Ergodicity in data assimilation methods

David Kelly

Andy Majda Xin Tong

Courant Institute
New York University
New York NY
www.dtbkelly.com

November 11, 2016

University of Delaware

Part I: What is data assimilation?

What is data assimilation?

Suppose u satisfies an evolution equation

$$\frac{du}{dt} = F(u)$$

with some **unknown** initial condition $u_0 \sim \mu_0$.

There is a true trajectory of u that is producing *partial, noisy* observations at times $t = h, 2h, \ldots, nh$:

$$y_n = Hu_n + \xi_n$$

where H is a linear operator (think low rank projection), $u_n = u(nh)$, and $\xi_n \sim N(0, \Gamma)$ iid.

The aim of **data assimilation** is to characterize the conditional distribution of u_n given the observations $\{y_1, \ldots, y_n\}$

The conditional distribution is updated via the **filtering cycle**.

Illustration (Initialization)

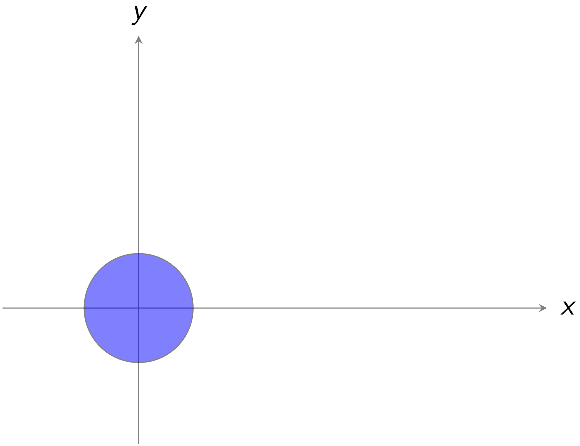


Figure: The blue circle represents our initial uncertainty $u_0 \sim \mu_0$.

Illustration (Forecast step)

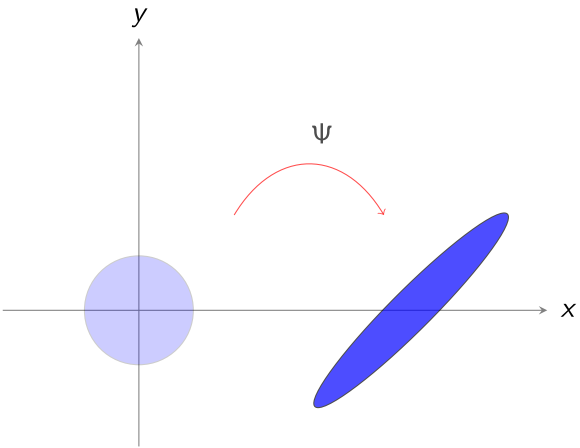


Figure: Apply the time h flow map Ψ. This produces a new probability measure which is our forecasted estimate of u_1. This is called the forecast step.

Illustration (Make an observation)

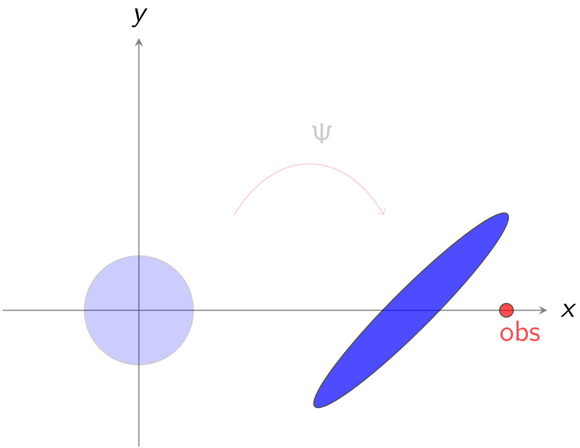


Figure: We make an observation $y_1 = Hu_1 + \xi_1$. In the picture, we only observe the x variable.

