What do we know about EnKF?

David Kelly
Kody Law   Andrew Stuart
Andrew Majda   Xin Tong

Courant Institute
New York University
New York, NY

April 10, 2015

CAOS seminar, **Courant**.

Talk outline

1. What is EnKF?

2. What is known about EnKF?

3. Can we use stochastic analysis to better understand EnKF?

The data assimilation problem

We have a **model**

$$\frac{dv}{dt} = F(v) \quad \text{with } v_0 \sim \mu \,,$$

with a flow $v(t) = \Psi_t(v_0)$. Think of this as **very high dimensional**, **nonlinear** and possibly **stochastic**.

We want to **estimate** $v_n = v(nh)$ for some $h > 0$ and $n = 0, 1, 2, \dots$ given the **observations**

$$y_n = Hv_n + \xi_n \quad \text{for } \xi_n \text{ iid } N(0, \Gamma).$$

To formulate a solution to this problem, we write down the conditional density using **Bayes' formula**.

Bayes' formula filtering update

Let $Y_n = \{y_0, y_1, \ldots, y_n\}$. We want to compute the conditional density $\mathbf{P}(v_{n+1}|Y_{n+1})$, using $\mathbf{P}(v_n|Y_n)$ and y_{n+1}.

By Bayes' formula, we have

$$\mathbf{P}(v_{n+1}|Y_{n+1}) = \mathbf{P}(v_{n+1}|Y_n, y_{n+1}) \propto \mathbf{P}(y_{n+1}|v_{n+1})\mathbf{P}(v_{n+1}|Y_n)$$

But we need to compute the integral

$$\mathbf{P}(v_{n+1}|Y_n) = \int \mathbf{P}(v_{n+1}|Y_n, v_n)\mathbf{P}(v_n|Y_n)dv_n \ .$$

For high dimensional nonlinear systems, this is computationally infeasible.

The Kalman Filter

For linear models, this integral is Gaussian and can be computed explicitly. The conditional density is characterized by its mean and covariance

$$
\begin{aligned}
m_{n+1} &= \widehat{m}_n - G_{n+1}(H\widehat{m}_n - y_{n+1}) \\
C_{n+1} &= (I - G_{n+1}H)\widehat{C}_{n+1} \, ,
\end{aligned}
$$

where

- $(\widehat{m}_{n+1}, \widehat{C}_{n+1})$ is the **forecast** mean and covariance.
- $G_{n+1} = \widehat{C}_{n+1}H^T(\Gamma + H\widehat{C}_{n+1}H^T)^{-1}$ is the **Kalman gain**.

The procedure of updating $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$ is known as the **Kalman filter**.

When applied to nonlinear models, this is called the **Extended Kalman filter**.

For high dimensional non-linear systems, calculations are expensive. A better idea is to **sample**.

The **Ensemble Kalman Filter** (EnKF) is a low dimensional sampling algorithm. (Evensen '94)

EnKF generates an ensemble of **approximate samples** from the posterior.

For linear models, one can draw samples, using the **Randomized Maximum Likelihood** method.

RML method

Let $u \sim N(\widehat{m}, \widehat{C})$ and $\eta \sim N(0, \Gamma)$. We make an observation

$$y = Hu + \eta \; .$$

We want the conditional distribution of $u|y$. This is called an **inverse problem**.

RML takes a sample

$$\{\widehat{u}^{(1)}, \ldots, \widehat{u}^{(K)}\} \sim N(\widehat{m}, \widehat{C})$$

and turns them into a sample

$$\{u^{(1)}, \ldots, u^{(K)}\} \sim u|y$$

RML method: How does it work?

Along with the prior sample $\{\widehat{u}^{(1)}, \ldots, \widehat{u}^{(K)}\}$, we create **artificial observations** $\{y^{(1)}, \ldots, y^{(K)}\}$ where

$$y^{(k)} = y + \eta^{(k)} \qquad \text{where } \eta^{(k)} \sim N(0, \Gamma) \text{ i.i.d}$$

Then define $u^{(k)}$ using the **Kalman mean update**, with $(\widehat{u}^{(k)}, y^{(k)})$

$$u^{(k)} = \widehat{u}^{(k)} - G(\widehat{u})(H\widehat{u}^{(k)} - y^{(k)}) .$$

Where the Kalman gain $G(\widehat{u})$ is computing using the covariance of the prior $\widehat{u}$.

