Conditioning and Preconditioning of the Optimal State Estimation Problem

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

Abstract

Optimal state estimation is a method that requires minimising a weighted, nonlinear, least squares objective function in order to obtain the best estimate of the current state of a dynamical system. Often the minimisation is non-trivial due to the large scale of the problem, the relative sparsity of the observations and the nonlinearity of the objective function. To simplify the problem the solution is often found via a sequence of linearised objective functions. The condition number of the Hessian of the linearised problem is an important indicator of the convergence rate of the minimisation and the expected accuracy of the solution. In the standard formulation the convergence is slow, indicating an ill-conditioned objective function. A transformation to different variables is often used to ameliorate the conditioning of the Hessian by changing, or preconditioning, the Hessian. There is only sparse information in the literature for describing the causes of ill-conditioning of the optimal state estimation problem and explaining the effect of preconditioning on the condition number. This paper derives descriptive theoretical bounds on the condition number of both the unpreconditioned and preconditioned system in order to better understand the conditioning of the problem. We use these bounds to explain why the standard objective function is often ill-conditioned and why a standard preconditioning reduces the condition number. We also use the bounds on the preconditioned Hessian to understand the main factors that affect the conditioning of the system. We illustrate the results with simple numerical experiments.

Keywords. Optimal state estimation, variational data assimilation, nonlinear least squares, condition number, preconditioning, correlation matrices, circulant matrices

1 Introduction

In dynamical systems, the aim of state estimation is to find the most likely current or future state of the system, given noisy, possibly indirect, observations. In many applications, such as numerical weather prediction (NWP), the number of observations is sparse relative to the dimension of the state space and so additional information, such as a prior estimate of the initial state of the system, is often required to guarantee a unique solution. The optimal state, called the `analysis', minimises a weighted nonlinear least-squares objective function, measuring the distance between the state trajectory and the observations and between the initial state and the prior estimate, weighted by the covariance of the errors in the observations and the prior respectively. In the meteorology community this optimization problem is referred to as four-dimensional variational data assimilation or 4DVar [25]. The analysis is optimal in the sense that, under certain assumptions, it provides the maximum a posteriori Bayesian estimate of the state of the system [22], [18]. Once the analysis is obtained the dynamical model is applied to predict future states.

The model and the observation operator, which maps model states to observations, are often

2 Optimal State Estimation

The aim of state estimation is to find the best estimate of the initial state of the system, $\mathbf{x}_0 \in \mathrm{R}^N$ (called *the analysis*), at time t_0, given a prior estimate $\mathbf{x}_0^b$ (called *the background*) and measurements $\mathbf{y}_i \in \mathrm{R}^{p_i}$ at time t_i ($i = 0, \ldots, n$), taken within a time window $[t_0, t_n]$, and subject to the state space equations

$$\mathbf{x}_i = \mathcal{M}(t_i, t_0, \mathbf{x}_0), \tag{1}$$

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

for $i = 0, \ldots, n$. The notation is as follows:

- the N model states at time t_i are denoted by the vector $\mathbf{x}_i \in \mathbb{R}^N$;
- the non-linear operator $\mathcal{M}(t_i, t_0, .) : \mathbb{R}^N \to \mathbb{R}^N$, describes the evolution of the states from time t_0 to time t_i;
- the non-linear operatori

2.2 Preconditioning

A common method for reducing the condition number of the objective function (4) is to use a linear transformation to change the variables [9]. The process of changing the condition number of the system is known as *preconditioning*. The condition number is minimised when the square root of the inverse of the Hessian is used as the change of variables transformation. However this is generally not practical due to the dimension of the problem and the complexity of the $\mathbf{B}, \mathbf{R}$ and $\mathbf{H}$ matrices. Instead, the symmetric square root of the covariance matrix of the errors in the prior estimates, $\mathbf{B}^{1/2}$, is often used [2], [19], [16]. The errors in the new variables $\ _0 = \mathbf{B}^{-1/2} \ _0$; are now uncorrelated, with unit variances, giving a prior error covariance matrix equal to the identity matrix.

