Data assimilation in high dimensions

David Kelly

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

February 12, 2015

Graduate seminar, CIMS

What is data assimilation?

Suppose u satisfies

$$du = F(u)dt + dW$$

with some **unknown** initial condition u_0. We are most interested in geophysical models, so think high dimensional, nonlinear, stochastic.

Suppose we make *partial, noisy* observations at times $t = h, 2h, \ldots, nh, \ldots$

$$y_n = Hu_n + \xi_n$$

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 say something about the conditional distribution of u_n given the observations $\{y_1, \ldots, y_n\}$

Outline

1 - The basics: Bayes, Kalman etc.

2 - What to do for nonlinear models?

3 - What to do in high dimensions?

Illustration (Initialization)

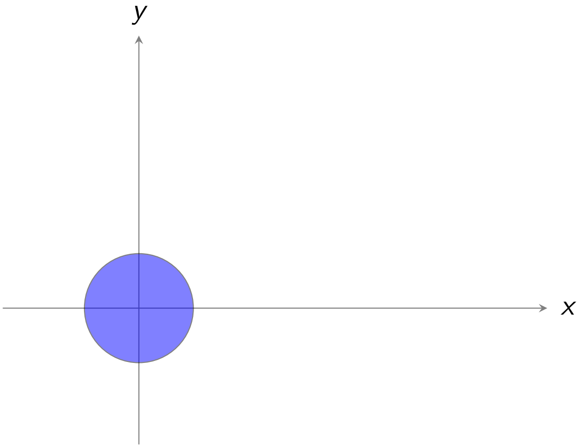


Figure: The blue circle represents our guess of u_0. Due to the uncertainty in u_0, this is a probability measure.

Illustration (Forecast step)

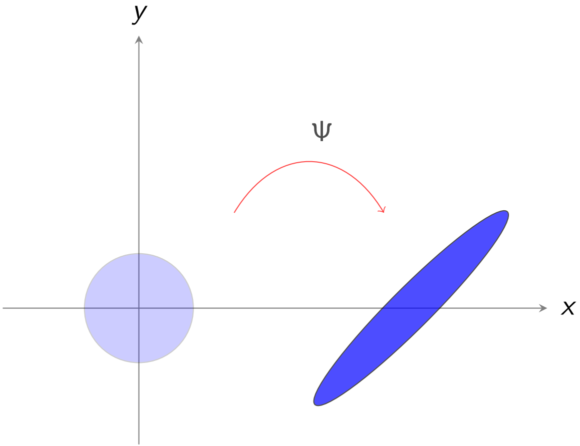


Figure: Apply the time h flow map Ψ. This produces a new probability measure which is our forecasted estimate of u_1. This is called the forecast step.

Illustration (Make an observation)

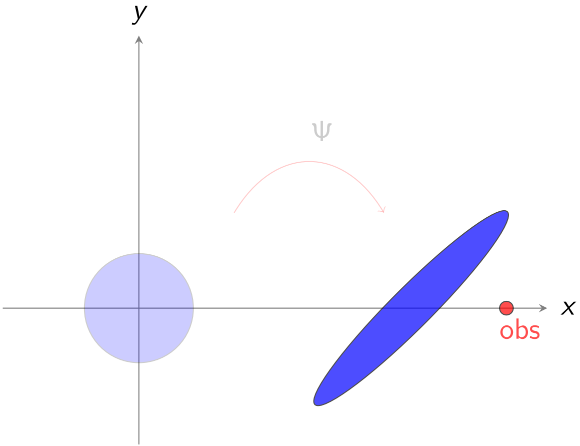


Figure: We make an observation $y_1 = Hu_1 + \xi_1$. In the picture, we only observe the x variable.

Illustration (Analysis step)

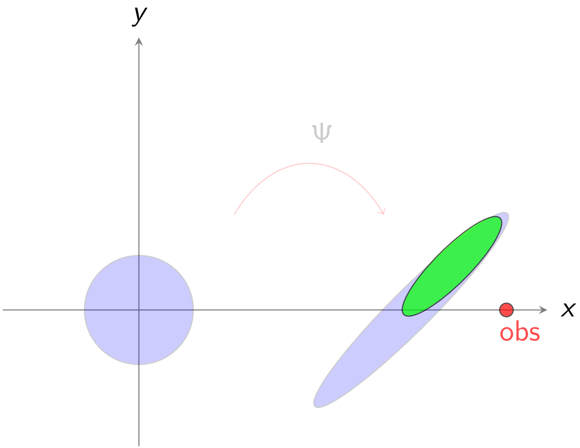


Figure: Using Bayes formula we compute the conditional distribution of $u_1|y_1$. This new measure (called the posterior) is the new estimate of u_1. The uncertainty of the estimate is reduced by incorporating the observation. The forecast distribution steers the update from the observation.

