$? Solution: $\pi_t(x \mid Y_\tau)$, $\tau > t$.
- **Parameter estimation:** What are the estimates of the static parameters, θ, given the measurements from the past to the present time? Solution: $\pi_t(\theta \mid Y_t)$.

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, $\pi_t(x \mid Y_t)$.
- The evolution of the moments—from which we can obtain the mean and the covariances.

These have already been obtained for the system (2).

Probabilistic/Bayesian framework

The system (2) can be rewritten as follows.

$$\begin{aligned} \text{Signal:} \quad & dx_t/dt = f(x_t, \theta) + G(x_t)v_t; \quad t_0 \le t, && \text{(3a)} \\ \text{Measurement:} \quad & dy_t/dt = h(x_t) + w_t, \quad t_0 \le t, && \text{(3b)} \end{aligned}$$

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

Conditional density

Theorem 2.1: Evolution of conditional density

For the system (3), the conditional density evolves as follows.

$$d\pi_t = \mathcal{L}(\pi_t)dt + (h - \hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt)\pi_t, \qquad (4)$$

where

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

Equation (4) is called **Kushner-Stratonovich equation**. In the absence of observations, that is, when $R \equiv 0$, we obtain the so-called **Kolmogorov's forward equation**—which is also known as the **Fokker-Plank equation**

Lemma 2.1: Evolution of moments

For the system (3), the moments, $\psi(x_t \mid Y_t)$, under the conditional density, π_t, evolve as follows.

$$d\widehat{\psi} = \widehat{\psi_x^{\mathrm{T}} f}\, dt + \frac{1}{2}\mathrm{tr}(GQG^{\mathrm{T}}\psi_{xx})dt + (\widehat{\psi h} - \hat{\psi}\hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt), \tag{5}$$

- From (5), 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 2.2: Evolution of the mean and covariance

For the system (3), the mean **mean**, $\hat{x}_t$, and the **covariance**, P_t, satisfy the following equations.

$$d\hat{x}_t = \hat{f}\,dt + (\widehat{x_t h^{\mathrm{T}}} - \hat{x}_t \hat{h}^{\mathrm{T}}) R^{-1}(t)(dy_t - \hat{h}dt), \tag{6}$$

$$\begin{aligned}(dP_t)_{ij} = (\widehat{x_i f_j} - \hat{x}_i \hat{f}_j)dt + (\widehat{f_i x_j} - \hat{f}_i \hat{x}_j)dt+ \\ (GQG^{\mathrm{T}})_{ij}dt - (\widehat{x_i h} - \hat{x}_i \hat{h})^{\mathrm{T}} R^{-1}(\widehat{h x_j} - \hat{h}\hat{x}_j)dt+ \\ (\widehat{x_i x_j h} - \widehat{x_i x_j}\hat{h} - \hat{x}_i \widehat{x_j h} - \hat{x}_j \widehat{x_i h}+ \\ 2\hat{x}_i \hat{x}_j \hat{h})^{\mathrm{T}} R^{-1}(t)(dy_t - \hat{h}dt),\end{aligned} \tag{7}$$

Expansion of the functions in equations (6) and (7) by Taylor series yield approximate filters.

Theorem 2.3: Second-order Approximate Filter

$$
\begin{aligned}
d\hat{x}_t = {} & f(\hat{x}_t)dt + \frac{1}{2} \triangle [f](\hat{x}_t) : P_t dt \\
& + P_t \boldsymbol{\nabla}[h]^{\mathrm{T}}(\hat{x}_t) R^{-1}(t) \\
& (dy_t - (h(\hat{x}_t) + \frac{1}{2} \triangle [h](\hat{x}_t) : P_t) dt),
\end{aligned} \tag{8a}
$$

$$
\begin{aligned}
dP_t = {} & P_t \boldsymbol{\nabla}[f]^{\mathrm{T}}(\hat{x}_t)dt + \boldsymbol{\nabla}[f](\hat{x}_t) P_t dt \\
& + g(\hat{x}_t) g^{\mathrm{T}}(\hat{x}_t) dt + \boldsymbol{\nabla}[g(\hat{x}_t)] \boldsymbol{\nabla}[g(\hat{x}_t)]^{\mathrm{T}} P_t dt \\
& + (\triangle[g](\hat{x}_t) : P_t) g^{\mathrm{T}}(\hat{x}_t) dt - P_t \boldsymbol{\nabla}[h]^{\mathrm{T}}(\hat{x}_t) R^{-1} \\
& \times \boldsymbol{\nabla}[h](\hat{x}_t) P_t dt + \frac{1}{2} P_t : \triangle[h]^{\mathrm{T}}(\hat{x}_t) R^{-1}(t) \\
& \left(dy_t - (h(\hat{x}_t) + \frac{1}{2} \triangle [h](\hat{x}_t) : P_t) dt \right) P_t.
\end{aligned} \tag{8b}
$$