In terms of the new variables, we aim to minimize the transformed objective function

$$\hat{J}(\ _0) = \frac{1}{2}(\ _0 - \ _0^b)^T(\ _0 - \ _0^b) + \frac{1}{2}(\mathbf{HB}^{1/2}\ _0 - \ )^T\mathbf{R}^{-1}(\mathbf{HB}^{1/2}\ _0 - \ ); \qquad (9)$$

with respect to $\ _0$, where $\ _0^b = \mathbf{B}^{-1/2} \ _0^b$.

The effect of the variable transform is symmetrically to precondition the Hessian (7) with the square root of the error covariance matrix of the prior. The Hessian of the preconditioned objective function (9) is now given by

$$= \mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HB}^{1/2}; \qquad (10)$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix throughout the paper.

In general there are fewer observations than states of the system and therefore the matrix $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{HB}^{1/2}$ is not of full rank, but is positive semi-definite. It follows that the smallest eigenvalue of (10) is unity and the condition number of the preconditioned Hessian is equal to its largest eigenvalue.

The aim of this paper is to prove theoretical bounds on the condition number of the unpreconditioned and preconditioned Hessians (7) and (10) respectively. The bounds enable the conditioning of the unpreconditioned and preconditioned Hessians to be compared and help to identify the main factors that affect the conditioning of the objective functions. This work extends the theoretical results presented in earlier work [12] [13], which examine the case of

For a periodic single-variable system discretized on a one-dimensional domain with equal spacing between grid points, many covariance and linear forecast models have a circulant structure. The eigenvalues for circulant matrices have a convenient form which makes them, and hence the condition number, simple to calculate. We exploit this useful property for producing our theoretical bounds. In more general cases, where the domain is not periodic, the autocovariance matrices will be Toeplitz instead. However when the dimension of the state space N is large these Toeplitz matrices and their properties can be approximated by circulant matrices [11], [15].

A circulant matrix has the form of a Toeplitz matrix where each row is a cyclic permutation of the previous row. Let $\mathbf{c} = [c_0, c_1, c_2, \ldots, c_{N-1}]$ denote the top row of a $N \times N$ circulant matrix $\mathbf{C}$. Then the eigenvalues of $\mathbf{C}$ are equal to the discrete Fourier transforms of the coefficients of the first row of the matrix [11] and can be written

$$\nu_m = \sum_{k=0}^{N-1} c_k e^{-2\pi i m k/N}. \tag{11}$$

The corresponding eigenvectors are given by the discrete exponential function,

$$\psi_m = \frac{1}{\sqrt{N}}(1, e^{-2\pi i m/N}, \ldots, e^{-2\pi i m(N-1)/N})^T. \tag{12}$$

Since circulant matrices are normal matrices we can explicitly calculate the condition number of $\mathbf{C}$

which completes the proof. $\square$

An alternative lower bound, which is easier to calculate explicitly, can be obtained using more restrictive assumptions

A3. The observation operator is the same at each time step, that is, $\mathbf{H}_i = \mathbf{H} \in \mathbb{R}^{q\times N}$, where $p_i = q$, for $i = 0, \ldots, n$, and all observations are direct observations of individual states.

A4. The forecast model is assumed to be time invariant with $\mathbf{M}_i := \mathbf{M}^i$ for $i = 1, \ldots, n$, for some circulant matrix $\mathbf{M} \in \mathbb{R}^{N\times N}$.

A5. The symmetric positive-definite error covariance matrix $\mathbf{B} \in \mathbb{R}^{N\times N}$, and hence also its inverse, are circulant.

A consequence of assumption A3 is that $\mathbf{H}_i^T\mathbf{H}_i = \mathbf{H}^T\mathbf{H} \in \mathbb{R}^{N\times N}$ for $i = 0, \ldots, n$ is a diagonal matrix with the k^{th} diagonal entry equal to one if the k^{th} position is observed or zero otherwise. Assumptions A4 and A5 mean that we can explicitly calculate the updated lower bound in the where $\mathbf{A}$ has eigenvalues $\lambda_N(\mathbf{A}) \leq \lambda_{N-1}(\mathbf{A}) \leq \cdots \leq \lambda_2(\mathbf{A}) \leq \lambda_1(\mathbf{A})$. Consider the Rayleigh quotient of and combine the bounds (29) and (30) to give

$$\kappa(\quad) \geq \left(\frac{1 + \frac{q}{N} \frac{{}^{2}_{b}}{{}^{2}_{o}} \lambda_{\min}(\mathbf{C}) \gamma_{\min}}{1 + \frac{q}{N}} \right.$$

