Variational data assimilation and the ensemble Kalman filter

Amos S. Lawless
Data Assimilation Research Centre
University of Reading
a.s.lawless@reading.ac.uk
@amoslawless
Summary from lecture 1

Given

- a prior estimate of the state $x_b$ with error covariance matrix $P$
- observations $y$ with error covariance matrix $R$

our best estimate of the state can be found by minimising

$$J(x) = (x - x_b)^T P^{-1} (x - x_b) + (y - H(x))^T R^{-1} (y - H(x))$$

If $H$ is linear (i.e., $H$) then this has exact solution

$$x = x_b + P^T H^T (HPH^T + R)^{-1}(y - Hx_b)$$
Variational data assimilation – the idea

In variational data assimilation we seek the solution that maximises the \emph{a posteriori} probability $p(x|y)$.

Since

$$ p(x|y) \propto \exp\left\{ -\frac{1}{2} \{(x - x^b)^T B^{-1} (x - x^b) + (H(x) - y)^T R^{-1} (H(x) - y)\} \right\} $$

we will have the maximum probability when $x$ minimises

$$ J(x) = \frac{1}{2} (x - x^b)^T B^{-1} (x - x^b) + \frac{1}{2} (H(x) - y)^T R^{-1} (H(x) - y) $$
We consider two main algorithms

- Three-dimensional variational assimilation (3D-Var)
  - Where we consider 3 space dimensions.

- Four-dimensional variational assimilation (4D-Var)
  - Where we consider 3 space dimensions plus time as the 4th dimension.
  - In this case we can consider the observation operator to include the dynamical model.

We will present 4D-Var first and 3D-Var as a variant of this.
Four-dimensional variational assimilation (4D-Var)

Aim: Find the best estimate of the true state of the system (analysis), consistent with both observations distributed in time and the system dynamics.
4D-Var cost function

Minimize

\[ J(x_0) = \frac{1}{2}(x_0 - x^b)^T B^{-1}(x_0 - x^b) + \frac{1}{2} \sum_{i=0}^{N} (H_i(x_i) - y_i)^T R_i^{-1}(H_i(x_i) - y_i) \]

with respect to \( x_0 \), subject to

\[ x_{i+1} = M_i(x_i), \]

\( x^b \) - a priori (background) state – Size of order \( 10^8 - 10^9 \)

\( y_i \) - Observations – Size of order \( 10^6 - 10^7 \)

\( H_i \) - Observation operator

\( B \) - Background error covariance matrix

\( R_i \) - Observation error covariance matrix

For \( N=0 \) we have 3D-Var
Numerical minimization - Gradient descent methods

Iterative methods, where each successive iteration is based on the value of the function and its gradient at the current iteration.

\[ x_0^{(k+1)} = x_0^{(k)} - \alpha \varphi(x_0^{(k)}) \]

where \( \alpha \) is a step length and \( \varphi \) is a direction that depends on \( J(x_0^{(k)}) \) and its gradient.

**Problem:** How do we calculate the gradient of \( J(x_0^{(k)}) \) with respect to \( x_0^{(k)} \)?
We introduce Lagrange multipliers $\lambda_i$ at time $t_i$ and define the Lagrangian

$$
\mathcal{L}(x_i, \lambda_i) = \mathcal{J}(x_0) + \sum_{i=0}^{N-1} \lambda_{i+1}^T(x_{i+1} - \mathcal{M}_i(x_i))
$$

Then necessary conditions for a minimum of the cost function subject to the constraint are found by taking variations with respect to $\lambda_i$ and $x_i$.

Variations with respect to $\lambda_i$ simply give the original constraint.
Variations with respect to $x_i$ give the adjoint equations

$$
\lambda_i = M_i^T \lambda_{i+1} - H_i^T R_i^{-1} (H_i(x_i) - y_i)
$$

with boundary condition $\lambda_{N+1} = 0$.

Then at initial time we have

$$
\nabla \mathcal{J}(x_0) = -\lambda_0 + B^{-1} (x_0 - x^b).
$$
An aside – What are the linear operators $H$ & $M$?

