Conditioning of the 3DVAR Data Assimilation Problem

S. A. Haben, A. S. Lawless and N. K. Nichols
Department of Mathematics
University of Reading

September 2009

Abstract

Variational data assimilation schemes are commonly used in major numerical weather prediction (NWP) centres around the world. The convergence of the variational scheme and the sensitivity of the analysis to perturbations are dependent on the conditioning of the Hessian of the linearized least-squares variational equation. The problem is ill-conditioned and hence is di±cult to solve quickly and accurately. To make the scheme operationally feasible, NWP centres perform a control variable transform with the aim of preconditioning the problem to reduce the condition number of the Hessian. In this paper we investigate the conditioning of the 3DVar problem for a single periodic system parameter. We give bounds on the condition number of both the original and preconditioned 3DVar problems and demonstrate the reasons for the superior performance of the preconditioned system. We also exhibit the e®ect of the observation error variances and the positions of the observations on the conditioning of the system.

1 Introduction

Variational data assimilation is popularly used in numerical weather and ocean forecasting to combine observations with a model forecast in order to produce a best estimate of the current state of the system and enable accurate prediction of future states. The estimate minimizes a weighted nonlinear least-squares measure of the error between the model forecast and the available observations and is found using an iterative optimization algorithm. Under certain statistical assumptions the solution to the variational data assimilation problem, known as the *analysis*, yields the *maximum a posteriori* Bayesian estimate of the state of the system [15].

In practice an incremental version of the variational scheme is implemented in many operational centres, including the Met O±ce [19] and the European Centre for Medium-Range Weather Forecasting (ECMWF) [18]. This method solves a sequence of linear approximations to the nonlinear least-squares problem and is equivalent to an approximate Gauss-Newton method for determining the analysis [14]. Each approximate linearised least-squares problem is solved using an inner gradient iteration method, such as the conjugate gradient method, and the linearization state is then updated in an outer iteration loop. Generally only a very few outer loops are performed. For use in operational forecasting the complete iteration scheme must produce an accurate solution to the variational problem rapidly, in *real* time.

The rate of convergence of the inner loop of the variational scheme and the sensitivity of the solution to perturbations are largely determined by the condition number, that is, the ratio of the largest and smallest eigenvalues, of the Hessian of the linear least-squares objective function [8]. Experimental results indicate that in operational systems the Hessian is ill-conditioned and that this is a result of the ill-conditioning of the background error covariance matrix [16]. In practice the system is preconditioned by transforming the state variables to new variables where the errors are assumed to be approximately uncorrelated [2]. Experimental comparisons have demonstrated that the preconditioning significantly improves the speed and accuracy of the assimilation scheme [7], [16].

Explanations are offered in the literature for the ill-conditioning of the variational assimilation problem and for the benefits of preconditioning in the operational system [1],[20],[16]. In [1], an analysis of the preconditioned system in a simplified 3DVar system with only 2 grid points shows that the conditioning of the preconditioned Hessian is dependent on the accuracy and density of observations. In their paper, Andersson *et al.* take p observations at each grid point with error variance σ_o^2 and a background with error variance σ_b^2 and find an approximation to the condition number given by

$$\kappa \approx 2p\left(\frac{\sigma_b^2}{\sigma_o^2}\right) + 1. \qquad (1)$$

This approximation is supported experimentally in the ECMWF operational 4DVar system in [20], where it is shown that for dense surface observations, the conditioning of the problem improves as the observations become less accurate. The causes for poor conditioning for dense observations are thus attributed to accurate observations, increasing number of observations (larger p) and large background error variances.

In this paper we examine the conditioning and preconditioning of a more general 3DVar problem theoretically. We derive expressions for the eigenvalues and hence bounds on the conditioning of the Hessian of the problem in the case of a single, periodic, spatially-distributed system parameter. We consider three questions: how does the condition number of the Hessian depend on the length-scale in the correlation structures; how does preconditioning compare with the conditioning of the original Hessian; and how do the error variances of the observations and the distances between observations affect the conditioning of the Hessian.