The set $\{u^{(1)}, \ldots, u^{(K)}\}$ are exact samples from $u|y$.

EnKF uses the same method, but with an **approximation** of the covariance in the Kalman gain.

The set-up for EnKF

Suppose we are given the ensemble $\{u_n^{(1)}, \dots, u_n^{(K)}\}$. For each ensemble member, we create an **artificial observation**

$$y_{n+1}^{(k)} = y_{n+1} + \xi_{n+1}^{(k)} \quad , \quad \xi_{n+1}^{(k)} \text{ iid } N(0, \Gamma).$$

We update each particle using the **Kalman update**

$$u_{n+1}^{(k)} = \Psi_h(u_n^{(k)}) - G(u_n) \left(H\Psi_h(u_n^{(k)}) - y_{n+1}^{(k)} \right) ,$$

where $G(u_n)$ is the "**Kalman gain**" computed using the **forecasted ensemble covariance**

$$\widehat{C}_{n+1} = \frac{1}{K} \sum_{k=1}^{K} (\Psi_h(u_n^{(k)}) - \overline{\Psi_h(u_n)})^T (\Psi_h(u_n^{(k)}) - \overline{\Psi_h(u_n)}) .$$

What do we know about EnKF?

Not much.

Theorem : For linear forecast models,

ENKF $\rightarrow$ KF as $K \rightarrow \infty$

(Le Gland et al / Mandel et al. 09').

Ideally, we would like results with a finite ensemble size.

1 - Filter divergence

2 - Filter stability

3 - Continuous time scaling limit

1 - Filter divergence

In certain situations, it has been observed ($\star$) that the ensemble can **blow-up** (ie. reach machine-infinity) in **finite time**, even when the model has nice bounded solutions.

This is known as **catastrophic filter divergence**.

Does this have a **dynamical justification** or is it a **numerical artefact**?

$\star$ Harlim, Majda (2010), Gottwald (2011), Gottwald, Majda (2013).

Assumptions on the model

We make a **dissipativity** assumption on the model. Namely that

$$\frac{dv}{dt} + Av + B(v, v) = f$$

with A linear elliptic and B bilinear, satisfying certain estimates and symmetries.

This guarantees the **absorbing ball property** (the system has a Lyapunov function).

Eg. 2d-Navier-Stokes, Lorenz-63, Lorenz-96.

Discrete time results

Fix an observation frequency $h > 0$. Let $e_n^{(k)} = u_n^{(k)} - v_n$.

Theorem (K, Law, Stuart 14')

If $H = Id$, $\Gamma = Id$ then there exists constant $\beta > 0$ such that

$$\mathbf{E}|e_n^{(k)}|^2 \leq e^{2\beta nh}\mathbf{E}|e_0^{(k)}|^2 + 2K\gamma^2 \left(\frac{e^{2\beta nh} - 1}{e^{2\beta h} - 1}\right)$$

Rmk. Thm (Tong, Majda, K 15) $\sup_{n\geq 1} \mathbf{E}|u_n^{(k)}|^2 < \infty$
The ensemble inherits the absorbing ball property from the model.

Discrete time results with variance inflation

Suppose we replace

$$\widehat{C}_{n+1} \mapsto \alpha^2 I + \widehat{C}_{n+1}$$

at each update step. This is known as **additive variance inflation**.

Theorem (K, Law, Stuart 14')

If $H = Id$, $\Gamma = \gamma^2 Id$ then there exists constant $\beta > 0$ such that

$$\mathbf{E}|e_n^{(k)}|^2 \leq \theta^n \mathbf{E}|e_0^{(k)}|^2 + 2K\gamma^2 \left(\frac{1-\theta^n}{1-\theta}\right)$$

where $\theta = \frac{\gamma^2 e^{2\beta h}}{\alpha^2+\gamma^2}$. In particular, if we pick α large enough (so that $\theta < 1$) then

$$\lim_{n\to\infty} \mathbf{E}|e_n^{(k)}|^2 \leq \frac{2K\gamma^2}{1-\theta}$$

2 - Filter stability

Is the filter **stable** with respect to its **initial data** $(u_0^{(1)}, \ldots, u_0^{(K)})$? Will initialization errors **dissipate** or **propagate** over time?

