Mathematics and numerics for data assimilation and state estimation – Lecture 16

Summer semester 2020

Overview

1. Extended Kalman filtering
2. Ensemble Kalman filtering
3. Approximation errors for Gaussian-based nonlinear filter methods
4. Efficient implementation of EnKF and extensions to nonlinear observations

Summary lecture 15 and plan for today

- Described two approximate filtering methods for the nonlinear problem

$$V_{j+1} = \Psi(V_j) + \xi_j, \qquad \xi_j \overset{iid}{\sim} N(0, \Sigma)$$

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

i.e., 3DVAR and Extended Kalman filtering.

Plan for today:

- More on Extended Kalman filtering
- Approximation error and study of why the filter distribution typically is non-Gaussian when Ψ is nonlinear
- The Ensemble Kalman filtering method.
- EnKF applied to nonlinear observations.

Key variational princple for extenstions of Kalman filtering

We recall that for Kalman filtering, we have the posterior

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

which implies that the filtering iteration $m_j \mapsto m_{j+1}$ can be described by the variational principle

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

3DVAR

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

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

...which by the derivations for Kalman filtering yield

$$\begin{aligned} \hat{m}_{j+1} &= \Psi(m_j) \\ K &= \hat{C}H^T(H\hat{C}H^T + \Gamma)^{-1} \\ m_{j+1} &= (I - KH)\hat{m}_{j+1} + Ky_{j+1}. \end{aligned} \tag{3}$$

Overview

1. Extended Kalman filtering
2. Ensemble Kalman filtering
3. Approximation errors for Gaussian-based nonlinear filter methods
4. Efficient implementation of EnKF and extensions to nonlinear observations

Filtering setting

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

$$
\begin{aligned}
V_{j+1} &= \Psi(V_j) + \xi_j, \\
Y_{j+1} &= HV_{j+1} + \eta_{j+1},
\end{aligned}
\tag{4}
$$

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

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 $y_{1:j}$), we obtain the approximation

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

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

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

Remarks on errors of ExKF and 3DVAR

- It generally does hold that

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

- 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.

Overview

Ensemble Kalman filtering

We again consider the problem with $V_0 \sim N(m_0, C_0)$ and for $j = 0, 1, \ldots$

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

and Gaussian noise assumptions as before.

EnKF initial condition is ensemble of iid "particles" $v_0^{(i)} \overset{iid}{\sim} \mathbb{P}_{V_0}$ for $i = 1, 2, \ldots, M$ and whose empirical measure approximates the true initial distribution:

$$\mathbb{P}_{V_0}(dv) \approx \frac{1}{M} \sum_{i=1}^{M} \delta_{v_0^{(i)}}(dv)$$

EnKF Prediction at time $j = 1$

To approximate the prediction $\mathbb{P}_{V_1}$, all particles are simulated one step ahead:

$$\hat{v}_1^{(i)} = \Psi(v_0^{(i)}) + \xi_1^{(i)}, \quad i = 1, 2, \ldots, M$$

where $\{\xi_j^{(i)}\}$ are iid $N(0, \Sigma)$-distributed and

$$\mathbb{P}_{V_1}(dv) \approx \frac{1}{M} \sum_{i=1}^{M} \delta_{\hat{v}_1^{(i)}}(dv).$$

Sample prediction mean and covariance

$$\hat{m}_1 := \frac{1}{M} \sum_{i=1}^{M} \hat{v}_1^{(i)}, \qquad \hat{C}_1 := \frac{1}{M-1} \sum_{i=1}^{M} (\hat{v}_1^{(i)} - \hat{m}_1)(\hat{v}_1^{(i)} - \hat{m}_1)^T.$$

EnKF analysis at time $j = 1$

- The Kalman gain is computed using the $\hat{C}_1$:

$$K_1 = \hat{C}_1 H^T (H \hat{C}_1 H^T + \Gamma)^{-1}$$

- and the observation y_1 is assimilated into each particle by

$$\left.\begin{aligned} y_1^{(i)} &= y_1 + \eta_1^{(i)} & \textbf{perturbed observations} \\ v_1^{(i)} &= (I - K_1 H)\hat{v}_1^{(i)} + K_1 y_1^{(i)} & \end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, M,$$

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

- As before, the empirical measure of $\{v_1^{(i)}\}$ approximates $V_1 | Y_1 = y_1$:

$$\mathbb{P}_{V_1 | Y_1 = y_1}(dv) \approx \frac{1}{M} \sum_{i=1}^{M} \delta_{v_1^{(i)}}(dv)$$

