Nonlinear ensemble data assimilation in high-dimensional spaces

Peter Jan van Leeuwen

Javier Amezcua, Mengbin Zhu, Melanie Ades

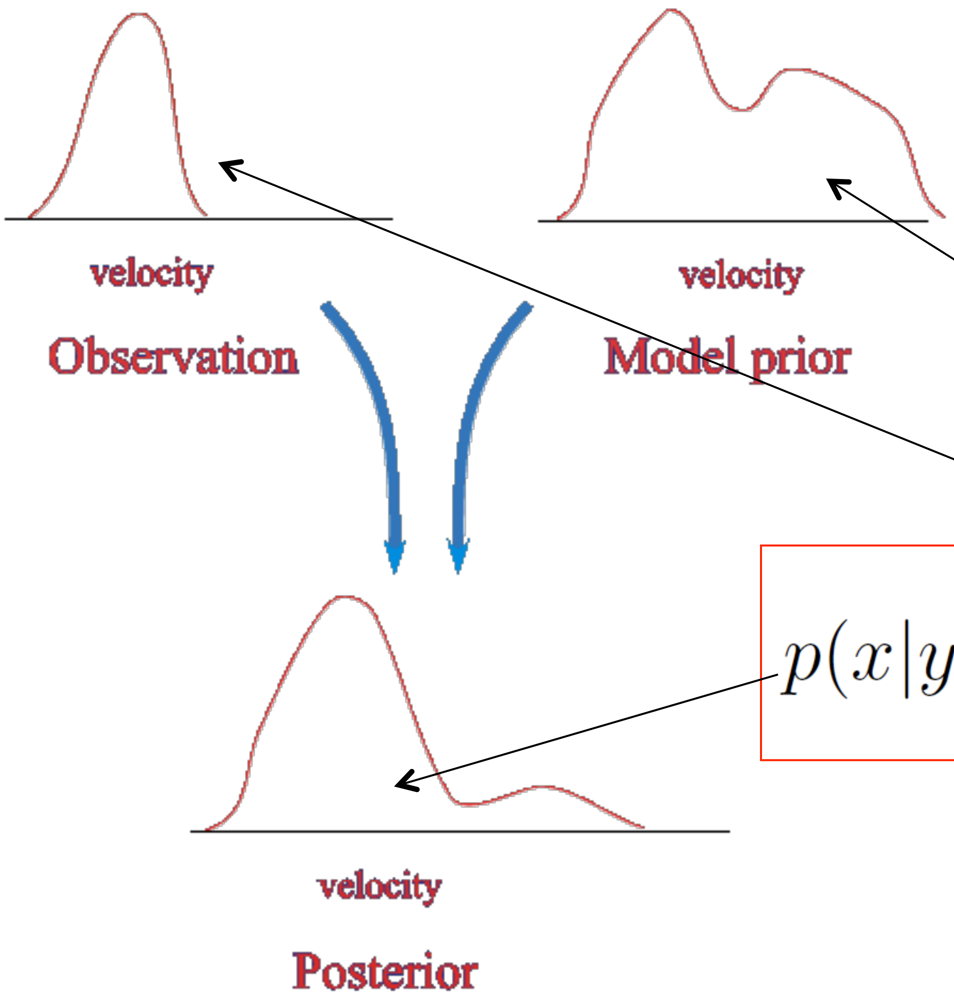
Data Assimilation Research Centre

European Research Council
Established by the European Commission

University of Reading

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Data assimilation: general formulation



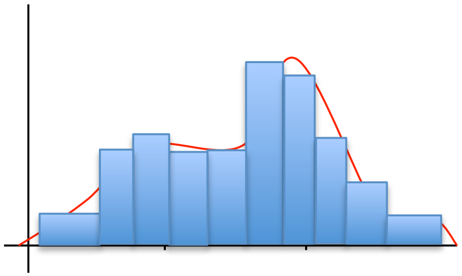
Bayes theorem:

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\; dx}$$

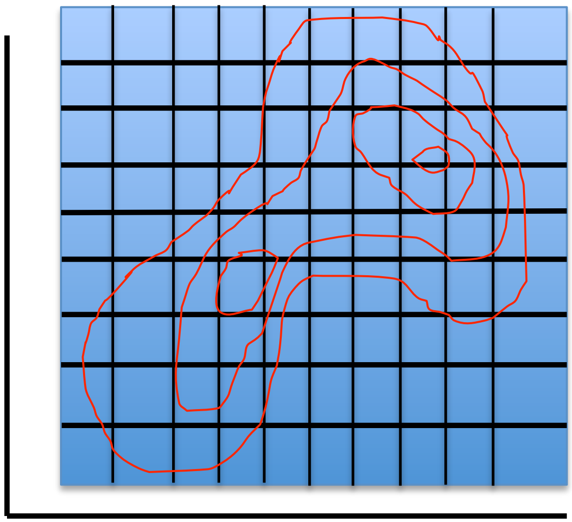
The solution is a pdf!

No inversion!

How big is the Data-Assimilation problem?



Store 10 numbers



Store 100 numbers

A model of 1,000,000 variables need storage of $10^{1,000,000}$ numbers

Estimated number of atoms in the whole universe 10^{80}…

The data assimilation problem is larger then the universe !

