Mathematics and numerics for data assimilation and state estimation – Lecture 15

Summer semester 2020

Overview

1. 3DVAR
2. Extended Kalman filtering

On the course's oral exam

- Time and place: between 10:00 and 18:00 on July 31, Kackertstrasse 9, room C301,
- Preparation: Will give you a list of 20-25 topics for you to prepare on on July 17.
- The exam: Will randomly draw 5 topics from list which you will be asked to expand upon.
- Duration: Roughly 20 minutes.

Information on the student presentation

- Presentations planned on Thursday 02.07 and Friday 03.07.
- Structure: Roughly 20 minutes presentation, most likely over Zoom, either alone or in pairs.
- Please email me before 21.06 with information on:
  1. What paper/topic you would like to present
  2. your preferred time for presenting
  3. and possibly, whom you'd like to present together with.
- I will try to avoid multiple presentations on the same topic, so email me early if you have found an interesting topic.

Summary lecture 14 and plan for today

- For a linear-Gaussian filtering problem

$$V_{j+1} = \Psi(V_j) + \xi_j,$$
$$Y_{j+1} = h(V_{j+1}) + \eta_{j+1}, \quad j = 1, 2, \dots,$$

  we described iterative formulas for the pdf of $V_n | Y_{1:n} = y_{1:n}$.

- Plan for today: develop Approximate Gaussian filters for settings where Ψ is nonlinear.

Summary of Kalman filtering

For linear-Gaussian dynamics

$$\begin{aligned} V_{j+1} &= AV_j + \xi_j, \quad j = 0, 1, \ldots \\ V_0 &\sim N(m_0, C_0) \end{aligned} \tag{1}$$

with $\xi_j \overset{iid}{\sim} N(0, \Sigma)$.

Observations:

$$Y_j = HV_j + \eta_j, \quad j = 1, 2, \ldots, \tag{2}$$

with $\eta_j \overset{iid}{\sim} N(0, \Gamma)$.

And independence assumptions:

$$\{\eta_j\} \perp \{\xi_j\} \perp \{V_0\}$$

We derived the ...

Kalman filtering algorithm

Given any sequence $y_1, y_2, \ldots$ and $V_j | Y_{1:j} = y_{1:j} \sim N(m_j, C_j)$ the next-time filtering distributions are iteratively determined by

Prediction step

$$\begin{aligned}\hat{m}_{j+1} &= A m_j \\ \hat{C}_{j+1} &= A C_j A^T + \Sigma\end{aligned}$$

Analysis step

$$\begin{aligned}K_{j+1} &= \hat{C}_{j+1} H^T (H \hat{C}_{j+1} H^T + \Gamma)^{-1} \qquad \textbf{Kalman gain} \\ m_{j+1} &= (I - K_{j+1} H) \hat{m}_{j+1} + K_{j+1} y_{j+1} \\ C_{j+1} &= (I - K_{j+1} H) \hat{C}_{j+1}\end{aligned}$$

Alternative Bayesian view of Kalman filtering

In Lecture 14, using that

$$V_{j+1}|Y_{1:j} = y_{1:j} \sim N(\hat{m}_{j+1}, \hat{C}_{j+1})$$

and

$$Y_{j+1}|V_{j+1} = v_{j+1} \sim N(h(v_{j+1}), \Gamma)$$

the analysis step of Kalman filtering was derived throught the posterior

$$\begin{aligned} \pi(v_{j+1}|y_{1:j+1}) &\propto \pi(y_{j+1}|v_{j+1})\pi(v_{j+1}|y_{1:j}) \\ &\propto \exp\Big(-\frac{1}{2}|y_{j+1} - Hv_{j+1}|_{\Gamma}^2 - \frac{1}{2}|v_{j+1} - \hat{m}_{j+1}|_{\hat{C}_{j+1}}^2 \Big). \end{aligned} \tag{3}$$