This can be answered by verifying **ergodicity**.

Geometric ergodicity

In addition to the **dissipativity** assumption, assume the model is **stochastic** with a positive density everywhere.

Theorem (Tong, Majda, K 15)

If $H = Id$ then the signal-ensemble process $(v_n, u_n^{(1)}, \dots, u_n^{(K)})$ is **geometrically ergodic**. *That is, there exists unique stationary measure ρ, $\theta \in (0,1)$ such that, given any initialization μ*

$$|P^n \mu - \rho| \leq C\theta^n$$

where $P^n\mu$ is the distribution $(v_n, u_n^{(1)}, \dots, u_n^{(K)})$ initialized with μ.

Rmk. The $H = Id$ is not really needed. Sufficient to have a Lyapunov function for $(v_n, u_n^{(1)}, \dots, u_n^{(K)})$.

3 - Scaling limit

Can we learn anything from the $h \to 0$ **scaling limit** of the algorithm?

The EnKF equations look like a discretization

Recall the ensemble update equation

$$
\begin{aligned}
u_{n+1}^{(k)} &= \Psi_h(u_n^{(k)}) + G(u_n)\left(y_{n+1}^{(k)} - H\Psi_h(u_n^{(k)})\right) \\
&= \Psi_h(u_n^{(k)}) + \widehat{C}_{n+1}H^T(H^T\widehat{C}_{n+1}H + \Gamma)^{-1}\left(y_{n+1}^{(k)} - H\Psi_h(u_n^{(k)})\right)
\end{aligned}
$$

Subtract $u_n^{(k)}$ from both sides and divide by h

$$
\begin{aligned}
\frac{u_{n+1}^{(k)} - u_n^{(k)}}{h} &= \frac{\Psi_h(u_n^{(k)}) - u_n^{(k)}}{h} \\
&\quad + \widehat{C}_{n+1}H^T(hH^T\widehat{C}_{n+1}H + h\Gamma)^{-1}\left(y_{n+1}^{(k)} - H\Psi_h(u_n^{(k)})\right)
\end{aligned}
$$

Clearly we need to rescale the noise (ie. Γ).

Continuous-time limit

If we set $\Gamma = h^{-1}\Gamma_0$ and substitute $y_{n+1}^{(k)}$, we obtain

$$\frac{u_{n+1}^{(k)} - u_n^{(k)}}{h} = \frac{\Psi_h(u_n^{(k)}) - u_n^{(k)}}{h} + \widehat{C}_{n+1} H^T (h H^T \widehat{C}_{n+1} H + \Gamma_0)^{-1} \left(Hv + h^{-1/2}\Gamma_0^{1/2}\xi_{n+1} + h^{-1/2}\Gamma_0^{1/2}\xi_{n+1}^{(k)} - H\Psi_h(u_n^{(k)}) \right)$$

But we know that

$$\Psi_h(u_n^{(k)}) = u_n^{(k)} + O(h)$$

and

$$\widehat{C}_{n+1} = \frac{1}{K}\sum_{k=1}^{K} (\Psi_h(u_n^{(k)}) - \overline{\Psi_h(u_n)})^T (\Psi_h(u_n^{(k)}) - \overline{\Psi_h(u_n)})$$

$$= \frac{1}{K}\sum_{k=1}^{K} (u_n^{(k)} - \overline{u_n})^T (u_n^{(k)} - \overline{u_n}) + O(h) = C(u_n) + O(h)$$

Continuous-time limit

We end up with

$$\frac{u_{n+1}^{(k)} - u_n^{(k)}}{h} = \frac{\Psi_h(u_n^{(k)}) - u_n^{(k)}}{h} - C(u_n)H^T\Gamma_0^{-1}H(u_n^{(k)} - v_n)$$
$$+ C(u_n)H^T\Gamma_0^{-1}\left(h^{-1/2}\xi_{n+1} + h^{-1/2}\xi_{n+1}^{(k)}\right) + O(h)$$

