Nonlinear Data Assimilation and Particle Filters

Peter Jan van Leeuwen
Data Assimilation Ingredients

- Prior knowledge, the Stochastic model:
  \[ x^n = f(x^{n-1}) + \beta^{n-1} \]

- Observations:
  \[ y^n \]

- Relation between the two:
  \[ y^n = H(x^n_{\text{truth}}) + \epsilon^n \]

- With \( x^n, x^n_{\text{truth}} \in \mathbb{R}^{N_x} \) and \( y^n \in \mathbb{R}^{N_y} \)
Data assimilation: general formulation

Bayes theorem:

\[ p(x|y) = \frac{p(y|x)}{p(y)} p(x) \]

Solution is pdf!

NO INVERSION !!!
Big Data

• How big is the nonlinear data-assimilation problem?
• Assume we need 10 frequency bins for each variable to build the joint pdf of all variables.
• Let’s assume we have a modest model with a million variables.
• Then we need to store $10^{1000,000}$ numbers.
• The total number of atoms in the universe is estimated to be about $10^{80}$.
• So the data-assimilation problem is larger than the universe...
Nonlinear filtering: Particle filter

\[ p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x) \, dx} \]

Use ensemble

\[ p(x) = \sum_{i=1}^{N} \frac{1}{N} \delta(x - x_i) \]

\[ p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i) \]

with

\[ w_i = \frac{p(y|x_i)}{\sum_j p(y|x_j)} \]

the weights.
What are these weights?

• The weight $w_i$ is the normalised value of the pdf of the observations given model state $x_i$.

• For Gaussian distributed variables is is given by:

$$w_i \propto p(y|x_i)$$
$$\propto \exp \left[ -\frac{1}{2} (y - H(x_i)) R^{-1} (y - H(x_i)) \right]$$

• One can just calculate this value

• That is all !!!
Standard Particle filter
A closer look at the weights

Assume particle 1 is at 0.1 standard deviations $s$ of $M$ independent observations.
Assume particle 2 is at 0.2 $s$ of the $M$ observations.

The weight of particle 1 will be

$$w_1 \propto \exp \left[ -\frac{1}{2} \ (y - H(x_i)) \ R^{-1} \ (y - H(x_i)) \right] = \exp(-0.005M)$$

and particle 2 gives

$$w_2 \propto \exp \left[ -\frac{1}{2} \ (y - H(x_i)) \ R^{-1} \ (y - H(x_i)) \right] = \exp(-0.02M)$$
A closer look at the weights II

The ratio of the weights is

\[
\frac{w_2}{w_1} = \exp(-0.015M)
\]

Take \(M=1000\) to find

\[
\frac{w_2}{w_1} = \exp(-15) \approx 3 \times 10^{-7}
\]

Conclusion: the number of independent observations is responsible for the degeneracy in particle filters.
How can we make particle filters useful?

The joint-in-time prior pdf can be written as:

\[ p(x^n, x^{n-1}) = p(x^n | x^{n-1}) p(x^{n-1}) \]

So the marginal prior pdf at time \( n \) becomes:

\[ p(x^n) = \int p(x^n | x^{n-1}) p(x^{n-1}) \, dx^{n-1} \]

We introduced the transition densities

\[ p(x^n | x^{n-1}) \]
Transition densities $p(x^n | x^{n-1})$

$p(x^n | x^{n-1}) = \delta(x^n - f(x^{n-1}))$

$\begin{align*}
\text{Deterministic model} & \quad \text{Deterministic model} \\

\begin{array}{c}
\text{t} \\
\hline
x^{n-1} \\
\hline
\text{t+1} \\
\hline
x^n = f(x^{n-1}) \\
\end{array}
\end{align*}$
Transition densities $p(x^n | x^{n-1})$

$p(x^n | x^{n-1}) = N(f(x^{n-1}), Q)$

Stochastic model

$t+1$

$f(x^{n-1})$

$t$

$x^{n-1}$
Bayes Theorem now becomes:

\[ p(x^n | y^n) = \frac{p(y^n | x^n)p(x^n)}{p(y)} = \frac{p(y^n | x^n)}{p(y)} \int p(x^n | x^{n-1}) p(x^{n-1}) \, dx^{n-1} \]

We have a set of particles at time \( n-1 \) so we can write

\[ p(x^{n-1}) = \frac{1}{N} \sum_{i=1}^{N} \delta(x^{n-1} - x_{i}^{n-1}) \]

and use this in the equation above to perform the integral:
The magic: the proposal density

Performing the integral over the sum of delta functions gives:

\[
p(x^n | y^n) = \frac{p(y^n | x^n)}{p(y^n)} \frac{1}{N} \sum_{i=1}^{N} p(x^n | x^{n-1}_i)
\]

The posterior is now given as a sum of transition densities. In the standard particle filter we use these to draw particles at time \( n \), which, remember, is running the stochastic model from time \( n-1 \) to time \( n \). We know that is degenerate.