In the next section we introduce the incremental variational assimilation method. In Section 3 we look at the conditioning of two particular background error covariance matrices. We consider the conditioning of the Hessian and the preconditioned Hessian in Sections 4 and 5. In Section 6 we investigate how the position of observations affects the conditioning and in Section 7 we summarize the conclusions.

2 Variational Data Assimilation

The aim of the variational assimilation scheme is to find an optimal estimate for the initial state of the system x_0 (the *analysis*) at time t_0 given a *prior* estimate x_0^b (the *background*) and observations y_i, $i = 0, \ldots, n$, subject to the nonlinear forecast model given by

$$x_i = \mathcal{M}(t_i, t_{i-1}, x_{i-1}), \qquad (2)$$

$$y_i = \mathcal{H}_i(x_i) + \delta_i, \qquad (3)$$

for $i = 0, \ldots, n$. Here $\mathcal{M}$ and $\mathcal{H}_i$ denote the evolution and observation operators of the system. The errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ in the background and the errors δ_i in the observations are assumed to be random with mean zero and covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$, respectively. The assimilation problem is then to minimize, with respect to $\mathbf{x}_0$, the objective function

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=0}^{n} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i), \tag{4}$$

subject to the model forecast equations (2)–(3). If observations are given at several points $t_i, \ i = 0, 1, \ldots, n$ over a time window $[t_0, t_n]$ with $n > 0$, the assimilation scheme is known as the four-dimensional variational method (4DVar). If observations are given only at the initial time with $n = 0$, then the optimization problem reduces to the three-dimensional data assimilation problem (3DVar).

2.1 Incremental variational assimilation

In operational NWP centres, to reduce computational cost, a sequence of linear approximations to the nonlinear least-squares problem (4) is solved. Given the current estimate of the analysis $\mathbf{x}_0$, the nonlinear objective function is linearized about the corresponding model trajectory $\mathbf{x}_i, \ i = 1, \ldots, n$, satisfying the nonlinear forecast model. An increment $\delta\mathbf{x}_0$ to the current estimate of the analysis is then calculated by minimizing the linearized objective function subject to the linearized model equations. The linear minimization problem is solved in an inner loop by a gradient iteration method. The current estimate of the analysis is then updated with the computed increment and the process is repeated in the outer loop of the algorithm. This data assimilation scheme is known as *incremental variational assimilation* [5], [14].

2.2 Condition number

A measure of the accuracy and efficiency with which the data assimilation problem can be solved is given by the *condition number* of the Hessian matrix

$$A = (B^{-1} + \hat{H}^T \hat{R}^{-1} \hat{H}) \tag{8}$$

of the previous row

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

The eigenvalues of such a matrix are the discrete Fourier transform of the coefficients of the first row of the matrix [9] and are given by

$$\lambda_m = \sum_{k=0}^{N-1} c_k e^{-2\pi i m k/N}. \tag{10}$$

Similarly the corresponding eigenvectors are given by the discrete exponential function,

$$\mathbf{v}_m = \frac{1}{\sqrt{N}}(1, e^{-2\pi i m/N}, \dots, e^{-2\pi i m(N-1)/N})^T. \tag{11}$$

3.1 Conditioning of the Gaussian background error covariance matrix

We first consider the Gaussian correlation matrix $\mathbf{C}$ ([6], [11]) with entries given by

$$c_{i,j} = \rho^{|i-j|^2} \tag{12}$$

for $|i-j| < N/2$, where $\rho = \exp\left(\frac{-\Delta x^2}{2L^2}\right)$, and by periodicity for the remaining entries. The coefficient $c_{i,j}$ denotes the correlation between background errors at positions i and j, L is the correlation length-scale and determines the strength of the spatial error correlations, Δx is the grid spacing and N is the number of grid points. A large length-scale means that the errors are strongly correlated over the whole grid. The maximum eigenvalue of this correlation matrix is

$$\lambda_{\max}(\mathbf{C}) = \sigma_b^2 \sum_{k=0}^{N-1} \rho^{k^2}, \tag{13}$$

with corresponding eigenvector $\mathbf{v}_{\max} = \frac{1}{\sqrt{N}}(1, \dots, 1)^T$. Similarly the minimum eigenvector is

$$\lambda_{\min}(\mathbf{C}) = \sigma_b^2 \sum_{k=0}^{N-1} (-1)^k \rho^{k^2}, \tag{14}$$

with corresponding eigenvector $\mathbf{v}_{\min} = \frac{1}{\sqrt{N}}(1, -1, 1, \dots, -1)^T$.