Theorem 2.4: First-order Approximate Filter

The equations

$$d\hat{x}_t = f(\hat{x}_t)dt + P_t \nabla[h](\hat{x}_t) R^{-1}(t)(dy_t - h(\hat{x}_t)dt), \tag{9a}$$

$$\begin{aligned} dP_t = & P_t \nabla[f]^{\mathrm{T}}(\hat{x}_t)dt + \nabla[f](\hat{x}_t)P_t dt + g(\hat{x}_t)g^{\mathrm{T}}(\hat{x}_t)dt \\ & + \nabla[g(\hat{x}_t)]\nabla[g(\hat{x}_t)]^{\mathrm{T}} P_t dt \\ & - P_t \nabla[h]^{\mathrm{T}}(\hat{x}_t) R^{-1} \nabla[h](\hat{x}_t) P_t dt, \end{aligned} \tag{9b}$$

constitute the vector form of the first-order approximate non-linear filter.

Let $\{x^i_{t_0:t_N}, w^i_{t_N}\}_{i=1}^M$ be a particle-weight system which we want to use to characterise the full posterior density

$$\pi_{t_0:t_N}\left(x_{t_0:t_N} \mid \delta y_{t_1:t_N}\right),$$

where $x_{t_0:t_N} = (x^{\mathrm{T}}_{t_0}, x^{\mathrm{T}}_{t_1}, \dots x^{\mathrm{T}}_{t_N})^{\mathrm{T}}$ and $\delta y_{t_0:t_N} = (\delta y^{\mathrm{T}}_{t_0}, \delta y^{\mathrm{T}}_{t_1}, \dots \delta y^{\mathrm{T}}_{t_N})^{\mathrm{T}}$. Suppose we cannot draw samples, $\{x^i_{t_0:t_N}\}_{i=1}^M$, from the target distribution, $\pi_{t_0:t_N}\left(x_{t_0:t_N} \mid \delta y_{t_0:t_N}\right)$, with any convenience. Let $\gamma_{t_0:t_N}\left(x_{t_0:t_N} \mid \delta y_{t_0:t_N}\right)$ be a density—with semblance to the target density—from which we can conveniently draw samples. Furthermore, let the samples, $\{x^i_{t_0:t_N}\}_{i=1}^M$, have weights

$$w^i_{t_N} \propto \frac{\pi_{t_0:t_N}\left(x^i_{t_0:t_N} \mid \delta y_{t_0:t_N}\right)}{\gamma_{t_0:t_N}\left(x^i_{t_0:t_N} \mid \delta y_{t_0:t_N}\right)}; \qquad i = 1, 2, 3, \dots, M. \quad (10)$$

The computation of weights, $w^i_{t_N}$, as shown in (10) requires the entire posterior, $\pi_{t_0:t_N}(x^i_{t_0:t_N} \mid \delta y_{t_0:t_N})$, and a whole set of measurements, $\delta y_{t_0:t_N}$. To necessitate online (that is to say, at the receipt of each measurement) computation of weights, we take up a sequential approach in which, firstly, we so assemble the proposal that it factorizes as follows

$$\begin{aligned}\gamma_{t_0:t_N}\left(x^i_{t_0:t_N} \mid \delta y_{t_0:t_N}\right) &= \gamma_{t_N}\left(x^i_{t_N} \mid x^i_{t_{N-1}}, \delta y_{t_N}\right) \\ &\qquad \gamma_{t_0:t_{N-1}}\left(x^i_{t_0:t_{N-1}} \mid \delta y_{t_0:t_{N-1}}\right), \\ &= \gamma_{t_0}\left(x^i_{t_0} \mid \delta y_{t_0}\right) \prod_{n=1}^{N} \gamma_{t_n}\left(x^i_{t_n} \mid x^i_{n-1}, \delta y_{t_n}\right).\end{aligned} \tag{11}$$

