The Role of Inverse Problems and Optimisation in Uncertainty Quantification
ICMS, Edinburgh
17th – 18th June 2015

A workshop sponsored by:
The Turing Gateway to Mathematics, The Knowledge Transfer Network,
The International Centre for Mathematical Sciences, and the Smith Institute

# Optimisation in Uncertainty Quantification and Management

Chris L. Farmer
Mathematical Institute, Oxford

farmer@maths.ox.ac.uk

UNIVERSITY OF OXFORD
Mathematical Institute

Acknowledgements:

This presentation is based on work funded by:

- The Oxford-Martin School Programme on Resource Stewardship.

Thanks are due to colleagues, in Oxford and elsewhere,
for their comments and criticism.

University of Oxford
Mathematical Institute

# Topics for discussion

UNIVERSITY OF OXFORD

Mathematical Institute

## Uncertainty Quantification

- Inverse problems – simultaneous formulation
  – sequential assimilation

## Uncertainty Management

- The stochastic control problem
- Model predictive control
- Computing policies
- Discussion
- References

# UQ: The Forward model

State vector: $\varphi^n = (\varphi_1^n, \varphi_2^n, \ldots, \varphi_K^n)$ at time $t^n = n\tau$

'Forward dynamical model' $G_i(\varphi^n, \varphi^{n-1}, c^n) = 0$

where the $c^n$ = a vector of controls, supposed known for UQ

Suppose $\exists$ $F$ such that $\varphi^n = F(\varphi^{n-1}, c^n)$

and $F^{-1}$ such that $\varphi^{n-1} = F^{-1}(\varphi^n, c^n)$ (inverse function)

Note that $G_i(\varphi^n, \varphi^{n-1}, c^n) = \varphi_i^n - \varphi_i^{n-1}$ for parameters

# UQ: The Observation model

Simulation of the observing apparatus:

$$s_k^n = h_k(\varphi^n) + \sigma_k \xi_k^n$$ observe at discrete times

and $\xi_k^n \sim N(0,1) \quad iid$

Let $S^n = (s^0, s^1, s^2, \ldots, s^n)$

# UQ: The Observation model

$$\pi(s^n \mid \varphi^n) = z \exp\left( -\sum_k \frac{1}{2\sigma_k^2} (\varphi_k^n - s_k^n)^2 \right)$$

$z =$ a generic normalisation constant

# UQ: The main problems

The 'smoothing problem'

Given $\pi(\varphi^0)$ compute $\pi(\varphi^0 \mid S^N)$

The 'filtering problem'

Given $\pi(\varphi^0)$ compute $\pi(\varphi^N \mid S^N)$

The problems involve computing a posterior pdf given a prior, a model and observations.

# UQ: Formulation of the smoothing problem

Suppose that $\varphi^n = F(\varphi^{n-1}, c^n)$ implies $\varphi^n = F^{(n)}(\varphi^0, c^{1:n})$

for given controls with $F^{(0)}(\varphi^0) = \varphi^0$

$$\pi(\varphi^0, S^N) = z \prod_{m=0}^{N} \exp[-\sum_k \frac{1}{2\sigma_k^2}(h_k(\varphi^n) - s_k^n)^2]\, \pi(\varphi^0)$$

Gaussian approximation

$$J(\varphi^0) = -\ln(\pi(\varphi^0, S^N)) = \sum_k \frac{1}{2\sigma_k^2}(h_k(\varphi^n) - s_k^n)^2 - \ln(\pi(\varphi^0))$$

$$\breve{\varphi}^0 = \arg\min J(\varphi^0)$$

See publications by:
O. Ghattas & T. Bui-Thanh
UT Austin

$$L_{ij} = \left.\frac{\partial^2 J}{\partial \varphi_i^0 \partial \varphi_j^0}\right|_{\breve{\varphi}^0}$$

$C = L^{-1}$ for the covariance

# UQ: Formulation of filtering (in principle)

The 'particle filter'

$$\pi(\varphi^0) = \sum_{r=1}^{R} a_r^o \delta(\varphi^0 - \varphi_r^0) \quad \text{where} \quad \varphi_r^0 \sim \pi_0(\varphi^0) \quad \text{and} \quad a_r^o = \frac{1}{R}$$

Suppose $\pi(\varphi^{n-1} | S^{n-1}) = \sum_r a_r^{n-1} \delta(\varphi^{n-1} - \varphi_r^{n-1})$

$$\pi(\varphi^n, s^n | S^{n-1}) = \sum_r a_r^{n-1} \exp[-\sum_k \frac{1}{2\sigma_k^2}(h_k(\varphi^n) - s_k^n)^2] \delta(\varphi^n - \varphi_r^n)$$

where $\varphi_r^n = F(\varphi_r^{n-1}, c^n)$

$$\pi(\varphi^n | S^n) = \sum_r a_r^n \delta(\varphi^n - \varphi_r^n)$$

where $a_r^n = \dfrac{\breve{a}_r^n}{\sum_{r'} \breve{a}_{r'}^n}$ and $\breve{a}_r^n = a_r^{n-1} \exp[-\sum_k \frac{1}{2\sigma_k^2}(h_k(\varphi_r^n) - s_k^n)^2]$

See Evensen 2009

# UQ: Formulation of filtering (practical approach)

The 'ensemble variational filter'

Let $\pi_g(\varphi \mid \bar{\varphi}, L) = z\exp[-\frac{1}{2}(\varphi - \bar{\varphi})' L(\varphi - \bar{\varphi})]$

$$\pi(\varphi^0) = \sum_{r=1}^{R} a_r^o \pi_g(\varphi^0 \mid \varphi_r^0, RL_r^0) \qquad \text{where} \quad \varphi_r^0 \sim \pi_0(\varphi^0),$$

$$a_r^o = \frac{1}{R} \text{ and } L_r^0 = \text{inverse prior cov.}$$

Suppose $\pi(\varphi^{n-1} \mid S^{n-1}) = \sum_r a_r^{n-1} \pi_g(\varphi^{n-1} \mid \varphi_r^{n-1}, RL_r^{n-1})$