The condition number is given by the ratio of the maximum to minimum eigenvalues and is highly sensitive to changes in length-scale, as shown in Figure 1 for a grid spacing of $\Delta x = 0.1$ and $N = 500$ grid points. The matrix becomes very ill-conditioned as the length-scale increases,

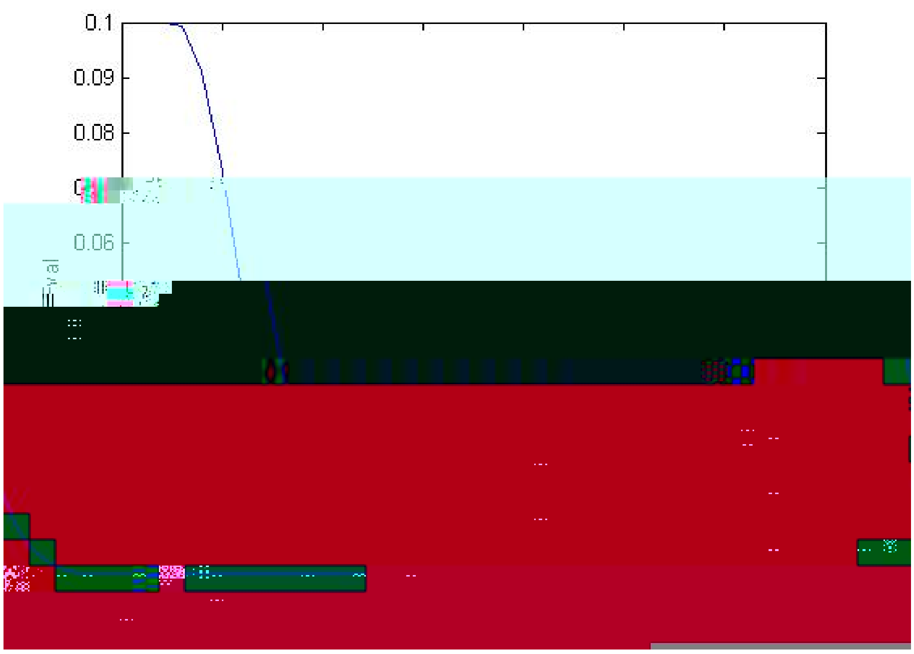

Figure 3: Change in smallest eigenvalue of the periodic Gaussian background error covariance matrix B with length-scale.

where the matrix $\mathbf{L}$ is given by

$$\mathbf{L} = \begin{pmatrix} -2 & 1 & 0 & 0 & \dots & 0 & 1 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ & \ddots & \ddots & & & & \\ \vdots & & \ddots & \ddots & & & 0 \\ 0 & & & \ddots & \ddots & & 1 \\ 1 & 0 & \dots & & & 1 & -2 \end{pmatrix};$$

and we define γ so that the maximum value of an element of $\mathbf{C}$ is unity. Using circulant theory we find the maximum eigenvalue to be

$$\lambda_{\max}(\mathbf{C}^{-1}) = \frac{1}{\gamma}\left(1 + 16\left(\frac{L^4}{2\Delta x^4}\right)\right); \tag{16}$$

with corresponding eigenvector $\mathbf{w}_{\max} = \frac{1}{\sqrt{N}}(1, -1, 1, \dots, -1)^T$. The smallest eigenvalue is

$$\lambda_{\min}(\mathbf{C}^{-1}) = \frac{1}{\gamma}; \tag{17}$$

with corresponding eigenvector $\mathbf{w}_{\min} = \frac{1}{\sqrt{N}}(1, 1, 1, \dots, 1)^T$. The conditioning of the Laplacian correlation matrix is therefore

$$\kappa(\mathbf{C}) = \left(1 + 16\frac{L^4}{2\Delta x^4}\right). \tag{18}$$

The conditioning grows in proportion to L^4 and hence is also quite poorly conditioned. However, as Figure 4 shows, the condition number is many orders of magnitude smaller than that of the Gaussian error covariance matrix at all length-scales.

