Computers & Fluids 46 (2011) 252–256

Contents lists available at ScienceDirect

Computers & Fluids

journal homepage: www.elsevier.com/locate/compfluid

Conditioning and preconditioning of the variational data assimilation problem

S.A. Haben [1], A.S. Lawless [2], N.K. Nichols [*,2]

Department of Mathematics, University of Reading, PO Box 220, Reading RG6 6AX, UK

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 22 August 2010
Accepted 24 November 2010
Available online 30 November 2010

Keywords:
Variational data assimilation
Error covariances
Least-squares optimization
Condition number
Hessian

ABSTRACT

Numerical weather prediction (NWP) centres use numerical models of the atmospheric flow to forecast future weather states from an estimate of the current state. Variational data assimilation (VAR) is used commonly to determine an optimal state estimate that miminizes the errors between observations of the dynamical system and model predictions of the flow. The rate of convergence of the VAR scheme and the sensitivity of the solution to errors in the data are dependent on the condition number of the Hessian of the variational least-squares objective function. The traditional formulation of VAR is ill-conditioned and hence leads to slow convergence and an inaccurate solution. In practice, operational NWP centres precondition the system via a control variable transform to reduce the condition number of the Hessian. In this paper we investigate the conditioning of VAR for a single, periodic, spatially-distributed state variable. We present theoretical bounds on the condition number of the original and preconditioned Hessians and hence demonstrate the improvement produced by the preconditioning. We also investigate theoretically the effect of observation position and error variance on the preconditioned system and show that the problem becomes more ill-conditioned with increasingly dense and accurate observations. Finally, we confirm the theoretical results in an operational setting by giving experimental results from the Met Office variational system.

1. Introduction

Variational data assimilation (VAR) 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 [7].

In practice an incremental version of VAR is implemented in many operational centres, including the Met Office [11] and the European Centre for Medium-Range Weather Forecasting (ECMWF) [10]. This method solves a sequence of linear least-squares approximations to the nonlinear least-squares problem and is equivalent to an approximate Gauss–Newton method for determining the analysis [6]. Each approximate linear 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.

The rate of convergence of the inner loop of the VAR iteration scheme and the sensitivity of the analysis to perturbations in the data of the problem are proportional to the condition number, that is, the ratio of the largest to the smallest eigenvalue, of the Hessian of the linear least-squares objective function [3]. Experimental results indicate that in operational systems the Hessian is ill-conditioned, with undesirable features [8]. Operationally the system is preconditioned by transforming the state variables to new variables where the errors are assumed to be approximately uncorrelated. Preconditioning reduces the sensitivity of the problem to be solved and hence enables a more accurate analysis to be computed [3]. Experimental comparisons have demonstrated that operationally the preconditioning significantly improves the speed and accuracy of the assimilation scheme [2,8].

A variety of explanations are offered in the literature for the ill-conditioning of the VAR problem and for the benefits of preconditioning in the operational system [9,1,12]. In this paper we examine the conditioning and preconditioning of the variational assimilation method theoretically. We give expressions for bounds on the conditioning of the Hessian of the problem in the case of a single, periodic, spatially-distributed system parameter. It is

* Corresponding author. Tel.: +44 0 118 378 8988; fax: +44 0 118 931 3423.
E-mail address: n.k.nichols@reading.ac.uk (N.K. Nichols).

[1] The work of this author was supported by an EPSRC CASE Award with the Met Office. The author is also grateful for support from the European Science Foundation OPTPDE Network on Optimization with PDE Constraints.

[2] The work of these authors was supported in part by the NERC National Centre for Earth Observation.

0045-7930/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compfluid.2010.11.025

assumed that the errors in the initial background states (model forecast) have a Gaussian correlation structure, although other forms of the error correlation structure can be analysed using the same theory. We consider three questions: (i) how does the condition number of the Hessian depend on the length-scale in the correlation structures? (ii) how does the variance of the observation errors affect the conditioning of the Hessian? and (iii) how does the distance between observations, or density of the observations, affect the conditioning of the Hessian?

In the next section we introduce the incremental variational assimilation method. In Section 3 we give bounds on the conditioning of the problem and examine our three questions. In Section 4 we present experimental results obtained using the Met Office Unified Model supporting the theory and in Section 5 we summarize the conclusions. In this paper we present results only for the 3D-variational method, but our techniques can be extended to the 4D-variational scheme and will be discussed in a subsequent paper.