Let $J_r^n(\varphi^n) = \sum_k \frac{1}{2\sigma_k^2}(h_k(\varphi^n) - s_k^n)^2] + \frac{R}{2}(F^{-1}(\varphi^n, c^n) - \varphi_r^{n-1})' L_r^{n-1}(F^{-1}(\varphi^n, c^n) - \varphi_r^{n-1})$

$$\varphi_r^n = \arg\min_{\varphi^n} J_r^n(\varphi^n), \quad \bar{\varphi}^n = \sum_{r=1}^{R} a_r^o \varphi_r^n, \quad \alpha_i^2 = \sum_{r=1}^{R} a_r^o (\bar{\varphi}_i^n - \varphi_{r,i}^n)^2$$

$$L_{r,ij}^n = (A_r' L_r^0 A_r)_{ij} + \frac{\delta_{ij}}{\alpha_i^2} \quad \text{where} \quad A_{r,ij} = \left.\frac{\partial F_i^{-1}(\varphi^n, c^n)}{\partial \varphi_j^n}\right|_{\varphi_r^n}$$

Main heuristic, clf 2014

# UM: Formulation of stochastic control

Suppose: $G_i(\varphi^n, \varphi^{n-1}, c^n) = 0$

$$s_k^n = h_k(\varphi^n) + \sigma_k^n \xi_k^n \quad \text{where} \quad \xi_k^n \sim N(0,1)$$

For given $c^{1:n}$, $\exists\ F^{(n)}$ s.t. $\varphi^n = F^{(n)}(\varphi^0, c^{1:n})$

Suppose $c^0$ and $\sigma^0$ are the initial settings of the controls

Let $\gamma^N = \sum_n [\gamma_\varphi(\varphi^n) + \gamma_c(c^n, c^{n-1}) + \gamma_\sigma(\sigma^n, \sigma^{n-1})]$

e.g. $\gamma_u = \kappa_u (c^n - c^{n-1})^2$ and $\gamma_\sigma = \dfrac{\kappa_{\sigma,0}}{\sigma^n} + \kappa_{\sigma,1}(\sigma^n - \sigma^{n-1})^2$

for some positive constants $\kappa_u, \kappa_{\sigma,0}$ and $\kappa_{\sigma,1}$

# UM: The stochastic control problem

Seek control *policy functions* $\tilde{u}^n, \tilde{\sigma}^n$ :

$$c^n = \tilde{c}^n(S^{n-1}), \quad \sigma^n = \tilde{\sigma}^n(S^{n-1})$$

Such that

$$\{\tilde{c}^n, \tilde{\sigma}^n\}_{n=0}^{N} = \underset{\{\tilde{c}^n, \tilde{\sigma}^n\}}{\arg\min}\ \bar{\gamma}^N$$

where

$$\bar{\gamma}^N =$$

$$\int \sum_{n=1,N} [\gamma_\varphi(\varphi^n) + \gamma_c(c^n, c^{n-1}) + \gamma_\sigma(\sigma^n, \sigma^{n-1})]\pi(\varphi^0)\pi(\xi^{1:N-1})d\varphi^0 d\xi^{1:N-1}$$

For a discussion of *policy functions* see:
W.B. Powell, 'Clearing the jungle of stochastic optimization'
http://castlelab.princeton.edu

# UM: Stochastic model predictive control

MPC: At *each* time $t^n$

Suppose that for times $1:(n-1)$

The policy *values* $\sigma^{1:n-1}, c^{1:n-1}$ have been applied

Solve the filtering problem for $\pi(\varphi^{n-1} \mid S^{n-1})$ and sample for

$\varphi_r^{n-1} \sim \pi(\varphi^{n-1} \mid S^{n-1})$

Then compute new policy *values* $u^{n:N}$ such that

$$c^{n:N} = \arg\min_{c^{n:N}} \sum_{m=n}^{N} \left\{ \sum_r \left( a_r^{n-1} \gamma_\varphi(\varphi_r^m) \right) + \gamma_c(c^m, c^{m-1}) \right\}$$

with $\varphi_r^m$ understood as functions of $\varphi_r^{n-1}$ and the controls $c^{n-1:N}$

Apply the controls $c^n$ *only* at time $n$ (i.e. discard the later controls) and make the observation of $s^n$.

# UM: Beyond model predictive control

Problem A:

Within MPC with constant control parameters, how do we find an optimal observation variance, $\sigma^n$?

Problem B: Compute linear control functions in a similar way

-- using an ensemble to estimate the mean cost

$$\tilde{c}^n = c_0^n + c_1^n . s^n$$

$$\tilde{\sigma}^n = \sigma_0^n + \sigma_1^n . s^n$$

and so on.

# Discussion

- Smoothing and filtering problems can be solved fairly well using Bayesian formulations and optimisation methods using adjoint techniques.

- Control problems are very challenging. We can formulate them but to go beyond MPC for large-scale engineering or geoscience problems is an outstanding challenge to us all.

We could also consider the topics of:

How to keep all matrices sparse and avoid any matrix inversions (e.g. clf 2007)
Stochastic dynamical systems
Robust control
Forecast evaluation
Model sensitivity
Model criticism and comparison
Multi-objective optimisation (Pareto fronts)
Adaptive management

# References

Mathematical Institute


- A. O'Hagan and J. Forster: Kendall's Advanced Theory of Statistics. Vol 2B, Bayesian Inference. 2nd Edition. Arnold Publishers, 2004.
- B.D.O. Anderson & J.B. Moore: Optimal Filtering, Prentice Hall 1979. (Dover 2005).
- A.H. Jazwinski: Stochastic Processes and Filtering Theory. Academic Press, 1970.
- G. Evensen: *Data Assimilation*: *The Ensemble Kalman Filter* (2nd Edition). Springer 2009.
- C.L. Farmer: Bayesian Field Theory Applied to Scattered Data Interpolation and Inverse Problems. pp. 147-166. Algorithms for Approximation. Editors. A. Iske and J. Levesley, Springer-Verlag, Heidelberg, 2007.
- Inverse Theory for Petroleum Reservoir Characterization and History Matching. D. S. Oliver, A. C. Reynolds and N. Liu. Cambridge University Press, 2008.
- Silverman, B.W. 1986 Density Estimation for Statistics and Data Analysis, Chapman and Hall.
- Stordal, A.S., Karlsen, H.A., Nævdal, G., Skaug, H.J., &Vallès, B. 2011 Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. Computational Geosciences, 15(2), pp 293-305.
- C. L. Farmer. "An ensemble variational filter for sequential inverse problems". Proceedings of 'Inverse Problems -- from Theory to Applications (IPTA 2014)', pp 164-168, Institute of Physics, 2014. http://ipta2014.iopconfs.org/IPTAProceedings
- W. Lee and C. L. Farmer. 2014. "Data Assimilation by Conditioning of Driving Noise on Future Observations". IEEE Transactions on Signal Processing, 62 (15), pp 3887-3896.