Notice that the weights can now be expressed as follows

$$\begin{aligned} w_{t_N}^i &\propto \frac{\pi_{t_0:t_N}\left(x_{t_0:t_N}^i \mid \delta y_{t_0:t_N}\right)}{\gamma_{t_0:t_N}\left(x_{t_0:t_N}^i \mid \delta y_{t_0:t_N}\right)}; \qquad i = 1,\, 2,\, 3,\, \ldots,\, M \\ &= w_{t_0}^i \prod_{n=1}^{N} w_{t_n}^i, \end{aligned} \tag{12a}$$

where

$$w_{t_n}^i \propto \frac{\pi_{t_n}\left(\delta y_{t_n} \mid x_{t_n}^i\right) \pi_{t_n}\left(x_{t_n}^i \mid x_{t_{n-1}}^i\right)}{\gamma_{t_n}\left(x_{t_n}^i \mid x_{t_{n-1}}^i,\, \delta y_{t_n}\right)} w_{t_{n-1}}^i. \tag{12b}$$

Interestingly, if we choose a proposal density

$$\gamma_{t_n}\left(x_{t_n}^i \mid x_{t_{n-1}}^i,\, \delta y_{t_n}\right) = \pi_{t_n}\left(x_{t_n}^i \mid x_{t_{n-1}}^i\right),$$

we obtain

$$w_{t_n}^i = \pi_{t_n}\left(\delta y_{t_n} \mid x_{t_n}^i\right) w_{t_{n-1}}^i, \tag{13}$$

which forms the weights of the most basic particle filter known as the *Bootstrap particle filter*.

Algorithm 1 Bootstrap particle filter

Require: δt, α, M, N, π_{t_0}, π_{t_n} and x_{t_0}.

Ensure: $\{x_{t_n}^i\}_{n,i=1}^{N,M}$.

Draw $x_{t_0}^i \sim \pi_{t_0}\left(x_{t_0}^i \mid x_{t_0}\right)$

Compute initial weights $w_{t_0}^i = \dfrac{1}{M}$

for $n = 1$ **to** N, $\delta t > 0$ **do**

  Draw $x_{t_n}^i \sim \pi_{t_n}\left(x_{t_n}^i \mid x_{t_{n-1}}^i\right)$

  Compute weights $w_{t_n}^i \sim \pi_{t_n}\left(\delta y_{t_n} \mid x_{t_n}^i\right) w_{t_{n-1}}^i$

  Normalise the weights to obtain $\{\tilde{w}_{t_n}^i\}_{i=1}^{M}$

  Compute $\mathrm{ESS}_{t_n} = \dfrac{1}{\sum_{i=1}^{M}(\tilde{w}_{t_n}^i)^2}$

  if $\mathrm{ESS}_{t_n} \leq \alpha$ **then**

    Resample the particles

    Set the weights to $w_{t_n}^i = \dfrac{1}{M}$

  end if

end for

Example 1: Lorentz 63 model

Stochastic Lorentz 63 model is given by,

$$dx_t = f(x_t)dt + G^{1/2}d\beta_t; \qquad t \in [t_0, T], \tag{14}$$

where x is a 3-dimensional column vector $[x_1, x_2, x_3]^{\mathrm{T}}$, $G = \alpha \mathrm{I}_{3\times 3}$, α is a constant, $\mathrm{I}_{3\times 3}$ is a 3-dimensional identity matrix,

$$f(x) = \begin{pmatrix} a(x_2 - x_1) \\ bx_1 - x_1x_3 - x_2 \\ x_1x_2 - cx_3 \end{pmatrix},$$

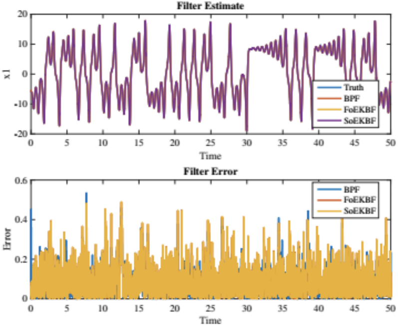
with $a = 10$, $b = 8/3$ and $c = 28$. $\{\beta_t, t \geq t_0\}$ is a 3-dimensional standard Brownian motion process. We use the following initial values $x_{t_0} = [-5.91652, -5.52332, 24.5723]^{\mathrm{T}}$.