Viewing the minus log-posterior as a cost/objective function,

$$\mathsf{J}(u) := \frac{1}{2}|y_{j+1} - Hu|_{\Gamma}^2 + \frac{1}{2}|u - \hat{m}_{j+1}|_{\hat{C}_{j+1}}^2$$

the analysis mean can be derived through variational principles:

$$m_{j+1} = \arg\min_{u \in \mathbb{R}^d} \mathsf{J}(u).$$

Kalman filter evolution of mean

In other words, the evolution of $m_j \mapsto m_{j+1}$ in Kalman filtering can be described by

$$\begin{aligned} \hat{m}_{j+1} &= \Psi(m_j) \\ \mathsf{J}(u) &:= \frac{1}{2}|y_{j+1} - Hu|_\Gamma^2 + \frac{1}{2}|u - \hat{m}_{j+1}|_{\hat{C}_{j+1}}^2 \\ m_{j+1} &= \arg\min_{u \in \mathbb{R}^d} \mathsf{J}(u). \end{aligned} \tag{4}$$

implicitly depending on $\hat{C}_{j+1}$ and y_{j+1}.

Equation (4) will be the basis for motivating three approximate Gaussian filtering algorithms.

Overview

1 3DVAR

2 Extended Kalman filtering

Filtering setting

Dynamics: Initial condition $V_0 \sim N(m_0, C_0)$ and for $j = 0, 1, \ldots$

$$\begin{aligned} V_{j+1} &= \Psi(V_j) + \xi_j, \\ Y_{j+1} &= HV_{j+1} + \eta_{j+1}, \quad j = 0, 1, \ldots \end{aligned} \tag{5}$$

with

$$\xi_j \overset{iid}{\sim} N(0, \Sigma), \quad \eta_j \overset{iid}{\sim} N(0, \Gamma) \quad \text{and} \quad \{\eta_j\} \perp \{\xi_j\} \perp \{V_0\}.$$

3DVAR: Fix the prediction covariance $\hat{C}_{j+1} := \hat{C}$ for all $j \geq 0$, and apply variational principle

$$\begin{aligned} \hat{m}_{j+1} &= \Psi(m_j) \\ \mathrm{J}(u) &:= \frac{1}{2}|y_{j+1} - Hu|_{\Gamma}^2 + \frac{1}{2}|u - \hat{m}_{j+1}|_{\hat{C}}^2 \\ m_{j+1} &= \arg\min_{u \in \mathbb{R}^d} \mathrm{J}(u). \end{aligned} \tag{6}$$

3DVAR

Alternatively, we may write,

$$\begin{aligned} \hat{m}_{j+1} &= \Psi(m_j) \\ K &= \hat{C}H^T(H\hat{C}H^T + \Gamma)^{-1} \\ m_{j+1} &= (I - KH)\hat{m}_{j+1} + Ky_{j+1} \end{aligned} \tag{7}$$

Properties:

- The gain K is time-independent!
- 3D – model physical space is typically three dimensional (v_n being a discretized representation of the state over 3D physical space, e.g. pressure, temperature, wind direction).
- VAR – method is derived from variational principle over 3D physical space.
- In numerical weather prediction, typcially $d \geq 10^6$, and "sparsification" from the true $\hat{C}_j$ to $\hat{C}$ is needed to construct a feasible filtering method.
- Gaussian approximation: $V_{j+1}|Y_{1:j} = y_{1:j} \sim N(\hat{m}_{j+1}, \hat{C})$ and $V_{j+1}|Y_{1:j+1} = y_{1:j+1} \sim N(m_{j+1}, (I - KH)\hat{C})$, with poor tracking of the covariance.

Example

Dynamics:

$$\begin{aligned} V_{j+1} &= 2.5\sin(V_j) + \xi_j \\ V_0 &\sim N(0,1) \end{aligned} \tag{8}$$

where $\xi_j \sim N(0, 0.09)$

Observations:

$$Y_j = V_j + \eta_j, \quad j = 1, 2, \dots,$$

with $\eta_j \sim N(0,1)$.

3DVAR: We have that $\Psi(v) = 2.5\sin(v)$, $H = 1$ and $\Gamma = 1$.

1. Fix $\hat{C} = 2$, for example, and off-line/pre compute

$$K = \hat{C}H^T(H\hat{C}H^T + \Gamma)^{-1} = \frac{2}{3}.$$

2. iterate $m_j \mapsto m_{j+1}$.