4 Conditioning of the Hessian

In this section we consider the conditioning of the Hessian of the 3DVar linearized least-squares problem

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \tag{19}$$

in the case of a single periodic system parameter with background error variance σ_b^2. We examine the Hessian for each of the two background error correlation matrices defined in Section 3. We write the observational error covariance matrix in the form $\mathbf{R} = \sigma_o^2\mathbf{I}_p$, where p is the number of observations. We assume that the observations are direct measurements of the state variables. Then $\mathbf{H}^T\mathbf{H}$ is a diagonal matrix, where the k^{th} diagonal element is unity if the k^{th} state variable is observed and is zero otherwise. Under these conditions we can prove the following bounds on the condition number of the Hessian matrix for the 3DVar problem

$$\left(\frac{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\min}(\mathbf{C})}{1 + \frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\max}(\mathbf{C})}\right)\kappa(\mathbf{C}) \leq \kappa(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \leq \left(1 + \left(\frac{\sigma_b^2}{\sigma_o^2}\right)\lambda_{\min}(\mathbf{C})\right)\kappa(\mathbf{C}), \tag{20}$$

where $\lambda_{\max}(\mathbf{C})$ and $\lambda_{\min}(\mathbf{C})$ are the largest and smallest eigenvalues of $\mathbf{C}$ respectively. A proof of this result is given in Appendix A.

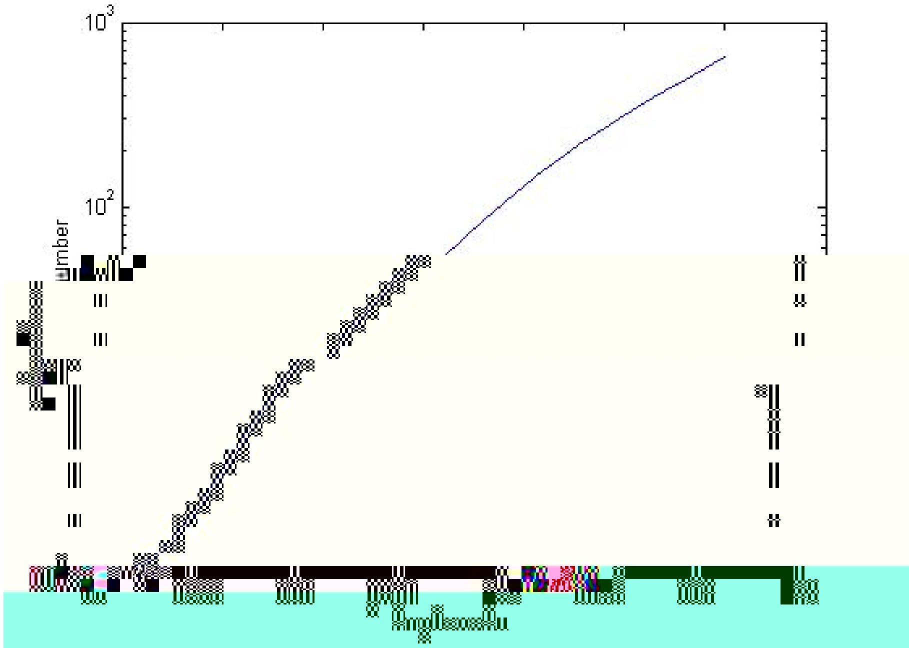

Figure 4: Condition number of Laplacian matrix B against lengthscale.