Suppose we observe the wind speed $w_s$.

Then we have $x = \begin{pmatrix} u \\ v \end{pmatrix}$, $y = w_s$ and $y = H(x)$

with

$$H(x) = \sqrt{u^2 + v^2}$$

Then

$$H = \begin{pmatrix} \frac{\partial H}{\partial u} & \frac{\partial H}{\partial v} \\ \sqrt{u^2 + v^2} & \sqrt{u^2 + v^2} \end{pmatrix} = \begin{pmatrix} \frac{u}{\sqrt{u^2 + v^2}} & \frac{v}{\sqrt{u^2 + v^2}} \end{pmatrix}$$
So where have we got to?

We wish to minimize

\[ J(x_0) = \frac{1}{2}(x_0 - x^b)^T B^{-1}(x_0 - x^b) + \frac{1}{2} \sum_{i=0}^{N} (H_i(x_i) - y_i)^T R_i^{-1}(H_i(x_i) - y_i) \]

with respect to \( x_0 \), subject to

\[ x_{i+1} = M_i(x_i). \]

On each iteration we have to calculate \( J \) and its gradient

- To calculate \( J \) we need to run the nonlinear model
- To calculate the gradient of \( J \) we need one run of the adjoint model (backward in time)
BUT this can be computationally expensive!
Incremental 4D-Var

Observation
Analysis
Background $x_0$

Time
Incremental 4D-Var

We solve a series of linearized minimization problems

\[
\tilde{J}(k)[\delta x_0^{(k)}] = \frac{1}{2} (\delta x_0^{(k)} - [x^b - x_0^{(k)}])^T B^{-1} (\delta x_0^{(k)} - [x^b - x_0^{(k)}]) \\
+ \frac{1}{2} \sum_{i=0}^{N} (H_i \delta x_i^{(k)} - d_i^{(k)})^T R_i^{-1} (H_i \delta x_i^{(k)} - d_i^{(k)})
\]

with

\[
d_i = y_i - \mathcal{H}_i[x_i^{(k)}] \\
\delta x_{i+1} = M_i \delta x_i
\]

and update using

\[
x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}
\]
Comments on incremental formulation

- Inner loop cost function is linear quadratic, so has a unique minimum.
- Can simplify the linear model (low resolution, simplified physics) in order to save computational time.
- Used in several operational centres, including ECMWF and Met Office.
3D-FGAT (First guess at appropriate time)

We solve a series of linearized minimization problems

\[
\tilde{J}^{(k)}[\delta x_0^{(k)}] = \frac{1}{2} (\delta x_0^{(k)} - [x^b - x_0^{(k)}])^T B^{-1} (\delta x_0^{(k)} - [x^b - x_0^{(k)}]) \\
+ \frac{1}{2} \sum_{i=0}^{N} (H_i \delta x_i^{(k)} - d_i^{(k)})^T R_i^{-1} (H_i \delta x_i^{(k)} - d_i^{(k)})
\]

with

\[
d_i = y_i - H_i [x_i^{(k)}] \\
\delta x_{i+1} = M_i \delta x_i
\]

and update using

\[
x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}
\]

Replace this equation with

\[
\delta x_{i+1} = \delta x_i
\]
Properties of 4D-Var

• Observations are treated at correct time.
• Use of dynamics means that more information can be obtained from observations.
• Covariances are implicitly evolved.
• Standard formulation assumes model is perfect. Weak-constraint 4D-Var being developed to relax this assumption.
• In practice development of linear and adjoint models may be complex, but can be done at level of code.
Ensemble Kalman filter
The basic idea

- In the Kalman filter we assimilate the observations sequentially, making use of the equation we found in the first lecture.

\[ x = x_b + P^T H^T (HPH^T + R)^{-1} (y - Hx_b) \]

- The background state comes from the forecast of the previous analysis.

- In the Kalman filter, the uncertainty on the background comes from a forecast of the uncertainty on the analysis.
Framework

• We assume a linear model and observation operator.