A guess for $\hat{C}$ may be motivated from Kalman filtering:

$$\hat{C}_{j+1} = A^T C_j A + \Sigma = A^T(I - KH)\hat{C}_j A + \Sigma, \quad AA^T \approx |\Psi'(v)|^2 \approx (2.5)^2/2?$$

Test

for one observation sequence $y_{1:J} = v^\dagger_{1:J} + \eta_{1:J}$ generated from synthetic data $v^\dagger_{1:J}$.

Error measure: MSE approximating the "truth" for different values of $\hat{C}$.

$$\frac{1}{J+1}\sum_{k=0}^{J} |v_k^\dagger - m_k|^2$$

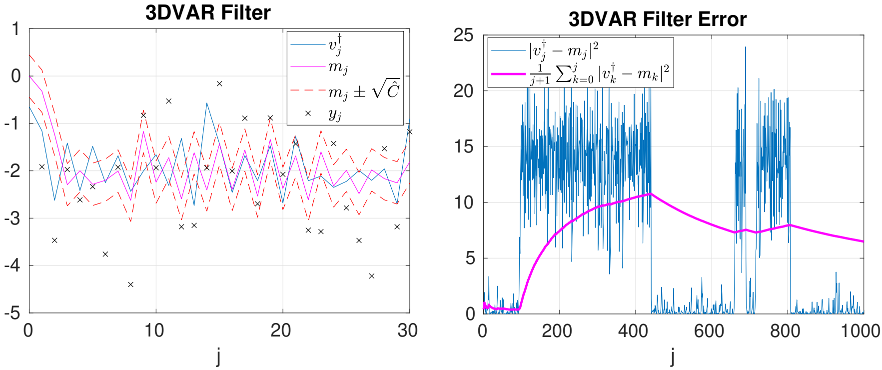
Implementation: The 3DVAR iteration

$$\begin{aligned} K &= \hat{C}H^T(H\hat{C}H^T + \Gamma)^{-1} \\ \hat{m}_{j+1} &= \Psi(m_j) \\ m_{j+1} &= (I - KH)\hat{m}_{j+1} + Ky_{j+1} \end{aligned} \tag{9}$$

becomes

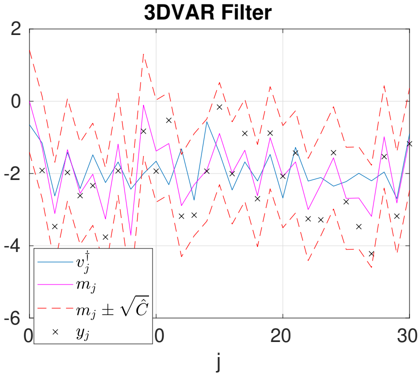
```
K=(cHat*H')/(H*cHat*H'+gamma^2);
for j=1:J
    mHat=2.5*sin(m(j));   m(j+1)=(1 - K*H)*mHat+K*y(j+1);
end
```