2. Variational data assimilation

The aim of the variational assimilation problem is to find an optimal estimate for the initial state of the system $\mathbf{x}_0$ (the *analysis*) at time t_0 given a *prior* estimate $\mathbf{x}_0^b$, (the *background*) and observations $\mathbf{y}_i$ at times t_i, subject to the nonlinear forecast model given by

$$\mathbf{x}_i = \mathcal{M}(t_i, t_{i-1}, \mathbf{x}_{i-1}), \tag{1}$$

$$\mathbf{y}_i = \mathcal{H}_i(\mathbf{x}_i) + \boldsymbol{\delta}_i, \tag{2}$$

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 $\boldsymbol{\delta}_i$ 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

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=0}^{n}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i), \tag{3}$$

subject to the model forecast Eqs. (1) and (2). 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).

In practice, to improve the computational efficiency of the variational assimilation procedure, a sequence of linear least-squares approximations to the nonlinear least-squares problem (3) 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 least-squares objective function subject to the linearized model equations. The linear least-squares 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.

In each outer loop the incremental scheme minimizes, with respect to $\delta\mathbf{x}_0$, the current linearized least-squares objective function, which may be written as

$$\begin{aligned}\tilde{\mathcal{J}}[\delta\mathbf{x}_0] = {} & \frac{1}{2}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)]^T \mathbf{B}^{-1}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)] \\ & + \frac{1}{2}(\widehat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}})^T \widehat{\mathbf{R}}^{-1}(\widehat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}}),\end{aligned} \tag{4}$$

subject to the linearized model equations

$$\delta\mathbf{x}_i = \mathbf{M}(t_i, t_{i-1})\delta\mathbf{x}_{i-1}, \tag{5}$$

where

$$\widehat{\mathbf{H}} = \left[\mathbf{H}_0^T, (\mathbf{H}_1\mathbf{M}(t_1, t_0))^T, \ldots, (\mathbf{H}_n\mathbf{M}(t_n, t_0))^T\right]^T,$$

$$\hat{\mathbf{d}}^T = \left[\mathbf{d}_0^T, \mathbf{d}_1^T, \ldots, \mathbf{d}_n^T\right], \quad \text{with} \quad \mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i).$$

The matrices $\mathbf{M}(t_i, t_0)$ and $\mathbf{H}_i$ are linearizations of the evolution and observation operators $\mathcal{M}(t_i, t_0, \mathbf{x}_0)$ and $\mathcal{H}_i(\mathbf{x}_i)$ about the current estimated state trajectory $\mathbf{x}_i$, $i = 0, \ldots, n$ and $\widehat{\mathbf{R}}$ is a block diagonal matrix with diagonal blocks equal to $\mathbf{R}_i$.

The minimizer of (4) is also the solution to $\nabla\tilde{\mathcal{J}} = 0$, which may be written explicitly as the linear system

$$(\mathbf{B}^{-1} + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\delta\mathbf{x}_0 = \mathbf{B}^{-1}(\mathbf{x}_0^b - \mathbf{x}_0) + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\hat{\mathbf{d}}. \tag{6}$$

Iterative gradient methods are used to solve (4), or equivalently (6). The gradients are found by an adjoint procedure.

3. Conditioning of the assimilation problem

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

$$\mathbf{A} = (\mathbf{B}^{-1} + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) \tag{7}$$

of the linearized objective function (4). Our aim here is to present explicit bounds on the condition number of $\mathbf{A}$ and investigate its properties in terms of the background and observation error covariance matrices $\mathbf{B}$ and $\widehat{\mathbf{R}}$.

The condition number of the Hessian, which is a square, symmetric, positive definite matrix, is defined in the L_2-norm by

$$\kappa(\mathbf{A}) = ||\mathbf{A}||_2||\mathbf{A}^{-1}||_2 \equiv \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \tag{8}$$

where $\lambda(\mathbf{A})$ denotes an eigenvalue of the matrix. The condition number measures the sensitivity of the solution to the linearized least-squares problem (4), or equivalently the solution to the gradient Eq. (6), to perturbations in the data of the problem. If the condition number of the Hessian, $\kappa(\mathbf{A})$, is very large, the problem is 'ill-conditioned' and, even for small perturbations to the system, the relative error in the solution may be extremely large. For the gradient methods that are commonly used to solve the problem, such as the conjugate gradient method, the rate of convergence then may also be very slow.