This looks like a **numerical scheme** for **Itô S(P)DE**

$$\frac{du^{(k)}}{dt} = F(u^{(k)}) - C(u)H^T\Gamma_0^{-1}H(u^{(k)} - v) \qquad (\bullet)$$
$$+ C(u)H^T\Gamma_0^{-1/2}\left(\frac{dB}{dt} + \frac{dW^{(k)}}{dt}\right).$$

Nudging

$$\frac{du^{(k)}}{dt} = F(u^{(k)}) - C(u)H^T\Gamma_0^{-1}H(u^{(k)} - v) \qquad (\bullet)$$
$$+ C(u)H^T\Gamma_0^{-1/2}\left(\frac{dB}{dt} + \frac{dW^{(k)}}{dt}\right) .$$

1 - Extra dissipation term only sees **differences in observed space**

2 - Extra dissipation only occurs in the **space spanned by ensemble**

Kalman-Bucy limit

If F were **linear** and we write $m(t) = \frac{1}{K}\sum_{k=1}^{K} u^{(k)}(t)$ then

$$\frac{dm}{dt} = F(m) - C(u)H^T\Gamma_0^{-1}H(m - v) + C(u)H^T\Gamma_0^{-1/2}\frac{dB}{dt} + O(K^{-1/2}) \, .$$

This is the equation for the **Kalman-Bucy** filter, with empirical covariance $C(u)$. The remainder $O(K^{-1/2})$ can be thought of as a **sampling error**.

Continuous-time results

Theorem (K, Law, Stuart)

Suppose that $\{u^{(k)}\}_{k=1}^{K}$ *satisfy* ($\bullet$) *with* $H = \Gamma = Id$. *Let*

$$e^{(k)} = u^{(k)} - v \ .$$

Then there exists constant $\beta > 0$ *such that*

$$\frac{1}{K}\sum_{k=1}^{K} \mathbf{E}|e^{(k)}(t)|^2 \leq \left(\frac{1}{K}\sum_{k=1}^{K} \mathbf{E}|e^{(k)}(0)|^2\right) \exp(\beta t) \ .$$

Rmk. Thm (Tong, Majda, K 15) $\sup_{t\geq 0} \mathbf{E}|u^{(k)}(t)|^2 < \infty$
The ensemble inherits the absorbing ball property from the model.

Why do we need $H = \Gamma = Id$?

In the equation

$$\frac{du^{(k)}}{dt} = F(u^{(k)}) - C(u)H^T\Gamma_0^{-1}H(u^{(k)} - v) + C(u)H^T\Gamma_0^{-1/2}\left(\frac{dW^{(k)}}{dt} + \frac{dB}{dt}\right) .$$

The **energy** pumped in by the noise must be balanced by **contraction** of $(u^{(k)} - v)$. So the operator

$$C(u)H^T\Gamma_0^{-1}H$$

must be **positive-definite**.

Both $C(u)$ and $H^T\Gamma_0^{-1}H$ are pos-def, but this doesn't guarantee the same for the **product**.

Testing stability on the fly

Suppose we can actually measure the spectrum of the operator

$$C(u)H^T\Gamma_0^{-1}H$$

whilst the algorithm is running. If we know that it is pos-def, then the filter must not be blowing up.

If we knew that

$$C(u)H^T\Gamma_0^{-1}H \geq \lambda(t) > 0 \; .$$

Then we can say even more (eg. stability).

Summary + Future Work

(1) Cannot "prove" that **catastrophic filter divergence** is a numerical phenomenon, but idecent starting point.

(2) If the filter isn't blowing up, then it should be **stable**.

(3) Writing down an SDE/SPDE allows us to see the **important quantities** in the algorithm.

(1) Improve the condition on H? Seems hard. Change the algorithm instead. (*Ongoing work with Majda, Tong*)

Thank you!

Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time.
D. Kelly, K.Law, A. Stuart.
Nonlinearity 2014.

Stability and geometric ergodicity of ensemble based Kalman methods.
X. Tong, A. Majda, D. Kelly.

`www.dtbkelly.com`