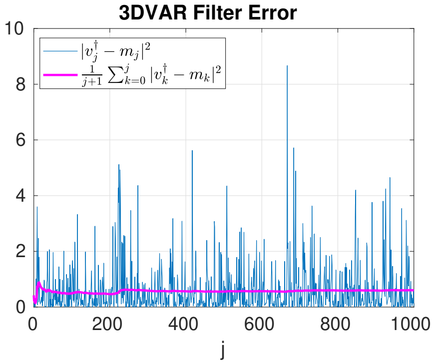
Test with $\hat{C} = 0.2$



$$\frac{1}{J+1}\sum_{k=0}^{J}|v_k^\dagger - m_k|^2 \approx 6.4866$$

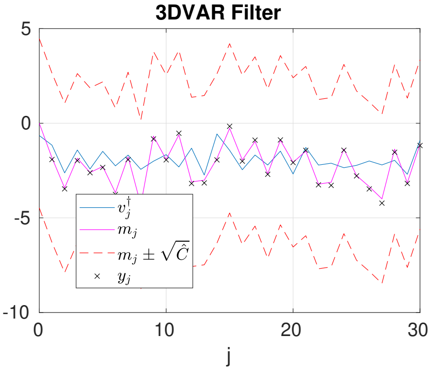
Numerical test, $\hat{C} = 2$

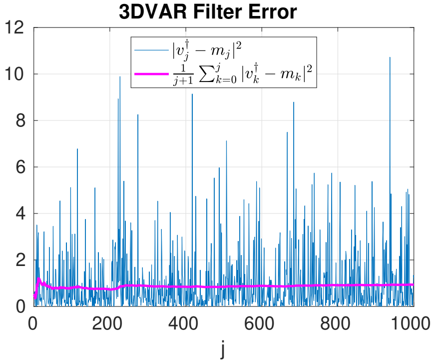




$$\frac{1}{J+1}\sum_{k=0}^{J} |v_k^\dagger - m_k|^2 \approx 0.6023$$

Test with $\hat{C} = 20$





$$\frac{1}{J+1}\sum_{k=0}^{J} |v_k^\dagger - m_k|^2 \approx 0.9373$$

Illustration of high dimensional filtering problem

Weather prediction[1]: for $(t, x) \in [0, T] \times \mathbb{R}^3$,

$$\frac{d\mathbf{v}}{dt} = -\alpha \nabla p - \nabla \phi + \mathbf{F} - 2\Omega \times \mathbf{v} \qquad \text{Cons. momentum}$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}) \qquad \text{Cons. mass}$$

$$p/\rho = RT \qquad \text{Eq. of state}$$

$$Q = C_p \frac{dT}{dt} - \rho^{-1} \frac{dp}{dt} \qquad \text{Cons. energy}$$

$$\frac{\partial \rho q}{\partial t} = -\nabla \cdot (\rho \mathbf{v} q) + \rho(E - C) \qquad \text{Cons. water vapor mixing ratio}$$

$\mathbf{v}(t, x)$ - wind velocity field, $\rho(t, x)$ - air density, p - pressure, T - temperature, q - vapor mixing ratio.

Observations:

$$Y(t_{n+1}) = h(v, \rho, p, T, q)(t_{n+1}) + \eta_{n+1}.$$

[1] E. Kalnay, Atmospheric data assimilation and applications.

Rough idea of numerical weather prediction

Introduce a mesh

$\mathcal{I} = \{(x_i, x_j, x_k) \in \mathbb{R}^3 \mid (x_i, x_j, x_k) \text{ is a point in (a subset of) the atmosphere}\}$

3DVAR prediction: Numerical solution of the weather model with filtering **conditional mean** $m_j \approx \mathbb{E}\left[\{(v, \rho, p, T, q)(t_j, x)\}_{x\in\mathcal{I}} \mid Y_{1:j} = y_{1:j}\right]$ as initial condition. That is,

$$m_j \overset{\bar{\Psi}(m_n)}{\mapsto} \hat{m}_{j+1} \approx \mathbb{E}\left[\{(v, \rho, p, T, q)(t_{j+1}, x)\}_{x\in\mathcal{I}} \mid Y_{1:j} = y_{1:j}\right].$$

Note: State-space dimension $d = |\mathcal{I}| \times 7$.

Analysis: Apply 3DVAR principle with a typically low-bandwith, fixed $\hat{C} \approx \mathrm{Cov}[\{(v, \rho, p, T, q)(t_{j+1}, x)\}_{\mathbf{x}\in\mathcal{I}} | Y_{1:j} = y_{1:j}]$,

$$m_{j+1} = (I - KH)\hat{m}_{j+1} + Ky_{j+1}.$$

Recovery of true signal by 3DVAR

Theorem 1 (LSZ 4.10)

Assume the true signal is given by

$$v_{j+1}^{\dagger} = \Psi(v_j^{\dagger})$$

and observations by

$$y_{j+1} = Hv_{j+1}^{\dagger} + \epsilon_{j+1}, \quad \textit{with} \sup_{j\geq 0} \|\epsilon_j\| \leq \epsilon.$$

For 3DVAR with any value of $m_0 \in \mathbb{R}^d$, if $\hat{C}$ is chosen such that it holds for all $u, v \in \mathbb{R}^d$ and some $a < 1$ that

$$\|(I - KH)\Psi(u) - (I - KH)\Psi(v)\| \leq a\|u - v\|,$$

then

$$\limsup_{j\geq 0} \|v_j^{\dagger} - m_j\| \leq \frac{\|K\|}{1-a}\epsilon.$$

Proof idea:

Write

$$m_{j+1} = (I - KH)\Psi(m_j) + K\underbrace{\left(H\Psi(v_j^\dagger) + \epsilon_{j+1}\right)}_{y_{j+1}}$$

$$v_{j+1}^\dagger = (I - KH)\Psi(v_j^\dagger) + KH\Psi(v_j^\dagger).$$

Then for

$$\begin{aligned}
\|m_{j+1} - v_{j+1}^\dagger\| &\leq \|(I - KH)\Psi(m_j) - (I - KH)\Psi(v_j^\dagger)\| + \|K\epsilon_{j+1}\| \\
&\leq a\|m_j - v_j^\dagger\| + \|K\|\|\epsilon_{j+1}\| \\
&\leq a\|m_j - v_j^\dagger\| + \|K\|\epsilon \\
&\leq \ldots \leq a^{j+1}\|m_0 - v_0^\dagger\| + \|K\|\epsilon\sum_{k=0}^{j} a^k
\end{aligned}$$

and $a^{j+1}\|m_0 - v_0^\dagger\| \to 0$ as $j \to \infty$.

Remarks on Theorem 1

- Note that the asymptotic tracking ability holds **regardless of the magnitude of** $\|m_0 - v_0^\dagger\|$ as long as $a < 1$.
- Not that interesting result if $H = I$, since if one were to choose the filtering approach of total reliance on observations: $m_j = y_j$, then one would anyway achieve

$$\|v_j^\dagger - m_j\| = \|\epsilon_j\| \leq \epsilon.$$

- Relevant in **partial observation** settings $H \in \mathbb{R}^{k \times d}$ with $k < d$. Then it shows that accurate observations of unstable components may lead to good tracking of the state of all components.
- (SST Theorem 9.2) extends result from deterministic upper bound on noise error $|\epsilon_j| < \epsilon$ to Gaussian random noise setting $y_j = Hv_j^\dagger + \epsilon_j$ with $\epsilon_j \sim N(0, \gamma^2 I)$, and

$$\limsup_{j \to \infty} \mathbb{E}\left[\|m_j - v_j^\dagger\| \right] \leq \frac{\|K\|}{1 - a} \gamma,$$

Choice of $\hat{C}$ guided by the preceding result.

- 3DVAR applied to a filtering problem with fixed $H = I$ and $\Gamma = \gamma^2 I$, and $\hat{C} = \sigma^2 I$ with adjustable parameter σ^2 yields

$$K = \frac{\gamma^2}{\sigma^2 + \gamma^2} I \quad \text{and} \quad (I - KH)\Psi(v) = \frac{(\gamma/\sigma)^2}{1 + (\gamma/\sigma)^2} \Psi(v)$$

- Choosing σ^2 so large that

$$\frac{(\gamma/\sigma)^2}{1 + (\gamma/\sigma)^2} \|D\Psi(v)\| < 1 \quad \forall v \in \mathbb{R}^d$$

  will lead stability in the form Theorem 1 (when other assumptions hold).
- In the example with $\Psi(v) = 2.5 \sin(v)$ and $\gamma^2 = 1$,

$$\frac{(\gamma/\sigma)^2}{1 + (\gamma/\sigma)^2} \|D\Psi\|_\infty < 1 \iff \frac{2.5}{\sigma^2 + 1} < 1 \iff \sigma^2 > 1.5.$$

- Interpretation: model variance inflation of σ^2 may help ensure stability of tracking (effectively it means putting more trust on observations).

Tracking of truth under partial observations

Consider now **partial** observations $H = (I_k, 0)^T \in \mathbb{R}^{k\times d}$, fixed $\Gamma = \gamma^2 I_k$ and $\hat{C} = \sigma^2 I_d$. Then

$$I_d - KH = \begin{bmatrix} \frac{\eta^2}{1+\eta^2} I_k & 0 \\ 0 & I_{d-k} \end{bmatrix}$$

with $\eta = \gamma/\sigma$.

For a linear dynamics mapping $\Psi(u) = Lu$ with

$$D\Psi = L = \begin{bmatrix} b_1 I_k & 0 \\ 0 & b_2 I_{d-k} \end{bmatrix}$$

we obtain

$$(I_d - KH)D\Psi = \begin{bmatrix} \frac{b_1\eta^2}{1+\eta^2} I_k & 0 \\ 0 & b_2 I_{d-k} \end{bmatrix}$$

Conclusion: $\|(I_d - KH)D\Psi\| < 1$ is only possible to achieve when $|b_2| < 1$. (This is a stability condition in dynamics of unobserved components.) Whatever the magnitude of $|b_1|$, on the other hand, this can be controlled by appropriately inflating σ^2.

