Improving the conditioning of estimated covariance matrices

Jemima M. Tabeart

Supervised by

Sarah L. Dance, Nancy K. Nichols, Amos S. Lawless, Joanne A. Waller (University of Reading) Sue Ballard, David Simonin (MetOffice@Reading)

Additional collaboration

Stefano Migliorini and Fiona Smith (Met Office, Exeter)

July 5, 2018

Motivation

- Including correlated observation error allows us to maximise the information content of observations.
- But diagnosed correlated covariance matrices have caused problems with convergence of the data assimilation minimisation procedure.
- Need to treat diagnosed matrices (symmetry, positive definiteness).

What is reconditioning?

- *Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.*
- Examples of methods:
    - Thresholding
    - Tapering
    - General regularisation methods.

What is reconditioning?

- *Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.*
- Examples of methods:
  - Thresholding
  - Tapering
  - General regularisation methods.
- We will focus on two methods that are used in NWP.
- Both work by altering eigenvalues of the covariance matrix.

What is reconditioning?

- *Methods which can be applied to matrices to reduce their condition number, while retaining underlying matrix structure.*
- Examples of methods:
    - Thresholding
    - Tapering
    - General regularisation methods.
- We will focus on two methods that are used in NWP.
- Both work by altering eigenvalues of the covariance matrix.

Reminder: If $\mathbf{S} \in \mathbb{R}^{p \times p}$ *is a symmetric and positive definite matrix with eigenvalues* $\lambda_1(\mathbf{S}) \geq \ldots \geq \lambda_p(\mathbf{S}) > 0$ *then we can write the condition number in the* L_2 *norm as*

$$\kappa(\mathbf{S}) = \frac{\lambda_1(\mathbf{S})}{\lambda_p(\mathbf{S})}.$$

If $\mathbf{S}$ *is singular, we take* $\kappa(\mathbf{S}) = \infty$.

The ridge regression (RR) and minimum eigenvalue (ME) methods

Both methods improve the condition number of a covariance matrix by altering their eigenvalues to yield a reconditioned matrix with a user-defined condition number κ_{max}.

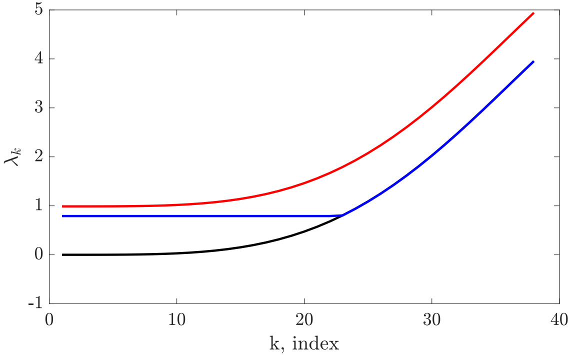


Figure: Illustration of recond methods: original spectrum (black), and spectrum reconditioned via ME and RR

Ridge regression method

Idea: Add a scalar multiple of identity to $\mathbf{R}$ *to obtain reconditioned* $\mathbf{R}_{RR}$ *with* $\kappa(\mathbf{R}_{RR}) = \kappa_{max}$.

Setting δ

- *Define* $\delta = \frac{\lambda_1(\mathbf{R}) - \lambda_p(\mathbf{R})\kappa_{max}}{\kappa_{max} - 1}$.
- *Set* $\mathbf{R}_{RR} = \mathbf{R} + \delta \mathbf{I}$

Ridge regression method

Idea: Add a scalar multiple of identity to $\mathbf{R}$ *to obtain reconditioned* $\mathbf{R}_{RR}$ *with* $\kappa(\mathbf{R}_{RR}) = \kappa_{max}$.

Setting δ

- *Define* $\delta = \frac{\lambda_1(\mathbf{R}) - \lambda_p(\mathbf{R})\kappa_{max}}{\kappa_{max} - 1}$.
- *Set* $\mathbf{R}_{RR} = \mathbf{R} + \delta\mathbf{I}$

- Similar to Steinian linear shrinkage [Ledoit and Wolf, 2004]
- Used at the Met Office [Weston et al, 2014].