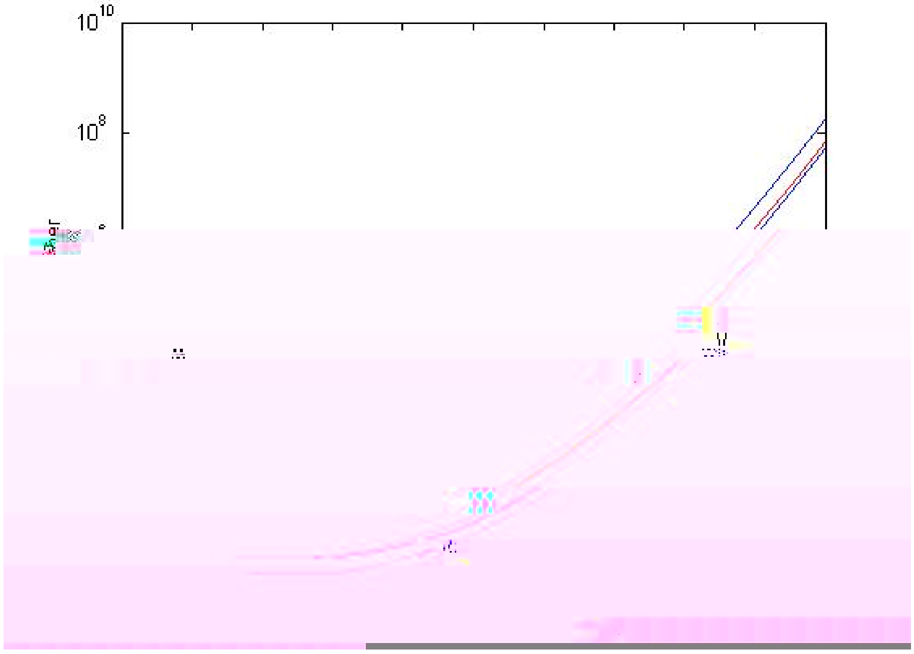

Figure 5: Condition number of the Hessian (red) and bounds (blue) against length-scale for Gaussian error covariance matrix **B**.

We see that with σ_b fixed, as σ_o increases and the observations become less accurate, the upper bound on the condition number of the Hessian decreases and both the upper and lower bounds converge to $\kappa(\mathbf{C}) = \kappa(\mathbf{B})$. As σ_o decreases, the lower bound goes to unity and, unless σ_o is much smaller than $\lambda_{min}(\mathbf{C})$, the upper bound remains of order $\kappa(\mathbf{C})$. We expect, therefore, that the conditioning of the Hessian will be dominated by the condition number of $\mathbf{C}$ as the correlation length-scales change in the background errors. We demonstrate this in Figure 5 for the Gaussian background covariance matrix with $\sigma_o^2 = \sigma_b^2 = 0.1$, $N = 500$ grid points and $p = 250$ observations. Similarly Figure 6 shows the conditioning of the Hessian for the same configuration but using the Laplacian background matrix. (Since the conditioning of the Laplacian is better than that of the Gaussian, a wider range of length-scales is shown in Figure 6.) In these cases including observations has little effect on the conditioning of the assimilation problem.

