THE UNIVERSITY OF READING

DEPARTMENT OF MATHEMATICS AND STATISTICS

Conditioning of the Weak-Constraint Variational Data Assimilation Problem for Numerical Weather Prediction

Adam El-Said

Thesis submitted for the degree of

Doctor of Philosophy

July 2015

Abstract

4-Dimensional Variational Data Assimilation (4DVAR) assimilates observations through the minimisation of a least-squares objective function, which is constrained by the model flow. We refer to 4DVAR as *strong-constraint* 4DVAR (sc4DVAR) in this thesis as it assumes the model is perfect. Relaxing this assumption gives rise to *weak-constraint* 4DVAR (wc4DVAR), leading to a different minimisation problem with more degrees of freedom. We consider two wc4DVAR formulations in this thesis, the model error formulation and state estimation formulation.

The 4DVAR objective function is traditionally solved using gradient-based iterative methods. The principle method used in Numerical Weather Prediction today is the Gauss-Newton approach. This method introduces a linearised 'inner-loop' objective function, which upon convergence, updates the solution of the non-linear 'outer-loop' objective function. This requires many evaluations of the objective function and its gradient, which emphasises the importance of the Hessian. The eigenvalues and eigenvectors of the Hessian provide insight into the degree of convexity of the objective function, while also indicating the difficulty one may encounter while iterative solving 4DVAR. The condition number of the Hessian is an appropriate measure for the sensitivity of the problem to input data. The condition number can also indicate the rate of convergence and solution accuracy of the minimisation algorithm.

This thesis investigates the sensitivity of the solution process minimising both wc4DVAR objective functions to the internal assimilation parameters composing the problem. We gain insight into these sensitivities by bounding the condition number of the Hessians of both objective functions. We also precondition the model error objective function and show improved convergence. We show that both formulations' sensitivities are related to error variance balance, assimilation window length and correlation length-scales using the bounds. We further demonstrate this through numerical experiments on the condition number and data assimilation experiments using linear and non-linear chaotic toy models.

Acknowledgments

بسمِ اللهِ الرحمانِ الرحيم

(In the name of God, the most beneficent, the most merciful)

I would like to thank my supervisors Prof. Nancy Nichols and Dr. Amos Lawless for all the valuable guidance and support they have given me over the course of the PhD. All the discussions and motivation you provided me with were a major contribution to me making it this far. I would also like to thank the Natural Environment Research Council (NERC) for the funding I have received.

On a more personal note I'd like to thank Prof. Mike Baines for helping me through difficulties encountered in my first year, not to mention the countless philosophical discussions about many other irrelevant things(!). I would also like to thank Dr. Roel Stappers for our numerous discussions and the amount of knowledge he let me siphon from him in the second year(!). Both Roel and Mike stand out as some of the most intelligent, polite and patient people I have ever encountered. It felt good to have colleagues and friends I could relate to especially in the lonely world of research. Both of your interests to help just for the sake of helping was very refreshing, Thanks again.

The real special thanks goes to my parents, Abdel-Monem, Fatin and my wife, Sarah. My parents played *the* role providing me with the environment to grow and evolve into who I am and do the things I have done. If I were to use every breath I had left to say 'Thank you' it would not be enough. My gorgeous wife Sarah, the beautiful soul that brightened up my life in the third year deserves a large portion of my thanks. Her love and unconditional support through my ups and downs remains unrecompensable. I hope with the arrival of our son, Nooh (Noah), we will continue to nurture our love for one another til death do us part.

The PhD journey for me has been a knowledge-seeking, humbling and a much needed self-finding experience. God knows the extent to which I asked to be guided to *the path* in a way suited to my philosophy. I asked Him for this using every iota that makes up my will and intention and He granted it to me along with the fruits that come with it. I am eternally and truly grateful to Him for guiding me with his mercy to the path leading back to *The Origin.*

سبحانَ ربِّك ربِّ العزَّةِ عمَّا يصفونَ وسلامٌ على المرسلينَ والحمدُ للهِ ربِّ العالمين

"Exalted is your Lord, the Lord of might, above what they describe,
And peace be upon the chosen envoys,
And praise be to God, Lord of the worlds."

(Quran 37:180-182)

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Adam El-Said

Contents

List of Figures

Chapter 1

Introduction and Motivation

The aim of data assimilation is to provide a statistically optimal estimate of the state of a system given a set of observations and a dynamical model. There are various data assimilation techniques used for a variety of problems in numerical weather prediction (NWP), earth sciences, oceanography, agriculture, ecology and the geo-sciences. The complexity of the data assimilation problem is related to the area of application, since the *size* and the *dynamics* of the system or model is dependent on the application.

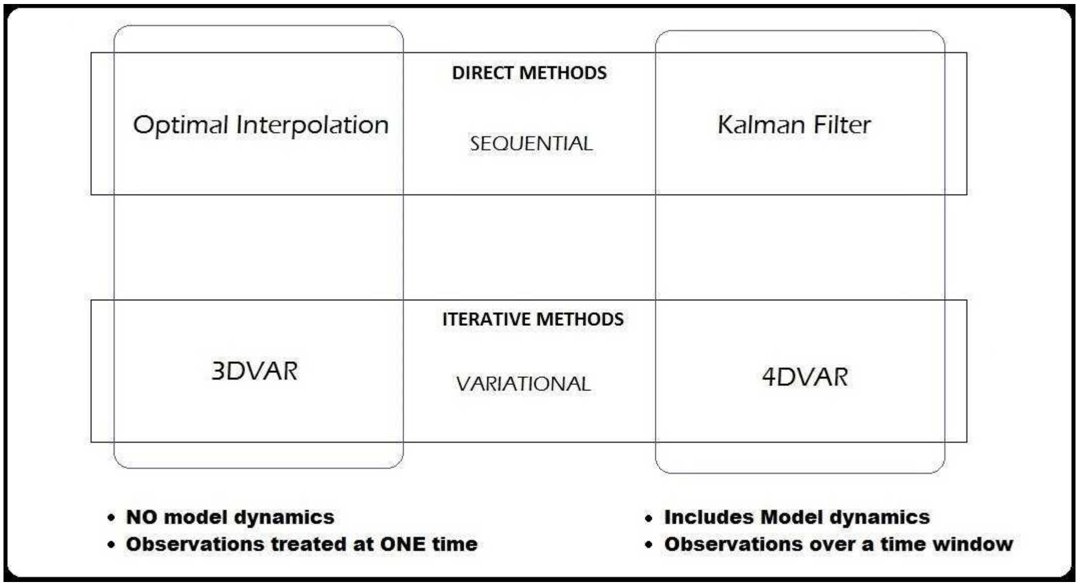


Figure 1.1: Classification of popular data assimilation techniques.

Figure 1.1 is diagrammatic representation of data assimilation techniques and their classification. Each technique has several sub-categories which we deliberately omit. For the remainder of the thesis we abbreviate the optimal interpolation technique as OI, 3-dimensional variational data assimilation as 3DVAR, 4-dimensional variational data assimilation as 4DVAR and the Kalman-filter equations as KF.

The standard 4DVAR approach seeks a statistically optimal fit to the observations, subject to the constraint of the flow, or the model of the physical process for which we are assimilating data. The statistical uncertainties are represented by the 4DVAR objective function, which aims to minimise the mismatch between the model trajectory and the background and observations. The errors in these two quantities are assumed to be independent of each other and possess Gaussian statistics with zero mean. The main assumption of 4DVAR is that the model describing the state contains no errors, which explains the occasional reference to the standard 4DVAR approach as strong-constraint 4DVAR (sc4DVAR). The 4DVAR objective function is traditionally minimised using gradient-based iterative techniques. In the context of NWP, the Gauss-Newton approach is used, introducing a series of linearised 'inner-loop' objective functions. Minimising the objective function is an *optimisation* problem, which in an NWP context requires several evaluations of both the objective function and its gradient to converge on a suitable solution.

The research in this thesis focuses on a more general form of 4DVAR known as weak-constraint 4DVAR (wc4DVAR). The wc4DVAR problem relaxes the strong model constraint by allowing for errors in the model. This modifies the objective function slightly and increases the degrees of freedom of the problem, while also introducing a more complicated optimisation problem than sc4DVAR. The primary focus of the thesis is identifying the sensitivities of the minimisation process to the input data composing the data assimilation problem. We explain this in more detail as the thesis unfolds.

We begin by briefly introducing the evolution of the data assimilation field up to where it is today, followed by current research involving relevant applications of wc4DVAR. We then state the aims of our research and then give a chapter overview of the thesis.

1.1 Brief Historical Background

In the 1950s there was significant theoretical research progress around the weather forecasting problem, which led to a variety of mathematically similar yet differently formulated ideas, forming the basis of data assimilation. The first marked attempt was by Gilchrist and Cressman, [33], where they use a least-squares method to fit a second degree polynomial presented by their interpretation of a simplified meteorological system. A serially successive correction technique was introduced by Bergthorsen and Döös, [8], where they added statistically weighted increments to a prior estimate. Variational data assimilation was theoretically suggested by Sasaki in the late 1950s in the same era as the OI and KF techniques, [76], [77]. The KF [48] and OI [29] techniques eventually made their way into the weather forecasting arena by the 1960's. The variational techniques at this time were not receiving as much research attention as the OI or KF variants. The strength of variational techniques was not yet realised.

Sasaki formally defined 'Variational formalism with weak constraint' as early as 1970, [78]. The weak-constraint variational formulation of the data assimilation problem has received increased attention in the last two decades, [38], [39], [72], [5], [83], [56], [14]. Weak-constraint 4DVAR is most useful when used with observations of a dynamical system or process that perhaps is not yet well-understood.

Notable distinctions and advantages of the variational techniques is the inclusion of model dynamics and feasibility for very large problems such as those in NWP. 4DVAR became feasible for operational NWP centres in 1994, [13], with the introduction of 'Incremental 4DVAR', nearly 30 years after its theoretical formulation. It was implemented for the first time by the European Center for

Medium-Range Weather Forecasts (ECMWF) in 1997, documented in [74], [64] and [50]. The Met Office then followed with their operational implementation of 4DVAR in 2004, [75].

Operational NWP centres in the last 25 years have largely concentrated their efforts in implementing variational techniques for longer range forecasting due to their computational feasibility. Variational techniques are difficult to implement compared to KF or OI because one of the components required to calculate the gradient is a backward or 'adjoint' model. Writing adjoint code is one of the main sources of difficulty and it can take years for scientists to correctly code these for very large NWP models, [75], [74]. The KF technique is infeasible for large problems such as those in NWP because KF requires propagation of background error covariances, which is too computationally expensive. However, there are studies beginning to emerge showing that KF variants may be practicable for large NWP systems. Comparisons between ensemble 4DVAR (4DEnVAR) variants and NWP-applied ensemble KF (EnKF) variants highlight the ease of implementing EnKF over hybrid-4DVAR due to the absence of an adjoint, [59], [22] [60].

The most recent developments surrounding the variational techniques is the implementation of the hybrid 4DVAR technique. These techniques aim to remedy the weakness in sc4DVAR where the background matrix is unable to capture 'errors of the day'. At the Met Office, hybrid 4DVAR utilises a variable transformation technique to combine the conventional climatological estimates of the background error covariance matrix with data from the 23-member Met Office ensemble prediction system (MOGREPS). This has been implemented by the Met Office in their global model as of July 2012, [10]. The Met Office are also attempting to develop a hybrid 4DEnVAR technique, which if successful will alleviate the need for linearised and adjoint models. The difference between hybrid 4DVAR and hybrid 4DEnVAR is that 4DEnVAR uses a localised linear combination of non-linear forecasts, whereas hybrid 4DVAR uses the linearised model and its adjoint. A comparison between these two techniques shows that the currently operational hybrid 4DVAR method is still superior to the proposed hybrid 4DEnVAR, [60].

We now briefly highlight broader application areas of wc4DVAR related to the earth system as a whole.

1.2 Applications of Weak-Constraint 4DVAR

In oceanography wc4DVAR has been used to study the tropical ocean circulation with a simple coupled ocean-atmosphere model, [6] and [7], where the authors discuss how the implementation of this technique led to the improvement in part of the model physics describing this process. The authors refer to their weak-constraint 4DVAR formulation as the 'iterated indirect representer method' in these papers.

The US Naval Research Laboratory (NRL) initially trialled a wc4DVAR formulation using the Burgers' equation, with the aim of understanding how to obtain model error covariance statistics, [93]. They later implemented wc4DVAR both in 'primal' and 'dual' forms for assimilating ocean observations with the bigger Navy Coastal Ocean Model. They did this using both using synthetic observations [70], and real observations [71], and they discuss differences between the 4DVAR and wc4DVAR systems. They conclude that wc4DVAR has lower solution errors than sc4DVAR, when compared to the truth. The 'primal' form of 4DVAR is the standard approach which solves the problem in what is known as 'state space'. Whereas the 'dual' form of 4DVAR maps the problem into 'observation space', which is much smaller than state space.

The University of California in collaboration with some other universities and the Institute of Marine and Coastal Sciences have detailed their incremental sc4DVAR and wc4DVAR systems, both in primal and dual forms applied to their Regional Ocean Modeling System (ROMS) in a lengthy three-part paper, [68], [66], [67].

A discussion of the implementation of wc4DVAR to the upper stratosphere model at the ECMWF on their pre-operational Integrated Forecast System (IFS) can be found in [84]. The operational application is discussed in [56] and [27]. The ECMWF briefly implemented a bias-only corrective version of wc4DVAR, but this has been suspended due to numerical conditioning issues, which is an area we address in this thesis theoretically, [personal communications with Mike Fisher and Yannick Tremolet, 2013], [Poster by Stephen English, ECMWF Research Dept: `https://cimss.ssec.wisc.edu/itwg/itsc/itsc19/program/posters/nwp_3_english.pdf`].

Another growing area of research that has begun implementing wc4DVAR is earth and soil observation. The main problem in this area is that the current models are not an accurate representation of terrestrial ecosystems. There is also the issue of models not being coupled with each other. So for example in the event of a forest fire, abrupt changes in the state would take place in a separate radiative transfer model which will have an effect on the terrestrial model, however, the terrestrial model would not be able to detect this, [55].

The wc4DVAR approach has only gained proper research attention in the last decade. The application of wc4DVAR is suited to problems where the dynamical model of a given system is known to contain errors. The errors could be biases, random errors, model parameter errors or errors in the model physics. Realising the nature of these errors by allowing for their estimation could potentially improve understanding of the process being assimilated, so aswell as a forecasting tool it could be used to diagnose errors in the physics of the process being modeled.

We now detail the aims of our research in this thesis.

1.3 Aims of Research

The weak-constraint variational problem introduces many more degrees of freedom in comparison to sc4DVAR, which only estimates the initial state required to initialise the model. Wc4DVAR seeks an optimal estimate of the states across the assimilation window, given the error statistics in the background, observations *and* the model. The problem is fully *4-dimensional* since it seeks temporally evolving information, states or model errors, rather than just the initial conditions.

There are two formulations of the weak-constraint problem at the focus of this thesis. One formulation aims to estimate the initial state and the model errors for each time interval within the assimilation window. The alternative formulation aims to estimate *all* the states at each time interval within the assimilation window.

More specifically, the aim of the research is to:

- Investigate differences in the characteristics of the solution process between the wc4DVAR model error and state formulations with identical input data.
- Establish theoretical grounding to identify the data assimilation parameters that are the most influential on the solution process of both the wc4DVAR model error formulation and state formulations.
- Determine the scope of our findings by applying both wc4DVAR formulations to a non-linear chaotic model with similar error growth characteristics to full NWP models.

In this thesis we examine the theoretical condition numbers of the Hessians of the wc4DVAR objective functions. The condition number measures the sensitivity of non-linear functions to small changes in their input data. We use the condition number of the wc4DVAR objective functions' first-order Hessians to quantify their level of sensitivity to changes in the assimilation parameters governing the data assimilation problem. We use this as insight as to how the gradient-based iterative solvers will perform when used to solve the wc4DVAR objective functions.

The problem is said to be *well-conditioned* if the solution is not greatly effected by the initial input data, otherwise the problem is said to be *ill-conditioned*

The new main results that we show in the thesis are as follows:

- The condition number of the sc4DVAR Hessian is bounded above by the condition number of the wc4DVAR model error formulation Hessian.

- There are clear differences in the number of iterations required for convergence, solution error and numerical condition numbers of the wc4DVAR model error and state formulations using a simple 1-dimensional advection model. These differences are evident when subjecting both formulations to changes in assimilated error variances, correlation length-scales, spatial observation densities and assimilation window lengths.

- The condition number of the Hessian of the wc4DVAR model error formulation and hence the iterative solution process, is sensitive to longer correlation length-scales, increased observation density and assimilation window length. It is also sensitive to the balance of the specification of background, observation and model error variance ratios.

- Preconditioning the wc4DVAR model error formulation using the symmetric square-root of the background and model error covariance matrix improves the condition number of the Hessian and the convergence rate of the solution process of the model error formulation.

- We show that the condition number of the Hessian of the wc4DVAR state formulation and hence the iterative solution process is *very* sensitive to the background and model error covariance matrix, more so than the wc4DVAR model error formulation. This formulation also exhibits sharp sensitivity to the *decrease* in observation density. It also exhibits a sharp sensitivity to assimilation window length in the event of scarce observations, but as the observation density approaches full rank the state formulation is no longer effected by assimilation window length.

1.4 Thesis Overview

In Chapter 2 we present the variational data assimilation problem. We also discuss the incremental 4DVAR and control variable transform (CVT) techniques which are used to enable operational execution of the variational algorithm. We then introduce the two weak-constraint variational methods and extend the incremental and CVT techniques to wc4DVAR followed by a short discussion of the Hessian structures of the two wc4DVAR formulations. Finally, we review the current literature more closely linked to the wc4DVAR formulations at the focus of the thesis.

In Chapter 3 we introduce the definition of the condition number used in this thesis as a measure to quantify the sensitivities of the variational problem to changes in its input parameters. We then detail the iterative solvers used to solve the 4DVAR optimisation problem. This is followed by an overview of the particular class of matrix, which are shared by the two covariance structures in the experiments conducted in our research. We then discuss the mathematical techniques and theorems used to obtain the results in the thesis. We then introduce the two models used in our theory and experiments.

In Chapter 4 we detail the practical implementation considerations of both the model error and state estimation wc4DVAR problems. We then detail the experimental design and examine their numerical minimisation characteristics when applied to the 1-dimensional advection equation model.

In Chapter 5 we examine the condition number of the Hessian of the model error objective function. We derive new theoretical bounds on the condition number of the Hessian and derive theoretical insight from the bounds. We explore the sensitivities of the condition number to input data by demonstrating the bounds through numerical experiments, both on the condition number and the iterative solution process. We precondition the problem and derive similar theoretical results and demonstrate in a similar fashion that the overall conditioning of the preconditioned problem is improved as a result.

Chapter 6 is dedicated to examining the condition number of the Hessian of the state estimation objective function. We derive new theoretical bounds on the condition number of the Hessian and derive theoretical insight from the bounds. We examine and highlight certain properties of this Hessian that are uniquely different from the model error formulation Hessian. We demonstrate all our findings through numerical experiments on the condition number and the solution process of the state estimation problem.

In Chapter 7 we implement both weak-constraint formulations on the Lorenz-95 system and show that the sensitivities of both formulations obtained in Chapters 5 and 6 also hold for a non-linear chaotic model.

Chapter 8 concludes our work and discusses avenues for further work.

Chapter 2

Variational Data Assimilation

We introduce the Gauss-Newton ‘incremental’ and CVT techniques currently used for sc4DVAR. We then introduce the two wc4DVAR formulations. We then extend the theory of the Gauss-Newton and CVT concepts to both formulations and briefly discuss the structures of the two wc4DVAR Hessians. We conclude the chapter with a literature review of applications of wc4DVAR in NWP and current understanding of the conditioning of the wc4DVAR problem.

We begin by detailing the style of notation used in this thesis.

2.1 Notation and Assumptions

Matrices and Vectors

Bold upper-case letters denote partitioned matrices, meaning a matrix of matrices. In this thesis we refer to these partitioned matrices as 4-dimensional (4D) since they possess spatial and temporal information. Matrices with a normal font represent a standard $N \times N$ matrix as opposed to a partitioned 4D $Nn \times Nn$ matrix, for $N, n \in \mathbb{N}$, where N refers to the spatial dimension and n denotes the temporal dimension. Similarly, we represent 4D partitioned vectors with bold lower-case letters and normal vectors of size N are written in normal font.

Operators

This notation also interlinks between operators and matrices. We denote non-linear operators using calligraphic font whereas a non-linear operator which has been differentiated and linearised around a point is denoted with normal font, which can then also be represented as a matrix. This also applies to 4D operators, so a linearised 4D operator for example would be bold. Letters with standard font denote linear or linearised operators,which can be represented in matrix form.

Condition Number

The condition number used throughout this chapter is the 2-norm condition number, composed of the ratio of the largest and smallest eigenvalue of a symmetric positive-definite matrix. We formally introduce the condition number in Chapter 3 Section 3.1.

We now introduce the sc4DVAR problem.

2.2 Strong-Constraint 4DVAR

The aim of data assimilation is to merge the trajectory of a model with observational data from the process being modeled. In sc4DVAR the model is assumed to be perfect meaning each state is described exactly by the model equations. The errors therefore in the strong-constraint problem are the background, a previous forecast, and the observations. The objective is to seek the model initial conditions which minimises the distance between the model trajectory and the background and observations.

We begin by writing the model evolution of the states as

$$x_i = \mathcal{M}_{i,i-1}(x_{i-1}),\ i = 1, ..., n\ . \tag{2.1}$$

The model is a discrete non-linear operator $\mathcal{M}_{i,i-1} : \mathbb{R}^N \to \mathbb{R}^N$ evolving the model state $x_i \in \mathbb{R}^N$ from time t_{i-1} to time t_i on the closed time interval $[t_0, t_n]$ where $\mathcal{M}_{i,i} = I_N$. The model state can have several spatial points and contain additional parameters or boundary conditions that govern the behaviour of the model. In this thesis we only consider models initialised by their respective states without any additional parameters.

The model integrations can be factorised into smaller integrations using the subscript time-stepping notation as follows

$$\mathcal{M}_{n,0}(x_0) = \mathcal{M}_{n,n-1}...(\mathcal{M}_{2,1}(\mathcal{M}_{1,0}(x_0))). \tag{2.2}$$

We utilise this notation througout the thesis. Now that we have discussed the model, we briefly introduce the notion of observations in variational data assimilation related to NWP.

There is a wide network of observations gathered with the use of various instruments and methods for obtaining measurements in NWP. For example, radiosondes are attached to weather balloons, which are sent up through the layers of the atmosphere collecting data such as pressure, humidity and temperature. Observations are also obtained through satellite radiances, aircrafts and buoys in the ocean. The process of translating the observations into data which can be compared with the model presents its own inverse problem, but this is incorporated into the variational problem as we will see shortly. An example of such a complex problem is the translation of Atmospheric InfraRed Sounder (AIRS) radiance data, which involves characterising the errors in the measured radiances and the radiative-transfer model, [65]. In practice the number of the observations is $\sim \mathcal{O}(10^6)$ whereas the number of variables in the state is significantly larger $\sim \mathcal{O}(10^8)$, [51].

Let $y_i \in \mathbb{R}^p$ denote the raw observation value at time i and let $\mathcal{H}_i(x_i)$ denote the non-linear observation operator, which maps the model equivalent of y_i from state space to observation space such that $\mathcal{H}_i : \mathbb{R}^N \rightarrow \mathbb{R}^p$. Therefore we have

$$\mathcal{H}_i(x_i) - y_i = \epsilon_i^o,\ i = 0, ..., n\ , \tag{2.3}$$

where $\epsilon_i^o \in \mathbb{R}^p$ denotes the observation error at t_i. The errors in the observations are typically assumed to be uncorrelated with all other types of error, and of the form

$$\epsilon_i^o \sim N(0, R_i),\ i = 0, ..., n\ , \tag{2.4}$$

where $R_i \in \mathbb{R}^{p \times p}$ is the observation error covariance matrix and the mean is equal to zero. The assumption of a normal distribution allows the distributions to be defined by the mean and covariance, which simplifies the problem. The Gaussian assumption in (2.4) is still currently used by leading weather centres' 4DVAR implementations, such as the Met Office and the ECMWF, [74], [75], [13].

Next, we consider model trajectory errors. Initial conditions x_0, produce a model trajectory by utilising the non-linear model described in (2.1), with states at each time $(x_1, ..., x_n)$. The initial conditions that produce the previous forecast trajectory, is known as the 'background', denoted as x_0^b. The background is the solution of a previous 4DVAR application, since variational data assimilation is a cyclic process. We therefore have a background trajectory such that

$$x_i^b = \mathcal{M}_{i,i-1}(x_{i-1}^b),\ i = 1, ..., n\ , \tag{2.5}$$

with initial conditions x_0^b producing a trajectory $(x_1^b, ..., x_n^b)$. The error associated with the background is such that

$$x_0 - x_0^b = \epsilon_0^b, \tag{2.6}$$

where the error is such that

$$\epsilon_0^b \sim N(0, B_0). \tag{2.7}$$

The background error $\epsilon_0^b \in \mathbb{R}^N$ is assumed to be uncorrelated with all other types of error, have a zero mean and a background error covariance matrix such that $B_0 \in \mathbb{R}^{N \times N}$.

So the aim of the variational problem is to minimise the errors in (2.6) and (2.3) with respect to the states x_i for $i = 0, ..., n$, subject to the constraint of the *perfect model* (2.1).

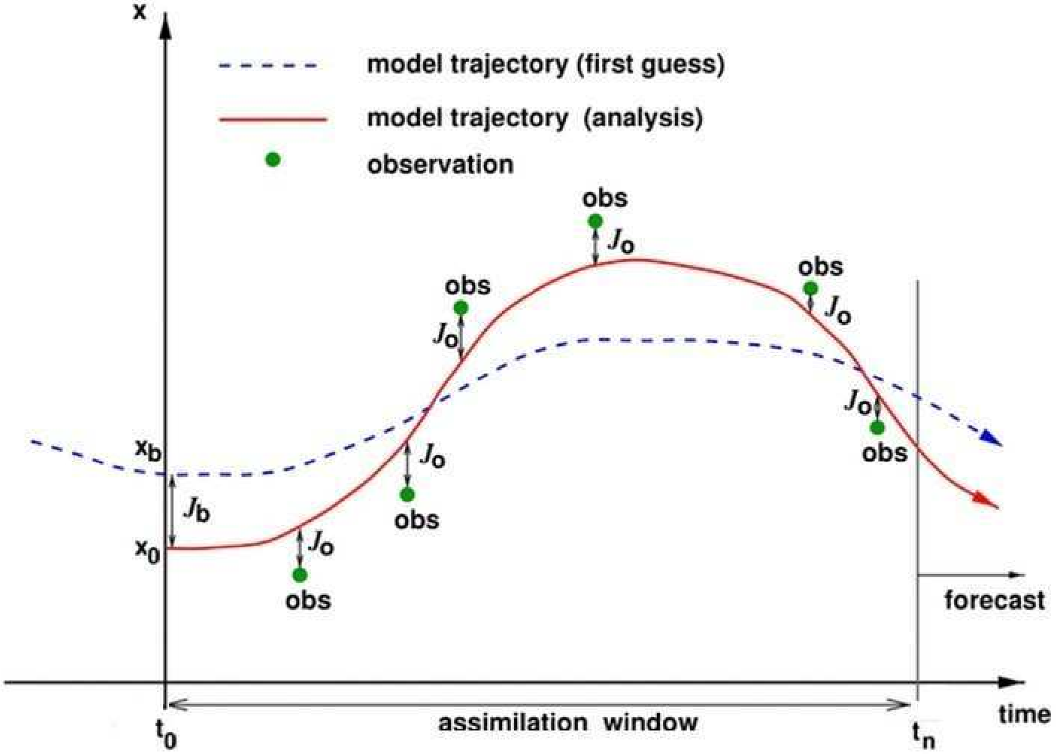


Figure 2.1: Strong-constraint 4DVAR assimilation window with following forecast trajectory. Background estimate (blue dotted line) and solution (red line). (Diagram template courtesy of ECMWF training course presentation by Phillipe Lopez)

Figure 2.1 is a pictorial representation of sc4DVAR. The aim is to find the model trajectory (red line), which minimises the distances between the background (blue dotted line) and the temporally distributed observations (green dots), within the assimilation window. Therefore, sc4DVAR seeks the initial model state x_0, which gives a trajectory that minimises the errors in the background and observations such that it minimises the following

$$\underset{x_0}{min}\ \mathcal{J}(x_0) = \frac{1}{2}\underbrace{(x_0 - x_0^b)^T B_0^{-1}(x_0 - x_0^b)}_{\mathcal{J}_b}$$
$$+\frac{1}{2}\sum_{i=0}^{n}\underbrace{(\mathcal{H}_i(\mathcal{M}_{i,0}(x_0)) - y_i)^T R_i^{-1}(\mathcal{H}_i(\mathcal{M}_{i,0}(x_0)) - y_i)}_{\mathcal{J}_o}, \tag{2.8}$$

where $\mathcal{J} : \mathbb{R}^N \to \mathbb{R}$. Solving the minimisation problem presented by the sc4DVAR objective function, (2.8), provides the initial conditions for the non-linear model $\mathcal{M}$, which minimises the errors in the background $\mathcal{J}_b$ and the observations $\mathcal{J}_o$.

The gradient equation is as follows

$$\nabla \mathcal{J}(x_0) = B_0^{-1}(x_0 - x_0^b) + \sum_{i=0}^{n} M_{0,i}^T H_i^T R_i^{-1}(\mathcal{H}_i(\mathcal{M}_{i,0}(x_0)) - y_i), \qquad (2.9)$$

where the Jacobian of $\mathcal{M}$ is denoted as M, which is known as the *tangent linear* or linearised model and M^T is traditionally known as the linearised *adjoint model.*

The first-order Hessian of (2.8) is

$$S = B_0^{-1} + \sum_{i=0}^{n} M_{0,i}^T H_i^T R_i^{-1} H_i M_{i,0}. \qquad (2.10)$$

The sc4DVAR problem is typically solved using gradient-based iterative procedures requiring evaluation of the objective function (2.8) and its gradient (2.9) numerous times. This fully non-linear form of 4DVAR is not directly practicable for the large problems in NWP. We now introduce the most prominent solution approach, which enabled 4DVAR to be practicable on NWP systems.

2.2.1 Incremental 4DVAR

The Gauss-Newton approach to the sc4DVAR problem, which is now known as incremental 4DVAR to the NWP community, was introduced in 1994 unlocking the operational practicality of 4DVAR for the first time, [13]. It was then introduced into the operational systems of leading weather centres around the world between 1997-2005, ECMWF (1997) [74], Japanese Meteorological Agency (2005) [47], Met Office (2007) [75] and Canadian Met Service (2007) [31].

We begin by introducing iterates, k, such that

$$x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}, \qquad (2.11)$$

where the first guesses for $k = 0$ are

$$x_0^{(0)} = x_0^b,$$
$$\delta x_0^{(0)} = 0.$$

We approximate the non-linear operators in (2.8) to first-order such that

$$\begin{aligned}\mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)})) &= \mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)} + \delta x_0^{(k)})),\\ &\approx \mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)})) + (\mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)})))'\delta x_0^{(k)},\\ &= \mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)})) + (H_i M_{i,0})_{x_0^{(k)}}\delta x_0^{(k)}. \end{aligned} \tag{2.12}$$

Thus an 'incremental objective function' can be written in terms of the increment $\delta x_0^{(k)}$,

$$\begin{aligned}\underset{\delta x_0^{(k)}}{min}\ J(\delta x_0^{(k)}) = \frac{1}{2}&(\delta x_0^{(k)} - (x_0^b - x_0^{(k)}))^T B_0^{-1}(\delta x_0^{(k)} - (x_0^b - x_0^{(k)}))\\ &+ \frac{1}{2}\sum_{i=0}^{n}(H_i M_{i,0}\delta x_0^{(k)} - d_i)^T R_i^{-1}(H_i M_{i,0}\delta x_0^{(k)} - d_i),\end{aligned} \tag{2.13}$$

where

$$d_i = y_i - (\mathcal{H}_i(\mathcal{M}_{i,0}(x_0^{(k)})). \tag{2.14}$$

Solving problem (2.13) is known as the 'inner-loop'. The inner-loop objective function (2.13) can be minimised directly using an iterative method, or by solving the gradient equation at the minimum ($\nabla J = 0$),

$$(B_0^{-1} + \sum_{i=0}^{n} M_{0,i}^T H_i^T R_i^{-1} H_i M_{i,0})\delta x_0^{(k)} = \sum_{i=0}^{n} M_{0,i}^T H_i^T R_i^{-1} d_i + B_0^{-1}(x_0^b - x_0^{(k)}). \tag{2.15}$$

We can see that (2.15) is simply the linearised sc4DVAR Hessian applied to δx_0, with the initial input data comprised of the errors in the background and observations on the right-hand side. The incremental 4DVAR Hessian of (2.13) is identical to the first-order Hessian of the non-linear objective function (2.10). Minimising the inner-loop objective function yields a new increment δx_0 to update the current guess for the outer-loop objective function via (2.11).