Theory of the ridge regression method

Effect of RR on standard deviations:

$$\mathbf{\Sigma}_{RR} = (\mathbf{\Sigma}^2 + \delta \mathbf{I}_p)^{1/2}. \quad (1)$$

i.e. variances are increased by the reconditioning constant, δ.

Theory of the ridge regression method

Effect of RR on standard deviations:

$$\mathbf{\Sigma}_{RR} = (\mathbf{\Sigma}^2 + \delta \mathbf{I}_p)^{1/2}. \tag{1}$$

i.e. variances are increased by the reconditioning constant, δ.

Effect of RR on correlations:

$$\text{For } i \neq j, |\mathbf{C}_{RR}(i,j)| < |\mathbf{C}(i,j)| \tag{2}$$

i.e. the magnitude of all off-diagonal correlations is strictly decreased.

Minimum eigenvalue method

Idea: Fix a threshold, T, below which all eigenvalues of the reconditioned matrix, $\mathbf{R}_{ME}$, are set equal to T to yield $\kappa(\mathbf{R}_{ME}) = \kappa_{max}$.

Setting T:

- *Set $\lambda_1(\mathbf{R}_{ME}) = \lambda_1(\mathbf{R})$*
- *Define $T = \lambda_1(\mathbf{R})/\kappa_{max}$.*
- *Set the remaining eigenvalues of $\mathbf{R}_{ME}$ via*

$$\lambda_k(\mathbf{R}_{ME}) = \begin{cases} \lambda_k(\mathbf{R}) & \text{if } \lambda_k(\mathbf{R}) > T \\ T & \text{if } \lambda_k(\mathbf{R}) \leq T \end{cases}. \tag{3}$$

- *We define $\mathbf{\Gamma}(k,k) = max\{T - \lambda_i, 0\}$.*

Minimum eigenvalue method

Idea: Fix a threshold, T, below which all eigenvalues of the reconditioned matrix, $\mathbf{R}_{ME}$, are set equal to T to yield $\kappa(\mathbf{R}_{ME}) = \kappa_{max}$.

Setting T:

- *Set $\lambda_1(\mathbf{R}_{ME}) = \lambda_1(\mathbf{R})$*
- *Define $T = \lambda_1(\mathbf{R})/\kappa_{max}$.*
- *Set the remaining eigenvalues of $\mathbf{R}_{ME}$ via*

$$\lambda_k(\mathbf{R}_{ME}) = \begin{cases} \lambda_k(\mathbf{R}) & \text{if } \lambda_k(\mathbf{R}) > T \\ T & \text{if } \lambda_k(\mathbf{R}) \leq T \end{cases}. \quad (3)$$

- *We define $\mathbf{\Gamma}(k,k) = max\{T - \lambda_i, 0\}$.*

A variant of this method is used at the European Centre for Medium-Range Weather Forecasts (ECMWF) [Bormann et al, 2016].

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\boldsymbol{\Sigma}_{ME}(i,i) = \left(\boldsymbol{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \boldsymbol{\Gamma}(k,k)\right)^{1/2} \tag{4}$$

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\boldsymbol{\Sigma}_{ME}(i,i) = \left(\boldsymbol{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \boldsymbol{\Gamma}(k,k) \right)^{1/2} \tag{4}$$

This can be bounded by

$$\boldsymbol{\Sigma}(i,i) \leq \boldsymbol{\Sigma}_{ME}(i,i) \leq \left(\boldsymbol{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R}) \right)^{1/2}. \tag{5}$$

Theory of the minimum eigenvalue method

Effect of ME on standard deviations:

$$\boldsymbol{\Sigma}_{ME}(i,i) = \left(\boldsymbol{\Sigma}(i,i)^2 + \sum_{k=1}^{p} \mathbf{V}_R(i,k)^2 \boldsymbol{\Gamma}(k,k) \right)^{1/2} \tag{4}$$

This can be bounded by

$$\boldsymbol{\Sigma}(i,i) \leq \boldsymbol{\Sigma}_{ME}(i,i) \leq \left(\boldsymbol{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R}) \right)^{1/2}. \tag{5}$$

Effect of ME on correlations:

It is not evident how correlation entries are altered in general.
This is due to the fact that the spectrum of **R** is not altered uniformly by this method.