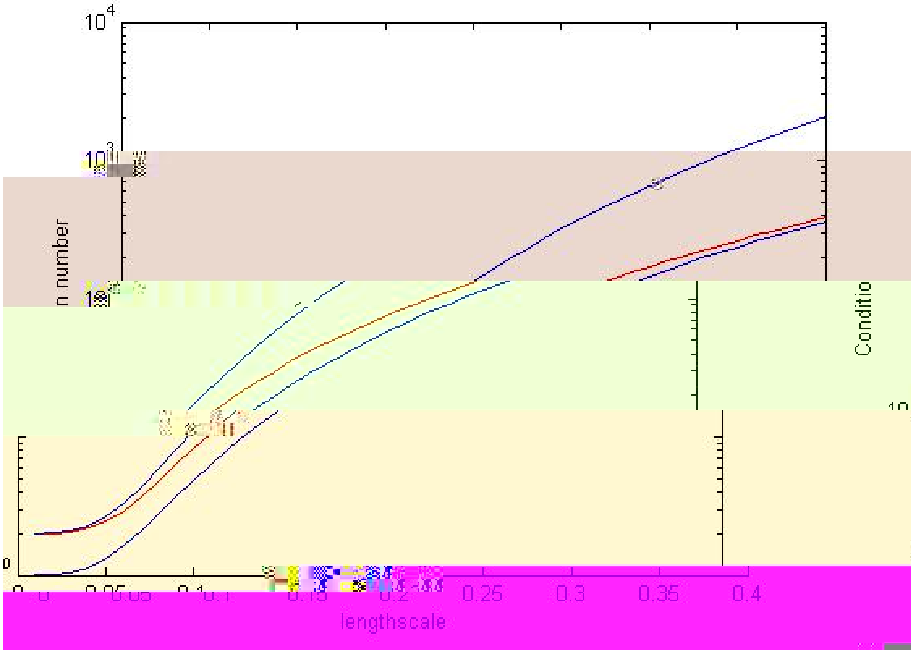

Figure 6: Condition number of the Hessian (red) and bounds (blue) against length-scale for Laplacian error covariance matrix B.

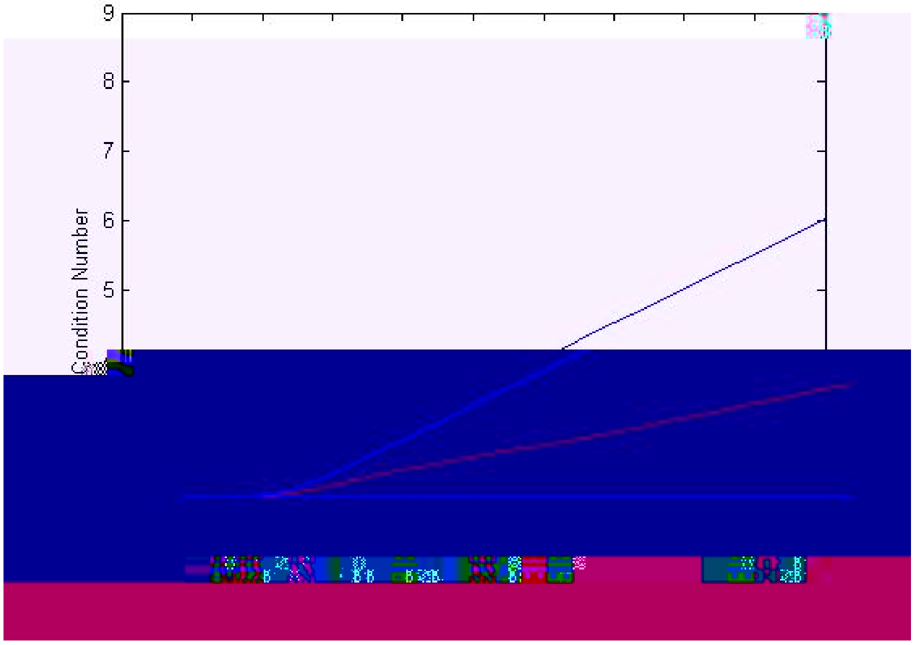

Figure 7: Condition number (red) and bounds (blue) against length-scale for the preconditioned Hessian with the Gaussian background error covariance matrix.