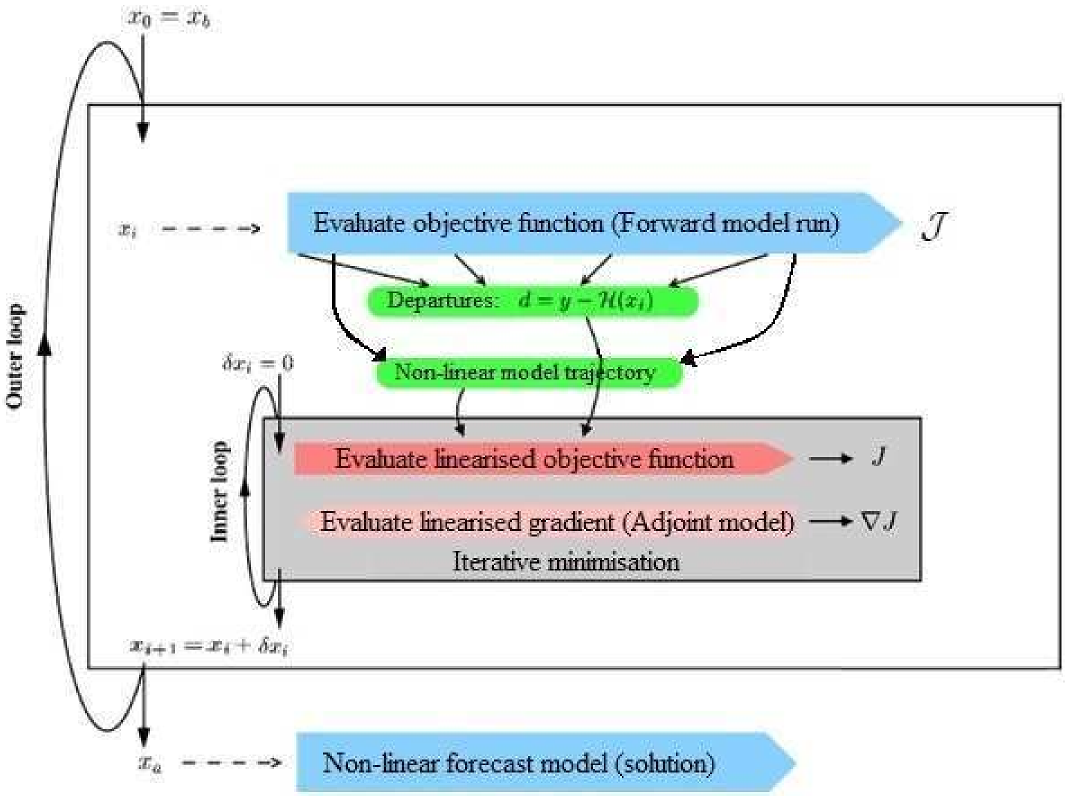

$x_0 = x_b$
x_i
Evaluate objective function (Forward model run)
$\mathcal{J}$
Departures: $d = y - \mathcal{H}(x_i)$
Non-linear model trajectory
Outer loop
$\delta x_i = 0$
Inner loop
Evaluate linearised objective function
J
Evaluate linearised gradient (Adjoint model)
∇J
Iterative minimisation
$x_{i+1} = x_i + \delta x_i$
x_a
Non-linear forecast model (solution)

Figure 2.2: Illustration of incremental sc4DVAR. (Diagram template: ECMWF presentation by Sebastien Lafont)

Figure 2.2 illustrates the incremental sc4DVAR algorithm. The initial guess to start the algorithm is $x_0 = x_b$, which is then used to evaluate the non-linear objective function $\mathcal{J}$. Evaluating the 'outer-loop' objective function, $\mathcal{J}$, yields the non-linear model trajectory and 'departures', as seen in Figure 2.2, which allows the linearised inner-loop to begin. The initial guess for the inner-loop objective function is $\delta x_i = 0$, then the iterative minimisation algorithm will solve using the linearised inner-loop objective function J and its gradient ∇J to provide the new δx_i increment which is added on to the previous guess x_i. This process is then repeated again until the desired convergence criterion is reached.

The Gauss-Newton approach detailed here is equivalent to solving the equations arising from the gradient equation (2.9), [52]. However, solving the gradient equation is not practicable operationally since it is deemed too computationally expensive, so we do not consider it in this thesis. In operational NWP most of the computational cost is associated with the minimisation of (2.13), [74]. The

ECMWF has the dominant super-computing capability in the NWP community and they perform ~ 50 inner-loop iterations with only ~ 3 outer-loop iterations.

The sc4DVAR problem is known to be ill-conditioned mainly due to the correlations in the background error covariance matrix B_0, [43], [41]. The matrix B_0 is also known to be very large due to the number of variables in the sc4DVAR problem, [4]. We now introduce a technique which is operationally used to deal with the background error covariance matrix.

2.2.2 The Control Variable Transform

The Control Variable Transform (CVT) technique has traditionally been used to deal with the ill-conditioning of the B_0 matrix in variational data assimilation, [58]. More recently the Met Office has utilised this technique to implement their hybrid 4DVAR and hybrid 4DEnVAR techniques, [60]. A change of variables is introduced which allows for the implicit treatment of B_0, therefore alleviating the need to store an explicit inverse of B_0. The two principal reasons for this transform are; the B_0 is too large to store or express explicitly, and it is known to be too ill-conditioned to find and represent its explicit inverse, [4]. We now discuss the CVT technique.

We introduce a change of variables such that

$$x_0^{(k)} = Uz^{(k)}, \tag{2.16}$$

where this change of variables also applies to increments defined in (2.11) and (2.14). Therefore (2.13) becomes

$$\hat{J}(\delta z^{(k)}) = \frac{1}{2}||\delta z^{(k)} - (z^b - z^{(k)})||^2_{U^T B_0^{-1} U} + \frac{1}{2}\sum_{i=0}^{n} ||H_i M_{i,0} U \delta z^{(k)} - d_i||^2_{R_i^{-1}}. \tag{2.17}$$

The ideal U-transform to aid in the conditioning of (2.13) is such that

$$U^T B_0^{-1} U = I. \tag{2.18}$$

In terms of data assimilation, equation (2.18) implies that in z co-ordinates the errors in elements of the background state vector are uncorrelated with each other and have variance equal to one. Solving (2.17) is equivalent to solving (2.13) as long as (2.18) holds. From (2.18) we require

$$B_0 = UU^T, \tag{2.19}$$

to hold. In practice U does not necessarily have to be square. The challenge is to find U and its adjoint U^T to be an optimum representation of B_0. Obtaining transforms for B_0 is an extensive area of current research, [4], which is not the focus of this thesis. We assume U is the unique *symmetric-square root* of B_0 in this thesis and thus $U = B_0^{1/2}$.

Therefore (2.17) becomes

$$\hat{J}(\delta z^{(k)}) = \frac{1}{2}\left(\delta z^{(k)} - (z^b - z^{(k)})\right)^T \left(\delta z^{(k)} - (z^b - z^{(k)})\right)^T \tag{2.20}$$

$$+ \frac{1}{2}\sum_{i=0}^{n}(H_i M_{i,0} B_0^{1/2}\delta z^{(k)} - d_i)^T R_i^{-1}(H_i M_{i,0} B_0^{1/2}\delta z^{(k)} - d_i), \tag{2.21}$$

with Hessian

$$\nabla^2 \hat{J}(\delta z) = I + \sum_{i=0}^{n} B_0^{1/2} M_{0,i}^T H_i^T R_i^{-1} H_i M_{i,0} B_0^{1/2}. \tag{2.22}$$

A paper by E.Andersson et al. [1] found the conditioning of (2.22) on a 2-grid point example, with q observations at each grid point to be

$$\kappa(\nabla^2 \hat{J}(\delta z)) = 2q\frac{\sigma_b^2}{\sigma_o^2} + 1, \tag{2.23}$$

where κ denotes the condition number of the preconditioned Hessian in the 2-norm. The two grid points are assumed to be close in proximity and therefore highly correlated. This suggests that for dense observations the conditioning of the system is dependent on the ratio of the background to observation errors.

The preconditioned Hessian matrix (2.22) has its smallest eigenvalue equal to one provided H_i is not full rank, which is true for most applications of data assimilation, especially in NWP. The preconditioned Hessian of sc4DVAR has been investigated more in-depth for more general cases in [41], [43]. The authors derive bounds on the condition number of (2.22) and showed that the convergence rate is much improved using $B_0^{1/2}$ as a preconditioner.

In the next section we introduce the two wc4DVAR formulations at the focus of the thesis.

2.3 Weak-Constraint 4DVAR

The weak-constraint problem arises from relaxing the perfect model assumption (2.1) allowing for model error. This implies the model is enforced as a *weak-constraint* and the control variable has now increased by an order of magnitude as we will see shortly. We revisit (2.1) now and find

$$x_i - \mathcal{M}_{i,i-1}(x_{i-1}) = \eta_i, \tag{2.24}$$

for $i = 1, ..., n$, where $\eta_i \in \mathbb{R}^N$, represents the model error. We assume the model errors are random with zero mean, Gaussian error statistics and a known covariance such that

$$\eta_i \sim N(0, Q_i), \tag{2.25}$$

for $i = 1, ..., n$, where $Q_i \in \mathbb{R}^{N \times N}$ represents the model error covariance matrix. We also assume that model errors are independent of the background and observation errors.

The additional model error now becomes a quantity for consideration and thus is incorporated into the objective function. One way of writing the objective function is in terms of the initial conditions x_0 and model errors η_i, such that

$$\begin{aligned} \underset{(x_0,\eta_1,...,\eta_n)}{min} \mathcal{J}(x_0, \eta_1, ..., \eta_n) = &\frac{1}{2}(x_0 - x_0^b)^T B_0^{-1}(x_0 - x_0^b) \\ &+ \frac{1}{2}\sum_{i=0}^{n}(\mathcal{H}_i(x_i) - y_i)^T R_i^{-1}(\mathcal{H}_i(x_i) - y_i) \\ &+ \frac{1}{2}\sum_{i=1}^{n}\eta_i^T Q_i^{-1}\eta_i, \end{aligned} \tag{2.26}$$

subject to the *weak model constraint* (2.24) .

The objective is to minimise the errors in the initial state, observations and the model by selecting the most appropriate initial condition and model error estimates. This formulation is more common in the literature than the alternative, implemented mainly on non-operational models, [94], [83], [84], [93]. An operational implementation of this formulation was functioning at the ECMWF, [56], until it was taken offline recently due to numerical conditioning issues.

Another way to consider the problem is in terms of the states x_i such that

$$\begin{aligned}\underset{(x_0,...,x_n)}{min} \mathcal{J}(x_0,...,x_n) = &\frac{1}{2}(x_0 - x_0^b)^T B_0^{-1}(x_0 - x_0^b)\\ &+ \frac{1}{2}\sum_{i=0}^{n}(\mathcal{H}_i(x_i) - y_i)^T R_i^{-1}(\mathcal{H}_i(x_i) - y_i)\\ &+ \frac{1}{2}\sum_{i=1}^{n}(x_i - \mathcal{M}_{i,i-1}(x_{i-1}))^T Q_i^{-1}(x_i - \mathcal{M}_{i,i-1}(x_{i-1})), \quad (2.27)\end{aligned}$$

where the constraint (2.24) is incorporated into the objective function. This formulation is not as common as (2.26) in the literature although there are some recent research contributions linking this formulation with particle filter and hybrid methods. In [2], the author studies the connection of (2.27) with implicit particle filters and demonstrates this using the Lorenz 63 model [61]. Another more recent paper shows the connection of (2.27) with 4DEnVAR and even proposes preconditioning strategies for the problem, [19]. An important feature of (2.27) is the potential for the resulting algorithm to be parallelised since NWP centres are constrained by computing power and time needed to produce a forecast, [27], we show why this is in Section .

In sc4DVAR the initial conditions alone could utilise the non-linear model (2.1) to produce an entire trajectory. By introducing model error the problem has become *fully 4-dimensional*. The forward model now requires initial conditions and additional forcing terms defined at each time step to obtain the states, as can be seen from equation (2.24).

We now introduce 4D notation as in [27]. We define the 4D state and model error vectors (respectively) as follows

$$\mathbf{p} = \begin{pmatrix} x_0 \\ \eta_1 \\ \vdots \\ \eta_n \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}. \quad (2.28)$$

Similarly the previous guess for the initial conditions and model errors produces a similar vector to $\mathbf{p}$, denoted as $\mathbf{p}^b \in \mathbb{R}^{N(n+1)}$, where the 'b' superscript denotes the background. We define the 4D model operator, $\boldsymbol{\mathcal{L}} : \mathbb{R}^{N(n+1)} \to \mathbb{R}^{N(n+1)}$ which enables us to map from 'state space' to 'model error space' such that

$$\boldsymbol{\mathcal{L}}(\mathbf{x}) = \mathbf{p}. \tag{2.29}$$

We can think of (2.29) as a 4D representation of (2.24), which links the two vectors $\mathbf{p}$ and $\mathbf{x}$ via (2.29). The operator $\boldsymbol{\mathcal{L}}$ is invertible, since we can determine $\mathbf{x}$ from $\mathbf{p}$ using (2.24).

We now define the following 4D spatial-temporal variables,

$$\mathbf{y} = \begin{pmatrix} y_0 \\ y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \tag{2.30}$$

$$\mathbf{D} = \begin{pmatrix} B_0 & & & \\ & Q_1 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & Q_n \end{pmatrix}, \mathbf{R} = \begin{pmatrix} R_0 & & & \\ & R_1 & & \\ & & \cdot & \\ & & & \cdot \\ & & & & R_n \end{pmatrix}. \tag{2.31}$$

We notice a few subtleties here. We have composed $\mathbf{D} \in \mathbb{R}^{N(n+1)\times N(n+1)}$ such that there are *no temporal correlations* between the initial conditions and model errors. This also applies to the observation error covariance matrix $\mathbf{R} \in \mathbb{R}^{p(n+1)\times p(n+1)}$ which is also assumed to be temporally uncorrelated.

We can now write the wc4DVAR objective function (2.26) in 4D form

$$\underset{\mathbf{p}}{min}\ \mathcal{J}(\mathbf{p}) = \frac{1}{2}||\mathbf{p} - \mathbf{p}^b||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p})) - \mathbf{y}||^2_{\mathbf{R}^{-1}}, \tag{2.32}$$

where $\boldsymbol{\mathcal{H}}$ is the 4D non-linear observation operator. The alternative formulation, (2.27), is as follows

$$\underset{\mathbf{x}}{min}\ \mathcal{J}(\mathbf{x}) = \frac{1}{2}||\boldsymbol{\mathcal{L}}(\mathbf{x}) - \mathbf{p}^b||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\boldsymbol{\mathcal{H}}(\mathbf{x}) - \mathbf{y}||^2_{\mathbf{R}^{-1}}. \tag{2.33}$$

Differentiating (2.32) yields

$$\nabla\mathcal{J}(\mathbf{p}) = \mathbf{D}^{-1}(\mathbf{p} - \mathbf{p}^b) + (\mathbf{H}_{\mathbf{x}}\mathbf{L}_{\mathbf{x}}^{-1})^T\mathbf{R}^{-1}(\boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p})) - \mathbf{y}), \tag{2.34}$$

where $\mathbf{H}_{\mathbf{x}}$ and $\mathbf{L}_{\mathbf{x}}^{-1}$ are Jacobians, linearised around the subscripted quantity. Similarly, by differentiating (2.33) we have

$$\nabla\mathcal{J}(\mathbf{x}) = \mathbf{L}_{\mathbf{x}}^T\mathbf{D}^{-1}(\boldsymbol{\mathcal{L}}(\mathbf{x}) - \mathbf{p}^b) + \mathbf{H}_{\mathbf{x}}^T\mathbf{R}^{-1}(\boldsymbol{\mathcal{H}}(\mathbf{x}) - \mathbf{y}). \tag{2.35}$$

The linearisation points in the subscripts of $\mathbf{H}$ and $\mathbf{L}$ are omitted herein since this is not the focus of the thesis. The different gradients (2.34) and (2.35) suggest that the minimisation characteristics of (2.32) and (2.33) will be different.

To be clear on the definition of each term in the gradients above, we write the operators $\mathbf{L}$ and $\mathbf{H}$ in matrix form

$$\mathbf{H} = \begin{pmatrix} H_0 & & & \\ & H_1 & & \\ & & \ddots & \\ & & & H_n \end{pmatrix}, \ \mathbf{L} = \begin{pmatrix} I & & & & \\ -M_{1,0} & I & & & \\ & -M_{2,1} & \ddots & & \\ & & \ddots & & \\ & & & -M_{n,n-1} & I \end{pmatrix}. \tag{2.36}$$

The inverse of $\mathbf{L}$ can be obtained from the weak-constraint equation (2.24), thus taking the following form

$$\mathbf{L}^{-1} = \begin{pmatrix} I & & & & & \\ M_{1,0} & I & & & & \\ M_{2,0} & M_{2,1} & I & & & \\ M_{3,0} & M_{3,1} & M_{3,2} & I & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ M_{n,0} & M_{n,1} & \dots & \dots & M_{n,n-1} & I \end{pmatrix}. \tag{2.37}$$

The linearised forward model of $\mathcal{M}$ is denoted by M, which is embedded in the operator $\mathbf{L}$. The adjoint operators are $\mathbf{L}^T$ and $\mathbf{L}^{-T}$, which have the linearised adjoint model M^T within them. We notice that $\mathbf{L}^{-1}$ is a lower triangular matrix meaning all its eigenvalues lie on its main diagonal, which all equal 1.

The Hessians of (2.32) and (2.33) are as follows,

$$\mathbf{S}_p = \nabla^2 \mathcal{J}(\mathbf{p}) = \mathbf{D}^{-1} + \mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}, \tag{2.38}$$

and

$$\mathbf{S}_x = \nabla^2 \mathcal{J}(\mathbf{x}) = \mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \tag{2.39}$$

We can already see at this point that the alternate minimimsation problems (2.32) and (2.33) are quite different, leading to different gradients and Hessians. Therefore it is natural to expect differences in their respective minimisation characteristics. Let us now examine the structure of the Hessians of $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$.

2.3.1 The Weak-Constraint 4DVAR Hessians

The Hessians are important since they provide information on the local curvature of the objective function. The structure of the Hessians give us insights into how each wc4DVAR formulation iteratively achieves its solution, as seen in (2.52) and (2.53).

We now illustrate the structure of the Hessian of $\mathcal{J}(\mathbf{p})$,

$$\mathbf{S}_p = \begin{pmatrix} B_0^{-1} & & & \\ & Q_1^{-1} & & \\ & & \ddots & \\ & & & Q_n^{-1} \end{pmatrix} +$$

$$\begin{pmatrix} \sum\limits_{i=0}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=1}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=2}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,2} & \dots & (H_nM_{n,0})^TR_n^{-1}H_n \\ \sum\limits_{i=1}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=1}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=2}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,2} & \dots & (H_nM_{n,1})^TR_n^{-1}H_n \\ \sum\limits_{i=2}^{n}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=2}^{n}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=0}^{n-2}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & (H_nM_{n,n-1})^TR_n^{-1}H_n \\ H_n^TR_n^{-1}H_nM_{n,0} & H_n^TR_n^{-1}H_nM_{n,1} & \dots & H_n^TR_n^{-1}H_nM_{n,n-1} & H_n^TR_n^{-1}H_n \end{pmatrix}. \tag{2.40}$$

The $\mathbf{S}_p$ structure is full block where each block is quite sparse in practice due to the observation operator having much lower dimension than the state.

The Hessian of $\mathcal{J}(\mathbf{x})$ possesses a block tri-diagonal structure,

$$\mathbf{S}_x = \begin{pmatrix} B_0^{-1}+M_1^TQ_1^{-1}M_1 & -M_1^TQ_1^{-1} & & & \\ -Q_1^{-1}M_1 & Q_1^{-1}+M_2^TQ_2^{-1}M_2 & -M_2^TQ_2^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & & \\ & & -Q_{n-1}^{-1}M_{n-1} & Q_{n-1}^{-1}+M_n^TQ_n^{-1}M_n & -M_n^TQ_n^{-1} \\ & & & -Q_n^{-1}M_n & Q_n^{-1} \end{pmatrix} +$$

$$\begin{pmatrix} H_0^TR_0^{-1}H_0 & & & \\ & H_1^TR_1^{-1}H_1 & & \\ & & \ddots & \\ & & & H_n^TR_n^{-1}H_n \end{pmatrix}. \tag{2.41}$$

These Hessians are both *symmetric positive-definite* matrices implying they possess a unique inverse. It is important to note that the Hessians of the incremental formulations (2.46) and (2.49) are identical to these first-order Hessians provided the linearisation state used to obtain these first-order Hessians is close to the solution of the non-linear objective functions. So our work in this thesis is relevant to both problems. We also notice that the Hessian of sc4DVAR, (2.8), is contained within (2.40), such that $S = \mathbf{S}_{p(1,1)}$.

The parallelism of (2.27) over (2.26) can be seen in the Hessian matrices (2.41) and (2.40) respectively. The separate blocks of (2.41) can be calculated much quicker than (2.40) since each block in (2.40) requires sequential model integration. Each single time-step block seen $\mathbf{S}_x$ can be allocated to a single processor, and with enough processors to cover each block in $\mathbf{S}_x$, the calculation can be obtained much quicker than $\mathbf{S}_p$. Each block in $\mathbf{S}_p$ requires the entire string of model time-step integrations to be completed, which in operational NWP can take a while.

We have discussed the structural differences in the Hessians of (2.32) and (2.33) in this section. We now introduce the Gauss-Newton incremental formulation of the weak-constraint problem.

2.3.2 Incremental Weak-Constraint 4DVAR

In this section we extend the Gauss-Newton incremental 4DVAR approach shown in Section 2.2.1 to the weak-constraint problem.

We derive the incremental formulation by defining an increment in $\mathbf{p}$ such that

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \delta\mathbf{p}^{(k)}. \tag{2.42}$$

We seek to re-write (2.32) in terms of the increment, $\delta\mathbf{p}^{(k)}$. Before doing so we approximate the non-linear operators to first-order

$$\begin{aligned}\boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k+1)})) &= \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k)} + \delta\mathbf{p}^{(k)})),\\ &\approx \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k)})) + (\boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k)})))'\delta\mathbf{p}^{(k)},\\ &= \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k)})) + \mathbf{H_x}\mathbf{L_x}^{-1}\delta\mathbf{p}^{(k)}. \end{aligned} \tag{2.43}$$

We also define

$$\mathbf{b^p} = \mathbf{p}^b - \mathbf{p}^{(k)}, \tag{2.44}$$

$$\mathbf{d^p} = \mathbf{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}^{(k)})), \tag{2.45}$$

where the superscripts denote the relevant formulation variable. We substitute (2.43), (2.44) and (2.45) into the non-linear objective function (2.32) giving us the incremental wc4DVAR inner-loop '$\delta\mathbf{p}$' function

$$\underset{\delta\mathbf{p}^{(k)}}{min}\ J(\delta\mathbf{p}^{(k)}) = \frac{1}{2}||\delta\mathbf{p}^{(k)} - \mathbf{b}^{\mathbf{p}}||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\mathbf{H}_{\mathbf{x}}\mathbf{L}_{\mathbf{x}}^{-1}\delta\mathbf{p}^{(k)} - \mathbf{d}^{\mathbf{p}}||^2_{\mathbf{R}^{-1}}, \tag{2.46}$$

which is now a quadratic function in $\delta\mathbf{p}^{(k)}$. Since all the operators have been linearised as in (2.43), the constraint (2.29) becomes

$$\mathbf{L}_{\mathbf{x}^{(k)}}\delta\mathbf{x}^{(k)} = \delta\mathbf{p}^{(k)}. \tag{2.47}$$

Solving the inner loop problem yields a new $\delta\mathbf{p}^{(k)}$ increment to update the old $\mathbf{p}^{(k)}$ as in (2.42).

We derive the incremental formulation for (2.33) in a similar fashion to (2.46) by approximating $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{L}}$ as in (2.43) and defining increment in $\mathbf{x}$ such that

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}, \tag{2.48}$$

similar to (2.42). We can now write an incremental $\delta\mathbf{x}$ formulation such that

$$\underset{\delta\mathbf{x}^{(k)}}{min}\ J(\delta\mathbf{x}^{(k)}) = \frac{1}{2}||\mathbf{L}_{\mathbf{x}}\delta\mathbf{x}^{(k)} - \mathbf{b}^{\mathbf{x}}||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\mathbf{H}_{\mathbf{x}}\delta\mathbf{x}^{(k)} - \mathbf{d}^{\mathbf{x}}||^2_{\mathbf{R}^{-1}}, \tag{2.49}$$

where

$$\mathbf{b}^{\mathbf{x}} = \mathbf{p}^{b} - \boldsymbol{\mathcal{L}}(\mathbf{x}^{(k)}), \tag{2.50}$$

$$\mathbf{d}^{\mathbf{x}} = \mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}^{(k)}). \tag{2.51}$$

Figure 2.3 illustrates the algorithmic schematic of wc4DVAR incremental formulation (2.46).

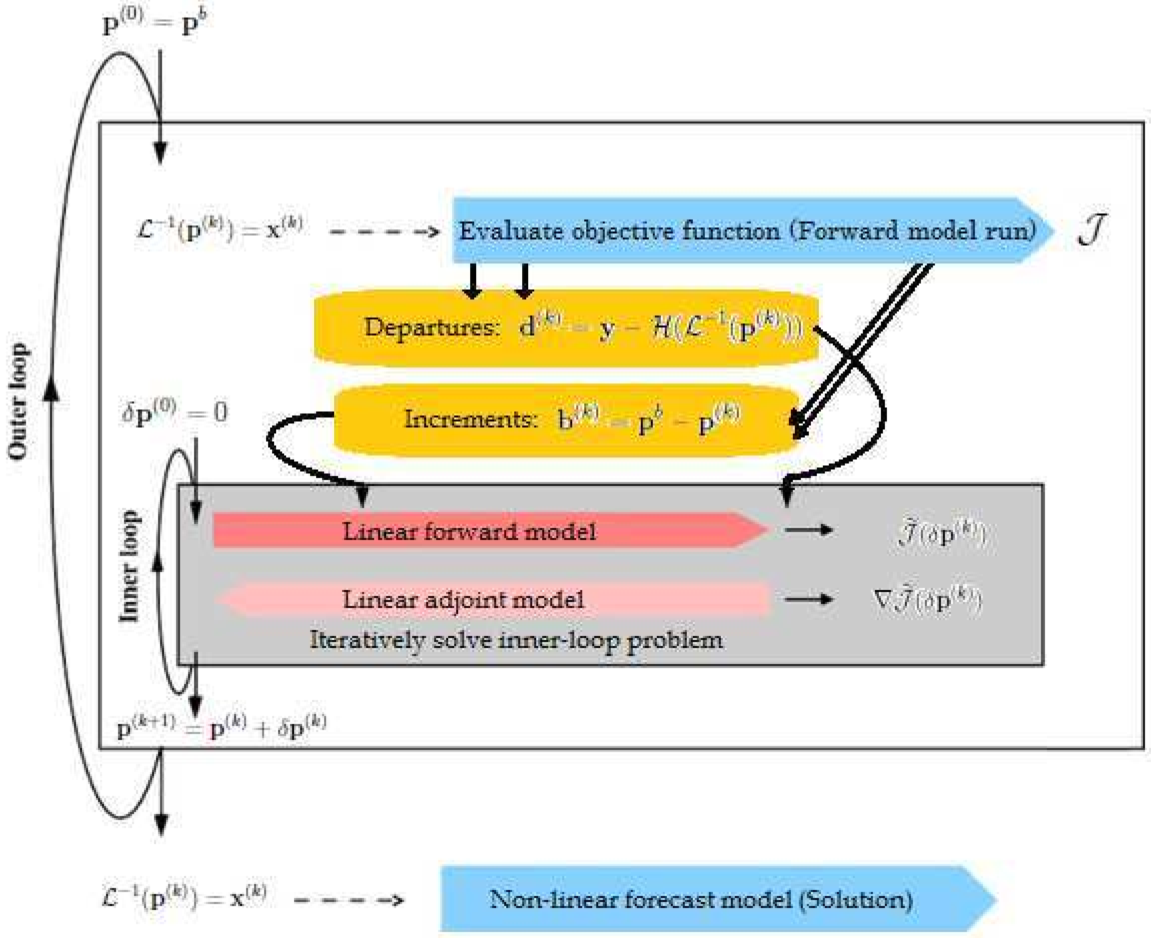

$\mathbf{p}^{(0)} = \mathbf{p}^b$
$\mathcal{L}^{-1}(\mathbf{p}^{(k)}) = \mathbf{x}^{(k)}$
Evaluate objective function (Forward model run)
$\mathcal{J}$
Departures: $\mathbf{d}^{(k)} = \mathbf{y} - \mathcal{H}(\mathcal{L}^{-1}(\mathbf{p}^{(k)}))$
Outer loop
$\delta\mathbf{p}^{(0)} = 0$
Increments: $\mathbf{b}^{(k)} = \mathbf{p}^b - \mathbf{p}^{(k)}$
Inner loop
Linear forward model
$\tilde{\mathcal{J}}(\delta\mathbf{p}^{(k)})$
Linear adjoint model
$\nabla\tilde{\mathcal{J}}(\delta\mathbf{p}^{(k)})$
Iteratively solve inner-loop problem
$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} + \delta\mathbf{p}^{(k)}$
$\mathcal{L}^{-1}(\mathbf{p}^{(k)}) = \mathbf{x}^{(k)}$
Non-linear forecast model (Solution)

Figure 2.3: Illustration of Weak-Constraint Incremental 4DVAR, $\delta\mathbf{p}$ formulation. (Diagram template courtesy of ECMWF presentation by Sebastien Lafont)

We notice how this algorithm is very similar to the incremental sc4DVAR algorithm shown in Figure 2.2. The same concept applies, except now the variables are much larger and represent both spatial and temporal information.

The algebraic linear system when the gradient is equal to zero is analogous to the sc4DVAR inner-loop gradient equation (2.15), for each of the incremental inner-loop objective functions (2.46) and (2.49) is thus

$$\left(\mathbf{D}^{-1} + \mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}\right)\delta\mathbf{p}^{(k)} = \mathbf{D}^{-1}\mathbf{b}^{\mathbf{p}} + \mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d}^{\mathbf{p}}, \tag{2.52}$$

$$\left(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\delta\mathbf{x}^{(k)} = \mathbf{L}^T\mathbf{D}^{-1}\mathbf{b}^{\mathbf{x}} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{d}^{\mathbf{x}}, \tag{2.53}$$

respectively. Solving (2.52) and (2.53) is equivalent to minimising (2.32) and (2.33) respectively. Solving the gradient equations and minimising the objective function directly have been shown to be equivalent for sc4DVAR, [52], we believe this is also true for wc4DVAR but this has yet to be proven. We see that the left hand side of both equations (2.52), (2.53) are the respective Hessians $\mathbf{S}_p$, $\mathbf{S}_x$, and the right-hand is the initial guess. The emphasis on the gradients and hence the Hessians $\mathbf{S}_p$ and $\mathbf{S}_x$ can be seen from these gradient equations.

In this section we have introduced the incremental wc4DVAR technique for both wc4DVAR formulations theoretically. We also emphasised the role of the Hessian in minimising the wc4DVAR problem. In the next section we introduce the CVT technique in the context of preconditioning wc4DVAR.

2.3.3 Preconditioning Weak-Constraint 4DVAR

The weak-constraint problem is a much larger problem then sc4DVAR since the matrix $\mathbf{D}$ encompasses B_0 and Q_i (for $i = 1, .., n$). At the time of writing this thesis there has been no real progress in preconditioning the $\mathcal{J}(\mathbf{x})$ formulation, but rather an alternative saddle-point formulation has been suggested by Fisher et al. [27]. The authors suggest preconditioning by finding a suitable low-cost approximation to $\mathbf{L}$, with some experiments to show minor improvements. We do not pursue the preconditioning of the $\mathcal{J}(\mathbf{x})$ formulation in this thesis. We now introduce the method we use to precondition the $\mathcal{J}(\mathbf{p})$ formulation using the $\mathbf{D}$ matrix.

We introduce a change of variables with the intention of alleviating ill-conditioning in (2.46) arising from $\mathbf{D}$,

$$\mathbf{p} = \mathbf{U}\mathbf{z}, \tag{2.54}$$

where this change of variables also applies to the background term and the increment (2.42) such that

$$\mathbf{p}^b = \mathbf{U}\mathbf{z}^b, \tag{2.55}$$

$$\delta\mathbf{p}^{(k)} = \mathbf{U}\delta\mathbf{z}^{(k)}. \tag{2.56}$$

Substituting (2.55) and (2.56) into (2.46) yields the following objective function

$$\underset{\delta\mathbf{z}^{(k)}}{min}\ \hat{J}(\delta\mathbf{z}^{(k)}) = \frac{1}{2}||\delta\mathbf{z}^{(k)} - (\mathbf{z}^b - \mathbf{z}^{(k)})||^2_{\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U}} + \frac{1}{2}||\mathbf{H}\mathbf{L}^{-1}\mathbf{U}\delta\mathbf{z}^{(k)} - \mathbf{d}||^2_{\mathbf{R}^{-1}}, \quad (2.57)$$

where ideal **U**-transform is such that

$$\mathbf{U}^T\mathbf{D}^{-1}\mathbf{U} = \mathbf{I}. \quad (2.58)$$

If a poor choice of **U** was chosen, the preconditioning would be inadequate and the iterative solver used to treat the wc4DVAR problem will not see an improvement in convergence rate. In practice the B_0 matrix which constitutes part of **D**, is obtained using various filtering techniques, [4]. The same methodology can be applied to the Q_i matrices inside **D**, but this has not had much research attention as of yet. We assume that **U** is the unique *symmetric square root* of **D** in this thesis.

So (2.57) becomes

$$\underset{\delta\mathbf{z}^{(k)}}{min}\ \hat{J}(\delta\mathbf{z}^{(k)}) = \frac{1}{2}||\delta\mathbf{z}^{(k)} - (\mathbf{z}^b - \mathbf{z}^{(k)})||^2_{\mathbf{I}} + \frac{1}{2}||\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2}\delta\mathbf{z}^{(k)} - \mathbf{d}^{\mathbf{p}}||^2_{\mathbf{R}^{-1}}. \quad (2.59)$$

Solving (2.59) is equivalent to solving (2.46) as long as (2.58) holds. The first order preconditioned Hessian of (2.59) is therefore

$$\nabla^2\hat{J}(\delta\mathbf{z}) = \hat{\mathbf{S}}_p = \mathbf{I} + \mathbf{D}^{1/2}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2}, \quad (2.60)$$

where **I** is the identity matrix of size $N(n+1) \times N(n+1)$.

We now detail the algorithm for solving (2.59)

Algorithm 2.1 Incremental Preconditioned Weak-Constraint 4DVAR $\hat{J}(\delta\mathbf{z}^{(k)})$

```
1: Initial guess $\delta\mathbf{p}^{(0)} = 0$.
2: Calculate innovation $\mathbf{d}^{\mathbf{p}}$ (2.45) using the full non-linear model.
3: Calculate $\delta\mathbf{z}$ via (2.56) using CVT.
4: Minimise $\hat{J}$ via (2.59) to obtain new $\delta\mathbf{z}$.
5: Update new increment $\delta\mathbf{p}^{(k)}$ via (2.56).
6: Update current outer-loop estimate $\mathbf{p}^{(k+1)}$ via (2.42).
7: Repeat steps 2 to 7 until desired iterative termination criterion (tolerance) is
   reached.
```

In this section we have introduced the method of preconditioning wc4DVAR (2.32) using the CVT technique, which is essential for wc4DVAR to be considered practicable operationally. This naturally extends from concepts used to implement sc4DVAR.