$$\mathbf{P}(u_1|y_1) \propto \mathbf{P}(y_1|u_1)\mathbf{P}(u_1)$$

Bayes' formula

Let $Y_n = \{y_0, y_1, \ldots, y_n\}$. 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

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

In the stochastic (Markovian) case we need to compute the integral

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

Animation 1

Suppose the model is $du = -\nabla V(u)dt + \sigma^{1/2}dW$ with $u = (u_x, u_y)$ and

$$V(x, y) = \frac{1}{2}(1 - x^2 - y^2)^2 \ .$$

We only observe the x-variable

$$y_n = u_x(nh) + \gamma^{1/2}\xi_n$$

with $\xi_n \sim N(0, 1)$ iid.

In geophysical models, we can have $u \in \mathbb{R}^N$ where $N = O(10^9)$. The rigorous Bayesian approach is computationally infeasible.

The Kalman Filter

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

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

where

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

Extended Kalman filter

Suppose we have a nonlinear model:

$$u_{n+1} = \Phi(u_n) + \Sigma^{1/2}\eta_n$$

where Φ is a nonlinear map, η_n Gaussian. The **Extended Kalman filter** is given by the same update formulas

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

where $\widehat{m}_{n+1} = \Phi(m_n)$ and $\widehat{C}_{n+1} = D\Phi(m_n)C_n D\Phi(m_n)^T + \Sigma$.

Thus we approximate the forecast distribution with a Gaussian.

Back to example

Suppose we approximate $du = -\nabla V(u)dt + \sigma^{1/2}dW$ with the discrete time formulation

$$u_{n+1} = \Phi(u_n) + \Sigma^{1/2}\eta_{n+1}$$

with η_n Gaussian.

Then it is easy to compute the Kalman update

$$m_{n+1} = \Phi(m_n) + (\gamma + \widehat{C}_{xx})^{-1}(y_n - \Phi_x(m_n)) \begin{bmatrix} \widehat{C}_{xx} \\ \widehat{C}_{xy} \end{bmatrix}$$

and $\widehat{C}_{n+1} = D\Phi(m_n)C_n D\Phi(m_n)^T + \Sigma$.

Computing $D\Phi(x)$ means evaluating Φ once for each degree of freedom. We want to get away with something cheaper.

Ensemble Kalman filter (Evensen 94)

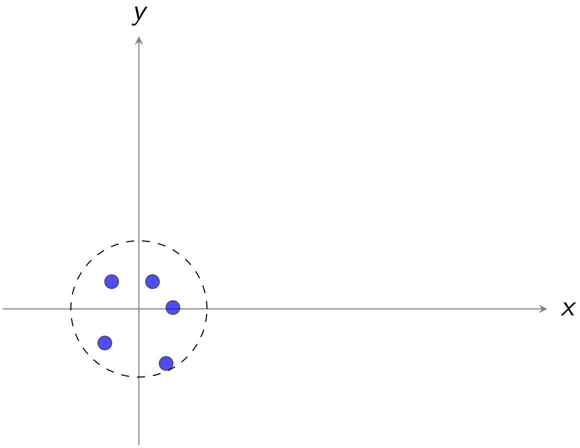


Figure: Start with K ensemble members drawn from some distribution. Empirical representation of u_0. The ensemble members are denoted $v_0^{(k)}$.

Only KN numbers are stored. Better than Kalman if $K < N$.

Ensemble Kalman filter (Forecast step)

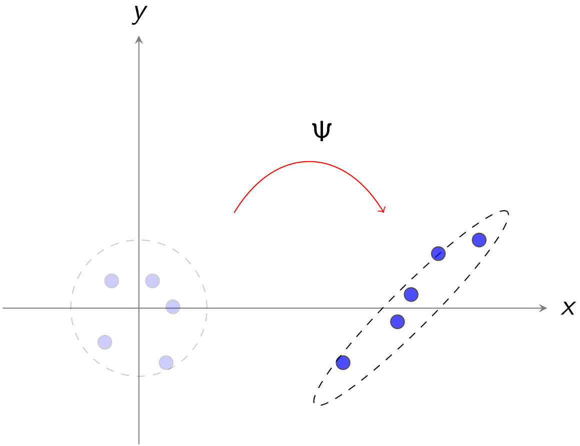


Figure: Apply the dynamics Ψ to each ensemble member.