For the purpose of this study, we introduce synthetic measurements

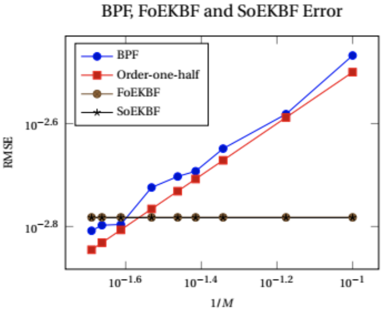
$$dy_t = h(x_t)dt + R^{1/2}d\eta_t, \tag{15}$$

where $\{\eta_t, t \geq t_0\}$ is a 3-dimensional standard Brownian motion process, $R = \sigma \mathrm{I}_{3\times 3}$, σ is a constant,

$$h(x) = \begin{pmatrix} a(x_2 - x_1) \\ bx_1 - x_1x_3 - x_2 \\ x_1x_2 - cx_3 \end{pmatrix}.$$



(a)



(b)

Figure: Filter estimates. (a) $M = 1000$, $\delta t = 0.01$, $\alpha = 0.20$, and $\sigma = 0.17$. Multinomial resampling is done when ESS$\leq 0.7M$ and the filter is ran for 5000 iterations. (b) Plot of RMSEs for $M = 10,\ 15,\ 22,\ 26,\ 29,\ 34,\ 46,\ 49$ against the reciprocal of ensemble sizes for 15000 iterations. RMSEs for FoEKBF$= 0.0017$, SoEKBF $= 0.0016$.

Critique of nonlinear filters

Particle degeneracy: loss of weights of the particles.

- Resampling -Hampers parallel computation
- Localization - To do away with spurious correlations in spatio-temporal models.
- Sampling new particles from the likelihood
- Sampling from a good proposal -It is difficult to construct a good proposal
- Moving the particles to desired position in the state space -This is the basis of exact filters: feedback particle filters, Benes filter, etc.

Linear Continuous Models

The linear version of the system (2) 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:

$$\text{Signal:} \quad dx_t = F(t, \theta)x_t dt + G(t)d\beta_t; \quad t_0 \leq t, \tag{16a}$$
$$\text{Measurement:} \quad dy_t = H(t)x_t dt + d\eta_t; \quad t_0 \leq t, \tag{16b}$$

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 3.1: Evolution of the mean and covariance

For the system (16), the **mean**, $\hat{x}_t$, and the **covariance**, P_t, satisfy the following equations.

$$d\hat{x}_t = F\hat{x}_t dt + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - H\hat{x}_t dt), \tag{17a}$$

$$dP_t = FP_t dt + P_t F^{\mathrm{T}} dt + GQG^{\mathrm{T}} dt - P_t H^{\mathrm{T}} R^{-1} HP_t dt. \tag{17b}$$

- The system (17) is the minimum variance (optimal) filter for the time-continuous system (16). It is known as **Kalman-Bucy filter**.

Theorem 3.2: Continuous Linear Smoother

For the system (16), the **smoothed estimate**, $\hat{x}_t^s$, and the **smoothed covariance**, P_t^s, satisfy the following equations.

$$d\hat{x}_t^s = F\hat{x}_t^s dt + GQG^{\mathrm{T}} P_t^{-1}(\hat{x}_t^s dt - \hat{x}_t dt), \tag{18a}$$

$$dP_t^s/dt = (F + GQG^{\mathrm{T}} P_t^{-1})P_t^s + P_t^s(F^{\mathrm{T}} + P_t^{-1} GQG^{\mathrm{T}}) - GQG^{\mathrm{T}} \tag{18b}$$

- The system (18) 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 (18a-18b) 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 (16) is of the following form.

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

$$dx_t^i = Fx_t^i dt + G(t)d\beta_t^i + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - 0.5H(x_t^i + \hat{X}_t)dt). \tag{19}$$

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

$$dx_t^i = Fx_t^i dt + G(t)d\beta_t^i + P_t H^{\mathrm{T}} R^{-1}(t)(dy_t - Hx_t^i dt + dw_t). \tag{20}$$

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

$$\hat{x}_t = \frac{1}{N}\sum_{i=1}^N x_t^i, \qquad P_t = \frac{1}{N-1}\sum_{i=1}^N (x_t^i - \hat{X}_t)(x_t^i - \hat{X}_t)^{\mathrm{T}}. \tag{21}$$

Example 1: Scalar SDE

Consider the following linear Gaussian Itô state space model.

$$dx_t = (0.2 - 0.2x_t)dt + Q^{1/2}dv_t; \qquad t_0 \leq t, \tag{22a}$$

$$dy_t = 1.01x_t dt + R^{1/2}dw_t; \qquad t_0 \leq t, \tag{22b}$$

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. Set $Q = 0.001$ and $R = 0.0001$. We seek an estimate of x_t.

