Mathematics and numerics for data assimilation and state estimation – Lecture 21

Summer semester 2020

Overview

1. Model error and model fitting
2. Kalman–Bucy filter
3. Continuous-time limit of discrete-time filtering
4. The Kushner–Stratonovich equation
5. Filtering in high/infinite-dimensional state space

Summary lecture 20

- Fokker-Planck equation, numerical integration of SDE and applications in filtering problems. Filtering methods for continuous-time dynamics and discrete-time observations.

- Plan for today: Model error and fitting. Filtering in continuous-time dynamics and observations, and filtering in high-dimensional state-space.

Overview

1. Model error and model fitting
2. Kalman–Bucy filter
3. Continuous-time limit of discrete-time filtering
4. The Kushner–Stratonovich equation
5. Filtering in high/infinite-dimensional state space

Model uncertainty

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

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

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

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

Question: given a collection $y_{1:J}$ and the true observation model, how can we estimate model errors to compare different 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 $\pi_{Y_{1:J}}(dy_{1:J})$.

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 error 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

$$\begin{aligned}\pi_{Y_{1:J}|\alpha}(y_{1:J}|\alpha) &= \int \pi_{Y_{1:J},V_{1:J}|\alpha}(y_{1:J}, v_{1:J}|\alpha)dv_{1:J} \\ &= \int \pi_{Y_{1:J}|V_{1:J},\alpha}(y_{1:J}|v_{1:J},\alpha)\pi_{V_{1:J}|\alpha}(v_{1:J}|\alpha)dv_{1:J} \\ &= \int \pi_{Y_{1:J}|V_{1:J}}(y_{1:J}|v_{1:J})\pi_{V_{1:J}|\alpha}(v_{1:J}|\alpha)dv_{1:J}.\end{aligned}$$

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

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

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

Toy problem

Dynamics

$$V_{j+1} = \alpha V_j, \qquad V_0 = 1,$$

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

Observations

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

and given obs sequence $y_j = (-1)^j$ for $j = 1, 2, \ldots, J$.

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

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

We conclude that

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

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].

Overview

Continuous-time observations

We now shift to studying filtering problems with dynamics

$$dV_t = b(V_t)dt + \sigma(V_t)dW_t, \qquad t \geq 0$$

and **continuous-time observations**

$$Y_t = h(V_t) + \gamma(V_t)\dot{U}_t, \qquad t \geq 0,$$

where W and U are independent Wiener processes (and $\dot{U}$ is white noise: in this case, the formal derivative of a Wiener process).

For mathematical convenience (to go from white noise to Itô SDE), one rather consider the observation

$$Z_t = \int_0^t Y_s ds = \int_0^t h(V_s)ds + \int_0^t \gamma(V_s)\frac{dU_s}{\cancel{ds}}\cancel{ds}$$

or, equivalently,

$$dZ_t = h(V_t)dt + \gamma(V_t)dU_t, \qquad Z_0 = 0.$$

1D linear filtering problem

Dynamics

$$dV_t = LV_t dt + \sigma dW_t, \quad V_0 \sim N(m_0, C_0)$$

and observations

$$dZ_t = HV_t dt + \gamma dU_t, \quad Z_0 = 0,$$

with scalars H, L, σ, γ and $\gamma > 0$, and $\{W_t\} \perp \{U_t\} \perp V_0$.

Theorem 1 (1D Kalman-Bucy filter)

Both V_t and Z_t are Gaussian processes, and $V_t | Z_{[0,t]} = z_{[0,t]} \sim N(m_t, C_t)$ with

$$dm_t = \left(L - \frac{H^2 C_t}{\gamma^2}\right) m_t dt + \frac{HC_t}{\gamma^2} dz_t, \qquad m_0 = \mathbb{E}\left[V_0 \right],$$

and C_t solving the Riccati equation

$$\dot{C}_t = 2LC_t - \frac{H^2}{\gamma^2} C_t^2 + \sigma^2, \quad C_0 = \mathrm{Var}[V_0],$$

Remark: The result is typically presented as SDE for $\hat{V}_t = \mathbb{E}\left[V_t | Z_{[0,t]} \right]$.