So we introduce another transition density, the proposal.
Optimal proposal density

\[ p(x^n | y^n) = \sum_i \frac{1}{N} \frac{p(y^n | x^n)}{p(y^n)} p(x^n | x_i^{n-1}) \]

Now use the identity

\[ p(y^n | x^n)p(x^n | x_i^{n-1}) = p(x^n | x_i^{n-1}, y^n)p(y^n | x_i^{n-1}) \]

to find

\[ p(x^n | y^n) = \sum_i \frac{1}{N} \frac{p(y^n | x_i^{n-1})}{p(y^n)} p(x^n | x_i^{n-1}, y^n) \]
Optimal proposal density

\[
p(x^n | y^n) = \sum_i \frac{1}{N} \frac{p(y^n | x_i^{n-1})}{p(y^n)} p(x^n | x_i^{n-1}, y^n)
\]

The optimal proposal density generates new particles by drawing from \( p(x^n | x_i^{n-1}, y^n) \) for each \( i \).
This leads to weights

\[
w_i^n \propto p(y^n | x_i^{n-1})
\]

Ades and Van Leeuwen (2012) have shown that \(-\log(\text{var}(w_i))\) is proportional to the number of independent observations \( M \) so this is degenerate for \( M \) large.
Optimal proposal density

Snyder, Bengtsson and Morzfeld (2015) showed that the variance in the weights when using a proposal density is bounded as:

$$\text{var}(w_k^*) \geq -1 + \int \frac{p(x_{k-1} \mid y_k)^2}{p(x_{k-1})} \, dx_{k-1},$$

with the minimum found for the optimal proposal density.

This would be bad news, as the optimal proposal density is not good enough...
Let us explore the relation with Weare and VandenEijnden.
Explore the optimal proposal density

Bayes Theorem can be written as:

\[
p(x^n | y^n) = \frac{p(y^n | x^n)p(x^n)}{p(y)} \times \int p(x^n | x^{n-1})p(x^{n-1}) \, dx^{n-1}
\]

Multiply and divide this expression by a proposal transition density \( q \):

\[
p(x^n | y^n) = \frac{p(y^n | x^n)}{p(y)} \times \int \frac{p(x^n | x^{n-1})}{q(x^n | x^{n-1}, y^n)} q(x^n | x^{n-1}, y^n) \, dx^{n-1}
\]
Exploring the proposal density

Start with the particle description of the conditional pdf at \( n-1 \) (assuming weighted particles):

\[
p(x^{n-1}) = \sum_{i=1}^{N} w_i \delta(x^{n-1} - x_i^{n-1})
\]

Leading to:

\[
p(x^n | y^n) = \sum_i w_i^{n-1} \frac{p(y^n | x_n)}{p(y^n)} \frac{p(x^n | x_i^{n-1})}{q(x^n | x_i^{n-1}, y^n)} q(x^n | x_i^{n-1}, y^n)
\]

Now use again:

\[
p(y^n | x^n) p(x^n | x_i^{n-1}) = p(x^n | x_i^{n-1}, y^n) p(y^n | x_i^{n-1})
\]
Exploring the proposal density

To find:

\[
p(x^n | y^n) = \sum_i w_i^{n-1} \frac{p(y^n | x_i^{n-1})}{p(y^n)} \frac{p(x^n | x_i^{n-1}, y^n)}{q(x^n | x_i^{n-1}, y^n)} q(x^n | x_i^{n-1}, y^n)
\]

Now assume we make a draw from the proposal density. The weights are then given by:

\[
w_i^n = w_i^{n-1} \frac{p(y^n | x_i^{n-1})}{p(y^n)} \frac{p(x^n_i | x_i^{n-1}, y^n)}{q(x^n_i | x_i^{n-1}, y^n)}
\]

Ideally all weights are equal to 1/N so let’s enforce that, to find for the proposal:
Exploring the proposal density

\[
q(x^n_i | x^{n-1}_i, y^n) = N w^{n-1}_i \frac{p(y^n | x^{n-1}_i)}{p(y^n)} p(x^n_i | x^{n-1}_i, y^n)
\]