We have introduced the two wc4DVAR formulations at the focus of this thesis and briefly highlighted differences in the minimisation problems that ensue just by viewing the different gradients and Hessians. We have also extended the theory of the incremental and CVT techniques from sc4DVAR to wc4DVAR. We now discuss the literature around the wc4DVAR problem both in its application and any relevant research related to the conditioning of the problem.

2.4 Literature Review

This chapter so far has been dedicated to introducing all the background material relevant to the work in this thesis.

We review the current literature in this section, with the intention of placing the research in this thesis adequately within the current body of research. This section is divided into two parts. We summarise the relevant literature with regards to the application of wc4DVAR, mainly the model error estimation formulation, in the first part. The second part reviews the literature more relevant to the subject of the thesis namely the conditioning of the wc4DVAR problem.

2.4.1 Applications Of Weak-Constraint 4DVAR

The sc4DVAR problem has had more time under research focus than wc4DVAR since it became operationally viable in the early 90's, [45], [38], [39], [18]. This can be seen as a necessary stepping stone required to begin to understand the weak-constraint problem, since the sc4DVAR is just a simplification of wc4DVAR, by assuming the model is perfect. There have been numerous suggestions in the literature that wc4DVAR holds an advantage over the sc4DVAR, [84], [16], [17], which we will now discuss. It is important to note that the weak-constraint formulation considered in the *majority* of the literature refers to the $\mathcal{J}(\mathbf{p})$ formulation.

A study by Zupanski [94] examined the application of the both wc4DVAR and sc4DVAR on the regional National Centre for Environmental Prediction (NCEP) model. The author highlights that in the presence of model error, the sc4DVAR method provides a solution with incorrect initial conditions since it attempts to correct errors while enforcing the constraint of a perfect model. However wc4DVAR will average these errors out across the assimilation window yielding state estimates that are more inline with the truth. This means that the solution increment for the initial conditions from wc4DVAR is not as severe as sc4DVAR. She concludes that there is a need for considering wc4DVAR over the sc4DVAR. She also concedes that wc4DVAR is computationally expensive and ill-conditioned, and proposes looking at the lower-dimensional observation-space dual formulation of the problem.

A climate application of wc4DVAR in Korea using satellite data for heavy rainfall simulation was documented in [54]. The authors detail a study where they use both sc4DVAR and wc4DVAR and they clearly show that wc4DVAR provided much improved initial conditions for their model compared to sc4DVAR.

In 2004, Vidard et al. showed that wc4DVAR gives a marked improvement over sc4DVAR when applied to a non-linear one-layer two-dimensional shallow water model, [86]. The model error in this case was a systematic bias, but nevertheless it does serve as a good guide for a more complex setting. The authors conclude that the weak-constraint formulation provides a better solution both over the assimilation window and in the forecast phase.

An article by Lindskog et al. [56] details the implementation of the weak-constraint model error formulation to correct for known biases in the upper stratosphere on the ECMWF operational system. The paper highlights potential issues from a more practical perspective, but this often provides well-informed directions for the requirement of theoretical understanding. They conclude that careful consideration is required when specifying the model error covariance matrix Q, and 'understanding the role of balance descriptions'. By balance descriptions they are referring to the combination of the model error correlations with the weighting functions of the satellite radiance measurements in the upper stratosphere. In other words, the specification of the model error against the observation error correlations is important. This is a *conditioning issue*, which we investigate in this thesis. They also allude to the requirement of considering time-correlated model error, which is a more complex problem for the future. Even with all the issues they have highlighted in their paper, they do show that the overall solution provided by wc4DVAR is less spurious than sc4DVAR, and the solution increments are not as severe. This is interpreted as an improvement over the current sc4DVAR.

A paper on the equivalence of the Kalman-smoother (KS) to the wc4DVAR problem, [26], is motivational with regards to developing the weak-constraint problem. Fisher et al. show that the solution of the Kalman-filter for large time intervals is equivalent to the solution provided by KS at the end of the interval, for linear models. They then show that for a sufficiently long enough assimilation window the solution of KF is identical to the wc4DVAR solution. This suggests that 'wc4DVAR may be a viable algorithm for implementing unapproximated KF equations'. They demonstrate that wc4DVAR gave a similar quality solution to that of the KS through experiments on the Lorenz 95 model. They explain the reason it is not exact is due to the linearisation states of the linearised model. The paper also mentions they have not investigated the *numerical conditioning* of wc4DVAR, more specifically with regards to the assimilation window length and the choice of control variable, ie. the difference between wc4DVAR formulations (2.26) and (2.27). Another similar paper by Desroziers et al. discusses the link between the wc4DVAR state estimation formulation (2.27) and hybrid 4DVAR, [19].

More recently, an article discussing the differences between sc4DVAR and wc4DVAR from a more theoretical perspective was presented by Cullen, [14]. The author compares cycled sc4DVAR to wc4DVAR by simplifying the problem down to a scalar case. He concludes that wc4DVAR must be interpreted as a smoother since it allows the control of error growth throughout the assimilation window. It is shown that where cycled sc4DVAR remains close to the observations, the solution in the scalar case converges to that of a long-window wc4DVAR equivalent. This is true if the regularisation of wc4DVAR, through the Q matrix, is identical to the regularisation of the cycled sc4DVAR method's B matrix at the beginning of each assimilation window cycle.

A. Moore et al. at the University of California discuss their Regional Ocean Modeling System (ROMS) implementation in a lengthy three-part paper, [68], [66] and [67]. The detailed implementation of both the original state-space primal form and lower dimensional observation-space dual form are detailed in [66]. The authors state that wc4DVAR is too large and computationally infeasible when considering the full primal problem. It is suggested that the dual formulation is a sensible step towards an operationally feasible implementation of wc4DVAR. They also discuss methods on error-covariance modeling and suggest preconditioners that have not been fully trialled yet. They conclude that the forecast skill of wc4DVAR is improved over sc4DVAR.

The collective flavour of the literature indicates that wc4DVAR is superior to sc4DVAR. The minimisation problem that ensues from the wc4DVAR approach requires further study, since more degrees of freedom and a larger problem needs careful consideration. Some pieces of literature point in the direction of the dual formulation as a remedy for the size of the problem, [12]. However, we are not concerned with dual problem in this thesis.

A few pieces of literature produced by the ECMWF suggest they are actively developing their implementation of wc4DVAR, [27], [83], [84], [26]. Their intention is to tackle the more practical issues since their operational wc4DVAR implementation detailed in [56] has been put off-line (`https://cimss.ssec.wisc.edu/itwg/itsc/itsc19/program/posters/nwp_3_english.pdf`) due to numerical conditioning issues (conversation with Mike Fisher, ECMWF training course, 2013).

We now review the literature that is more closely related to the conditioning and preconditioning of the wc4DVAR problem.

2.4.2 Conditioning and Preconditioning of Weak-Constraint 4DVAR

At this moment, there are only a few select articles that are directly related to the conditioning or preconditioning of the wc4DVAR problem. They are not related to the study of the condition number, but the areas of research seem to be pointing in the direction of trying to understanding the minimisation process that arises from the wc4DVAR problem.

In [83], the author broadly summarises the variational approaches to the data assimilation problem in the presence of model error. An illustrative example in this paper alludes to the 'Laplacian-like' nature of the first term of $\mathbf{S}_x$ under simplistic assumptions ($M = I$ and $B = Q = I$) and using $\mathbf{Q} = diag\{Q, ..., Q\} = diag\{I, ..., I\}$ to precondition.

$$\mathbf{S}_{\mathbf{x}}^{precond} = \mathbf{A} + \mathbf{Q}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Q}^{1/2}, \tag{2.61}$$

where

$$\mathbf{A} = \begin{pmatrix} 2I & -I & & & \\ -I & 2I & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I & -I \\ & & & -I & I \end{pmatrix}, \tag{2.62}$$

where the other bold-faced matrices are block-diagonal partitions of their own respective matrices similar to $\mathbf{Q}$. If $M \neq I$, then the preconditioner would need to be composed in such a way as to remove the influence of M from the first part of the Hessian $\mathbf{S}_{\mathbf{x}}^{precond}$. This leads into the next part of the research efforts by the ECMWF to find a preconditioner which approximates $\mathbf{L}$ well, since $\mathbf{L}$ contains the model M.

An internal ECMWF report, [27], suggests that the Hessian $\mathbf{S_x}$ is sensitive to the choice of preconditioner. Fisher et al. introduce an alternative saddle-point formulation of the problem. A disadvantage of the saddle-point system to be solved is that it will be *at least* double the size of the weak-constraint problem, which is already considerably larger than the strong-constraint problem. In addition to this the Hessian matrix proposed in the saddle-point formulation is symmetric indefinite. S. Gürol presented encouraging results at the University of Reading DARC series and NASA on the saddle-point wc4DVAR system, which has the advantage of avoiding the inversion of $\mathbf{D}$, [25]. However, the preconditioner is dependent on a good approximation of $\mathbf{L}$, which the authors state is a remaining challenge.

The iterative methods for the primal and dual formulations of the weak-constraint problem have been studied by A. El Akkraoui in her PhD thesis, [20]. She discusses the convergence characteristics of the dual formulation in observation space and finds it is sensitive to the iterative procedure used. She uses a minimum residual approach over the conventional conjugate gradient technique widely used for 4DVAR problems, and shows some improvement. She also investigates the effects of using singular vectors of the Hessian from a previous assimilation window to precondition the Hessian of the following assimilation window, [21].

Previous work on the conditioning of sc4DVAR by Haben et al. [43], [42], [41] increased understanding of the sensitivity of the Hessian condition number to certain aspects of the assimilation. The authors investigated the effects of varying the observation configuration and specifying the accuracy of the observations via the observation variance parameter. The authors also explored the effect of observation thinning on the condition number and the overall solution of the problem on the Met Office operational system. The authors also show that observation thinning and preconditioning indeed provide accelerated convergence rates.

This concludes the literature review. We now summarise this chapter.

2.5 Summary

In this chapter we have introduced the strong-constraint and weak-constraint variational data assimilation problems. We introduced concepts such as the Gauss-Newton incremental approach and the CVT technique for both sc4DVAR and wc4DVAR. We also discussed the structures of the weak-constraint Hessians. This was then followed by a review of the current literature detailing the applications and conditioning of the weak-constraint problem.

We now introduce the mathematical framework required to understand and solve the 4DVAR problem and the necessary tools used to obtain the results in this thesis.

Chapter 3

Mathematical Theory

The variational data assimilation problem is statistical in its formulation but obtaining a solution from the non-linear objective function is an optimisation problem. In this chapter we introduce the necessary material and mathematical tools required to understand and solve the wc4DVAR problems. We remind the reader of the model error formulation,

$$\underset{\mathbf{p}}{min}\ \mathcal{J}(\mathbf{p}) = \frac{1}{2}||\mathbf{p} - \mathbf{p}^b||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p})) - \mathbf{y}||^2_{\mathbf{R}^{-1}}, \tag{3.1}$$

and the state estimation formulation,

$$\underset{\mathbf{x}}{min}\ \mathcal{J}(\mathbf{x}) = \frac{1}{2}||\boldsymbol{\mathcal{L}}(\mathbf{x}) - \mathbf{p}^b||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\boldsymbol{\mathcal{H}}(\mathbf{x}) - \mathbf{y}||^2_{\mathbf{R}^{-1}}. \tag{3.2}$$

We begin by introducing the condition number, followed by the numerical optimisation techniques used to solve wc4DVAR problems (3.1), (3.2). We then detail matrix norm properties required to analyse the condition number of the Hessians of (3.1), (3.2). Finally we introduce the models we use in our data assimilation experiments to put into context the sensitivities of the bounds and their effect on the performance of the optimisation problem.

3.1 Condition Number

The condition number measures sensitivities of the solution to perturbations in the input data. The input data for the data assimilation problem in this thesis is governed by the wc4DVAR objective functionals (3.1), (3.2). We examine the effect of perturbing input data on the wc4DVAR problem in this section to show the importance of the condition number, using a similar argument to that used in [34], (pages 302-304).

We assume the wc4DVAR objective functional has a solution, which we denote as $\mathbf{x}^*$. We then perturb the input data by perturbing $\mathcal{J}$ and denote the perturbed objective function as $\widetilde{\mathcal{J}}$. The perturbed objective function has the solution

$$\hat{\mathbf{x}} = \mathbf{x}^* + h\delta\mathbf{x}, \tag{3.3}$$

where $h = ||\hat{\mathbf{x}} - \mathbf{x}^*||$ and $||\delta\mathbf{x}|| = 1$. We assume that the perturbation in the objective function is small enough to satisfy the following

$$|\widetilde{\mathcal{J}}(\mathbf{x}^*) - \mathcal{J}(\mathbf{x}^*)| \leq |\mathcal{J}(\hat{\mathbf{x}}) - \mathcal{J}(\mathbf{x}^*)| \leq \epsilon. \tag{3.4}$$

The difference in the perturbed and original objective functions at $\mathbf{x}^*$ is assumed to be bounded above by the difference in the original objective functions evaluated at the original solution $\mathbf{x}^*$ and the perturbed solution $\hat{\mathbf{x}}$. We make this assumption to understand some of the factors influencing solution accuracy. We expand $\mathcal{J}$ using the Taylor series

$$\mathcal{J}(\hat{\mathbf{x}}) = \mathcal{J}(\mathbf{x}^* + h\delta\mathbf{x}) = \mathcal{J}(\mathbf{x}^*) + \frac{1}{2}h^2\delta\mathbf{x}^T\nabla^2\mathcal{J}(\mathbf{x}^*)\delta\mathbf{x} + \mathcal{O}(h^3) + \ldots, \tag{3.5}$$

and approximate to second order. Therefore

$$2|\mathcal{J}(\hat{\mathbf{x}}) - \mathcal{J}(\mathbf{x}^*)| \approx ||\hat{\mathbf{x}} - \mathbf{x}^*||^2\delta\mathbf{x}^T\nabla^2\mathcal{J}(\mathbf{x}^*)\delta\mathbf{x}. \tag{3.6}$$

Using $\frac{1}{|\delta\mathbf{x}^T A\delta\mathbf{x}|} \leq \frac{||A^{-1}||}{|\delta\mathbf{x}^T\delta\mathbf{x}|}$, we have

$$||\hat{\mathbf{x}} - \mathbf{x}^*||^2 \approx \frac{2\epsilon\kappa}{||\nabla^2\mathcal{J}(\mathbf{x}^*)||}, \tag{3.7}$$

where we define the condition number as

$$\kappa = ||(\nabla^2 \mathcal{J})^{-1}||.||\nabla^2 \mathcal{J}||. \tag{3.8}$$

We see from the expression (3.7) that the growth of the squared difference of the original and perturbed solutions is proportional to the condition number of the Hessian and the objective function differences. The relationship in (3.7) shows that the condition number of the Hessian is an appropriate measure of the sensitivity of the solution to small perturbations in the input data, and hence the objective function. However, the limitation of this assumption is that the perturbation in the objective function $\mathcal{J}$ must be small enough for (3.4) to hold and for the condition number κ seen in (3.7) to be considered a good measure. Another limitation is that the condition number of the Hessian here is linearised *at* the solution, which is not known in practice.

The specific condition number we use in this thesis is using the 2-norm. Therefore

$$\kappa = \left| \frac{\lambda_{max}(\nabla^2 \mathcal{J})}{\lambda_{min}(\nabla^2 \mathcal{J})} \right|, \tag{3.9}$$

since the first-order Hessians of both wc4DVAR objective functions are symmetric and hence normal.

In this section we have shown and justified our reasoning for using the condition number of the Hessian of the wc4DVAR objective functions as the measure which quantifies the sensitivities of the wc4DVAR objective functions to changes in the input data. We now introduce the numerical optimisation techniques used to solve wc4DVAR in this thesis.

3.2 Numerical Optimisation

This section is dedicated to introducing the iterative gradient techniques used to solve the full non-linear problems (3.1), (3.2) and linearised problems (2.46), (2.49). We begin this section by introducing the popular conjugate gradient method.

3.2.1 The Linear Conjugate Gradient Method

Conjugate gradient methods first appeared in 1952 when Hestenes and Stiefel proposed the idea as an iterative method for solving large linear systems with positive definite coefficient matrices, [44]. Conjugate gradient can be adapted to solve non-linear optimisation problems which we introduce later in Section 3.2.3. The conjugate gradient method can be used both as an algorithm for solving linear systems or an iterative technique for minimising convex quadratic functions such as (2.13), (2.46) and (2.49). We use it in this thesis to solve the inner-loop quadratic problem.

Consider the quadratic problem

$$\underset{x}{min}\ J(x) = \frac{1}{2}x^T Ax - x^T b + c, \tag{3.10}$$

which is identical to the linear incremental problems (2.13), (2.46), (2.49) with x being the control vector and appropriate choices of A, b and c. We set the residual to be the gradient of (3.10) and introduce iterates such that

$$r^{(k)} = Ax^{(k)} - b. \tag{3.11}$$

The linear conjugate gradient (LCG) code used in the work in this thesis is the pre-coded Matlab CG procedure. The LCG algorithm, which we use to solve the incremental formulations (2.46) and (2.49), is as follows

Algorithm 3.1 Linear Conjugate Gradient

```
Counter $k = 0$.
Initial guess $x^{(0)} = 0$, if initial data does not exist,
Set residual $r^{(0)} = Ax^{(0)} - b^{(0)}$,
Set search direction $p^{(0)} = -r^{(0)}$,
While $||r^{(k)}|| > \tau$, where $\tau$ denotes tolerance;
    $\alpha^{(k)} = \frac{(r^{(k)})^T r^{(k)}}{(p^{(k)})^T A p^{(k)}};$   (3.12)
    $x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)};$   (3.13)
    $r^{(k+1)} = r^{(k)} + \alpha^{(k)} A p^{(k)};$   (3.14)
    $\beta^{(k)} = \frac{(r^{(k+1)})^T r^{(k+1)}}{(r^{(k)})^T r^{(k)}};$   (3.15)
    $p^{(k+1)} = -r^{(k+1)} + \beta^{(k)} p^{(k)};$   (3.16)
    $k = k + 1;$
End while.
```

In theory if there are no numerical errors of any kind the CG method will converge in at most N iterations for the sc4DVAR problem, or $N(n+1)$ iterations for the wc4DVAR problem. We state a useful upper bound for the convergence rate of CG,

$$||x^{(k)} - x^*||_A \leq 2||x^{(0)} - x^*||_A \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k , \tag{3.17}$$

where $k = 0$ denotes the initial data and $*$ denotes the solution. The bound shows the dependance of the convergence rate of CG on the condition number of the system of equations being solved, [35].

We now briefly introduce the preconditioned version of CG.

3.2.2 Preconditioned Conjugate Gradient

Preconditioned Conjugate Gradient (PCG) is used to speed up the convergence rate of CG by lowering the condition number of the system being solved. The cost of preconditioning must be cheap and reduce the condition number enough to achieve a considerable reduction in iterates. Let P denote the symmetric positive-definite. The algorithm is as follows

Algorithm 3.2 Preconditioned Conjugate Gradient

```
Counter $k = 0$.
Initial guess $x^{(0)} = 0$, if initial data does not exist,
Set residual $r^{(0)} = Ax^{(0)} - b^{(0)}$,
For the first iteration compute $z^{(0)} = Pr^{(0)}$
Set $p^{(0)} = z^{(0)}$,
While $||r^{(k)}|| > \tau$;
    $\alpha^{(k)} = \frac{(r^{(k)})^T z^{(k)}}{(p^{(k)})^T Ap^{(k)}};$   (3.18)
    $x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)};$   (3.19)
    $r^{(k+1)} = r^{(k)} + \alpha^{(k)} Ap^{(k)};$   (3.20)
    $z^{(k+1)} = Pr^{(k+1)}$   (3.21)
    $\beta^{(k)} = \frac{(r^{(k+1)})^T z^{(k+1)}}{(r^{(k)})^T z^{(k)}};$   (3.22)
    $p^{(k+1)} = -z^{(k+1)} + \beta^{(k)} p^{(k)};$   (3.23)
    Counter $k = k + 1$;
End while.
```

A full discussion of this method can be found in [35]. The preconditioned conjugate gradient technique used in our work is the pre-coded Matlab procedure.

We now introduce a non-linear conjugate gradient technique for iteratively solving the non-linear problem directly.

3.2.3 The Polak-Ribiere Conjugate Gradient Method

We use the Polak-Ribiere CG (PRCG) method as an alternative to the linear CG method in later chapters to demonstrate links between iteratively solving the full non-linear problem and the iterative treatment of the Gauss-Newton approach to the 4DVAR problem.

Fletcher and Reeves extended the linear CG method to non-linear functions by making two simple changes, [28]. Firstly, in the LCG algorithm, line (3.12) requires the replacement of the step length $\alpha^{(k)}$, which minimises $\mathcal{J}$ along the search direction $p^{(k)}$. We require a line search that identifies an approximate minimum of the non-linear function along $p^{(k)}$. Secondly, the residual $r^{(k)}$ must be replaced by the gradient of the non-linear objective function.

There are many variants of the Fletcher-Reeves CG method, mainly differing in the choice of the parameter $\beta^{(k)}$. The PRCG variant defines this parameter as

$$\beta^{(k)} = \frac{(\nabla \mathcal{J}^{(k+1)})^T(\nabla \mathcal{J}^{(k+1)} - \nabla \mathcal{J}^{(k)})}{||\nabla \mathcal{J}^{(k)}||^2}. \tag{3.24}$$

In addition to this, the PRCG method imposes conditions on the step length $\alpha^{(k)}$ to ensure that every step direction $p^{(k)}$ is indeed a descent direction for the function $\mathcal{J}$. These conditions are known as the *strong* Wolfe conditions, [91]. The Wolfe conditions are a set of inequalities that ensure an inexact line search is performed. If these conditions are enforced 'strongly' then the step length, $\alpha^{(k)}$ is forced close to a critical point. These conditions are as follows

$$\mathcal{J}(x^{(k)} + \alpha^{(k)}p^{(k)}) \leq \mathcal{J}(x^{(k)}) + c_1\alpha^{(k)}(\nabla \mathcal{J}(x^{(k)}))^T p^{(k)}, \tag{3.25}$$

$$\nabla \mathcal{J}(x^{(k)} + \alpha^{(k)}p^{(k)})^T p^{(k)} \leq -c_2(\nabla \mathcal{J}(x^{(k)}))^T p^{(k)}, \tag{3.26}$$

where $0 < c_1 < c_2 < \frac{1}{2}$. These techniques are discussed in more depth in [73]. We use this method to solve the wc4DVAR non-linear objective functions (3.1) and (3.2) directly, without the need for inner or outer loops. The PRCG code used in this thesis was obtained from `http://learning.eng.cam.ac.uk/carl/code/minimize/minimize.m`, written by C.E Rasmussen (University of Cambridge).

In an operational NWP setting there is not enough time or computing power to execute the amount of iterations required to solve the problem completely. Therefore an iterative stopping criterion is required. In the next section we briefly discuss the iterative stopping criterion used in our work.

3.2.4 Iterative Stopping Criterion

The purpose of iterative stopping criteria is to enable the user to stop the iterative solver when certain criterion are met, for example when it reaches a certain tolerance or a certain number of iterations. We use an iterative tolerance criterion derived by Lawless et al. in [53] which uses the gradient norm such that

$$\frac{||\nabla\mathcal{J}^{(m)}||_2}{||\nabla\mathcal{J}^{(0)}||_2} < \tau, \tag{3.27}$$

where m is the final iterate. Ideally, if the iterations are making progress the norm of the gradient as the iterates progress should decrease until the final iteration, which should be smaller than the initial gradient-norm. Decreasing the tolerance demands more accurate convergence with respect to the gradient norm.

The authors specified this criterion specifically for the inner-loop incremental 4DVAR objective function to guarantee convergence of the outer-loops. In this thesis we use this iterative stopping criterion for the convergence of both formulations of the inner-loop wc4DVAR functions. In chapter 7 we deviate from the authors intended use of the criterion slightly by using it with the PRCG technique presented in Section 3.2.3.

We now introduce matrix properties, norms and special matrix systems used in the thesis.

3.3 Matrices

We begin by defining 'positive-definiteness'.

Definition 3.3.1 *A matrix $A \in \mathbb{R}^{N \times N}$ is positive-definite if and only if*

$$x^T A x > 0, \tag{3.28}$$

for non-zero $x \in \mathbb{R}^N$.

Furthermore if the matrix A is positive-definite then all the eigenvalues of A are real and if symmetric then the eigenvalues are positive.

Definition 3.3.2 *The eigenvalues of symmetric positive-definite matrix A are solutions of the eigenvalue equation*

$$Av_i = \lambda_i v_i, \tag{3.29}$$

where $\lambda_i \in \mathbb{R}$ is the eigenvalue of A and $v_i \in \mathbb{R}^N$ is the corresponding eigenvector. We write the eigenvalues in order on the interval $\lambda(A) \in [\lambda_1(A), \lambda_N(A)]$ such that

$$\lambda_N(A) > ... > \lambda_k > ... > \lambda_1(A), \tag{3.30}$$

where $\lambda_N = \lambda_{max}$ and $\lambda_1 = \lambda_{min}$.

3.3.1 Norms

Norms permit the concept of a distance or more formally a metric space to be applied to vectors and matrices. We use $||.||$ to denote a vector or matrix norm.

Definition 3.3.3 (See [35], Sec 2.3) *The family of vector p-norms on $\mathbb{R}^N$ is such that*

$$||x||_p = \left(\sum_{i=1}^{N} |x|^p \right)^{\frac{1}{p}}, \tag{3.31}$$

for $x \in \mathbb{R}^N$, $p \geq 1$.

Definition 3.3.4 (See [35], Sec 2.3) *The family of matrix p-norms on $\mathbb{R}^{N\times M}$ is such that*

$$||C||_p = \underset{x\neq 0}{sup}\frac{||Cx||_p}{||x||_p}, \tag{3.32}$$

for $C \in \mathbb{R}^{N\times M}$ and $x \in \mathbb{R}^M$.

In this thesis we use the 1-norm, 2-norm and ∞-norm. For explicit definitions of these norms please refer to [35], Section 2.3.

We now state some useful norm relations which are used in cases where the norms may be difficult to calculate explicitly.

Theorem 3.3.5 (See [3], Sec A.1) *For matrices $A, B \in \mathbb{R}^{N\times N}$ the following statements hold:*

$$||AB|| \leq ||A||\ ||B||, \tag{3.33}$$

$$||A+B|| \leq ||A|| + ||B||. \tag{3.34}$$

The first statement is also known as the Cauchy-Schwarz inequality while the second statement can be derived using the triangle inequality.

Another useful norm equivalence is the one which involves the 1-norm, 2-norm and ∞-norm

Theorem 3.3.6 (See [3], Sec A.1) *For $A \in \mathbb{R}^{N\times N}$ the following inequality holds:*

$$||A||_2 \leq \sqrt{||A||_1||A||_\infty}. \tag{3.35}$$

We now introduce a particular family of matrices with some interesting properties used in our research.

3.3.2 Toeplitz Matrices

We use covariance matrices with a special structure in our research, which fall under a class of matrices known as Toeplitz matrices. So we begin this section by introducing the Toeplitz matrix, which gets its name from the German mathematician Otto Toeplitz. He was the first person to work with Toeplitz operators in 1911, [82]. A Toeplitz matrix is such that

$$T = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \dots & \dots & t_{-(N-1)} \\ t_1 & t_0 & t_{-1} & \ddots & & \vdots \\ t_2 & t_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & t_{-1} & t_{-2} \\ \vdots & & \ddots & t_1 & t_0 & t_{-1} \\ t_{N-1} & \dots & \dots & t_2 & t_1 & t_0 \end{pmatrix}$$

where $T \in \mathbb{R}^{N \times N}$ and the entries $a_{i,j}$ follow the rule $a_{i,j} = a_{i+1,j+1} = a_{i-j}$. Toeplitz matrices are a special case of an even larger family matrices called *persymmetric matrices.*

We are interested in a special type of Toeplitz matrix known as the *circulant* matrix. Circulant matrices are composed of a single row of elements which is permuted periodicly from one row to the next.

Definition 3.3.7 (See [37], Chapter 3) *A circulant matrix $C \in \mathbb{R}^{N \times N}$ takes the following form*

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

.

The matrix is composed of cyclic permutations of the first row. A useful property of a circulant matrix is that the eigenvalues and eigenvectors can be written as Fourier transforms of the top row explicitly. The eigenvalues and eigenvectors of circulant matrices are explicitly known.

Theorem 3.3.8 (See [37], Section 3.1) *The eigenvalues of* C *denoted* $\lambda_m(C) \in \mathbb{C}$ *are such that*

$$\lambda_m(C) = \sum_{k=0}^{N-1} c_k e^{\frac{-2\pi imk}{N}}, \tag{3.36}$$

with corresponding eigenvectors

$$\mathbf{v}_m = \frac{1}{\sqrt{N}}(1, e^{\frac{-2\pi im}{N}}, ..., e^{\frac{-2\pi im(N-1)}{N}})^T, \tag{3.37}$$

for $m = 0, ..., N-1$ *and* $i = \sqrt{-1}$.

Another useful property of circulant matrices is they have a convenient eigendecomposition using Fourier matrices. We formally define the Fourier matrix first.

Definition 3.3.9 *A Fourier matrix* $F \in \mathbb{C}^{N\times N}$ *is such that*

$$F = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \dots & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ \vdots & \omega^2 & \omega^4 & \ddots & \omega^{2(N-1)} \\ \vdots & \dots & \ddots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)^2} \end{pmatrix}$$

where $\omega = e^{\frac{-2\pi i}{N}}$.

A convenient property of Fourier matrices is that they are unitary. Therefore the inverse of a Fourier matrix is equal to its Hermitian matrix,

$$FF^H = I. \tag{3.38}$$

We now state the eigendecomposition of circulant matrices.

Theorem 3.3.10 (See [37]) *Circulant matrices have the following eigendecomposition:*

$$C = F\Lambda_C F^H \tag{3.39}$$

where $\Lambda_C = diag(\lambda_1(C), ..., \lambda_n(C))$ *and* F *is a Fourier matrix as in Definition 3.3.9.*

We conclude this section by highlighting the ease of matrix operations on circulant matrices. Powers and matrix multiplications are conveniently simple due to their eigendecomposition.

Theorem 3.3.11 (See [37]) *The inverse, square root and product of circulant matrices are obtained by taking the inverse, square root or product of* Λ_C *such that*

$$C^{-1} = F\Lambda_C^{-1}F^H, \tag{3.40}$$

$$C^{1/2} = F\Lambda_C^{1/2}F^H, \tag{3.41}$$

$$C_1C_2 = F\Lambda_{C_1}\Lambda_{C_2}F^H. \tag{3.42}$$

In this section we have introduced a particular class of matrix used on numerous occasions throughout the thesis. We now introduce the fundamental theory of covariance matrices since these are very commonly used in NWP data assimilation applications.

3.3.3 Covariance Matrices

The covariance matrix arises from covariances functions. Covariance functions describe the measure of how one variable's statistics effect another. A function $f(x, y)$ of 2 random variables x,y, is the covariance function of a random field $X : \mathbb{R}^N \to \mathbb{R}^N$ if

$$f(x, y) =< X(x)- < X(x) >, X(y)- < X(y) >>, \tag{3.43}$$

for $x, y \in \mathbb{R}^N$. The expected value of a random field is denoted as $<>$. A direct consequence of (3.43) is the function is symmetric $f(x, y) = f(y, x)$. Uncorrelated variables will have covariance equal to zero, which also comes as a consequence of (3.43). Using the same notation, we also understand that $f(x, x)$ is the *variance* of the random variable x and the square root of this is the *standard deviation*. Normalising the covariance with the standard deviations gives us a *correlation function*

$$\rho(x, y) = \frac{f(x, y)}{\sqrt{f(x, x)f(y, y)}}, \tag{3.44}$$

where the diagonal of the correlation function has unit variances. The variances of the variables are assumed to be non-zero so that ρ is well-defined. If there are no correlations between different parameters f is an *auto-covariance* function and ρ an *auto-correlation* function. We assume the errors of the parameter in this thesis are homogeneous. This means that the correlations only depend on the distance between the errors and not their position [4],

$$\rho(x, y) = \hat{\rho}(|x - y|), \tag{3.45}$$

where the distance between x and y is characterised by $\hat{\rho}$. We can verify the validity of a correlation function with the following Theorem.

Theorem 3.3.12 (See [30]) *Let $\hat{\rho}$ be an even continuous function on $\mathbb{R}$ with $\hat{\rho}(0) = 1$ and*

$$\int_{\mathbb{R}} |\hat{\rho}| dx < \infty, \tag{3.46}$$

then $\rho(x, y) = \hat{\rho}(|x - y|)$ is a homogeneous correlation function on $\mathbb{R}$ if and only if the Fourier transform of $\hat{\rho}$ is everywhere non-negative.

Now let us consider a set of correlated points $p_1, p_2, ..., p_N \in \mathbb{R}$ with an auto-correlation function $\rho(x, y)$. We define a positive-definite symmetric auto-correlation matrix $C \in \mathbb{R}^{N \times N}$ such that

$$C_{i,j} = \rho(p_i, p_j), \tag{3.47}$$

for $i, j = 1, ..., N$.

We now discuss the background and model error covariances more specific to the work in this thesis.

3.3.4 Background and Model Error Covariance Matrices

We define the background error covariance matrix such that

$$B = \Sigma_b C_B \Sigma_b, \tag{3.48}$$

where C_B is the background error correlation matrix as in (3.47). The diagonal matrix Σ_b contains the positive background error standard deviations along its diagonal [49], Section 5.4. In this thesis we assume the background variance is equal to σ_b^2 for all variables, from which it follows,

$$B = \sigma_b^2 C_B. \tag{3.49}$$

We also assume the application of this methodology to compose the model error covariance matrix, $Q = \sigma_q^2 C_Q$.

We now introduce two valid correlation functions on the real line. We also discuss details of extending these to a periodic domain.