Example [Oksendal 6.2.9]

Noisy observations of a constant process

$$\begin{aligned} dV_t &= 0, & V_0 &\sim N(m_0, C_0) \\ dZ_t &= V_t dt + \gamma dU_t, & Z_0 &= 0, \end{aligned}$$

Equations for moments in $V_t | Z_{[0,t]} = z_{[0,t]} \sim N(m_t, C_t)$:

$$dm_t = -\frac{C_t}{\gamma^2} m_t dt + \frac{HC_t}{\gamma^2} dz_t,$$

and

$$\dot{C}_t = -\frac{1}{\gamma^2} C_t^2 \implies C_t = \frac{C_0 \gamma^2}{\gamma^2 + C_0 t}$$

Plugging C_t into equation for m_t gives

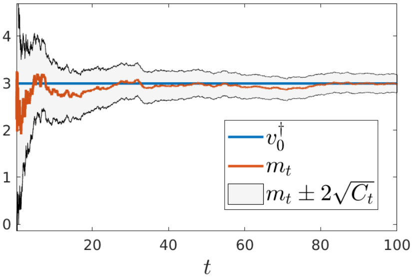
$$m_t = \frac{\gamma^2}{\gamma^2 + C_0 t} m_0 + \frac{C_0}{\gamma^2 + C_0 t} z_t$$

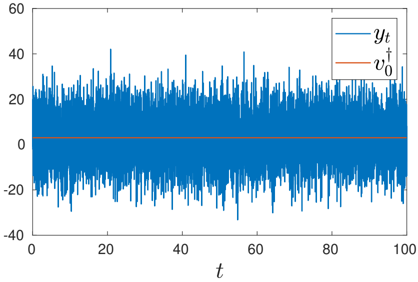
Hence $m_t \approx z_t / t$ for $t \gg 1$.

Illustration

With $\gamma = 1$, $V_0 \sim N(m_0 = 1, C_0 = 4)$ and $v_0^\dagger = 3$ and

$$z_t = v_0^\dagger t + u_t^\dagger, \qquad \text{(and numerical approx of)} \quad y_t = \dot{z}_t$$





Example [Oksendal 6.2.10]

Noisy observation of a Wiener process

$$\begin{aligned} dV_t &= dW_t, \qquad V_0 \sim N(m_0 = 0, C_0 = 0) \\ dZ_t &= V_t dt + dU_t, \qquad Z_0 = 0 \end{aligned}$$

Yields that $V_t | Z_{[0,t]} = z_{[0,t]} \sim N(m_t, C_t)$ where

$$\frac{dC_t}{\sigma^2 - C_t^2} = dt \implies C_t = \frac{\exp(2t) - 1}{\exp(2t) + 1} = \tanh(t),$$

and

$$m_t = \frac{1}{e^t + e^{-t}} \int_0^t \left(e^s - e^{-s} \right) dz_s$$

Observation: the conditional mean weights recent observations more than the distant past:

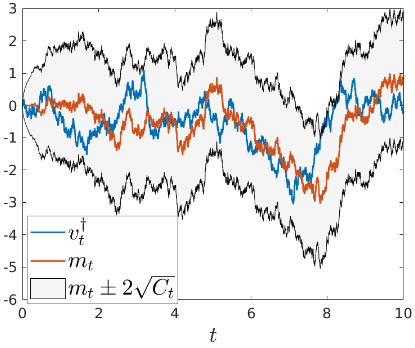
$$m_t \approx \int_0^t \frac{e^s}{e^t} dz_s = \int_0^t \frac{e^s}{e^t} y_s ds$$

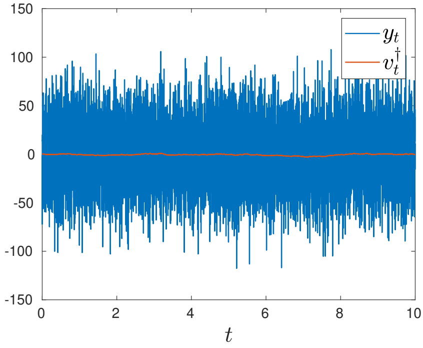
for $t \gg 1$, recalling that $z_t = \int_0^t y_s ds$.

