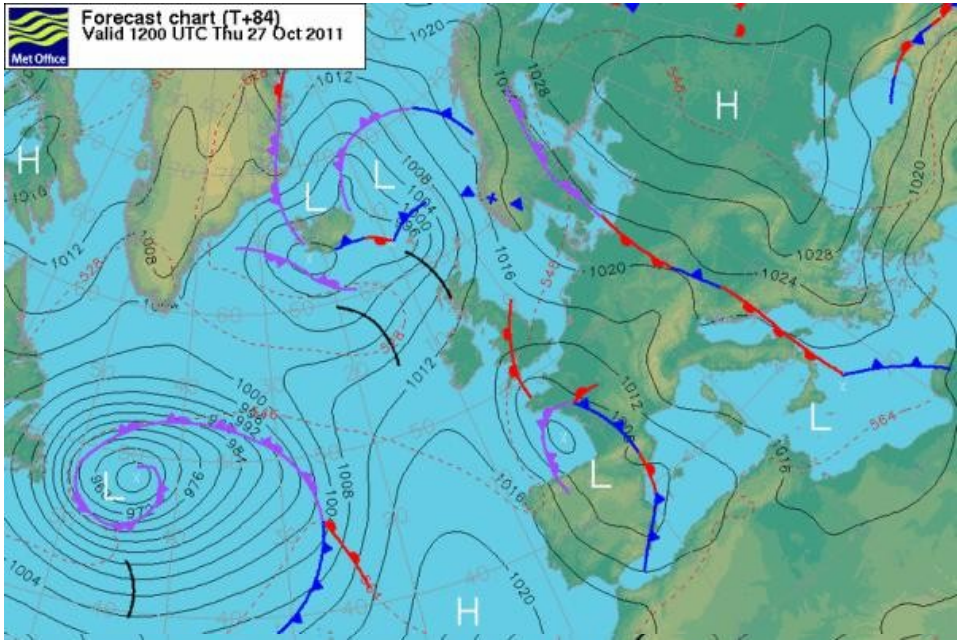
Conditioning and Preconditioning of the Optimal State Estimation Problem



Met Office Website

Nancy Nichols, Stephen Haben and Amos Lawless

University of Reading

University of Reading

Department of Mathematics

Outline

- Optimal State Estimation
- Conditioning
- Preconditioning
- Ill-posedness and Regularization
- Conclusions

1.