# References

Mathematical Institute


- J. D. Jansen: Adjoint-based optimization of multi-phase flow through porous media – A review. Computers and Fluids, v 46, pp 40-51, 2011.

- D. P. Bertsekas: *Dynamic Programming and Optimal Controli.* Vol. I, 3rd Edition, 2005.
  Vol. II, 4th Edition, 2012 (On Approximate Dynamic Programming**)**

- W.B. Powell: Approximate Dynamic Programming, 2nd Edition. Wiley, 2011.

- P. Whittle: *Optimal Control*: *Basics and Beyond*. Wiley 1996.

# UM: Extra slides

Analytical solution of a one-step stochastic control problem for the logistic map.

This is equivalent to a *decision problem* where we need to decide what measurement to make before making it, and then observing and then setting the control on the basis of the *feedback* from the measurement.

# Example: One-step stochastic control of the logistic map

Model: $x = 1 - u{x_0}^2, \quad x_0 \sim N(q, \lambda^2)$

Choose $u$, changing it from $u_{-1}$ with cost $= \dfrac{\beta}{2}(u(v,s) - u_{-1})^2$

to land near $z$ with an error cost $= \dfrac{\gamma}{2}(x - z)^2$

Observations of $x_0$ cost $\dfrac{\alpha}{2v^2}$; you need to choose $v$ too.

i.e. $s = x_0 + v\zeta, \quad \zeta \sim N(0,1)$

total cost $= \dfrac{\alpha}{2v^2} + \dfrac{\beta}{2}(u - u_{-1})^2 + \dfrac{\gamma}{2}(x - z)^2$

But as we do not know $x_0$ very well we do not know $x$.

So what should we do?

# Example: One-step stochastic control of the logistic map

Taking variations w.r.t. $u$ gives

$$\left.\frac{dJ(u+\varepsilon\delta u)}{d\varepsilon}\right|_{\varepsilon=0} = 0 =$$

$$\int\left[\beta(u(v,s)-u_{-1})\delta u-\gamma(1-ux_0{}^2-z)\delta u\, x_0{}^2\right]N(s-x_0,v^2)N(x_0-q,\lambda^2)dx_0ds$$

So

$$\int\left[\beta(u(v,s)-u_{-1})-\gamma(1-ux_0{}^2-z)\,x_0{}^2\right]N(s-x_0,v^2)N(x_0-q,\lambda^2)dx_0=0$$

and

$$u(v,s)=\frac{\int\left[\beta u_{-1}+\gamma(1-z)\,x_0{}^2\right]N(s-x_0,v^2)N(x_0-q,\lambda^2)dx_0}{\int\left[\beta+\gamma(1+x_0{}^4)\right]N(s-x_0,v^2)N(x_0-q,\lambda^2)dx_0}$$

## Example: One-step stochastic control of the logistic map

$$x = 1 - u{x_0}^2, \quad x_0 \sim N(q, \lambda^2)$$

$$s = x_0 + v\zeta, \quad \zeta \sim N(0,1)$$

$$\pi(x, s, x_0) = \delta(x - 1 + u{x_0}^2) N(s - x_0, v^2) N(x_0 - q, \lambda^2)$$

$$J(v) = \frac{\alpha}{2v^2} +$$

$$\int \left[ \frac{\beta}{2}(u(v,s) - u_{-1})^2 + \frac{\gamma}{2}(x - z)^2 \right] \delta(x - 1 + u{x_0}^2) N(s - x_0, v^2) N(x_0 - q, \lambda^2) dx_0 dx ds$$

so

$$J(v) = \frac{\alpha}{2v^2} +$$

$$\int \left[ \frac{\beta}{2}(u(v,s) - u_{-1})^2 + \frac{\gamma}{2}(1 - u{x_0}^2 - z)^2 \right] N(s - x_0, v^2) N(x_0 - q, \lambda^2) dx_0 ds$$

Find a control policy $u(v,s)$ and a measurement accuracy $v$.

$$J^*(v) = \inf_u J, \qquad J^* = \inf_v J^*(v)$$

# Example: One-step stochastic control of the logistic map

Using the calculus of variations one can show that:

$$u(v,s) = \frac{\beta u_{-1} + \gamma(1-z)(\kappa^2 + a^2)}{\beta + \gamma(a^4 + 6a^2\kappa^2 + 3\kappa^4)}$$