Illustration (Analysis step)

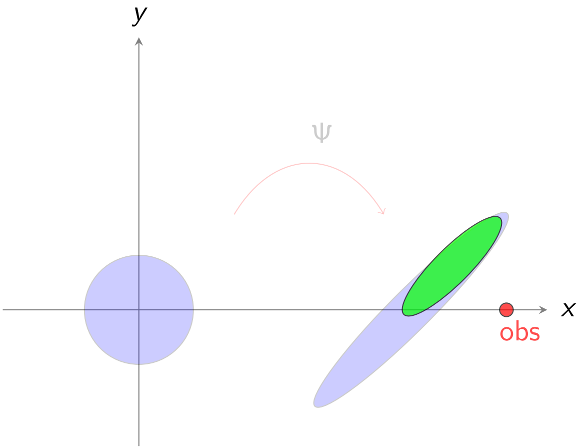


Figure: Using Bayes formula we compute the conditional distribution of $u_1|y_1$. This new measure (called the posterior) is the new estimate of u_1. The uncertainty of the estimate is reduced by incorporating the observation. The forecast distribution steers the update from the observation.

Bayes' formula filtering update

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

By Bayes' formula, we have

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

But we need to compute the integral

$$\mathbf{P}(u_{n+1}|Y_n) = \int \mathbf{P}(u_{n+1}|Y_n, u_n)\mathbf{P}(u_n|Y_n)du_n \ .$$

The 'optimal filtering' framework is **impractical** for high dimensional models, as the integrals are impossible to compute and densities impossible to store.

In **numerical weather prediction**, we have $O(10^9)$ variables for ocean-atmosphere models (discretized PDEs).

The **Ensemble Kalman filter** (EnKF) is a low dimensional, empirical approximation of the posterior measure $\mathbf{P}(u_n|Y_n)$.

EnKF builds on the idea of a **Kalman filter**.

The Kalman Filter

For a linear model $u_{n+1} = Mu_n + \eta_{n+1}$, the Bayesian integral is Gaussian and can be computed explicitly. The conditional density is proportional to

$$\exp\left(-\frac{1}{2}|\Gamma^{-1/2}(y_{n+1} - Hu)|^2 - \frac{1}{2}|\widehat{C}_1^{-1/2}(u - \widehat{m}_{n+1})|^2\right)$$

where $(\widehat{m}_{n+1}, \widehat{C}_{n+1})$ is the **forecast** mean and covariance. We obtain

$$m_{n+1} = (1 - K_{n+1}H)\widehat{m}_n + K_{n+1}y_{n+1}$$
$$C_{n+1} = (I - K_{n+1}H)\widehat{C}_{n+1} \ ,$$

where $K_{n+1} = \widehat{C}_{n+1}H^T(\Gamma + H\widehat{C}_{n+1}H^T)^{-1}$ 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**.

EnKF is an '**approximate linearized sampling**' procedure.

Ensemble Kalman filter (Evensen 94)

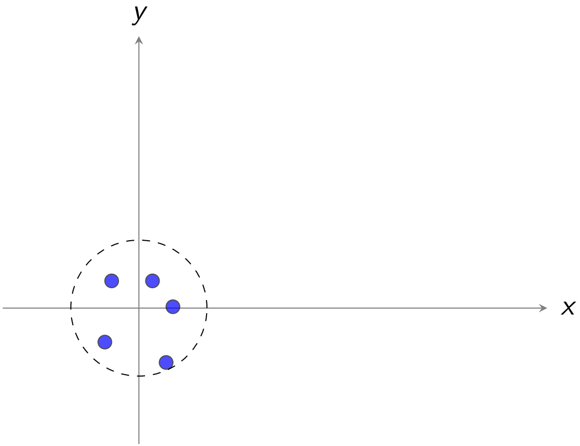


Figure: Start with K ensemble members drawn from some distribution. Empirical representation of u_0. The ensemble members are denoted $u_0^{(k)}$.

Only KN numbers are stored. For the covariance, better than Kalman if $K < N$.