Data-assimilation is a problem of finding the best approximation

Nonlinear filtering: Particle filter

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\ dx}$$

Use ensemble

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

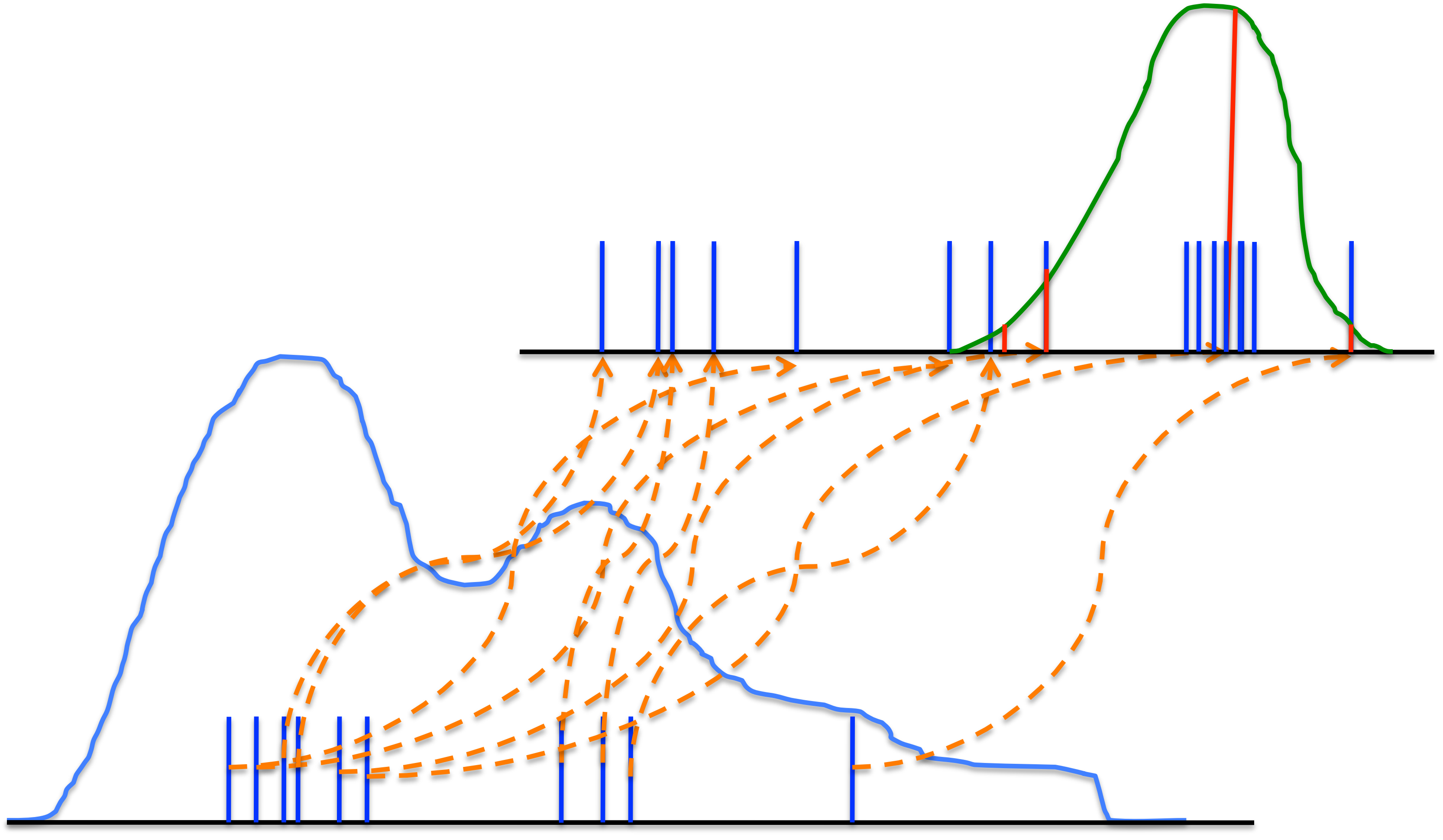
$$p(x|y) = \sum_{i=1}^{N} w_i \delta(x - x_i)$$

with

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

the weights.

Standard Particle filter



How to make particle filters useful?

1. Introduce localisation to reduce the number of observations.
2. Combine Particle Filters and Ensemble Kalman Filters or Gaussian Mixtures
3. Use proposal-density freedom.

3. Exploring the proposal density freedom

The evolution equation for the prior pdf can be written as:

$$p(x^n) = \int p(x^n|x^{n-1})p(x^{n-1})\ dx^{n-1}$$

Use this in Bayes Theorem to find:

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

Now consider the particles at time *n-1* :

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

Bayes Theorem and the proposal density

to find for the posterior pdf:

$$p(x^n|y^n) = \frac{p(y^n|x^n)}{p(y^n)} \frac{1}{N} \sum_{i=1}^{N} p(x^n|x_i^{n-1})$$

Now use

$$p(y^n|x^n)p(x^n|x_i^{n-1}) = p(y^n|x_i^{n-1})p(x^n|x_i^{n-1}, y^n)$$

To find:

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

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

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

One can show that the least degenerate proposal of the form $q(x^n|x_i^{n-1}, y^n)$ is the optimal proposal.

A better proposal density

So, Particle Filters can never work? They can, it is easy to come up with a counter example. Recall the posterior

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

This can be seen as a so-called mixture density, a weighted sum of densities. To draw directly from that density:

1. Draw *i* from the weight distribution w_i.
2. Say we draw *i=8.* Then draw a sample from $p(x^n|x_8^{n-1}, y^n)$
3. Do this *N times.*