Comparison of both methods

- Both methods increase (or maintain) standard deviations

Comparison of both methods

- Both methods increase (or maintain) standard deviations
- We can show that $T - \lambda_p(\mathbf{R}) < \delta$ which yields:

$$\mathbf{\Sigma}_{ME}(i,i) \leq \left(\mathbf{\Sigma}(i,i)^2 + T - \lambda_p(\mathbf{R})\right)^{1/2} < (\mathbf{\Sigma}(i,i)^2 + \delta)^{1/2} = \mathbf{\Sigma}_{RR}(i,i)$$

- Therefore RR increases standard deviations more than ME

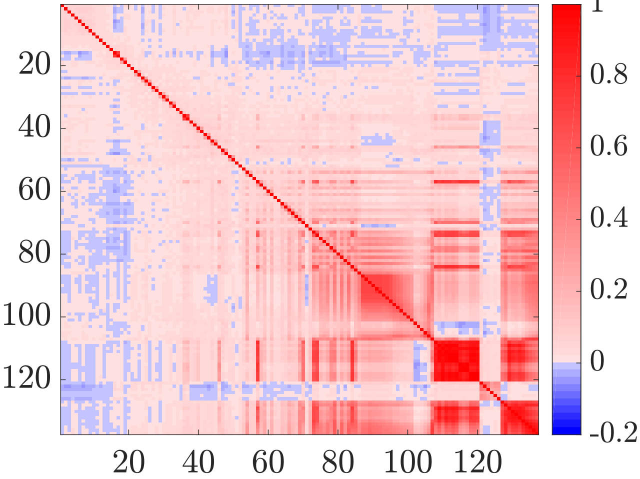
IASI matrix - framework

Interchannel correlations for a covariance matrix of satellite observation errors

- The UK Met Office diagnosed a correlated observation error covariance matrix in 2011.
- This was extremely ill-conditioned and crashed the system when used directly.
- Recondition 137 channels.

Original condition number: 27703.

Diagnosed IASI correlation matrix



IASI - change to standard deviations

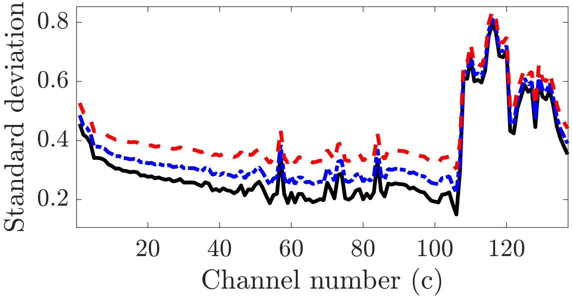


Figure: Standard deviations $\mathbf{\Sigma}$ (solid), $\mathbf{\Sigma}_{RR}$ and $\mathbf{\Sigma}_{ME}$ for $\kappa_{max} = 100$. Recall $\kappa(\mathbf{R}) = 27703.45$ for the original IASI matrix.