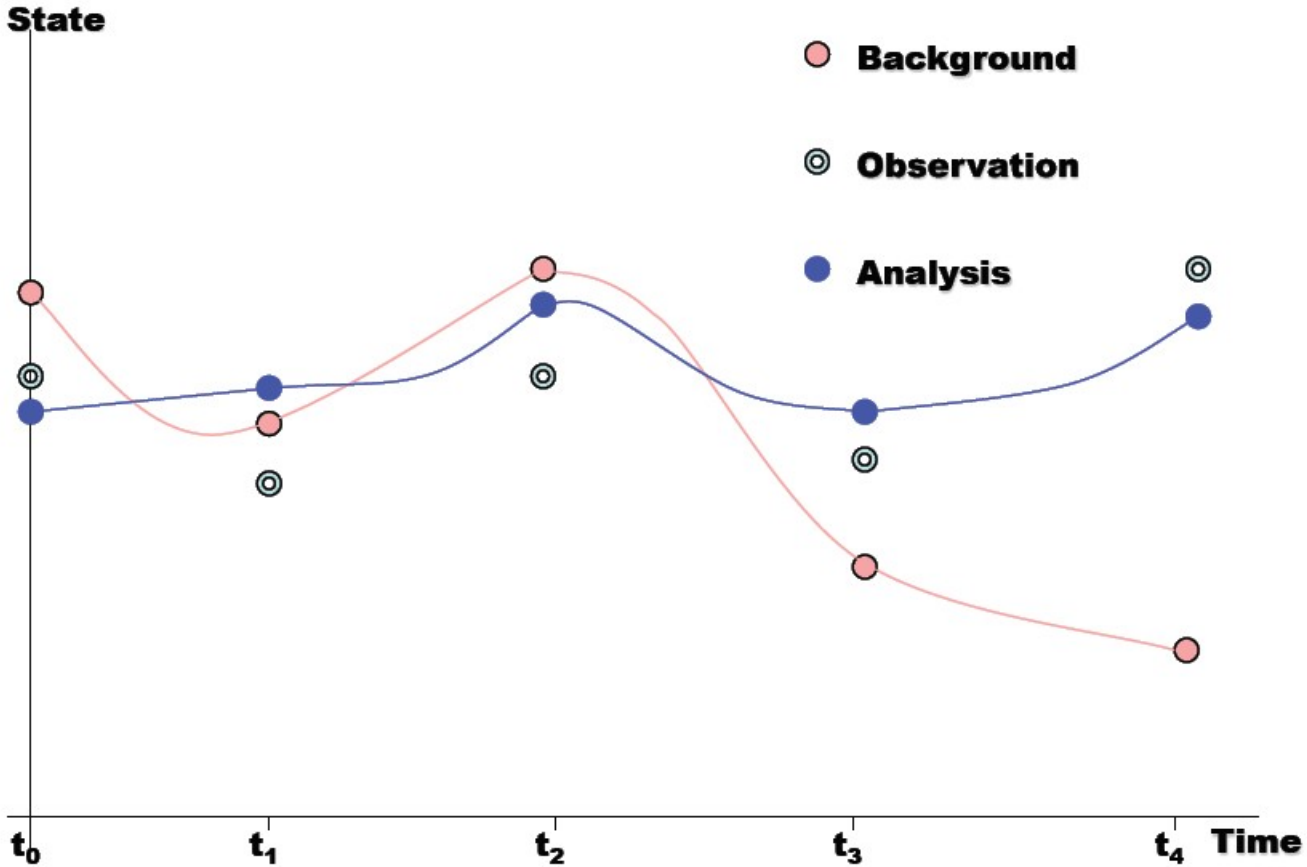
Optimal State Estimation

State Estimation / Data Assimilation

Aim: Find the optimal estimate (analysis) of the expected states of a system, consistent with both observations and the system dynamics given:

- Numerical prediction model (PDE)
- Observations of the system (over time)
- Background state (prior estimate)
- Estimates of the error statistics

State Estimation / Data Assimilation



Optimal Bayesian Estimate

$$\min \; J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b)$$

$$+ \sum_{i=0}^{n} (\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i)$$

subject to $\mathbf{x}_i = \mathcal{M}(t_i, t_0, \mathbf{x}_0)$

$\mathbf{x}_0^b$ - Background state (prior estimate)

$\mathbf{y}_i$ - Observations

$\mathcal{H}_i$ - Observation operator

$\mathbf{B}$ - Background error covariance matrix

$\mathbf{R}_i$ - Observation error covariance matrix

Significant Properties:

- Very large number of unknowns (10^7 – 10^8)
- Few observations (10^5 – 10^6)
- System nonlinear unstable/chaotic
- Multi-scale dynamics

Solution of the Problem

- Solve iteratively by an approximate Gauss-Newton (incremental variational) method
- Solves a sequence of linear least-squares problems
- Use conjugate gradient/quasi-Newton iteration method to solve inner linear least-squares problem.
- Must solve in real time

Conditioning of the Problem

Accuracy/rate of convergence depend on the condition number = $\lambda_{max} / \lambda_{min}$ of the Hessian:

$$\boxed{\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}}}$$

where

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_n \mathbf{M}_{0,n} \end{pmatrix} \qquad \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_n \end{pmatrix}$$

$$\mathbf{M}_{0,k} = \frac{\partial \mathcal{M}_{0,k}}{\partial \mathbf{x}}|_{\mathbf{x}_0} \qquad \mathbf{H}_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}|_{\mathcal{M}_{0,k}(\mathbf{x}_0)}$$

Three Questions

- What is the effect of the covariance structure on the conditioning of the Hessian
- What is the effect of preconditioning on the problem
- How does the observational system affect the conditioning of the problem

Reference: (HLN)

Haben, Lawless and Nichols, *UoR Maths Rpt 3/2009*

2.

Conditioning of the Hessian

Optimal Bayesian Estimate – 3DVar

Minimizes with respect to initial state $\mathbf{x}_0$:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}(\mathcal{H}(\mathbf{x}_0) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}_0) - \mathbf{y}),$$

- Hessian is:

$$\mathbf{A} = \mathbf{B}^{-1} + (\mathbf{H})^T \mathbf{R}^{-1} \mathbf{H}$$

- $\mathbf{B}$ and $\mathbf{R}$ are covariance matrices that depend on the correlation length-scales of the errors; $\mathbf{H}$ is the Jacobian of $\mathcal{H}$.

Optimal Bayesian Estimate – 3DVar

Minimizes with respect to initial state $\mathbf{x}_0$:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}(\mathcal{H}(\mathbf{x}_0) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}_0) - \mathbf{y}),$$

- Hessian is:

$$\boxed{\mathbf{A} = \mathbf{B}^{-1} + (\mathbf{H})^T \mathbf{R}^{-1} \mathbf{H}}$$

- $\mathbf{B}$ and $\mathbf{R}$ are covariance matrices that depend on the correlation length scales of the errors; $\mathbf{H}$ is the Jacobian of $\mathcal{H}$.

Background Covariance Matrix

Define $\mathbf{B} = \sigma_b^2 \mathbf{C}$ on a 1D periodic domain by a Gaussian correlation structure:

$$(\mathbf{B})_{i,j} = \sigma_b^2 exp\left(\frac{-r_{i,j}^2}{2L^2}\right)$$

- $r_{i,j}$ distance between points at position i and j
- L correlation lengthscale
- σ_b^2 background error variance
- Circulant