Iterated EnKF formulas

Given any $y_1, y_2, \ldots$ and $\{v_j^{(i)}\}_{i=1}^M$, the EnKF iterations are

Prediction step

$$\hat{v}_{j+1}^{(i)} = \Psi(v_j^{(i)}) + \xi_j^{(i)}, \quad i = 1, 2, \ldots, M$$

$$\hat{C}_{j+1} = \underbrace{\frac{1}{M-1}\sum_{i=1}^{M}(\hat{v}_{j+1}^{(i)} - \hat{m}_{j+1})(\hat{v}_{j+1}^{(i)} - \hat{m}_{j+1})^T}_{=:\mathrm{Cov}_M[\hat{v}_{j+1}^{(\cdot)}]}, \qquad \hat{m}_{j+1} = \underbrace{\frac{1}{M}\sum_{i=1}^{M}\hat{v}_{j+1}^{(i)}}_{=:E_M[\hat{v}_{j+1}^{(\cdot)}]}$$

Analysis step

$$K_{j+1} = \hat{C}_{j+1}H^T(H\hat{C}_{j+1}H^T + \Gamma)^{-1}$$

and

$$\left.\begin{aligned} y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{(i)} &= (I - K_{j+1}H)\hat{v}_{j+1}^{(i)} + K_{j+1}y_{j+1}^{(i)} \end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, M,$$

Comments

- In settings when $\hat{C}_j$ is non-singular, the analysis step can be viewed as the variational principle

$$v_j^{(i)} := \arg\min_{u\in\mathbb{R}^d} \frac{1}{2}|y_j^{(i)} - Hu|_\Gamma^2 + \frac{1}{2}|u - \hat{m}_j|_{\hat{C}_j}^2$$

  (see [SST Chp 9] for an extension of this argument when $\hat{C}_j$ is singular).

- A random perturbation $\eta_j^{(i)}$ is added to the observation in the analysis step for each particle for the purpose of consistency: in the setting with linear dynamics $\Psi(v) = Av$,

$$\lim_{M\to\infty} \mathbb{E}\left[\,\hat{C}_j^{EnKF}\right] \begin{cases} < \hat{C}_j^{Kalman} & \text{without perturbed obs} \\ = \hat{C}_j^{Kalman} & \text{with perturbed obs} \end{cases}$$

  [Ubung 8].

- **It can be shown that $v_{j+1}^{(i)} \in$ Span$(\{\hat{v}_{j+1}^{(i)}\}_{i=1}^M)$: motivate**

Comments

- The EnKF empirical measure is of course an approximation of but the method has obvious advantages over other in terms of robustness and storage.

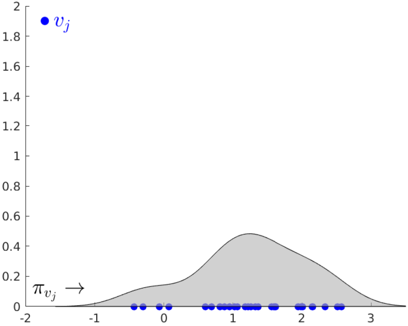
- Storage: EnKF needs to store $\mathcal{O}(M \times d)$ values $(v_j^{(1)}, \ldots, v_j^{(M)} \in \mathbb{R}^d)$. The Kalman filter needs to store $\mathcal{O}(d \times d)$ (the covariance $C_j \in \mathbb{R}^{d \times d}$).

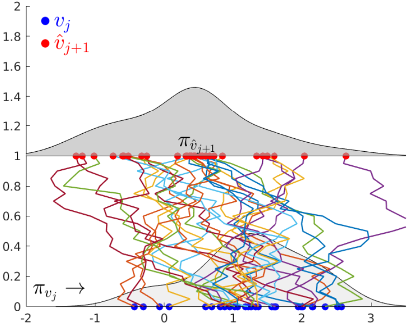
  If the true dimension of problem is much smaller than d, then EnKF is often successful in tracking the truth at a storage constraint than $d \times d$.

- EnKF is more directly applicable to nonlinear problems than ExKF, and better at handling nonlinearities than both ExKF and 3DVAR.