The following panels show the results.

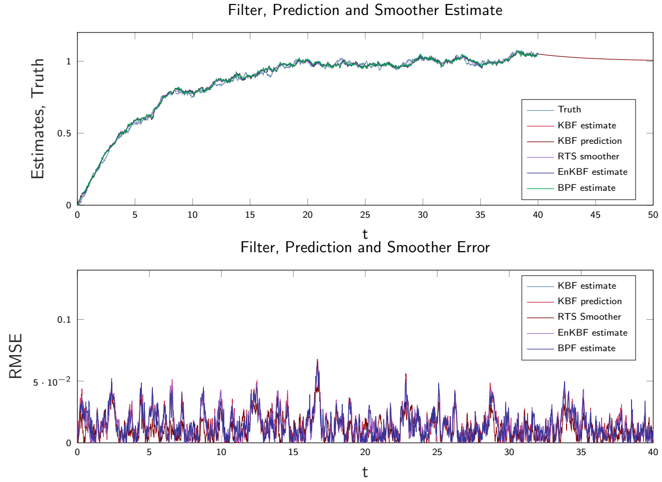


Figure: Filtering, prediction and smoothing estimates for. The ensemble size used in both the EnKBF and the BPF is $M = 1000$. Prediction is shown beyond $T = 40$. A uniform time step-size of $\delta t = 0.02$ is used. The prior used is $\pi_{t_0} = \mathcal{N}(0, 0.001)$.

	KBF pred.	**KBF est.**	**RTS est.**	**EnKBF est.**	**BPF est.**
Time-averaged RMSE	0.0129	0.0127	0.0104	0.0127	0.0131

Table: A table showing time-averaged RMSE for the prediction and estimate of KBF, the RTS smoother estimate, EnKBF and BPF estimates shown in after a burn-in of 1000 iterations.

The RMSE for the smoother is obtained by

$$\mathrm{RMSE}_t = \sqrt{(\hat{x}_t^s - x_t)^2}. \tag{23}$$

Feedback Particle filter

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 controlled **Statonovich SDE** below.

$$dx_t^i = f(x_t^i, \theta)dt + G(x_t^i)d\beta_t^i + K(x_t^i)(dy_t - 0.5H(x_t^i + \hat{X}_t)dt). \quad (24)$$

where $K = \nabla\phi$, the gain, where $\phi(x)$ satisfies the following Poisson equation

$$\nabla \cdot (\pi_t(x \mid Y_t)\nabla\phi(x)) = -(h(x_t, t) - \hat{h})\pi_t(x \mid Y_t)R^{-1}(t), \quad (25a)$$

$$\int \phi(x_t)\pi_t(x \mid Y_t)dx = 0, \quad (25b)$$

where $\pi_t(x \mid Y_t)$ is the conditional density of the particle, x_t^i, given measurements $Y_\tau = \{y_\tau : t_0 \le \tau \le t\}$.

Sinkhorn Particle filter

Schrödinger problem: Find two functions $\hat{\phi}_{t_n}(x_{t_n})$ and $\phi_{t_{n+1}}(x_{t_{n+1}})$, satisfying the following equations

$$\pi_{t_n}(x_{t_n} \mid Y_{t_n}) = \pi^{\phi}_{t_n}(x_{t_n} \mid Y_{t_n})\hat{\phi}_{t_n}(x_{t_n}), \tag{26a}$$

$$\pi_{t_{n+1}}(x_{t_{n+1}} \mid Y_{t_{n+1}}) = \pi^{\phi}_{t_{n+1}}(x_{t_{n+1}} \mid Y_{t_{n+1}})\phi_{t_{n+1}}(x_{t_{n+1}}), \tag{26b}$$

$$\pi^{\phi}_{t_{n+1}}(x_{t_{n+1}} \mid Y_{t_{n+1}}) = \int \pi_{t_{n+1}}(x_{t_{n+1}} \mid x_{t_n})\pi^{\phi}_{t_n}(x_{t_n} \mid Y_{t_n})dx_{t_n}, \tag{26c}$$

$$\hat{\phi}_{t_n}(x_{t_n}) = \int \pi_{t_{n+1}}(x_{t_{n+1}} \mid x_{t_n})\phi_{t_{n+1}}(x_{t_{n+1}})dx_{t_{n+1}}, \tag{26d}$$

where $\pi_{t_n}(x_{t_n} \mid Y_{t_n})$ and $\pi_{t_{n+1}}(x_{t_{n+1}} \mid Y_{t_{n+1}})$ are the marginal filtering distributions at time t_n and t_{n+1}, respectively.