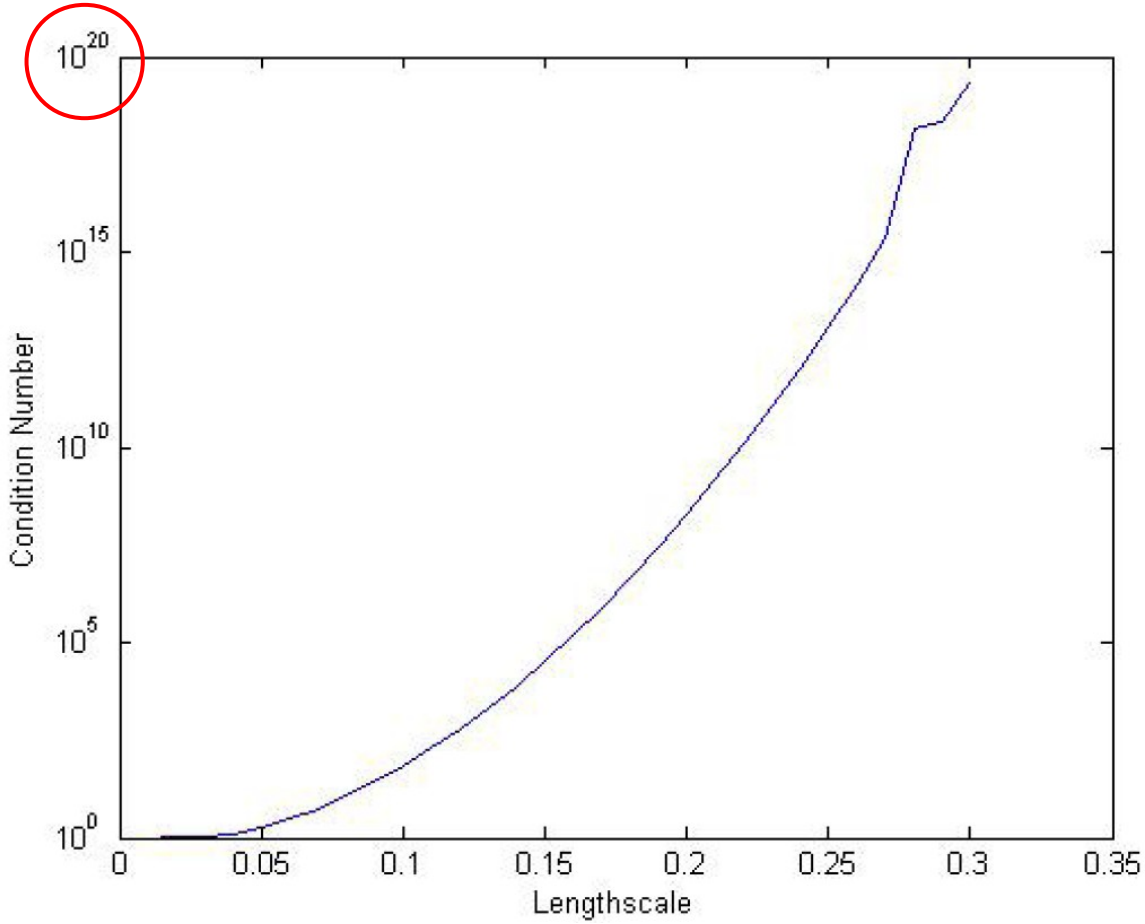
Circulant Matrices

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & \cdots & c_{N-2} & c_{N-1} \\ c_{N-1} & c_0 & c_1 & c_2 & \cdots & c_{N-3} & c_{N-2} \\ c_{N-2} & c_{N-1} & c_0 & & & & \vdots \\ & & \ddots & \ddots & & & \\ \vdots & & & & \ddots & & c_2 \\ c_2 & & & & \ddots & c_0 & c_1 \\ c_1 & c_2 & \cdots & & c_{N-2} & c_{N-1} & c_0 \end{pmatrix}$$

Eigenvalues

$$\lambda_m = \sum_{k=0}^{N-1} c_k e^{-2\pi i m k/N}.$$

Condition of B - Gaussian



Condition of Hessian

$$\mathbf{A} = \mathbf{B}^{-1} + (\mathbf{H})^T \mathbf{R}^{-1} \mathbf{H}$$

- $\mathbf{B} = \sigma_b^2 \mathbf{C}$, where $\mathbf{C}$ is a correlation matrix
- Assume observations are at grid points
- Assume observation errors uncorrelated
- σ_0^2 variance observation errors
- $\mathbf{R} = \sigma_0^2 \mathbf{I}$
- $\mathbf{H}$ linear.
- $\mathbf{H}^T \mathbf{H}$ diagonal.

Condition Number of Hessian

Assume **C** has a circulant covariance structure.

Bounds on the conditioning of the Hessian are:

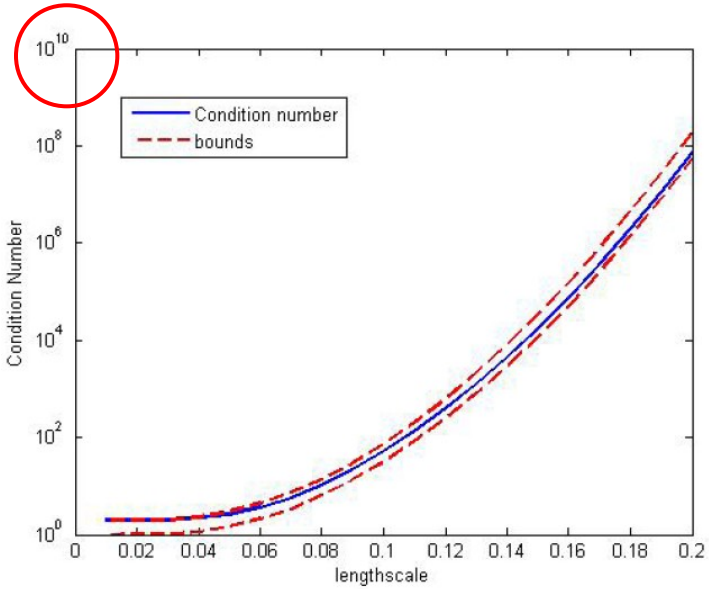
$$\alpha\kappa(\mathbf{C}) \leq \kappa(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \leq \left(1 + \left(\frac{\sigma_b^2}{\sigma_o^2}\right)\lambda_{\min}(\mathbf{C})\right)\kappa(\mathbf{C})$$