Illustration

Numerical approximations of

$$z_t = \int_0^t v_s^\dagger \, ds + u_t^\dagger, \qquad \text{and} \quad y_t = \dot{z}_t$$





Overview

From discrete to continuous time

The continuous-time filtering problem

$$dV_t = b(V_t)dt + \sqrt{\Sigma_0} dW_t$$
$$dZ_t = h(V_t) + \sqrt{\Gamma_0} dU_t$$

with positive definite Γ_0, Σ_0 can be formally derived as the continuous-time limit of

$$\begin{aligned} V_{\tau_{j+1}} &= V_{\tau_j} + b(V_{\tau_j})\Delta\tau + \sqrt{\Sigma}\xi_j \\ Y_{\tau_{j+1}} &= h(V_{\tau_{j+1}}) + \sqrt{\Gamma}\eta_j \end{aligned} \qquad (1)$$

where ξ_j, η_k iid standard Gaussians.

First introduce the discrete primitive

$$\frac{Z_{\tau_{j+1}} - Z_{\tau_j}}{\Delta\tau} = Y_{\tau_{j+1}} \quad \text{(recalling that for continuous problem} \quad \dot{Z}_t = Y_t)$$

and the scaling

$$\Sigma = \Delta\tau\Sigma_0, \quad \Gamma = \frac{1}{\Delta\tau}\Gamma_0$$

Then

$$V_{\tau_{j+1}} - V_{\tau_j} = b(V_{\tau_j})\Delta\tau + \sqrt{\Sigma_0 \Delta\tau}\xi_j$$
$$Z_{\tau_{j+1}} - Z_{\tau_j} = h(V_{\tau_{j+1}})\Delta\tau + \Delta\tau\sqrt{\frac{\Gamma_0}{\Delta\tau}}\eta_j$$

Recalling that

$$W_{\tau_{j+1}} - W_{\tau_j} \sim N(0, \Delta\tau) \quad \text{and} \quad U_{\tau_{j+1}} - U_{\tau_j} \sim N(0, \Delta\tau),$$

we rewrite

$$V_{\tau_{j+1}} - V_{\tau_j} = b(V_{\tau_j})\Delta\tau + \sqrt{\Sigma_0}\Delta W_j$$
$$Z_{\tau_{j+1}} - Z_{\tau_j} = h(V_{\tau_{j+1}})\Delta\tau + \sqrt{\Gamma_0}\Delta U_j$$

and obtain in the limit $\Delta\tau \downarrow 0$,

$$dV_t = b(V_t)dt + \sqrt{\Sigma_0}dW_t$$
$$dZ_t = h(V_t)dt + \sqrt{\Gamma_0}dU_t$$

A second look at the Kalman-Bucy filter

Theorem 2 (Multidimensional Kalman-Bucy filter [LSZ Thm 8.1])

Consider

$$dV_t = LV_t dt + \sqrt{\Sigma_0} dW_t, \qquad V_0 \sim N(m_0, C_0)$$
$$dZ_t = HV_t dt + \sqrt{\Gamma_0} dU_t, \qquad Z_0 = 0,$$

with $L, \Sigma_0 \in \mathbb{R}^{d\times d}$, $H \in \mathbb{R}^{k\times d}$ *and* $\Gamma_0 \in \mathbb{R}^{k\times k}$, *positive definite* Γ_0, Σ_0, *and independence* $\{W_t\} \perp \{U_t\} \perp V_0$.
Then V_t *and* Z_t *are Gaussian processes, and* $V_t | Z_{[0,t]} = z_{[0,t]} \sim N(m_t, C_t)$ *with*

$$dm_t = Lm_t dt + C_t H^T \Gamma_0^{-1}(dz_t - Hm_t dt), \qquad m_0 = \mathbb{E}\left[V_0 \right],$$

and (the matrix ODE)

$$\dot{C}_t = LC_t + C_t L + \Sigma_0 - C_t H^T \Gamma_0^{-1} HC_t, \qquad C_0 = \mathrm{Var}[V_0],$$