4. In the special case where the prior errors are all uncorrelated and $\mathbf{C} = \mathbf{I}_N$, with $\mathbf{I}_N$ being the $N \times N$ identity matrix, then all eigenvalues of $\mathbf{C}$ are unity and the exact condition number $\kappa(\ ) = 1 + \frac{\sigma_b^2}{\sigma_o^2}\lambda_{\max}(\mathbf{H}^T\mathbf{H})$, which is equal to the upper bounds in (14) and (20). In this case the upper bound on the conditioning of the Hessian is strict.

In conclusion, the conditioning of the state estimation problem (4) is strongly dependent on the conditioning of the prior error covariance matrix. With commonly arising prior error covariance matrices, it was shown in [14] that for large correlation length-scales, these matrices are very ill-conditioned and lead to a poorly conditioned Hessian of (4). This is consistent with previous results on variational data assimilation that suggest that the error covariances of the prior estimates are the cause of slow convergence in the minimization of the objective function [19]. In Section 4 we further illustrate the effect of an ill-conditioned prior error covariance matrix on the conditioning of the optimal state estimation problem using simplified numerical experiments.

3.3 Conditioning of the Preconditioned System

In this section we consider the effect of preconditioning the Hessian with the square root of the error covariance matrix of the prior estimate. The following theorem derives new theoretical bounds on the condition number of the preconditioned Hessian (10).

4 *Let* $\mathbf{B} = \sigma_b^2\mathbf{C} \in \mathbb{R}^{N\times N}$ *and* $\mathbf{R} = diag(\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_n) \in \mathbb{R}^{r\times r}$ *be the prior and observation error covariance matrices, respectively, satisfying assumptions A1 and A2. Additionally let* $\mathbf{H} \in \mathbb{R}^{r\times N}$ *be the observation operator defined by (6). Then the following bounds hold on the condition number of the preconditioned Hessian* $= \mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$:

$$1 + \frac{1}{r}\frac{\sigma_b^2}{\sigma_o^2}\sum_{k,l=1}^{r}\{\mathbf{HCH}^T\}_{k,l} \leq \kappa(\ ) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2}||\mathbf{HCH}^T||_\infty, \tag{35}$$

where $\{\mathbf{A}\}_{k,l}$ *represents the* $(k,l)^{th}$ *entry of the matrix* $\mathbf{A}$.

Proof. Since there are fewer observations than variables in the state space ($r < N$), the Hessian is just a low rank update of the identity matrix and its smallest eigenvalue is unity. The condition number of the Hessian is then equal to the largest eigenvalue of . Let $\mathbf{E} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2}$. The matrices $\mathbf{E}^T\mathbf{E} = \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ and 1E

The lower bound is established by applying the Rayleigh quotient of $\mathbf{G}$ with the unit vector $\ = \frac{1}{\sqrt{r}}(1; 1; \ldots; 1)^T \in \mathsf{R}^r$,

$$R_{\mathbf{G}}(\ ) = \ ^T\mathbf{G}\ \ = 1 + \frac{1}{r}\frac{\sigma_b^2}{\sigma_o^2}\sum_{k;l=1}^{r}\{\mathbf{HCH}^T\}_{k;l}. \tag{38}$$

Since $\lambda_{\max}(\mathbf{G}) \geq R_{\mathbf{G}}(\ )$ for any $\ \in \mathsf{R}^{r\times r}$, this completes the proof. $\square$

The bounds on the condition number of the preconditioned Hessian for the special case of observations at only one time step, derived in [13], can be found in the following corollary to Theorem 4.