We want to have no constants in front of the density in \(x^n_i\) so the weight at previous time should be:

\[
\omega^{n-1}_i = \frac{1}{N} \frac{p(y^n)}{p(y^n | x^{n-1}_i)}
\]

This weight is set by the proposal density for time \(n-1\).
The proposal density at time $n-1$

The weight at time $n-1$ is related to the proposal density via:

$$w_{i}^{n-1} = w_{i}^{n-2} \frac{p(x_{i}^{n-1} | x_{i}^{n-2})}{q(x_{i}^{n-1} | x_{i}^{n-2}, y^{n})}$$

Equating this to:

$$w_{i}^{n-1} = \frac{1}{N} \frac{p(y^{n})}{p(y^{n} | x_{i}^{n-1})}$$

We find for the proposal density:

$$q(x_{i}^{n-1} | x_{i}^{n-2}, y^{n}) = N w_{i}^{n-2} \frac{p(y^{n} | x_{i}^{n-1})p(x_{i}^{n-1} | x_{i}^{n-2})}{p(y^{n})}$$
The proposal density at time \( n-1 \)

Using again:

\[
p(y^n | x_i^{n-1}) \ p(x_i^{n-1} | x_i^{n-2}) = p(x_i^{n-1} | x_i^{n-2}, y^n) \ p(y^n | x_i^{n-2})
\]

we find for the proposal density:

\[
q(x_i^{n-1} | x_i^{n-2}, y^n) = Nw_i^{n-2} \ \frac{p(y^n | x_i^{n-2})}{p(y^n)} p(x_i^{n-1} | x_i^{n-2}, y^n)
\]

Again, we don’t want the factors in front so the weight at \( n-2 \) should be:

\[
\omega_i^{n-2} = \frac{1}{N} \ \frac{p(y^n)}{p(y^n | x_i^{n-2})}
\]
So we find the sequence

for the proposal density:

\[
q(x_{i}^{n-m} | x_{i}^{n-m-1}, y^n) = p(x_{i}^{n-m} | x_{i}^{n-m-1}, y^n)
\]

and for the weights:

\[
\omega_{i}^{n-m} = \frac{1}{N} \frac{p(y^n)}{p(y^n | x_{i}^{n-m})}
\]
What does this mean?

The proposal density, for each m:

\[ q(x_i^{n-m} | x_i^{n-m-1}, y^n) = p(x_i^{n-m} | x_i^{n-m-1}, y^n) \]

This is 4DVar-like for one time step:
- Initial condition given,
- Model error has to be included
- We need a draw, not the mode…

And we have to repeat that for each time step until the observation time.
What does this mean?

value of state variable

m-1 m

time
And the weights?

\[ w_{i}^{n-m} = \frac{1}{N} \frac{p(y^{n})}{p(y^{n} | x_{i}^{n-m})} \]

Assume linear model and H and Gaussian observation and model errors:

\[ w_{i}^{n-m} \propto \exp \left( -\frac{1}{2} (y^{n} - HF^{n-m:n} x_{i}^{n-m})^T C^{-1} (y^{n} - HF^{n-m:n} x_{i}^{n-m}) \right) \]

with

\[ C = \sum_{j=n-m}^{n} HQ_{j} H^{T} + R \]

So these weights become similar as \( m \) grows!
So?

We have generated a particle filter with 4DVar-like proposal densities for each time step and weights converging to each other when the number of time steps between observations $m$ is large. Large means

$$\frac{mHQ_jH^T}{B^{n-m}} \approx M$$

Note that weights are equal if all particles start from one single particle at time $n-m$.

$$w_{i}^{n-m} = \frac{1}{N} \frac{p(y^n)}{p(y^n | x_i^{n-m})}$$
Optimalised optimal proposal density

\[
p(x^n | y^n) = \sum_i \frac{1}{N} \frac{p(y^n | x_{i}^{n-1})}{p(y^n)} p(x^n | x_{i}^{n-1}, y^n)
\]

However, we can see this as a mixture density and we can draw from this density by first drawing from the mixture coefficients:

\[
\frac{1}{N} \frac{p(y^n | x_{i}^{n-1})}{p(y^n)}
\]

And then make a draw from the chosen density in the mixture.

