Mathematics and numerics for data assimilation and state estimation – Lecture 20

Summer semester 2020

Overview

1. The Fokker-Planck equation
2. Numerical integration of SDE
3. Filtering problems with SDE dynamics
4. Examples using Euler–Maruyama integration
5. Model error and model fitting

Summary lecture 20

- Itô integrals and theory of stochastic differential equations (SDE)

$$V_t = V_0 + \int_0^t b(V_s)ds + \int_0^t \sigma(V_s)dW_s$$

- Plan for today: Fokker-Planck equation, numerical integration of SDE and applications in filtering problems.

Overview

1 The Fokker-Planck equation

2 Numerical integration of SDE

3 Filtering problems with SDE dynamics

4 Examples using Euler–Maruyama integration

5 Model error and model fitting

The kernel density for SDE

Our plan is to study filtering problems

$$V_{j+1} = \Psi(V_j) := V_j + \int_0^1 b(V_{j+s})ds + \int_0^1 \sigma(V_{j+s})dW_s^{(j)}$$
$$Y_{j+1} = h(V_{j+1}) + \eta_{j+1}$$

where $W^{(j)}$ are independent Wiener processes.

The Bayes filter for this problem takes the form

$$\pi(v_{j+1}|y_{1:j+1}) \propto \pi(y_{j+1}|v_{j+1}) \int_{\mathbb{R}^d} \pi(v_{j+1}|v_j)\pi(v_j|y_{1:j})dv_j$$

with $\pi_{V_{j+1}|V_j}(x|y)$ equal to the kernel density for $t \in (0,1]$,

$$p(t,x|y) = \frac{\mathbb{P}(V_{j+t} \in dx|V_j \in dy)}{dx} = \frac{\mathbb{P}(V_t \in dx|V_0 \in dy)}{dx}$$

(due to the time-independent coefficients the SDE is stationary).

The density of an SDE

Consider the 1D SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 \sim p(0,x)$$

and assume that the density $p(t,x) = \mathbb{P}(X_t \in dx)/dx$ exists for any $t > 0$.

Recall that for any $f \in C_C^2(\mathbb{R})$ (mapping with compact support),

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 = (f'b + \sigma^2/2f'')dt + f'\sigma dW_t$$

By integration,

$$f(X_t) - f(X_0) = \int_0^t (bf' + \frac{\sigma^2}{2}f'')(X_s)ds + \int_0^t (f'\sigma)(X_s)dW_s.$$

Recalling that Itô integrals have mean-zero,

$$\mathbb{E}\left[f(X_t) - f(X_0)\right] = \int_0^t \mathbb{E}\left[(bf' + \frac{\sigma^2}{2}f'')(X_s)\right] ds$$

Note: expectation is here wrt the density $p(s,x)$

Fokker-Planck equation

$$\int_{\mathbb{R}} f(x)(p(t,x)-p(0,x))dx = \int_0^t \int_{\mathbb{R}} \left[b(x)f'(x)+\sigma^2(x)\frac{f''(x)}{2}\right]p(s,x)dxds$$

Integration by parts, using the compact support of f (and its derivatives), we obtain

$$\int_0^t \int_{\mathbb{R}} f(x)p_t(s,x)dxds$$
$$= \int_0^t \int_{\mathbb{R}} f(x)\Big[-\partial_x\Big(b(s)p(s,x)\Big)+\partial_{xx}\Big(\frac{\sigma^2(x)}{2}p(s,x)\Big)\Big]dxds \quad \forall f \in C_C^2(\mathbb{R})$$

Conclusion: The density $p(t,x) = \mathbb{P}(X_t \in dx)/dx$ must satisfy the **Fokker-Planck** PDE

$$\begin{aligned} p_t &= \partial_x(-bp) + \partial_{xx}(\frac{\sigma^2}{2}p) \quad (t,x) \in [0,T]\times\mathbb{R} \\ p(t,x)|_{t=0} &= p(0,x). \end{aligned} \tag{1}$$

If the SDE coefficients are sufficiently smooth and $\sigma > 0$, then (1) is well-posed and a unique classical solution exists for all $t > 0$.