4DVAR

Is an extension of 3DVAR in the analyisis step applying the variational principle over both 3D space and **time** (i.e., allowing for measurements scattered also over a time window)

Given dynamics:

$$V_{j+1} = \Psi(V_j) + \xi_j,$$

with $\xi_j \overset{iid}{\sim} N(0, \Sigma)$ and observations $y_{1:J}$ as before,

w4DVAR weak constraint 4DVAR is for stochastic dynamics $\Sigma > 0$. Then assimilation is done over the time window $0 : J$ in one step:

$$m_{0:J} = \arg \min_{v_{0:J} \in \mathbb{R}^{d(J+1)}} \frac{1}{2} |v_0 - m_0|^2_{C_0} + \frac{1}{2} \sum_{j=0}^{J-1} |v_{j+1} - \Psi(v_j)|^2_{\Sigma} + \frac{1}{2} \sum_{j=1}^{J} |y_j - Hv_j|^2_{\Gamma}$$

If Ψ is bounded and continuous, then a minimizer $m_{0:J}$ exists and corresponds to a MAP estimator for the very same smoothing problem over the same time-window [SST 9.3].

4DVAR is for settings with deterministic dynamics, i.e., $\Sigma = 0$, when w4DVAR turns into a minimization problem

$$m_{0:J} = \arg\min_{v_{0:J}} \frac{1}{2}|v_0 - m_0|^2_{C_0} + \frac{1}{2}\sum_{j=1}^{J} |y_j - Hv_j|^2_{\Gamma}$$

subject to the strong constraint

$$v_{j+1} = \Psi(v_j), \quad j = 0, 1, \ldots, J-1.$$

Comparisons 4DVAR vs 3DVAR

- 4DVAR is a minimization problem in typically higher-dimensional space than 3DVAR
- Both methods originally developed for numerical weather prediction, with emphasis on an efficient method for high-dimensional state space analysis/update.
- We focus here on one, but there exist many hybrid versions of 3D/4DVAR combined with other filtering techniques for the prediction step, cf., E. Kalnay "Atmospheric modeling, data assimilation and predictability".

Overview

1 3DVAR

2 Extended Kalman filtering

Filtering setting

Initial condition $V_0 \sim N(m_0, C_0)$ and for $j = 0, 1, \ldots$

$$
\begin{aligned}
V_{j+1} &= \Psi(V_j) + \xi_j, \\
Y_{j+1} &= HV_{j+1} + \eta_{j+1},
\end{aligned}
\tag{10}
$$

and Gaussian noise assumptions as before.

Extened Kalman filtering (ExKF): At time j and given state (m_j, C_j), linearize dynamics around m_j:

$$\Psi_L(v; m_j) := \Psi(m_j) + D\Psi(m_j)(v - m_j).$$

And apply Kalman filtering one prediction-update step to the linearized dynamics

$$V_{j+1} = \Psi(m_j) + D\Psi(m_j)(V_j - m_j) + \xi_j,$$

Extended Kalman filtering algorithm

Given any sequence $y_1, y_2, \ldots$ and $V_j | Y_{1:j} = y_{1:j} \sim N(m_j, C_j)$,

Prediction step

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

Analysis step

$$K_{j+1} = \hat{C}_{j+1} H^T (H \hat{C}_{j+1} H^T + \Gamma)^{-1}$$
$$m_{j+1} = (I - K_{j+1} H) \hat{m}_{j+1} + K_{j+1} y_{j+1}$$
$$C_{j+1} = (I - K_{j+1} H) \hat{C}_{j+1}$$

Motiation for prediction step: We have the following approximations:

$$m_j \approx \mathbb{E}\left[V_j | Y_{1:j} = y_{1:j} \right], \quad C_j \approx \mathbb{E}\left[(V_j - m_j)(V_j - m_j)^T | Y_{1:j} = y_{1:j} \right]$$

Note further that the ExKF moments m_j and C_j are **not random** (given $y_{1:j}$).