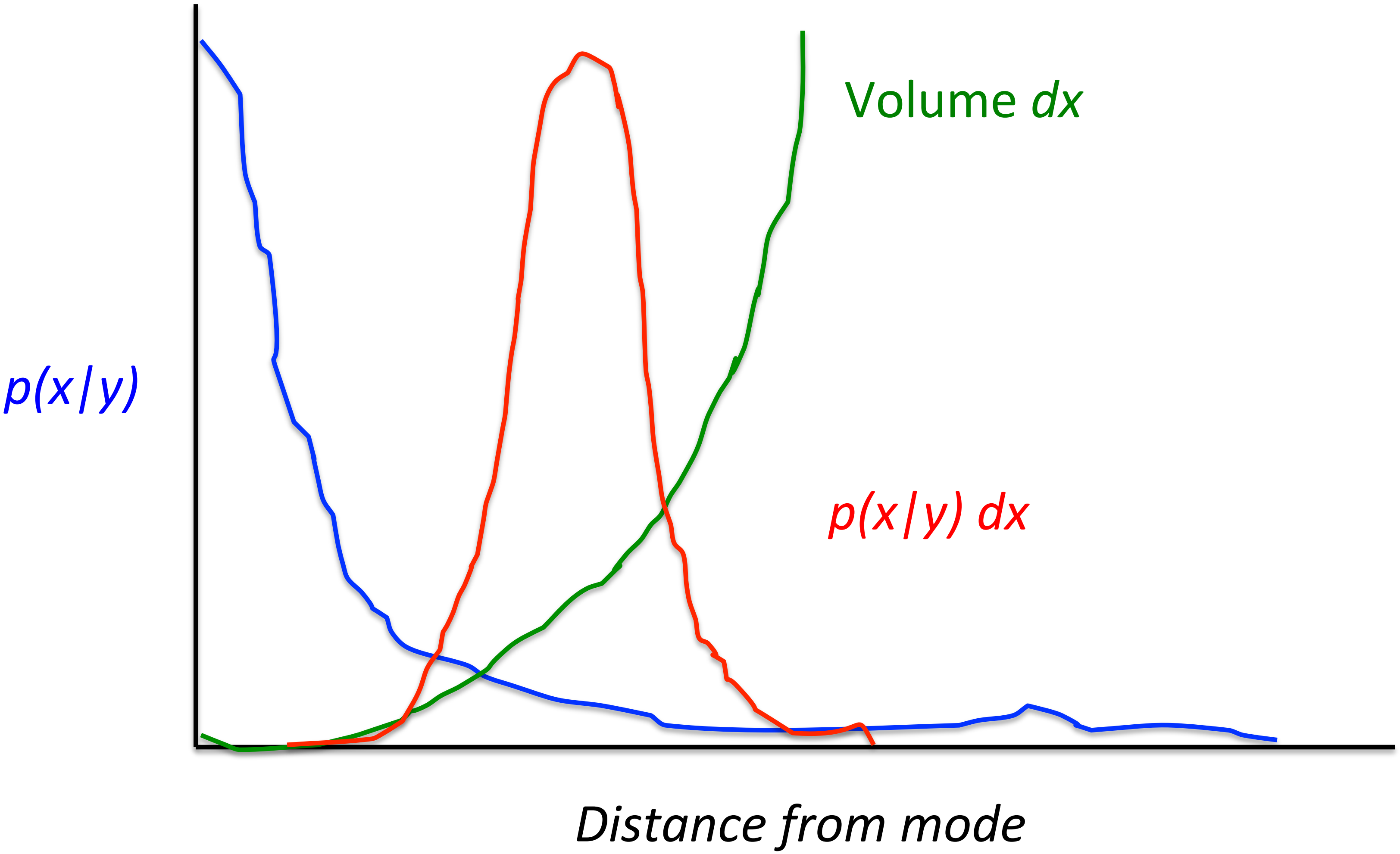
This PF has equal weights on the particles, but is not efficient for low *N.*

How to save the Particle Filter for low N?

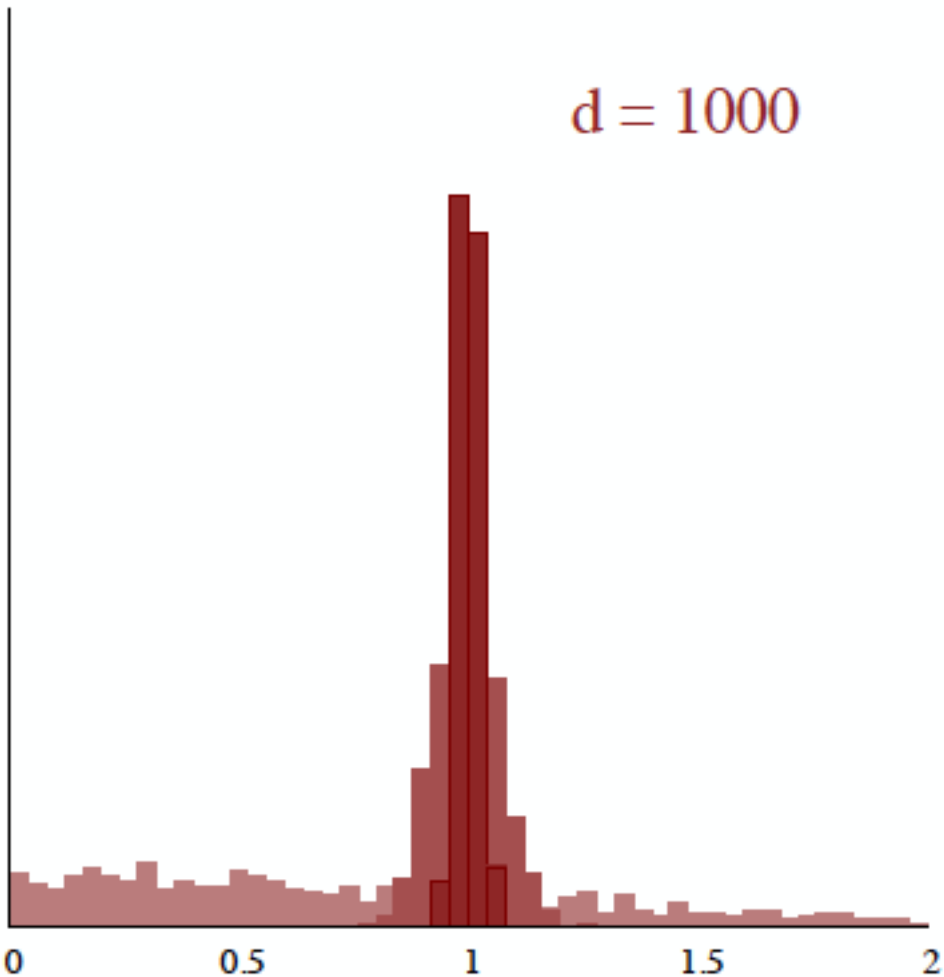
For these high dimensional systems interaction among particles is essential.

- Interaction through resampling is not strong enough as the weights will be degenerate..
- Extended-space proposals densities that enforce equal weights via stronger interactions can perhaps save the particle filter.
- Try to understand these methods and improve on them explore the typical set and Hamiltonian Monte Carlo.

Typical Set: high probability mass



Probability mass concentrates on a hypersurface surrounding the mode



Distance from centre normalised by sqrt(dimension)

(‘stolen’ from Michael Betancourt)

Is there always a typical set?