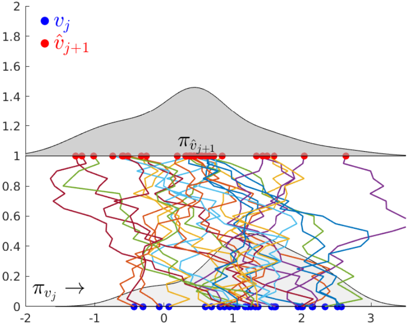
- As for other nonlinear filtering methods, $\mathbb{P}_{V_0}$ need not be Gaussian for EnKF.

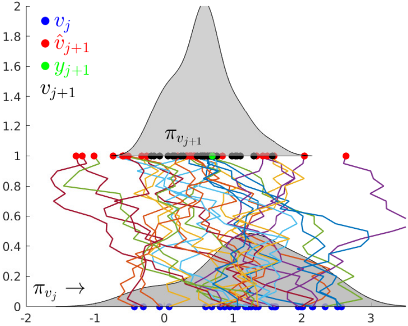
Animation of EnKF





Animation of EnKF





Example implementation of EnKF

Dynamics:

$$\begin{aligned} V_{j+1} &= 2.5\sin(V_j) + \xi_j \\ V_0 &\sim N(0,1) \end{aligned} \tag{6}$$

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

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

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

EnKF:

1. Sample iid $v_0^{(i)} \sim N(0, 1)$ for $i = 1, 2, \ldots,$
2. Simulate $\hat{v}_1^{(i)} = 2.5\sin(v_0^{(i)}) + \xi_0^{(i)}$ for $i = 1, 2, \ldots, M$.

EnKF continued

EnKF:

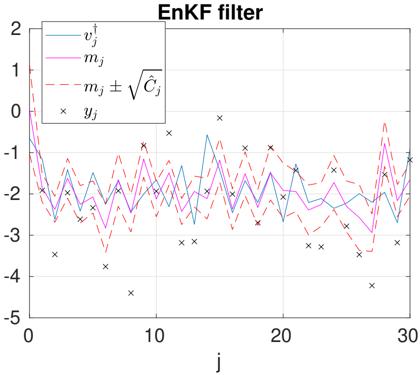
3. Compute
$$\hat{C}_1 = \mathrm{Cov}_M[\hat{v}_1^{(\cdot)}]$$
and
4. $$K_1 = \hat{C}_1 H^T (H \hat{C}_1 H^T + \Gamma)^{-1}$$
and
$$\left.\begin{aligned} y_1^{(i)} &= y_1 + \eta_1^{(i)} \\ v_1^{(i)} &= (I - K_1 H)\hat{v}_1^{(i)} + K_1 y_1^{(i)} \end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, M,$$
5. Simulate
$$\hat{v}_2^{(i)} = 2.5 \sin(v_1^{(i)}) + \xi_1^{(i)} \quad \text{for} \quad i = 1, 2, \ldots, M,$$
and so forth.

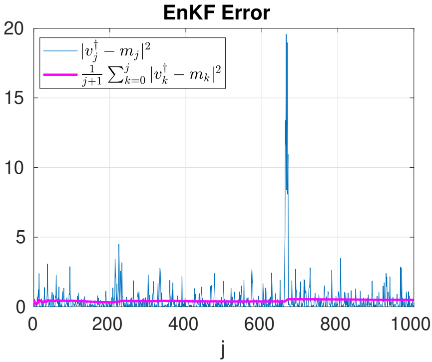
Matlab code:

```
Psi  = @(v) 2.5*sin(v);
v = m0 + sqrt(C0)*randn(M,1); %initial condition
m(1) = mean(v); C(1) = cov(v);

for j=1:J

    % EnKF filtering
    vHat        = Psi(v) + sqrt(Sigma)*randn(M,1);
    cHat        = cov(vHat);
    K           = (cHat*H')/(H*cHat*H'+Gamma);
    yPerturbed  = y(j) + sqrt(Gamma)*randn(M,1);
    v           = (1-K*H)*vHat+K*yPerturbed;

    % for plotting puropses
    m(j+1)      = mean(v);  C(j+1)= cov(v);
end
```