Ensemble Kalman filter (Make obs)

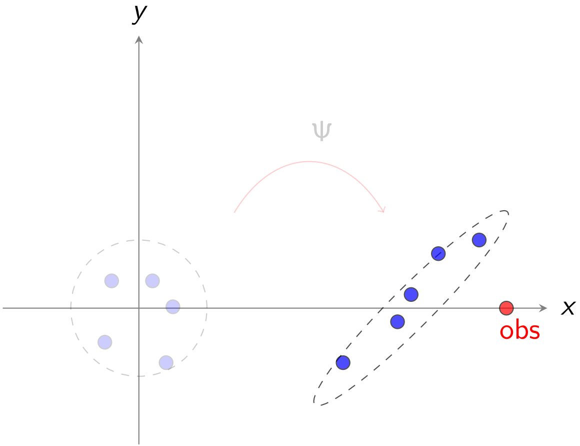


Figure: Make an observation.

Ensemble Kalman filter (Perturb obs)

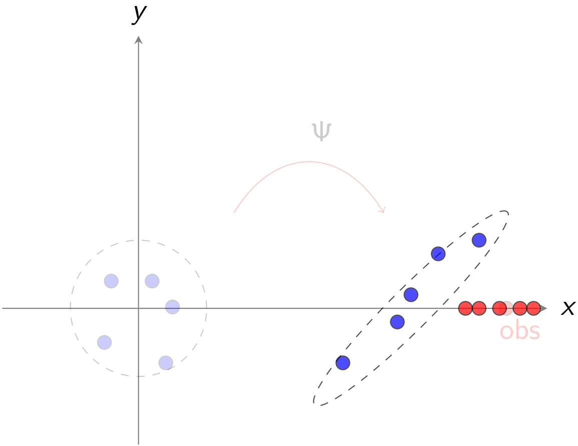


Figure: Turn the observation into K artificial observations by perturbing by the same source of observational noise.

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

Ensemble Kalman filter (Analysis step)

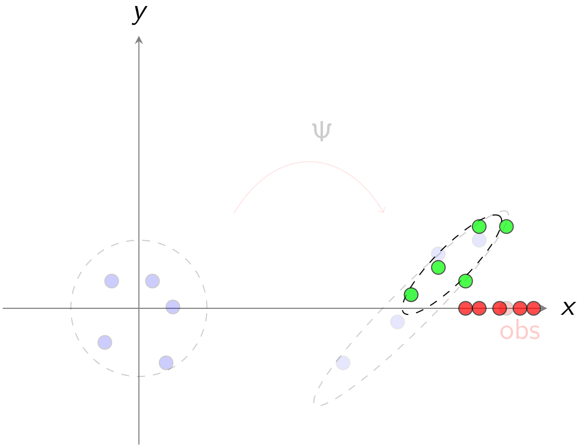


Figure: Update each member using the Kalman update formula. The Kalman gain K_1 is computed using the ensemble covariance.

$$v_1^{(k)} = (1 - K_1 H)\Psi(v_0^{(k)}) + K_1 H y_1^{(k)} \quad K_1 = \widehat{C}_1 H^T (\Gamma + H \widehat{C}_1 H^T)^{-1}$$

$$\widehat{C}_1 = \frac{1}{K-1} \sum_{k=1}^{K} (\Psi(v_0^{(k)}) - \overline{\Psi(v_0)})(\Psi(v_0^{(k)}) - \overline{\Psi(v_0)})^T$$

Ensemble Kalman filter

The conditional distribution is represented **empirically** using an ensemble $\{v_n^{(k)}\}_{k=1}^K$.

When an observation is made, it is perturbed by an iid copy of the observational noise

$$y_{n+1}^{(k)} = y_{n+1} + \xi_{n+1}^{(k)} \; .$$

Each ensemble member is updated using the 'Kalman update' formula

$$v_{n+1}^{(k)} = (1 - K_{n+1} H)\Psi(v_n^{(k)}) + K_{n+1} H y_{n+1}^{(k)}$$

and the Kalman gain is computed using the ensemble covariance

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

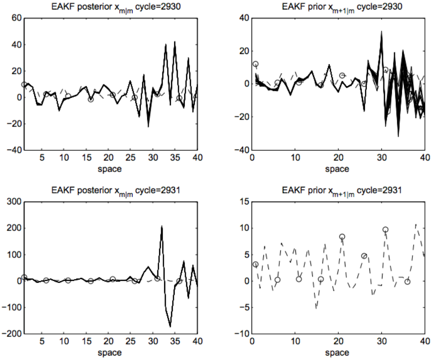
There are many **good** justifications for this algorithm:

- When the model is linear and K is large, the ensemble members are exact samples from the conditional distribution (Monte Carlo Kalman filter).

