School of Mathematics, Meteorology
and Physics

Department of Mathematics

Preprint MPS_2010_16

21 April 2010

Conditioning and Preconditioning of the Variational Data Assimilation Problem

by

S.A. Haben, A.S. Lawless, N.K. Nichols

Conditioning and Preconditioning of the Variational Data Assimilation Problem

S.A. Haben*, A.S. Lawless, N.K. Nichols

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

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 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 establish the

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 O ce [11] and the European Centre for Medium-Range Weather Forecasting (ECMWF) [10]. This method solves a sequence of linear approximations to the nonlinear least-squares problem of the error correlation structure can be analysed using the same theory. We consider three questions: how does the condition number of the Hessian depend on the length-scale in the correlation structures; how does the variance of the observation errors a ect the conditioning of the Hessian; and how does the distance between observations, or density of the observations, a ect the conditioning of the Hessian.

In the next section we introduce the incremental variational assimilation method. In Section 3 we derive bounds on the conditioning of the problem and examine our three questions. In Section 4 we present experimental results obtained using the Met O ce 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 x_0 (the *analysis*) at time t_0 given a *prior* estimate x_0 (the *background*) and observations y at times t, subject to the nonlinear forecast model given by

$$\mathrm{x} = \mathcal{M}(t, t_{-1}, \mathrm{x}_{-1}), \tag{1}$$

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

for $i = 0, \ldots, n$. Here $\mathcal{M}$ and $\mathcal{H}$ denote the evolution and observation operators of the system. The errors $(\mathrm{x}_0 - \mathrm{x}_0)$ in the background and the errors $\boldsymbol{\delta}$ in the observations are assumed to be random with mean zero and covariance matrices $\mathbf{B}$ and $\mathbf{R}$, In practice, to improve the computational e ciency of the variational assimilation procedure, a sequence of linear approximations to the nonlinear

3. Conditioning of the Assimilation Problem

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

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

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

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

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

where $\lambda(\mathrm{A})$ 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 equation (6), to perturbations in the data of the problem. If the condition number of the Hessian, $\kappa(\mathrm{A})$, 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 the conditioning of the 3DVar linearized least-squares problem in the case of a single periodic system parameter with background error variance σ^2 and uncorrelated observation errors with variance σ^2.

3.1. Conditioning of the background error covariance matrix

We write the background error covariance in the form $\mathrm{B} = \sigma^2 \mathrm{C}$, where C denotes the correlation structure of the background errors. The condition number $\kappa(\mathrm{B})$ then equals the condition number $\kappa(\mathrm{C})$. 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 $c_{,}$ of the correlation matrix $\mathbf{C}$ are then given by

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

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, primarily due to a rapid reduction in its smallest eigenvalue.

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}

3.2. Conditioning of the Hessian

We write the observational error covariance matrix in the form $\mathbf{R} = \sigma^2 \mathbf{I}$, 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 diagonal element is unity if the k state variable is observed and is zero otherwise. Under these conditions we can prove [5] the following bounds on the condition number of the Hessian matrix for the

3DVar problem

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

where $\lambda_{\max}(\mathbf{C})$ and $\lambda_{\min}(\mathbf{C})$ are the largest and smallest eigenvalues of $\mathbf{C}$ respectively.

We see that with σ fixed, as σ 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 decreases, the lower bound goes to unity and, unless σ is much smaller than λ $(\mathbf{C})$, 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 Figure 1a for the Gaussian background covariance matrix with $\sigma^2 = \sigma^2 = 0.1$, $N = 500$ grid points and $p = 250$ observations. In this case the observations have little e ect 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 symmetrically using the 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

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

where $\mathbf{z}_0 = \mathbf{B}^{-1/2}\mathbf{x}_0$ 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.

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 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

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

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 Figure 2.

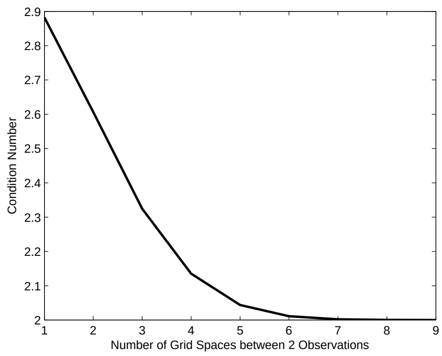

Figure 2: Conditioning of the preconditioned Hessian for two observations as the grid-point separation is increased.

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 e ect 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 di erent values for the observation error variance. The condition number of the Hessian for these di erent 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{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 e ect 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 e ect of observation density,

Observation and by an EPSRC CASE Award with the Met O ce. We are grateful to Dr T. Payne from the Met O ce for his contribution to this research.

References

[1] Andersson, E., M. Fisher, R. Munro and A. McNally, 2000: 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.*, **126**, 1455{1472.

[2] Gauthier, P., C.Charette, L. Fillion, P. Koclas and S. Laroche, 1999: Implementation of a 3D Variational Data Assimilation System at the Canadian Meteorological Centre. Part 1: The Global Analysis, *Atmosphere-Ocean*, **37**, 103{156.

[3] Golub, G.H. and C.F. Van Loan, 1996: *Matrix Computations*,(3rd ed.), Johns Hopkins University Press.

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

[5] Haben, S.A., A.S. Lawless and N.K. Nichols, 2009: Conditioning of the 3DVar Data Assimilation Problem, University of Reading, Dept. of Mathematics, Math Report Series 3/2009.

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

[7] Lorenc, A., 1988: Optimal nonlinear objective analysis, *Quart. J. Roy. Met. Soc.*, **114**, 205{240.

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

[9] Lorenc, A.C., 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, 2000: The Met O ce global three-dimensional variational data assimilation scheme, *Quart. J. Roy. Met. Soc.*, 126, 2991{3012.

[10] Rabier, F., H. Jarvinen, E. Klinker, J. F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Quart. J. Roy. Met. Soc.*, 126, 1143{1170.

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

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