where $a = \dfrac{s\lambda^2 + qv^2}{\lambda^2 + v^2}$

and $\kappa = \dfrac{\lambda v}{\sqrt{\lambda^2 + v^2}}$

Remember: $N(s - x_0, v^2) \sim$ likelihood

$N(x_0 - q, \lambda^2) \sim$ prior

# Control policy for the logistic map

$\lambda = 0.5$ ~ prior standard deviation

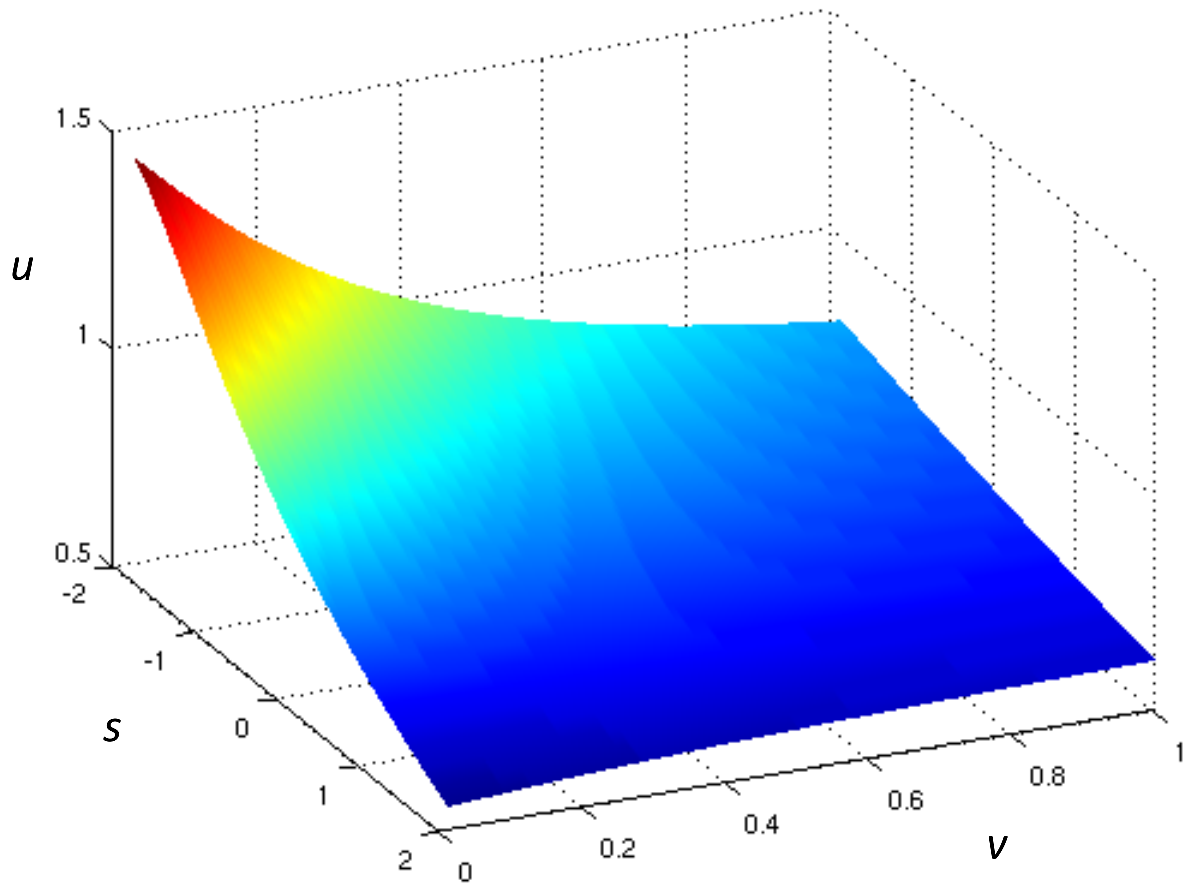
$q = 0.6$ ~ prior mean

$z = 0.1$

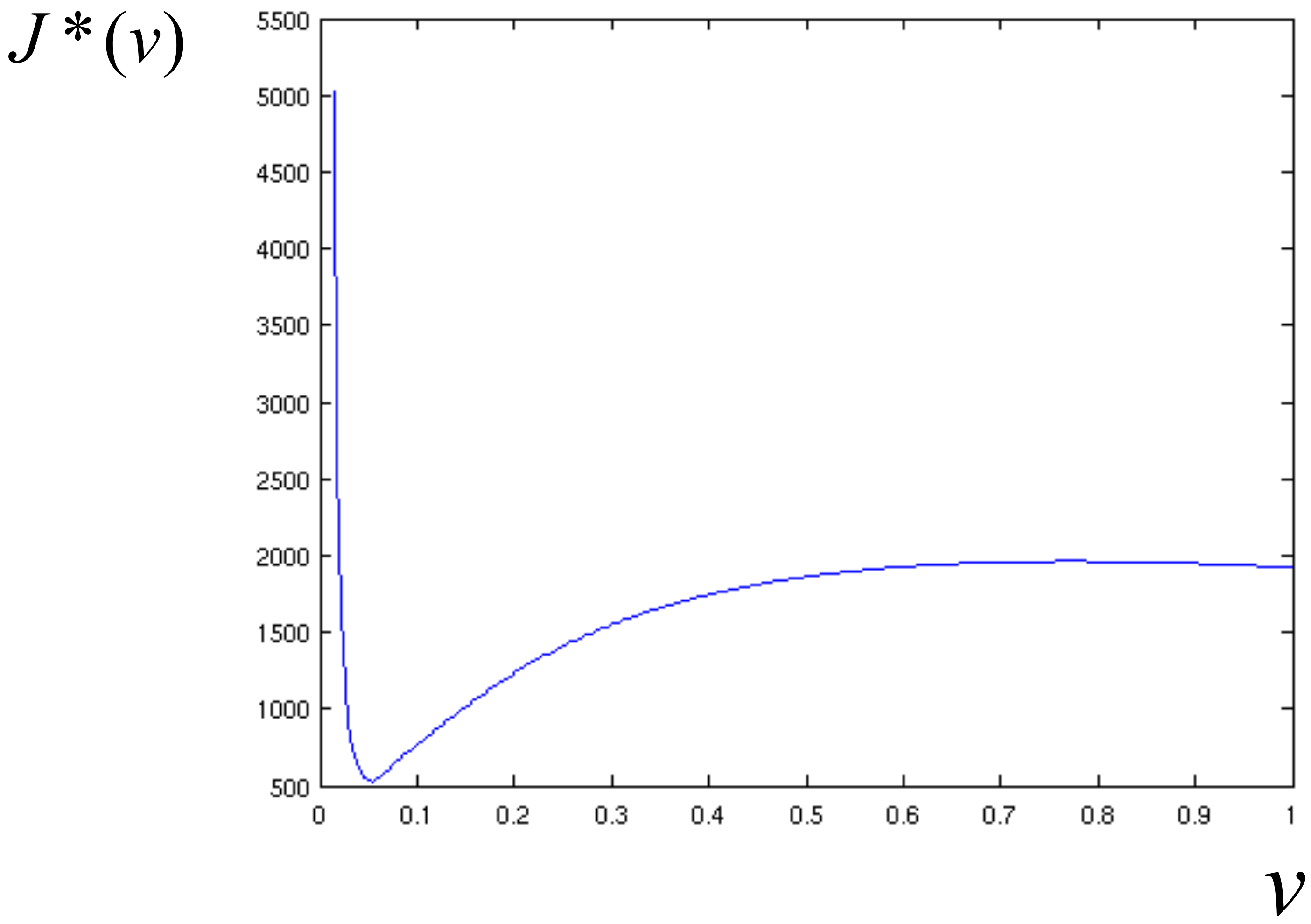
$\alpha = 1.0$

$\beta = 10^{-3}$

$\gamma = 10^{4}$



# Cost-to-go for the logistic map by numerical integration




$(v^*, J^*)$ is at the minimum of the curve

# UQ: Extra slides

Application of the ensemble variational filter to the filtering problem on the Lorenz-96 equations