where $\alpha = \left(\dfrac{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\min}(\mathbf{C})}{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\max}(\mathbf{C})}\right)$

and p = number of observations

Conditioning of Hessian

Condition Number of $(B^{-1} + HR^{-1}H^T)$ vs Length Scale



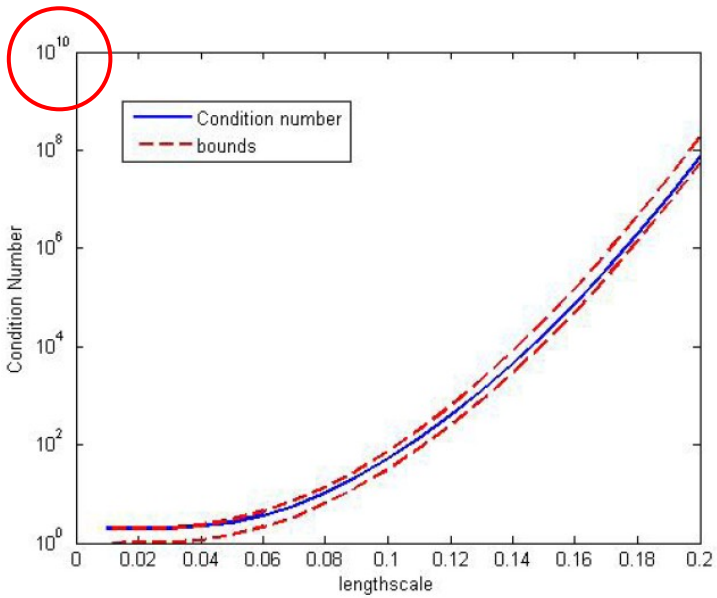
Periodic Gaussian Exponential

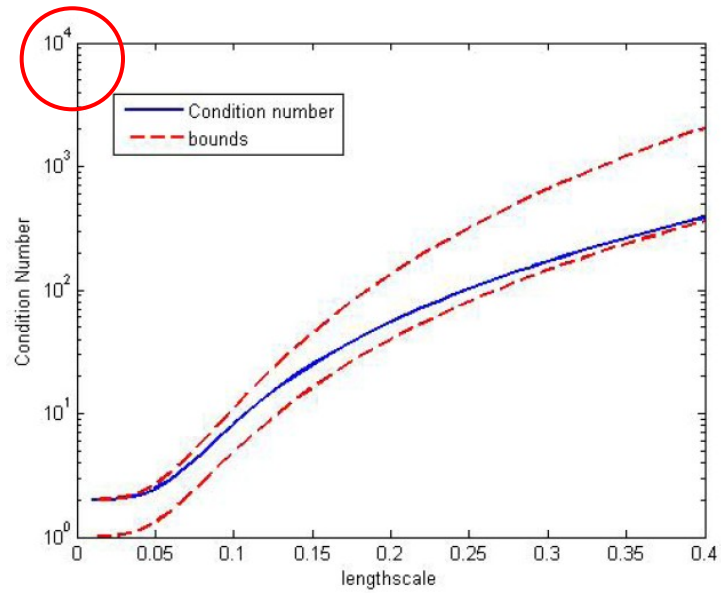
$$\mathbf{B}_{ij} = \sigma_b^2 \exp\left(\frac{-r_{i,j}^2}{2L^2}\right)$$

Blue = condition number Red = bounds

Conditioning of Hessian

Condition Number of $(B^{-1} + HR^{-1}H^T)$ vs Length Scale





Periodic Gaussian Exponential

$$\mathbf{B}_{ij} = \sigma_b^2 \exp\left(\frac{-r_{i,j}^2}{2L^2}\right)$$

Laplacian 2nd Derivative

$$\mathbf{B}^{-1} = \gamma^{-1}\left(\mathbf{I} + \frac{l^4}{2\Delta x^4}(\mathbf{L})^2\right)$$

Blue = condition number Red = bounds

3.

Preconditioning of the Hessian

Control Variable Transform

To improve conditioning transform to new variable :

- $\mathbf{z} = \mathrm{B}^{1/2}(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}})$
- Uncorrelated variable
- Equivalent to preconditioning by $\mathrm{B}^{1/2}$
- Hessian of transformed problem is

$$\boxed{\mathrm{I} + \mathrm{B}^{1/2}\hat{\mathrm{H}}^T\hat{\mathrm{R}}^{-1}\hat{\mathrm{H}}\mathrm{B}^{1/2}}$$

Preconditioned Hessian

Bounds on the conditioning of the preconditioned Hessian are:

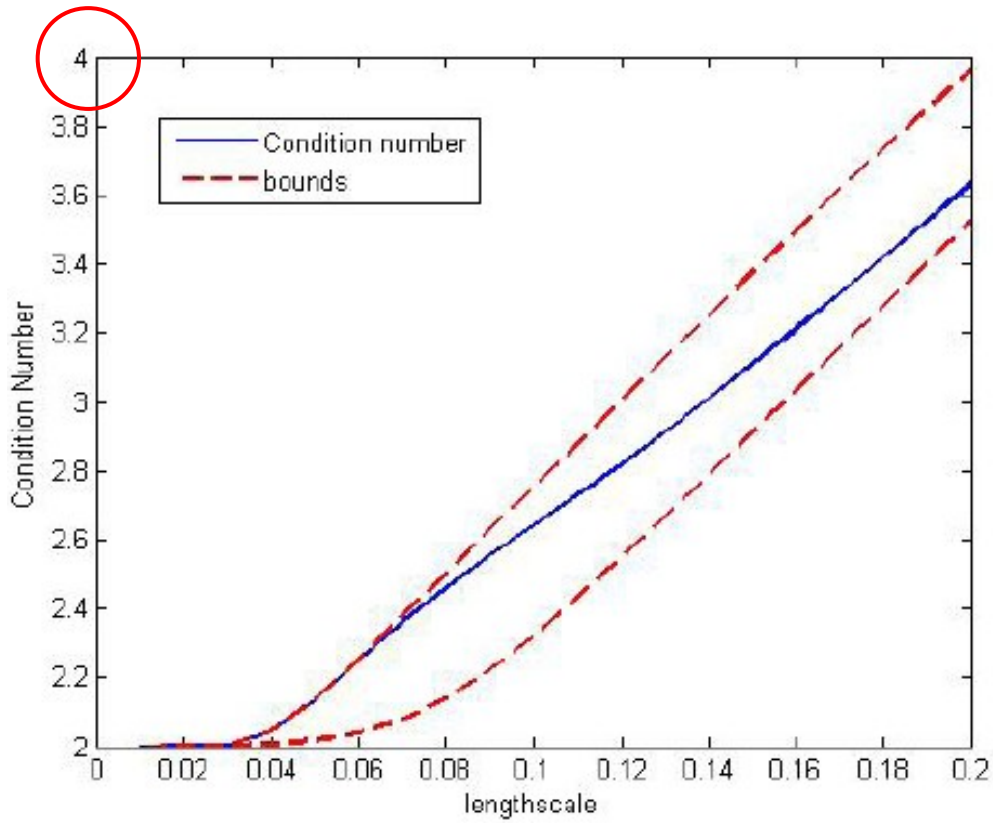
$$1 + \frac{\sigma_b^2}{\sigma_o^2}\gamma \leq \kappa(\mathbf{I} + \sigma_o^{-2}(\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{H}\mathbf{B}^{1/2})) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2}\nu_0$$

where

$$\nu_0 = \| \mathbf{HCH^T} \|_\infty \ , \quad \gamma = \frac{1}{p}\sum_{i,j\in J} c_{i,j} \ .$$

ν_0 changes slowly as a function of length scale.

Preconditioned Hessian - Gaussian



Preconditioned (blue)
Bounds (red)

Condition number as a function of length scale

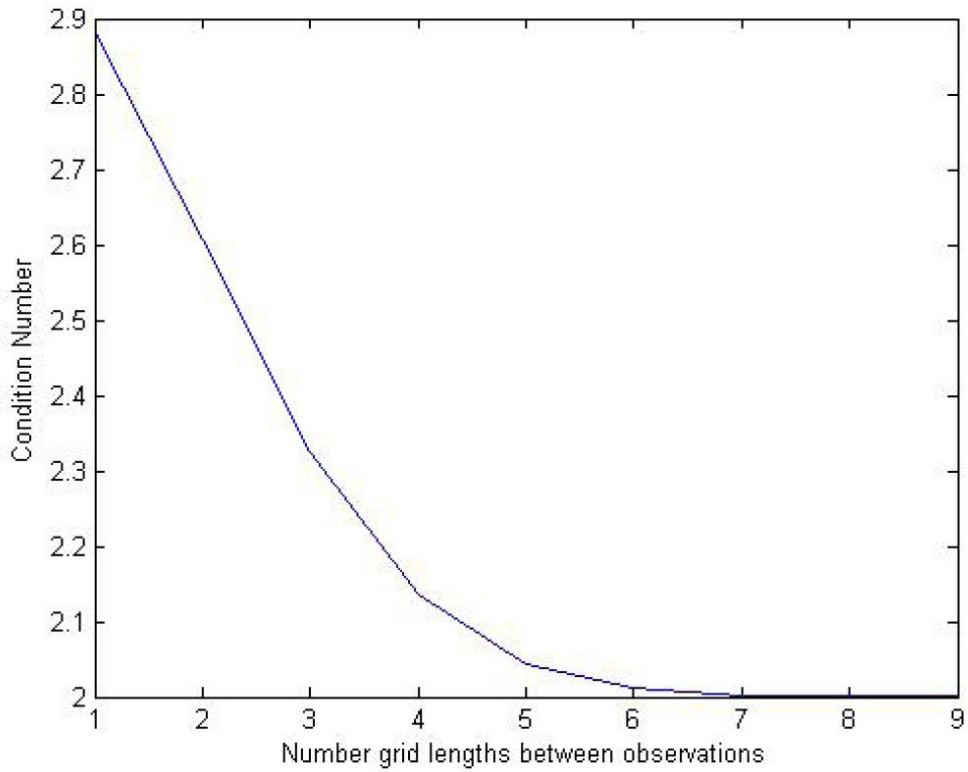
Preconditioned Hessian - Gaussian

Assume two observations at k^{th} and m^{th} grid points

$$\kappa(\mathbf{I} + \sigma_o^{-2}(\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{H}\mathbf{B}^{1/2})) = 1 + \frac{\sigma_b^2}{\sigma_o^2} + \frac{\sigma_b^2}{\sigma_o^2}\,|c_{k,m}|$$