Now, suppose that we have $L = kM$, where $k \in \mathbb{N}$, samples $\{x_{t_n}^j\}_{j=1}^L$ drawn from $\pi_{t_n}(x_{t_n} \mid Y_{t_n})$. We obtain a bi-stochastic matrix $Q \in \mathbb{R}^{L\times M}$, which approximates the Markov process defined by $\pi(x_{t_{n+1}} \mid x_{t_n})$. The Schrödinger problem can then be recast as follows: Find two non-negative vectors $u \in \mathbb{R}^L$ and $v \in \mathbb{R}^M$ so that,

$$P^* = \mathrm{diag}(u) Q \mathrm{diag}(v)^{-1}, \tag{27}$$

given that P^* belong to a polytope

$$U := \left\{ P \in \mathbb{R}^{L\times M} : P \geq 0, \sum_{j=1}^{L} p_{ji} = p_1, \ \sum_{i=1}^{M} p_{ji} = p_0 \right\}.$$

P^* is also a solution to the optimization problem defined by minimizing the distance between all possible bi-stochastic matrices P and Q; that is,

$$P^* = \arg\min_{P\in U} \mathrm{KL}(P \| Q), \tag{28}$$

where KL is the Kullback Leibler divergence between $P \in U$ and Q; that is,

$$\mathrm{KL}(P \| Q) := \sum_{j,i=1}^{L,M} p_{ji} \log \frac{p_{ji}}{q_{ji}}, \tag{29}$$

where p_{ji} and q_{ji} are the elements of, respectively, matrices P and Q in row j and column i.

Now the feedback particle filter in the Schrödinger formulation is as follows: we first obtain an M-sized ensemble of states $\{\bar{x}^i_{t_n}\}_{i=1}^M$ using a forecast distribution, $\pi_{t_n}(\bar{x}^i_{t_n} \mid Y_{t_n})$, with respect to which each particle evolves according to the weak form of the Fokker-Planck equation, that is

$$\bar{x}^i_{t_n} = \bar{x}^i_{t_{n-1}} + f(\bar{x}^i_{t_{n-1}})\delta t. \tag{30}$$

Secondly, $L = kM$ ensembles are obtained as follows:

$$x^j_{t_n} = \bar{x}^j_{t_n} + g(t_n) d\beta^j_{t_n}, \tag{31}$$

where $\{\bar{x}^j_{t_n}\}_{j=1}^L$ are obtained by replicating each particle $\bar{x}^i_{t_n}$ k times. The weights are obtained using

$$w^j_{t_n} = \exp\Big(-\frac{1}{2\delta t}(-2\delta y^{\mathrm{T}}_{t_n} h(x^j_{t_n})\delta t + h^{\mathrm{T}}(x^j_{t_n})h(x^j_{t_n})\delta t^2)\Big) w^j_{t_{n-1}}. \tag{32}$$

This gives a particle weight system, $\{x^j_{t_n},\, \tilde{w}^j_{t_n}\}_{j=1}^L$, where $\tilde{w}^j_{t_n}$ signifies the normalised $w^j_{t_n}$—normalisation done according to

$$\tilde{w}^j_{t_n} = \frac{w^j_{t_n}}{\sum_{j=1}^L w^j_{t_n}}.$$

By $Q \in \mathbb{R}^{L\times M}$, we denote and understand a matrix whose elements are

$$q_{ji} = \exp\left(-\frac{1}{2g^2\delta t}\|x^j_{t_n} - \bar{x}^i_{t_n}\|^2\right). \tag{33}$$

Then we solve the Schrödinger problem defined by (26), from whence we obtain P^*. P^* can be obtained, iteratively, by means of the Sinkhorn scheme stipulated in the following algorithm, whose implementation details are stipulated in Cuturi (2013).