3.3.4.1 The SOAR Covariance Matrix

The *Second-Order Auto-Regressive* (SOAR) correlation function [15], for points on the real line, separated by a distance $|r|$ is defined by

$$\rho_S(r) = \left(1 + \frac{|r|}{L}\right) exp\left(-\frac{|r|}{L}\right), \tag{3.50}$$

where $r \in \mathbb{R}$ and $L > 0$ denotes the correlation length scale. The SOAR function has been used by the Met Office to model the horizontal correlation of errors in the atmosphere [57], [46]. To enable this function to be a valid correlation function on the real line and on the periodic domain we replace the great circle distance r in (3.50) by the chordal distance

$$d = 2a \sin\left(\frac{\theta}{2}\right), \tag{3.51}$$

where θ is the angle between the two points on the circle and a is the radius. This substitution is necessary to allow any valid correlation model on the real line to also be a valid correlation model on the circle, [88], [92]. The SOAR error correlation matrix C_{SOAR} is such that

$$(C_{SOAR})_{i,j} = \left(1 + \frac{|2a \sin\left(\frac{\theta_{i,j}}{2}\right)|}{L}\right) exp\left(-\frac{|2a \sin\left(\frac{\theta_{i,j}}{2}\right)|}{L}\right), \tag{3.52}$$

for $i, j = 1, ..., N$, where $\theta_{i,j}$ is the angle between the points p_i and p_j on the circle. It has been previously shown that increasing the correlation length-scale L increases the condition number of the SOAR auto-covariance matrix, [41].

3.3.4.2 The Laplacian Covariance Matrix

The Laplacian correlation matrix is obtained from the explicit expression

$$C_{LAP}^{-1} = \gamma^{-1}\left(I + \frac{L^4}{2\Delta x^4} D_L^2\right), \tag{3.53}$$

where the great circle distance between grid points is denoted by Δx and $\gamma > 0$ is a constant that ensures that the maximum element of C_{LAP} is equal to one. The identity matrix I is size $N \times N$ and the second order derivative matrix is such that

$$D_L^2 = \begin{pmatrix} -2 & 1 & 0 & 0 & \dots & 0 & 1 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ & \ddots & \ddots & & & & \\ \vdots & & \ddots & \ddots & & & 0 \\ 0 & & & & & & 1 \\ 1 & 0 & \dots & & & 1 & -2 \end{pmatrix}. \tag{3.54}$$

The Laplacian covariance matrix is a valid correlation function on the periodic domain (for proof see [41]).

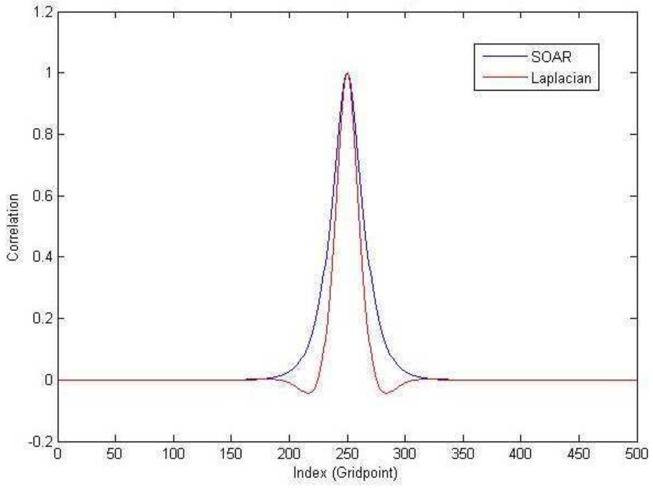


Figure 3.1: 250th row of the Laplacian (red line) and SOAR (blue line) correlation matrices. Model grid points $N = 500$, $L = 0.9$ for both Laplacian and SOAR.

The correlation structures of the SOAR and Laplcian covariance matrices are shown in Figure 3.1. The Laplacian covariance matrix has negative correlations whereas the SOAR matrix does not. We also notice that the SOAR correlations have a larger spread across the grid points in comparison to the Laplacian correlation structure.

We now introduce the apparatus we have employed in the thesis to bound the condition number of the Hessian of the wc4DVAR objective functions.

3.4 Mathematical Theorems

We aim to examine the condition number of Hessians (2.38) and (2.39), which are very large matrices. Therefore we need to introduce theory which will aid in bounding the eigenvalues of these large matrices, since the extreme eigenvalues compose the definition of the condition number we have chosen in this thesis.

3.4.1 Eigenvalue Bounds and Mathematical Results

We begin with the following determinant theorem.

Theorem 3.4.1 *For any given square matrices $A, B \in \mathbb{R}^{N\times N}$ of equal size we have*

$$Det(AB) = Det(A)Det(B). \tag{3.55}$$

One of the most useful eigenvalue bounds used on more than one occasion in our work is the following.

Theorem 3.4.2 *Courant-Fischer Theorem [See [35], Section 8.1].*
For any given symmetric *matrices $A, B \in \mathbb{R}^{N\times N}$ the k^{th} eigenvalue of the matrix sum $A + B$ satisfies*

$$\lambda_k(A) + \lambda_{min}(B) \leq \lambda_k(A + B) \leq \lambda_k(A) + \lambda_{max}(B). \tag{3.56}$$

We also have

Theorem 3.4.3 (See [35], Sec 8.6) *Let $E \in \mathbb{R}^{N\times M}$ such that $M < N$. Then the non-zero eigenvalues of EE^T and E^TE are equal and EE^T has N - M additional eigenvalues equal to zero.*

Another simple yet effective upper bound using norms is as follows:

Theorem 3.4.4 (See [3], Section A.1) *For a matrix $A \in \mathbb{R}^{N\times N}$ then the following is true:*

$$|\lambda_k(A)| \leq ||A||_p \tag{3.57}$$

for $p \geq 1$.

Finally,

Theorem 3.4.5 (See [11], Section 2.4 (p13-14)) *For finite $m, n \in \mathbb{Z}_{>0}$ and $p \in \mathbb{R}$, we have:*

$$\sum_{p=m}^{n} p = \frac{(n+1-m)(n+m)}{2} \tag{3.58}$$

$$\sum_{p=1}^{n} p^2 = \frac{n(n+1)(2n+1)}{6}. \tag{3.59}$$

We now introduce the Rayleigh Quotient.

3.4.2 Rayleigh Quotient

The Rayleigh Quotient is historically named after Baron Rayleigh (John William Strutt), an English physicist who received a Nobel prize in physics in 1904. This function is also known as the 'Rayleigh-Ritz ratio' in engineering, where it was also named after Walther Ritz, a Swiss theoretical physicist. The Rayleigh Quotient is a function which we use for the purpose of eigenvalue estimation in this thesis.

Definition 3.4.6 (See [3], Section 4.4) *The Rayleigh quotient of a symmetric matrix $A \in \mathbb{R}^{N\times N}$ is as follows:*

$$\mathcal{R}_A(\mathbf{x}) = \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}} \tag{3.60}$$

for $\mathbf{x} \in \mathbb{C}^N$, where $\mathbf{x}^H$ is the Hermitian of $\mathbf{x}$.

To find the smallest eigenvalue one would simply substitute the eigenvector that corresponds to the smallest eigenvalue,

$$\mathcal{R}_A(\mathbf{x}_{min}) = \frac{\mathbf{x}_{min}^H A\mathbf{x}_{min}}{\mathbf{x}_{min}^H\mathbf{x}_{min}} = \lambda_{min}(A). \tag{3.61}$$

The Rayleigh quotient is also bounded by the eigenvalue spectrum of the matrix.

Theorem 3.4.7 (See [81], Section 5.9) *Let $A \in \mathbb{R}^{N\times N}$ be a symmetric matrix. Then the Rayleigh quotient (3.4.6) is bounded such that:*

$$\lambda_{min}(A) \leq \mathcal{R}_A(\mathbf{x}) \leq \lambda_{max}(A). \tag{3.62}$$

3.4.3 The Block Analogue of Geršgorin's Circle Theorem

Semyon Aranovich Geršgorin introduced his theorem as early as the 1930's, [32], now known as the *scalar* Geršgorin's circle theorem. He introduced the notion of bounding the eigenvalues of a matrix by the sum of the row and/or column constituents in the following theorem.

Theorem 3.4.8 (See [85]) *Let $A \in \mathbb{C}^{N\times N}$. Then all eigenvalues λ of A satisfy*

$$|\lambda - a_{i,i}| \leq \sum_{j\neq i}^{N} |a_{i,j}|, \tag{3.63}$$

where $a_{i,j}$ denotes the entries of A on the i^{th} row and j^{th} column.

It is a well-known theorem with many applications in linear algebra and numerical analysis for estimating eigenvalue spectrums. Varga and Feingold extended this to encompass block matrices some 30 years later, [23].

Theorem 3.4.9 (See [23], Theorem 2) *Let $\mathbf{A} \in \mathbb{C}^{N(n+1)\times N(n+1)}$ be a partitioned matrix such that*

$$\mathbf{A} = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & & \\ & & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,n-1} & A_{n,n} \end{pmatrix}, \tag{3.64}$$

where each $A_{i,i} \in \mathbb{C}^{N\times N}$. Then each eigenvalue λ of $\mathbf{A}$ satisfies

$$||(A_{i,i} - \lambda I_i)^{-1}||^{-1} \leq \sum_{\substack{i\neq j \\ j=1}}^{n} ||A_{i,j}||, \tag{3.65}$$

for at least one i, $1 \leq i \leq n$. Remark*: If the partitioning of (3.64) is such that all the diagonal submatrices are 1×1 matrices and $||x|| = |x|$ (since they are now scalar), then Theorem 3.4.9 reduces to the Gershgorin Circle Theorem 3.4.8.*

This constitutes all the mathematical apparatus used in the rest of the thesis. We now introduce the models used in our experiments to demonstrate the sensitivities obtained from the theoretical bounds on the condition number of the Hessian.

3.5 Models

In this section we introduce the models used in this thesis to illustrate the theory we have derived.

The first model is a linear advection equation. This is a simplified model describing the transportation of a passive tracer through the atmosphere. In the atmosphere if we consider very small intervals of space and time, the movement of a passive tracer will be approximately linear, similar to that of the advection equation.

The second model is the non-linear chaotic Lorenz 95 system. The variables in this system simulate values of some atmospheric quantity in sectors of a latitude circle. The physics of the model possess useful weather-model-like characteristics such as external forcing, internal dissipation and advective terms. The error growth of this model is also similar to that of full NWP models.

The numerical discretisation of these models presents a set of calculations required to propagate the model from one time step to the next. These are represented in matrix form in the following sections. We now introduce the models used in this thesis.

3.5.1 The Advection Equation

The advection equation is a partial differential equation describing the flow of a scalar quantity, $u(x,t)$, through space, x with respect to time, t:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{3.66}$$

where the scalar quantity is moved through a vector field at a velocity of $a(x,t)$. We only consider the case where the speed is constant. Solutions of (3.66) are then of the form $u(x,t) = u(x - at)$. We also specify periodic boundary conditions so that the problem is well-posed and the unique solution depends continuously on the boundary [49]. The initial conditions we use throughout the thesis are of a Gaussian profile such that

$$u(x,0) = be^{-\frac{(x-c)^2}{2d^2}}, \tag{3.67}$$

where b, c and d are constants denoting height, peak centre and 'bell' width respectively.

We discretise (3.66) using the upwind numerical scheme, [69], Chapter 4. We have a uniform 1-dimensional domain which is divided into N equally spaced grid points of length $\Delta x = \frac{1}{N}$. We discretise time by dividing it into n equally spaced time steps of length $\Delta t = \frac{1}{n}$. Let $u_j^i = u(j\Delta x, i\Delta t)$ be the numerical approximation of $u(x,t)$ at the point $(j\Delta x, i\Delta t)$ for $j = 1, ..., N$, $i = 0, ..., n$. The upwind scheme uses a finite difference approximation which adapts according to the direction of velocity, a:

$$u_j^{i+1} = \begin{cases} u_j^i - a\frac{\Delta t}{\Delta x}(u_j^i - u_{j-1}^i) & \text{if } a > 0 \\ u_j^i - a\frac{\Delta t}{\Delta x}(u_{j+1}^i - u_j^i) & \text{if } a < 0 \end{cases}, \tag{3.68}$$

with periodic boundary conditions,

$$u_1^i = u_N^i \text{ if } a > 0 \tag{3.69}$$

$$u_N^i = u_1^i \text{ if } a < 0 \tag{3.70}$$

for all $i = 0, ..., n$. In this thesis we restrict ourselves to negative velocities $a < 0$ using the upwind scheme. Now let $\mu = a\frac{\Delta t}{\Delta x}$ denote the Courant number. We can write the model equations for $a < 0$ in matrix form as

$$\begin{pmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_N \end{pmatrix}^{i+1} = \begin{pmatrix} 1+\mu & -\mu & 0 & 0 & 0 \\ 0 & 1+\mu & -\mu & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \ddots & -\mu \\ -\mu & 0 & 0 & 0 & 1+\mu \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_N \end{pmatrix}^{i}. \tag{3.71}$$

For $-1 \leq \mu \leq 0$ the finite difference system (3.71) is consistent, stable and convergent, [69], Section 5.4.

We have introduced all the necessary properties of the advection model that we use in the thesis. We now discuss the non-linear chaotic Lorenz 95 model.

3.5.2 The Lorenz 95 Model

The Lorenz 95 model was pioneered by Edward Lorenz, making its first appearance in the article [62], in 1996. This later made its way into published format accompanied with a few more complex versions of the same model, with the aim of designing suitable models for weather prediction, [63]. The Lorenz 95 model has been used as a suitable experimental model in the data assimilation research community, [24], [9], [22], as well as the operational research community, [26].

To understand the relevance of using the Lorenz 95 system we must understand the essence of predictability. The rate of error growth in a system as the range of prediction increases is a highly influential factor in determining system predictability. Prediction error is simply the difference between the estimated state and the true state. This is hypothetical as it is not a quantity we can explicitly state, but it can be quantified. The long-term average factor by which an infinitesimal error will amplify per unit time is known as the *leading Lyapunov number*, named after Russian mathematician Aleksandr Mikhailovich Lyapunov. The logarithm of this quantity is known as the *leading Lyapunov exponent.* A common indication of a chaotic system is a positive leading Lyapunov exponent. A more common term used within the meteorological community is the 'doubling time' of a system, which is inversely proportional to the Lyapunov exponent.

The Lorenz 95 system variables can be thought of as values of some atmospheric quantity in N sectors of a latitude circle. The physics included in this model are external forcing, internal dissipation and advection. The quadratic advective terms are also designed to conserve energy. The growth of errors of the Lorenz

95 system are similar to that of full weather models, with a doubling time of 2.1 days, making it a suitable model to use for weather prediction purposes.

The Lorenz 95 ODE equations take the form

$$\frac{dX_j}{dt} = -X_{j-2}X_{j-1} + X_{j-1}X_{j+1} - X_j + F, \tag{3.72}$$

for $j = 1, 2, ..., N$. In this thesis we work with $N = 40$ variables, so that each sector of the latitude circle covers 9 degrees of longitude. We set the forcing term $F = 8$ to produce the chaotic behaviour desired. The scaling of the variables in the model dictates that one time unit is equivalent to 5 days. We use the 4th order Runge-Kutta method (RK4) with a time interval of $\Delta t = 0.025$, which is equivalent to a 3 hour time-step.

We have introduced both models used in the research in this thesis. We now summarise this chapter.

3.6 Summary

The aim of this chapter was to introduce the necessary mathematical framework required to obtain and demonstrate the theoretical results in this thesis.

In Section 3.1 we showed the direct relationship between the condition number of the first-order Hessian and solution error in the 4DVAR problem. In Section 3.2 we discussed the three CG-based methods: LCG, PCG and PRCG and we also discussed our rationale for using a relative-gradient norm iterative stopping criterion. We then discussed matrix properties and norms relevant to our work, along with the specific class of matrices and covariance structures used in our research in Section 3.3. Section 3.4 details the more specialist mathematical theorems required to analyse and bound the condition number of the Hessians of the wc4DVAR objective functions (3.1) and (3.2). Finally we introduce the models used in the thesis in Section 3.5.

In the next chapter we discuss the design considerations for the application of both formulations $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$ on the 1D advection model. We then compare the performance of both formulations of wc4DVAR when subjected to changes in the data assimilation parameters composing the problem.

Chapter 4

Weak-Constraint 4DVAR: 1D Advection Model

In the previous chapter we discussed all the relevant mathematics to enable us to deduce the results in this thesis. In this chapter we highlight the key differences between the model error estimation (2.32) and state estimation (2.33) problems via numerical experiments using the 1D advection equation model.

4.1 Experimental Design

In this section we detail the specifics of setting up the two resulting algorithms of the weak-constraint formulations (2.32) and (2.33) to carry out numerical experiments. We state the wc4DVAR design considerations such as the model parameters, wc4DVAR component tests and observation configuration in this section. Before doing so we clarify the difference between the model and the weather process for which we assimilate data.

4.1.1 The Imperfect Model

To carry out any credible experiments to test hypotheses and new theory in data assimilation, we require a model $\mathcal{M}_{i,i-1}$ as described in Section 2.2. We consider the weather process in its entirety as the perfect model and the imperfect model is the human approximation of this weather process.

The *weather process* is such that,

$$x_i^t = \mathcal{M}_{i,i-1}^t(x_{i-1}^t), \tag{4.1}$$

where $\mathcal{M}^t$ describes the *true* non-linear weather process and x^t is the true state. One way of gauging assimilation performance is measuring how closely the chosen wc4DVAR algorithm can quantify the true model error. In our experiments, we create this model error using a specified covariance matrix with zero mean and a specified variance.

The approximation, which attempts to match the perfect model, is the *imperfect model*,

$$x_i^t = \mathcal{M}_{i,i-1}(x_{i-1}) + \eta_i, \tag{4.2}$$

which we use in the wc4DVAR algorithms. The model error/shortfall is assumed to be additive Gaussian noise as in (2.24).

The true model error is created using a known mean, variance and covariance matrix. We gauge the performance of the assimilation algorithms by comparing how closely each algorithm was able to estimate the true model error.

We now discuss the model parameter settings used in the experiments in this chapter.

4.1.2 1D Advection Equation: Model Properties

The model is the 1-dimensional advection equation discretised using the upwind scheme, yielding the matrix as in (3.71), which we denote as M. In this chapter we use the following model settings unless otherwise stated. The spatial domain is size $N = 50$ with a spatial resolution of $\Delta x = 0.02$. We use time-intervals of $\Delta t = 0.02$ and a wave speed of $a = -1$, thus giving us a Courant number of $\mu = -1$.

For testing the wc4DVAR systems we set the total assimilation window time $n = 50$. We also make sure that the assimilation time period in these tests and each experiment is enough for at least one complete spatial-domain revolution of the Gaussian curve.

We now discuss testing the wc4DVAR systems.

4.1.3 The Weak-Constraint 4DVAR System

The sc4DVAR system has a forward model, (2.1), linearised model and adjoint model, which arise from calculating the gradient (2.9). Once all of the constituents of the gradient and objective function are validated, it follows that one must ensure that the coded gradient is in fact, the gradient of the objective function (2.8). It is therefore good practice to include the following tests in the design of the sc4DVAR assimilation system:

1. tangent linear test;
2. adjoint test;
3. objective function gradient test.

The wc4DVAR the model operator and its inverse come from (2.29), since it maps between model errors and model states. The wc4DVAR equivalent to the adjoint and tangent linear arise from linearising $\boldsymbol{\mathcal{L}}$. The input and output of the wc4DVAR operators are '4-dimensional', since they require inputs defined at several temporal points. The wc4DVAR model operator also has linearised inverses, $\mathbf{L}^{-1}$ and $\mathbf{L}^{-T}$, which constitute part of the wc4DVAR gradient calculations. So the additional tests required for wc4DVAR are to ensure that the mapping between model states and model errors is correct for non-linear $\boldsymbol{\mathcal{L}}$ and linearised $\mathbf{L}$ operators and their inverses.

We carry out four principal tests in the preceeding sections to ensure the that the wc4DVAR assimilation system is correctly coded. The first test is checking that the numerical mapping of; the $\boldsymbol{\mathcal{L}}$ operator, the linearised $\mathbf{L}$ operator and the linearised adjoint operator $\mathbf{L}^T$ are all correct. The second test ensures that the gradient of the $\boldsymbol{\mathcal{L}}$ operator and its inverse, are indeed correct. The third test checks that the adjoint of both $\mathbf{L}$ and its inverse are correct representations of the adjoint. Finally, the fourth test ensures that the coded objective function gradient is infact the gradient of $\mathcal{J}$.

We generate the test states in all these tests using pseudo-random values drawn from the normal distribution with arbitrary mean and variance values. The chosen state remains unchanged throughout the test. We now detail each of the model operator tests with numerical results to verify each test.

4.1.3.1 The Weak-Constraint Model Operator: Mapping Tests

The main purpose of this test is to ensure the numerical validity of the following:

1. non-linear model operator and inverse;

   (a) $||\boldsymbol{\mathcal{L}}(\mathbf{x}) - \mathbf{p}|| \approx 0$;

   (b) $||\boldsymbol{\mathcal{L}}^{-1}(\mathbf{p}) - \mathbf{x}|| \approx 0$.

2. Linearised model operator and inverse;

   (a) $||\mathbf{L}\delta\mathbf{x} - \delta\mathbf{p}|| \approx 0$;

   (b) $||\mathbf{L}^{-1}\delta\mathbf{p} - \delta\mathbf{x}|| \approx 0$.

3. Linearised adjoint model operator and inverse;

   (a) $||\mathbf{L}^T\delta\mathbf{x} - \delta\mathbf{p}|| \approx 0$;

   (b) $||\mathbf{L}^{-T}\delta\mathbf{p} - \delta\mathbf{x}|| \approx 0$.

The quantities in tests 1, 2 and 3 must equal exactly zero or be very close to machine precision $\sim \mathcal{O}(10^{-15})$. We choose the 2-norm for each test detailed above and ensure it is in the vicinity of machine precision.

Test	**Norm of the Difference**
1(a)	1.70E-014
1(b)	3.72E-015
2(a)	1.43E-015
2(b)	1.37E-015
3(a)	1.32E-015
3(b)	1.43E-015

Table 4.1: Mapping test results.

Table 4.1 shows that the results are all in the region of machine precision, therefore the numerical mapping tests are all numerically valid.

We now discuss the wc4DVAR equivalent of the tangent linear test.

4.1.3.2 The Linearised Weak-Constraint Model Operator: Correctness Tests

Taylor expansion of our non-linear operator to first-order yields the following approximated identities:

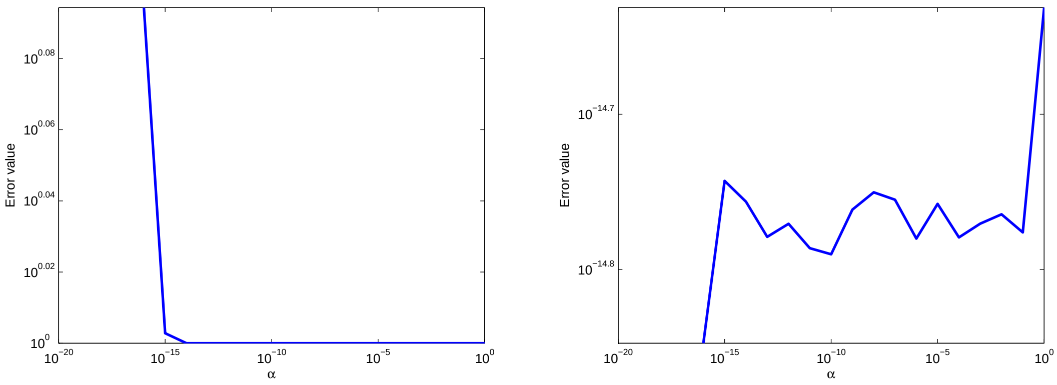
$$\frac{||\boldsymbol{\mathcal{L}}(\mathbf{x} + \alpha_i\delta\mathbf{x}) - \boldsymbol{\mathcal{L}}(\mathbf{x})||}{||\mathbf{L}\alpha_i\delta\mathbf{x}||} = 1 + \mathcal{O}(\alpha_i\delta\mathbf{x}), \tag{4.3}$$

$$||\boldsymbol{\mathcal{L}}(\mathbf{x} + \alpha_i\delta\mathbf{x}) - \boldsymbol{\mathcal{L}}(\mathbf{x}) - \mathbf{L}\alpha_i\delta\mathbf{x}|| \approx 0, \tag{4.4}$$

which should hold for small values of $\alpha_i \delta \mathbf{x}$. We vary α_i such that

$$\alpha_i = 10^{1-i}, \tag{4.5}$$

for $i = 1, ..., 16$. Since the advection model is linear, there should be no higher order terms in the expansions above. The purpose of these tests is to ensure the numerical validity correctness of the gradients of these two operators. We also test the inverse, $\boldsymbol{\mathcal{L}}$ in a similar manner.

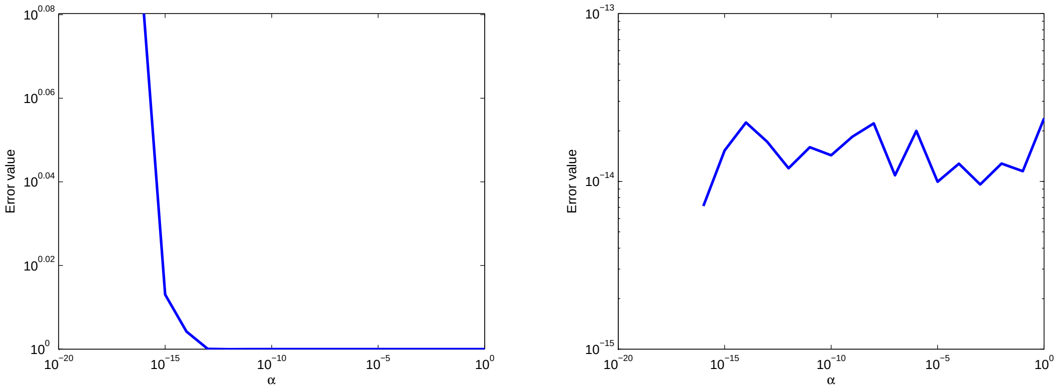


(a) Identity test (4.3).

(b) Identity test (4.4).

Figure 4.1: Correctness test plots for the **L** operator.

In Figure 4.1 we see (a) adheres to identity (4.3) since it equals one for all sizes of α up to machine precision. Figure 4.1(b) shows that identity (4.4) is equal to approximately $\mathcal{O}(10^{-15})$.



(a) Identity test (4.3).

(b) Identity test (4.4).

Figure 4.2: Correctness test plots for the $\mathbf{L}^{-1}$ operator.

Figure 4.2 shows that the correctness tests also hold for inverse operator, $\mathbf{L}^{-1}$.

We now discuss the final test with regards to the $\mathcal{L}$ operator. This is required for the calculation of the gradients of $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$.

4.1.3.3 The Linearised Weak-Constraint Adjoint Model Operator: Validity Tests

This test is equivalent to the sc4DVAR adjoint test. The aim is to test the validity of the inner products

$$< \delta\mathbf{y}, \mathbf{L}\delta\mathbf{x} > =< \mathbf{L}^T\delta\mathbf{y}, \delta\mathbf{x} >, \tag{4.6}$$

$$< \delta\mathbf{y}, \mathbf{L}^{-1}\delta\mathbf{x} > =< \mathbf{L}^{-T}\delta\mathbf{y}, \delta\mathbf{x} > . \tag{4.7}$$

These tests are done by executing each side of the respective equations numerically and comparing the results. We call the left-hand side of each equation (4.6) and (4.7) the 'forward product' and the right-hand side is called the 'adjoint product'.

	Forward Product	**Adjoint Product**	**Difference**
Test (4.6)	-45.484273829763183	-45.484273829763133	4.9738e-014
Test (4.7)	-216.363507105409070	-216.363507105409130	5.6843e-014

The difference of both products is in the range of machine precision, which concludes that the numerical adjoint operator is accurate to machine precision.

This concludes all the tests for the $\mathcal{L}$ operator. The $\mathcal{L}$ operator is required for both calculating the objective functions (2.32), (2.33) and the gradients of the objective functions (2.34) and (2.35). We now discuss the final test in the assimilation system, which tests the numerical validity of the coded objective function gradient.

4.1.3.4 Objective Function Gradient: Validity Tests

This test is similar to the tests in Section 4.1.3.2, but instead we check the numerical validity of the objective functions (2.32) and (2.33) and their respective gradient calculations (2.34) and (2.35). We verify that

$$\Phi(\alpha) = \frac{\mathcal{J}(\mathbf{x} + \alpha\delta\mathbf{x}) - \mathcal{J}(\mathbf{x})}{\alpha\delta\mathbf{x}^T\nabla\mathcal{J}(\mathbf{x})} = 1 + \mathcal{O}(\alpha), \tag{4.8}$$

is accurate for sufficiently small perturbations $\alpha\delta\mathbf{x}$.

The gradient test for the objective function is different to the gradient test in Section 4.1.3.2 because the operators are different. The operator in Section 4.1.3.2 is such that $\boldsymbol{\mathcal{L}} : \mathbb{R}^{N(n+1)} \rightarrow \mathbb{R}^{N(n+1)}$, which is why norms were used. The weak-constraint objective functions (2.32) and (2.33) are such that $\mathcal{J} : \mathbb{R}^{N(n+1)} \rightarrow \mathbb{R}$, so no norms are required.

For perturbations that are too large the identity (4.8) will not hold since the higher order terms will increase and the approximation made in (4.8) is to first-order. If the perturbations are too close to machine precision the identity (4.8) will not hold because the denominator of (4.8) will approach zero.

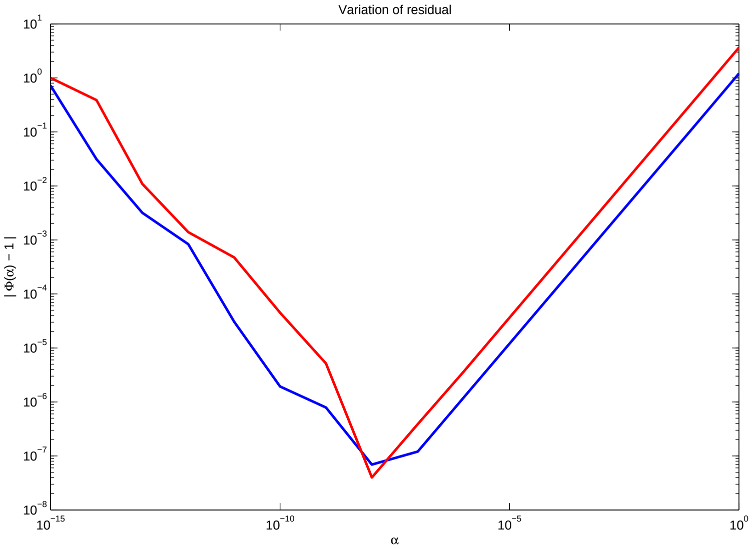


Figure 4.3: Objective function gradient test. The red line shows the gradient test (4.8) for $\mathcal{J}(\mathbf{p})$. The blue line shows the gradient test (4.8) for $\mathcal{J}(\mathbf{x})$.

Figure 4.3 shows that for sufficiently small perturbations the identity (4.8) holds for both $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$.

This concludes all the tests to ensure mathematical and numerical accuracy of both wc4DVAR assimilation systems for solving $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$. The second consideration to discuss is the nature of the observations we use to observe the truth.

4.1.4 Observations

The observations $\mathbf{y}$ are generated using the truth trajectory plus additive Gaussian noise such that

$$\mathbf{y} = \mathbf{y}^t + \mathbf{y}^e, \tag{4.9}$$

where $\mathbf{y}^t$ is the unchanged true state at the appropriate spatio-temporal grid-points, and $\mathbf{y}^e \sim N(0, \sigma_o^2 \mathbf{I})$. The observation error variance is stated before each experiment.

We take the observations directly at the grid points with regular intervals in space, *where the first spatial point is always observed.* We also observe at regular intervals in time, *where the first temporal point is always observed.* We let the temporal observation interval (also referred to in this thesis as an 'assimilation step') be every Δq model steps. We observe the same grid-points at every assimilation step, thus the observation operator H_i is linear and time invariant. So for example if we observe 5 out of 10 spatial points, then grid-points 1,3,5,7,9 are observed. This means *every* 2^{nd} *gridpoint* is observed. This also applies temporally.

We also note that it is possible to take combinations of observations in space and time such that the observations can miss some of the *characteristic lines* due to our chosen regular spatial-temporal observational spacing regime.

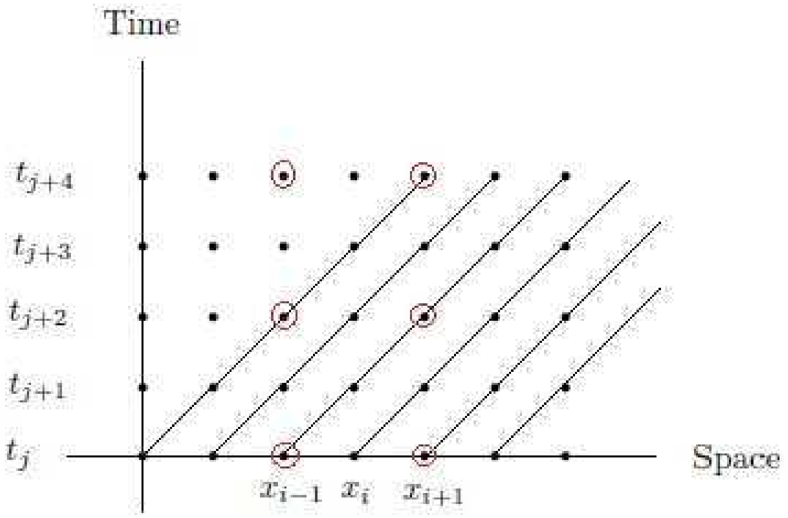


Figure 4.4: Advection Equation characteristic curves. The black lines are the advection equation characteristic lines, and the red circles are observation points.

In Figure 4.4 we see that if we were to observe every other temporal and spatial point, some of the characteristic lines will be missed. Even with a periodic domain, the same characteristic lines will remain unobserved for an indefinite time period. We ensure that the temporal and spatial spacing of the observations is such that none of the characteristic lines are missed.

In this section we have discussed our choice of observation configuration. We now state how our background trajectory is created.