Condition number decreases as the separation of the observations increases

Preconditioned Hessian - Gaussian



Condition number as a function of observation spacing

Extension to 4DVar

Convergence depends on condition number of

$$\boxed{\mathbf{I} + \mathbf{B}^{1/2}\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}\mathbf{B}^{1/2}}$$

where

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_n\mathbf{M}_{0,n} \end{pmatrix} \qquad \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_n \end{pmatrix}$$

$$\mathbf{M}_{0,k} = \frac{\partial \mathcal{M}_{0,k}}{\partial \mathbf{x}}\Big|_{\mathbf{x}_0} \qquad \mathbf{H}_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}\Big|_{\mathcal{M}_{0,k}(\mathbf{x}_0)}$$

Preconditioned Hessian

Bounds on the conditioning of the preconditioned Hessian are:

$$1 + \frac{1}{p(n+1)} \frac{\sigma_b^2}{\sigma_o^2} \sum_{i,j=1}^{p(n+1)} (\hat{\mathbf{H}}\mathbf{C}\hat{\mathbf{H}}^T)_{i,j} \leq \kappa(\mathbf{A}_p) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2} ||\hat{\mathbf{H}}\mathbf{C}\hat{\mathbf{H}}^T||_\infty$$

where

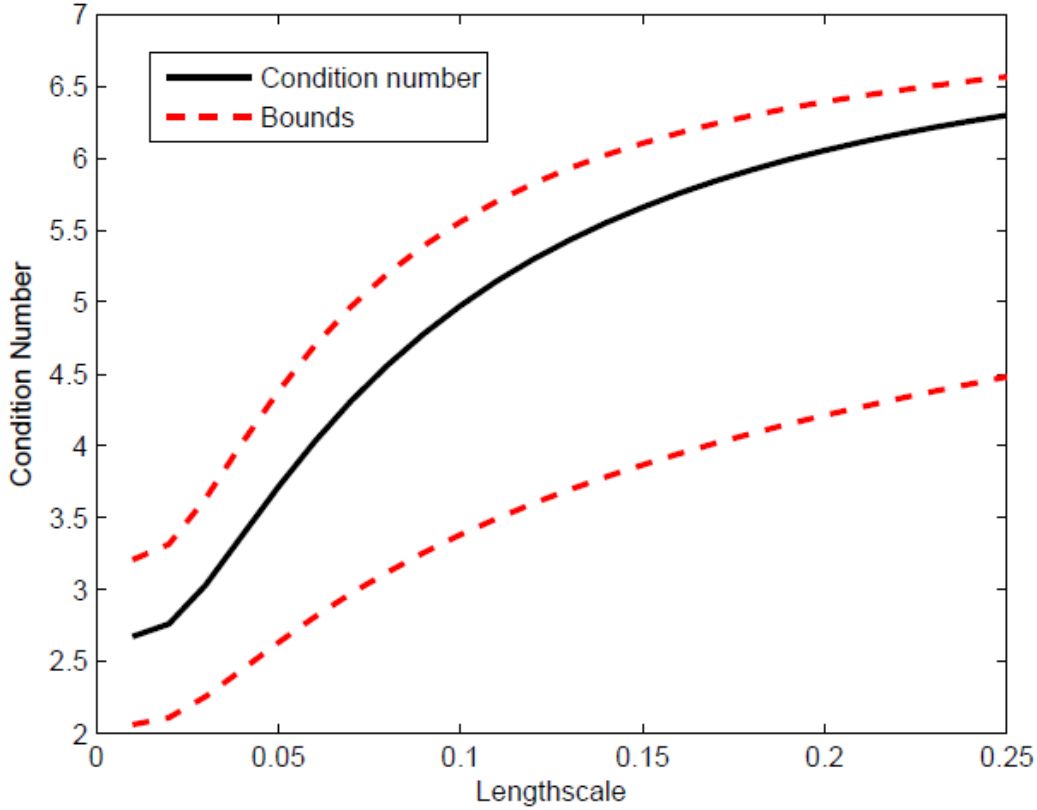
- $\mathbf{B} = \sigma_b^2 \mathbf{C}$, $\mathbf{C}$ is correlation matrix
- $\mathbf{R}_k = \sigma_0^2 \mathbf{I}$ for $k = 0, \ldots, n$
- Advection model discretized using upwind scheme

4D Background Correlation Matrix

$$\begin{pmatrix} \mathbf{B} & \mathbf{B}\mathbf{M}_1^T & \mathbf{B}\mathbf{M}_2^T & \cdots & \mathbf{B}\mathbf{M}_n^T \\ \mathbf{M}_1\mathbf{B} & \mathbf{M}_1\mathbf{B}\mathbf{M}_1^T & \mathbf{M}_1\mathbf{B}\mathbf{M}_2^T & \cdots & \mathbf{M}_1\mathbf{B}\mathbf{M}_n^T \\ \mathbf{M}_2\mathbf{B} & \mathbf{M}_2\mathbf{B}\mathbf{M}_1^T & \mathbf{M}_2\mathbf{B}\mathbf{M}_2^T & & \mathbf{M}_2\mathbf{B}\mathbf{M}_n^T \\ & & \ddots & \ddots & \\ \vdots & & & \ddots & \\ & & & \ddots & \\ \mathbf{M}_n\mathbf{B} & \mathbf{M}_n\mathbf{B}\mathbf{M}_1^T & \mathbf{M}_n\mathbf{B}\mathbf{M}_2^T & & \mathbf{M}_n\mathbf{B}\mathbf{M}_n^T \end{pmatrix}.$$

Each block describes a covariance matrix between the background errors at different times.