For a multivariate standard Gaussian the Central Limit Theorem gives

$$\xi^T \xi = N_x \pm \sqrt{2N_x}$$

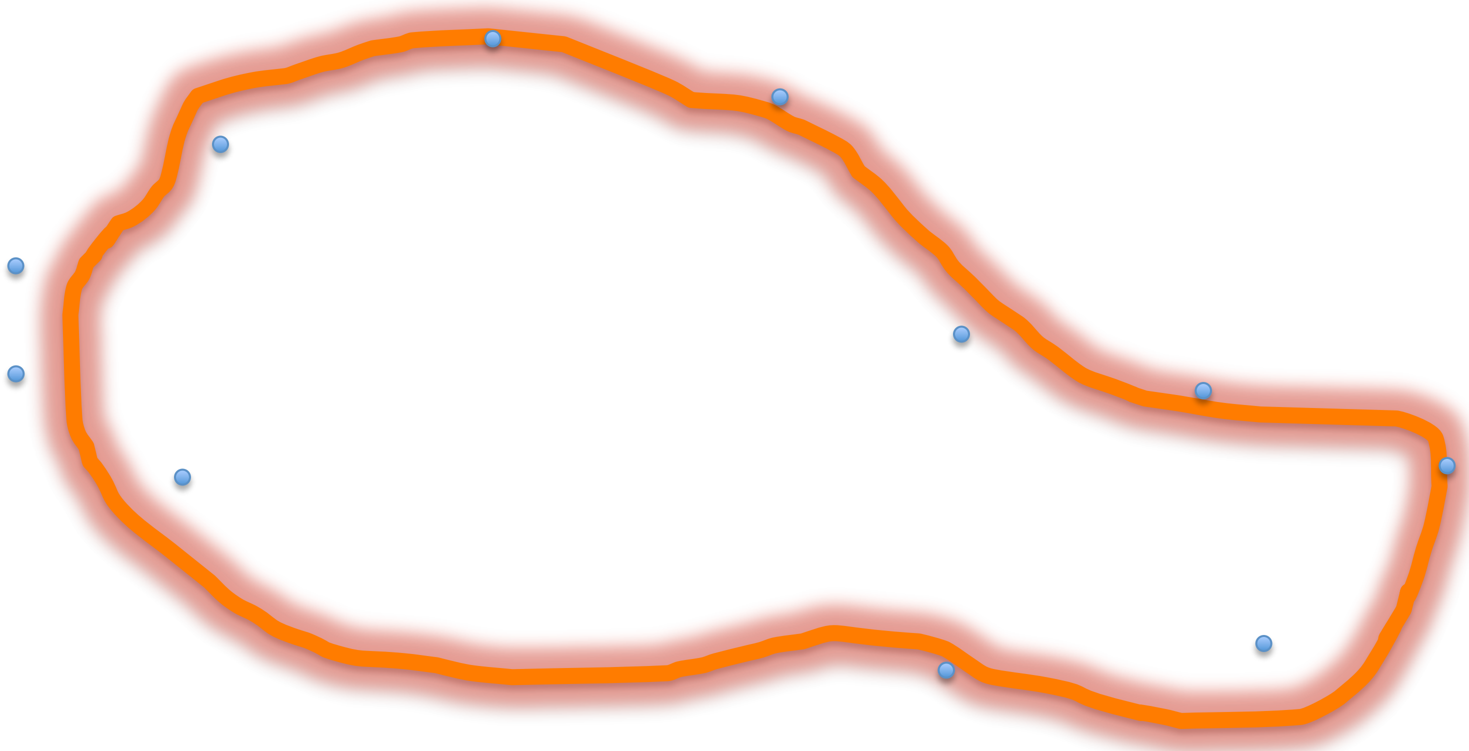
hence

$$|\xi| = \sqrt{N_x} \pm \frac{1}{2}$$

So any particle drawn from the Gaussain will be far away from the mode!

Unclear for 'arbitrarily shaped' pdfs, but intuition and numerical experimentation suggest there is a typical set.

Generate particles close to typical set



But e.g optimal proposal weights vary enormously.

Can we come up with a particle filter that has all particles on typical set and has equal weights ?

Hamiltonian Monte Carlo

Metropolis-Hastings (MH) on an extended space:

- View state as position variable of a physical system
- Introduce velocity variables to move particles in state space
- Form Hamiltonian as

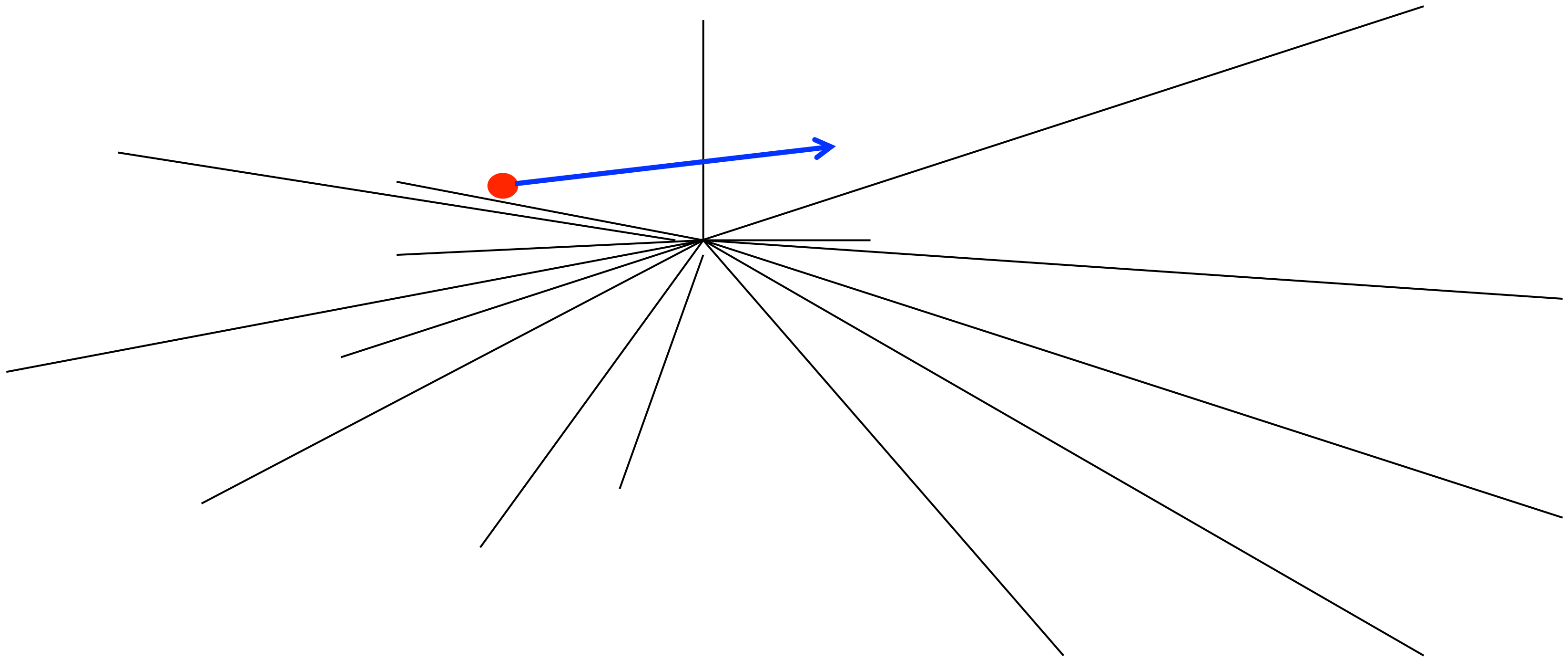
$$p(x, v) = p(x)p(v) \propto \exp\left[-H(x, v)\right]$$