• The model may be imperfect.

\[
x_{i+1}^t = M_i x_i^t + \eta_i
\]

with

\[
\langle \eta_i \rangle = 0
\]

and

\[
\langle \eta_i \eta_i^T \rangle = Q_i
\]
Kalman filter - Illustration

- Observation
- Forecast
- Analysis
We have the following steps:

- Kalman gain computation

\[ K_i = P_i^f H_i^T (H_i P_i^f H_i^T + R_i)^{-1} \]

- State analysis

\[ x_i^a = x_i^f + K_i (y_i - H_i x_i^f) \]

- Error covariance of analysis

\[ P_i^a = (I - K_i H_i) P_i^f \]
• State forecast

\[ x_i^f = M_{i-1} x_{i-1}^a \]

• Error covariance forecast

\[ P_i^f = M_{i-1} P_{i-1}^a M_{i-1}^T + Q_{i-1} \]
Ensemble Kalman filter

In the ensemble Kalman filter (EnKF) the error covariance forecast is approximated by an ensemble of model runs.

Uncertainty at analysis time

Uncertainty at forecast time with covariance $\mathbf{P}$ (Gaussian)
There are 2 main types of EnKF:

I. Perturbed observation EnKF
II. Deterministic EnKF

We consider both of them.
I. Perturbed observation EnKF

Prediction step:

1. Evolve each ensemble member using the nonlinear model

\[ x^{(i),f} = M(x^{(i),a}) + \eta \]

\[ \eta \sim \mathcal{N}(0, Q) \]

2. Form the ensemble mean

\[ \overline{x^f} = \frac{1}{N} \sum_{i=1}^{N} x^{(i),f} \]
3. Form the perturbation matrix and reconstruct the forecast error covariance matrix

$$X_f = \frac{1}{\sqrt{N-1}} [x^{(1),f} - \overline{x^f}, x^{(2),f} - \overline{x^f}, \ldots, x^{(N),f} - \overline{x^f}]$$

$$P^f = X_f (X_f)^T$$
Analysis step:
Update each ensemble member, perturbing the observations

\[ x^{(i),a} = x^{(i),f} + K(y + \epsilon_y - Hx^{(i),f}) \]

with

\[ \epsilon_y \sim \mathcal{N}(0, R) \]

and

\[ K = P^f H^T (H P^f H^T + R_e)^{-1} \]
Notes

• Perturbing the observations is necessary to ensure

\[ P^a = (I - KH)P^f \]

• However this introduces extra sampling noise.

• There is also the problem of needing to invert a low rank matrix, for example using the pseudo-inverse.
II. Deterministic EnKF

The idea here is to create an ensemble consistent with

\[ P^a = (I - KH)P^f \]

The prediction step is the same as for the perturbed observation EnKF.

The analysis step then proceeds as follows:
Analysis step:
1. Transform the forecast ensemble to observation space
   \[ y^{(i),f} = H x^{(i),f} \]
2. Compute the mean \( \overline{y^f} \) and a perturbation matrix \( Y^f \)
3. Compute the analysis
   \[ \overline{x^a} = \overline{x^f} + K (y - \overline{y^f}) \]
   with
   \[ K = X^f (Y^f)^T (Y^f (Y^f)^T + R)^{-1} \]
   and
   \[ X^a = X^f T \]
The matrix $T$ is chosen such that

$$
\begin{align*}
P^a &= X^a(X^a)^T = (X^f T)(X^f T)^T \\
&\approx (I - KH)P^f
\end{align*}
$$

Different choices of $T$ lead to different versions of the EnKF.
EnKF issues

The small ensemble size relative to the size of the system leads to 2 problems that must be faced:

1. The ensemble collapses, i.e. the matrix $P_f$ does not contain enough spread.

**Solution:** Covariance inflation

$$P_f^f = (1 + \rho)P_e^f$$
EnKF issues

2. The ensemble covariance matrix $P_f$ is low rank, which leads to spurious long-range correlations

Solution: Covariance localization

$$P_f = L \circ P_e$$

where $L$ is a matrix that ensures long-range correlations are zero.
References – Variational methods


References - EnKF