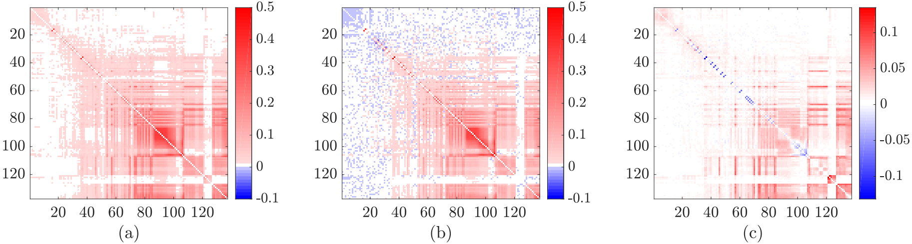


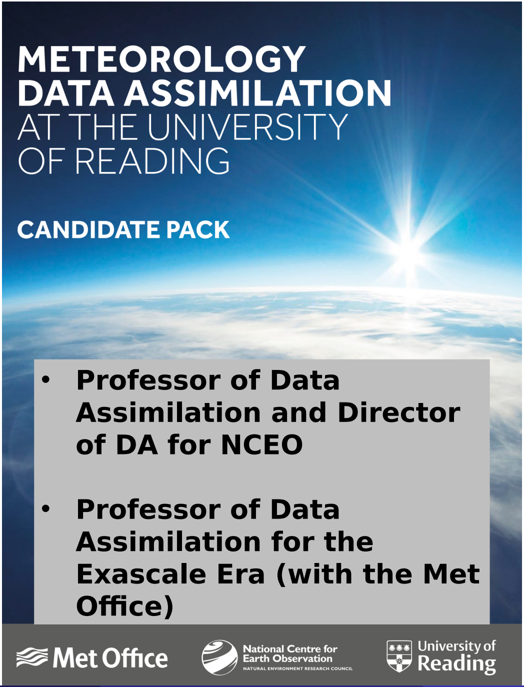
Figure: Difference in correlations (a) $(\mathbf{C} - \mathbf{C}_{RR}) \circ sign(\mathbf{C})$, (b) $(\mathbf{C} - \mathbf{C}_{ME}) \circ sign(\mathbf{C})$, and (c) $(\mathbf{C}_{ME} - \mathbf{C}_{RR}) \circ sign(\mathbf{C})$. The colorscale is the same for (a) and (b) but different for (c).

Ongoing work

- Focused on improving conditioning and the relationship between condition number and convergence of a conjugate gradient type minimisation.
- But by reconditioning we are altering the problem we are solving!
- Studying impact of reconditioning on an operational system - 1D-Var QC procedure used at the Met Office.
- Focussed on RR - run experiments comparing uncorrelated $\mathbf{R}$ with different values of δ.
- Increasing δ improves convergence - better than current operational choice of $\mathbf{R}$!
- But difficult to assess impacts on retrieved variables:
  Differences between retrieved values mostly small for ctrl - expt
  But some very large differences which are hard to explain

Conclusions and ongoing work

- Developed new theory showing how two reconditioning methods alter covariance matrices theoretically.
- The ridge regression method increases standard deviations more than the minimum eigenvalue method.
- The ridge regression method moves all correlations closer to zero, whereas the minimum eigenvalue method can increase correlations.
- The ridge regression method changes most correlation entries by a larger amount than the minimum eigenvalue method in numerical testing.


METEOROLOGY
DATA ASSIMILATION
AT THE UNIVERSITY
OF READING
CANDIDATE PACK
• Professor of Data Assimilation and Director of DA for NCEO
• Professor of Data Assimilation for the Exascale Era (with the Met Office)
Met Office
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading

References I

J. M. Tabeart, S. L. Dance, S. A. Haben, A. S. Lawless, N. K. Nichols, and J. A. Waller (2018)
The conditioning of least squares problems in variational data assimilation.
Numerical Linear Algebra with Applications http://dx.doi.org/10.1002/nla.2165

P. Weston, W. Bell and J. R. Eyre (2014)
Accounting for correlated error in the assimilation of high-resolution sounder data
Q. J. R Met Soc 140, 2420 – 2429.

Niels Bormann, Massimo Bonavita, Rossana Dragani, Reima Eresmaa, Marco Matricardi, and Anthony McNally (2016)
Enhancing the impact of IASI observations through an updated observation error covariance matrix
doi: 10.1002/qj.2774

S. Rainwater, C. H. Bishop and W. F. Campbell (2015)
The benefits of correlated observation errors for small scales
Q. J. R. Met. Soc. 141, 3439–3445

References II

O. Ledoit and M. Wolf (2004)
A well-conditioned estimator for large-dimensional covariance matrix.
J. Multivariate Anal. 88:365-411

M. Tanaka and K. Nakata (2014)
Positive definite matrix approximation with condition number constraint.
Optim. Lett. 8:939947.