Fokker-Planck for kernel densities

The PDE extends to kernel densities $p(t,x|y) = \mathbb{P}(X_t \in dx | y \in dy)/dx$:

$$\begin{aligned} p_t(\cdot,\cdot|y) &= \partial_x(-bp(\cdot,\cdot|y)) + \partial_{xx}(\frac{\sigma^2}{2}p(\cdot,\cdot|y)) \quad (t,x) \in [0,T] \times \mathbb{R} \\ p(0,x|y) &= \delta_y(x). \end{aligned} \tag{2}$$

Remarks: The operator

$$(\mathcal{L}^* p)(x) := \partial_x(-bp)(x) + \partial_{xx}(\frac{\sigma^2}{2}p)(x)$$

may be associated to the transition function of Markov chains (here denoted P):

$$p(t+\Delta t, x) \approx p(t,x) + \Delta t(\mathcal{L}^* p)(x),$$

vs

$$\pi_i^{n+1} = \sum_{j=1}^{N} P_{ji}\pi_j^n = \pi_i^n + \Big((P-I)^T\pi^n\Big)_i$$

And just like Markov chains, SDE may have stationary distributions:

$$\mathcal{L}^* p = 0 \iff p \text{ stationary }, \qquad (P-I)^T\pi = 0 \iff \pi \text{ stationary.}$$

Application in filtering

Returning to the filtering problem

$$V_{j+1} = \Psi(V_j) := V_j + \int_0^1 b(V_{j+s})ds + \int_0^1 \sigma(V_{j+s})dW_s^{(j)}$$
$$Y_{j+1} = h(V_{j+1}) + \eta_{j+1}$$

the iterative Bayes filter equation

$$\pi(v_{j+1}|y_{1:j+1}) \propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j})$$

can be written

$$\pi(v_{j+1}|y_{1:j+1}) \propto \pi(y_{j+1}|v_{j+1})p(1, v_{j+1})$$

where p solves

$$p_t = \mathcal{L}^* p \qquad (t,x) \in [0,T] \times \mathbb{R}$$
$$p(t,x)|_{t=0} = \pi_{V_j|Y_{1:j}}(x|y_{1:j})$$

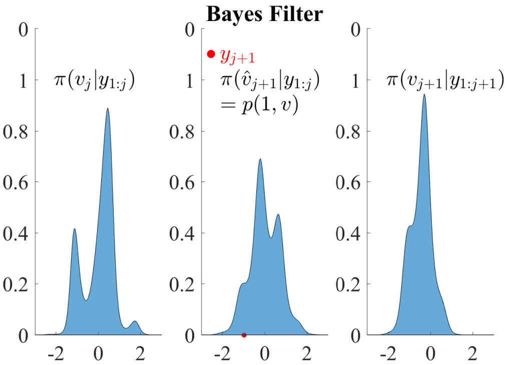
Conclusion: In principle we can solve these filtering problems exactly!

Example

Filtering problem:

$$V_{j+1} = V_j + \int_0^1 U'(V_{j+s})ds + \int_0^1 dW_s^{(j)}$$
$$Y_{j+1} = V_{j+1} + \eta_{j+1}$$

with $U(x) = x^2/2 + 0.15\sin(2\pi x)$ and for some j, we have set $\pi(v_j|y_{1:j}) \propto \exp\big(-2U(v_j) + \sin(4v_j)\big)$.



Overview

1 The Fokker-Planck equation

2 Numerical integration of SDE

3 Filtering problems with SDE dynamics

4 Examples using Euler–Maruyama integration

5 Model error and model fitting

Euler–Maruyama scheme

For the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \quad t \in [0, T], \qquad X_t|_{t=0} = X_0,$$

the Euler-Maruyama scheme on a uniform mesh $t_j = j\Delta t$

$$\bar{X}_{t_{j+1}} = \bar{X}_{t_j} + b(\bar{X}_{t_j})\Delta t + \sigma(\bar{X}_{t_j})\Delta W_j$$

where $\Delta W_j = W_{t_{j+1}} - W_{t_j}$ and $\bar{X}_0 = X_0$.

Motivation:

$$\begin{aligned} X_{t_{j+1}} - X_{t_j} &= \int_{t_j}^{t_{j+1}} b(X_t)dt + \int_{t_j}^{t_{j+1}} \sigma(X_t)dW_t \\ &\approx \int_{t_j}^{t_{j+1}} b(X_{t_j})dt + \int_{t_j}^{t_{j+1}} \sigma(X_{t_j})dW_t \end{aligned}$$

Let $\bar{X}_t := LinInterp(t; \{(t_j, \bar{X}_{t_j})\}_{j=0}^{T/\Delta t})$.

Strong convergence rate for Euler–Maruyama

Under the regularity assumptions in Thm 4, Lecture 19, most importantly

$$
\begin{aligned}
|b(x) - b(y)| + |\sigma(x) - \sigma(y)| &\leq K|x - y| \\
|b(x)|^2 + |\sigma(x)|^2 &\leq K(1 + |x|^2),
\end{aligned}
$$

the Euler–Maruyama method converges strongly with rate $1/2$.

$$
\sqrt{\max_{t \in [0,T]} \mathbb{E}\left[\,|\bar{X}_t - X_t|^2\right]} \leq C \Delta t^{1/2}
$$

for some $C > 0$.