Ensemble Kalman filter (Forecast step)

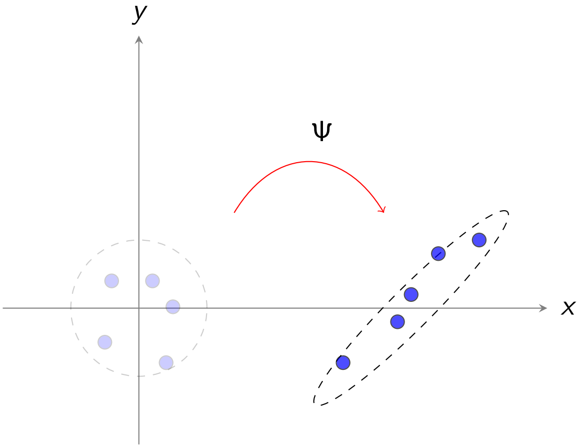


Figure: Apply the dynamics Ψ to each ensemble member.

Ensemble Kalman filter (Make obs)

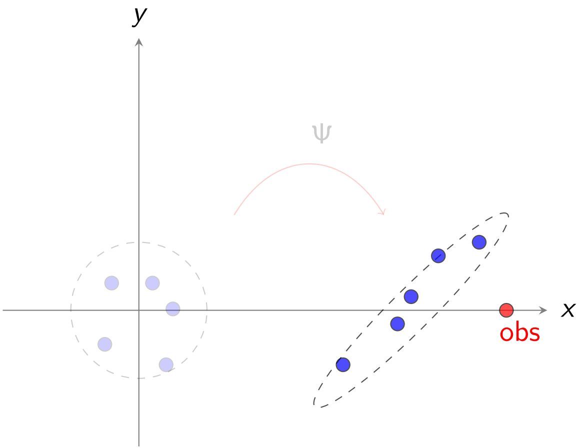


Figure: Make an observation.

Ensemble Kalman filter (Analysis step)

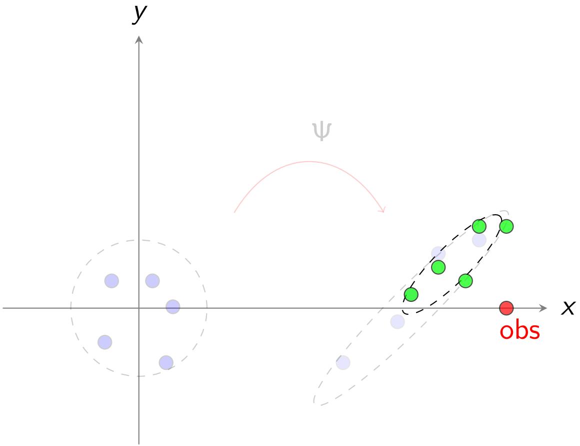


Figure: Approximate the forecast distribution with a Gaussian. Fit the Gaussian using the empirical statistics of the ensemble.

How to implement the Gaussian approximation

The posterior is *approximately sampled* using the **Randomized Maximum Likelihood** (RML) method: Draw the sample $u_{n+1}^{(k)}$ by minimizing the functional

$$\frac{1}{2}|\Gamma^{-1/2}(y_{n+1}^{(k)} - Hu)|^2 + \frac{1}{2}|\widehat{C}_{n+1}^{-1/2}(u - \Psi^{(k)}(u_n^{(k)}))|^2$$

where $y_{n+1}^{(k)} = y_{n+1} + \xi_{n+1}^{(k)}$ is a perturbed observation and $\widehat{C}_{n+1}$ is the empirical covariance of $\{\Psi^{(k)}(u_n^{(k)})\}_{k=1}^{K}$

In the linear case $\Psi(u_n) = Mu_n + \eta_n$, this produces iid Gaussian samples with mean and covariance satisfying the Kalman update equations, with $\widehat{C}_{n+1}$ in place of the true forecast covariance.

We end up with $u_{n+1}^{(k)} = (1 - K_{n+1}H)\Psi^{(k)}(u_n^{(k)}) + K_{n+1}Hy_{n+1}^{(k)}$.