Numerical results EnKF for $M = 10$

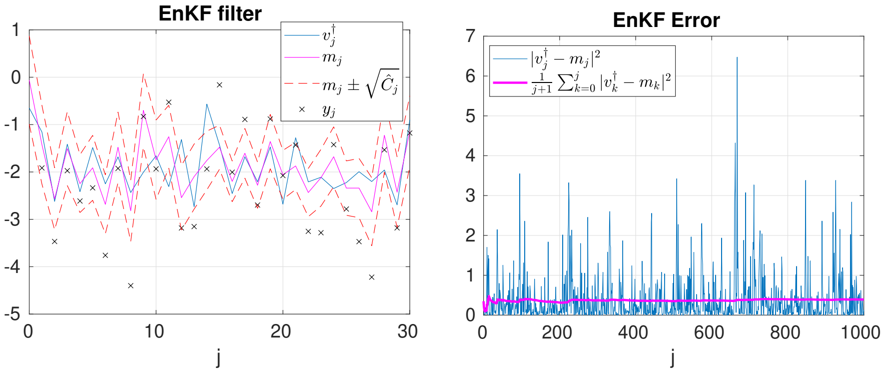
An observation sequence $y_{1:J} = v^\dagger_{1:J} + \eta_{1:J}$ is generated from synthetic data for $J = 1000$.





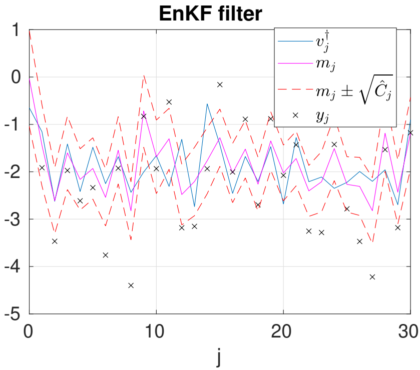
$$\frac{1}{1001}\sum_{k=0}^{1000} |v_k^\dagger - m_k|^2 \approx 0.4950 \quad \text{and}$$

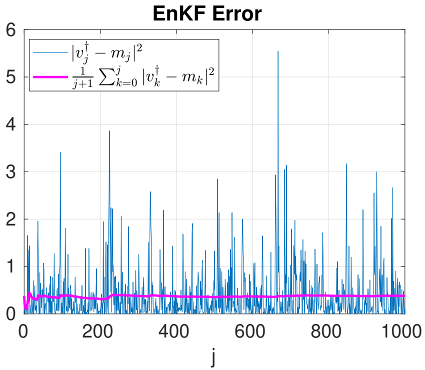
Numerical results EnKF for $M = 100$



$$\frac{1}{1001}\sum_{k=0}^{1000} |v_k^\dagger - m_k|^2 \approx 0.3902 \quad \text{and}$$

Numerical results EnKF for $M = 1000$ (very similar to $M = 100$)





$$\frac{1}{1001}\sum_{k=0}^{1000}|v_k^\dagger - m_k|^2 \approx 0.3799$$

Why does not the error converge towards 0?

Comparison of time-averaged errors

EnKF $M = (10, 100, 1000)$:

$$\frac{1}{1001}\sum_{k=0}^{1000}|v_k^\dagger - m_k|^2 \approx (0.4950, 0.3902, 0.3799),$$

ExKF

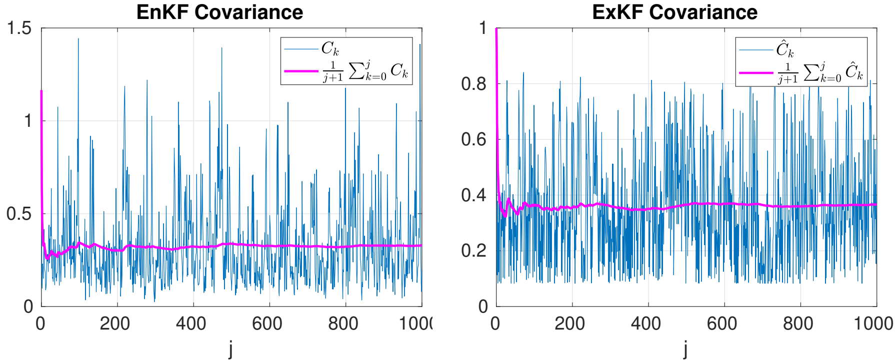
$$\frac{1}{1001}\sum_{k=0}^{1000}|v_k^\dagger - m_k|^2 = .9969$$

3DVAR (best try, with $\hat{C} = 2$)

$$\frac{1}{1001}\sum_{k=0}^{1000}|v_k^\dagger - m_k|^2 = 0.6023.$$

Comparison of covariances

EnKF with ensemble size $M = 10$



Variation in ExKF covariance relates to linearization around different points m_j in prediction step: $\hat{C}_{j+1} = D\Psi(m_j)C_jD\Psi(m_j)^T + \Sigma$

Variation in EnKF covariance relates to variations in the ensemble flow: $\hat{C}_{j+1} = \mathrm{Cov}_M[\hat{v}_j^{(\cdot)}]$.