4.1.5 Background Trajectory

The background trajectory, $\mathbf{p}^b$, is created using the truth trajectory plus additive Gaussian noise such that

$$\mathbf{p}^b = \mathbf{p}^t + \mathbf{p}^e, \tag{4.10}$$

where $\mathbf{p}^e \sim N(0, \mathbf{D}^e)$, where $\mathbf{D}^e$ is the covariance of the 'true' background and model errors added to the truth. The background and model error covariance matrices, B_0 and Q are stated before each experiment.

We now detail the method we use to calculate the solution accuracy.

4.1.6 Solution Error

The relative solution errors are calculated at each time t_i such that

$$re_i = \frac{||x_i^t - x_i||_2}{||x_i^t||_2}, \tag{4.11}$$

where $x_i \in \mathbb{R}^N$ is the solution vector resulting from the assimilation, which describes the state at time t_i and the superscript denotes 'truth'. The total relative error is simply the L_2 norm calculation of the vector containing the values of re_i for $i = 1, ..., n+1$.

We now state our choice of iterative solver.

4.1.7 Iterative Solver and Stopping Criterion

We use the LCG method detailed in Section 3.2.1 for both (2.32) and (2.33). Both objective functions are linear because the 1D advection model is linear. We also use the relative reduction in the gradient norm as in Section 3.2.4 and specify a tolerance, τ, in each experiment.

This concludes the experimental design for our experiments in this chapter. We now show results on the comparison between (2.32) and (2.33) when applied to the 1D advection equation.

4.2 Results

In this section we compare the performance of the minimisation of (2.32) and (2.33) applied to the 1D advection model. We aim to demonstrate the different minimisation characteristics exhibited by both weak-constraint formulations when subjected to a change in the following assimilation parameters:

1. number of observations;
2. length of the assimilation window;
3. correlation length-scales;
4. background, model and observation error variances.

We gauge the performance of the weak-constraint minimisation problems by examining:

1. the relative error within the assimilation window between the truth and the solution. We compare the generated truth to the state estimates obtained using the $\mathcal{J}(\mathbf{x})$ formulation. We also compare the generated 'true' model errors to the model error estimates obtained from the $\mathcal{J}(\mathbf{p})$ formulation;
2. the number of iterations required to achieve the desired tolerance;
3. the numerical condition number.

The covariances and error variances used to generated the truth are identical to those used in the assimilation experiments. We now present our experimental results.

4.2.1 Experiment 1: Observation Density

The aim of this experiment is to highlight the effect of number of observations on the solution process of both wc4DVAR formulations. We choose all other parameters in this experiment such that the only possible contribution to any rise in condition number must be the number of observations. So we choose low correlation length-scales, short assimilation windows and error variance ratios which are close to 1.

4.2.1.1 Experiment 1a: Half Spatial Domain Observed

The experiment settings are as follows. We choose the the background error, $B_0 = \sigma_b^2 C_{SOAR}$, such that the correlation length-scale $L = 2\Delta x = 0.04$ and $\sigma_b = 0.1$. The model error, $Q_i = \sigma_q^2 C_{LAP}$ is such that the correlation length-scale $L = \Delta x = 0.02$ and $\sigma_q = 0.05$. The observation error is such that $R_i = \sigma_o^2 I$, where $\sigma_o = 0.05$. We take observations every $\Delta q = 5$ model time-steps, $n = 10$ in total, with 25 equally spaced observed grid-points out of the $N = 50$ grid-points per assimilation step. The iterative tolerance is set to $\tau = 10^{-4}$.

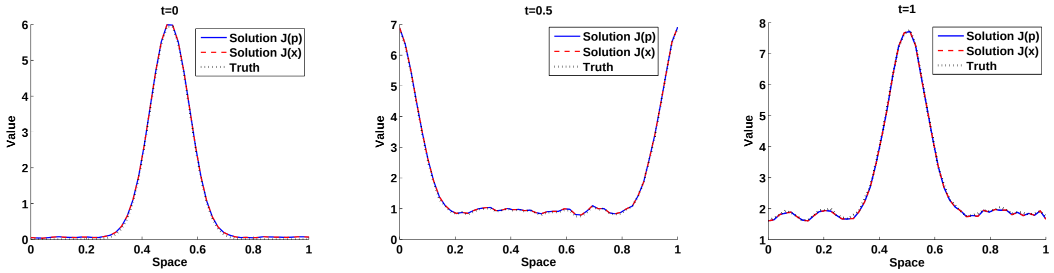


Figure 4.5: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.5 we see the time series plot of the truth and the solutions of both wc4DVAR algorithms. We can see that visually the solutions are in close agreement with the truth.

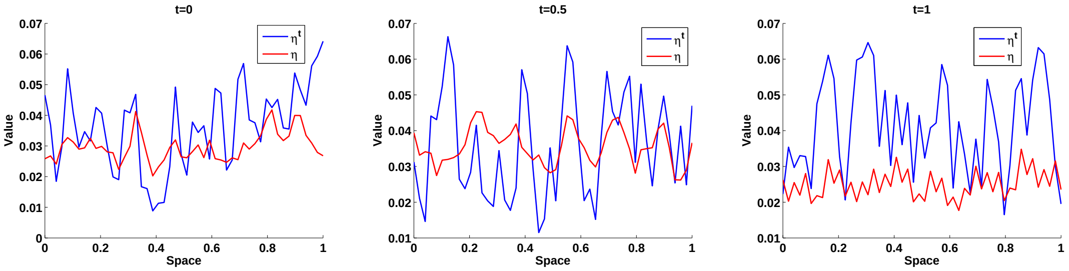


Figure 4.6: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

In Figure 4.6 we see the time series plot of the true model error vs the estimated model error at the end of the minimisation using $\mathcal{J}(\mathbf{p})$. The variance of the estimated model error is of the same order of magnitude as the true model error for all times, but variance of the estimated model error is consistently under-estimated. We also see that the estimated model error at final time is considerably worse than other times.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	278	25
$\mathbf{S}_x$	766	93
$\mathbf{D}$	837	-

Table 4.2: Numerical condition numbers and iteration count of respective objective function minimisations.

In Table 4.2 we see that the number of iterations required for the minimisation of $\mathcal{J}(\mathbf{x})$ to converge is over 3 times more than $\mathcal{J}(\mathbf{p})$. We also see that the numerical condition number of $\mathbf{S}_x$ is approximately 3 times as much as the numerical condition number of $\mathbf{S}_p$. It is expected to find an increase in iterations with the increase in condition number but the similar proportional increase in condition number and iterations in Table 4.2 is coincidental.

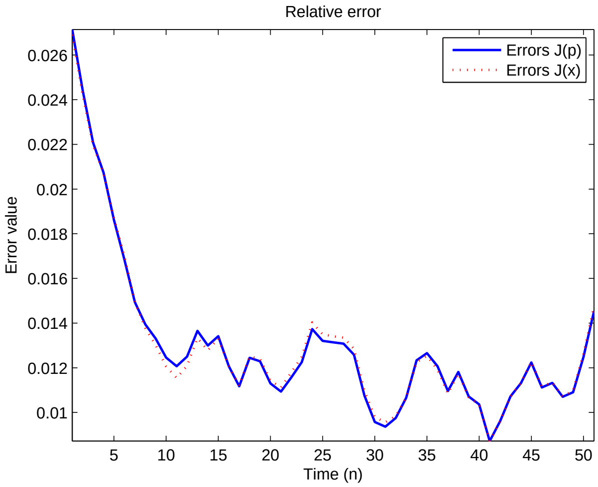


Figure 4.7: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.7 we see that the relative errors are of the same order of magnitude throughout the assimilation window, with the largest errors being at initial time. The total relative errors are identical, 0.096 for both $\mathcal{J}(\mathbf{x})$ and $\mathcal{J}(\mathbf{p})$. The low solution errors are to be expected since both Hessians are not ill-conditioned.

4.2.1.2 Experiment 1b: Sparse Spatial Observations

The experiment settings are identical to those in Experiment 1a, except that there are 7 out of 50 spatial points observed per assimilation time step. We also use the same background trajectory generated in Experiment 1a.

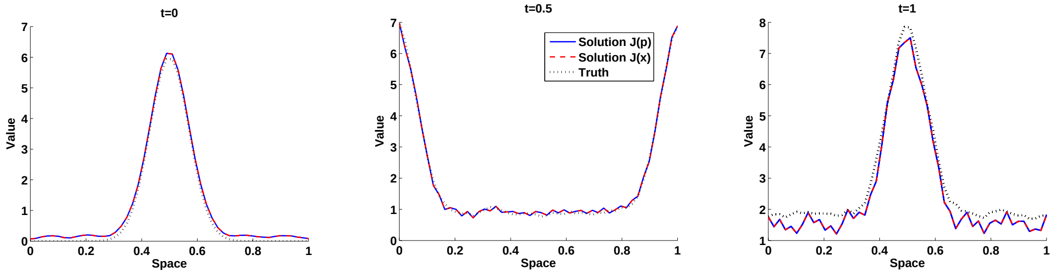


Figure 4.8: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.8 we see the time series plot of the truth and the solutions of both wc4DVAR algorithms. The solutions are now noticeably missing the truth due to the greatly reduced number of observations, from 250 in Experiment 1a to 70 in this experiment. The final time-step state estimations of both algorithms is also quite poor, since most small-scale features of the truth are missed.

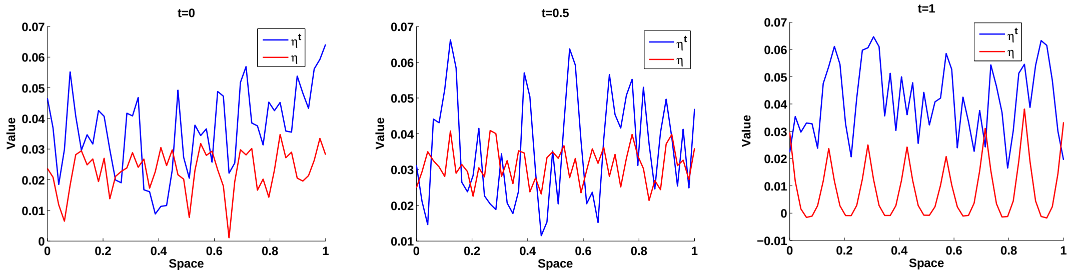


Figure 4.9: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

In Figure 4.9 we see the time series plot of the true model error vs the estimated model error at the end of the minimisation using $\mathcal{J}(\mathbf{p})$. The variance of the estimated model errors remains within the same order of magnitude as the true model errors. We also see that the mean of the estimated model errors is good for

$t = 0$ and $t = n/2$ with a noticeable under-estimation of the variance. We also see evidence of poor model error estimation at the final time step in Figure 4.9. The final time step estimated model error mean is incorrect, however the variance has been well estimated.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	575	43
$\mathbf{S}_x$	1663	412
$\mathbf{D}$	837	-

Table 4.3: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.3 shows that minimisation of $\mathcal{J}(\mathbf{x})$ takes 10 times more iterations than $\mathcal{J}(\mathbf{p})$, as well as an increase in Hessian condition number. These condition numbers are still not particularly indicative of any serious ill-conditioning. We believe the condition number of $\mathbf{D}$ is not the main contributor of ill-conditioning in this experiment since it remains the same as Experiment 1a, while the only change we have introduced is a decrease in the number of observations. The observations are associated with the second term of both Hessians $\mathbf{S}_p$ and $\mathbf{S}_x$, where $\mathbf{D}$ is the first term.

We also see that the condition numbers of $\mathbf{S}_p$ and $\mathbf{S}_x$ have both roughly doubled, compared to Experiment 1a, while the condition number of $\mathbf{S}_x$ remains approximately 3 times higher than the numerical condition number of $\mathbf{S}_p$. It is possible that the $\mathcal{J}(\mathbf{x})$ formulation is sensitive to the decrease in spatial observations, due to the increase in condition number and iterations exhibited in this experiment.

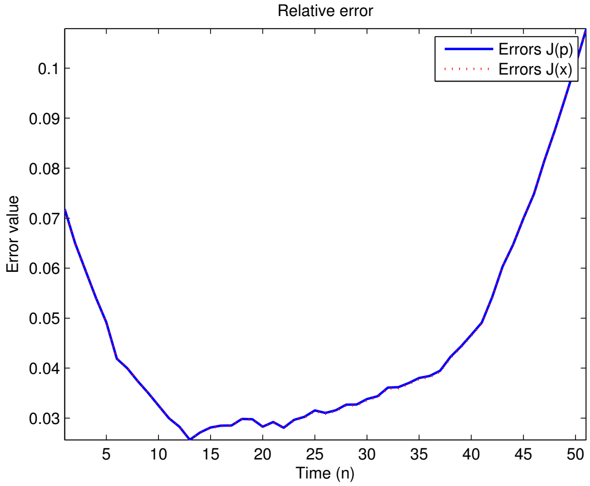


Figure 4.10: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.10 we see that the errors look the same throughout the assimilation window, with total relative errors of 0.356 for $\mathcal{J}(\mathbf{x})$ and 0.357 for $\mathcal{J}(\mathbf{p})$. We also see that the errors are distributed at the beginning and mostly the end of the assimilation window, showing that both solutions failed to correctly specify the initial conditions and the model errors at the end of the assimilation window.

4.2.1.3 Summary

The number of observations affects the assimilation problem in that there is less information to fit. In this experiment we see two pieces of evidence, which show the sensitivities of $\mathcal{J}(\mathbf{x})$ to the number of observations: the increase in numerical condition number and the number of iterations required for convergence. The errors in the solution remain the same as they should, since we solve both wc4DVAR problems to the same accuracy.

We conclude that the $\mathcal{J}(\mathbf{x})$ formulation has increased sensitivity to fewer spatial observations than $\mathcal{J}(\mathbf{p})$ in this experiment. Both formulations exhibit an increase in iterations and condition numbers when there are less observations, but $\mathcal{J}(\mathbf{x})$ is more pronounced. This is based on changes seen in the iterations and condition numbers in Tables 4.2 and 4.3.

4.2.2 Experiment 2: Error Variance Ratios

The aim of this experiment is to highlight the effect of changing the error variances $(\sigma_b^2, \sigma_q^2, \sigma_o^2)$ on the minimisation of $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$. We choose all other parameters to ensure that any change in condition number or iterations comes solely from the error variances. The iterative tolerance is changed to $\tau = 10^{-10}$ to ensure high solution accuracy. The iterative solver will reach the solution before the tolerance is reached, but we are ensuring that each algorithm yields its respective optimal solution. The iterations after reaching the solution are not important and the algorithm that reaches its solution in the least number of iterations will still take fewer iterations than the other algorithm.

We begin by investigating the effect of changing the background error variance.

4.2.2.1 Experiment 2a (i): Large Background Error Variance

The experiment settings here are exactly the same as Experiment 1a, except we change the background standard deviation from $\sigma_b = 0.1$ to $\sigma_b = 10$.

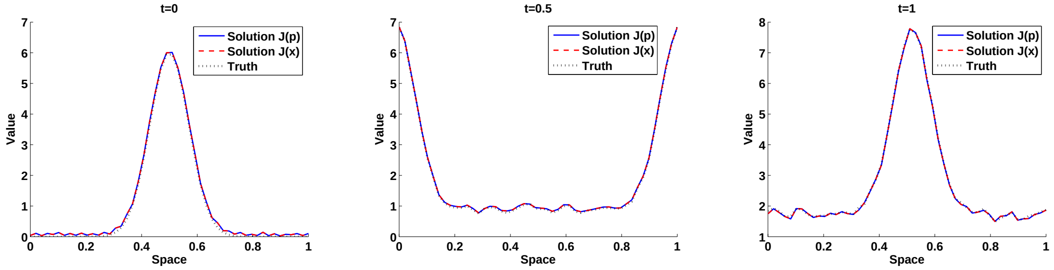


Figure 4.11: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.11 we see the time series plot of the truth and the solutions of both wc4DVAR algorithms. The error graph will perhaps be more telling of the shortfall in data matching between both formulations and the truth, but at this scale they look relatively accurate.

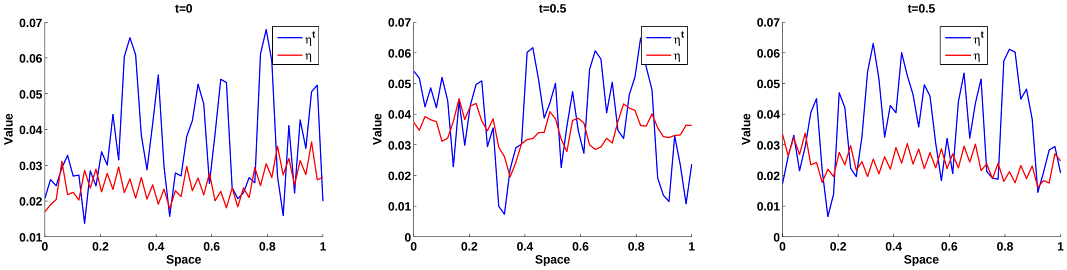


Figure 4.12: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

In Figure 4.12 we see the variance of the estimated model error is approximately 0.02 smaller than the variance of the true model error. The estimated model errors are still of the same order of magnitude as the true model errors, which is why the estimated trajectories in Figure 4.11 are closely resembling the truth trajectory. It is interesting that the variance of the estimated model errors is not as large as for the true model errors, even though identical model error statistics were used in the assimilation and to generate the truth. This under-estimation of model errors has also been seen in previous experiments.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	462	121
$\mathbf{S}_x$	742	217
$\mathbf{D}$	1.41×10^6	-

Table 4.4: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.4 shows the minimisation of $\mathcal{J}(\mathbf{x})$ requiring just under twice as many iterations as $\mathcal{J}(\mathbf{p})$ to achieve the same gradient tolerance respective to each objective function. The numerical condition number of $\mathbf{S}_x$ is approximately 2 times higher than that of $\mathbf{S}_p$. We see that the numerical condition number of $\mathbf{D}$ is of order $\mathcal{O}(10^6)$, which is about 4 orders of magnitude larger than the Hessians. It is interesting how this does not effect the condition number of the wc4DVAR Hessians, since we can see just from the structures of both Hessians (2.40) and (2.41) that $\mathbf{D}$ should be influential on the spread of eigenvalues in both Hessians.

Ratio	Value
σ_b/σ_q	200
σ_b/σ_o	200
σ_q/σ_o	1

Table 4.5: Assimilation error variance ratios.

The σ_b/σ_q ratio in Table 4.5 explains the large condition number of $\mathbf{D}$ since this ratio increases the difference between the largest and smallest eigenvalue of the matrix $\mathbf{D}$. The ratio between the background and observation error variances are of the same magnitude, both $\mathcal{O}(10^2)$ bigger than the σ_q/σ_o ratio.

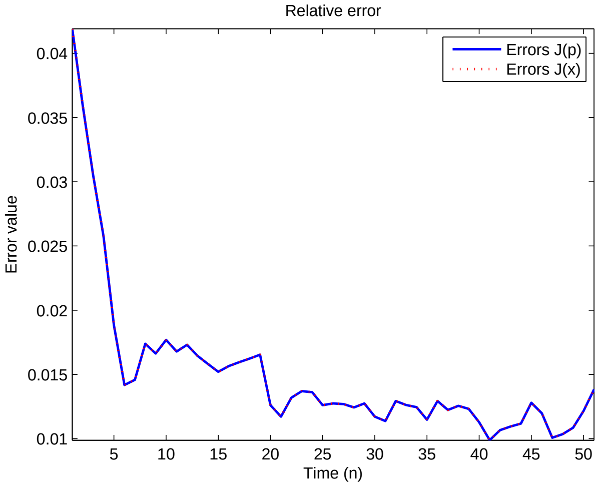


Figure 4.13: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.13 we see that the errors are identical, with the largest errors distributed at the beginning of the assimilation window. The total relative errors have identical values for both formulations, 0.116. the solution error at initial time is still the greatest for both algorithms. We see here that even with the large σ_b chosen for the assimilation and truth, the solution obtained from both algorithms has large errors at the beginning of the window. This is similar to the under-specified variance of model errors in Experiment 1b, Figures 4.9 and 4.10.

We now decrease the value of the background error and study its effects on the solution process of both wc4DVAR algorithms.

4.2.2.2 Experiment 2a (ii): Small Background Error Variance

In this experiment we use the same parameters as the previous experiment except we change the background standard deviation from $\sigma_b = 10$ to $\sigma_b = 2.5 \times 10^{-4}$ so that it is now 200 times smaller than σ_q, as opposed to being 200 times bigger as in Experiment 2a (i). We only show results related to the performance of the minimisation of both $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	8.53×10^6	635
$\mathbf{S}_x$	1.00×10^8	1756
$\mathbf{D}$	8.53×10^6	-

Table 4.6: Numerical condition numbers and iteration count of respective objective function minimisations.

In Table 4.6 we see that the minimisation of $\mathcal{J}(\mathbf{x})$ requires just under 3 times as many iterations as $\mathcal{J}(\mathbf{p})$ to achieve the same gradient tolerance respective to each objective function. The numerical condition number of $\mathbf{S}_x$ is $\mathcal{O}(10^2)$ higher than $\mathbf{S}_p$. This complements the higher number of iterations seen for $\mathcal{J}(\mathbf{x})$ over $\mathcal{J}(\mathbf{p})$. We also see that the numerical condition number of $\mathbf{S}_p$ is of the same order of magnitude as $\mathbf{D}$.

Ratio	Value
σ_b/σ_q	5×10^{-3}
σ_b/σ_o	5×10^{-3}
σ_q/σ_o	1

Table 4.7: Assimilation error variance ratios.

The small σ_b/σ_q is the reason for the large condition number of $\mathbf{D}$. The large condition numbers of $\mathbf{S}_p$ and $\mathbf{S}_x$ follow the large condition number of $\mathbf{D}$ in this experiment, with $\mathbf{S}_x$ exhibiting more sensitivity.

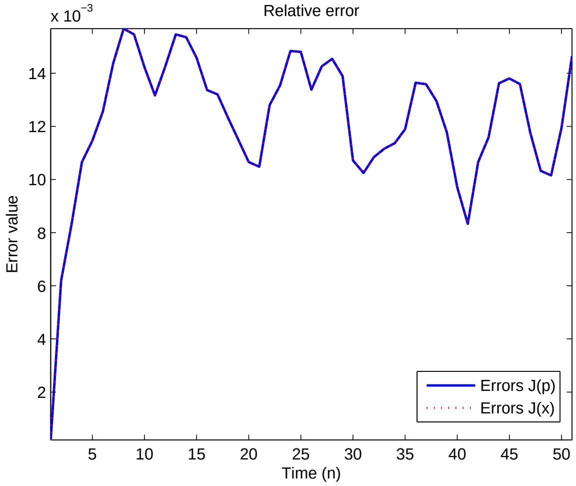


Figure 4.14: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.14 we see that the errors are identical again. The total relative errors have identical values for both formulations, 0.089. In this experiment we see the errors in the beginning of the assimilation window are at their lowest, while the rest of the errors are spread across the rest of the assimilation window. The low value of σ_b and high values of σ_q, both in the assimilation and the truth, are responsible for this. We observed this behaviour in Experiment 2a, Figure 4.13, where the high value of σ_b and low value of σ_q caused the errors to be spread inversely to what is shown in Figure 4.14 here.

4.2.2.3 Experiment 2b (i): Large Model Error Variance

The experiment settings are identical to the previous experiment except that we set $\sigma_b = 0.1$ and $\sigma_q = 10$. The model error variance in comparison to the background error variance in this experiment is large, which is not a likely situation that would arise in NWP. The purpose is to highlight the sensitivities of the minimisation problems (2.52) (2.53) to these parameter settings.

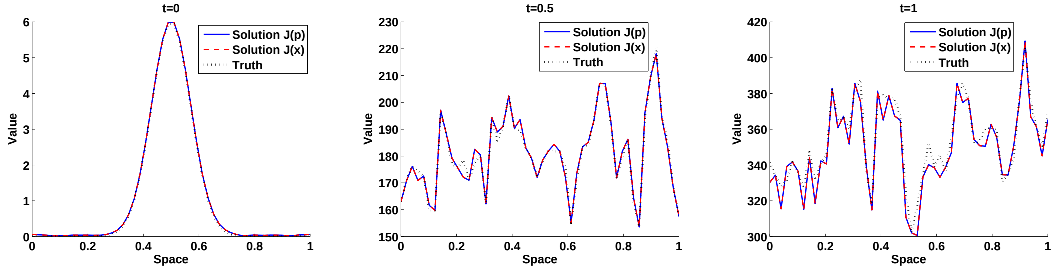


Figure 4.15: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.15 shows that the solutions are of similar quality. The problem is more demanding since the variance of the model errors are much larger now. Even with the power of wc4DVAR to closely match the trajectory inside the assimilation window, both solutions are noticeably missing the truth because the true model errors are considerably large.

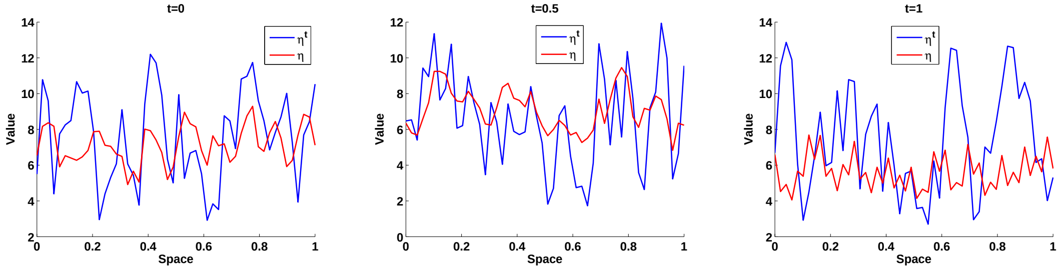


Figure 4.16: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

In Figure 4.16 we see the variance of the estimated model error is again not quite as large as the true model error. On the final time step the variance of the true model error is more than twice as large as the range of the estimated model error.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	1.09×10^7	341
$\mathbf{S}_x$	1.88×10^7	972
$\mathbf{D}$	2.13×10^6	-

Table 4.8: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.8 shows the numerical condition number of $\mathbf{S}_x$ to be nearly double that of

$\mathbf{S}_p$. Similarly, the minimisation of $\mathcal{J}(\mathbf{x})$ requires more than double the number of iterations compared to $\mathcal{J}(\mathbf{p})$.

Ratio	**Value**
σ_b/σ_q	10^{-2}
σ_b/σ_o	2
σ_q/σ_o	200

Table 4.9: Assimilation error variance ratios.

The numerical condition number of $\mathbf{D}$ is of order $\mathcal{O}(10^6)$ as in Experiment 2a, which is expected since the ratio σ_b/σ_q is the same as the inverse of the ratio used in Experiment 2a. The effect of this ratio on the largest and smallest eigenvalues of $\mathbf{D}$ is identical. We note however that the condition numbers of both Hessians and the ratio σ_q/σ_o are large. Although the numerical condition number of both Hessians are large, the number of iterations of $\mathcal{J}(\mathbf{p})$ is not as heavily affected as $\mathcal{J}(\mathbf{x})$.

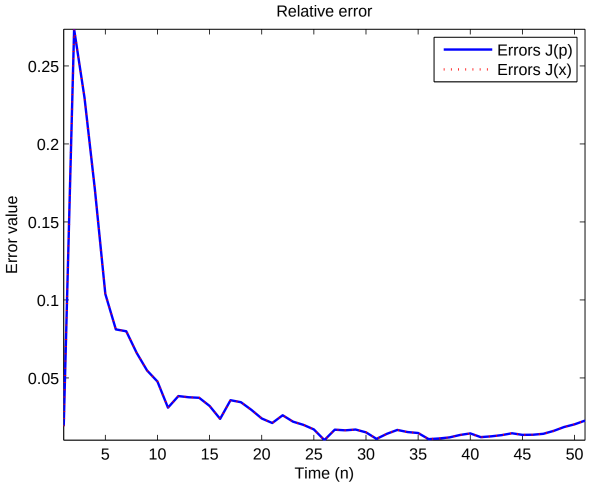


Figure 4.17: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.17 shows that the errors in both algorithms are identical, with total relative errors of 0.457. We also observe a noticeable spike in the errors near the beginning of the window. The errors here are quite large in comparison to previous experiments so far.

We now reduce the model error variance and examine its effect on the minimisation of both wc4DVAR problems.

4.2.2.4 Experiment 2b (ii): Small Model Error Variance

In this experiment we use the same parameters as the previous experiment except we change the model standard devation from $\sigma_q = 10$ to $\sigma_q = 5 \times 10^{-4}$. We now discuss the effect this has on the assimilation.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	7.85×10^3	182
$\mathbf{S}_x$	1.57×10^6	2693
$\mathbf{D}$	1.41×10^6	-

Table 4.10: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.10 shows the minimisation of $\mathcal{J}(\mathbf{x})$ requiring over 15 times as many iterations as $\mathcal{J}(\mathbf{p})$. The numerical condition number of $\mathbf{S}_x$ and $\mathbf{D}$ are 3 orders of magnitude higher than $\mathbf{S}_p$, which complements the difference in the number of iterations. We also see that the numerical condition number of $\mathbf{S}_x$ is of the same order of magnitude as $\mathbf{D}$.

Ratio	Value
σ_b/σ_q	200
σ_b/σ_o	2
σ_q/σ_o	0.01

Table 4.11: Assimilation error variance ratios.

The high σ_b/σ_q value is the reason for the high condition number of $\mathbf{D}$, since they increase the distance between the extrema eigenvalues of $\mathbf{D}$.

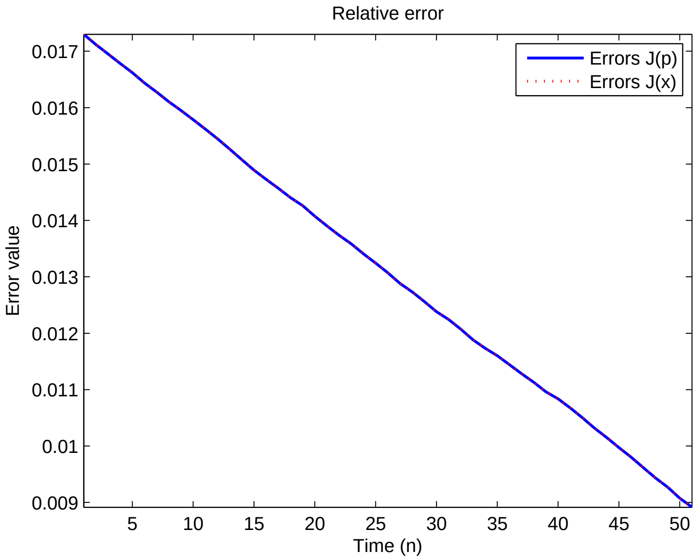


Figure 4.18: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.18 we see that the errors are identical again with total relative error values for both formulations at 0.095, while the distribution of errors is linear and differs from the previous experiment, Figure 4.17. The bulk of the errors are in the beginning of the assimilation window, which linearly decrease until final time. The errors are largest at the beginning of the window because the size of the background error variance σ_b is large relative to σ_q.

We now examine the effects of the observation error variance.

4.2.2.5 Experiment 2c (i): Large Observation Error Variance

The experiment settings identical to the previous experiment with the exception of, the background standard deviation, $\sigma_b = 0.1$, model standard deviation, $\sigma_q = 0.05$ and increased observation standard deviation $\sigma_o = 10$, thus yielding the following error variance ratios:

Ratio	Value
σ_b/σ_q	2
σ_b/σ_o	0.01
σ_q/σ_o	5×10^{-3}

Table 4.12: Assimilation error variance ratios.

We now present the time series plots of the solution with the truth

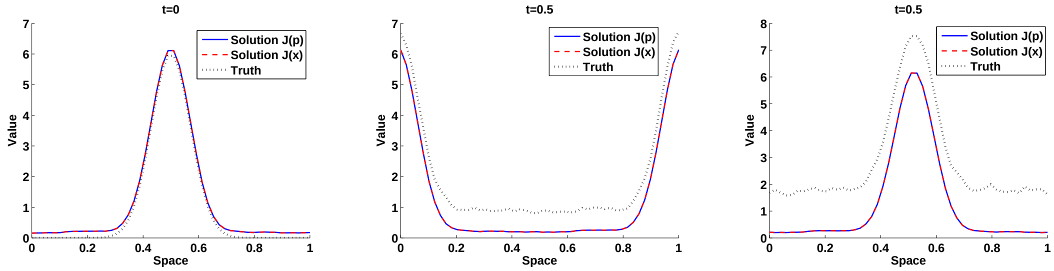


Figure 4.19: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.19 shows that both solutions are showing visually noticeable shortfalls at this scale, even with the truth and assimilation error settings being identical. This is mainly due to the σ_q parameter being too restrictive and not allowing for manoeuvrability of solution choice for each algorithm to fit the observation data.

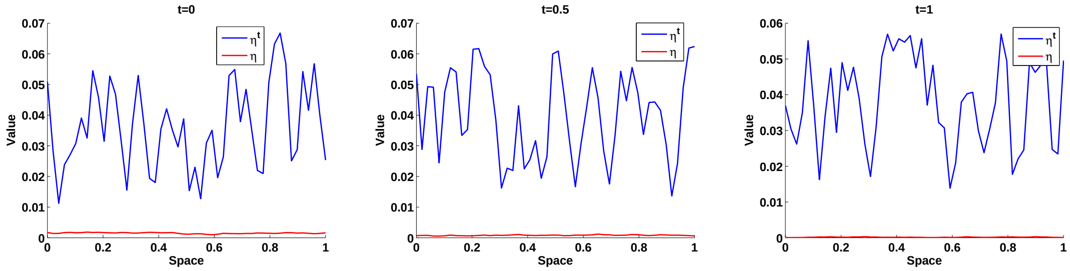


Figure 4.20: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

Figure 4.20 shows that the $\mathcal{J}(\mathbf{p})$ formulation was unable to quantify reasonable η_i's because the model error variance was too restrictive, even though it reflected the model error variance selected to generate the truth.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	834	87
$\mathbf{S}_x$	1.11×10^5	2821
$\mathbf{D}$	838	-

Table 4.13: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.13 shows that the minimisation of $\mathcal{J}(\mathbf{x})$ requires nearly 25 as many iterations as $\mathcal{J}(\mathbf{p})$ to converge on a solution of similar quality. The increases in condition numbers are similar to the increase in iterations, with the condition number of $\mathbf{S}_x$ being 3 orders of magnitude higher.