Ensemble Kalman filter (Perturb obs)

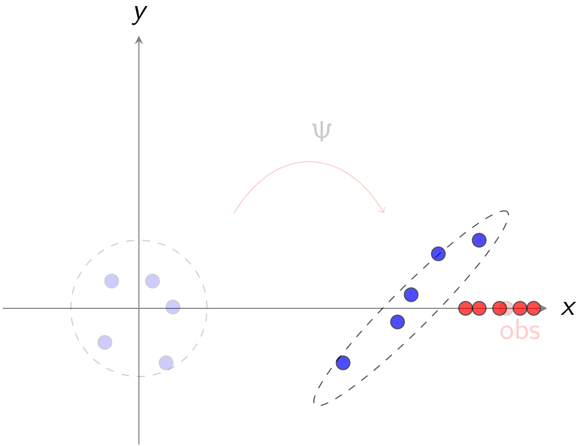


Figure: Turn the observation into K artificial observations by perturbing by the same source of observational noise.

$$y_1^{(k)} = y_1 + \xi_1^{(k)}$$

Ensemble Kalman filter (Analysis step)

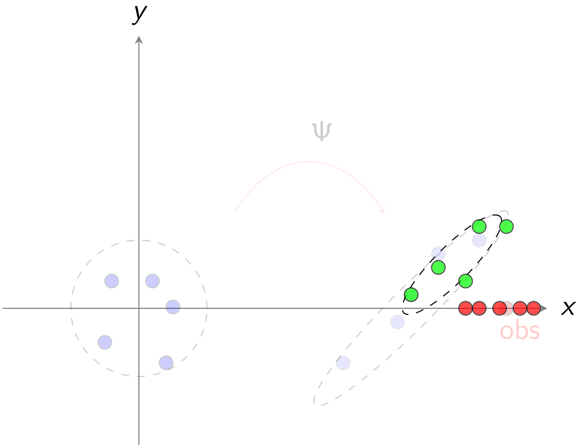


Figure: Update each member using the Kalman update formula. The Kalman gain K_1 is computed using the ensemble covariance.

$$u_1^{(k)} = (1 - K_1 H)\Psi(u_0^{(k)}) + K_1 H y_1^{(k)} \quad K_1 = \widehat{C}_1 H^T (\Gamma + H \widehat{C}_1 H^T)^{-1}$$

$$\widehat{C}_1 = \frac{1}{K-1} \sum_{k=1}^{K} (\Psi(u_0^{(k)}) - \widehat{m}_1)(\Psi(u_0^{(k)}) - \widehat{m}_1)^T$$

Part II : Ergodicity for EnKF

What are we interested in understanding?

1− Stability with respect to initializations (signal-filter ergodicity):

Statistics of $(u_n, u_n^{(1)}, \dots, u_n^{(K)})$ *converge to invariant statistics as* $n \to \infty$.

2− Accuracy of the approximation (conditional ergodicity):

Laws of $(u_n^{(1)}, \dots, u_n^{(K)}) | Y_n$ *with different intializations* $(u_0^{(1)}, \dots, u_0^{(K)})$ *converge to eachother (and hopefully the posterior) as* $n \to \infty$.

Rmk. All results are in the K fixed regime.

Animation

We have the two-dimensional model

$$du = -\nabla V(u)dt + \sigma dW$$

where $V(x, y) = (1 - x^2 - y^2)^2$ and we only observe the x variable.
EnKF is **signal-filter ergodic**, as the marginals converge to uniform measure on circle. But also **conditionally ergodic**, the law is close to the posterior, regardless of initialization. **Not** close to posterior, ensemble clusters on one of the two modes.

Today we focus on geometric ergodicity for the **signal-filter** process $(u_n, u_n^{(1)}, \dots, u_n^{(K)})$.

The theoretical framework

A Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on a state space $\mathcal{X}$ is called **geometrically ergodic** if it has a unique invariant measure π and for any initialization x_0 we have

$$\sup_{|f|\leq 1}\left|\mathbf{E}_{x_0} f(X_n) - \int f(x)\pi(dx)\right| \leq C(x_0) r^n$$

for some $r \in (0,1)$ and any m-ble bdd f.