Overview

1. Extended Kalman filtering
2. Ensemble Kalman filtering
3. Approximation errors for Gaussian-based nonlinear filter methods
4. Efficient implementation of EnKF and extensions to nonlinear observations

Exact vs approximate filtering methods

For the nonlinear filtering problem

$$V_{j+1} = \Psi(V_j) + \xi_j, \qquad \xi_j \overset{iid}{\sim} N(0, \Sigma)$$

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

with same independence assumptions as before, we derived in Lecture 14 that if we know the pdf of $V_j | Y_{1:j} = y_{1:j}$ then

Prediction step

The prediction rv $V_{j+1} | Y_{1:j} = y_{1:j}$ equals rv $\Psi(V_j) + \xi_j | Y_{1:j} = y_{1:j}$.

3DVAR: Approximated by $N(\Psi(m_j), \hat{C})$.

ExKF: Approximated by $N(\Psi(m_j), \hat{C}_j)$, linearized covariance.

EnKF: Approximated by empirical distribution of $\{\Psi(v_j^{(i)}) + \xi_j^{(i)}\}_{i=1}^M$.

Will be a good approximation asymptotically (provided $\{v_j^{(i)}\}_{i=1}^M$ is a good approximation of analysis distribution at time j).

Analysis step:

$$\pi(v_{j+1}|y_{1:j+1}) \propto \exp\Big(-\frac{1}{2}|y_{j+1} - Hv_{j+1}|_{\Gamma}^{2}\Big)\pi(v_{j+1}|y_{1:j})$$
$$\propto \pi_{N(0,\Gamma)}(y_{j+1} - Hv_{j+1})\pi(v_{j+1}|y_{1:j})$$

3DVAR and ExKF: The analysis step for these methods is, after linearization, a carbon copy of Kalman filtering. Using that $V_{j+1}|Y_{1:j} = y_{1:j} \sim N(\Psi(m_j), \hat{C}_j)$ for these methods, we have that

$$\pi(v_{j+1}|y_{1:j+1}) \propto \pi_{N(0,\Gamma)}(y_{j+1} - Hv_{j+1})\pi_{N(\Psi(m_j),\hat{C}_{j+1})}(v_{j+1})$$

(with $\hat{C}_j = \hat{C}$ for 3DVAR).

Conclusion: Approximation errors enter in prediction step for these two methods.

EnKF: Is more subtle to study as the particles correlate/mix in the analysis step. We will look at the simplified setting when $M = \infty$.

Mean-field limit

$$\text{Pr}\begin{cases}\hat{v}_{j+1}^{(i)} &= \Psi(v_j^{(i)}) + \xi_j^{(i)} \\ \hat{C}_{j+1} &= \mathrm{Cov}_M[v_{j+1}^{(\cdot)}]\end{cases} \quad \text{Anl}\begin{cases}K_{j+1} &= \hat{C}_{j+1}H^T(H\hat{C}_{j+1}H^T + \Gamma)^{-1} \\ y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{(i)} &= (I - K_{j+1}H)\hat{v}_{j+1}^{(i)} + K_{j+1}y_{j+1}^{(i)}\end{cases}$$

$M = \infty$ yields iid mean-field EnKF (MFEnKF) particles with dynamics

$$\text{Pr}\begin{cases}\hat{v}_{j+1}^{\mathrm{MF,(i)}} &= \Psi(v_j^{\mathrm{MF,(i)}}) + \xi_j^{(i)} \\ \hat{C}_{j+1}^{\mathrm{MF}} &= \mathrm{Cov}[v_{j+1}^{\mathrm{MF}}]\end{cases} \quad \text{Anl}\begin{cases}K_{j+1}^{\mathrm{MF}} &= \hat{C}_{j+1}^{\mathrm{MF}}H^T(H\hat{C}_{j+1}^{\mathrm{MF}}H^T + \Gamma)^{-1} \\ y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{\mathrm{MF},(i)} &= (I - K_{j+1}^{\mathrm{MF}}H)\hat{v}_{j+1}^{\mathrm{MF},(i)} + K_{j+1}^{\mathrm{MF}}y_{j+1}^{(i)}\end{cases}$$

Note: $v_{j+1}^{\mathrm{MF},(i)}$ are all iid.