Motivation for the ExKF algorihtm

Using that $\Psi(m_j)$ and $D\Psi(m_j)$ are deterministic (given $y_{1:j}$), the following approimations motivate the ExKF prediciton steps:

$$\begin{aligned}
\hat{m}_{j+1} &= \mathbb{E}\left[\, \Psi(m_j) + D\Psi(m_j)(V_j - m_j) + \xi_j | Y_{1:j} = y_{1:j}\right] \\
&= \Psi(m_j) + D\Psi(m_j)\Big(\mathbb{E}\left[\, V_j | Y_{1:j} = y_{1:j}\right] - m_j\Big) \\
&\approx \Psi(m_j)
\end{aligned}$$

and (similar derivation as for Kalman filtering with $A = D\Psi(m_j)$),

$$\begin{aligned}
\hat{C}_{j+1} &= \mathrm{Cov}[\Psi(m_j) + D\Psi(m_j)(V_j - m_j) + \xi_j | Y_{1:j} = y_{1:j}] \\
&= \mathrm{Cov}[D\Psi(m_j)(V_j - m_j) + \xi_j | Y_{1:j} = y_{1:j}] \\
&= D\Psi(m_j)\mathbb{E}\left[(V_j - m_j)(V_j - m_j)^T | Y_{1:j} = y_{1:j}\right] D\Psi(m_j)^T + \Sigma \\
&\approx D\Psi(m_j) C_j D\Psi(m_j)^T + \Sigma.
\end{aligned}$$

Example

Dynamics:

$$\begin{aligned} V_{j+1} &= 2.5\sin(V_j) + \xi_j \\ V_0 &\sim N(0,1) \end{aligned} \tag{11}$$

where $\xi_j \sim N(0, 0.09)$ **Observations:**

$$Y_j = V_j + \eta_j, \quad j = 1, 2, \ldots,$$

with $\eta_j \sim N(0,1)$.

ExKF: linearized dynamics mapping becomes

$$\Psi_L(v; m_j) = 2.5\sin(m_j) + 2.5\cos(m_j)(v - m_j),$$

Starting with $(m_0, C_0) = (0, 1)$ apply linearized mapping $\Psi_L(v; 0)$ and Kalman filtering to transition $(m_0, C_0) \mapsto (m_1, C_1)$, apply linearized mapping $\Psi_L(v; m_1)$ to and KF to transition $(m_1, C_1) \mapsto (m_2, C_2)$ ...

Test

for one observation sequence $y_{1:J} = v^{\dagger}_{1:J} + \eta_{1:J}$ generated from synthetic data $v^{\dagger}_{1:J}$.

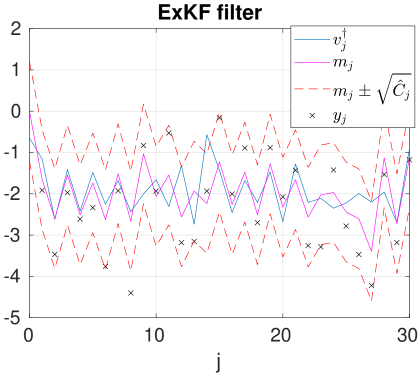
Implementation: The ExKF given m_j and C_j:

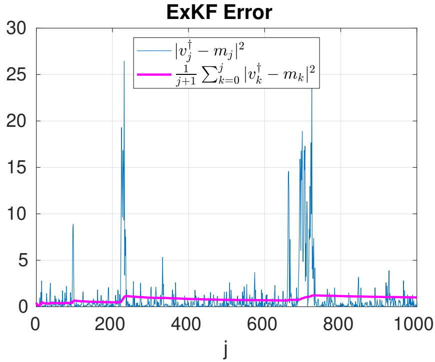
```
Psi  = @(v) 2.5*sin(v); %Dynamics mapping
DPsi = @(v) 2.5*cos(v); %Jacobian

for j=1:J

    %ExKF filtering
    mHat = Psi(m(j));
    cHat = DPsi(m(j))*C(j)*DPsi(m(j))' + Sigma;
    K      = (cHat*H')/(H*cHat*H'+Gamma);
    m(j+1) = (1-K*H)*mHat+K*y(j);
    C(j+1) = (1-K*H)*cHat;

end
```