Here we consider specifically the conditioning of the 3DVar linearized least-squares problem. In this case observations are given at only one point in time and $\widehat{\mathbf{H}} = \mathbf{H} \equiv \mathbf{H}_0$. We consider in theory the case of a single periodic system parameter with background error variance σ_b^2 and uncorrelated observation errors with variance σ_o^2.

3.1. Conditioning of the background error covariance matrix

We write the background error covariance in the form $\mathbf{B} = \sigma_b^2\mathbf{C}$, where $\mathbf{C}$ denotes the correlation structure of the background errors. The condition number $\kappa(\mathbf{B})$ then equals the condition number $\kappa(\mathbf{C})$. We assume that the correlation structure is homogeneous, where the correlations depend only on distance between states and not position. Under these conditions the correlation matrices used commonly in practice have a circulant structure [4], which we exploit to obtain our theoretical bounds. For example, the Gaussian, Markov and SOAR correlation matrices have this structure, as do those based on Laplacian smoothing. A circulant matrix is a special case of a Toeplitz matrix and has the essential property that each row is a cyclic permutation of the previous row. The eigenvalues are given simply by the discrete Fourier transform of the first row of the matrix and the eigenvectors are given by the discrete exponential function.

For example, we consider a Gaussian correlation structure for a one-dimensional system parameter on a uniform grid of N points. The elements $c_{i,j}$ of the correlation matrix $\mathbf{C}$ are then given by

$$c_{i,j} = \exp\left(\frac{-\Delta x^2|i-j|^2}{2L^2}\right) \tag{9}$$

for $|i-j| < N/2$, and by periodicity for the remaining elements, where Δx is the grid step and L is the correlation length-scale. Explicit expressions for the conditioning of $\mathbf{C}$ can then be derived [5]. The condition numbers for increasing length-scale L are shown in Table 1 for a grid with step size $\Delta x = 0.1$ and $N = 500$. It can be seen that the correlation matrix becomes highly ill-conditioned as the length-scale increases. The reason is primarily due to a rapid reduction in its smallest eigenvalue with length-scale, rather than a rapid increase in the largest eigenvalue (see [5]).

3.2. Conditioning of the Hessian

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 kth diagonal element is unity if the kth state variable is observed and is zero otherwise. Under these conditions we can show the following bounds on the condition number of the Hessian matrix for the 3DVar problem

$$\kappa(\mathbf{C})\left(\frac{1+\frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\min}(\mathbf{C})}{1+\frac{p}{N}\frac{\sigma_b^2}{\sigma_o^2}\lambda_{\max}(\mathbf{C})}\right) \leqslant \kappa(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \leqslant \kappa(\mathbf{C})\left(1+\left(\frac{\sigma_b^2}{\sigma_o^2}\right)\lambda_{\min}(\mathbf{C})\right), \tag{10}$$

where $\lambda_{\max}(\mathbf{C})$ and $\lambda_{\min}(\mathbf{C})$ are the largest and smallest eigenvalues of $\mathbf{C}$ respectively. The proof is given in [5] and relies on basic eigenvalue inequalities for symmetric matrices and on the use of the Rayleigh quotient [3]. In the case where the correlation matrix is not circulant, the same upper bound can be established, together with a weaker lower bound [5].

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 $\kappa(\mathbf{B})$. We expect, therefore, that the conditioning of the Hessian will be dominated by the condition number of $\mathbf{B}$ as the correlation length-scales change in the background errors. We demonstrate this in Fig. 1a for the Gaussian background covariance matrix with $\sigma_o^2 = \sigma_b^2 = 0.1$, $N = 500$ grid points and $p = 250$ observations. In this case the observations have little effect on the conditioning of the assimilation problem.

3.3. Preconditioned variational data assimilation

A well-known technique for improving the convergence of an iterative method for solving a linear least-squares problem is to apply a linear transformation to 'precondition' the system and thus reduce the condition number of the Hessian [3]. The strategy used in many forecasting centres is to precondition the Hessian using the symmetric square root of the background error covariance matrix $\mathbf{B}^{1/2}$. The preconditioning is implemented using a control variable transform to new variables $\delta\mathbf{z} = \mathbf{B}^{-1/2}\delta\mathbf{x}_0$, which are thus uncorrelated. In terms of the new control variable the problem is to minimize

$$\hat{\mathcal{J}}[\delta\mathbf{z}] = \frac{1}{2}[\delta\mathbf{z}-(\mathbf{z}_0^b-\mathbf{z}_0)]^T[\delta\mathbf{z}-(\mathbf{z}_0^b-\mathbf{z}_0)] + \frac{1}{2}(\mathbf{H}\mathbf{B}^{1/2}\delta\mathbf{z}-\mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}^{1/2}\delta\mathbf{z}-\mathbf{d}), \tag{11}$$