We use the **coupling** approach: Let X'_n and X''_n be two copies of X_n, such that $X'_0 = x_0$ and $X''_0 \sim \pi$, and are coupled in such a way that $X'_n = X''_n$ for $n \geq T$, where T is the first hitting time $X'_T = X''_T$. Then we have

$$\|P^n \delta_{x_0} - \pi\|_{TV} \leq 2\mathbf{P}(T > n)$$

The **Doeblin** / **Meyn-Tweedie** approach is to verify two assumptions that guarantee the coupling can be constructed with $\mathbf{P}(T > n) \lesssim r^n$:

1- Lyapunov function / Energy dissipation: $\mathbf{E}_n|X_{n+1}|^2 \leq \alpha|X_n|^2 + \beta$ with $\alpha \in (0,1)$, $\beta > 0$.

2- Minorization: Find compact $C \subset \mathcal{X}$, measure ν supported on C, $\kappa > 0$ such that $P(x, A) \geq \kappa\nu(A)$ for all $x \in C$, $A \subset \mathcal{X}$.

To construct the coupling (within C) we let $\tilde{X}_{n+1} = \tilde{f}(\tilde{X}_n, \omega)$ describe the Markov chain with kernel $\tilde{P}(x, A) = \frac{1}{1-\kappa}(P(x, A) - \kappa\nu(A))$ and let

$$X'_{n+1} = \phi\tilde{f}(X'_n, \omega) + (1 - \phi)\xi$$

where $\phi \sim \mathrm{Bernoulli}(\kappa)$ and $\xi \sim \nu$. Easy to see that $\mathbf{P}_x(X'_1 \in A) = P(x, A)$, so this is a copy of the original Markov chain.

Theorem (K., Majda, Tong 15 Nonlinearity)

Let $X_n = (u_n, u_n^{(1)}, \dots, u_n^{(K)})$ *and suppose* u_n *satisfies* **observable energy dissipation**:

$$\mathbf{E}_n |H u_{n+1}|^2 \leq \alpha |H u_n|^2 + \beta \; ,$$

for $\alpha \in (0,1)$, $\beta > 0$. *Then there exists a unique stationary measure* π *for the Markov chain* X_n *and moreover, there exists* $r \in (0,1)$ *such that*

$$\sup_{|f| \leq 1} \left| \mathbf{E}_{x_0} f(X_n) - \int f(x) \pi(dx) \right| \leq C(x_0) r^n$$

for any initialization $x_0 = (u_0, u_0^{(1)}, \dots, u_0^{(K)})$ *and any observable* f.

The **Observable energy dissipation** assumption

$$\mathbf{E}_n|Hu_{n+1}|^2 \leq \alpha|Hu_n|^2 + \beta \; .$$

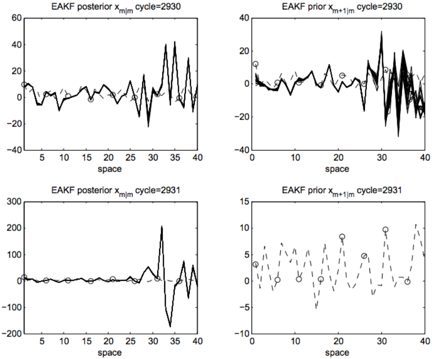
requires that the dissipation in the observed variables is controlled by the observed variables.

Guarantees 'inheritance of stability', the signal-ensemble process X_n has a Lyapunov function $\mathbf{E}_n|X_{n+1}|^2 \leq \alpha'|X_n|^2 + \beta'$.

This assumptions is **strong** - stability is not always inherited by the filter.

Catastrophic filter divergence

Lorenz-96: $\dot{u}_j = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F$ with $j = 1, \ldots, 40$. Periodic BCs. Observe every fifth node. *(Harlim-Majda 10, Gottwald-Majda 12)*



True solution in a bounded set, but filter **blows up** to machine infinity in finite time!

The rotate-and-lock map (K., Majda, Tong. PNAS 15.)