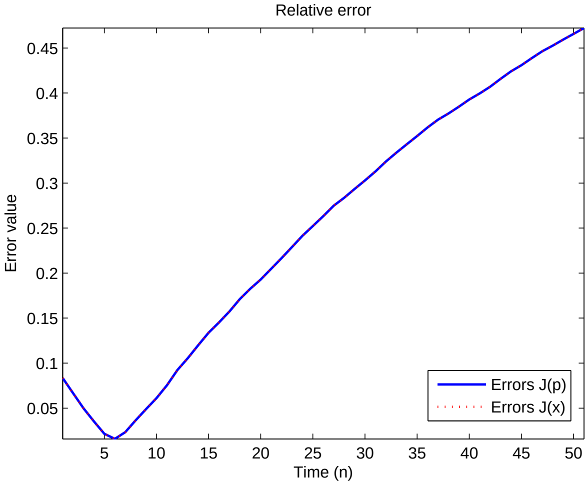


Figure 4.21: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.21 shows that the errors in both algorithms are identical and also very large in comparison to all previous experiments thus far, with total relative errors of 2.067. This is mainly due to the very large observation error variance $\sigma_o^2 = 100$. We can see from this experiment that if σ_o is large relative to σ_b and σ_q, then the wc4DVAR algorithms cannot yield solutions which fit the observations well.

4.2.2.6 Experiment 2c (ii): Small Observation Error Variance

In this experiment we use the same parameters except we change the observation standard devation from $\sigma_o = 10$ to $\sigma_o = 5 \times 10^{-4}$, yielding the following error variance ratios

Ratio	**Value**
σ_b/σ_q	2
σ_b/σ_o	200
σ_q/σ_o	100

Table 4.14: Assimilation error variance ratios.

We now discuss the effect this has on the assimilation.

Matrix	**Numerical Condition No.**	**No. of iterations**
$\mathbf{S}_p$	2.71×10^6	191
$\mathbf{S}_x$	1.84×10^5	176
$\mathbf{D}$	838	-

Table 4.15: Numerical condition numbers. The size of all square matrices in table: 2550.

In Table 4.15 we see that the number of iterations of both algorithms is similar, with the minimisation of $\mathcal{J}(\mathbf{x})$ requiring less iterations for the first time in our experiments. The condition numbers of both $\mathbf{S}_p$ and $\mathbf{S}_x$ are high and it is clear that $\mathbf{D}$ is not the contributor. We would have expected the difference in iterations between both algorithms to be higher since there is an order of magnitude of difference in their condition numbers. This is an example of the condition number not being an exact indication of the iterative performance of an algorithm.

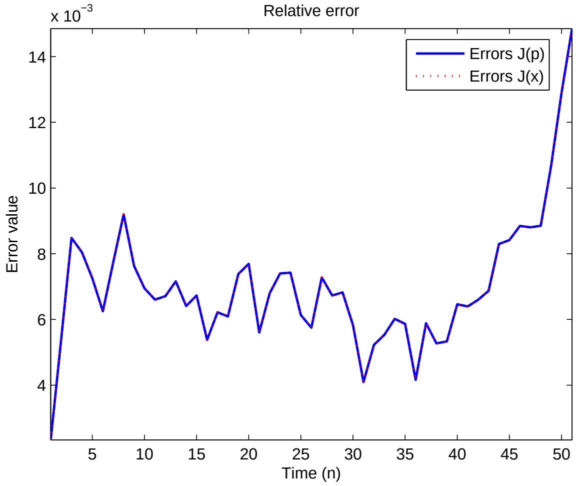


Figure 4.22: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

In Figure 4.22 we see that both algorithms have the same error distribution with total relative solution errors of 0.052. The relative errors are low and within the range of $\mathcal{O}(10^{-3})$ consistently in contrast to Experiment 2c(i), Figure 4.21, which has much higher errors due to the larger observation model error variance chosen. So even though the decrease in observation error variance has increased the condition number and number of iterations, the relative errors of both solutions have dropped.

We now summarise the error variance balance experiments.

4.2.2.7 Summary

A large background error does not affect the minimisation problem of the assimilation as much as a small one, which can be seen in Experiment 2a. We also see clear evidence that the minimisation of both $\mathcal{J}(\mathbf{x})$ and $\mathcal{J}(\mathbf{p})$ are sensitive to smaller background error, since the condition number is higher and the iterations dramatically increase. We also see evidence of $\mathcal{J}(\mathbf{x})$ being more sensitive to this change, exhibiting larger condition number by two orders of magnitude and taking almost 3 times as many iterations as $\mathcal{J}(\mathbf{p})$ to converge.

The experiments we have considered with model errors show that even with model errors larger than the background error, both algorithms can still solve the problem relatively well, as seen in Figure 4.18. However this comes at the cost of increased condition numbers and iterations for both $\mathcal{J}(\mathbf{x})$ and $\mathcal{J}(\mathbf{p})$, where $\mathcal{J}(\mathbf{x})$ exhibits the most sensitivity in terms of iterations to convergence. When the model error is small the problem becomes less demanding in general, and both algorithms solve to much improved accuracy as seen from Figure 4.18. But it is clearly evident that $\mathcal{J}(\mathbf{x})$ is far more sensitive to changes in σ_q, more so when σ_q is very small, which can be seen from the number of iterations required for convergence and condition number, Table 4.10 in Experiment 2b.

Inaccurate observations (large σ_o) results in an ill-conditioned $\mathbf{S}_x$ matrix and a well-conditioned $\mathbf{S}_p$ matrix. We also see this in the high iteration number for $\mathcal{J}(\mathbf{x})$ over $\mathcal{J}(\mathbf{p})$. The quality of the solutions are almost identical, albeit a more strenuous task for $\mathcal{J}(\mathbf{x})$, as shown by the number of iterations and condition number in Table 4.13. Although there is a clear difference in the minimisation iterations and condition numbers, as well as an obvious shortfall between the truth and the solutions provided by both algorithms, especially at the end of the assimilation window, as seen in Experiment 2c (i). For more accurate observations however, $\mathbf{S}_p$ has a condition number larger than that of $\mathbf{S}_x$ by an order of magnitude, as seen in Experiment 2c (ii), while $\mathcal{J}(\mathbf{x})$ requires slightly less iterations.

We now perform the final experiment which examines the effect of longer assimilation windows.

4.2.3 Experiment 3: Assimilation Window Length

Longer assimilation windows mean incorporating more observations and increasing the difficulty of the data assimilation problem. It is believed that longer assimilation windows are beneficial for longer-validity of weather forecasts, [84], [27], [26]. With this in mind we compose an appropriate experiment and examine the effect of a longer assimilation window.

In previous experiments in this chapter we had an assimilation window which allowed the advection model to propagate the Gaussian curve far enough through the domain so it passes by its original percevied position, we denote this as one period. In the following experiment we lengthen the assimilation window to allow for the Gaussian curve to pass its original starting position 5 times. We reduce the spatial resolution so that the Hessian matrix remains a reasonable size for an accurate numerical condition number calculation. The model settings are such that the spatial domain is size $N = 25$ with a spatial resolution of $\Delta x = 0.04$. The time-intervals are $\Delta t = 0.04$ and the wave speed is $a = -1$, yielding a Courant number of $\mu = -1$.

We choose the the background error, $B_0 = \sigma_b^2 C_{SOAR}$, such that the correlation length-scale $L = 2\Delta x = 0.08$ and $\sigma_b = 0.5$. The model error, $Q_i = \sigma_q^2 C_{LAP}$ is such that the correlation length-scale $L = \Delta x = 0.04$ and $\sigma_q = 0.3$. The observation error is such that $R_i = \sigma_o^2 I$, where $\sigma_o = 0.1$. We take observations every $\Delta q = 5$ model time-steps and the domain is *fully observed*. We take a longer assimilation window of $n = 100$ here, meaning we go through 500 model time-steps, since we observe every 5'th model time-step starting with the first time-step. The iterative tolerance is *reduced* to $\tau = 10^{-5}$.

We now present the time series plots.

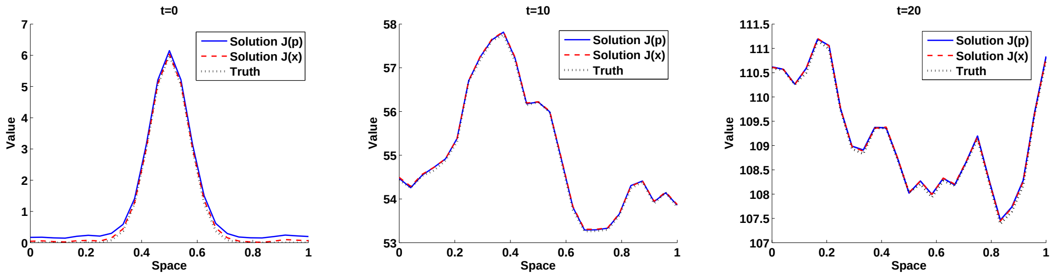


Figure 4.23: Assimilation window time series left to right, $t = 0$, $t = n/2$ and $t = n$. Truth (black-dashed line), wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.23 shows that both solutions are very closely matching the truth. We notice the Gaussian curve has moved upwards and deformed considerably over time, since the assimilation window is now much longer and the model has more time to evolve the initial state. We can also see that some finer details of the Gaussian curve structure have been missed by both solutions.

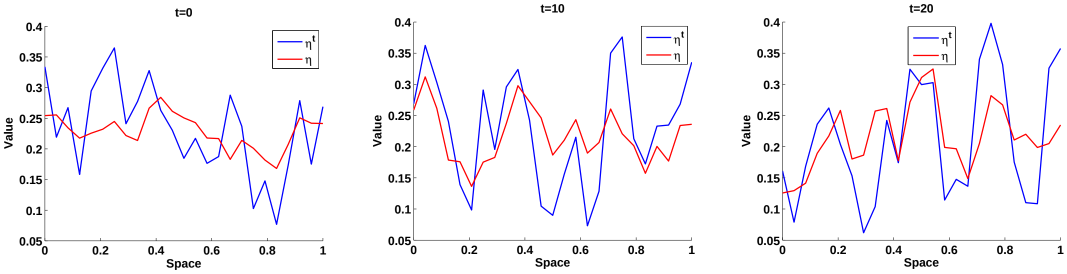


Figure 4.24: Model error time series left to right, $t = 0$, $t = n/2$ and $t = n$. Estimated model error (red line) using wc4DVAR $\mathcal{J}(\mathbf{p})$. True model error (blue line).

Figure 4.24 agrees with Figure 4.23 in that the $\mathcal{J}(\mathbf{p})$ formulation has mimicked the truth. The estimated model errors have a much improved error variance than in previous experiments. It is likely that the longer assimilation window has improved the estimates of the model error.

Matrix	Numerical Condition No.	No. of iterations
$\mathbf{S}_p$	6.13×10^4	71
$\mathbf{S}_x$	1.66×10^3	42
$\mathbf{D}$	878	-

Table 4.16: Numerical condition numbers and iteration count of respective objective function minimisations.

Table 4.16 shows that $\mathcal{J}(\mathbf{p})$ requires nearly twice as many iterations as $\mathcal{J}(\mathbf{x})$ to converge on an equivalent solution. The condition number of $\mathbf{S}_p$ is an order of magnitude higher than $\mathbf{S}_x$. This is not proportional to the increase in iterations, but we see a simultaneous increase in condition number and iteration count of $\mathcal{J}(\mathbf{p})$ over $\mathcal{J}(\mathbf{x})$, further reinforcing the possibility of $\mathcal{J}(\mathbf{p})$ being more sensitive to assimilation window length than $\mathcal{J}(\mathbf{x})$.

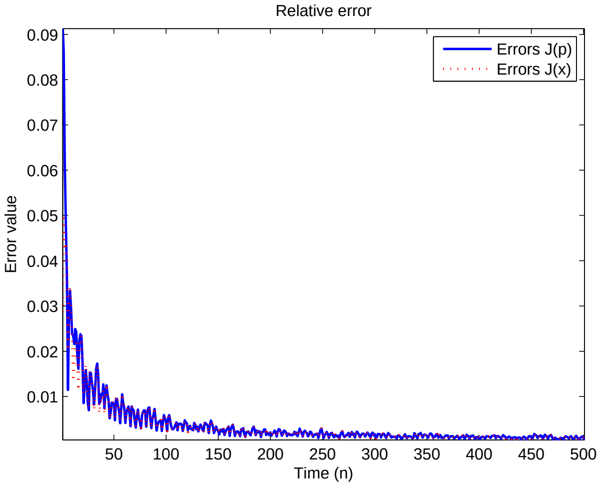


Figure 4.25: Assimilation relative error calculations. Errors in wc4DVAR $\mathcal{J}(\mathbf{x})$ solution (red line), wc4DVAR $\mathcal{J}(\mathbf{p})$ solution (blue line).

Figure 4.25 shows that the errors in $\mathcal{J}(\mathbf{p})$ are slightly higher, with a total relative error of 0.196, whereas $\mathcal{J}(\mathbf{x})$ has a total relative error of 0.153. The relative errors are low with the exception of the beginning of the assimilation window.

Summary

This experiment shows that the length of the assimilation window, while it affects both algorithms, has a more profound effect on the minimisation of $\mathcal{J}(\mathbf{p})$, through an increased Hessian condition number and iterations. The $\mathcal{J}(\mathbf{x})$ formulation performs better in this experiment in terms of condition number, number of iterations and relative solution error, with a fully observed domain.

4.3 Conclusions

In this chapter we detailed the design of the weak-constraint variational system along with the tests to ensure its numerical validity. We then explained our reasoning behind the choice of observation configuration and model setup to carry out the experiments. The experiments were carried out on a simple 1-dimensional linear system using correlated background and model error covariances and regular observation spacing to enable us to study the effects of different parameter settings on the minimisation process. The experiment results showed the following:

1. The $\mathcal{J}(\mathbf{x})$ formulation is more sensitive to lower observation density than $\mathcal{J}(\mathbf{p})$. The $\mathcal{J}(\mathbf{x})$ formulation takes longer to converge onto an identical quality solution to $\mathcal{J}(\mathbf{p})$ with the same settings. The Hessian condition number of $\mathcal{J}(\mathbf{x})$ is also higher than that of $\mathcal{J}(\mathbf{p})$. This is shown in Experiments 1a and 1b.

2. The $\mathcal{J}(\mathbf{x})$ formulation is more sensitive than $\mathcal{J}(\mathbf{p})$ to the balance of model errors with background errors. This can be seen from findings in Experiments 2a and 2b.

   (a) Experiment 2a shows that $\mathcal{J}(\mathbf{x})$ is sensitive to changes in the background error, more so when the background error is small. This is seen in the number of iterations only.

   (b) Experiment 2b shows the increased sensitivity of $\mathcal{J}(\mathbf{x})$ over $\mathcal{J}(\mathbf{p})$ for small model error variances σ_q. This is seen in the condition number and the number of iterations required for convergence.

   (c) Experiment 2c shows that a large observation error variance dramatically increases the number of iterations required by $\mathcal{J}(\mathbf{x})$ to converge. The condition number is also very large, of order 5 times larger than the condition number of $\mathbf{S}_p$. We see that for a small observation error variance, the $\mathcal{J}(\mathbf{x})$ formulation takes less iterations to converge than $\mathcal{J}(\mathbf{p})$ for the first time, albeit not by a significant amount.

3. The $\mathcal{J}(\mathbf{p})$ formulation is more sensitive than $\mathcal{J}(\mathbf{x})$ to assimilation window length where the spatial domain is fully observed, shown in Experiment 3.

4. Another more general conclusion about wc4DVAR is that the variance of the estimated model errors provided by the solutions of both $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$ were consistently under-estimated in comparison to the true model errors. This can be seen from Experiments 1a, 1b, 2a and 2b. However, the estimation

of the model error variance by both algorithms was noticeably improved in Experiment 3, with a longer assimilation window.

The aim is to gain a deeper theoretical understanding into the behaviour of both the minimisation problems presented by $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$. In the next chapter we bound the condition number of the Hessian of $\mathcal{J}(\mathbf{p})$ and analyse it more rigorously.

Chapter 5

Conditioning and Preconditioning of the Model Error Formulation: $\mathcal{J}(\mathbf{p})$

In the previous chapter we examined the effect that various assimilation parameters had on the iterative solution process of the wc4DVAR problem when applied to the 1D advection equation.

The results in this chapter extend the results in [41], where the author bounded the condition number of the 3DVAR Hessian and then the Hessian of the strong-constraint 4DVAR objective function, denoted as S. We have derived a general result linking the condition numbers of the sc4DVAR Hessian S and the wc4DVAR Hessian $\mathbf{S}_p$ such that,

$$\kappa(S) \leq \kappa(\mathbf{S}_p), \tag{5.1}$$

with no assumptions. This result shows that the condition number of the Hessian of sc4DVAR can *never* exceed the condition number of the Hessian of the wc4DVAR $\mathcal{J}(\mathbf{p})$ formulation for identical assimilation problems. Assuming that the condition number is a good measure for iterative performance, this result indicates that the iterative solution process of the wc4DVAR problem will only be as good as the solution process of the sc4DVAR. The proof of this result is contained in Appendix A.

In this chapter we present new theoretical bounds on the condition number of

$$\mathbf{S}_p = \begin{pmatrix} B_0^{-1} & & & \\ & Q_1^{-1} & & \\ & & \ddots & \\ & & & Q_n^{-1} \end{pmatrix} +$$
$$\begin{pmatrix} \sum\limits_{i=0}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=1}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=2}^{n}(H_iM_{i,0})^TR_i^{-1}H_iM_{i,2} & \dots & (H_nM_{n,0})^TR_n^{-1}H_n \\ \sum\limits_{i=1}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=1}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=2}^{n}(H_iM_{i,1})^TR_i^{-1}H_iM_{i,2} & \dots & (H_nM_{n,1})^TR_n^{-1}H_n \\ \sum\limits_{i=2}^{n}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,0} & \sum\limits_{i=2}^{n}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,1} & \sum\limits_{i=0}^{n-2}(H_iM_{i,2})^TR_i^{-1}H_iM_{i,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & (H_nM_{n,n-1})^TR_n^{-1}H_n \\ H_n^TR_n^{-1}H_nM_{n,0} & H_n^TR_n^{-1}H_nM_{n,1} & \dots & H_n^TR_n^{-1}H_nM_{n,n-1} & H_n^TR_n^{-1}H_n \end{pmatrix} \tag{5.2}$$

and its preconditioned counter-part

$$\hat{\mathbf{S}}_p = \mathbf{I} + \mathbf{D}^{1/2}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2}. \tag{5.3}$$

The eigenvalue spectrum of these matrices are not explicitly known and in practice they are too computationally expensive to calculate explicitly. So we take the route of estimating the condition number of the Hessian by bounding it in order to obtain information from the expressions yielded by the bounds. We utilise the bounds to gain insight into the Hessian condition number sensitivities of the objective function $\mathcal{J}(\mathbf{p})$ and its preconditioned counter-part.

We first derive bounds on the condition number of the Hessian $\mathbf{S}_p$ with some simple assumptions on the observations. The assumptions become more specific with each theorem. The first theorem assumes general correlation structures for the background, observation and model errors while assuming there are fewer observations than the dimension of state space. The second theorem derives bounds that are more specific to a particular class of covariance and model matrices, whereas the final theorem is specific to the advection equation. We then take the preconditioned Hessian of objective function $\mathcal{J}(\mathbf{p})$ and bound its condition number. We then show the improvement in overall conditioning and minimisation iteration rates of the preconditioned problem compared to the original problem.

The insight gained from the bounds are demonstrated through numerical experiments on the condition number. We also further demonstrate the condition number sensitivities obtained from the bounds by examining their effect on the convergence rate of the model error estimation an preconditioned model error estimation minimisation problems.

We now present the theoretical bounds.

5.1 Theoretical Results: Bounding the Condition Number of $\mathbf{S}_p$

The following result bounds the spectral condition number of $\mathbf{S}_p$

Theorem 5.1.1 *Let $B_0 \in \mathbb{R}^{N\times N}$ and $Q_i \in \mathbb{R}^{N\times N}$ for $i = 1,..,n$ be the background and model error covariance matrices respectively, so $\mathbf{D} \in \mathbb{R}^{N(n+1)\times N(n+1)}$. Suppose we take $q < N$ observations at each time interval t_i for $i = 0,...,n$ with observation error covariance matrix $R_i \in \mathbb{R}^{q\times q}$, so $\mathbf{R} \in \mathbb{R}^{q(n+1)\times q(n+1)}$. Let $H_i \in \mathbb{R}^{q\times N}$ for $i = 0,..,n$, be the observation operator, so $\mathbf{H} \in \mathbb{R}^{q(n+1)\times N(n+1)}$. Finally, let $M_{i,i-1} \in \mathbb{R}^{N\times N}$ for $i = 1,..,n$, represent the model operator and $\mathbf{L} \in \mathbb{R}^{q(n+1)\times N(n+1)}$ represent the 4D weak-constraint model propagator. Then the following bounds are satisfied by the spectral condition number of $\mathbf{S}_p$:*

$$\frac{\kappa(\mathbf{D})}{(1+\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})\lambda_{max}(\mathbf{D}))} \leq \kappa(\mathbf{S}_p)$$
$$\leq \kappa(\mathbf{D})\left(1+\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})\lambda_{min}(\mathbf{D})\right).$$

Proof: We use Theorem 3.4.2 to bound $\lambda_{min}(\mathbf{S}_p)$ and $\lambda_{max}(\mathbf{S}_p)$, yielding

$$\lambda_{min}(\mathbf{D}^{-1}) + \lambda_{min}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}) \leq \lambda_{min}(\mathbf{S}_p)$$
$$\leq \lambda_{min}(\mathbf{D}^{-1}) + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}) \quad (5.4)$$

and

$$\lambda_{max}(\mathbf{D}^{-1}) + \lambda_{min}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}) \leq \lambda_{max}(\mathbf{S}_p) \leq \lambda_{max}(\mathbf{D}^{-1}) + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}). \quad (5.5)$$

We then take the upper bound of $\lambda_{max}(\mathbf{S}_p)$ and lower bound of $\lambda_{min}(\mathbf{S}_p)$ giving us the following upper bound on the condition number,

$$\kappa(\mathbf{S}_p) \leq \frac{\lambda_{max}(\mathbf{D}^{-1}) + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})}{\lambda_{min}(\mathbf{D}^{-1}) + \lambda_{min}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})}. \quad (5.6)$$

Similarly for the lower bound we take the lower bound of $\lambda_{max}(\mathbf{S}_p)$ and upper bound of $\lambda_{min}(\mathbf{S}_p)$, which yields the following lower bound on the condition number,

$$\kappa(\mathbf{S}_p) \geq \frac{\lambda_{max}(\mathbf{D}^{-1}) + \lambda_{min}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})}{\lambda_{min}(\mathbf{D}^{-1}) + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})}. \quad (5.7)$$

Since we assumed fewer observations than the number of states, ie $q < N$, implying $\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}$ is a singular matrix with zero eigenvalues, since it is rank deficient. We also know that $(\lambda_{max}(\mathbf{D}^{-1}))^{-1} = \lambda_{min}(\mathbf{D})$. Now if we combine (5.6) and (5.7), we arrive at

$$\frac{\kappa(\mathbf{D})}{(1 + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})\lambda_{max}(\mathbf{D}))} \leq \kappa(\mathbf{S}_p) \leq \kappa(\mathbf{D})\left(1 + \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})\lambda_{min}(\mathbf{D})\right), \quad (5.8)$$

as required. ■

We observe the presence of the condition number of the background and model error covariance matrix **D** in both the upper and lower bounds. This is an early strong indication that the condition number of $\mathbf{S}_p$ will be heavily influenced by $\kappa(\mathbf{D})$. We cannot interpret anything further from the bounds in this theorem. So we now make our assumptions more specific in a bid to uncover more definitive expressions from the later bounds.

Theorem 5.1.2 *Let $B_0 = \sigma_b^2 C_B \in \mathbb{R}^{N\times N}$ be the background error covariance matrix, where C_B is a symmetric, positive-definite* circulant *correlation matrix and $\sigma_b^2 > 0$ is the background error variance. Let $Q_i = Q = \sigma_q^2 C_Q \in \mathbb{R}^{N\times N}$ be* *the* time invariant *model error covariance matrix, for $i = 1, ..., n$, where C_Q is a symmetric, positive-definite* circulant *correlation matrix and $\sigma_q^2 > 0$ is the model error variance. Assume $q < N$ observations are taken with the same error variance $\sigma_o^2 > 0$ at each time interval such that $R_i = R = \sigma_o^2 I_q$ for $i = 0, ..., n$, where I_q is a $q \times q$ identity matrix. Assume that observations of the parameter are made at the same grid points at each time interval such that $H_i^T H_i = H^T H \in \mathbb{R}^{N\times N}$, so $H^T H$ is a diagonal matrix with unit entries at observed points and zeros otherwise. Finally, we assume that $M_{i,i-1} = M \in \mathbb{R}^{N\times N}$ for $i = 1, .., n$ is a* circulant *matrix, and $M_{i,i} = I_N$. The following bounds are satisfied by the condition number of $\mathbf{S}_p$:*

$$\left(\frac{1+\frac{q}{N}\frac{min\left\{\sigma_b^2\lambda_{min}(C_B),\sigma_q^2\lambda_{min}(C_Q)\right\}}{\sigma_o^2}\psi_{min}}{1+\frac{q}{N}\frac{max\left\{\sigma_b^2\lambda_{max}(C_B),\sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}\psi_{max}}\right)\kappa(\mathbf{D}) \leq \kappa(\mathbf{S}_p)$$

$$\leq \kappa(\mathbf{D})\left(1+\frac{min\left\{\sigma_b^2\lambda_{min}(C_B),\sigma_q^2\lambda_{min}(C_Q)\right\}}{\sigma_o^2}\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1})\right), \quad (5.9)$$

where

$$\psi_l = \begin{cases} \sum\limits_{k=0}^{n}|\lambda_l|^{2k} & \text{if } \lambda_l(\mathbf{D}) = \lambda_l(B_0) \\ \frac{1}{n}\sum\limits_{i=1}^{n-1}\sum\limits_{j=0}^{i}\sum\limits_{k=0}^{n-i}\left(2Re(\lambda_l^j)-1\right)\cdot|\lambda_l|^{2k} & \text{if } \lambda_l(\mathbf{D}) = \lambda_l(Q) \end{cases}. \quad (5.10)$$

The eigenvalue of M is denoted by λ in (5.10) and the subscript l denotes the largest or smallest eigenvalue (max/min) respectively.

Proof: We begin by noticing that as a direct consequence of the assumptions, we have

$$\kappa(\mathbf{D}) = \frac{\lambda_{max}(\mathbf{D})}{\lambda_{min}(\mathbf{D})} = \frac{max\left\{\sigma_b^2\lambda_{max}(C_B),\sigma_q^2\lambda_{max}(C_Q)\right\}}{min\left\{\sigma_b^2\lambda_{min}(C_B),\sigma_q^2\lambda_{min}(C_Q)\right\}}. \quad (5.11)$$

Furthermore, we recognise that $\mathbf{D}$ has $N(n+1)$ eigenvalues and eigenvectors where exactly Nn of the eigenvalues are repeated since Q is time invariant. The eigenvectors of Q constitute the non-zero components of the Nn repeated eigenvectors of D. We also know that since the constituent matrices of $\mathbf{D}$ are circulant, and M is also circulant, they possess the same orthogonal eigenvectors as in Chapter 3, Theorem 3.37 equation (3.37).

With this in mind we choose a vector, $V_k \in \mathbb{R}^{N(n+1)}$ such that

$$V_k = \begin{pmatrix} v_k \\ v_k \\ \vdots \\ v_k \end{pmatrix}, \tag{5.12}$$

where $v_k \in \mathbb{R}^N$ is an arbitrary eigenvector of a circulant matrix. We apply the Rayleigh quotient using (5.12) to obtain the lower bound of $\mathbf{S}_p$. We begin by considering the second term of $\mathbf{S}_p$

$$\frac{1}{\sigma_o^2}\frac{V_k^H[\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}]V_k}{V_k^H V_k}, \tag{5.13}$$

while deliberately omitting $\mathbf{D}$ for now.

The denominator of (5.13) yields

$$V_k^H V_k = n + 1, \tag{5.14}$$

since the eigenvectors of a circulant matrix are orthogonal, Theorem 3.37. The computation in (5.13) requires v_k and v_k^H to multiply every matrix block inside $\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}$. Each block multiplication yields the following:

$$v_k^H (M^j)^T = v_k^H \bar{\lambda}_\alpha^j(M), \tag{5.15}$$

$$(M^j) v_k = \lambda_\alpha^j(M) v_k, \tag{5.16}$$

where $\lambda_\alpha^j(M)$ is some eigenvalue of M and $\bar{\lambda}_\alpha^j(M)$ is the corresponding complex conjugate eigenvalue of M. We write $\lambda_\alpha(M) = \lambda_\alpha$ for convenience.

Substituting (5.14), (5.15) and (5.16) into (5.13), we obtain the following series:

$$\frac{1}{n+1}\left[\sum_{i=0}^{n}\sum_{j=0}^{n-i}(\bar{\lambda}_\alpha)^j(\lambda_\alpha)^j v_k^H H^T H v_k + \sum_{i=1}^{n}\sum_{j=0}^{n-i}(\bar{\lambda}_\alpha)(\bar{\lambda}_\alpha)^j(\lambda_\alpha)^j v_k^H H^T H v_k \right.$$
$$\left. + \sum_{i=1}^{n}\sum_{j=0}^{n-i}(\lambda_\alpha)(\bar{\lambda}_\alpha)^j(\lambda_\alpha)^j v_k^H H^T H v_k + \cdots + \cdots + (\bar{\lambda}_\alpha)^n v_k^H H^T H v_k + (\lambda_\alpha)^n v_k^H H^T H v_k \right], \tag{5.17}$$

where the first term in the *geometric* series (5.17) comes from the main diagonal of (5.13). The second term of (5.17) is from the upper off-diagonal block entries of (5.13) and the third term is from the lower off-diagonal block entries. This pattern continues until the final term in the bottom right hand corner of (5.13), which coincides with the final term in (5.17).

We now compute each of the terms in the series above. We have

$$v_k^H H^T H v_k = \frac{q}{N}, \tag{5.18}$$

since circulant matrices have orthogonal eigenvectors and $H^T H$ is a square matrix with q unit entries on the main diagonal at positions of observation and 0 elsewhere. We also know the following to be true:

$$(\bar{\lambda}_\alpha)(\lambda_\alpha) = |\lambda_\alpha|^2. \tag{5.19}$$

Substituting (5.18) and (5.19) into (5.17) we arrive at

$$\frac{q}{N}\frac{1}{n+1}\sum_{i=0}^{n}\sum_{k=0}^{n-i}\sum_{j=0}^{i}\left(\lambda_\alpha^j + \bar{\lambda}_\alpha^j - 1\right)\cdot|\lambda_\alpha|^{2k}. \tag{5.20}$$

We define a new parameter such that

$$\psi_\alpha = \frac{1}{n+1}\sum_{i=0}^{n}\sum_{k=0}^{n-i}\sum_{j=0}^{i}\left(2Re(\lambda_\alpha^j) - 1\right)\cdot|\lambda_\alpha|^{2k}. \tag{5.21}$$

Now that (5.21) represents the general expression for the computation of (5.13), we reselect vector (5.12) to yield the largest possible value. Estimating the largest possible lower bound yields the most optimum estimate and therefore the tightest bound.

The extreme eigenvalues of $\mathbf{D}$ are related to B_0 and Q as in (5.11). Now let us consider a vector utilising the eigenvectors associated with $\lambda_{max/min}(\mathbf{D})$ such that

$$V_{max/min} = \begin{cases} [\boldsymbol{\beta}_{max/min}^T, 0, \ldots, 0]^T & \text{if } \lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(B_0) \\ [0, \boldsymbol{\xi}_{max/min}^T, \ldots, \boldsymbol{\xi}_{max/min}^T]^T & \text{if } \lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(Q). \end{cases}, \tag{5.22}$$

where $\boldsymbol{\beta}_{max/min}$ and $\boldsymbol{\xi}_{max/min}$ denote the eigenvector associated with the maximum or minimum eigenvalue of the respective matrix.

We consider the Rayleigh quotient as in (5.13) but for the vector $V_{max/min}$, since the Rayleigh quotient of $\mathbf{D}$ yields the respective extreme eigenvalues for $V_{max/min}$. The denominator of the Rayleigh quotient as in (5.13) will yield

$$V_{max/min}^{H} V_{max/min} = \begin{cases} 1 & \text{if } \lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(B_0) \\ n & \text{if } \lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(Q) \end{cases} . \tag{5.23}$$

It also follows that series (5.21) will have a reduced number of terms since the vector $V_{max/min}$ now has some zero entries, whereas the general vector chosen in (5.12) did not. We compute the two possible cases of series (5.21) below:

1. If $\lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(B_0)$ then vector-matrix multiplication in (5.13) will only yield the uppermost left corner block of $\mathbf{L}^{-T}\mathbf{H}^{T}\mathbf{H}\mathbf{L}^{-1}$, which by no coincidence is identical to the sc4DVAR Hessian term $\hat{\mathbf{H}}^{T}\hat{\mathbf{H}}$ as in [41] (Chapter 7, Theorem 7.1.2). Therefore (5.21) becomes

$$\psi_{l/m} = \sum_{k=0}^{n} |\lambda_{l/m}|^{2k}, \tag{5.24}$$

where the l and m are separate subscripts denoting the l^{th} and m^{th} eigenvalues of M.

2. If $\lambda_{max/min}(\mathbf{D}) = \lambda_{max/min}(Q)$ then we would obtain all the terms of $\mathbf{L}^{-T}\mathbf{H}^{T}\mathbf{H}\mathbf{L}^{-1}$ *excluding* the first row and first column blocks. So (5.21) yields

$$\psi_{g/h} = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{k=0}^{n-i} \sum_{j=0}^{i} \left(2Re(\lambda_{g/h}^{j}) - 1 \right) \cdot |\lambda_{g/h}|^{2k}, \tag{5.25}$$

where again, the g and h are separate subscripts denoting the g^{th} and h^{th} eigenvalues of M.