Sketch of proof

Let us look at the continuous-time limit of the Kalman filter

$$V_{\tau_{j+1}} - V_{\tau_j} = LV_{\tau_j}\Delta\tau + \sqrt{\Sigma_0\Delta\tau}\xi_j$$

$$\frac{Z_{\tau_{j+1}} - Z_{\tau_j}}{\Delta\tau} = HV_{\tau_{j+1}} + \sqrt{\frac{\Gamma_0}{\Delta\tau}}\eta_j$$

i.e., of

$$V_{\tau_{j+1}} = (I + L\Delta\tau)V_{\tau_j} + \sqrt{\Sigma_0\Delta\tau}\xi_j$$

$$Y_{\tau_{j+1}} = HV_{\tau_{j+1}} + \sqrt{\frac{\Gamma_0}{\Delta\tau}}\eta_j$$

With $V_{\tau_j}|Y_{\tau_{1:j}} = y_{\tau_{1:j}} \sim N(m_{\tau_j}, C_{\tau_j})$, Kalman filtering (with $A = (I + L\Delta\tau)$) yields the prediction

$$\hat{m}_{\tau_{j+1}} = (I + L\Delta\tau)m_{\tau_j} = m_{\tau_j} + Lm_{\tau_j}\Delta\tau$$

$$\begin{aligned}\hat{C}_{\tau_{j+1}} &= (I + L\Delta\tau)C_{\tau_j}(I + L\Delta\tau)^T + \Delta\tau\Sigma_0 \\ &= C_{\tau_j} + (LC_{\tau_j} + C_{\tau_j}L^T + \Sigma_0)\Delta\tau + o(\Delta\tau)\end{aligned}$$

And for the analysis

$$K = \hat{C}_{\tau_{j+1}} H^T (H \hat{C}_{\tau_{j+1}} H^T + \frac{\Gamma_0}{\Delta\tau})^{-1} = C_{\tau_j} H^T \Gamma_0^{-1} \Delta\tau + o(\Delta\tau)$$

and

$$y_{\tau_{j+1}} - H\hat{m}_{\tau_{j+1}} = y_{\tau_{j+1}} - H(m_{\tau_j} + L m_{\tau_j} \Delta\tau)$$

so that

$$\begin{aligned} m_{\tau_{j+1}} &= \hat{m}_{\tau_{j+1}} + K(y_{\tau_{j+1}} - H\hat{m}_{\tau_{j+1}}) \\ &= m_{\tau_j} + L m_{\tau_j} \Delta\tau + C_{\tau_j} H^T \Gamma_0^{-1} (y_{\tau_{j+1}} - H m_{\tau_j}) \Delta\tau + o(\Delta\tau). \end{aligned}$$

And, recalling that $\hat{C}_{\tau_{j+1}} = C_{\tau_j} + (L C_{\tau_j} + C_{\tau_j} L^T + \Sigma_0)\Delta\tau + o(\Delta\tau)$,

$$\begin{aligned} C_{\tau_{j+1}} &= (I - KH)\hat{C}_{\tau_{j+1}} \\ &= C_{\tau_j} + (L C_{\tau_j} + C_{\tau_j} L^T + \Sigma_0)\Delta\tau - C_{\tau_j} H^T \Gamma_0^{-1} H C_{\tau_j} \Delta\tau + o(\Delta\tau). \end{aligned}$$

Next, truncate $o(\Delta\tau)$ terms and rewrite as follows

Up to order $\Delta\tau$,

$$\begin{aligned} m_{\tau_{j+1}} - m_{\tau_j} &= Lm_{\tau_j}\Delta\tau + C_{\tau_j}H^T\Gamma_0^{-1}(y_{\tau_{j+1}} - Hm_{\tau_j})\Delta\tau \\ &= Lm_{\tau_j}\Delta\tau + C_{\tau_j}H^T\Gamma_0^{-1}(z_{\tau_{j+1}} - z_{\tau_j} - Hm_{\tau_j}\Delta\tau), \end{aligned}$$

where we used that $y_{\tau_{j+1}}\Delta\tau = z_{\tau_{j+1}} - z_{\tau_j}$, and

$$C_{\tau_{j+1}} - C_{\tau_j} = (LC_{\tau_j} + C_{\tau_j}L^T + \Sigma_0)\Delta\tau - C_{\tau_j}H^T\Gamma_0^{-1}HC_{\tau_j}\Delta\tau.$$

Taking the limit $\Delta\tau \downarrow 0$ leads to Kalman-Bucy equations:

$$dm = Lmdt + CH^T\Gamma_0^{-1}(dz - Hmdt),$$

and

$$dC = \Big(LC + CL^T + \Sigma_0 - CH^T\Gamma_0^{-1}HC\Big)dt.$$

Nonlinear filtering methods

For the nonlinear filtering problem

$$\begin{aligned} dV_t &= b(V_t)dt + \sqrt{\Sigma_0}dW_t \\ dZ_t &= HV_t + \sqrt{\Gamma_0}dU_t \end{aligned} \tag{2}$$

there exist, as in the discrete-time setting, approximate Gaussian filtering methods: 3DVAR, ExKF, EnKF, and (also non-Gaussian methods, e.g., particle filters) [LSZ Chapter 8.2].