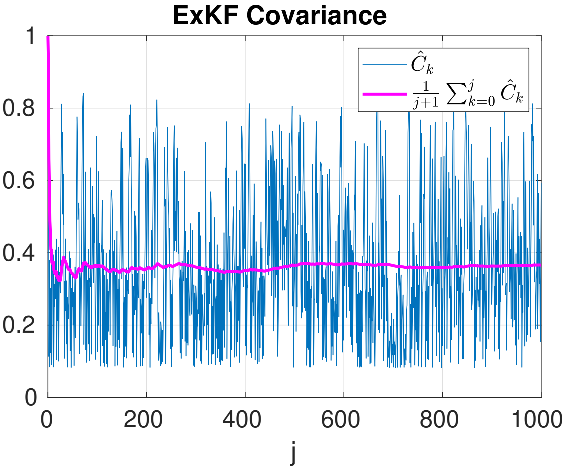


$$\frac{1}{1001}\sum_{k=0}^{1000} |v_k^\dagger - m_k|^2 \approx 0.9969 \quad \text{and} \quad \frac{1}{10001}\sum_{k=0}^{10000} |v_k^\dagger - m_k|^2 \approx 0.6169.$$

(MSE ≈ 0.9969 is not very impressive, this is roughly same error as one would get for

$$\frac{1}{J}\sum_{k=1}^{J} |v_k^\dagger - y_k|^2, \quad \text{since} \quad y_k = v_k^\dagger + \eta_k \text{ and } \eta_k \sim N(0,1).)$$

For comparison with the 3DVAR fixed prediction covariance $\hat{C}$, plot of evolution of $\hat{C}_j$ for ExKF:



Remarks on errors of ExKF and 3DVAR

- It generally does hold that

$$\mathbb{E}[\Psi(V)+\xi] = \Psi(\mathbb{E}[V]) \implies \hat{m}_j = \Psi(m_j) \overset{\textit{in general}}{\neq} \mathbb{E}[\Psi(V_j)|Y_{1:j} = y_{1:j}].$$

- Nor does it generally hold that $V_j|Y_{1:j} = y_{1:j}$ is Gaussian when Ψ is nonlinear, and the analysis step, being derived under the assumption of Gaussian posterior

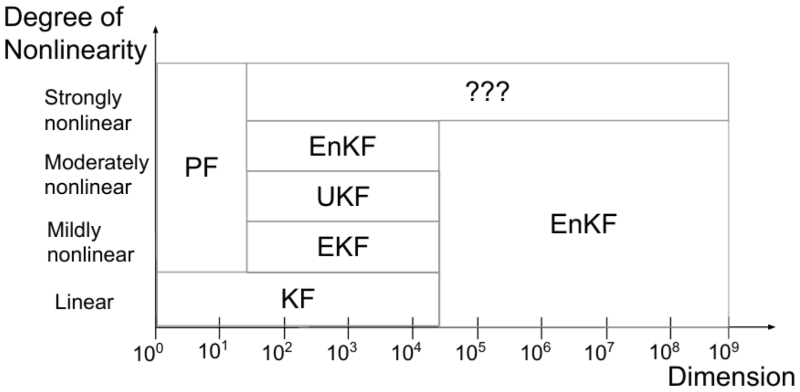
$$\pi(v_j|y_{1:j}) \propto \exp\Big(-\frac{1}{2}|y_{j+1} - Hv_{j+1}|_\Gamma^2 - \frac{1}{2}|v_{j+1} - \hat{m}_{j+1}|_{\hat{C}_{j+1}}^2\Big).$$

which, may only approximately hold, and the consecutive variational principle

$$m_{j+1} = \arg\min_{u\in\mathbb{R}^d} \frac{1}{2}|y_{j+1} - Hu|_\Gamma^2 + \frac{1}{2}|u - \hat{m}_{j+1}|_{\hat{C}_{j+1}}^2$$

is thus also only an approximation.

Best filtering method measured in terms of accuracy and efficiency



KF = Kalman filter; PF = particle filter; EKF = extended KF; UKF = unscented KF; EnKF = ensemble KF

Figure from talk by Mattias Katzfuss on "Extended ensemble Kalman filters for high-dimensional hierarchical state-space models".

Plan for next lecture

- Implementation and convergence properties of EnKF in large ensemble limit.
- Particle filtering.