Weak convergence rate Euler–Maruyama

Under more restrictive regularity conditions, the Euler–Maruyama converges weakly with rate 1.

$$\max_{t\in[0,T]} |\mathbb{E}\left[f(\bar{X}_t) - f(X_t)\right]| \leq C_f \Delta t$$

for any mapping $f \in C_P^\infty(\mathbb{R}^d, \mathbb{R})$ with $C_f > 0$ depending on f. [1]

Remark: See [ELV-E 7, 8] for more on results in higher-dimensional state space, and on higher order numerical methods.

[1] C_p^∞ is set of functions with at most polynomial growth in any partial derivative: $|\partial_\alpha f(x)| \leq C_\alpha |x|^{p_\alpha}$ for any $\alpha \in \mathbb{N}^d$, some $p_\alpha \in \mathbb{N}$ and all $x \in \mathbb{R}^d$.

Example - geometric Brownian motion

Consider the SDE

$$dX_t = X_t dt + X_t dW, \quad X_0 = 1,$$

and let us approximate: $\mathbb{E}[X_1] = e^t$ [Ubung 9].

Monte Carlo strategy:

1. Fix $\Delta t = 1/N$ and generate M numerical solutions of the SDE $\bar{X}_1^{(i)}$ by the EM scheme

$$\bar{X}_{t_{j+1}}^{(i)} = X_{t_j} + \bar{X}_{t_j}^{(i)} \Delta t + \bar{X}_{t_j}^{(i)} \Delta W_j^{(i)}, \quad j = 0, 1, \ldots, N-1,$$

   with independent Wiener paths $W^{(i)}$ and $X_0^{(i)} = 1$.

2. And apply the Monte Carlo method:

$$\mathbb{E}[X_1] = E_M[\bar{X}_1^{(\cdot)}] = \frac{1}{M} \sum_{i=1}^{M} \bar{X}_1^{(i)}$$

Illustration of approximation

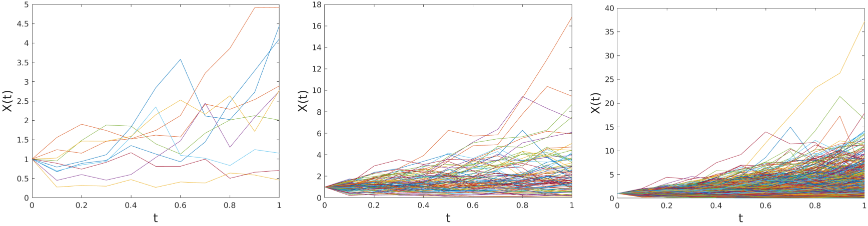


Figure: From left to right $(M, N) = (10, 10),\ (100, 10),\ (1000, 10)$.

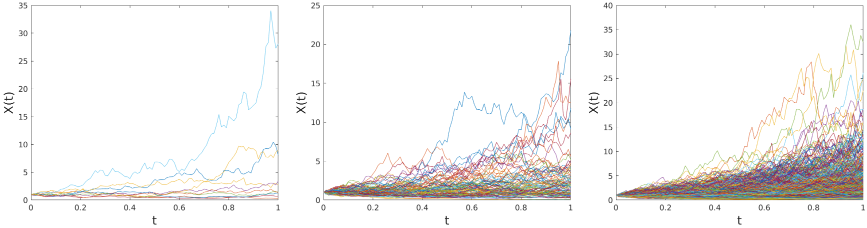


Figure: From left to right $(M, N) = (10, 100),\ (100, 100),\ (1000, 100)$

Approximation error

For any $f \in C_P(\mathbb{R}^d, \mathbb{R})$,

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[f(X_1)] - E_M[f(\bar{X}_1)])^2\right] &= \mathbb{E}\left[\left(\mathbb{E}[f(X_1)] \pm \mathbb{E}\left[f(\bar{X}_1)\right] - E_M[f(\bar{X}_1)]\right)^2\right] \\
&\le \mathbb{E}\left[(\mathbb{E}[f(X_1)] - \mathbb{E}\left[f(\bar{X}_1)\right])^2\right] \\
&+ 2\mathbb{E}\left[(\mathbb{E}[f(X_1)] - \mathbb{E}\left[f(\bar{X}_1)\right])(\mathbb{E}\left[f(\bar{X}_1)\right] - E_M[f(\bar{X}_1)])\right] \\
&+ \mathbb{E}\left[(\mathbb{E}\left[f(\bar{X}_1)\right] - E_M[f(\bar{X}_1)])^2\right] \\
&= (\mathbb{E}[f(X_1)] - \mathbb{E}\left[f(\bar{X}_1)\right])^2 + \mathbb{E}\left[(\mathbb{E}\left[f(\bar{X}_1)\right] - E_M[f(\bar{X}_1)])^2\right] \\
&\le C(N^{-2} + M^{-1}).
\end{aligned}$$