The model $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ is a composition of two maps $\Psi(x,y) = \Psi_{lock}(\Psi_{rot}(x,y))$ where

$$\Psi_{rot}(x,y) = \begin{pmatrix} \rho\cos\theta & -\rho\sin\theta \\ \rho\sin\theta & \rho\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and Ψ_{lock} rounds the input to the nearest point in the grid

$$\mathcal{G} = \{(m,(2n+1)\varepsilon) \in \mathbb{R}^2 : m,n \in \mathbb{Z}\} \,.$$

It is easy to show that this model has an **energy dissipation principle**:

$$|\Psi(x,y)|^2 \leq \alpha|(x,y)|^2 + \beta$$

for $\alpha \in (0,1)$ and $\beta > 0$.

(a)

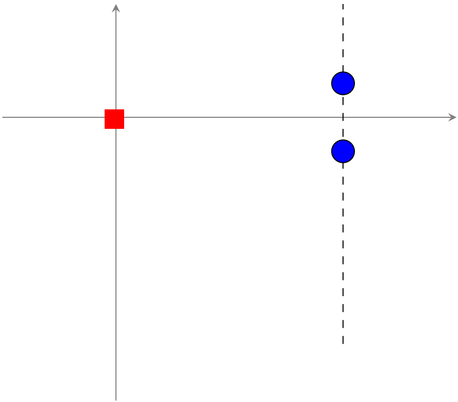


Figure: The red square is the trajectory $u_n = 0$. The blue dots are the positions of the forecast ensemble $\Psi(u_0^+)$, $\Psi(u_0^-)$. Given the locking mechanism in Ψ, this is a natural configuration.

(b)

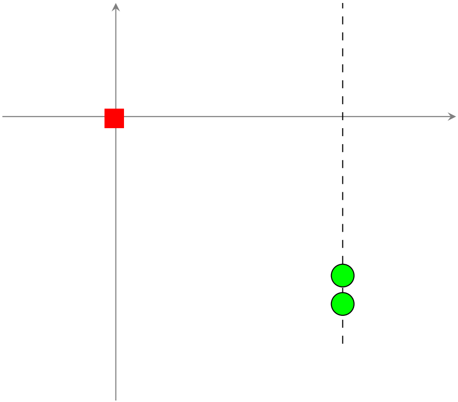


Figure: We make an observation (H shown below) and perform the analysis step. The green dots are u_1^+, u_1^-.

Observation matrix

$$H = \begin{bmatrix} 1 & 0 \\ \varepsilon^{-2} & 1 \end{bmatrix}$$

Truth $u_n = (0, 0)$.

The filter is certain that the x-coordinate is $\hat{x}$ (the dashed line). The filter thinks the observation must be $(\hat{x}, \varepsilon^{-2}\hat{x} + u_{1,y})$, but it is actually $(0, 0)$. The filter concludes that $u_{1,y} \approx -\varepsilon^{-2}\hat{x}$.

(c)

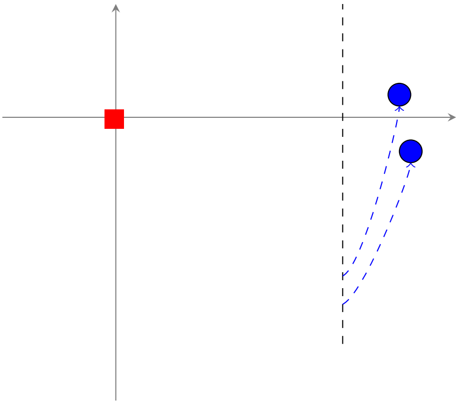


Figure: Beginning the next assimilation step. Apply Ψ_{rot} to the ensemble (blue dots)

(d)

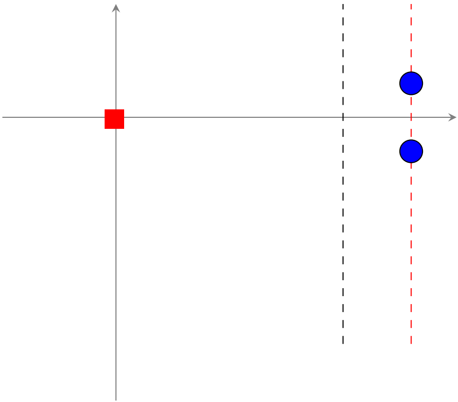


Figure: Apply Ψ_{lock}. The blue dots are the forecast ensemble $\Psi(u_1^+)$, $\Psi(u_1^-)$. Exact same as frame 1, but higher energy orbit. The cycle repeats leading to **exponential growth**.