with $H(x, v) = E(x) + K(v)$

and $p(x|y) \propto \exp[-E(x)]$ and $p(v) \propto \exp\left[-\frac{1}{2}v^T v\right]$

- Perform MH on this extended space.
- Because Hamiltonian is conserved almost all moves in this extended space are accepted, even very large moves.

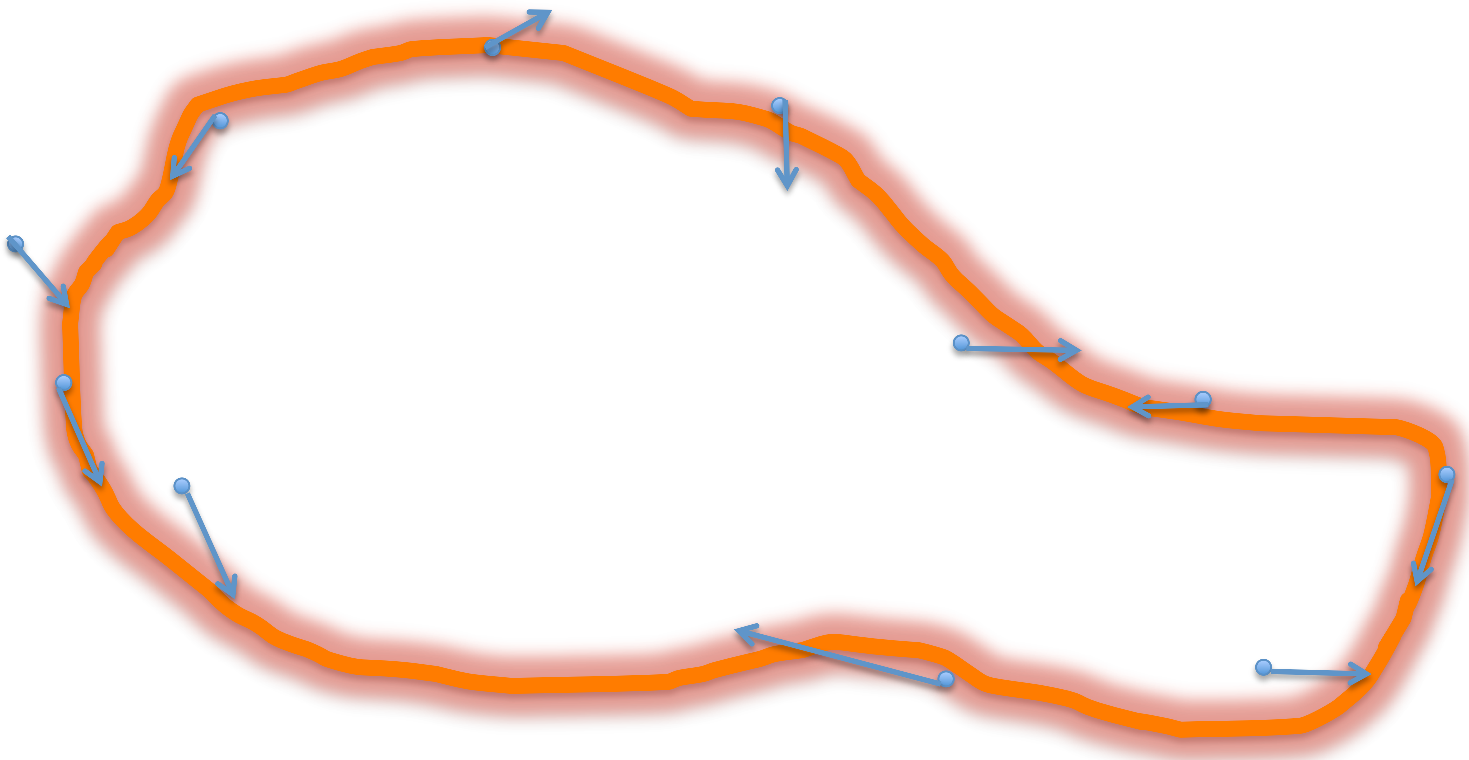
Why does HMC move stay on typical set?



HMC move can be written as $$x_i^n = x_i^* + \epsilon_i P^{1/2} v_i - \frac{\epsilon_i^2}{2}\frac{dE}{dx}$$

v is drawn iid at every step, so in high-dimensional systems the component parallel to gradient is small.
So HMC 'moves around the mode'.

HMC in proposal density of particle filter



1. Use smart particle filter to find particles close to typical set
2. Use HMC to move these particles to equal weight positions on the set.

Extended state 2-step proposal:

This is a proposal on an extended space (space different from HMC!!!):

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

where we just multiplied and divided by a proposal *q(…)* which can depend on *all* previous particles, and with

$$q(x^n x^*|x_{i;1:N}^{n-1}, y^n) = q(x^n|x^*, x_{1:N}^{n-1}) q(x^*|x_i^{n-1}, y^n)$$

This leads to a whole class of particle filters not hampered by classical proofs of degeneracy.

Example non-degenerate PF

The following particle filter results in equal weights but is also efficient for small ensemble sizes.

1. For each *i* draw $x_i^* \sim p(x^n | x_i^{n-1}, y^n)$
2. For each *i* draw $\xi_i \sim N(0, P)$ with $P^{-1} = Q^{-1} + H^T R^{-1} H$
3. For each *i* write $x_i^n = x_i^* + \alpha_i P^{1/2} \xi_i$

And solve for α_i in

$$w_i(\alpha_i) = \frac{p(y|x_i^{n-1}) p(x_i^n | x_i^{n-1}, y^n)}{q(x_i^n x_i^* | x_{i;1:N}^{n-1}, y^n)} = w_{target}$$

The Bias

- The scheme is biased (or inconsistent) because of the target weight construction.
- It is difficult to quantify the bias because the limit $N \to \infty$ is not relevant in practice.
- If the bias is smaller than the sampling error it is of no concern.
- Sampling error estimates are typically not that useful as they tend to be of the form

$$C/\sqrt{N}$$

where *C* is unknown.