*In this way the filter is never degenerate, even if we need to draw all particles from the same density in the mixture!*
Optimalised optimal proposal density

\[ p(x^n | y^n) = \sum_i \frac{1}{N} \frac{p(y^n | x_i^{n-1})}{p(y^n)} p(x^n | x_i^{n-1}, y^n) \]

What we have done is allowing particle \( j \) be drawn from a proposal that originates at particle \( i \), in which \( j \) is not \( i \). This opens up a different class of particle filters that do not necessarily suffer from the degeneracy proved in Snyder et al 2015.

Can we generalise this?. 
The proposal transition density

Multiply numerator and denominator with a proposal density $q$:

$$p(x^n | y^n) = \frac{p(y^n | x^n)}{p(y^n)} \frac{1}{N} \sum_{i=1}^{N} \frac{p(x^n | x_{i}^{n-1})}{q(x^n | x_{1:N}^{n-1}, y^n)} q(x^n | x_{1:N}^{n-1}, y^n)$$

Note that 1) the proposal depends on the future observation, and 2) the proposal depends on all previous particles, not just one.

1) Ensures that the particles end up close to the observations because they know where the observations are.
2) Allows for an equal-weight filter, as the performance bounds suggested by Snyder, Bickel, and Bengtsson do not apply.
What does this all mean?

• The standard Particle Filter propagates the original model by drawing from $p(x^n|x^{n-1})$.

• Now we draw from $q(x^n|x^{n-1}_{1:N}, y^n)$, so we propagate the state using a different model.

• This model can be anything, e.g.

$$x^n = g(x^{n-1}, y^n) + \hat{\beta}^n$$
How are the weights affected?

Draw samples from the proposal transition density $q$, to find:

$$p(x^n | y^n) = \frac{p(y^n | x^n_i)}{p(y^n)} \frac{1}{N} \sum_{i=1}^{N} \frac{p(x^n_i | x^{n-1}_i)}{q(x^n_i | x^{1:N-1}, y^n)} \delta(x^n - x^n_i)$$

Which can be rewritten as:

$$p(x^n | y^n) = \sum_{i=1}^{N} w_i \delta(x^n - x^n_i)$$

with weights

$$w_i = \frac{p(y^n | x^n_i)}{N p(y^n)} \frac{p(x^n_i | x^{n-1}_i)}{q(x^n_i | x^{1:N-1}, y^n)}$$
Example: EnKF as proposal

EnKF update:

\[ x_i^n = x_i^* + K^e (y^n + \epsilon_i - H(x_i^*)) \]

Use model equation:

\[ x_i^n = f(x_i^{n-1}) + \beta_i^n + K^e \left( y^n + \epsilon_i - H \left( f(x_i^{n-1}) + \beta_i^n \right) \right) \]

Regroup terms:

\[ x_i^n = f(x_i^{n-1}) + K^e \left( y^n - H \left( f(x_i^{n-1}) \right) \right) + (1 - K^e H) \beta_i^n + K^e \epsilon_i \]

Leading to:

\[ x_i^n = g(x_i^{n-1}, y^n) + \hat{\beta}_i^n \]
Algorithm

• Generate initial set of particles
• Run proposed model conditioned on next observation
• Accumulate proposal density weights $p/q$
• Calculate likelihood weights
• Calculate full weights and resample
• Note, the original model is never used directly.
Equal-weight Particle filtering

Define an *implicit map* as follows:

\[ x^n_i = x^a_i + \alpha_i \beta_i \]

\( x^a_i \) is the mode of the optimal proposal density, which is given by \( p(x^n | x^{n-1}_i, y^n) \)

\( \beta_i \) is a random draw from the density \( N(0,P) \), with the \( P \) covariance of the optimal proposal density,

\( \alpha_i \) is chosen such that all particles have equal weight (using the expression for the weights)
What is this optimal proposal density?

Using Bayes theorem we find:

\[
p(x^n | x^{n-1}_i, y^n) = \frac{p(y^n | x^n)}{p(y^n | x^{n-1}_i)} p(x^n | x^{n-1}_i)
\]

Assuming Gaussian model errors and Gaussian observation errors this is proportional to

\[
\propto \exp \left[ -\frac{1}{2} (y - H(x^n))^T R^{-1} (y - H(x^n)) - \frac{1}{2} (x^n - f(x^{n-1}_i))^T Q^{-1} (x^n - f(x^{n-1}_i)) \right]
\]

If \( H \) is linear the mean is the maximum and given by:

\[
x^a_i = f(x^{n-1}_i) + QH^T (HQH^T + R)^{-1} (y - Hf(x^{n-1}_i))
\]