Computational cost of the Euler–Maruyama+Monte Carlo approach is of the order

$$Cost = M \times N.$$

Minimization of the erorr as a function of the cost:

$$M = \mathcal{O}(N^2) = \mathcal{O}(\Delta t^{-2})$$

Then

$$\|\mathbb{E}[f(X_1)] - E_M[f(\bar{X}_1)]\|_{L^2(\Omega)} = \mathcal{O}(\Delta t) = Cost^{-1/3}.$$

Overview

1 The Fokker-Planck equation

2 Numerical integration of SDE

3 Filtering problems with SDE dynamics

4 Examples using Euler–Maruyama integration

5 Model error and model fitting

Filtering problem

$$V_{\tau_{j+1}} = \Psi(V_{\tau_j}) := V_{\tau_j} + \int_0^{\Delta\tau} b(V_{\tau_j+t})dt + \int_0^{\Delta\tau} \sigma(V_{\tau_j+t})dW_t^{(j)}$$
$$Y_{\tau_{j+1}} = h(V_{\tau_{j+1}}) + \eta_{j+1}$$

where $\Delta\tau$ denotes the observation time interval, and $\eta_j \overset{iid}{\sim} N(0,\Gamma)$, $\Gamma > 0$ and $k \times k$ matrix, $\{W^{(j)}\}$ are independent Wiener processes and

$$V_0 \perp \{\eta_j\} \perp \{W^{(j)}\}$$

Shorthand notation: To align with previous notation we write $V_j := V_{\tau_j}$ and $Y_j := Y_{\tau_j}$.

Objective: Approximate the Bayes filter $\pi_{V_{\tau_j}|Y_{\tau_{1:j}}}$.

Exact model EnKF method

1. Set $\tau_j = 0$ and sample initial distribution $v_j^{(i)} \overset{iid}{\sim} \mathbb{P}_{V_0}$.

   and for $j = 0, 1, \ldots$:
2. **Prediction:** Simulate particles

$$\hat{v}_{\tau_{j+1}}^{(i)} = \Psi(v_{\tau_j}^{(i)}) = v_{\tau_j}^{(i)} + \int_0^{\Delta\tau} b(v_{\tau_j+t}^{(i)})dt + \int_0^{\Delta\tau} \sigma(v_{\tau_j+t}^{(i)})dW_t^{(j,i)},$$

   for $i = 1, 2, \ldots, M$, where $\{W^{(i,j)}\}_{i,j}$ are independent Wiener processes.
3. **Analysis:**

$$v_{\tau_{j+1}}^{(i)} = \hat{v}_{\tau_{j+1}}^{(i)} + K_{j+1}(y_{\tau_{j+1}}^{(i)} - h(\hat{v}_{\tau_{j+1}}^{(i)})),$$

   for $i = 1, 2, \ldots, M$ where

$$y_{\tau_{j+1}}^{(i)} = y_{\tau_{j+1}} + \eta_{j+1}^{(i)}, \qquad \eta_{j+1}^{(i)} \overset{iid}{\sim} N(0, \Gamma)$$

   and

$$K_{j+1} = \mathrm{Cov}_M[v_{\tau_{j+1}}^{(\cdot)}, h(v_{\tau_{j+1}}^{(\cdot)})](\mathrm{Cov}_M[h(v_{\tau_{j+1}}^{(\cdot)})] + \Gamma)^{-1}.$$

Problem: In many cases Ψ must be approximated by numerical integration.

Artificial example

Consider the Ornstein-Uhlenbeck process

$$\Psi(V_{\tau j}) = V_{\tau_j} - \int_0^{\Delta\tau} \theta V_{\tau_j + t} dt + \int_0^{\Delta\tau} dW_s \stackrel{D}{=} e^{-\theta\Delta\tau} V_{\tau_j}$$

with $\theta, \sigma > 0$.
We can solve this exactly:

$$\Psi(V_{\tau j}) \stackrel{D}{=} AV_{\tau_j} + \xi_j$$

where $\xi_j \sim N(0, \Sigma_{\Delta\tau})$ and we are in the familiar the linear-Gaussian setting. [see Ubung 9]

Approximation of the stochastic integrator

Let Ψ^N be the Euler-Maruyama approximation of

$$\Psi(V_{\tau_j}) = V_{\tau_j} + \int_0^{\Delta\tau} b(V_{\tau_j+t})dt + \int_0^{\Delta\tau} \sigma(V_{\tau_j+t})dW_t^{(j)}$$

using a uniform timestep $\Delta t = \Delta\tau/N$.