Condition of Preconditioned 4DVar – using SOAR Correlation Matrix



Convergence Rates of CG in 4DVar – using SOAR Correlation Matrix

	Iterations	
Lengthscale	Unprecond	Precond
0.01	8	8
0.1	54	11
0.2	187	12
0.3	361	12

Haben et al, 2011

4.

Regularization

Preconditioned Hessian – 4DVar

Let $\mathbf{B} = \sigma_b^2 \mathbf{C}_B \qquad \hat{\mathbf{R}} = \sigma_o{}^2 \mathbf{I} \qquad \mu^2 = \dfrac{\sigma_o^2}{\sigma_b^2}$

Then the preconditioned optimal state estimation problem may be written in the form of a classical Tikhonov regularized problem:

$$\boxed{\hat{J} = \mu^2 ||\mathbf{z}||_2^2 + ||\hat{\mathbf{H}}\mathbf{C}_B^{1/2}\mathbf{z} - \hat{\mathbf{d}}||_2^2}$$

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_n\mathbf{M}_{0,n} \end{pmatrix} \qquad \hat{\mathbf{d}} = \begin{pmatrix} \mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0^b) \\ \mathbf{y}_1 - \mathcal{H}_1(\mathbf{x}_1^b) \\ \vdots \\ \mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n^b) \end{pmatrix}$$

Information Content

Rewrite the 4D Var solution as:

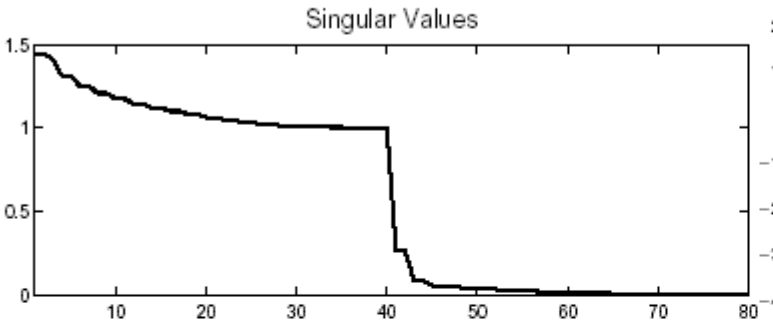
$$\mathbf{z} = \sum_i \frac{\lambda_i^2}{\mu^2 + \lambda_i^2} \frac{\mathbf{u}_i^T \hat{\mathbf{d}}}{\lambda_i} \mathbf{v}_i$$

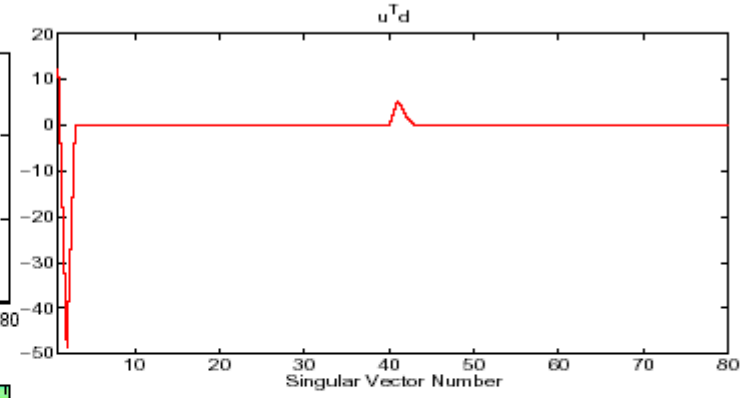
where $\mu^2 = \frac{\sigma_o^2}{\sigma_b^2}$ and λ_i, $\mathbf{v}_i$, $\mathbf{u}_i$ are

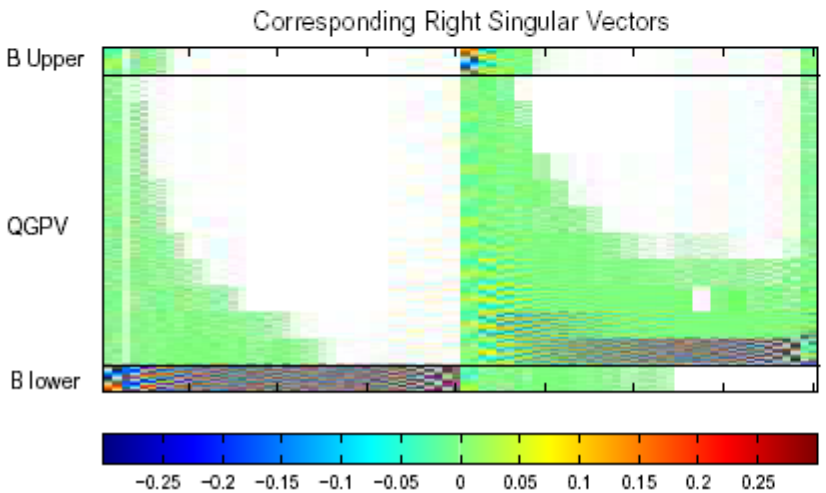
the singular values and right and left singular vectors of $\hat{\mathbf{H}}\mathbf{C}_B^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$.