Definition 3 (Continuous-time ExKF)

The distribution of $V_t | Z_{[0,t]} = z_{[0,t]}$ is approximated by $N(m_t, C_t)$ where

$$dm = b(m_t)dt + C_t H^T \Gamma_0^{-1}(dz_t - Hm_t dt), \qquad m_0 = \mathbb{E}\left[V_0 \right],$$

and

$$\dot{C}_t = Db(m_t)C_t + C_t(Db(m_t))^T + \Sigma_0 - C_t H^T \Gamma_0^{-1} HC_t, \qquad C_0 = \mathrm{Var}[V_0],$$

with Db denoting the Jacobian of $b : \mathbb{R}^d \to \mathbb{R}^d$.

Derivation of continuous-time ExKF:

- Apply ExKF on discrete-time approximation of (2),

$$\begin{aligned} V_{\tau_{j+1}} &= V_{\tau_j} + b(V_{\tau_j})\Delta\tau + \sqrt{\Sigma}\xi_j \\ Y_{\tau_{j+1}} &= h(V_{\tau_{j+1}}) + \sqrt{\Gamma}\eta_j \end{aligned} \tag{3}$$

  leading to a system of difference equations for m_{τ_j} and C_{τ_j}.

- Taking the continuous-time limit $\Delta\tau \downarrow 0$ leads to the ExKF system of differential equations for m and C, similarly as for Kalman-Bucy.

Example – Nonlinear filtering

Consider the following stochastic version of Lorenz 63

$$\begin{aligned} dv_1 &= \alpha(v_1 - v_2)dt + \sigma dW_1(t) \\ dv_2 &= -\Big(\alpha v_v + v_2 + v_1 v_3\Big)dt + \sigma dW_2(t) \\ dv_3 &= v_1 v_2 - b v_3 + b(r + \alpha) + \sigma dW_3(t) \end{aligned} \tag{4}$$

with $\sigma = 2$ and standard coefficient values $(\alpha, b, r) = (10, 8/3, 28)$.

On compact form, with $v = (v_1, v_2, v_3)^T$ and $W = (W_1, W_2, W_3)$, we rewrite

$$dv = f(v)dt + \sigma dW.$$

Observations:

$$dz = Hvdt + \gamma dU, \qquad z(0) = 0,$$

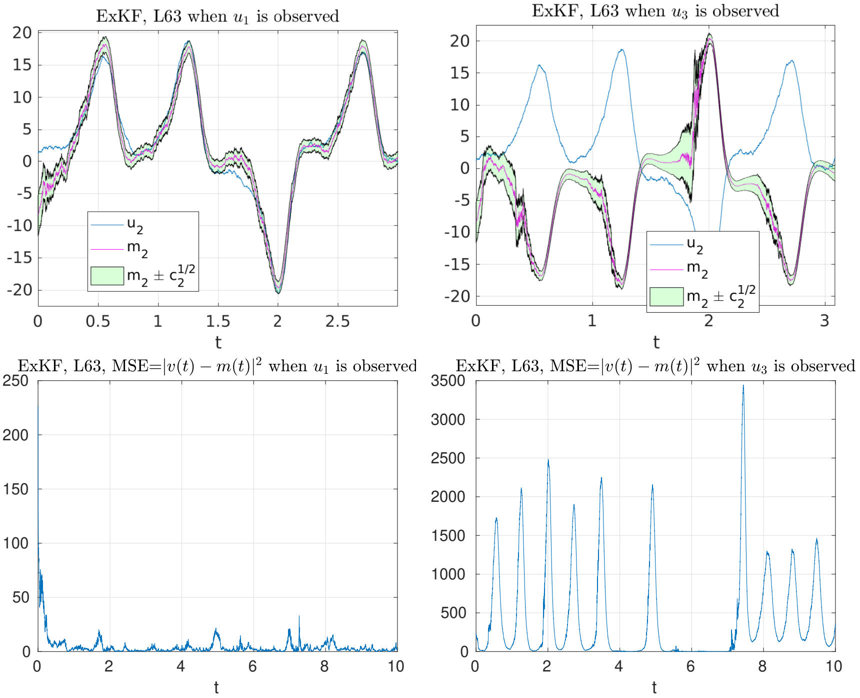
with either $H = (0, 0, 1)$ or $(1, 0, 0)$ and $\gamma = 1/2$.