Corollary 5 *Let $n = 0$, and let $\mathbf{B}$, $\mathbf{R} \equiv \mathbf{R} = \sigma_o^2\mathbf{I}_q$ and $\mathbf{H} \equiv \mathbf{H}$ satisfy assumptions A1-A5. Then the following bounds on the condition number of $\ = \mathbf{I}_N + \mathbf{B}^{1=2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1=2}$ hold*

$$1 + \frac{1}{q}\frac{\sigma_b^2}{\sigma_o^2}\sum_{i;j\in K}\{\mathbf{C}\}_{i;j} \leq \kappa(\ ) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2}||\mathbf{HCH}^T||_\infty; \tag{39}$$

where K are indices of the state variables that are observed.

Proof. From Theorem 4 with $n = 0$ we obtain the bounds

$$1 + \frac{1}{q}\frac{\sigma_b^2}{\sigma_o^2}\sum_{k;\,l=1}^{q}\{\mathbf{HCH}^T\}_{k;\,l} \leq \kappa(\ ) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2}||\mathbf{HCH}^T||_\infty. \tag{40}$$

Since $\mathbf{HCH}^T$ is simply the matrix $\mathbf{C}$ with rows and columns removed at the unobserved positions it follows that

$$\sum_{k;\,l=1}^{q}\{\mathbf{HCH}^T\}_{k;\,l} = \sum_{i;j\in K}\{\mathbf{C}\}_{i;j}; \tag{41}$$

where K are indices of the state variables that are observed. $\square$

Before discussing the implications of the bounds we first note that the matrix $\mathbf{HCH}^T$ which appears in the upper and lower bounds (35) can also be written in the form $\mathbf{HCH}^T = \sigma_b^{-2}\mathbf{HBH}^T$, where $\mathbf{H}$ is the block diagonal matrix consisting of $n+1$ blocks equal to $\mathbf{H}_i;\ i = 0; \ldots; n;$ and $\mathbf{B} = \sigma_b^2\mathbf{C}$ is the four-dimensional error covariance matrix associated with the background state vector $(\mathbf{x}_0^{b\,T}; \mathbf{x}_1^{b\,T}; \ldots; \mathbf{x}_n^{b\,T})^T$. Here $\mathbf{x}_i^b$ denotes the state vector at time t_i, $i = 1; \ldots; n$ evolved from the prior state estimate, $\mathbf{x}_0^b$, using the dynamical model (5) [14]. Since $\mathbf{HCH}^T$ is simply $\mathbf{C}$ with rows and columns deleted at positions that are unobserved, we refer to this as the *reduced* error covariance matrix.

The reduced error covariance matrix $\mathbf{HCH}^T$ plays a key role in the condition number of the Hessian (10). In particular, the lower bound in (35) is linearly related to the *average* row sum $\frac{1}{r}\sum_{k;\,l=1}^{r}\{\mathbf{HCH}^T\}_{k;l}$ whereas the upper bound is related to the absolute maximum row sum $||\mathbf{HCH}^T||_\infty$. In fact the lower and upper bounds are identical if all entries of $\mathbf{HCH}^T$ are positive and its row sums are identical. The dependence on the reduced error covariance matrix implies further details about the condition number of the preconditioned Hessian.

1. The number and positions of the observations are important to the conditioning of the preconditioned problem. In particular, if we assume that the correlations in the prior error covariance matrix decrease with increased distance between grid points and also that the linear model $\mathbf{M}$ acts to ensure that the coefficients of the correlation matrix $\mathbf{C}$ remain positive and decrease monotonically with distance, then increasing the distance between observations will imply smaller entries in the reduced error covariance matrix and thus smaller sums in the upper and lower bounds in (35) and, potentially, a smaller condition number. The assumptions apply, for instance, in the case where the model is an advection equation and the prior error covariance has a Gaussian or SOAR structure [14].

2. Additionally, under the same assumptions, if we have fewer observations at fewer time steps, then there will be fewer entries in the reduced error covariance matrix, implying smaller sums in the bounds and hence a smaller condition number of the Hessian (10).

3. Finally, it follows from the dependence of the bounds (35) on the ratio σ_b^2/σ_o^2 that the accuracy of the observations is also important to the conditioning of the problem. In particular, increasing the accuracy of the observations, where $\sigma_o^2 \longrightarrow 0$ while the other variables remain fixed, implies an increase in the bounds and a potential increase in the conditioning of the Hessian. In the limit, the model trajectories must fit exactly to the data, as in the unpreconditioned case, and the problem becomes much harder to solve and hence more ill-posed.