Bayes filter vs mean-field EnKF

Assuming that for some $j \geq 0$,

$$\pi_{v_j^{\mathrm{MF},(i)}} = \pi_{V_j | Y_{1:j} = y_{1:j}}$$

then, since

$$v_{j+1}^{\mathrm{MF}} = \Psi(v_j^{\mathrm{MF}}) + \xi_j \overset{D}{=} \Psi(V_j) + \xi_j | (Y_{1:j} = y_{1:j}) \ = \hat{V}_{j+1} | Y_{1:j} = y_{1:j}$$

the next-time prediction pdfs of BF and MFEnKF will agree:

$$\pi_{\hat{v}_{j+1}^{\mathrm{MF},(i)}} = \pi_{V_{j+1} | Y_{1:j} = y_{1:j}}$$

$$\text{However, by} \quad v_{j+1}^{\mathrm{MF},(i)} = \hat{v}_{j+1}^{\mathrm{MF},(i)} + \underbrace{K_{j+1}^{\mathrm{MF}} \left(y_{j+1}^{(i)} - H \hat{v}_{j+1}^{\mathrm{MF},(i)} \right)}_{Y}$$

we obtain

$$\pi_{v_{j+1}^{\mathrm{MF},(i)}}(v) = \int \rho_{Y | v_{j+1}^{\mathrm{MF},(i)}}(v - x) \pi_{v_j^{\mathrm{MF},(i)}}(x) \, dx = \pi_{Y | v_{j+1}^{\mathrm{MF},(i)}} * \pi_{v_j^{\mathrm{MF},(i)}}(v).$$

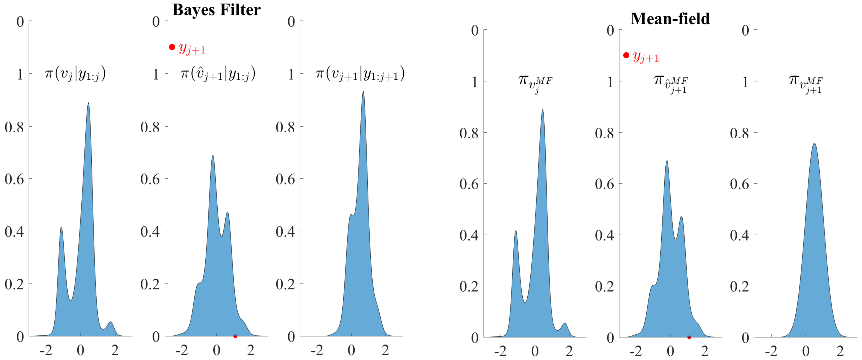
with

$$Y | v_{j+1}^{\mathrm{MF},(i)} = K_{j+1}^{\mathrm{MF}} \left(y_{j+1}^{(i)} - H \hat{v}_{j+1}^{\mathrm{MF},(i)} \right) | v_j^{\mathrm{MF},(i)} \sim K_{j+1}^{\mathrm{MF}} N(y_{j+1} - H \hat{v}_{j+1}^{\mathrm{MF},(i)}, \Gamma).$$

Bayes filter vs mean-field measure

$$\text{BF:} \quad \pi(v_{j+1}|y_{1:j+1}) \propto \pi_{N(y_{j+1},\Gamma)}(v_{j+1})\pi(v_{j+1}|y_{1:j})$$

$$\text{MFEnKF:} \quad \pi_{\pi_{v_{j+1}^{\mathrm{MF}}}}(v_{j+1}) \propto \pi_{K_{j+1}^{\mathrm{MF}} N(y_{j+1}-H\hat{v}_{j+1}^{\mathrm{MF}},\Gamma)} * \pi_{v_j^{\mathrm{MF}}}(v_{j+1}).$$



Conclusion: EnKF has two types of approximation errors:

1. Prediction error due to a finite ensemble, and
2. analysis error due to the particle-wise Gaussian variational principle.

Convergence of EnKF

Notation: Let

$$\pi_j^{\mathrm{EnKF,M}}(dv) := \frac{1}{M}\sum_{i=1}^{M} \delta_{v_j^{(i)}}(dv),$$

and let π_j^{MF} denote the distribution for a mean-field particle at time j:

$$v_j^{\mathrm{MF},(i)} \sim \pi_j^{\mathrm{MF}} \quad \text{and} \quad \pi_j^{\mathrm{MF}}[f] = \mathbb{E}^{\pi_j^{\mathrm{MF}}}[f].$$

For a QoI $f : \mathbb{R}^d \to \mathbb{R}$, let

$$\pi_j^{\mathrm{EnKF,M}}[f] := \frac{1}{M}\sum_{i=1}^{M} f(v_j^{(i)}) = \mathbb{E}^{\pi_j^{\mathrm{EnKF,M}}}[f]$$

and

$$\pi_j^{\mathrm{MF}}[f] := \mathbb{E}^{\pi_j^{\mathrm{MF}}}[f].$$

We describe two kinds of large-ensemble limit types of convergence:

- convergence of EnKF to the Kalman filter when Ψ is linear, and
- $\pi_j^{\mathrm{EnKF,M}}[f] \to \pi_j^{\mathrm{MF}}[f]$ when Ψ is nonlinear.