- So we have to rely on numerical tests until progress is made on the maths...

Experiments on Lorenz 1963 model

- 10,000 independent models Lorenz 1963 models
- 30,000 variables, 10,000 parameters
- 10 particles
- Observations:
  every 20 time steps,
  first two variables
- Observation errors Gaussian
- HMC step up to $O(\epsilon)$

Sequential parameter estimation

- SPDE $$x^n = f(x^{n-1}, \theta) + \beta^n$$

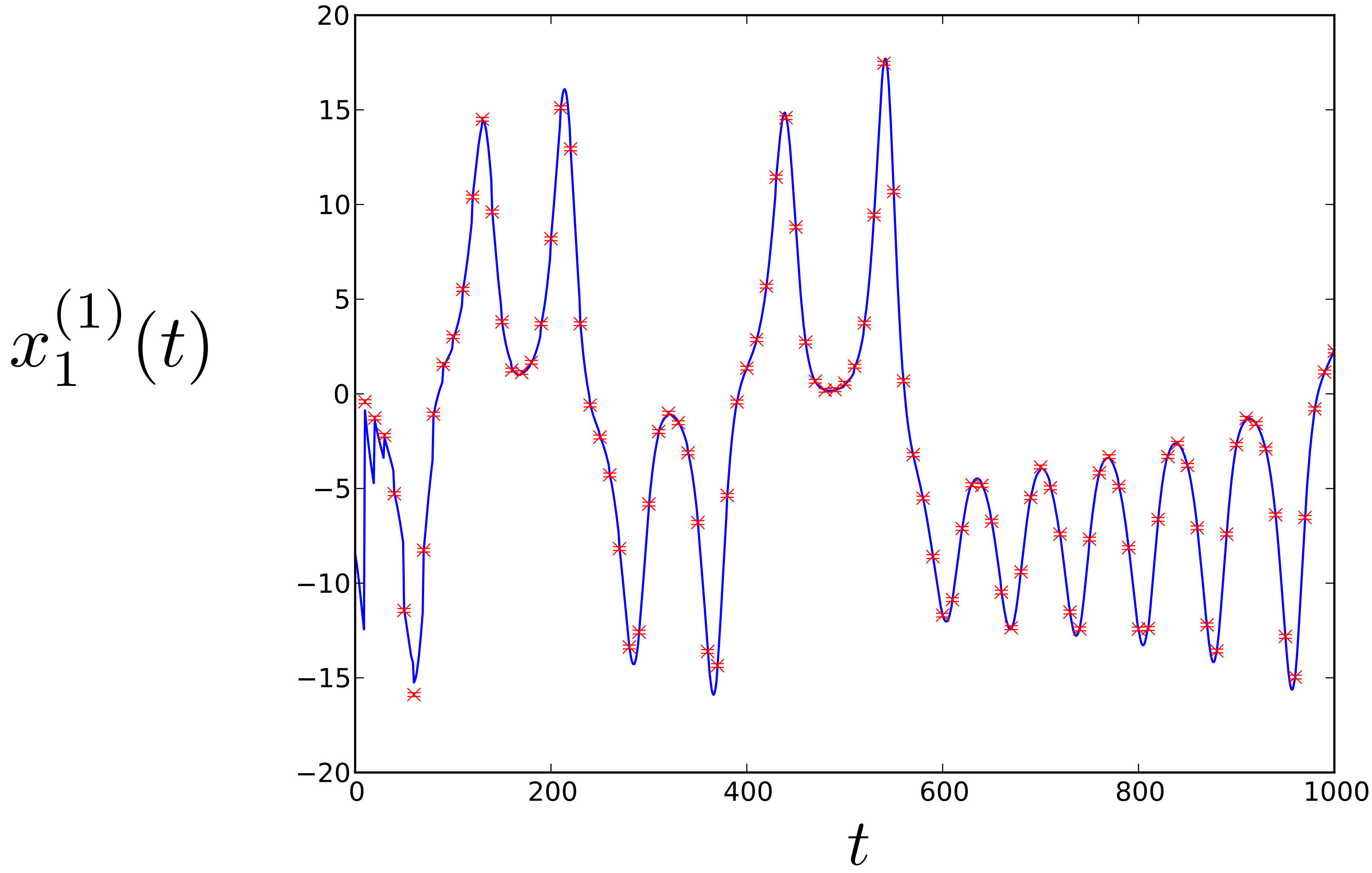
- Unknown parameter

$$x^n = f(x^{n-1}, \theta_0) + \frac{\partial f}{\partial \theta}(\theta - \theta_0) + \beta$$