The bounds (35) and (39) are quite general and do not require the more restrictive assumptions A3-A5 used in the unpreconditioned case. Additionally the bounds do not depend on the condition number of the background error covariance matrix but simply on a summation of the coefficients of a four-dimensional background error covariance matrix. In [12] it was shown that, in the case where observations are only made at a single time step, preconditioning brought a dramatic reduction in the condition number of the Hessian compared to the unpreconditioned case. Contrary to intuition, however, the bounds also show that in the preconditioned case, as well as in the unpreconditioned case, increasing the accuracy and density of observations is likely to make the conditioning of the problem increase and the estimation problem harder to solve accurately.

4 Numerical Experiments

In this section we illustrate the effect of varying different parameters and properties of the state estimation problem on the condition number of the unpreconditioned (7) and preconditioned (10) Hessians. We apply the theoretical bounds derived in Section 3.2 and 3.3, respectively, to explain these effects. Throughout this Section we consider a dynamical system where the state vector consists of a single periodic variable discretized at equally spaced grid points on a one-dimensional domain. As shown in Section 3.2 the prior error covariance plays a influential role in the conditioning of the preconditioned and unpreconditioned Hessian and so we first introduce and describe some of the properties of a common prior error covariance matrix.

4.1 Condition Number of Error Covariance Matrices

In this section we assume that the prior error covariance matrix is of the form $\mathbf{B} = \sigma_b^2 \mathbf{C} \in \mathbb{R}^{N \times N}$ where $\mathbf{C}$ denotes the error correlation matrix and σ_b^2 is the error variance. By definition (8), the condition number $\kappa(\mathbf{B}) = \kappa(\mathbf{C})$. We use the second-order auto-regressive correlation (SOAR) function [4], defined by

$$\rho_S(r) = \left(1 + \frac{|d|}{L}\right) \exp\left(-\frac{|d|}{L}\right), \tag{42}$$

to model the correlation structure, where $L > 0$ is the correlation length-scale and $0 \leq d \in \mathbb{R}$ is the distance between two points on the real line. The SOAR function is commonly used to define correlations in meteorological applications [4]. For a periodic variable we identify the values of the variable at two points $-D$ and D. However, the function (42), which defines a valid correlation function on the real line, may no longer define valid correlation models on the finite interval, since the corresponding Fourier transforms are not necessarily positive [29], [8], [28]. We transform to a valid correlation model on the circle by replacing the distance along the great circle by the chordal distance

$$d = 2a \sin(\theta/2), \tag{43}$$

where θ is the angle between two points on the circle and a is the radius. This guarantees that the corresponding correlation matrix is positive definite [30, Sec. 22.5]. Applying the transform (43) to the SOAR correlation function and sampling at evenly spaced points on the circle s_i, $i = 1, \dots, N$, produces the SOAR correlation matrix $\mathbf{C}_S$ on the circle with elements given by

$$(\mathbf{C}_S)_{i,j} = \left(1 + \frac{|2a\sin(\theta_{i,j}/2)|}{L}\right) \exp\left(-\frac{|2a\sin(\theta_{i,j}/2)|}{L}\right) \tag{44}$$

where $i, j = 1, \dots, N$ and $\theta_{i,j}$ is the angle between the points s_i and s_j on the circle. We note that the resultant correlation matrix is circulant and therefore has eigenvalues given by (11).

Length-scale	0.05	0.1	0.15	0.2	0.25	0.3	0.35
Condition Number	5.96	58.1	265	807	1963	3978	7328

Table 1: The condition number of the SOAR correlation matrix as a function of different correlation length-scales.