# UQ: Numerical example: Lorenz 96

Mathematical Institute

$$\frac{du_i}{dt} = 0 \quad : \text{'reality' and forward model}$$

$$\frac{dv_i}{dt} = (v_{i+1} - v_{i-2})v_{i-1} - v_i(\kappa + e^{u_i}) + W_i \quad : \text{'reality' and forward model}$$

$$\varphi = (u, v)$$

Reality:

$u_i = 0.5 + 0.2 \sin(0.3\, i) + 0.02\,(0.5 - \zeta'_i)$; $\quad W_i = 10.0$

$v_i(0) = 10 + 2\zeta''_i \qquad \zeta'_i, \zeta''_i \sim N(0,1) \quad iid$

No parameters are observed. 1000 equilibration steps before observing

Observe variables with $\sigma_i^2 = 0.01$

Initialised with random sampling. Window: $w(R) = R^{0.1}$

Cor. lengths. 4.0 (parms) & 0.1 (vars) Init. var 1.0 parm and 10.0 var

Implicit Euler, time step = 0.01 between obs and for dynamics

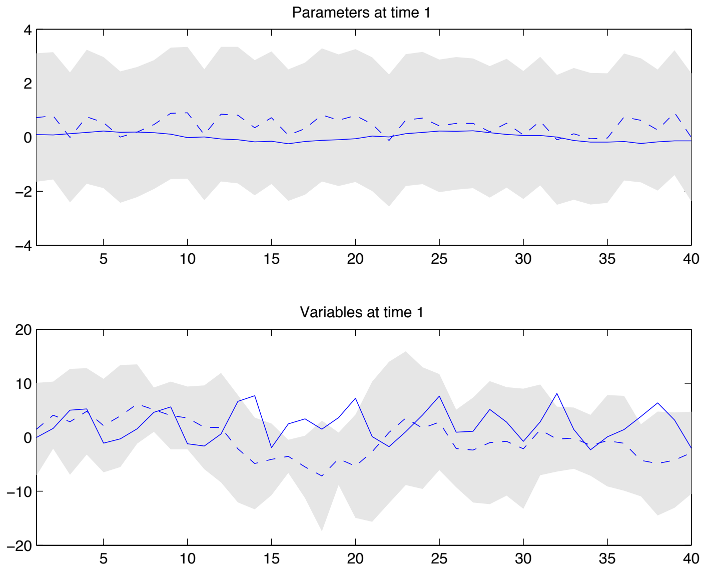
See Yang & DelSole 2009, 2010 for related work

Yang, X., and DelSole, T. 2009 Using the ensemble Kalman Filter to estimate multiplicative model parameters. Tellus, 61, pp 601-609.
DelSole, T. and Yang, X. 2010 State and parameter estimation in stochastic dynamical models. Physica D, 239, pp 1781-1788.

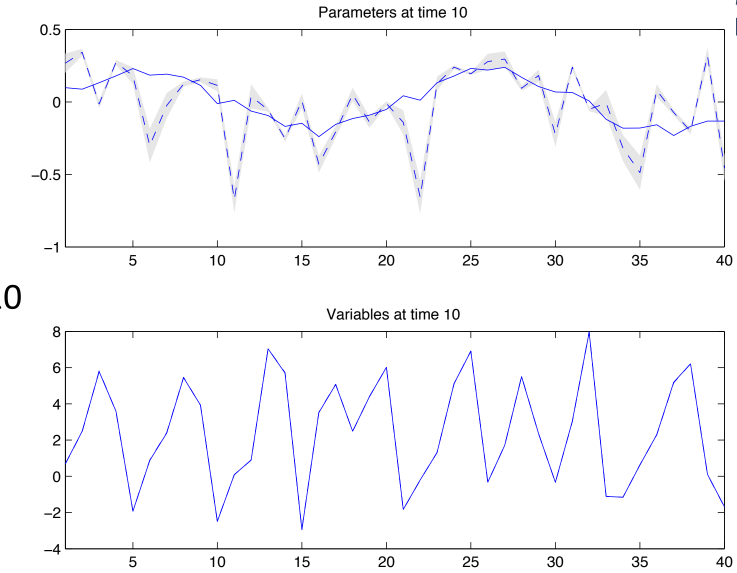
# ‘Validation’: $R$ = 8. Observe all variables