We utilise the eigenvalue range of the Rayleigh Quotient from Theorem 3.4.7 to bound the condition number

$$\begin{aligned} \lambda_{max}(\mathbf{S}_p) \geq \frac{V_{max}^{H}\mathbf{S}_p V_{max}}{V_{max}^{H} V_{max}} &= \frac{V_{max}^{H}\mathbf{D}^{-1}V_{max} + \sigma_o^{-2} V_{max}^{H}\mathbf{L}^{-T}\mathbf{H}^{T}\mathbf{H}\mathbf{L}^{-1}V_{max}}{V_{max}^{H} V_{max}} \\ &= \lambda_{max}(\mathbf{D}^{-1}) + \frac{q}{N}\frac{1}{\sigma_o^2}\psi_{\alpha} \\ &\geq \lambda_{max}(\mathbf{D}^{-1}) + \frac{q}{N}\frac{1}{\sigma_o^2}\psi_{min}, \end{aligned} \tag{5.26}$$

which bounds the largest eigenvalue. Similarly for the smallest eigenvalue,

$$\lambda_{min}(\mathbf{S}_p) \leq \frac{V_{min}^H \mathbf{S}_p V_{min}}{V_{min}^H V_{min}} \leq \lambda_{min}(\mathbf{D}^{-1}) + \frac{q}{N}\frac{1}{\sigma_o^2}\psi_{max} \tag{5.27}$$

where $\psi_{max/min}$ is as computed in (5.24) and (5.25)

$$\psi_l = \begin{cases} \sum\limits_{k=0}^{n} |\lambda_l|^{2k} & \text{if } \lambda_l(\mathbf{D}) = \lambda_l(B_0) \\ \frac{1}{n}\sum\limits_{i=1}^{n-1}\sum\limits_{j=0}^{i}\sum\limits_{k=0}^{n-i}\left(2Re(\lambda_l^j) - 1\right)\cdot|\lambda_l|^{2k} & \text{if } \lambda_l(\mathbf{D}) = \lambda_l(Q) \end{cases}, \tag{5.28}$$

where the l subscript denotes the largest or smallest eigenvalue (min/max) respectively.

Combining these eigenvalue bounds yields,

$$\kappa(\mathbf{S}_p) \geq \frac{\lambda_{max}(\mathbf{D}^{-1}) + \frac{q}{N}\frac{1}{\sigma_o^2}\psi_{min}}{\lambda_{min}(\mathbf{D}^{-1}) + \frac{q}{N}\frac{1}{\sigma_o^2}\psi_{max}}. \tag{5.29}$$

For the next step we recall $\lambda_{max}(A^{-1})^{-1} = \lambda_{min}(A)$ for $A \in \mathbb{R}^{n\times n}$, then take a factor of $\kappa(\mathbf{D})$ out and substitute (5.11) into (5.29), arriving at

$$\kappa(\mathbf{S}_p) \geq \kappa(\mathbf{D})\left(\frac{1 + \frac{q}{N}\frac{min\left\{\sigma_b^2\lambda_{min}(C_B),\sigma_q^2\lambda_{min}(C_Q)\right\}}{\sigma_o^2}\psi_{min}}{1 + \frac{q}{N}\frac{max\left\{\sigma_b^2\lambda_{max}(C_B),\sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}\psi_{max}}\right), \tag{5.30}$$

which establishes the lower bound. For the upper bound, we substitute $\mathbf{R}^{-1} = \sigma_o^{-2}\mathbf{I}_{q(n+1)}$ and $\lambda_{min}(\mathbf{D})$ from (5.11) and thus

$$\kappa(\mathbf{S}_p) \leq \kappa(\mathbf{D})\left(1 + \frac{min\left\{\sigma_b^2\lambda_{min}(C_B),\sigma_q^2\lambda_{min}(C_Q)\right\}}{\sigma_o^2}\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1})\right), \tag{5.31}$$

as required. ■

The upper and lower bounds in Theorem 5.1.2 clearly show that the condition number of $\mathbf{S}_p$ is dependent on the condition number of $\mathbf{D}$. The components governing the condition number of $\mathbf{D}$, shown in (5.11), are as follows:

1. The background and model error variance ratio σ_b/σ_q.

2. The background and the model error covariance matrices.

As the ratio σ_b/σ_q approaches zero, or diverges away from 1, the condition number of $\mathbf{D}$ and hence the condition number of $\mathbf{S}_p$ will grow. This means if the model error variance were to be too small, or too large, in comparison to the background error variance, the condition number of $\mathbf{S}_p$ will be large. This argument also applies to the background error variance. Secondly, as the correlation length-scales in the background and the model error covariance matrices grows, the condition number of $\mathbf{D}$ and hence the condition number of $\mathbf{S}_p$ will also grow. The upper bound in Theorem 5.1.2 also shows that as the observation accuracy (decreasing σ_o) increases, then the upper bound will increase. The lower bound will also increase as σ_o decreases, provided $\psi_{min} << \psi_{max}$ is true. So both bounds suggest that the condition number of $\mathbf{S}_p$ may grow as σ_o decreases.

We now use the 1D advection equation as described in Section 3.5.1 to derive more specific bounds to investigate $\kappa(\mathbf{S}_p)$ further.

5.1.1 The 1D Advection Equation

Theorem 5.1.3 *In addition to the assumptions in Theorem 5.1.2, let M be matrix (3.71), which is the advection equation discretised using the upwind scheme. Then for Courant number $\mu \in [-1, 0]$ we have the following bounds on $\kappa(\mathbf{S}_p)$:*

$$\kappa(\mathbf{D}) \left(\frac{1 + \frac{q}{N} \frac{min\left\{\sigma_b^2 \lambda_{min}(C_B), \sigma_q^2 \lambda_{min}(C_Q)\right\}}{\sigma_o^2} \psi_{min}^{adv}}{1 + \frac{q}{N} \frac{max\left\{\sigma_b^2 \lambda_{max}(C_B), \sigma_q^2 \lambda_{max}(C_Q)\right\}}{\sigma_o^2} \psi_{max}^{adv}} \right) \leq \kappa(\mathbf{S}_p)$$

$$\leq \kappa(\mathbf{D}) \left(1 + \frac{min\left\{\sigma_b^2 \lambda_{min}(C_B), \sigma_q^2 \lambda_{min}(C_Q)\right\}}{\sigma_o^2} (n+1)^2 \right), \quad (5.32)$$

where

$$\psi_{min}^{adv} \begin{cases} = \frac{1}{n} \sum_{i=1}^{n-1} \left[2 \left(\frac{1-(1+2\mu)^{(i+1)}}{1-(1+2\mu)} \right) - 1 \right] \cdot \left[\frac{1-|1+2\mu|^{2(n+1-i)}}{1-|1+2\mu|^2} \right] & \textit{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(Q) \\ \geq \frac{1-|1+2\mu|^{2(n+1)}}{1-|1+2\mu|^2} & \textit{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(B_0) \end{cases} \quad (5.33)$$

and

$$\psi_{max}^{adv} = \begin{cases} \frac{n^2}{3} + \frac{3}{2}n - \frac{5}{6} - \frac{1}{n} & \textit{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(Q) \\ (n+1) & \textit{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(B_0) \end{cases}. \quad (5.34)$$

Proof: We require results on the minimum and maximum eigenvalues of M to obtain bounds for $\kappa(\mathbf{S}_p)$. We use similar methodology as in [41], where the author obtained the extreme eigenvalues of a matrix similar to (3.71). Since M is circulant with entries as shown in (3.71), by Theorem 3.3.8 the eigenvalues take the following form,

$$\lambda_m = 1 + \mu - \mu e^{-\frac{2\pi i m}{N}} \tag{5.35}$$

for $m = 0, ..., N-1$ where $i = \sqrt{-1}$. We also have

$$|\lambda_m|^2 = (\lambda_m)(\bar{\lambda}_m) = (1+\mu)^2 - 2\mu(1+\mu)\cos(\frac{2\pi m}{N}) + \mu^2. \tag{5.36}$$

Let $f(m) = |\lambda_m|^2$ be a continuous function of $m \in [0, N)$. We can find the minimum and maximum of this function by differentiation:

$$f'(m) = 2\mu(1+\mu)(\frac{2\pi}{N})\sin(\frac{2\pi m}{N}), \tag{5.37}$$

$$f''(m) = 2\mu(1+\mu)(\frac{2\pi}{N})^2\cos(\frac{2\pi m}{N}). \tag{5.38}$$

Now we see that $f'(m) = 0$ implies the extrema occur at $m = 0, \frac{N}{2}$. It follows that $f''(0) < 0$ and $f''(\frac{N}{2}) > 0$ for all permissible values of $\mu \in (-1, 0)$. Therefore, for N even, it is trivial to see that

$$\lambda_{max}(M) = \lambda_0(M) \qquad = 1 + \mu - \mu(e^0) = 1, \tag{5.39}$$

$$\lambda_{min}(M) = \lambda_{\frac{N}{2}}(M) \qquad = 1 + \mu - \mu(e^{-\pi i}) = 1 + 2\mu \text{ (N even)}, \tag{5.40}$$

$$\lambda_{min}(M) = \lambda_{\frac{(N-1)}{2}}(M) = 1 + \mu - \mu(e^{-\frac{(N-1)\pi i}{N}}) \geq 1 + 2\mu \text{ (N odd)}. \tag{5.41}$$

Therefore, for values $\mu \in (-1, 0)$, M has the following minimum and maximum eigenvalues

$$|\lambda_{max}(M)|^2 = 1, \tag{5.42}$$

$$|\lambda_{min}(M)|^2 \geq (1 + 2\mu)^2, \tag{5.43}$$

where we have equality in (5.43) if N is even.

Now that we have computed the minimum and maximum eigenvalues of M, we compute $\psi_{min/max}$, which we will denote as ψ^{adv}_{min} and ψ^{adv}_{max} respectively.

Substituting the minimum and maximum eigenvalues of M, we find:

$$\psi^{adv}_{max} = \begin{cases} \frac{1}{n}\sum\limits_{i=1}^{n-1}\sum\limits_{j=0}^{i}\sum\limits_{k=0}^{n-i}((1)^j + (1)^j - 1)\cdot|1|^{2k} & \text{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(Q) \\ \sum\limits_{k=0}^{n}|1|^{2k} & \text{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(B_0), \end{cases} \tag{5.44}$$

and

$$\psi^{adv}_{min} = \begin{cases} \frac{1}{n}\sum\limits_{i=1}^{n-1}\sum\limits_{j=0}^{i}\sum\limits_{k=0}^{n-i}(2(1+2\mu)^j - 1)\cdot|1+2\mu|^{2k} & \text{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(Q) \\ \sum\limits_{k=0}^{n}|\lambda_{min}|^{2k} \geq \sum\limits_{k=0}^{n}|1+2\mu|^{2k} & \text{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(B_0). \end{cases} \tag{5.45}$$

We now compute ψ^{adv}_{max}. We utilise Theorem 3.4.5 to simplify the arising summative expressions in the proceeding computations. For the case $\lambda_{max}(\mathbf{D}) = \lambda_{max}(B_0)$,

$$\sum_{k=0}^{n}|1|^{2k} = (n+1). \tag{5.46}$$

For the case $\lambda_{max}(\mathbf{D}) = \lambda_{max}(Q)$, we compute an expression by first recognising that these are arithmetic sums governed by an outer sum. We begin with the inner-most sum:

$$\sum_{k=0}^{n-i}|1|^{2k} = (n+1-i) \tag{5.47}$$

and the second inner sum yields:

$$\sum_{j=0}^{i}(1^j + 1^j - 1) = (2i+1). \tag{5.48}$$

Since both sums are dependent on the index i, which is governed by the first sum, it follows that i remains in these expressions. Combining (5.47) and (5.48) we now have:

$$\psi^{adv}_{max} = \frac{1}{n}\sum_{i=1}^{n-1}(2i+1).(n+1-i). \tag{5.49}$$

Computing (5.49) we find:

$$\begin{aligned}
\psi_{max}^{adv} &= \frac{1}{n}\left(\sum_{i=1}^{n-1} 2i(n+1-i) + \sum_{i=1}^{n-1}(n+1-i)\right) \\
&= \frac{1}{n}\left(2\left[\sum_{i=1}^{n-1} i(n+1) - i^2\right] + \left[(n+1)(n-1) - \sum_{i=1}^{n-1} i\right]\right) \\
&= \frac{1}{n}\left(2\left[\frac{(n+1)(n-1)(n)}{2} - \frac{(n-1)(n)(2n-1)}{6}\right] + \left[(n+1)(n-1) - \frac{(n-1)(n)}{2}\right]\right) \\
&= \frac{n-1}{n}\left[(n+1)(n) - \frac{n(2n-1)}{3} + (\frac{n}{2}+1)\right] \\
&= \frac{n-1}{n}(\frac{n}{6}(2n+11)+1) \\
&= \frac{n^2}{3} + \frac{3}{2}n - \frac{5}{6} - \frac{1}{n}. \qquad (5.50)
\end{aligned}$$

Therefore,

$$\psi_{max}^{adv} = \begin{cases} \frac{n^2}{3} + \frac{3}{2}n - \frac{5}{6} - \frac{1}{n} & \text{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(Q) \\ (n+1) & \text{if } \lambda_{max}(\mathbf{D}) = \lambda_{max}(B_0) \end{cases}. \qquad (5.51)$$

It remains to find ψ_{min}^{adv}. For the case $\lambda_{min}(\mathbf{D}) = \lambda_{min}(B_0)$ in (5.45), we recognise that this is a geometric sum:

$$\sum_{k=0}^{n} |1+2\mu|^{2k} = \frac{1-|1+2\mu|^{2(n+1)}}{1-|1+2\mu|^2}. \qquad (5.52)$$

For the case $\lambda_{min}(\mathbf{D}) = \lambda_{min}(Q)$ in (5.45) we have:

$$\begin{aligned}
\psi_{min}^{adv} =& \sum_{i=1}^{n-1}\sum_{j=0}^{i}\left(2(1+2\mu)^j - 1\right) \cdot \left[\frac{1-|1+2\mu|^{2(n+1-i)}}{1-|1+2\mu|^2}\right] \\
=& \sum_{i=1}^{n-1}\left[2\left(\frac{1-(1+2\mu)^{(i+1)}}{1-(1+2\mu)}\right) - 1\right] \cdot \left[\frac{1-|1+2\mu|^{2(n+1-i)}}{1-|1+2\mu|^2}\right]. \qquad (5.53)
\end{aligned}$$

Therefore,

$$\psi_{min}^{adv} \begin{cases} = \sum_{i=1}^{n-1}\left[2\left(\frac{1-(1+2\mu)^{(i+1)}}{1-(1+2\mu)}\right) - 1\right] \cdot \left[\frac{1-|1+2\mu|^{2(n+1-i)}}{1-|1+2\mu|^2}\right] & \text{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(Q) \\ \geq \frac{1-|1+2\mu|^{2(n+1)}}{1-|1+2\mu|^2} & \text{if } \lambda_{min}(\mathbf{D}) = \lambda_{min}(B_0) \end{cases} \qquad (5.54)$$

as required.

For the upper bound in Theorem 5.1.2, we begin by recognising that

$$\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}) = ||\mathbf{H}\mathbf{L}^{-1}||_2^2 \leq ||\mathbf{H}||_2^2||\mathbf{L}^{-1}||_2^2, \qquad (5.55)$$

by using the definition of the 2-norm and norm relationship in Theorem 3.3.6.

We now briefly discuss the 2-norm of the observation operator $\mathbf{H} \in \mathbb{R}^{p(n+1)\times N(n+1)}$. The main assumption states that there are fewer observations than state space, so from Definition 3.3.4, we have

$$||\mathbf{H}||_2 = \underset{\mathbf{x}\neq 0}{sup}\left(\frac{|x_1|^2 + |x_3|^2 + ... + |x_{q(n+1)}|^2}{|x_1|^2 + |x_2|^2 + ... + |x_{N(n+1)}|^2}\right), \tag{5.56}$$

where $\mathbf{x} \in \mathbb{R}^{N(n+1)}$, such that

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N(n+1)} \end{pmatrix}. \tag{5.57}$$

It is obvious that the numerator can never exceed the denominator because $q < N$. To illustrate this, let us assume every other point in the state is observed, therefore it is obvious that

$$\frac{|x_1|^2 + |x_3|^2 + ... + |x_{q(n+1)}|^2}{|x_1|^2 + |x_2|^2 + ... + |x_{N(n+1)}|^2} \leq 1. \tag{5.58}$$

We have assumed a particular instance, which adheres to the original assumption of $q < N$. In general, the number of observations being less than the state means the denominator in (5.58) *can never exceed* the numerator. Therefore the supremum of (5.58) is

$$||\mathbf{H}||_2 = 1. \tag{5.59}$$

To calculate $||\mathbf{L}^{-1}||_2$ we use the inequality

$$||\mathbf{L}^{-1}||_2 \leq ||\mathbf{L}^{-1}||_1 ||\mathbf{L}^{-1}||_\infty, \tag{5.60}$$

while also noting that the infinity-norm and 1-norm of $\mathbf{L}^{-1}$ are equal, which can be seen by quick inspection of $\mathbf{L}^{-1}$, (2.37). The matrix $\mathbf{L}^{-1}$ can be written as a power series such that,

$$\mathbf{L}^{-1} = \mathbf{I} + \mathbf{M} + \mathbf{M}^2 + ... + \mathbf{M}^n, \tag{5.61}$$

$$\mathbf{L}^{-1} = \begin{pmatrix} I & & & \\ & I & & \\ & & \ddots & \\ & & & I \end{pmatrix} + \begin{pmatrix} \mathbf{0} & & & & \\ M_1 & \mathbf{0} & & & \\ & M_2 & \mathbf{0} & & \\ & & \ddots & \ddots & \\ & & & M_n & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & & & & \\ \mathbf{0} & \mathbf{0} & & & \\ M_2M_1 & \mathbf{0} & \mathbf{0} & & \\ & M_3M_2 & \mathbf{0} & \mathbf{0} & \\ & & \ddots & \ddots & \ddots \\ & & M_nM_{n-1} & \mathbf{0} & \mathbf{0} \end{pmatrix} +$$

$$... + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ M_n...M_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{5.62}$$

The ECMWF write the $\mathbf{L}^{-1}$ operator as a *Neumann series* to approximate $\mathbf{L}$, [27]. The authors hope to approximate $\mathbf{L}$ to precondition the state estimation formulation (2.33). We write it as a series here with the intent of approximating it to have a more comprehensive expression for the bounds on the condition number. It follows that

$$\begin{aligned}
||\mathbf{L}^{-1}||_1 &= ||\mathbf{I}+\mathbf{M}+...+\mathbf{M}^n||_1,\\
&\leq ||\mathbf{I}||_1+||\mathbf{M}||_1+...+||\mathbf{M}^n||_1,\\
&= 1+||M||_1+...+||M^n||_1,\\
&\leq 1+||M||_1+...+||M||_1^n = \sum_{k=0}^{n}||M||_1^k. \qquad (5.63)
\end{aligned}$$

Since M is linear, then $M_{i,i-1} = M$ for $i = 1,...,n$ is true, and therefore the following statement holds $||\mathbf{M}||_1 = ||M||_1$. We also know that the absolute row and column sums of a circulant matrix are equal. Computing the norm for the advection equation we have

$$||M||_1 = (1+\mu)+(-\mu) = 1, \qquad (5.64)$$

and since $||\mathbf{L}^{-1}||_\infty = 1$ by the same argument, therefore

$$\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}) \leq (n+1)^2. \qquad (5.65)$$

By substituting the ψ expressions, (5.51), (5.54) and (5.65) into the bounds in Theorem 5.1.2, we arrive at the bounds in Theorem 5.1.3, which completes the proof. ■

We can now see the following new elements in the bounds in Theorem 5.1.3:

1. the parameters $\psi^{adv}_{min/max}$, which are specific to the advection equation;

2. the presence of the assimilation window length, n, in the upper bound.

In the lower bound of Theorem 5.1.3 we can see that ψ^{adv}_{max} will increase as the assimilation window length increases, whereas ψ^{adv}_{min} divulges no definitive information. We can see that the assimilation window length, n, has a quadratic influence from the ψ^{adv}_{max} expression in (5.34). The upper bound of Theorem 5.1.3 shows the quadratic influence of the assimilation window length n. Both the upper and lower bounds suggest that the assimilation window length will have an influence on the condition number of $\mathbf{S}_p$.

This concludes the derivation of our bounds on $\mathbf{S}_p$. We now briefly compare the bounds on the condition number of $\mathbf{S}_p$ to the bounds on the condition number of the sc4DVAR Hessian, before demonstrating the bounds numerically.

5.1.2 Comparison to Strong-Constraint 4DVAR

The bounds in Theorem 5.1.2 bear some similarities to the bounds derived on the condition number of the Hessians of the sc4DVAR and 3DVAR problems as shown in [41] (Theorem 6.1.2 and Theorem 7.1.2). The influence of the condition number of B_0 on the condition number of the sc4DVAR Hessian is similar to the influence of the condition number of $\mathbf{D}$ on the condition number of $\mathbf{S}_p$. The B_0 matrix was influenced only by the condition number of the background error covariance matrix C_B, whereas $\mathbf{D}$ is influenced by C_B, C_Q and the ratio of σ_b/σ_q. We further illustrate this by taking a simplified scenario as an example.

Assume the background and model errors are uncorrelated in space such that

$$\mathbf{D} = \begin{pmatrix} \sigma_b^2 I & & & & \\ & \sigma_q^2 I & & & \\ & & \sigma_q^2 I & & \\ & & & \ddots & \\ & & & & \sigma_q^2 I \end{pmatrix}. \tag{5.66}$$

We also assume that the background error variance is larger than the model error variance, $\sigma_b > \sigma_q$. The background error variance is representative of the errors in the previous assimilation window in its entirety, which normally consists of several model time steps. The model error variance represents the errors in one model time step. It is intuitive to believe the error variance in one time step is less than multiple model time steps. Allowing for more model time steps between error corrections implies that the model error variance will grow such that $\sigma_q \to \sigma_b$.

The condition number of $\mathbf{D}$ becomes

$$\kappa(\mathbf{D}) = \left(\frac{\sigma_b}{\sigma_q}\right)^2. \tag{5.67}$$

We now briefly analyse the wc4DVAR bounds from Theorem 5.1.2 in light of these additional arguments. We have

$$\frac{\left(\frac{\sigma_b}{\sigma_q}\right)^2 + \frac{q}{N}\left(\frac{\sigma_b}{\sigma_o}\right)^2 \psi_{min}}{1 + \frac{q}{N}\left(\frac{\sigma_b}{\sigma_o}\right)^2 \psi_{max}} \leq \kappa(\mathbf{S}_p) \leq \left(\frac{\sigma_b}{\sigma_q}\right)^2 + \left(\frac{\sigma_b}{\sigma_o}\right)^2 \lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}), \tag{5.68}$$

which can be compared directly to the sc4DVAR bounds in [41] (Theorem 7.1.2), with the same assumptions:

$$\frac{1 + \frac{q}{N}\left(\frac{\sigma_b}{\sigma_o}\right)^2 \gamma_{min}}{1 + \frac{q}{N}\left(\frac{\sigma_b}{\sigma_o}\right)^2 \gamma_{max}} \quad \leq \quad \kappa(S) \quad \leq \quad 1 \quad + \quad \left(\frac{\sigma_b}{\sigma_o}\right)^2 \lambda_{max}(\hat{\mathbf{H}}^T\hat{\mathbf{H}}), \tag{5.69}$$

where S is the sc4DVAR first order Hessian and γ is the sc4DVAR equivalent to ψ, (5.21). We see the added dimension of the background and model error variance covariance matrix represented by the ratio $\frac{\sigma_b}{\sigma_q}$ playing a significant role in the conditioning of $\mathbf{S}_p$. We also see the contribution of the maximum eigenvalue of the terms $\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}$ and $\hat{\mathbf{H}}^T\hat{\mathbf{H}}$, which is linked to the length of the assimilation window and observation operator.

We showed in Theorem (5.1.3) that $\lambda_{max}(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1})$ can be approximated to $(n+1)^2$, where the author in [41] showed that $\lambda_{max}(\hat{\mathbf{H}}^T\hat{\mathbf{H}})$ for sc4DVAR reduces to $(n+1)$. So the effect of the assimilation window on the bounds from sc4DVAR to wc4DVAR is greater by an order of magnitude.

In this section we have demonstrated the inherent similarities between the condition numbers of $\mathbf{S}_p$ and S. In the next section we demonstrate the sensitivities shown by the bounds in the Theorems on the condition number of $\mathbf{S}_p$.

5.1.3 Numerical Results

We now demonstrate the bounds through numerical experiments. We also highlight sensitivities of the condition number of $\mathbf{S}_p$ with respect to assimilation parameters, which have been revealed by the theorems in Section 5.1.

We let M be the linear advection model as in (3.71), with a one-dimensional domain of size $N = 500$ grid points and spatial intervals of $\Delta x = 0.1$. We use temporal intervals of $\Delta t = 0.1$ and wave speed $a = -0.3$. We let $n = 2$, so we have a total of three model time levels including initial time, all of which are observed. We let $q = 20$ spatial observations at the grid points with equal spacing, so $q(n+1) = 60$. The temporal observations are made every 3 model time steps, so at $t_0 = 0$, $t_1 = 3\Delta t$ and $t_2 = 6\Delta t$. We assume no spatial correlations for the observation errors whereas the background and model errors are spatially correlated (as in Sections 3.3.4.1 and 3.3.4.2), $B_0 = \sigma_b^2 C_{SOAR}$, $Q_i = Q = \sigma_q^2 C_{LAP}$, $R = \sigma_o^2 I_q$ where $\sigma_b = \sigma_q = \sigma_o = 1$ unless otherwise stated. We denote the correlation length-scale of a covariance matrix C as $L(C)$ (Section 3.3.3).

5.1.3.1 Experiment 1: Correlation Length-Scales

We first examine the effects of increasing the background correlation length-scale on the condition number of $\mathbf{S}_p$.

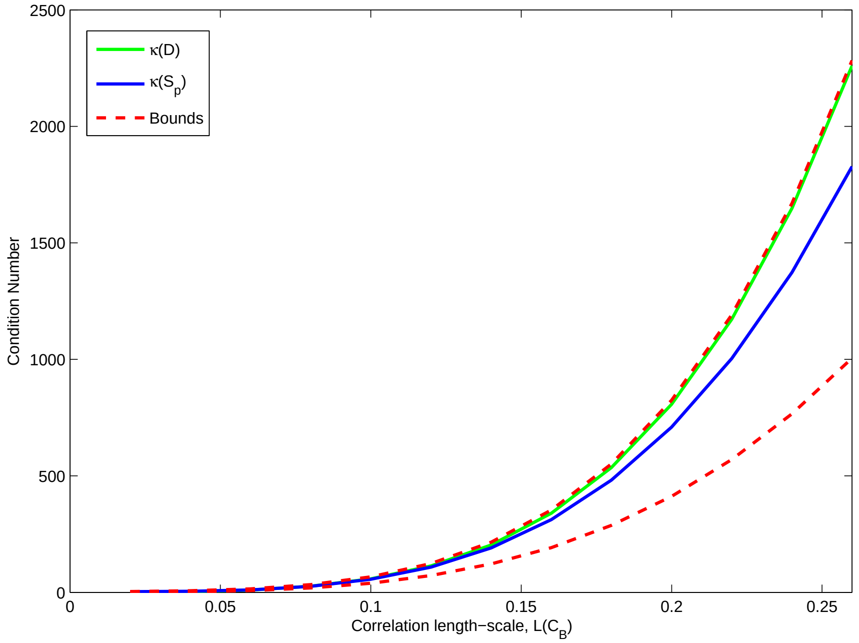


Figure 5.1: $\kappa(\mathbf{S}_p)$ (blue line), $\kappa(\mathbf{D})$ (green line) and theoretical bounds (red-dotted line) as a function of $L(C_B)$. Model error correlation length-scale $L(C_Q) = \Delta x/5$.

Figure 5.1 shows the bounds from Theorem 5.1.2 with the condition numbers of $\mathbf{S}_p$ and $\mathbf{D}$. We see the dependence of $\kappa(\mathbf{S}_p)$ on $\kappa(\mathbf{D})$, which rises as a result of the correlation length-scale increasing in the background error covariance matrix, [41]. The bounds prove to be a good estimate of the overall behaviour of the condition number when varying length-scales in B_0 and hence $\mathbf{D}$.

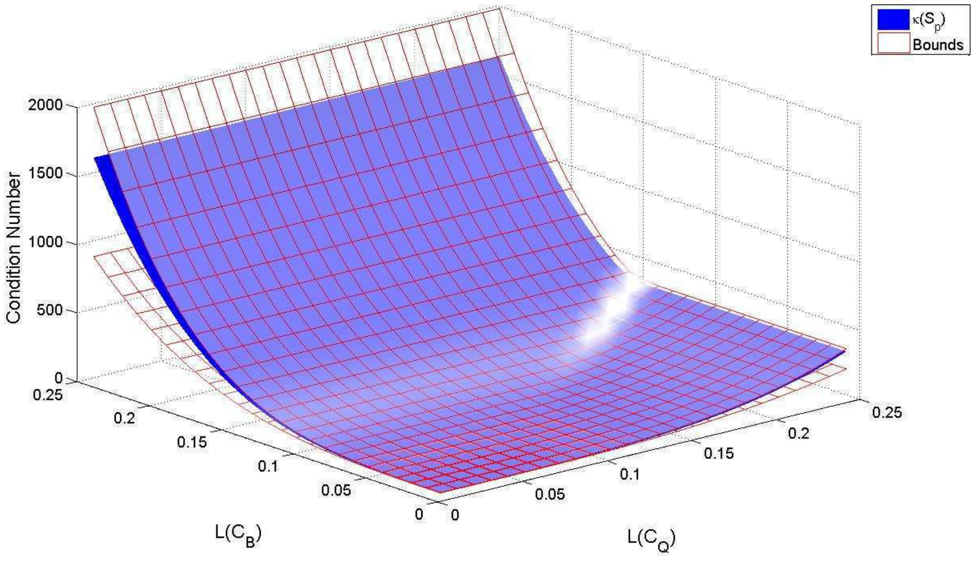


Figure 5.2: $\kappa(\mathbf{S}_p)$ (blue-surface) and bounds (red-mesh surface) as a function of $L(C_B)$ and $L(C_Q)$.

In Figure 5.2 we show that the increasing the model error correlation length-scale does not affect the condition number as much as the increase in length-scale in the B matrix. This is due to the Laplacian covariance matrix being better conditioned than the SOAR covariance matrix in general, [41], Chapter 5. We see evidence of this in this experiment: with correlation length-scales of $L(C_B) = L(C_Q) = 2.5\Delta x$, the condition numbers of the SOAR and Laplacian matrices are $\kappa(C_{SOAR}) = 1973$ and $\kappa(C_{LAP}) = 359$.

Figures 5.1 and 5.2 demonstrate the following:

1. The sensitivity of the condition number of the Hessian $\mathbf{S}_p$ to the condition number of the background and model error covariance matrix $\mathbf{D}$, as shown initially in Theorem 5.1.1. We specifically showed the sensitivity of $\kappa(\mathbf{S}_p)$ to the increase in both the background and model error correlation length-scales.

2. The sensitivity of the condition number of $\mathbf{S}_p$ to the condition number of $\mathbf{D}$ from Theorem 5.1.2. The condition number of $\mathbf{D}$ is sensitive to

the correlation length-scales in the covariance matrices C_B and C_Q, which influences the condition number of $\mathbf{S}_p$.

3. The bounds accurately and closely estimate the true condition number when varying the correlation length-scales of C_B and C_Q in these experiments.

We now demonstrate the bounds and Hessian condition number sensitivities to the error variance ratios.

5.1.3.2 Experiment 2: Error Variance Ratios

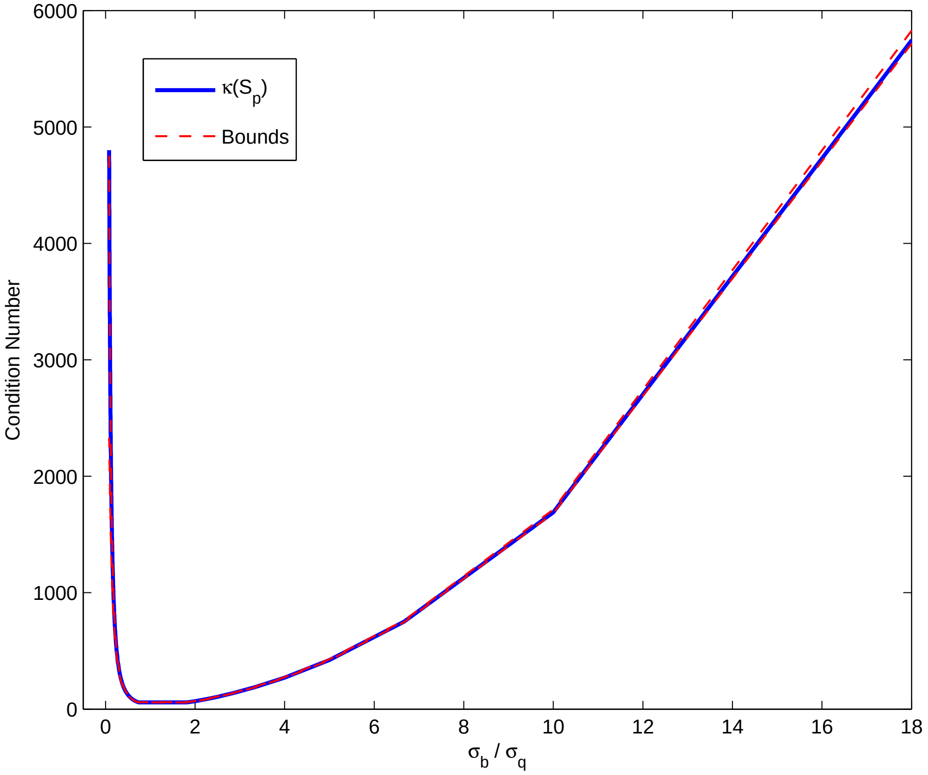


Figure 5.3: $\kappa(\mathbf{S}_p)$ (blue line) and theoretical bounds (red-dotted line) as a function of ratio σ_b/σ_q. $L(C_B) = L(C_Q) = 1\Delta x$.