$\bar{V}_{\tau_{j+1}} = \Psi^N(\bar{V}_{\tau_j})$ is computed as follows

1. **Input:** $\bar{V}_{\tau_j}$.
2. For $k = 0 : N-1$, compute

$$\begin{aligned}\bar{V}_{\tau_j+(k+1)\Delta t} = \bar{V}_{\tau_j+k\Delta t} &+ b(\bar{V}_{\tau_j+k\Delta t})\Delta t \\ &+ \sigma(\bar{V}_{\tau_j+k\Delta t})\Big(W_{\tau_j+(k+1)\Delta t} - W_{\tau_j+k\Delta t}\Big)\end{aligned}$$

3. **Output:** $\bar{V}_{\tau_{j+1}} = \bar{V}_{\tau_j+N\Delta t}$

EnKF method using a numerical integrator

1. Set $\tau_j = 0$ and sample initial distribution $v_j^{(i)} \overset{iid}{\sim} \mathbb{P}_{V_0}$.

   and for $j = 0, 1, \dots$:

2. **Prediction:** Simulate particles

$$\hat{v}^{(i)}_{\tau_{j+1}} = \Psi^N(v^{(i)}_{\tau_j}),$$

   for $i = 1, 2, \dots, M$, where $\{W^{(i,j)}\}_{i,j}$ are independent Wiener processes used in Ψ^N.

3. **Analysis:**

$$v^{(i)}_{\tau_{j+1}} = \hat{v}^{(i)}_{\tau_{j+1}} + K_{j+1}(y^{(i)}_{\tau_{j+1}} - h(\hat{v}^{(i)}_{\tau_{j+1}})),$$

   for $i = 1, 2, \dots, M$ where

$$y^{(i)}_{\tau_{j+1}} = y_{\tau_{j+1}} + \eta^{(i)}_{j+1}, \qquad \eta^{(i)}_{j+1} \overset{iid}{\sim} N(0, \Gamma)$$

   and

$$K_{j+1} = \mathrm{Cov}_M[\hat{v}^{(\cdot)}_{\tau_{j+1}}, h(\hat{v}^{(\cdot)}_{\tau_{j+1}})](\mathrm{Cov}_M[h(\hat{v}^{(\cdot)}_{\tau_{j+1}})] + \Gamma)^{-1}.$$

Overview

1 The Fokker-Planck equation

2 Numerical integration of SDE

3 Filtering problems with SDE dynamics

4 Examples using Euler–Maruyama integration

5 Model error and model fitting

Ornstein-Uhlenbeck process

$$
\begin{aligned}
V_{\tau_{j+1}} &= \Psi(V_{\tau_j}) := V_{\tau_j} + -\frac{1}{4}\int_0^{\Delta\tau} V_{\tau_j+t}dt + \frac{1}{4}\int_0^{\Delta\tau} dW_t^{(j)} \\
Y_{\tau_{j+1}} &= V_{\tau_{j+1}} + \eta_{j+1}
\end{aligned}
$$

with $V_0 = 1$, $\Delta\tau = 1/2$ and $\eta_j \sim N(0, \Gamma)$.

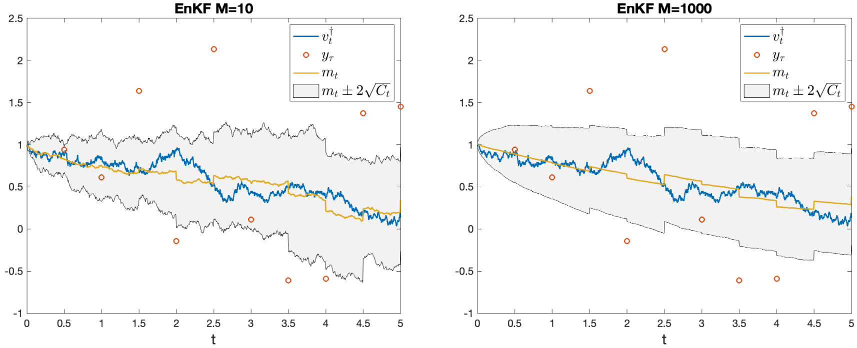
We generate an observation sequence for $y_{\tau_{1:J}}$ for $J = 10$ from synthetic data $v^\dagger_{\tau_{1:J}}$.