Algorithm 2 Sinkhorn iteration

Require: p_0 and p_1.
Ensure: P.
1: Compute Q and set $P^0 = Q$
2: **while** $k > 1$ **do**
3: Compute: $u^{k+1} = \mathrm{diag}(\sum_{m=1}^{M} P^k)^{-1} p_1$.
4: Compute: $v^{k+1} = \mathrm{diag}(\sum_{m=1}^{M} p_0)^{-1} \sum_{l=1}^{L} (\mathrm{diag}(u^{k+1}) P^k)^{\mathrm{T}}$.
5: Compute: $P^{k+1} = \mathrm{diag}(u^{k+1}) P^k \mathrm{diag}(v^{k+1})^{-1}$.
6: **end while**

Filtered particles are eventually obtained thus

$$\tilde{x}_{t_n}^i = \sum_{j=1}^{L} x_{t_n}^j p_{ji}^* + g(\delta t)^{1/2} \xi_{t_n}^i, \qquad \xi_{t_n}^i \sim \mathcal{N}(\bar{0}_{n\times 1}, \mathrm{I}_{n\times n}), \qquad (34)$$

where $\bar{0}_{n\times 1}$ and $\mathrm{I}_{n\times n}$ are, respectively, the null vector and the identity matrix of sizes as indicated in the subscripts. This yields the proposed Sinkhorn particle filter (SPF).

Algorithm 3 Sinkhorn particle filter

Require: $x^i_{t_0}$, $w^i_{t_0} = 1/M \ \forall i \in \{1, 2, ..., M\}$, k, and $\delta y_{[t_0, t_T]}$.
Ensure: $\hat{x}_{[t_0, t_T]}$.

```
for n = 1 to N, δt > 0 do
    for i = 1 to M do
        Obtain x̄^i_{t_n} using (31)
        for j = 1 to L do
            Replicate x̄^i_{t_n} k times and obtain x^j_{t_n} using (31)
            Compute weights w^j_{t_n} using (33)
            Compute q_{ji} using (34)
        end for
        Calculate P* by solving the the Schrödinger problem via Alg. 2.
        Compute the filtered particles x̃^i_{t_n} using (35)
    end for
    Compute x̂_{t_n} = (1/M) Σ_{i=1}^{M} x̃^i_{t_n}
end for
```

for $n = 1$ **to** N, $\delta t > 0$ **do**
- **for** $i = 1$ **to** M **do**
  - Obtain $\bar{x}^i_{t_n}$ using (31)
  - **for** $j = 1$ **to** L **do**
    - Replicate $\bar{x}^i_{t_n}$ k times and obtain $x^j_{t_n}$ using (31)
    - Compute weights $w^j_{t_n}$ using (33)
    - Compute q_{ji} using (34)
  - **end for**
  - Calculate P^* by solving the the Schrödinger problem via Alg. 2.
  - Compute the filtered particles $\tilde{x}^i_{t_n}$ using (35)
- **end for**
- Compute $\hat{x}_{t_n} = \dfrac{1}{M} \sum_{i=1}^{M} \tilde{x}^i_{t_n}$

end for

Alternatively to obtaining the particles by means of (35), we can resample $\{x_{t_n}^j\}_{j=1}^L$ such that

$$\mathbb{P}(x_{t_n}^i = x_{t_n}^j) = p_{ji}^*, \qquad \forall i = 1, 2, ..., M. \tag{35}$$

This results in the resampling Sinkhorn Particle filter (RSPF). Notice that all these formulations in this section are done on a discrete setting. Time-continuous settings are arrived at by passing to the formal limit as $\delta t \to 0$. We now consider an example to test the performance of the SPF and RSPF in comparison to the BPF.

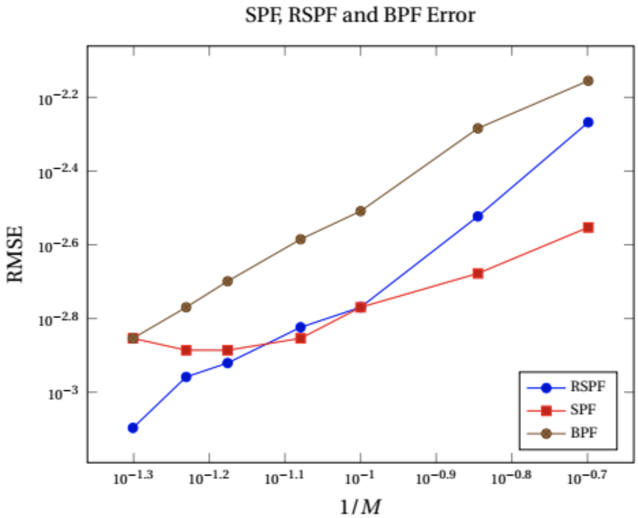


Figure: Plot of RMSE resulting from the estimates of Lorentz 63 using optimal filters. The settings are:– $dt = 0.01$, $P_{t_0} = \mathrm{I}_{3\times3}$ and $T = 4,000,000$; $M = 5,\ 7,\ 10,\ 12,\ 15,\ 17,\ 20.$