UNIVERSITY OF OXFORD
Mathematical Institute

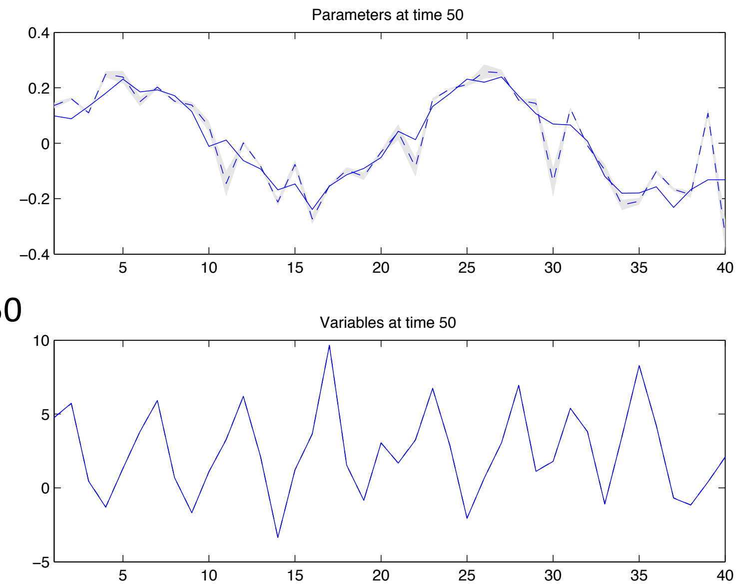
$t = 1$



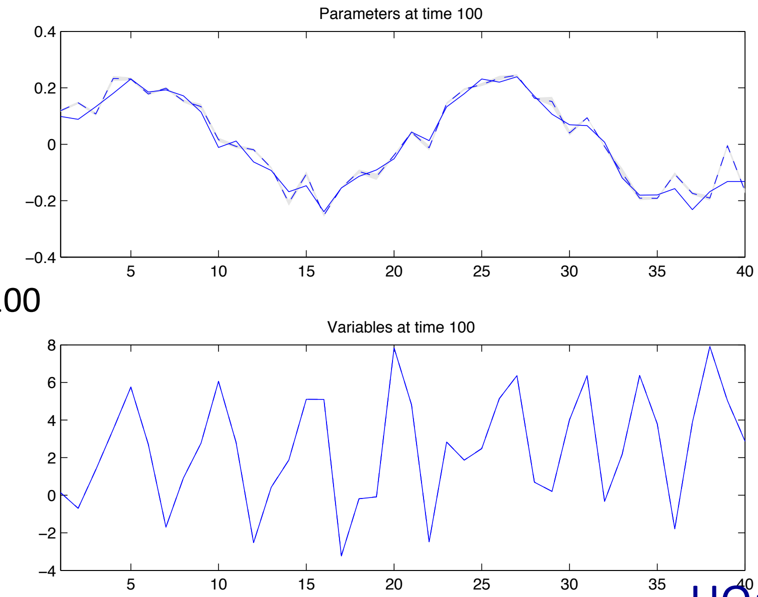

$t = 10$




$t = 50$



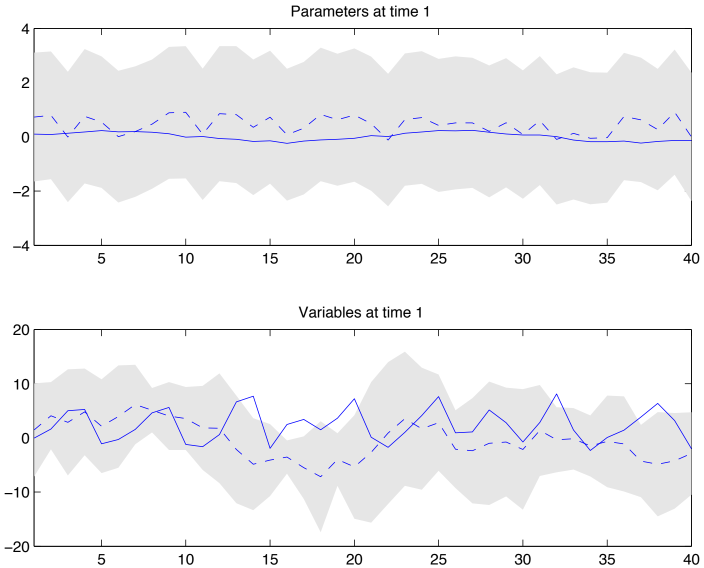

$t = 100$



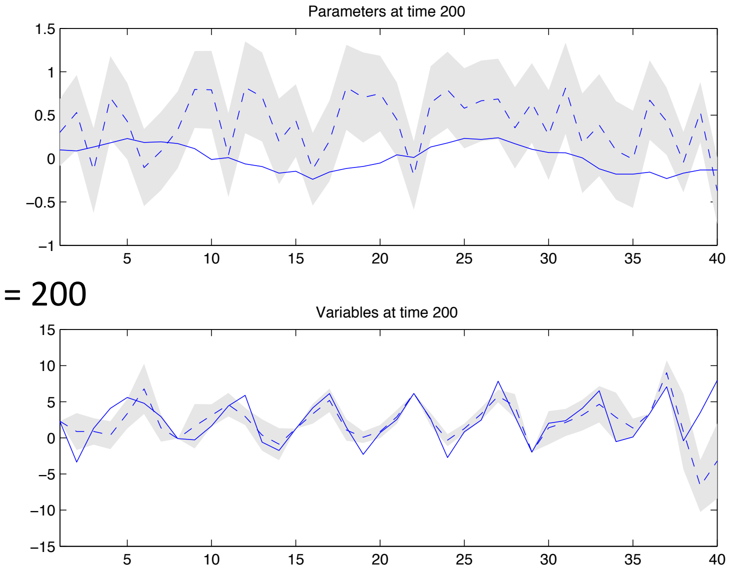
# $R$ = 8. Observe each 7th variable


UNIVERSITY OF OXFORD
Mathematical Institute

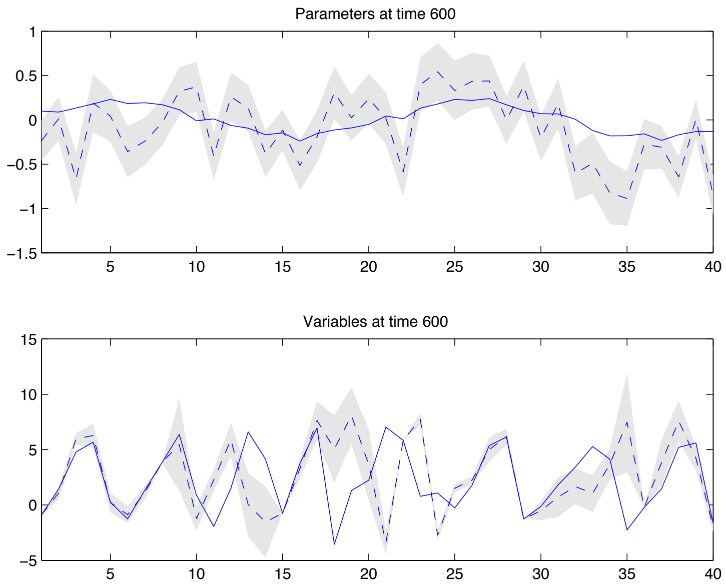

$t = 1$



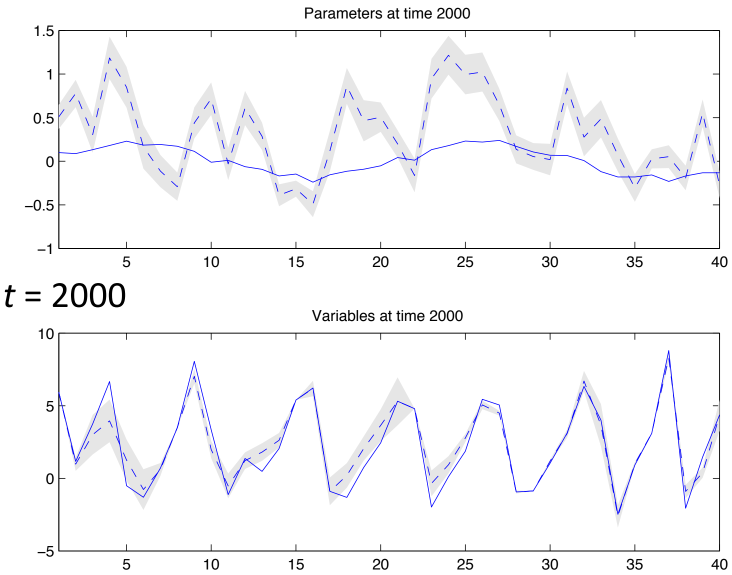

$t = 200$




$t = 600$



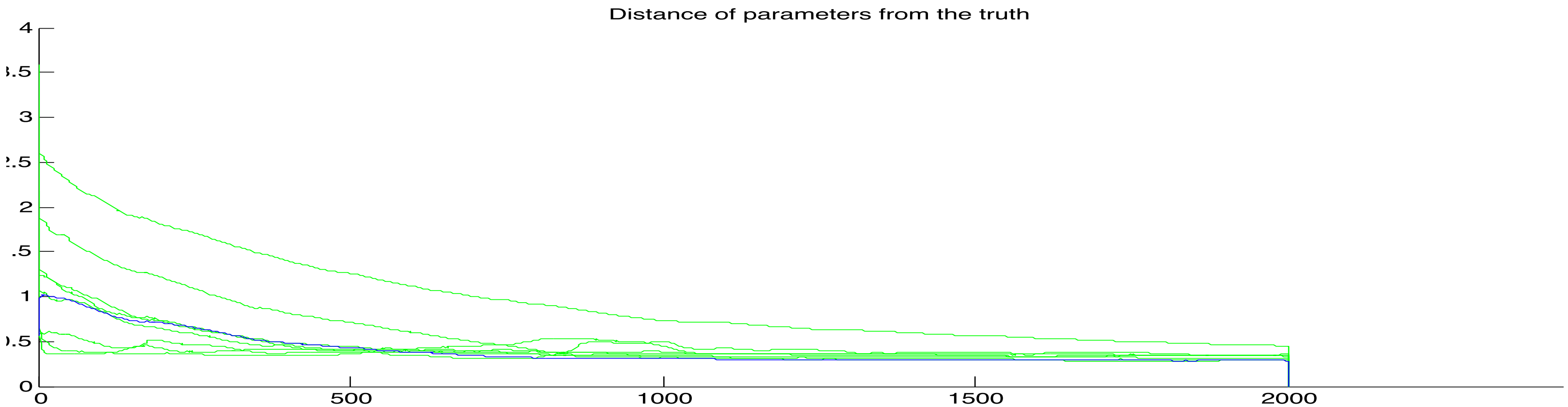

$t = 2000$

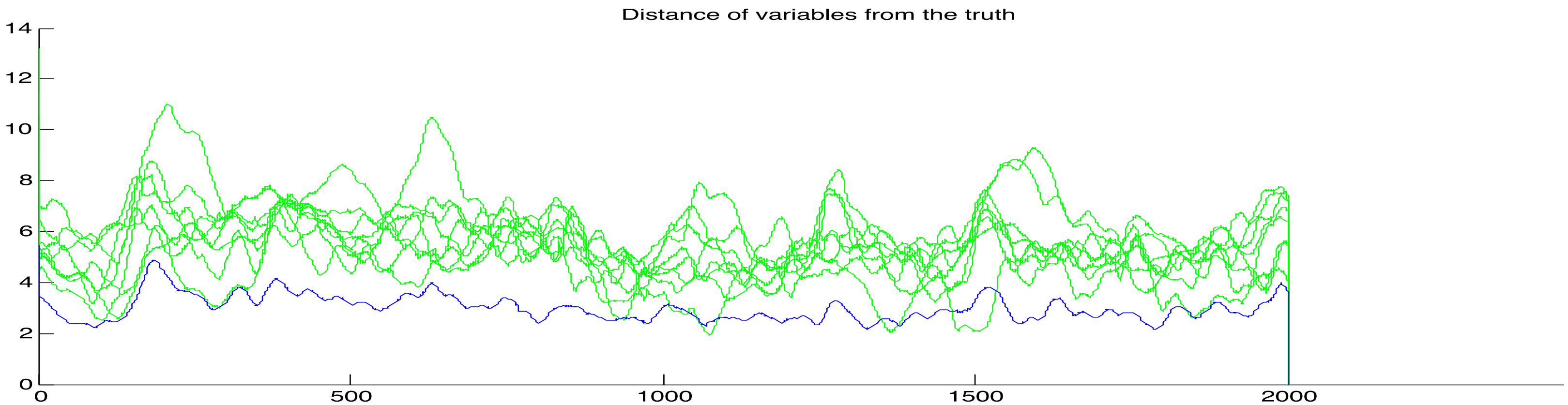


# $R$ = 8. Observe each $7^{th}$ variable

UNIVERSITY OF OXFORD

Mathematical Institute




Root mean square distance of ensemble members (green) and the mean (blue) from the truth plotted against time

# The ensemble variational filter

## The value of Implicit Euler time discretisation

$$\frac{d\varphi_i}{dt} = f_i(\varphi, t)$$

time step $\tau$

$$\varphi^n = \varphi^{n\text{-}1} + \tau f(\varphi^n)$$

One finds:

$$F^{-1}(\varphi^n) = \varphi^n - \tau f(\varphi^n)$$

Also:

$$A_{ij}^n = \delta_{ij} - \tau \frac{\partial f_i(\varphi^n)}{\partial \varphi_j^n}$$