Objective: "continuous-time" ExKF filtering with estimates of m_0 and C_0 given (in practice, fine-timestep numerical integration of associated discrete-time filtering problem).

Main steps in code (more details in [LSZ p16c.m])

```
%Initial data
m0=zeros(3,1); C0=eye(3);% prior initial condition covariance
v(:,1)=m0+sqrtm(C0)*randn(3,1);% initial truth
m(:,1)=10*randn(3,1);% initial mean/ESTIMATE
c(:,:,1)=10*C0;% initial covariance operator/ESTIMATE
H=[1,0,0];% observation operator
tau=1e-4;% time discretization is tau

%% solution % assimilate!
for j=1:J
    % truth
    v(:,j+1)=v(:,j)+tau*f(v(:,j))+sigma*sqrt(tau)*randn(3,1);
    z(:,j+1)=z(:,j)+tau*H*v(:,j+1) + gamma*sqrt(tau)*randn;% observa
    mhat=m(:,j)+tau*f(m(:,j));% estimator predict
    chat=(I+tau*Df(m(:,j)))*c(:,:,j)* ...
        (I+tau*Df(m(:,j)))'+sigma^2*tau*I;% covariance predict
    d=(z(j+1)-z(j))/tau-H*mhat;% innovation
    K=(tau*chat*H')/(H*chat*H'*tau+gamma^2);% Kalman gain
    m(:,j+1)=mhat+K*d;% estimator update
    c(:,:,j+1)=(I-K*H)*chat;% covariance update
end
```



ExKF, L63 when u_1 is observed
ExKF, L63 when u_3 is observed
u_2
m_2
$m_2 \pm c_2^{1/2}$
t
ExKF, L63, MSE=$|v(t) - m(t)|^2$ when u_1 is observed
ExKF, L63, MSE=$|v(t) - m(t)|^2$ when u_3 is observed

Overview

In linear-Gaussian settings, the continuous-time filtering problem can be solved exactly.

Under sufficient regularity, we have the following extension to nonlinear settings:

Theorem 4 (Kushner–Stratonovich equation)

Consider the filtering problem

$$
\begin{aligned}
dV_t &= b(V_t)dt + \sqrt{\Sigma_0}dW_t, \quad V_0 \sim p(0,x) \\
dZ_t &= h(V_t)dt + \sqrt{\Gamma_0}dU_t, \quad Z_0 = 0.
\end{aligned}
$$

If b and h are sufficiently smooth, then there exists a pdf $p(t,x)$ for $V_t | Z_{[0,t]} = z_{[0,t]}$, and it is the solution of

$$
p_t(t,x) = \mathcal{L}^* p(t,x) + p(t,x)\left(h(x) - \int_{\mathbb{R}^d} h(y)p(t,y)dy\right)\Gamma_0^{-1}\left(\frac{dz}{dt} - \int_{\mathbb{R}^d} h(y)p(t,y)dy\right)
$$

over $(x,t) \in \mathbb{R}^d \times [0,T]$, and (cf. the Fokker-Planck equation)

$$
\mathcal{L}^* p(t,x) = -\nabla_x \cdot (b(x)p(t,x)) + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{x_i x_j}(\Sigma_{0,ij}\, p(t,x)).
$$

Overview

Problem description

Filering problem

$$\left.\begin{array}{ll}\text{Dynamics} & u_{\tau_{j+1}} = \Psi(u_{\tau_j}) \\ \text{Observations} & Y_{\tau_{j+1}} = Hu_{\tau_{j+1}} + \eta_{j+1}\end{array}\right\} \qquad j = 0, 1, \ldots$$

With

- u_{τ_j} almost surely belongs to an infinite-dimensional Hilbert space $\mathcal{H}$, e.g., $u(\tau_j, \cdot) \in L^2(\mathbb{R})$.
- dynamics is possibly non-linear $\Psi : L^2(\Omega, \mathcal{H}) \to L^2(\Omega, \mathcal{H})$,
- linear operator $H : \mathcal{H} \to \mathbb{R}^k$.
- hence **finite-dimensional observations** y_{τ_j}, and $\eta_j \overset{iid}{\sim} N(0, \Gamma)$ with $\eta_j \perp \{u_{\tau_j}\}$, and $\Gamma \in \mathbb{R}^{k \times k}$.