Approximation method: EnKF with numerical integrator Ψ^N with $N = 100$ and $\Delta t = \Delta\tau/N$.

Note: Continuum dynamics makes it possible to also estimate the filtering distribution for times between observation times.

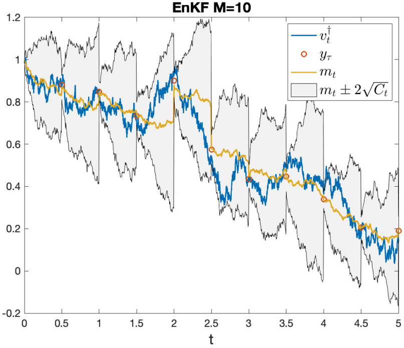
Numerical results

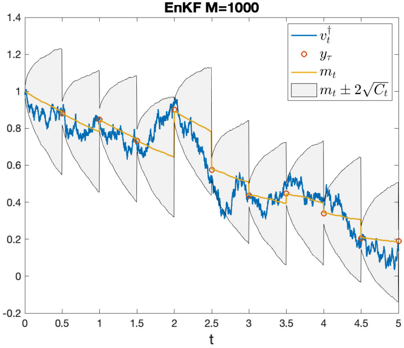
Large uncertainty in observations, $\Gamma = 1$, yields small correction at observation times:



Numerical results

Small uncertainty in observations, $\Gamma = 1/1000$, yields small correction at observation times:





Langevin equation

$$dX_t = V_t dt$$
$$dV_t = (-0.25 V_t - U'(X_t))dt + 0.5 dW_t$$

with $(X_0, V_0) = (0, 1)$.

Observations

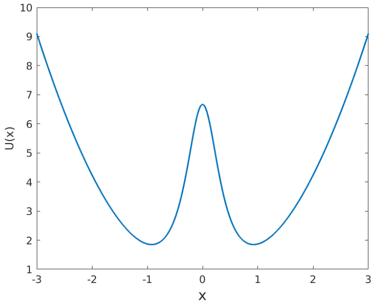
$$Y_{\tau_k} = V_{\tau_k} + \eta_k, \quad k = 1, 2, \ldots$$

with $\eta \sim N(0, \Gamma)$.

The state X_t will oscillate between local minima of $U(x)$.

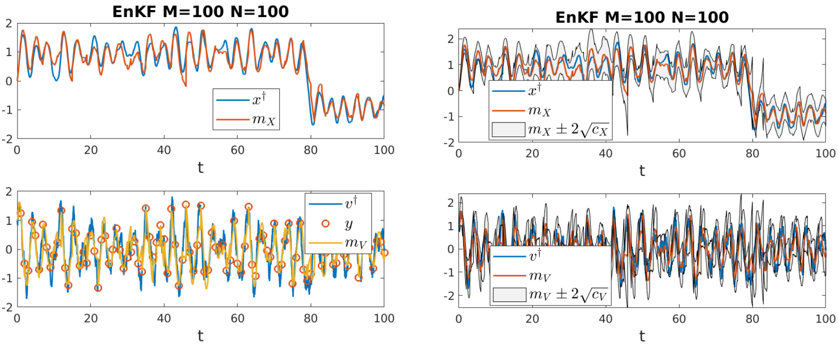
Can we infer the pseudo-stable state of X_t from observing V_t?

Potential:
$U(X) = X^2 + 1/(0.15 + X^2)$



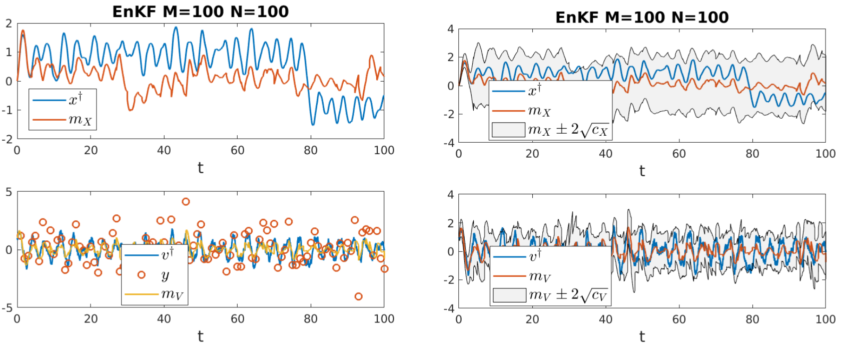
Numerical results

Small observation noise $\Gamma = 1/100$ and infrequent observations $\Delta\tau = 1$,