Laura Stewart (2010)
Correlated observation errors in data assimilation
PhD thesis University of Reading

SOAR matrix - simple framework

Spatial correlations for a simple numerical experiment

- Use a second-order auto-regressive function to construct a circulant correlation matrix - i.e. a matrix fully defined by its first row.
- We fix the standard deviations to be constant for all variables.
- In this framework we have update the standard deviations for the minimum eigenvalue method via:

$$\mathbf{\Sigma}_{ME}(i,i) = \left(\mathbf{\Sigma}(i,i)^2 + \frac{1}{p}\sum_{k=1}^{p} \mathbf{\Gamma}(k,k) \right)^{1/2}. \tag{6}$$

Original condition number: 81121.

SOAR matrix - results

Table: Change to standard deviations of the SOAR matrix.

κ_{max}	σ	σ_{RR}	% change RR	σ_{ME}	% change ME
1000	2.23606	2.26471	+1.28%	2.25439	+0.82%
500	2.23606	2.29340	+2.56%	2.27599	+1.79%
100	2.23606	2.51306	+12.39%	2.45737	+9.90%

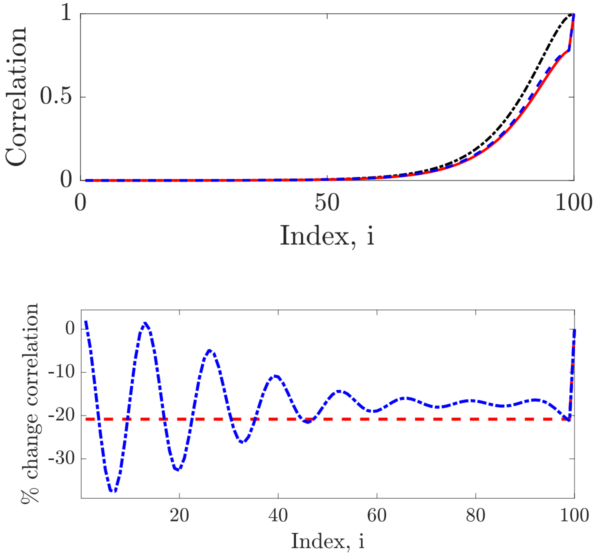
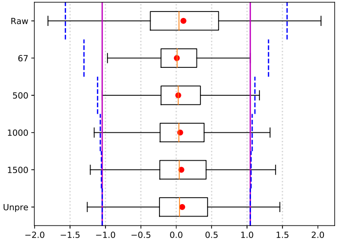
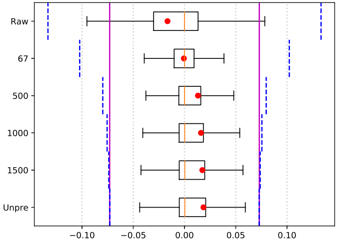


Figure: Changes to correlations between the 100th row of the original SOAR matrix and the reconditioned matrices (for $\kappa_{max} = 100$). (a) $\mathbf{C}$ (black), $\mathbf{C}_{RR}$, $\mathbf{C}_{ME}$ (b) $100 \times \frac{\mathbf{C}-\mathbf{C}_{RR}}{\mathbf{C}}$ and $100 \times \frac{\mathbf{C}-\mathbf{C}_{ME}}{\mathbf{C}}$.

Changes to retrievals in 1D-Var





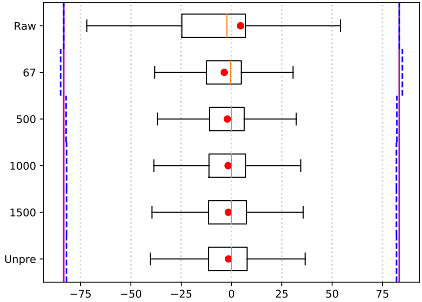


Figure: (a) Skin temperature, (b) cloud fraction, (c) cloud top pressure