Objective: Approximate pdf of $u_{\tau_j} | Y_{\tau_{1:j}} = y_{\tau_{1:j}}$.

Forward model approximation: $\Psi \approx \Psi^{N_t, N_x}$ with discretization both time and space. Something like $\Delta t = \Delta\tau / N_t$ and $\Delta x = \mathcal{O}(N_x^{-1})$.

Problem description

Filtering problem

$$\left.\begin{array}{ll}\text{Dynamics} & u_{\tau_{j+1}} = \Psi(u_{\tau_j}) \\ \text{Observations} & Y_{\tau_{j+1}} = Hu_{\tau_{j+1}} + \eta_{j+1}\end{array}\right\} \qquad j = 0, 1, \ldots$$

With

- u_{τ_j} almost surely belongs to an infinite-dimensional Hilbert space $\mathcal{H}$, e.g., $u(\tau_j, \cdot) \in L^2(\mathbb{R})$.
- dynamics is possibly non-linear $\Psi : L^2(\Omega, \mathcal{H}) \to L^2(\Omega, \mathcal{H})$,
- linear operator $H : \mathcal{H} \to \mathbb{R}^k$.
- hence **finite-dimensional observations** y_{τ_j}, and $\eta_j \overset{iid}{\sim} N(0, \Gamma)$ with $\eta_j \perp \{u_{\tau_j}\}$, and $\Gamma \in \mathbb{R}^{k \times k}$.

Objective: Approximate pdf of $u_{\tau_j} | Y_{\tau_{1:j}} = y_{\tau_{1:j}}$.

Forward model approximation: $\Psi \approx \Psi^{N_t, N_x}$ with discretization both time and space. Something like $\Delta t = \Delta\tau / N_t$ and $\Delta x = \mathcal{O}(N_x^{-1})$.

Numerical example, 1D SPDE

1D stochastic reaction-diffusion equation

$$\begin{aligned} \mathrm{d}u &= ((\Delta - I)u + f(u))\,\mathrm{d}t + \mathrm{d}W \qquad (t,x) \in [0,\infty) \times (0,1), \\ u(0,x) &= 4(x-1/2)^2 \\ u(t,0) &= u(t,1), \qquad \forall t \in [0,\infty). \end{aligned}$$

- operator A is spectral decomposition of $\Delta - I$

$$A = \sum_{j=1}^{\infty} \lambda_j \phi_j \otimes \phi_j, \quad \text{with} \quad \lambda_j \eqsim -j^2$$

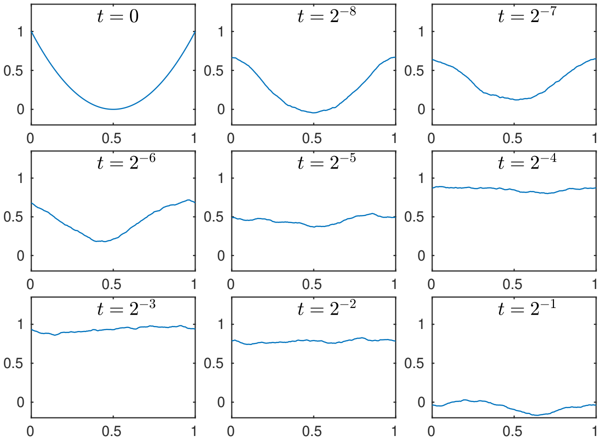
  where $\phi_j(x)$ are Fourier series functions $\{1, \sin(2\pi x), \cos(2\pi x), \ldots\}$.
- With space-time colored noise

$$W(t,x) = \sum_{j \in \mathbb{N}} j^{-1} W^{(j)}(t) \phi_j(x)$$

- consider mild solutions

$$u_{\tau_{j+1}} = \Psi(u_{\tau_j}) := e^{A\Delta\tau} u_{\tau_j} + \int_0^{\Delta\tau} e^{A(T-s)} f(u_{\tau_j+s}) ds + \int_0^{\Delta\tau} e^{A(t-s)} dW_{\tau_j+s}$$

Simulation of SPDE



Simulation of the SPDE with reaction term $f(u) = \sin(\pi u)$ over one observation-time interval $\Delta\tau = 1/2$, by Ψ^{N_t, N_x} with $N_x = N_t = 2^{12}$. See arxiv preprint A. Chernov et al. “Multilevel ensemble Kalman filtering for spatio-temporal processes” for more details.