# The ensemble variational filter

UNIVERSITY OF OXFORD
Mathematical Institute

Heuristic for updating the weights

$$\breve{a}_r^n = a_r^{n-1} \frac{|L_r^{n-1}|^{1/2}}{|L_r^n|^{1/2}} \exp(-J_r^n(\varphi_r^n))$$

$$a_r^n = \varepsilon_w \frac{1}{R} + (1-\varepsilon_w) \frac{\breve{a}_r^n}{\sum_r \breve{a}_r^n}$$

Stordal et al 2011

# Variational Smoothing Filter



## Updating the precision matrices

1. $$L^n_{ij,r} = \left.\frac{\partial^2 J^n_r(\varphi^n)}{\partial\varphi^n_i \partial\varphi^n_j}\right|_{\varphi^n_r}$$

$$= p_i\delta_{ij} + \sum_{mk}(A^n_{im,r}L^{n-1}_{mk,r}A^n_{kj,r} - \tau\frac{\partial^2 f_m(\varphi^n_i)}{\partial\varphi^n_{i,r}\partial\varphi^n_{j,r}}L^{n-1}_{mk,r}(\varphi^n_{k,r} - \tau f_k(\varphi^n_r) - \varphi^{n-1}_{k,r}))$$

All of the component matrices are sparse

2. Approximate $\frac{|L^{n-1}_r|^{1/2}}{|L^n_r|^{1/2}} \approx 1$

3. Reset $L^n_{ij,r} = (A^{nT}_r L^0_r A^n_r)_{i,j} + \frac{w(R)\,\delta_{ij}}{\varepsilon + \text{var}_i}$ every time step

Main heuristic

where $w(R)$ is a slowly increasing function of $R$ and

$\text{var}_i =$ empirical, ensemble variance of the ensemble $\{\varphi^n_r\}$

# Choosing $L^0$ via local random fields – for sparsity

clf 2007

$$Q(\psi) = \frac{1}{2}\int [a\psi^2 + b(\nabla\psi)^2 + c(\nabla^2\psi)^2]d\omega \quad = \quad \frac{1}{2}\int [\psi L \psi]d\omega$$

where $L = a - b\nabla^2 + c\nabla^2\nabla^2$

Helmholtz Green's functions

$$g(r) = \frac{e^{-r\sqrt{\frac{a}{b}}}}{4\pi r b} \quad \text{3D}$$

$$g(r) = \frac{e^{-r\sqrt{\frac{a}{b}}}}{4\sqrt{ab}} \quad \text{1D}$$

$$\pi(\varphi) = z\exp(-Q(\varphi - \overline{\varphi}))$$

$$g(x-y) := \langle \varphi(x)\varphi(y) \rangle$$

*Theorem*: $Lg(x-y) = \delta(x-y)$

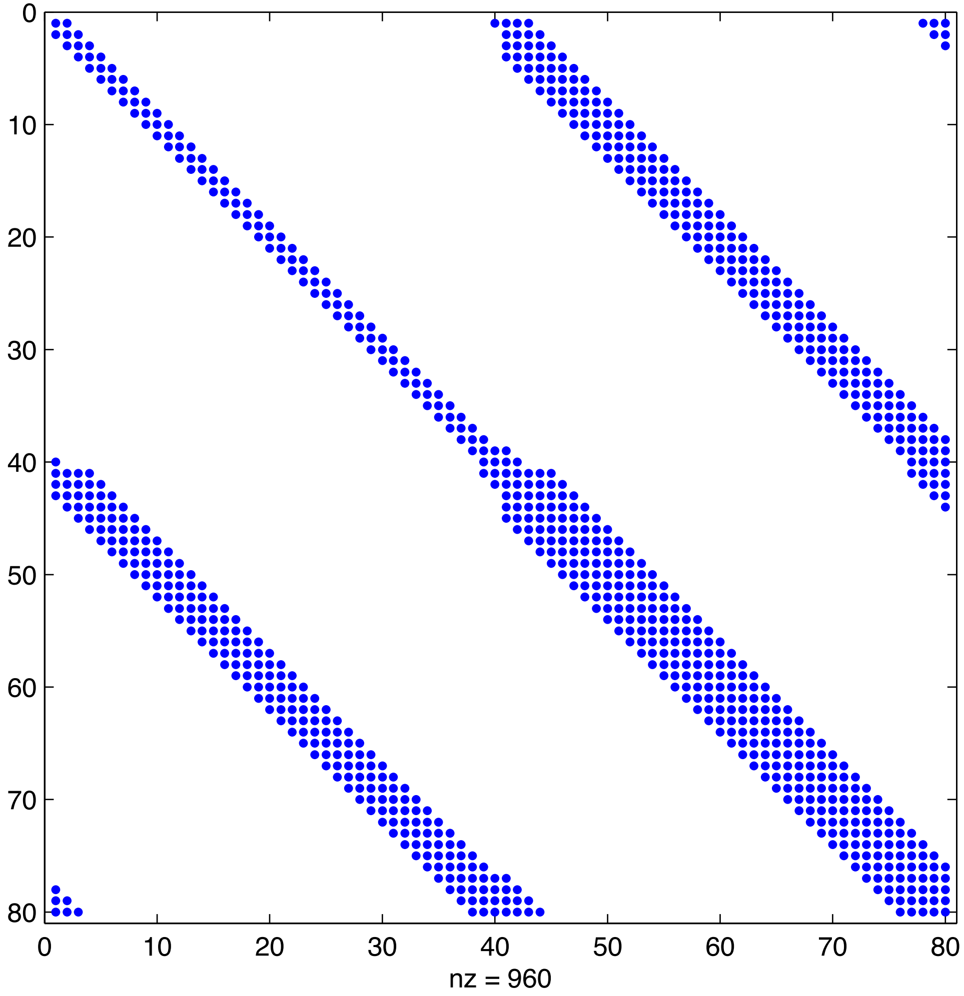
After discretisation $L_{ij}$ is sparse and $C = (L)^{-1}$ is the covariance matrix

$L = a - b\nabla^2$ - the 'Helmholtz precision matrix' - is particularly convenient and was used in the numerical experiments on Lorenz-96

Set $L_r^0 = wL$ for some 'sharpness control' $w$ that increases with $R$

# ‘Validation’: Observe all variables

UNIVERSITY OF OXFORD
Mathematical Institute




Approximate sparse precision matrix