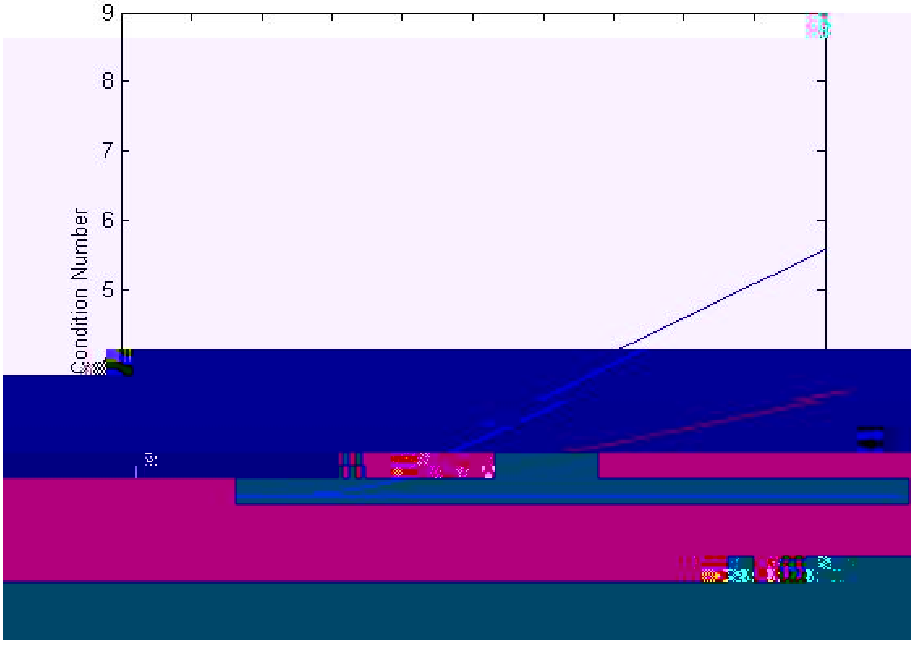

Figure 8: Condition number (red) and bounds (blue)against length-scale for the preconditioned Hessian with the Laplacian background error covariance matrix.

(23) depend on sums of the elements of the matrix $\mathbf{HCH}^T$; which can be viewed as a `reduced'

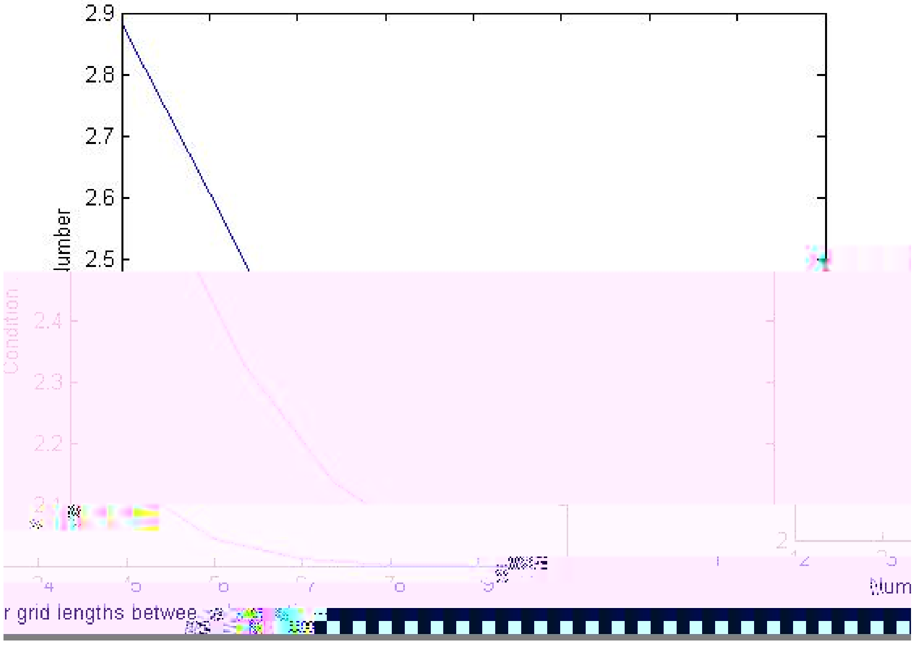

Figure 9: Condition number of the preconditioned Hessian for two observations as the grid-point separation is increased. The background error covariance matrix is Gaussian with a length-scale of 0.2.

found in [1] for a two grid-point observing system, which shows that the conditioning of the preconditioned system reduces as the accuracy of the observations is decreased.

Experiments in the Met Office operational variational assimilation system support the theoretical results presented here and confirm that they hold in a more general case. These results will be published in a forthcoming report.