where $\mathbf{z}_0^b = \mathbf{B}^{-1/2}\mathbf{x}_0^b$ and $\mathbf{z}_0 = \mathbf{B}^{-1/2}\mathbf{x}_0$. The Hessian of the preconditioned objective function is now given by

$$\mathbf{I}+\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}. \tag{12}$$

In general there are fewer observations than states of the system and therefore the matrix $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ is not of full rank, but is positive semi-definite. It follows that the smallest eigenvalue of (12) is unity and the condition number of the preconditioned Hessian is equal to its largest eigenvalue. We can then establish [5] that the condition number satisfies

$$1+\frac{\sigma_b^2}{\sigma_o^2}\gamma \leqslant \kappa(\mathbf{I}_n+\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}) \equiv \lambda_{\max}(\mathbf{I}_p+\frac{\sigma_b^2}{\sigma_o^2}\mathbf{H}\mathbf{C}\mathbf{H}^T) \leqslant 1+\frac{\sigma_b^2}{\sigma_o^2}\|\mathbf{H}\mathbf{C}\mathbf{H}^T\|_\infty, \tag{13}$$

where $\gamma = \frac{1}{p}\sum_{i,j\in J}c_{i,j}$, and J is the set of indices of the variables that are observed. The proof does not require the correlation matrix to have a circulant structure.

We see that the upper bound on the condition number is significantly reduced by preconditioning. For the Gaussian background covariance matrix we demonstrate this in Fig. 1b for the case with the same data as in Fig. 1a. The condition number of the preconditioned problem is shown to be of order unity and to increase roughly linearly with length-scale. In comparison to the case without preconditioning, there is a dramatic reduction in the condition number from order 10^7 to size $\cong 3.5$ at length-scale $L = 0.2$. Similar, although less dramatic results, are obtained with other choices of the background covariance structures (see [5]).

Furthermore, from (13) we see that with σ_b fixed, as σ_o decreases and the observations become more accurate, then both the lower and upper bounds on the condition number of the preconditioned Hessian increase and the problem becomes more ill-conditioned.

3.4. Spacing of observations

We now examine the condition number of the preconditioned Hessian as a function of the density or separation of the observations. From the definition, the coefficients $c_{i,j}$ of the correlation matrix $\mathbf{C}$ are expected to decrease as the distance $|i-j|$ increases. The upper bound given by (13) on the condition number of the

Table 1
Variation of the condition number of the background error covariance matrix with length-scale.

Length-scale	0.05	0.1	0.15	0.2	0.25	0.3
Condition number	1.74	69.5	3.32×10^4	1.87×10^8	1.24×10^{13}	7.45×10^{17}