But there are no **great** justifications ...

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!

For complicated models, only heuristic arguments offered as explanation.

Can we **prove** *it for a simpler constructive model?*

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)

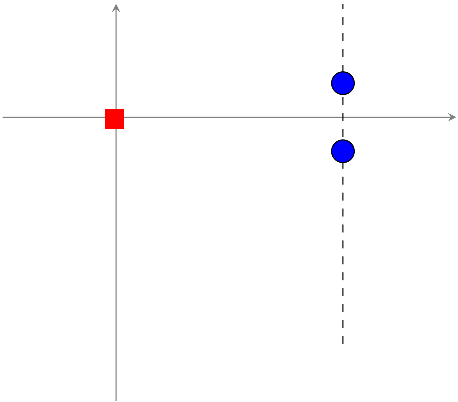


Figure: The red square is the trajectory $u_n = 0$. The blue dots are the positions of the forecast ensemble $\Psi(v_0^+)$, $\Psi(v_0^-)$. Given the locking mechanism in Ψ, this is a natural configuration.

(b)

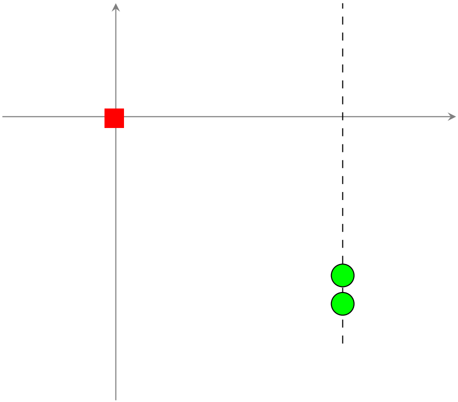


Figure: We make an observation (H shown below) and perform the analysis step. The green dots are v_1^+, v_1^-.

$$H = \begin{pmatrix} 1 & 0 \\ \varepsilon^{-2} & 1 \end{pmatrix} \quad y_1 = (\xi_{1,x}, \xi_{1,y} + \varepsilon^{-2}\xi_{1,x})$$

$$v_1^{\pm} \approx (\hat{x}, \pm\varepsilon - 2\hat{x}/(1+2\varepsilon^2))$$

(c)

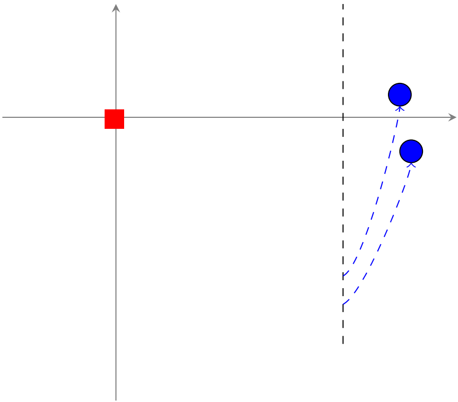


Figure: Beginning the next assimilation step. Apply Ψ_{rot} to the ensemble (blue dots)

(d)

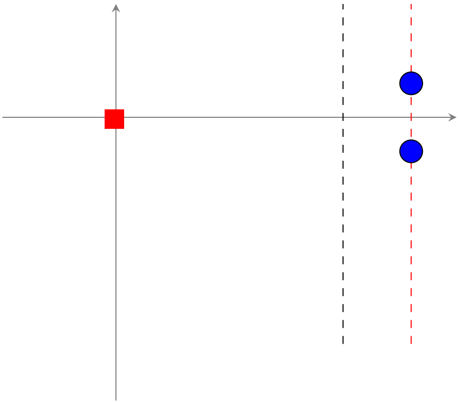


Figure: Apply Ψ_{lock}. The blue dots are the forecast ensemble $\Psi(v_1^+)$, $\Psi(v_1^-)$. Exact same as frame 1, but higher energy orbit. The cycle repeats leading to **exponential growth**.

Theorem (K.-Majda-Tong 15 PNAS)

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

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

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.

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** (2015).

4 - X. Tong, A. Majda & D. Kelly. *Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation*. To appear in **Comm. Math. Sci.** (2015).

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