Theorem 1 (Mandel et al. "On the convergence of the ensemble Kalman filter" (2011))

Consider the linear-Gaussian filter problem

$$V_{j+1} = AV_j + \xi_j, \quad \xi_j \sim N(0, \Sigma),$$
$$Y_{j+1} = HV_{j+1} + \eta_{j+1}, \quad \eta_{j+1} \sim N(0, \Gamma),$$

and assume that $V_0 \sim N(m_0, C_0)$.
Then, for any observation sequence $y_1, y_2, \ldots$, *it holds that*

$$\pi_j^{\mathrm{MF}} = \mathbb{P}_{V_j | Y_{1:j} = y_{1:j}} = N(m_j, C_j)$$

with (m_j, C_j) *determined through the Kalman filtering iterative formulas, and as* $M \to \infty$,

$$E_M[v_j^{(\cdot)}] \overset{L^2(\Omega)}{\to} m_j, \quad \mathrm{Cov}_M[v_j^{(\cdot)}] \overset{L^2(\Omega)}{\to} C_j.$$

Application: EnKF may be a sound choice in linear-Gaussian settings when $d \gg 1$, because then Kalman filtering becomes infeasible due to storage constraints.

Theorem 2 (Le Gland et al., (2009))

Consider the dynamics and observations,

$$
\begin{aligned}
V_{j+1} &= \Psi(V_j) + \xi_j, \quad \xi_j \sim N(0, \Sigma), \\
V_{j+1} &= HV_{j+1} + \eta_{j+1}, \quad \eta_{j+1} \sim N(0, \Gamma),
\end{aligned}
$$

and assume that $V_0 \in L^p(\Omega)$ *for any order* $p \geq 1$, *and that for the drift mapping* Ψ *and a QoI* $f : \mathbb{R}^d \to \mathbb{R}$,

$$
\max(|f(x) - f(y)|, |\Psi(x) - \Psi(y)|) \leq C|x-y|(1+|x|^s+|u|^s), \text{ for some } s \geq 0.
$$

Then, for any fixed observation sequence $y_1, y_2, \ldots$, *it holds for any* $p \geq 1$ *that*

$$
\|\pi_j^{EnKF,M}[f] - \pi_j^{\mathrm{MF}}[f]\|_{L^p(\Omega)} \leq \frac{C(p, j, y_{1:j})}{\sqrt{M}},
$$

(which also can be written

$$
\left(\mathbb{E}\left[\left| \sum_{i=1}^{M} \frac{f(v_j^{(i)})}{M} - \int_{\mathbb{R}^d} f(x)\, \pi_j^{\mathrm{MF}}(dx) \right|^p \right] \right)^{1/p} \leq \frac{C(p, j, y_{1:j})}{\sqrt{M}}).
$$

Overview

Computing sample moments in the ambient space $\mathbb{R}^k$

A crucial step in the EnKF iteration is the computation of the prediction sample covariance:

$$\hat{C}_j = \mathrm{Cov}_M[v_j^{(\cdot)}].$$

and its usage in the Kalman gain:

$$K_j = \hat{C}_j H^T (H \hat{C}_j H^T + \Gamma)^{-1}.$$

Note that rather than the full matrix $\hat{C}_j$, what one needs for computing the gain is

$$\begin{aligned}
H \hat{C}_j H^T &= H \Big(\frac{1}{M-1} \sum_{i=1}^{M} (\hat{v}_j^{(i)} - \hat{m}_j)(\hat{v}_j^{(i)} - \hat{m}_j)^T \Big) H^T \\
&= \frac{1}{M-1} \sum_{i=1}^{M} H(\hat{v}_j^{(i)} - \hat{m}_j) \Big(H(\hat{v}_j^{(i)} - \hat{m}_j) \Big)^T \\
&= \mathrm{Cov}_M[H v_j^{(\cdot)}] \in \mathbb{R}^{k \times k}.
\end{aligned}$$

and

$$\hat{C}_j H^T = \mathrm{Cov}_M[v_j^{(\cdot)}, H v_j^{(\cdot)}] \in \mathbb{R}^{d \times k}.$$