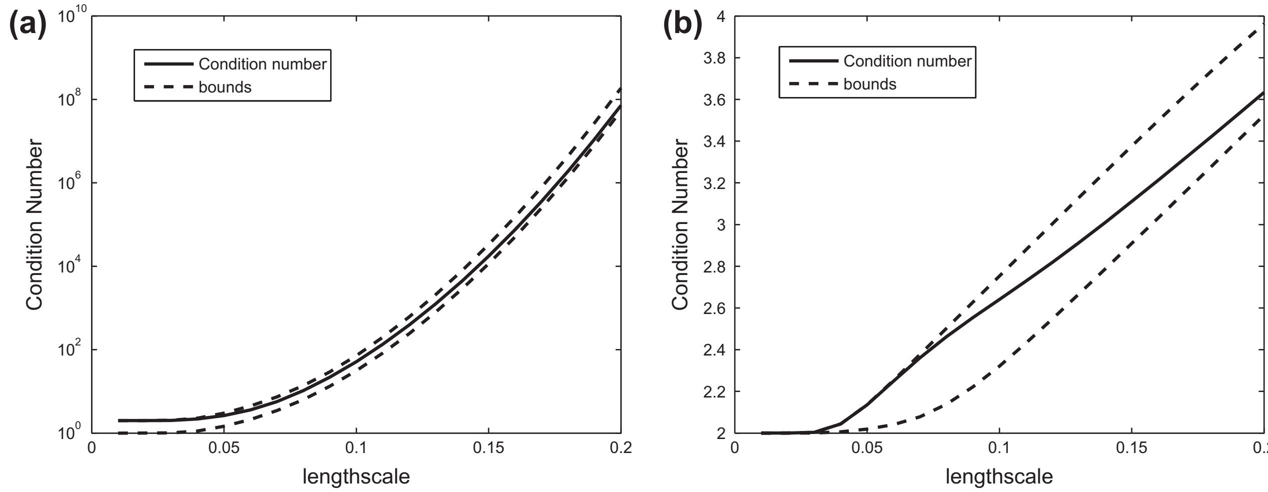

Fig. 1. Conditioning of (a) the Hessian matrix and (b) the preconditioned Hessian matrix against background correlation length-scale in the Gaussian case. Solid line — denotes condition number and dashed line - - - denotes the upper and lower bounds.

preconditioned Hessian takes the form $1 + \max_{i \in J} \sum_{j \in J} |c_{i,j}|$, where J is the set of indices of the variables that are observed. We therefore expect the conditioning of the problem to decrease as the separation of the observations increases or as the density decreases.

In the case of only two observations at positions k and m, for example, we find that the condition number of the preconditioned Hessian is exactly equal to

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

and the conditioning changes in proportion to the background error correlations. In the Gaussian case, as the points get further apart the condition number decays exponentially, as shown in Fig. 2.

3.5. Discussion

For the preconditioned variational assimilation problem, we conclude that as the accuracy and number/density of the observations increases, the computed solution to the problem is expected to become less accurate. This result seems counterintuitive, but can be understood by recognizing that the problem is then more highly constrained and hence more difficult to solve. This theory provides a generalization of the results of Andersson et al. [1], which show that for a 2 grid-point system with multiple observations, the condition number of the preconditioned system is proportional to the number of observations and to the inverse of the observation variance. The theory presented here includes this result as a special case and demonstrates that the same conclusions apply in more general cases. The theory given here can also be extended directly to the 4D-variational assimilation problem, although the bounds on the condition number then depend on the model evolution of the background correlation matrix.on the model evolution of the background correlation matrix.

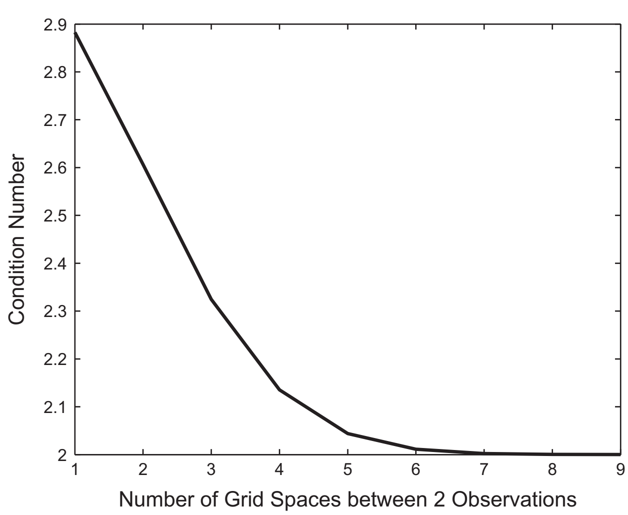

Fig. 2. Conditioning of the preconditioned Hessian for two observations as the grid-point separation is increased.

4. Numerical experiments with the Met Office Unified Model

In order to demonstrate the theoretical results in an operational system we show results obtained using the Unified Model of the Met Office. The Met Office incremental 3DVar system [9] is run for one outer loop at a spatial resolution of N108 (approximately 1.46×1.11 degrees) for both the inner and outer loops. The operational configuration of the Met Office data assimilation system already includes a preconditioning by the square root of the background error covariance matrix and so the objective function is of the form given by (11), with Hessian of the form (12). Thus, as explained in Section 3.3, the smallest eigenvalue is unity and the condition number is equal to the largest eigenvalue. This eigenvalue is calculated during the minimization process by a Lanczos algorithm.

We first investigate the effect of the observation error variance on the conditioning of the 3DVar assimilation problem. We define a set of 16 pseudo-observations of pressure at the lowest model level, arranged in a 4×4 square over the UK with a constant spacing between observations in the latitudinal and longitudinal directions. Since the objective function is specified in the incremental formulation (11), the pseudo-observations are defined by the innovations $\mathbf{d}$ rather than the actual observations. The value of these observed innovations are taken to be 1 Pa. The 3DVar assimilation is then run using these observations with different values for the observation error variance. The condition number of the Hessian for these different experiments is shown in Table 2. We see that as the observations become more accurate the condition number of the problem increases. This supports the theory of Sections 3.2 and 3.3, which shows that the bound on the condition number is inversely proportional to the observation error variance.