- Model as $$\theta^n = \theta^{n-1} + \eta^n$$

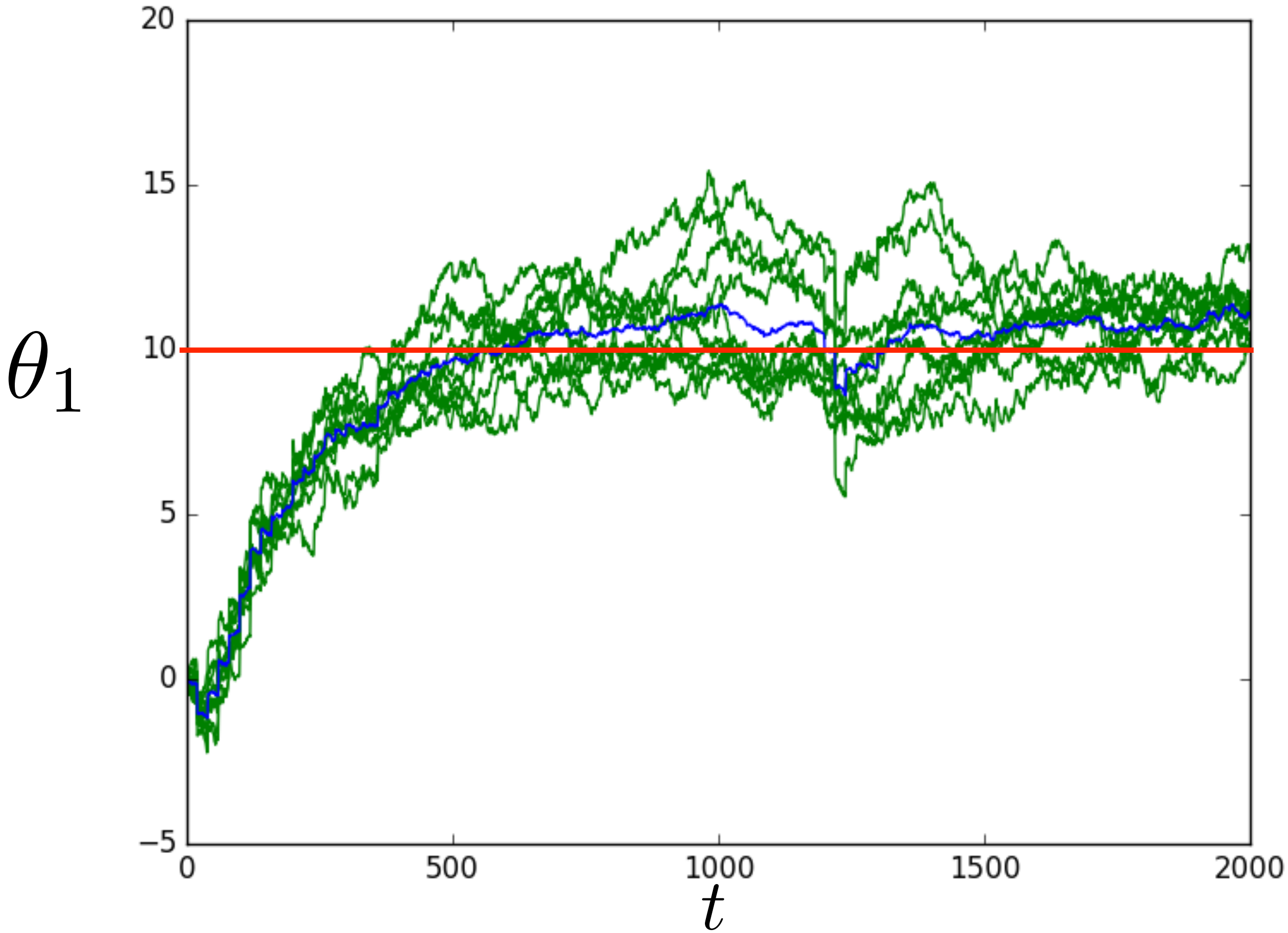
hence model error $$Q_{xx} = Q_\beta + \frac{\partial f}{\partial \theta} Q_\eta \frac{\partial f}{\partial \theta}^T$$

40,000 dimensional system (30,000 variables, 10,000 parameters).



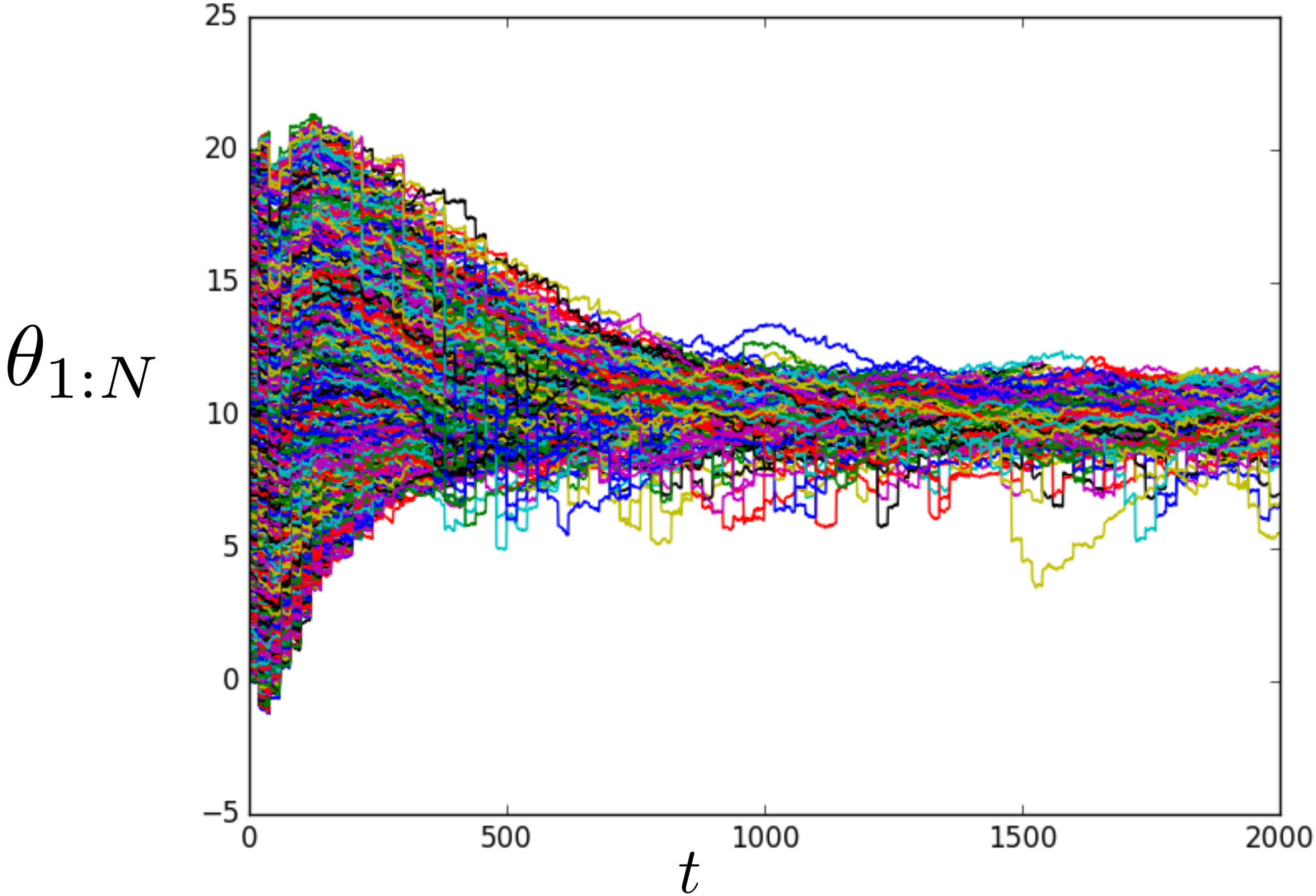
Time evolution mean of first variable system 1, starting 10 lower than true value.