Table 1 shows the condition number of $\mathbf{C} = \mathbf{C}_S$, for different length-scales, L, where the correlation function is sampled at $N = 500$ equally spaced grid points on the interval $[-25, 25]$. The table shows that the condition number of the correlation matrix increases as a function of the correlation length-scale. As shown there is a large increase in the condition number as the length-scale increases. An increase [illegible] matrix are positive, we find from (11) that the largest eigenvalue satisfies

$$\lambda_{\max}(\mathbf{C}) = ||\mathbf{C}||_{\infty} = \sum_{k=0}^{N-1} c_k; \tag{45}$$

and therefore increases slowly as a function of L and is bounded by $N = 500$ since $|c_k| \leq 1$. It is the decrease in the smallest eigenvalue that causes the increase in condition number of the correlation matrix [14, Chap. 5]. For the

Since $\mathbf{C}$ and $\mathbf{M}$

For the preconditioned Hessian, the condition number is much smaller than the absolute upper bound predicted by (47) and is much better conditioned than the unpreconditioned Hessian. For instance, at length-scale $L = 0.25$ the condition number of the unpreconditioned Hessian is approximately 1900, whereas for the preconditioned Hessian it is around 6. The conditioning of the preconditioned system increases as the length-scale increases, which can be explained by the increase in the bounds (35). The larger length-scale increases the coefficients of the matrix $\mathbf{C}$ and therefore the size of the row sums of the coefficients of $\mathbf{HCH}^T = \mathbf{HCH}^T$ in the upper and lower bounds (35).

4.3 The Effect of Observations on the Conditioning of the Preconditioned Hessian

We now consider the conditioning of the preconditioned Hessian for the numerical advection forecast model in more detail. The bounds for the preconditioned Hessian (35) and (39) identify the accuracy and positioning of observations as important to the conditioning of the preconditioned objective function.

Assuming the same data as for the experiment shown in Figure 2, we consider the effect of changing the observation accuracy on the condition number of the Hessian. We use the SOAR correlation matrix and fix the correlation length-scale to $L = 0.2$, but vary the observation variance. Table 2 shows the effect of changing the observation accuracy on the condition number of the preconditioned Hessian. As demonstrated in section 3.3, the bounds (35) are linearly related to the inverse of the observation variance and hence we expect the condition number of the Hessian to increase as the observation variance decreases and the accuracy of the observations increases. This is confirmed by the results of the numerical experiment, as seen in Table 2. For instance, a doubling in the accuracy of the observations from a variance of 0.1 to 0.05 roughly doubles the condition number of the Hessian from 51.55 to 102.11. Similar results also hold where other common prior error covariance matrices and observation locations are used (see [14]).

Obs Variance	0.01	0.05	0.10	0.50	1.00	2.00	5.00	10.00
Condition Number	506.53	102.11	51.55	11.11	6.06	3.53	2.01	1.51

Table 2: The condition number of the preconditioned Hessian as a function of the observation error variance using SOAR correlation matrices.

We now consider the condition number of the preconditioned Hessian as a function of the separation of the observations. From the definition of the correlation matrix (44) the coefficients in each block of $\{\mathbf{C}\}_{i,j}$ monotonically decrease as the distance $|i - j|$ increases, as shown in Figure 1. The upper and lower bounds on the Hessian (10) depend on sums of the elements of the matrix $\mathbf{HCH}^T$; which is viewed as a `reduced' covariance matrix. The reduced matrix is simply the 4D covariance matrix $\mathbf{C}$ with all non-observed rows and columns deleted. As the separation of the observations increases, the elements of the reduced matrix become smaller in magnitude due to the decrease in the coefficients (or covariance) with distance. We therefore expect the conditioning of the problem to decrease as the separation of the observations increases or the density decreases. We illustrate this with our numerical model.

We fix the observation error variances to $\sigma_o^2 = 1$ and assume that $q = 20$ observations are made at grid points at each of the time steps $t_0 = 0; t_1 = 3\Delta t$ and $t_2 = 6\Delta t$ with uniform spacing between adjacent observations. We consider the condition number as the uniform spacing is increased. Table 3 shows the results of the experiment. As expected from the theoretical bounds (14), increasing the spacing between the observations reduces the size of the condition number of the Hessian. Since the coefficients of the covariance matrix $\mathbf{C}$, given by (44), decrease with an increase in the distance between sampling points, the condition number of the preconditioned Hessian becomes smaller with larger distances and decreased density of observations. Additionally, as predicted, the condition number is larger for larger length-scales corresponding to the increase in the size of the coefficients of $\mathbf{C}$. Similar results hold for the preconditioned Hessians using other common prior error covariance matrices (See [14, Chap. 7]).