Extension to nonlinear filtering settings

The resulting EnKF formulas

$$\text{Prediction} \begin{cases} \hat{v}_{j+1}^{(i)} &= \Psi(v_j^{(i)}) + \xi_j^{(i)} \end{cases}$$

$$\text{Analysis} \quad \begin{cases} K_{j+1} &= \mathrm{Cov}_M[v_j^{(\cdot)}, Hv_j^{(\cdot)}](\mathrm{Cov}_M[Hv_j^{(\cdot)}] + \Gamma)^{-1} \\ y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{(i)} &= \hat{v}_{j+1}^{(i)} + K_{j+1}\left(y_{j+1}^{(i)} - H\hat{v}_{j+1}^{(i)}\right) \end{cases}$$

may also be viewed as a motivation for the following extension to nonlinear observation mappings[1] $h : \mathbb{R}^d \to \mathbb{R}^k$:

$$\text{Prediction} \begin{cases} \hat{v}_{j+1}^{(i)} &= \Psi(v_j^{(i)}) + \xi_j^{(i)} \end{cases}$$

$$\text{Analysis} \quad \begin{cases} K_{j+1} &= \mathrm{Cov}_M[v_j^{(\cdot)}, h(v_j^{(\cdot)})](\mathrm{Cov}_M[h(v_j^{(\cdot)})] + \Gamma)^{-1} \\ y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{(i)} &= \hat{v}_{j+1}^{(i)} + K_{j+1}\left(y_{j+1}^{(i)} - h(\hat{v}_{j+1}^{(i)})\right). \end{cases}$$

[1] Evensen, "Data Assimilation, The Ensemble Kalman Filter", (2009).

Rough idea of alternative approach to nonlinear observations in EnKF

$$\text{Prediction}\begin{cases}\hat{v}_{j+1}^{(i)} &= \Psi(v_j^{(i)}) + \xi_j^{(i)} \\ \hat{m}_{j+1} &= E_M[v_{j+1}^{(\cdot)}] \\ \hat{C}_{j+1} &= \mathrm{Cov}_M[v_{j+1}^{(\cdot)}]\end{cases}$$

And solve the following minimization problem by iterated solver for each particle $i = 1, 2, \ldots, M$ [2]:

$$\text{Analysis}\begin{cases}y_{j+1}^{(i)} &= y_{j+1} + \eta_{j+1}^{(i)} \\ v_{j+1}^{(i)} &= \arg\min_{u\in\mathbb{R}^d} \frac{1}{2}|y_j^{(i)} - h(u)|_\Gamma^2 + \frac{1}{2}|u - \hat{m}_j|_{\hat{C}_j}^2\end{cases}$$

[2] Oliver and Gu, "An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation" (2007)

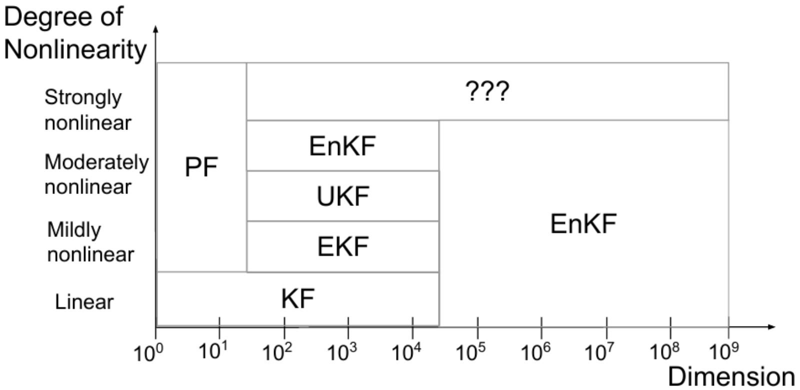
Summary

- We have introduced three nonlinear filtering methods based on Gaussian approximation in the update step (3DVAR, ExKF and EnKF).

- The methods do not generally converge to the Bayes filter when Ψ is nonlinear, but should not for that reason alone be excluded from practical use.

- EnKF offers the most robust prediction-step approach, it converges in weak sense to the mean-field EnKF when h is linear, and it may be extended to settings with nonlinear h.

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".