EnKF filtering in high-dimensions

Approximate $\Psi \approx \Psi^{N_t,N_x}$, with elements $u_j \in \mathcal{H}^{N_x} \subset \mathcal{H}$ (let us here assume $\mathcal{H}^{N_x}$ is an N_x-dimensional state-space).

Sample iid $v_0^{(i)} \sim \text{Projection}_{\mathcal{H}^{N_x}} \mathbb{P}_{u_0}$ for $i = 1, 2, \ldots, M$ and (using the shorthand $v_j^{(i)} := v_{\tau_j}^{(i)}$ below)

Prediction

$$\hat{v}_{j+1}^{(i)} = \Psi^{N_t,N_x}(v_j^{(i)}), \quad i = 1, 2, \ldots, M, \qquad \text{and}$$

$$\hat{C}_{j+1} = \underbrace{\mathrm{Cov}_M[\hat{v}_{j+1}^{(\cdot)}]}_{\in \mathbb{R}^{N_x \times N_x}}, \quad \text{or rather} \quad \hat{C}_{j+1} H^* = \underbrace{\mathrm{Cov}_M[\hat{v}_{j+1}^{(\cdot)}, \hat{v}_{j+1}^{(\cdot)}]}_{\in \mathbb{R}^{N_x \times k}},$$

Analysis

$$v_{j+1}^{(i)} = \hat{v}_{j+1}^{(i)} + K(y_{j+1}^{(i)} - H\hat{v}_{j+1}^{(i)}), \quad \text{where} \quad K = \hat{C}_{j+1} H^* (H \hat{C}_{j+1} H^* + \Gamma)^{-1}.$$

Localization for EnKF

- When state-space dimension $N_x \gg 1$, the sample-covariance

$$\hat{C}_j^M = \mathrm{Cov}_M[\hat{v}_j^{(\cdot)}],$$

tend to have "spurious correlations", meaning that for some ℓ, m

$$|\hat{C}_{j,\ell,m}^M - \hat{C}_{j,\ell,m}^\infty| \gg \hat{C}_{j,\ell,m}^\infty, \quad \text{with} \quad \hat{C}_j^\infty \text{ denoting the reference cov.}$$

- Not a problem that can be easily solved through increasing M, as there is a cost associated to that.
- Two diametrically opposite approaches for dealing with this (cf. Reich and Cotter chp 8.2-8.3)
  - **Variance inflation:** To increase the magnitude of components of $\hat{C}_{j,\ell,m}^M$ where you expect it to be underestimated, inflate for instance the model uncertainty.
  - **Covariance localization:** To reduce the magnitude of components of $\hat{C}_{j,\ell,m}^M$ where you expect it to be overestimated, do for instance

    $$\text{spatial/spectral localiazation: } \textbf{replace} \quad \hat{C}_j^M \quad \textbf{by} \quad \rho \circ \hat{C}_j^M$$

    where $\circ$ is the element-wise product and $\rho_{\ell m} = \mathbb{1}_{|\ell - m| \leq n}$ is a banded filter matrix.

Summary

- Have treated filtering in the continuous-time dynamics and observations setting.
- The Kalman-Bucy filter solves the linear-Gaussian filtering problem, while the Kushner-Stratonovich equation applies more generally.
- Described extension of EnKF to filtering problems with high/infinite-dimensional state-space.
- Still unclear how well particle filtering can perform in high-dimensional filtering. The number of particles required to avoid degeneracy is conjectured to typically scale exponentially with state-space dimension.
- Next time: Presentations by Dmitry Kabanov on Ensemble Kalman Inversion applied to machine learning, and by Luis Espath on Bayesian optimal experimental design.