There are two natural extensions to the work presented here. The first is to extend the results for the preconditioned system to encompass more general systems. One approach analogous to our treatment here is to use the dual form of the Hessian [3]

$$I_p + R^{-1/2} H B H^T R^{-1/2}. \qquad (26)$$

The largest eigenvalue of this matrix is the condition number of the preconditioned Hessian and a general upper bound can be found, as in Appendix B, to be

$$\kappa(I_N + B^{1/2} H^T R^{-1} H B^{1/2}) \leq 1 + ||R^{-1/2} H B H^T R^{-1}$$

An improvement in the bounds can be achieved using the *Rayleigh quotient*, $R_{\mathbf{A}}(\mathbf{v})$, which, for a Hermitian matrix $\mathbf{A}$ and non-zero vector $\mathbf{v}$, is defined to be

$$R_{\mathbf{A}}(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}}. \tag{33}$$

The maximum and minimum eigenvalues of $\mathbf{A}$ are the maximum and minimum values of $R_{\mathbf{A}}(\mathbf{v})$ respectively, where $\mathbf{v}$ is the corresponding eigenvector. We first consider the eigenvector corresponding to the largest eigenvalue of $\mathbf{B}^{-1}$. For the the Gaussian covariance matrix this is $\mathbf{v}_{\min}$ and for the Laplacian covariance this is $\mathbf{w}_{\max}$ (see Section 3). In both cases the Rayleigh quotient with respect to $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} = \sigma_o^{-2} \mathbf{H}^T \mathbf{H}$ is simply $\sigma_o^{-2} p/N$, where p is the number of observations and N is the number of grid points. Then an improved lower bound on the maximum eigenvalue of the Hessian is given by

$$\lambda_{\max}(\mathbf{A}) = \max_{\mathbf{v} \in \mathbb{R}^n} \left(\frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \right) \geq \mathbf{z}_{\max}^T \mathbf{A} \mathbf{z}_{\max} = \lambda_{\max}(\mathbf{B}^{-1}) + \sigma_o^{-2} \frac{p}{N}; \tag{34}$$

where $\mathbf{z}_{\max}$ is either $\mathbf{v}_{\min}$ or $\mathbf{w}_{\max}$. Similarly we consider $\mathbf{v}_{\max}$ and $\mathbf{w}_{\min}$;-1 63TD[(min)]TJ/F710.

A lower bound can be achieved by considering the Rayleigh quotient on (38). We de¯ne a unit vector $\mathbf{v} \in \mathrm{R}^p$

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations, third edition,* Johns Hopkins University Press, 1996.

[9] R.M. Gray, *Toeplitz and Circulant matrices: A Review*, Foundations and Trends® on Communications and Information Theory, **2**, no.3, 2006, pp.155{239.

[10] C.Johnson, B.J. Hoskins, N.K. Nichols, *A singular vector perspective of 4D-Var:Filtering and interpolation*, Q. J. R. Met Soc., **131**, 2005, 1-19.

[11] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability,* Cambridge University Press, 2006.

[12] A.B.Kostinski and A.C. Koivunen, *On the Condition Number of Gaussian Sample-Covariance Matrices*, IEEE Transactions on Geoscience and Remote Sensing, **38**, 2000, 329-332.

[13] A.C. Koivunen and A.B.Kostinski, *The Feasibility of Data Whitening to Improve Performance of Weather Radar*, J. Applied Meteorology, **38**, 1999, 741-749.

[14] A.S. Lawless, S. Gratton and N.K. Nichols, *An investigation of incremental 4D-Var using non-tangent linear models*, Q. J. R. Met. Soc., **131**, 2005, 459-476.

[15] A. Lorenc, *Optimal nonlinear objective analysis*, Q. J. R. Met. Soc., **114**, 1988, 205{240.

[16] A. Lorenc, *Development of an Operational Variational Assimilation Scheme*, J. Met. Soc. of Japan, **75**, 1997, 339{346.

[17] A.C. Lorenc, S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne and F. W. Saunders, *The Met. O±ce global three-dimensional variational data assimilation scheme*, Q. J. R. Met. Soc., **126**, 2000, 2991{3012.

[18] F. Rabier, H. Jarvinen, E. Klinker, J. F. Mahfouf and A. Simmons, *The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simpli¯ed physics*, Q. J. R. Met Soc., **126**, 2000, 1143{1170.

[19] F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc and T.J. Payne, *The Met O±ce global four-dimensional variational data assimilation scheme*, Q. J. R. Met Soc., **133**, 2007, 347{362.

[20] Y. Tremolet, *Incremental 4D-Var convergence study*, Tellus A, **59**, 2007, 706{718.