The state vector is augmented with the vector of parameters to form an extended state-space and then the filter is run forward in time for an update of both the state and the parameters. The parameters are induced with artificial dynamics, or are made to assume a random walk; that is, respectively,

$$dz_t = \zeta_t; \quad t_0 \leq t, \tag{36}$$

where

$$dz_t = \begin{pmatrix} dx_t \\ d\theta_t \end{pmatrix} \text{ and } \zeta_t = \begin{pmatrix} f(x_t)dt + g(x_t)d\beta_t \\ 0 \end{pmatrix}, \tag{37a}$$

or where

$$\zeta_t = \begin{pmatrix} f(x_t)dt + g(x_t)d\beta_t \\ \sigma d\chi_t \end{pmatrix}, \tag{37b}$$

in which $\{\chi_t,\ t > t_0\}$ is a $d-$dimensional standard Brownian motion vector process and σ is a small constant. A filter is then implemented with the augmented state z_t in the place of x_t.

Example 1: Scalar SDE

Consider the following linear Gaussian Itô state space model.

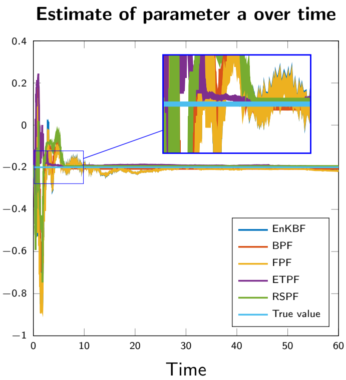
$$dx_t = (ax_t + b)dt + Q^{1/2}dv_t; \qquad t_0 \leq t, \tag{38a}$$

$$dy_t = cx_t dt + R^{1/2}dw_t; \qquad t_0 \leq t, \tag{38b}$$

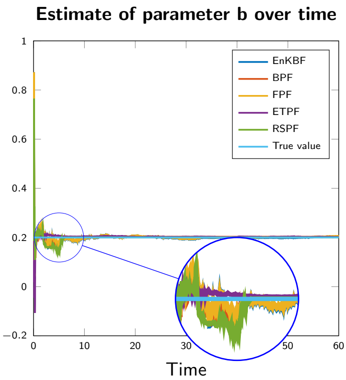
where $\{v_t\}$ and $\{w_t\}$ are standard Brownian motion processes with, respectively,

$$\mathbb{E}\{dv_t dv_t^T\} = dt \text{ and } \mathbb{E}\{dw_t dw_t^T\} = dt.$$

Let the state, 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. We take $a = -0.2$, $b = 0.2$, $c = 1.01$, $Q = 0.001$, $R = 0.0001$ and proceed to simultaneously estimate the state, x_t, and the parameters, a and b using different filters.

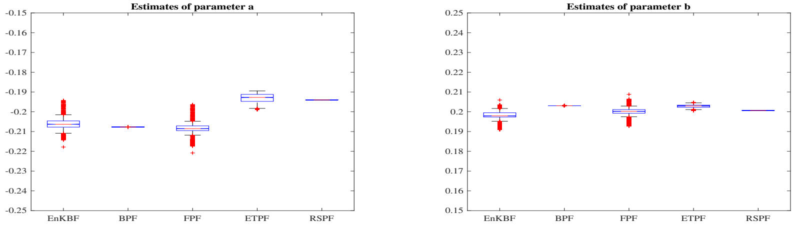


(a) Parameter a



(b) Parameter b

Figure: Plots showing:- (a) estimates of parameter a and (b) estimates of parameter b over time using EnKBF, BPF, FPF, ETPF and RSPF. The true parameter values are, respectively, $a = -0.2$ and $b = 0.2$. The plots indicate that all the filters converge to the true parameter values. The time step used is $\delta t = 0.02$.



(a) Parameter *a*

(b) Parameter *b*

Figure: Box-plots (a) and (b) showing, respectively, the distribution of estimates of parameters *a* and *b* obtained using EnKBF, BPF, FPF, ETPF and RSPF. The gain in the FPF is computed using the kernel based gain approximation method. In both cases, the EnKBF and FPF register more dispersive results with more outliers than the BPF, ETPF and RSPF.

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 RSPF 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.

Thank you.