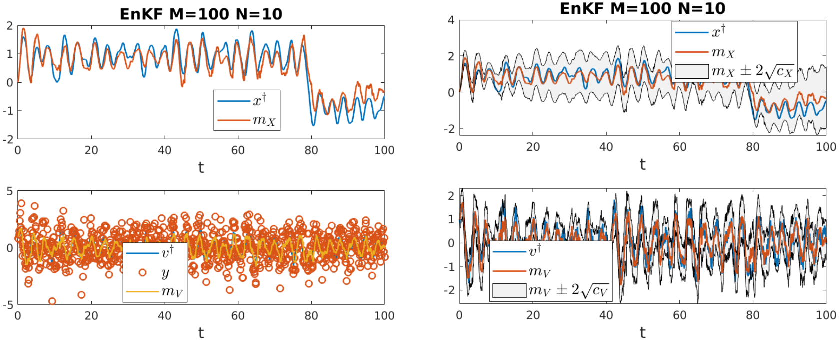
Numerical results

Large observation noise $\Gamma = 1$ and infrequent observations $\Delta\tau = 1$,



Numerical results

Large observation noise $\Gamma = 1$ and frequent observations $\Delta\tau = 0.1$,



Model and approximation error for EnKF

Let $\pi_j^{M,N}$ denote the EnKF empirical measure at time τ_j with ensemble size M and timestep $\Delta t = \Delta\tau/N$ in the Euler–Maruyama integrator. Then, under sufficient regularity it holds for QoI f that

$$\|\pi_j^{M,N}[f] - \pi_j^{\infty,\infty}[f]\|_{L^p(\Omega)} \leq C_{p,j,f}(M^{-1/2} + N^{-1}).$$

$\pi_j^{\infty,\infty}$ – mean-field large-ensemble limit with $N = \infty$ exact-model integration. [Hoel, Law, Tempone (2016)].

Rule of thumb configuration of degrees of freedom in EnKF with Euler–Maruyama: $M = \mathcal{O}(N^2)$.

The error may be split into/bounded from above by

$$\|\pi_j^{M,N}[f] - \pi_j^{\infty,\infty}[f]\|_p \leq \underbrace{\|\pi_j^{M,N}[f] - \pi_j^{\infty,N}[f]\|_p}_{\text{bias error}} + \underbrace{\|\pi_j^{M,N}[f] - \pi_j^{\infty,N}[f]\|_p}_{\text{statistical error}}$$

Bias error is a particular kind of model error, using Ψ^N rather than the exact model Ψ as solver.

Overview

1 The Fokker-Planck equation

2 Numerical integration of SDE

3 Filtering problems with SDE dynamics

4 Examples using Euler–Maruyama integration

5 Model error and model fitting

Model uncertainty

Assume that we are given a sequence of observations $y_{1:J}$, or a collection of such, that satisfy

$$Y_j = h(V_j) + \eta_j.$$

The exact dynamics for V_j, which we denote Ψ, is unknown, but we can sample from a set of approximate dynamics $\{\Psi_\alpha\}_{\alpha \in \mathcal{M}o}$. That is

$$\textbf{Unknown dyn: } V_{j+1} = \Psi(V_j), \quad \textbf{known approx dyn } V_{j+1}^\alpha = \Psi_\alpha(V_j^\alpha).$$

Question: given the collection of observations $y_{1:J}$ and the true observation model, how can we estimate model errors and compare models?

Strategy: Estimate error in the data space rather than in the state space.

Non-Bayesian approach

Assume the setting of exact observations

$$Y_j = h(V_j).$$

Given a collection of M_0 observation sequences $\{y_{1:J}^{(i)}\}_{i=1}^{M_O}$, we associate it to an empirical measure $\pi_{Y_{1:J}}(y_{1:J})$.

Computing the error for Ψ_α:

- Generate M_D path realizations of the dynamics $\{v_{1:J}^{\alpha,(i)}\}_{i=1}^{M_D}$.
- Associate each of these paths to observation sequences $y_{1:J}^{\alpha,(i)} = h(v_{1:J}^{\alpha,(i)})$.
- Approximate the error/divergence etc with the relevant measure in the data space. For instance, root-mean-square error,

$$RMSE(\alpha) = \| Y_{1:J}^{\alpha} - \mathbb{E}[\, Y_{1:J}] \|_{L^2(\Omega)} \approx \sqrt{\frac{1}{M_D}\sum_{i=1}^{M_D} |y_{1:J}^{\alpha,(i)} - E_{M_O}[y_{1:J}^{(\cdot)}]|^2}$$

- Best model: $\alpha^* = \arg\min_{\alpha \in \mathcal{M}_O} RMSE(\alpha)$.