40000 dimensional system (30000 variables, 10000 parameters).



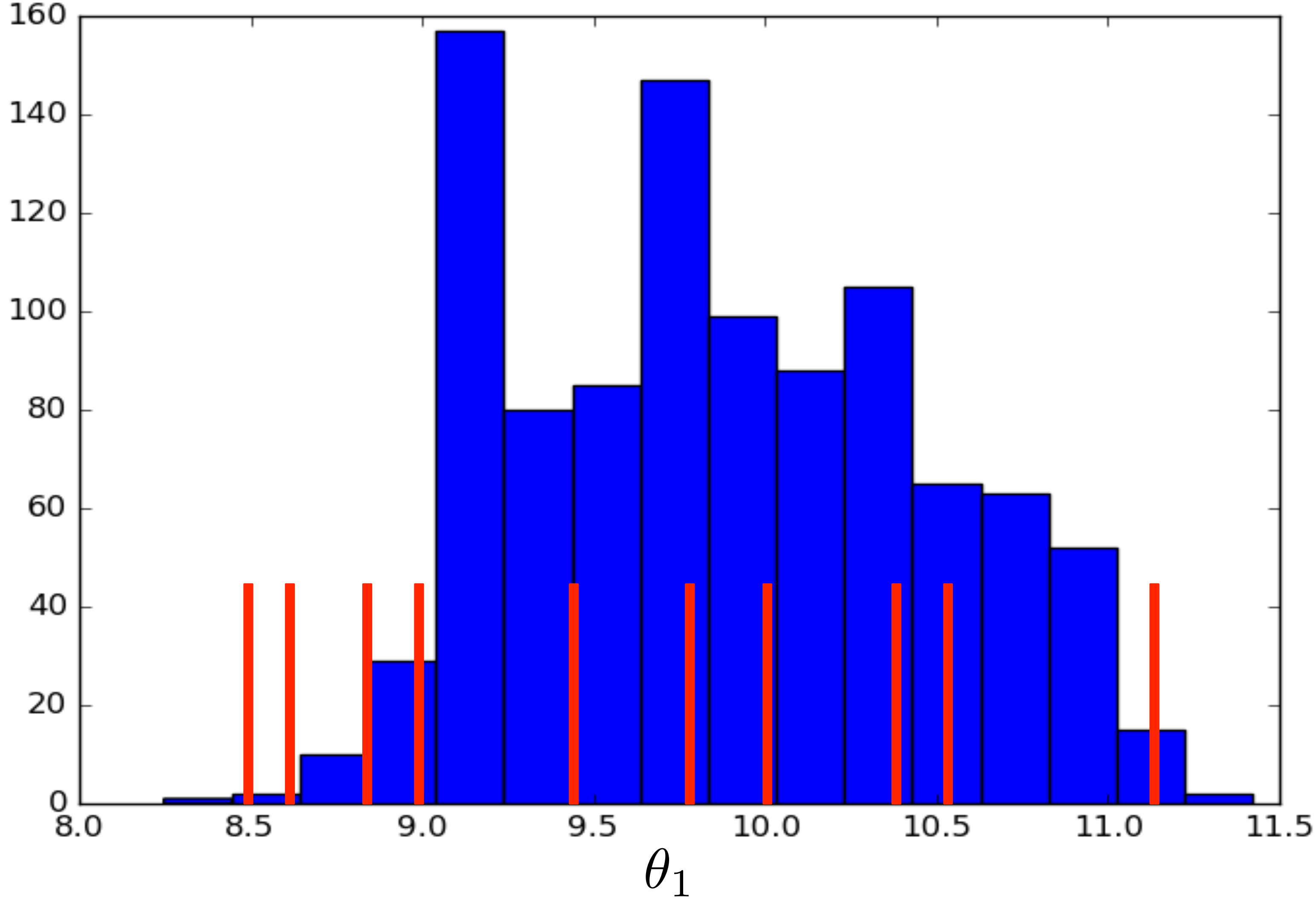
Time evolution mean of parameter system 1, starting 10 lower than true value.

Parameter mean values (dim=10,000)



Time evolution mean values parameter all 10,000 systems

Histogram parameter system 1, t=2000



Blue: histogram SIR 1000 members, red IEWPF 10 members

Conclusions

- Fully nonlinear non-degenerate particle filters for systems with arbitrary dimensions (but with bias) have been derived.
- The example can be viewed as an optimal proposal step to move particles to typical set, followed by an HCM step.
- Proposal-density freedom needs further exploration
- We need good estimate of *Q*…
- We need efficient weak-constraint 4DVar with fixed initial condition (for small time window). Could use EnsVar ?
- Need to explore bias versus MC variance.
- Needs mathematical back up…

- Implicit Equal-weights Particle Filter Zhu, M, P.J. van Leeuwen, and J. Amezcua, Q J Royal Meteorol. Soc., doi: 10.1002/qj.2784, 2015
- Aspects of Particle Filtering in high-dimensional spaces. Van Leeuwen, P.J., in: Dynamic Data-Driven Environmental System Science, LNCS 8964, 251, Springer, doi:10.1007/978-3-319-25138-7, 2015.
- Nonlinear Data Assimilation. Van Leeuwen, P.J., Y. Cheng, and S. Reich., Springer, doi:10,1007/978-3-319-18347-3, 2015.
- Twin experiments with the Equivalent-Weights Particle Filter and HadCM3. Browne, P.A., and P.J. van Leeuwen, Q.J.Royal.Meteorol. Soc., 141, doi: 10.1002/qj.2621, 2015.
- The effect of the Equivalent-weights particle Filter on model balances in a primitive equation model Ades M., P.J. van Leeuwen Monthly Weather Rev. 143, 581-596, doi:10.1175/MWR-D-14-00050.1, 2015.
- The Equivalent-weights Particle Filter in a high-dimensional system Ades M., P.J. van Leeuwen Q.J.Royal.Met.Soc.,141,484-503,doi:10.1002/qj.2370, 2015.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.