Johnson et al, 2005a,b

SVD







2 pairs of RSVs contribute towards the analysis increment. The second pair lie in the near-null space.

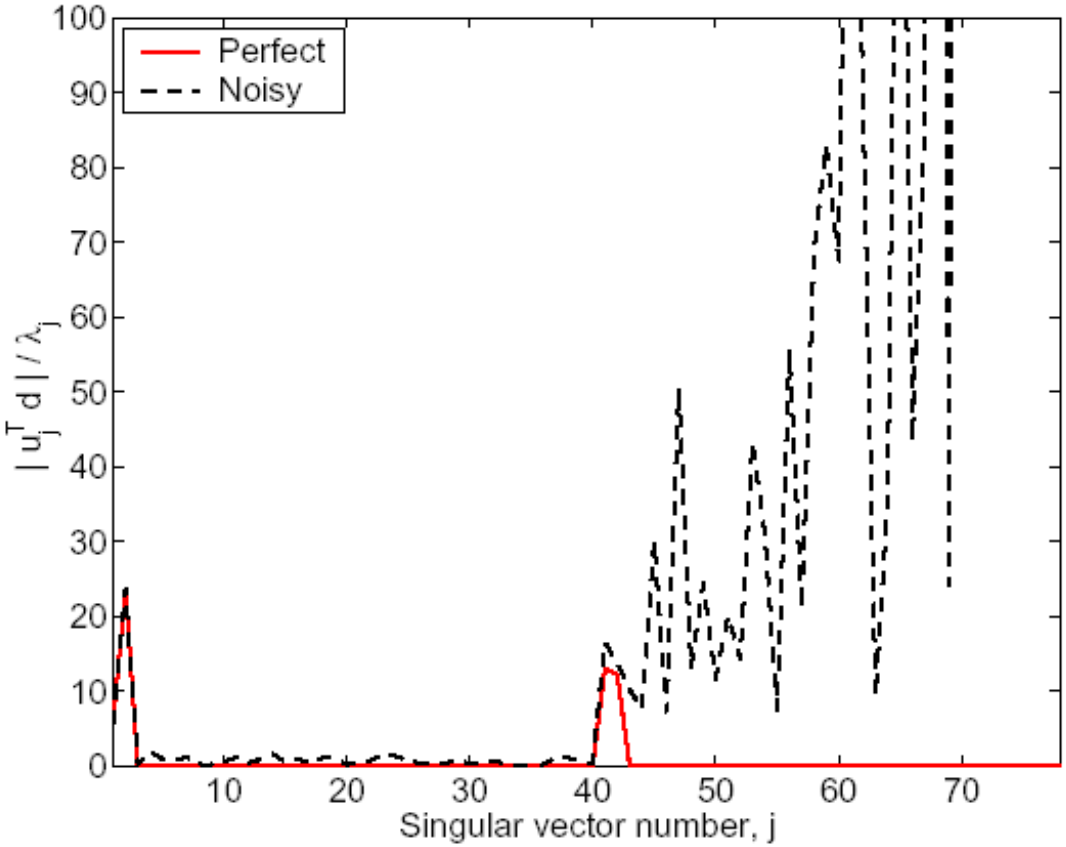
Johnson et al, 2005a,b

Effect of Observational Noise

Values of

$$\left| \frac{\mathbf{u}_j^T \hat{\mathbf{d}}}{\lambda_j} \right|$$

for perfect and noisy observations.



Johnson et al, 2005a,b

5.

Conclusions

Conclusions

- The condition number of the Hessian is sensitive to lengthscale
- Preconditioning with $\mathbf{B}^{1/2}$ generally reduces the condition number and improves the convergence rates
- The condition number of the Hessian increases as the observations become more accurate and as their separation decreases.

References

S.A. Haben, A.S. Lawless and N.K. Nichols, Conditioning of incremental variational data assimilation, with application to the Met Office system, *Tellus,* **63A**, 2011, 782 – 792.

S.A. Haben, A.S. Lawless and **N.K. Nichols**, Conditioning and preconditioning of the variational data assimilation problem, *Computers & Fluids*, **46**, 2011, 252 – 256.

S.A. Haben, Conditioning and preconditioning of the minimisation problem in variational data assimilation, University of Reading, Dept of Mathematics, PhD Thesis, 2011.

C. Johnson, B.J. Hoskins and **N.K. Nichols**, A singular vector perspective of 4-DVar: Filtering and interpolation, *Q. J. Roy. Met. Soc.,* **131**, 2005, 1-20.

C. Johnson, **N.K. Nichols** and B.J. Hoskins, Very large inverse problems in atmosphere and ocean modelling, *Internat. J. for Numer. Methods Fluids*, **47**, 2005, 759-771.

Future

Many more challenges left!