[See RC 4.4] for more on scoring rules.

Bayesian approach to model selection

Assume we are given one observation sequence $Y_{1:J} = y_{1:J}$ from the noisy observation model

$$Y_{1:J} = h(V_{1:J}) + \eta_{1:J}$$

where we assume the "truth" $V^{\dagger}_{1:J}$ that produced the observation was generated from a model Ψ_α for some $\alpha \in \mathcal{M}o$.

Bayesian framework:

1. Assign a prior pdf π_α to the model space.
2. and Bayesian inversion yields

$$\pi_{\alpha|Y_{1:J}}(\alpha|y_{1:J}) \propto \pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha)\pi_\alpha(\alpha)$$

3. Select model for instance by

$$\alpha^* = MAP(\pi_{\alpha|Y_{1:J}}(\cdot|y_{1:J}).$$

Problem: evaluating $\pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha)$ may not be straightforward.

Approximating the likelihood

Note that

$$\begin{aligned}
\pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha) &= \int \pi_{Y_{1:J},V_{1:J}|\alpha}(y_{1:J}, v_{1:J}|\alpha) dv_{1:J} \\
&= \int \pi_{Y_{1:J}|V_{1:J},\alpha}(y_{1:J}|v_{1:J}, \alpha) \pi_{V_{1:J}|\alpha}(v_{1:J}|\alpha) dv_{1:J} \\
&= \int \pi_{Y_{1:J}|V_{1:J}}(y_{1:J}|v_{1:J}) \pi_{V_{1:J}|\alpha}(v_{1:J}|\alpha) dv_{1:J} \\
&= \mathbb{E}\left[\pi_{Y_{1:J}|V_{1:J}}(y_{1:J}|V_{1:J})|\alpha \right]
\end{aligned}$$

Hence, the likelihood can be approximated by the Monte Carlo method:

$$\pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha) \approx \sum_{i=1}^{M} \frac{\pi_{Y_{1:J}|V_{1:J}}(y_{1:J}|V_{1:J}^{(i)})}{M}$$

where $V_{1:J}^{(i)} \overset{iid}{\sim} \pi_{V_{1:J}|\alpha}(\cdot|\alpha)$.

Toy problem

Dynamics

$$V_{j+1} = \alpha V_j,$$

with $V_0 = 1$ and prior $\pi_\alpha(\alpha) = \mathbb{1}_{[-1,1]}(\alpha)$ Observations

$$Y_{j+1} = V_{j+1} + \eta_{j+1}, \qquad \eta_j \overset{iid}{\sim} N(0,1).$$

Observation sequence $y_j = (-1)^j$ for $j = 1, 2, \dots, J$.

Since $V_j = \alpha^j$ (each α leads to a unique dynamics), we derive that

$$\pi_{\alpha|Y_{1:J}}(\alpha|y_{1:J}) \propto \pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha)\pi_\alpha(\alpha) \propto \mathbb{1}_{[-1,1]} \exp\left(-\frac{1}{2}\sum_{j=1}^{J}\left((-1)^j - \alpha^j\right)^2\right)$$

We conclude that

$$MAP\left(\pi_{\alpha|Y_{1:J}}(\cdot|y_{1:J})\right) = -1.$$

Model parameter estimation/selection through filtering

Consider the parameter dependent dynamics

$$V_{\tau_{j+1}} = \Psi_\alpha(V_{\tau_j})$$

and a sequence of observations

$$Y_{\tau_{j+1}} = h(V_{\tau_{j+1}}) + \eta_{j+1}$$

Filtering strategy to parameter estimation: Augment the state space with α. New dynamics $(V_{\tau_j}, \alpha_{\tau_j})$:

$$V_{\tau_{j+1}} = \Psi_{\alpha_{\tau_j}}(V_{\tau_j})$$
$$\alpha_{\tau_{j+1}} = \alpha_{\tau_j} + \nu_j$$

where ν_j is noise. (Adding noise may improve the exploration of possible α but, unless careful, it may also render the dynamics unstable!)

Can be implemented using for instance EnKF or particle filtering with the goal that $\alpha_{\tau_j} \to \alpha_{true}$. [See ubung 9].

Summary

- The density of SDE is described by the Fokker-Planck equation.

- Have introduced the Euler–Maruyama numerical scheme for SDE studied applications of EnKF+Euler–Maruyama model approximation.

- Similarly, one may combine particle filtering/3DVar/ExKF and Euler–Maruyama (and more).

- Next time: continuous-time filtering for linear-coefficient SDEs.