If \( H \) is nonlinear this maximum has to be found by iteration.
How to find $\alpha_i$

Remember the new particle is given by:

$$x_i^n = x_i^a + \alpha_i \beta_i$$

in which $x_i^a$ and $\beta_i$ are known. Use this expression in the weights and set all weights equal to a target weight:

$$w_i = \frac{p(y_n | x_i^n) p(x_i^n | x_i^{n-1})}{N p(y^n)} \frac{p(x_i^n | x_i^{n-1})}{q(x_i^n | x_1^{n-1}, y^n)} = w_{\text{target}}$$

and solve for $\alpha_i$. Now all weights are equal by construction!
Experiments, model error and observation errors Gaussian, H linear

- Linear model of Snyder et al. 2008.
- 1000 dimensional independent Gaussian linear model
- 20 particles
- Observations every time step

- Lorenz 1996
- 1000 dimensional
- 20 particles
- Observations every 5 time steps, half of state
Linear model: Rank histogram
1000 time steps, 20 particles
Normalised pdf 1000 time steps
20 particles

![Graph showing a normalised probability density function (PDF) with 1000 time steps and 20 particles, comparing Sample PDF Start From 8 and True PDF of Timestep 8.](image)
Normalised pdf 1000 time steps
1000 particles -> Convergence!

Sample PDF Start From 8
True PDF of Timestep 8
Lorenz 1996  1000 dimensional
20 particles
Rank histograms 10000 time steps
20 particles, land-sea configuration

Observed gridpoint

Unobserved gridpoint
Conclusions

• Large number of ‘nonlinear’ filters and smoothers available
• Best method will be system dependent
• Fully nonlinear equal-weight particle filters for systems with arbitrary dimensions that converge to the true posterior pdf do exist.
• Proposal-density freedom needs further exploration
• Example shown for 1000 dimensional system, but methods have been applied to 2.3 million dimensional climate model too.
A closer look at the weights IV

• The volume of a hyperball of radius $r$ in an $M$ dimensional space is

$$ V \propto \frac{r^M}{\Gamma(M/2 - 1)} $$

• Taking for the radius $r \approx 3\sigma_y$ we find, using Stirling:

$$ V \propto \left[ \frac{9\sigma_y}{M/2} \right]^{M/2} $$

• So very small indeed.
The volume in hyperspace occupied by observations

$\log_{10}$ of Volume of hyperball of radius 1.
How to calculate $p/q$ in the weights?

Let’s assume that the original model has Gaussian distributed model errors:

$$p(x^n|x^{n-1}) = N\left(f(x^{n-1}), Q\right)$$

To calculate the value of this term realise it is the probability of moving from $x_{i}^{n-1}$ to $x_{i}^{n}$. Since $x_{i}^{n}$ and $x_{i}^{n-1}$ are known from the proposed model we can calculate directly:

$$p(x_{i}^{n}|x_{i}^{n-1}) \propto \exp \left[ -\frac{1}{2} \left(x_{i}^{n} - f(x_{i}^{n-1})\right)^T Q^{-1} \left(x_{i}^{n} - f(x_{i}^{n-1})\right) \right]$$
Example calculation of $p$

• Assume the proposed model is

$$x^n = f(x^{n-1}) + \hat{\beta}^n + K (y^n - H(x^{n-1}))$$

• Then we find

$$p(x^n_i|x_i^{n-1}) \propto \exp \left[ -\frac{1}{2} \left( K(y^n - H(x_i^{n-1})) + \hat{\beta}^n \right)^T Q^{-1} \left( K(y^n - H(x_i^{n-1})) + \hat{\beta}^n \right) \right]$$

• We know all the terms, so this can be calculated
And $q$ ...

- The deterministic part of the proposed model is:

\[ x^n = f(x^{n-1}) + K \left( y^n - H(x^{n-1}) \right) \]

- So the probability becomes

\[ q(x^n | x_{1:N}^{n-1}, y^n) \propto \exp \left[ -\frac{1}{2} \beta_i^T Q^{-1} \beta_i \right] \]

- We did draw the stochastic terms, so we know what they are, so this term can be calculated too.
The weights

• We can calculate $p/q$ and we can calculate the likelihood so we can calculate the weights:

$$w_i = \frac{p(y^n|x^n_i)}{Np(y^n)} \cdot \frac{p(x^n_i|x^{n-1}_i)}{q(x^n_i|x^{n-1}_{1:N}, y^n)}$$