Spacing	1	2	3	4	5	6	7	8	9	10
Condition Number ($L = 0{:}2$)	22.0	12.5	8.9	6.9	5.8	5.1	4.6	4.3	4.0	3.9
Condition Number ($L = 0{:}3$)	29.6	17.6	12.5	9.8	8.1	7.0	6.2	5.6	5.1	4.8
Condition Number ($L = 0{:}5$)	39.8	26.3	19.3	15.2	12.6	10.8	9.4	8.4	7.6	7.0

Table 3: The condition number of the preconditioned Hessian as a function of the number of spaces between observations for different correlation length-scales L.

The results of this section indicate that less accurate and less dense observations reduce the conditioning of the preconditioned Hessian and hence may increase the rate of convergence of the iterative solver used to find the optimal state estimate. These results appear to be condition number of the Hessian of the objective function in both the unpreconditioned and preconditioned forms. The bounds derived identify the main sources of ill-conditioning in both systems and explain how preconditioning can improve the conditioning of the problem. In particular, we found that the condition number of the unpreconditioned Hessian is proportional to the condition number of the prior error covariance matrix. Hence an ill-conditioned prior error covariance matrix can produce an ill-conditioned Hessian. The bounds on the preconditioned system showed that preconditioning using the prior error covariance matrix can produce a significant reduction in the condition number of the Hessian. Additionally, the distribution, quantity and accuracy of the observations play key roles in the conditioning of the preconditioned Hessian, with more accurate and dense observations creating a more ill-conditioned problem.

We presented results from numerical experiments in order to demonstrate the e ect of the various factors on the condition number of the Hessians, as indicated by the bounds. We presented the SOAR covariance matrix, which is commonly used in variational data assimilation, and showed that the conditioning of this matrix becomes very ill-conditioned for only relatively small increases in correlation length-scale. We then demonstrated that this prior error covariance matrix resulted in the ill-conditioning of the unpreconditioned Hessian and that preconditioning dramatically reduced the conditioning, as predicted by the theoretical bounds. We also illustrated the reduction in the conditioning of the preconditioned system as we increased the separation between observations and reduced the accuracy of the observations, as expected from our theoretical results. We remark that the conclusions derived from the theory presented here have also been found to hold for experimental data from the high-dimensional, multi-variable Met O ce Numerical Weather Prediction data assimilation system [13].

A simple, natural extension to this problem would be to consider more general observation operators which incorporate interpolation and to introduce correlations into the observation errors. Very recently, extra preconditioning, in addition to the variable transform via the matrix $\mathbf{B}$, has been considered [27], [6] for use in optimal state estimation. Further exploration and analysis of these, and other, preconditioning techniques, following the theoretical approach presented here, may be valuable in order to produce further improvements in the conditioning of the problem.

Ac . . This research has been supported in part by the National Centre for Earth Observation, the UK Engineering and Physical Sciences Research Council and the Met O ce.

References

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

[2] R. Bannister, *A Review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics*, Q. J. R. Met. Soc., **134**, 1971{1996, 2008.

[3] P. Courtier, J.-N. Thépaut and A. Hollingsworth, *A strategy for operational implementation of 4D-Var, using an incremental approach,* Q. J. R. Met. Soc., **120**, 1367{1387, 1994.

[4]

[21] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Di erential Equations*, Cambridge University Press, 1994.

[22] N.K. Nichols, *Data Assimilation: Aims and Basic Concepts* In: R. Swinbank, V. Shutyaev and W.A. Lahoz, (ed.) Data Assimilation for the Earth System. Kluwer Academic, pp. 9-20, 2003.

[23] S. Pajic, *Power System State Estimation and Contingency Constrained Optimal Power Flow - A Numerically Robust Implementation*, PhD Thesis, Worcester Polytechnic Institute, 2007.

[24] 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**, 347{362, 2007.

[25] O. Talagrand and P. Courtier, *Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory*, Q. J. R. Met. Soc., **113**, 1311{1328, 1987.

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

[27] J. Tshimanga, S. Gratton, A. T. Weaver and A. Sartenaer *Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation*, Q. J. R. Met. Soc., **134**, 751-769, 2008.

[28]