Theorem (K., Majda, Tong 15 PNAS)

For any $N > 0$ and any $p \in (0,1)$ there exists a choice of parameters such that

$$\mathbf{P}\left(|u_n^{(k)}| \geq M_n \ \text{ for all } n \leq N\right) \geq 1 - p$$

where M_n is an exponentially growing sequence.

ie - The filter can be made to grow exponentially for an arbitrarily long time with an arbitrarily high probability.

Ensemble alignment can cause EnKF to gain energy, eventually leading to finite time **blow-up**

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

Can we get around the problem by **tweaking** the algorithm?

Adaptive Covariance Inflation (Tong, Majda, K. 15)

We modify algorithm by introducing a **covariance inflation** :

$$\widehat{C}_{n+1} \mapsto \widehat{C}_{n+1} + \lambda_{n+1} I$$

where

$$\lambda_{n+1} \propto \Theta_{n+1} \mathbf{1}(\Theta_{n+1} > \Lambda)$$

$$\Theta_{n+1} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} |y_{n+1}^{(k)} - H\Psi(u_n^{(k)})|^2}$$

and Λ is some constant threshold. If the predictions are near the observations, then there is no inflation.

Thm. The modified EnKF inherits an energy principle from the model.

$$\mathbf{E}_x |\Psi(x)|^2 \le \alpha |x|^2 + \beta \Rightarrow \mathbf{E}_n |u_{n+1}^{(k)}|^2 \le \alpha' |u_n^{(k)}|^2 + \beta'$$

Consequently, the modified EnKF is signal-filter ergodic.

Adaptive inflation allows us to use **cheap** integration schemes in the forecast dynamics. These would usually lead to numerical blow-up.

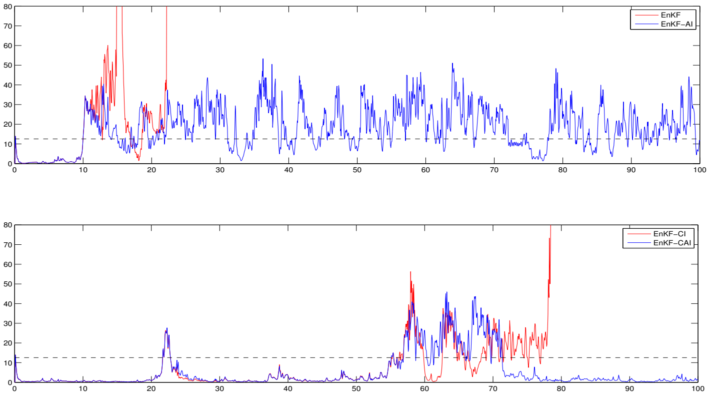


Figure: RMS error for EnKF on 5d Lorenz-96 with sparse obs (1 node), strong turbulence regime. Euler method with course step size. Lower panel has additional constant inflation which helps accuracy.

Applicable to more sophisticated geophysical models, such as 2-layer QG with course graining (Lee, Majda 16').

References

1 - D. Kelly, K. Law & A. Stuart. *Well-Posedness And Accuracy Of The Ensemble Kalman Filter In Discrete And Continuous Time.* **Nonlinearity** (2014).

2 - D. Kelly, A. Majda & X. Tong. *Concrete ensemble Kalman filters with rigorous catastrophic filter divergence*. **Proc. Nat. Acad. Sci.** (2015).

3 - X. Tong, A. Majda & D. Kelly. *Nonlinear stability and ergodicity of ensemble based Kalman filters*. **Nonlinearity** (2016).

4 - X. Tong, A. Majda & D. Kelly. *Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation*. To appear in **Comm. Math. Sci.** (2016).

All my slides are on my website (www.dtbkelly.com) **Thank you**!