Table 2
Variation of condition number with observation error variance.

Observation error variance	0.01	0.25	0.5	1	2
Condition number	15,153,612	606,145	303,062	151,537	75,771

Next we investigate the effect of observation density on the conditioning. We again define a set of 16 pseudo-observations of pressure on model grid points as in the previous experiment. With a spacing of one grid step between observations, the condition number of the Hessian is found to be 151,537. When the spacing between observations is increased to two grid steps then the condition number falls to 115,355. A further spread of the observations, to a 15-grid-point spacing, results in an even lower condition number of 24,434. Thus the pseudo-observation experiment confirms the theory of Section 3.4 that the condition number of the Hessian decreases as the observations are spread further apart.

Finally, in order to investigate further the effect of observation density, we determine the conditioning of the 3DVar assimilation system with the operational observations. We use an example test case from 12 UTC on 27th October 2008. The condition number of the Hessian, including all observations, is calculated to be 3658. An examination of the leading eigenvector of the Hessian shows it to be concentrated over Europe. Previous studies have suggested that this is associated with the dense observation network of surface observations over Europe [1,12]. To test this hypothesis for the Met Office system we repeat our experiment, but first assimilating only surface observations, and then assimilating all except the surface observations. With only the surface observations the condition number is still high, with a value of 3345. When surface observations are omitted from the assimilation then the condition number reduces to 1431. As for other systems, it thus appears that the conditioning of the Met Office assimilation system is dominated by the high density surface observations.

5. Conclusions

We have examined the conditioning of the incremental variational data assimilation problem in the case of a single periodic system parameter. We have given expressions for the conditioning of a class of background error correlation structures and demonstrated that in the Gaussian case, the correlation matrix becomes extremely ill-conditioned as the length-scale increases. We also presented bounds on the condition number of the Hessian in the 3DVar case and showed that this is dominated by the background error correlations as the length-scales increase. The theory shows also that for a fixed background variance, as the variance of the observation errors decreases, the conditioning of the Hessian increases. We presented theoretical bounds on the preconditioned Hessian (preconditioned by the symmetric square root of the background covariance matrix) and demonstrated that the preconditioning provides a dramatic reduction in the condition number of the problem. For the preconditioned system we also showed theoretically that the conditioning of the problem improves as the separation between observations is increased and the density reduced. Finally, we have described results of experiments using the Unified Model of the Met Office that confirm in practice the theory developed here.

Acknowledgement

We are grateful to Dr. T. Payne from the Met Office for his contribution to this research.

References

[1] Andersson E, Fisher M, Munro R, McNally A. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. Quart J Roy Met Soc 2000;126:1455–72.
[2] Gauthier P, Charette C, Fillion L, Koclas P, Laroche S. Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part 1: the global analysis. Atmos Ocean 1999;37:103–56.
[3] Golub GH, Van Loan CF. Matrix Computations. third ed. Johns Hopkins University Press; 1996.
[4] Gray RM. Toeplitz and circulant matrices: a review. Found Trends® Commun Inform Theory 2006;2(3):155–239.
[5] Haben SA, Lawless AS, Nichols NK. Conditioning of the 3DVar data assimilation problem. University of Reading, Dept. of Mathematics, Math Report Series 3/2009; 2009. <http://www.reading.ac.uk/maths-and-stats/research/maths-report-series.aspx>.
[6] Lawless AS, Gratton S, Nichols NK. An investigation of incremental 4D-Var using non-tangent linear models. Quart J Roy Met Soc 2005;131:459–76.
[7] Lorenc A. Optimal nonlinear objective analysis. Quart J Roy Met Soc 1988;114:205–40.
[8] Lorenc AC. Development of an operational variational assimilation scheme. J Met Soc Jpn 1997;75:339–46.
[9] Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, et al. The Met Office global three-dimensional variational data assimilation scheme. Quart J Roy Met Soc 2000;126:2991–3012.
[10] Rabier F, Jarvinen H, Klinker E, Mahfouf JF, Simmons A. The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. Quart J Roy Met Soc 2000;126:1143–70.
[11] Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, et al. The Met Office global four-dimensional variational data assimilation scheme. Quart J Roy Met Soc 2007;133:347–62.
[12] Tremolet Y. Incremental 4D-Var convergence study. Tellus A 2007;59:706–18.