We now examine the effect of the background and model error variance ratio on the condition number of $\mathbf{S}_p$. In Figure 5.3, we see that as the ratio σ_b/σ_q tends to 0 and increases from 1, the condition number of $\mathbf{S}_p$ also increases. This is due to the condition number of $\mathbf{D}$ increasing as the ratio of σ_b/σ_q tends to 0 and increases from 1.

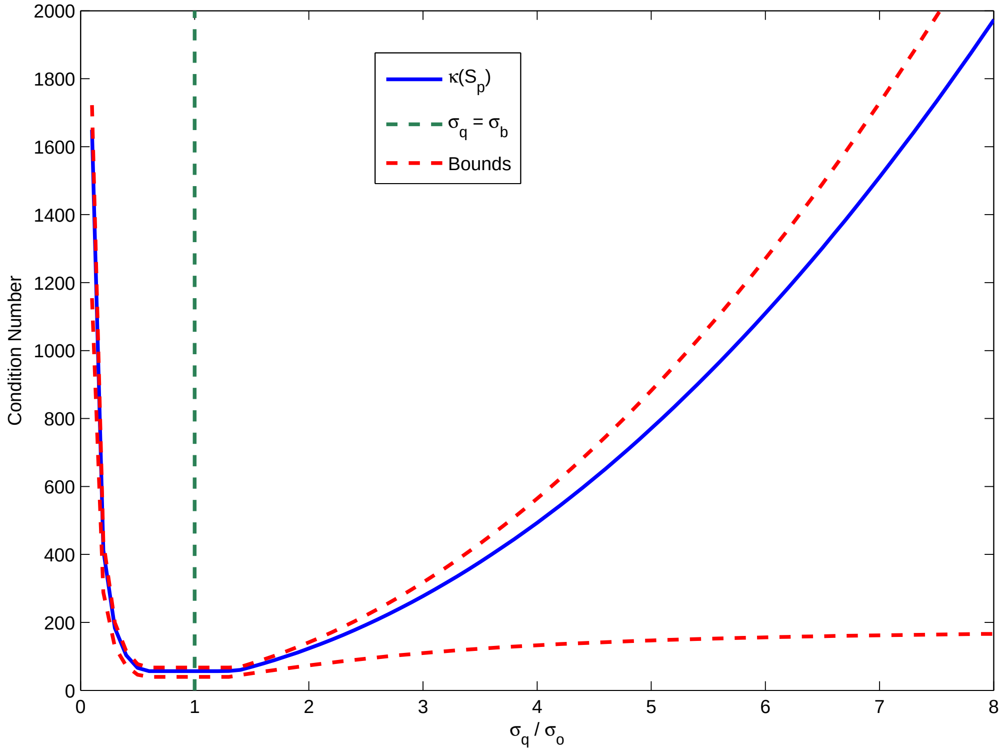


Figure 5.4: $\kappa(\mathbf{S}_p)$ (blue line) and theoretical bounds (red-dotted line) as a function of ratio σ_q/σ_o. $L(C_B) = L(C_Q) = 1\Delta x$. Green dotted line at the point $\sigma_q = \sigma_b$

Figure 5.4 shows similar behaviour of the σ_q/σ_o to the ratio σ_b/σ_q shown in Figure 5.3. We show the point at which $\sigma_q = \sigma_b$, where $\sigma_b = 1$ on this graph to emphasise the importance of the actual ratio $max/min \left\{\sigma_b^2\lambda_{min}(C_B), \sigma_q^2\lambda_{min}(C_Q)\right\}/\sigma_o$, from both bounds in Theorem 5.1.2. As soon as one of the extreme eigenvalues of B_0 or Q takes precedence over the other and grows further away from σ_o, the condition number of $\mathbf{S}_p$ increases.

Figures 5.3 and 5.4 demonstrate the sensitivities of the condition number $\mathbf{S}_p$ to the ratio in error variances within the wc4DVAR problem. More specifically we have shown:

1. As the the ratio $\sigma_b/\sigma_q \to 0, \infty$ from 1, the condition number of the Hessian

$\mathbf{S}_p$ increases.

2. As the ratio $max/min\left\{\sigma_b^2\lambda_{min}(C_B), \sigma_q^2\lambda_{min}(C_Q)\right\}/\sigma_o^2 \rightarrow 0, \infty$ the condition number of $\mathbf{S}_p$ increases.

3. The bounds estimate the true condition number well when varying the background and model σ_b/σ_q error variance ratios in these experiments. The upper bound is also tight for the model and observation error variance ratio whereas the lower bound is a poor estimate of the σ_q/σ_o ratio.

We now demonstrate the bounds and Hessian condition number sensitivities to the length of the assimilation window.

5.1.3.3 Experiment 3: Assimilation Window Length

We now examine the effects of assimilation window length on the condition number of $\mathbf{S}_p$.

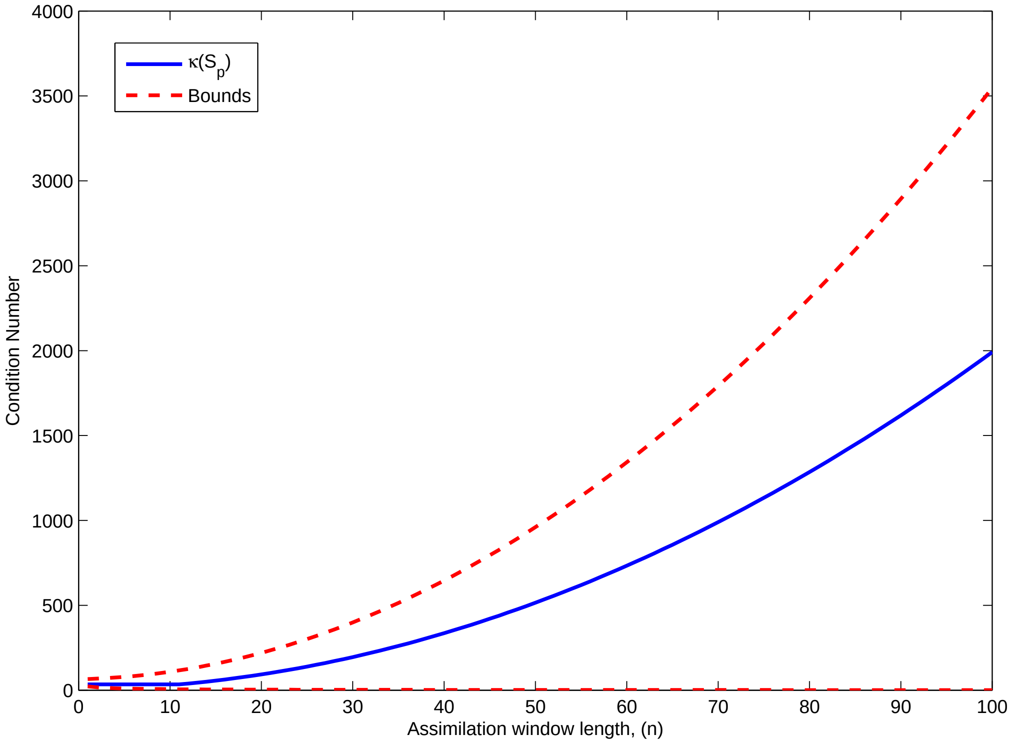


Figure 5.5: $\kappa(\mathbf{S}_p)$ as a function of assimilation window length, n. $L(C_B) = L(C_Q) = \Delta x$.

Figure 5.5 demonstrates the bounds in Theorem 5.1.3. The upper bound has the term $(n+1)^2$, which shows that the bound is quadratically influenced by the assimilation window length. We see that the actual condition number of $\mathbf{S}_p$ does increase quadratically as the assimilation window length increases, for example doubling the window from 50 to 100 sees approximately 4 times the increase in the condition number of $\mathbf{S}_p$ from ~ 500 to ~ 2000. The upper bound has similar behaviour which can be seen from the shape of the graph but it is not exactly quadratic, doubling the window from 50 to 100 increases the upper bound from ~ 1000 to ~ 3500. The lower bound is uninformative.

5.1.4 Summary

We have obtained new general bounds on the condition number of the wc4DVAR $\mathcal{J}(\mathbf{p})$ formulation. We then developed the bounds by making simple assumptions about the observations, the nature of the model and the covariance matrices. This was then extended to the specific case where the model is a 1D advection equation, which is of relevance in NWP since advection is a physical process occurring in numerous models describing atmospheric systems.

The theorems in this section extend the work of Haben et al. [41] on the condition number of the standard 3DVAR and 4DVAR systems. We briefly discussed and compared $\mathcal{J}(\mathbf{p})$ to the conventional sc4DVAR approach (2.8) in Section 5.1.2 using the lower and upper bounds derived in [41] and the bounds we have derived in Theorem 5.1.2. In sc4DVAR, $\frac{\sigma_b}{\sigma_o}$ is the only error variance ratio, which means if the observations are accurate and/or the background error variance is large then the condition number of the of Hessian of the sc4DVAR problem would rise. We showed that for wc4DVAR there is an intricate balance to be considered for the combination of the three ratios, $\frac{\sigma_b}{\sigma_q}$, $\frac{\sigma_b}{\sigma_o}$ and $\frac{\sigma_q}{\sigma_o}$. We showed that the magnitude (whether small or large) of the difference between the error variances in wc4DVAR directly effects the condition number of $\mathbf{S}_p$.

The bounds in Theorem 5.1.2 also indicated the sensitivity of $\kappa(\mathbf{S}_p)$ to correlation length-scales of the background and model error covariance matrices since these have a direct influence on $\kappa(\mathbf{D})$ and hence $\kappa(\mathbf{S}_p)$. We have also shown for the advection equation in Theorem 5.1.3, that the assimilation window length, n, influences the condition number of $\mathbf{S}_p$.

We now examine the preconditioned problem.

5.2 Theoretical Results: Bounding the Condition Number of $\hat{\mathbf{S}}_p$

We recall the preconditioned $\mathbf{S}_p$ Hessian as in Chapter 4, Section 2.3.3 equation (2.60),

$$\hat{\mathbf{S}}_p = \mathbf{I} + \mathbf{D}^{1/2}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2}. \tag{5.70}$$

The following result bounds the condition number of $\hat{\mathbf{S}}_p$,

Theorem 5.2.1 *Let $B_0 \in \mathbb{R}^{N\times N}$ and $Q_i \in \mathbb{R}^{N\times N}$ for $i = 1, .., n$ be our background and static model error covariance matrices respectively. We assume q observations are taken such that $q < N$ with covariance $R_i \in \mathbb{R}^{q\times q}$ thus $\mathbf{R} \in \mathbb{R}^{q(n+1)\times q(n+1)}$. Let $H_i = H \in \mathbb{R}^{q\times N}$ for $i = 0, .., n$, be the time invariant observation operator. Finally, let $M_{i,i-1} = M \in \mathbb{R}^{N\times N}$ for $i = 1, .., n$, represent the time invariant model equations. Then the following bounds are satisfied by the condition number of the Hessian $\hat{\mathbf{S}}_p$:*

$$1 + \frac{1}{q(n+1)}\sum_{i,j=1}^{q(n+1)}\left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1/2}\right)_{i,j} \leq \kappa(\hat{\mathbf{S}}_p)$$
$$\leq 1 + \frac{\lambda_{max}(\mathbf{D})}{\lambda_{min}(\mathbf{R})}\lambda_{max}(\mathbf{L}^{-T}\mathbf{L}^{-1}) \tag{5.71}$$

where $\mathbf{R}^{-1/2}$ is the symmetric square root of $\mathbf{R}^{-1}$.

Proof: Let $\mathbf{E} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2}$. We remember that since $\mathbf{H}$ is not full rank,

$\lambda_{min}(\mathbf{E}^T\mathbf{E}) = 0$. Therefore

$$\kappa(\hat{\mathbf{S}}_p) = \frac{\lambda_{max}(\hat{\mathbf{S}}_p)}{\lambda_{min}(\hat{\mathbf{S}}_p)} = \frac{1+\lambda_{max}(\mathbf{E}^T\mathbf{E})}{1+\lambda_{min}(\mathbf{E}^T\mathbf{E})} = 1+\lambda_{max}(\mathbf{E}^T\mathbf{E}) = \lambda_{max}(\hat{\mathbf{S}}_p), \tag{5.72}$$

meaning the condition number of $\hat{\mathbf{S}}_p$ is equal to its largest eigenvalue. Following on from (5.72) and by Theorem 3.3.5 we deduce that

$$\begin{aligned}\kappa(\hat{\mathbf{S}}_p) &= 1 + ||\mathbf{E}||_2^2 \\ &\leq 1 + ||\mathbf{R}^{-1/2}||_2^2||\mathbf{H}||_2^2||\mathbf{L}^{-1}||_2^2||\mathbf{D}^{1/2}||_2^2.\end{aligned} \tag{5.73}$$

We know by the same argument in Section 5.1, equation (5.59) that

$$||\mathbf{H}||_2 = 1, \tag{5.74}$$

and

$$||\mathbf{D}^{1/2}||_2^2 = \lambda_{max}(\mathbf{D}), \tag{5.75}$$

$$||\mathbf{R}^{-1/2}||_2^2 = \lambda_{max}(\mathbf{R^{-1}}) = \frac{1}{\lambda_{min}(\mathbf{R})}, \tag{5.76}$$

since $\mathbf{D}$ and $\mathbf{R}$ are both symmetric. The upper bound is therefore

$$\kappa(\hat{\mathbf{S}}_p) \leq 1 + \frac{\lambda_{max}(\mathbf{D})}{\lambda_{min}(\mathbf{R})}\lambda_{max}(\mathbf{L}^{-T}\mathbf{L}^{-1}), \tag{5.77}$$

as required.

To obtain the lower bound we define

$$\widetilde{\mathbf{S}}_p = \mathbf{I} + \mathbf{E}\mathbf{E}^T, \tag{5.78}$$

which is also known as the preconditioned Hessian for the wc4DVAR *dual space* formulation. This preconditioned Hessian is related to the lower-dimensional alternative formulation for solving $\mathcal{J}(\mathbf{p})$. The wc4DVAR dual space formulation has had recent research attention with respect to the iterative solvers and preconditioners, [36].

By Theorem 3.4.3 we know that $\hat{\mathbf{S}}_p \in \mathbb{R}^{N(n+1)\times N(n+1)}$ possesses the same *non-unit* eigenvalues as $\widetilde{\mathbf{S}}_p \in \mathbb{R}^{q(n+1)\times q(n+1)}$, therefore $\kappa(\hat{\mathbf{S}}_p) = \lambda_{max}(\widetilde{\mathbf{S}}_p)$. Additionally, $\mathbf{E}^T\mathbf{E}$ will have $(N-q)(n+1)$ eigenvalues equal to zero. We obtain the lower bound on the condition number by applying the Rayleigh quotient to $\widetilde{\mathbf{S}}_p$ using a unit vector $\mathbf{y} \in \mathbb{R}^{q(n+1)}$, such that,

$$\mathbf{y} = \frac{1}{\sqrt{q(n+1)}}(1, 1, \dots, 1). \tag{5.79}$$

The Rayleigh Quotient is bounded by Theorem 3.4.7, so it follows that

$$\kappa(\hat{\mathbf{S}}_p) = \lambda_{max}(\widetilde{\mathbf{S}}) \geq \mathcal{R}_{\widetilde{\mathbf{S}}_p}(\mathbf{y}), \tag{5.80}$$

where $\mathcal{R}_{\widetilde{\mathbf{S}}_p}(\mathbf{y})$ denotes the Rayleigh Quotient of $\widetilde{\mathbf{S}}_p$ using the vector $\mathbf{y}$. Therefore

$$\begin{aligned} \kappa(\hat{\mathbf{S}}_p) \geq \mathcal{R}_{\widetilde{\mathbf{S}}_p}(\mathbf{y}) = \mathbf{y}^T \widetilde{\mathbf{S}}_p \mathbf{y}, && (5.81) \\ = 1 + \frac{1}{q(n+1)} \sum_{i,j=1}^{q(n+1)} \left(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1/2}\right)_{i,j}, && (5.82) \end{aligned}$$

which completes the proof. ■

The aim of preconditioning with $\mathbf{D}^{1/2}$ is to remedy the ill-conditioning that arises from $\mathbf{D}$. We see in the upper bound that preconditioning using $\mathbf{D}$ has alleviated the dominating effect of $\kappa(\mathbf{D})$ on the condition number of $\hat{\mathbf{S}}_p$. A new dependance has been introduced, $\frac{\lambda_{max}(\mathbf{D})}{\lambda_{min}(\mathbf{R})}$, which can be seen this by comparing the bounds in Theorem 5.2.1 to Theorem 5.1.1. The ratio $\frac{\lambda_{max}(\mathbf{D})}{\lambda_{min}(\mathbf{R})}$ shown in the upper bound indicates that if the observation errors are small with respect to the background and model errors or vice versa, then the bound will also increase. We also see there is a contribution of the eigenvalues of $\mathbf{L}$ although it is not yet clear how it influences the condition number exactly.

We now make our assumptions more specific to obtain more informative bounds on the condition number of $\hat{\mathbf{S}}_p$.

Theorem 5.2.2 *Let $B_0 = \sigma_b^2 C_B \in \mathbb{R}^{N \times N}$ and $Q = \sigma_q^2 C_Q \in \mathbb{R}^{N \times N}$ be the background and model error covariance matrices where C is a valid error correlation matrix on the unit circle and $\sigma_b^2, \sigma_q^2 > 0$ denote the respective error variances. We assume $q < N$* direct *observations are taken with the same error variance at each time step t_i so $R_i = \sigma_o^2 I \in \mathbb{R}^{q \times q}$. Let $H_i \in \mathbb{R}^{q \times N}$ such that*

$H_iH_i^T = I_q$ and $M_{i,i-1} \in \mathbb{R}^{N\times N}$ denote the observation and model operators respectively and $M_{i,i} = I_N$. We then have the following bounds on the condition number of $\hat{\mathbf{S}}_p$:

$$1 + \frac{1}{q(n+1)}\left(\frac{\sigma_b^2}{\sigma_o^2}\sum_{i,j=1}^{q(n+1)}(\mathbf{H}\widetilde{\mathbf{C}}_B\mathbf{H}^T)_{i,j} + \frac{\sigma_q^2}{\sigma_o^2}\sum_{i,j=1}^{q(n+1)}(\mathbf{H}\widetilde{\mathbf{C}}_Q\mathbf{H}^T)_{i,j}\right) \leq \kappa(\hat{\mathbf{S}}_p)$$
$$\leq 1 + \frac{max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}\left(\sum_{k=0}^{n}||M||_\infty^k\right)\left(\sum_{k=0}^{n}||M||_1^k\right) \qquad (5.83)$$

where

$$\widetilde{\mathbf{C}}_B = \begin{pmatrix} C_B & C_BM_{1,0}^T & \dots & C_BM_{n,0}^T \\ M_{1,0}C_B & M_{1,0}C_BM_{1,0}^T & \dots & M_{1,0}C_BM_{n,0}^T \\ & & M_2C_BM_2^T & \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,0}C_B & M_{n,0}C_BM_{1,0}^T & \dots & M_{n,0}C_BM_{n,0}^T \end{pmatrix}, \qquad (5.84)$$

$$\widetilde{\mathbf{C}}_Q = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & C_Q & \dots & C_QM_{n-1,0}^T \\ \vdots & & M_{1,0}C_QM_{1,0}^T + C_Q & \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & M_{n-1,0}C_Q & \dots & \sum_{i=0}^{n-1} M_{i,0}C_QM_{i,0}^T \end{pmatrix}, \qquad (5.85)$$

and $\mathbf{0}$ is a zero matrix of appropriate size.

Proof: We let $\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-T} = \sigma_b^2\widetilde{\mathbf{C}}_B + \sigma_q^2\widetilde{\mathbf{C}}_Q$, thus allowing us to write the dual formulation preconditioned Hessian as

$$\widetilde{\mathbf{S}}_p = \mathbf{I} + \frac{1}{\sigma_o^2}\mathbf{H}(\sigma_b^2\widetilde{\mathbf{C}}_B + \sigma_q^2\widetilde{\mathbf{C}}_Q)\mathbf{H}^T. \qquad (5.86)$$

As in the previous proof, we apply the Rayleigh Quotient to $\widetilde{\mathbf{S}}_p$ using the unit vector (5.79) to obtain the following expression

$$\kappa(\hat{\mathbf{S}}_p) \geq \mathcal{R}_{\widetilde{\mathbf{S}}_p}(\mathbf{y}) = 1 + \frac{1}{q(n+1)}\frac{1}{\sigma_o^2}\sum_{i,j=1}^{q(n+1)}(\mathbf{H}(\sigma_b^2\widetilde{\mathbf{C}}_B + \sigma_q^2\widetilde{\mathbf{C}}_Q)\mathbf{H}^T)_{i,j},$$
$$= 1 + \frac{1}{q(n+1)}\left(\frac{\sigma_b^2}{\sigma_o^2}\sum_{i,j=1}^{q(n+1)}(\mathbf{H}\widetilde{\mathbf{C}}_B\mathbf{H}^T)_{i,j} + \frac{\sigma_q^2}{\sigma_o^2}\sum_{i,j=1}^{q(n+1)}(\mathbf{H}\widetilde{\mathbf{C}}_Q\mathbf{H}^T)_{i,j}\right), \qquad (5.87)$$

which by the bounds of the Rayleigh Quotient as in (5.81), establishes the lower bound.

For the upper bound we know

$$\lambda_{max}(\mathbf{D}) = max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}, \tag{5.88}$$

and

$$\lambda_{min}(\mathbf{R}) = \sigma_o^2, \tag{5.89}$$

which leaves us with

$$\lambda_{max}(\mathbf{L}^{-T}\mathbf{L}^{-1}) = ||\mathbf{L}^{-1}||_2^2 \leq ||\mathbf{L}^{-1}||_1||\mathbf{L}^{-1}||_\infty. \tag{5.90}$$

Using the argument in Section 5.1 equation (5.63) it follows that

$$||\mathbf{L}^{-1}||_1||\mathbf{L}^{-1}||_\infty \leq \left(\sum_{k=0}^{n}||M||_\infty^k\right)\left(\sum_{k=0}^{n}||M||_1^k\right), \tag{5.91}$$

which completes the proof. ■

The upper bound shows that if M is a contraction with respect to the 1-norm and ∞-norm, $||M|| < 1$, then for long assimilation windows the geometric series in the upper bound of Theorem 5.2.2 will tend to $\frac{1-M^{n+1}}{1-M}$. However if $||M|| \geq 1$ then the series will increase, which will increase the upper bound as the assimilation windows get longer. We also see that the lower and upper bounds are no longer influenced by the condition number of $\mathbf{D}$. The lower bound shows that the constituents of the evolved error covariance matrices $\widetilde{\mathbf{C}}_B$ and $\widetilde{\mathbf{C}}_Q$ may contribute to the magnitude of the lower bound. We can therefore see that

1. Increasing the number of observations q increases the number of summation terms. With the increase in observations the $\mathbf{H}$ operator will change and incorporate more terms from the evolved error covariance matrices which *could* increase the lower bound, depending on the entries of the evolved error covariance matrices.

2. The lower bound will increase if the ratios $\frac{\sigma_b}{\sigma_o}$ and $\frac{\sigma_q}{\sigma_o}$ increase. We notice in comparison to the bounds of the unpreconditioned Hessian in Theorem 5.1.1, the ratio $\frac{\sigma_q}{\sigma_b}$ is no longer present.

3. If the size of the entries in both the background and model error evolved covariance matrices are large and positive, this will also increase the lower bound.

4. Longer assimilation windows will increase the summation terms in the upper bound. This increase will be more noticeable if the one and infinity norms of M are larger than one.

We now derive bounds in the case where the model is a circulant matrix to obtain more informative bounds.

Theorem 5.2.3 *In addition to the assumptions in Theorem 5.2.2, we assume the model operator $M_{i,i-1} \in \mathbb{R}^{N \times N}$ is a* circulant *matrix with $M_{i,i} = I$. The following bounds on the condition number of $\hat{\mathbf{S}}_p$ then hold:*

$$1 + \tfrac{q}{N(n+1)}\tfrac{1}{\sigma_o^2}\left(\sigma_b^2\lambda_{max}(C_B)\gamma_{min} + \sigma_q^2\lambda_{max}(C_Q)\omega_{min} + \sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\phi_{min}\right)$$
$$\leq \kappa(\hat{\mathbf{S}}_p) \leq 1 + \frac{max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}\left(\sum_{k=0}^{n} ||M||_\infty^k\right)^2, \tag{5.92}$$

where

$$\gamma_{min} = \sum_{i=0}^{n} |\lambda_{min}(M)|^{2i}, \tag{5.93}$$

$$\phi_{min} = \sum_{i=1}^{n}\sum_{j=0}^{n-i} |\lambda_{min}(M)|^{2j}.(2Re((\lambda_{min}(M))^{i})), \tag{5.94}$$

$$\omega_{min} = \sum_{l=1}^{2}\sum_{i=l}^{n}\sum_{j=0}^{n-i} |\lambda_{min}(M)|^{2j}.(2Re((\lambda_{min}(M))^{(l-1)i} - 1). \tag{5.95}$$

Proof: We begin by computing the Rayleigh Quotient of the preconditioned Hessian

$$\mathcal{R}_{\hat{\mathbf{S}}_p}(V_{max}) = \frac{V_{max}^T\hat{\mathbf{S}}_p V_{max}}{V_{max}^T V_{max}}, \tag{5.96}$$

where $V_{max} \in \mathbb{R}^{N(n+1)}$ is a vector of eigenvectors which correspond to the largest eigenvalues of B and Q such that

$$V_{max} = \begin{pmatrix} \boldsymbol{\beta}_{max} \\ \boldsymbol{\xi}_{max} \\ \vdots \\ \boldsymbol{\xi}_{max} \end{pmatrix}, \tag{5.97}$$

where $\boldsymbol{\beta}_{max}$ and $\boldsymbol{\xi}_{max}$ refer to the eigenvector corresponding to the largest eigenvalue of B and Q respectively. We now compute

$$V_{max}^T[\frac{1}{\sigma_o^2}\mathbf{D}^{-1/2}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{-1/2}]V_{max}, \tag{5.98}$$

in segments. We refer to the blocks of $\mathbf{D}^{-1/2}\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{-1/2}$ as $\mathbf{A}_{i,j}$, where i refers to the block row and j refers to the block column. We recall the structure of $\mathbf{L}^{-T}\mathbf{H}^T\mathbf{H}\mathbf{L}^{-1}$,

$$\begin{pmatrix} \sum_{i=0}^{n}(HM^i)^THM^i & \sum_{i=0}^{n-1}(HM^{i+1})^THM^i & \sum_{i=0}^{n-2}(HM^{i+2})^THM^i & \dots & (HM^n)^TH \\ \sum_{i=0}^{n-1}(HM^i)^THM^{i+1} & \sum_{i=0}^{n-1}(HM^i)^THM^i & \sum_{i=0}^{n-2}(HM^{i+2})^THM^{i+1} & \ddots & \vdots \\ \sum_{i=0}^{n-2}(HM^i)^THM^{i+2} & \sum_{i=0}^{n-2}(HM^{i+1})^THM^{i+2} & \sum_{i=0}^{n-2}(HM^i)^THM^i & \ddots & (HM^2)^TH \\ \vdots & \ddots & \ddots & \ddots & (HM)^TH \\ H^THM^n & \dots & H^THM^2 & H^THM & H^TH \end{pmatrix}. \tag{5.99}$$

Block $\mathbf{A}_{1,1}$ yields

$$\begin{aligned} \boldsymbol{\beta}_{max}^T[B^{1/2}\sum_{i=0}^{n}(HM^i)^THM^iB^{1/2}]\boldsymbol{\beta}_{max} =& \frac{q}{N}\lambda_{max}(B)\sum_{i=0}^{n}|\lambda_k|^{2i}, && (5.100) \\ =& \frac{q}{N}\sigma_b^2\lambda_{max}(C_B)\gamma_k, && (5.101) \end{aligned}$$

where λ_k in this proof explicitly refers to the k^{th} eigenvalue of the matrix M. We denote eigenvalues of other matrices as $\lambda_k(A)$, where A is the relevant matrix. The calculation (5.100) is similar to calculations in the proof for Theorem 5.1.2, seen in equations (5.15) and (5.16).

Since B and Q are symmetric, positive-definite and circulant we know,

$$\boldsymbol{\beta}_{max}^TB^{1/2} = \sigma_b\boldsymbol{\beta}_{max}^T\sqrt{\lambda_{max}(C_B)}, \tag{5.102}$$

$$B^{1/2}\boldsymbol{\beta}_{max} = \sigma_b\sqrt{\lambda_{max}(C_B)}\boldsymbol{\beta}_{max}, \tag{5.103}$$

by the circulant matrix eigendecomposition in Theorem 3.3.10. The other blocks of (5.98) will yield expressions similar to (5.101) with mixed $B^{1/2}$ and $Q^{1/2}$ terms on either side of the observation and model operator matrices. We collate the terms emerging from the computation of (5.98) by computing the first block row and first block column together while omitting $\mathbf{A}_{1,1}$ computed in (5.101). The first block row and column computation is as follows

$$\boldsymbol{\xi}_{max}^T[Q^{1/2}\sum_{i=0}^{n-1}(HM^i)^T HM^{i+1}B^{1/2}]\boldsymbol{\beta}_{max} = \lambda_k \frac{q}{N}\sqrt{\lambda_{max}(B)\lambda_{max}(Q)}\sum_{i=0}^{n-1}|\lambda_k|^{2i}, \tag{5.104}$$

$$\boldsymbol{\beta}_{max}^T[B^{1/2}\sum_{i=0}^{n-1}(HM^{i+1})^T HM^{i}Q^{1/2}]\boldsymbol{\xi}_{max} = \bar{\lambda}_k \frac{q}{N}\sqrt{\lambda_{max}(B)\lambda_{max}(Q)}\sum_{i=0}^{n-1}|\lambda_k|^{2i}, \tag{5.105}$$

where (5.104) refers to block $\mathbf{A}_{1,2}$ and (5.105) refers to block $\mathbf{A}_{2,1}$. To represent the emerging summation arising from the first row and column blocks, $\sum\limits_{i=1}^{n+1}\mathbf{A}_{i,2}$ and $\sum\limits_{j=1}^{n+1}\mathbf{A}_{2,j}$, we write

$$\begin{aligned}
&\frac{q}{N}\sqrt{\lambda_{max}(B)\lambda_{max}(Q)}\left(\sum_{i=1}^{n}\sum_{j=0}^{n-i}|\lambda_k|^{2j}.((\bar{\lambda}_k)^i + (\lambda_k)^i)\right)\\
=&\frac{q}{N}\sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\left(\sum_{i=1}^{n}\sum_{j=0}^{n-i}|\lambda_k|^{2j}.(2Re(\lambda_k)^i)\right)\\
=&\frac{q}{N}\sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\phi_k.
\end{aligned} \tag{5.106}$$

Now we compute the remaining blocks $\sum\limits_{i,j=2}^{n+1}\mathbf{A}_{i,j}$. Block $\mathbf{A}_{2,2}$ yields

$$\boldsymbol{\xi}_{max}^T[Q^{1/2}\sum_{i=0}^{n-1}(HM^i)^T HM^{i}Q^{1/2}]\boldsymbol{\xi}_{max} = \frac{q}{N}\sigma_q^2\lambda_{max}(C_Q)\sum_{i=0}^{n-1}|\lambda_k|^{2i}, \tag{5.107}$$

while block $\mathbf{A}_{2,3}$ yields

$$\boldsymbol{\xi}_{max}^T[Q^{1/2}\sum_{i=0}^{n-2}(HM^{i+2})^T HM^{i+1}Q^{1/2}]\boldsymbol{\xi}_{max} = \bar{\lambda}_k\frac{q}{N}\sigma_q^2\lambda_{max}(C_Q)\sum_{i=0}^{n-2}|\lambda_k|^{2i}. \tag{5.108}$$

Finally we have block $\mathbf{A}_{3,2}$,

$$\boldsymbol{\xi}_{max}^T[Q^{1/2}\sum_{i=0}^{n-2}(HM^{i+1})^T HM^{i+2}Q^{1/2}]\boldsymbol{\xi}_{max} = \lambda_k\frac{q}{N}\sigma_q^2\lambda_{max}(C_Q)\sum_{i=0}^{n-2}|\lambda_k|^{2i}. \tag{5.109}$$

We write the summation that encompasses the main block diagonal, $\sum\limits_{i,j=2}^{n+1} \mathbf{A}_{i,j}$ for i=j while excluding the first block, as

$$\frac{q}{N}\sigma_q^2\lambda_{max}(C_Q)\sum_{i=1}^{n}\sum_{j=0}^{n-i}|\lambda_k|^{2j}. \tag{5.110}$$

For the remaining blocks, we examine the sub and super diagonals that sequentially emanate from the main diagonal, which exclude the first block row and column since they have been computed above,

$$\left(\begin{array}{c|cccc|}
\mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} & \cdots & \mathbf{A}_{1,n+1} \\ \cline{2-5}
\mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{3,2} & \cdots & \mathbf{A}_{2,n+1} \\
\mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} & \cdots & \mathbf{A}_{3,n+1} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\mathbf{A}_{n+1,1} & \cdots & \cdots & \cdots & \mathbf{A}_{n+1,n+1} \\ \cline{2-5}
\end{array}\right).$$

A geometric progression in M and M^T manifests itself, which when computing the Rayleigh quotient presents a geometric progression in the eigenvalues of M, similar to that in Section 5.1 equation (5.17). This can be seen in the first terms of the super and sub diagonals (5.108), (5.109) respectively. Summing together the sums that arise from the super and sub-diagonals emanating from the main diagonal consecutively, we arrive at the following expression:

$$\frac{q}{N}\sigma_q^2\lambda_{max}(C_Q)\sum_{i=2}^{n}\sum_{j=0}^{n-i}|\lambda_k|^{2j}.\left(2Re((\lambda_k)^i)\right). \tag{5.111}$$

We now combine (5.110) and (5.111) since they have the same coeffecients which yields

$$\omega_k = \sum_{l=1}^{2}\sum_{i=k}^{n}\sum_{j=0}^{n-i}|\lambda_k|^{2j}.\left(2Re((\lambda_k)^{(l-1)i}) - 1\right). \tag{5.112}$$

Combining (5.101), (5.106), (5.112) and knowing the denominator of the computation (5.98) is equal to $(n+1)$, we have

$$\mathcal{R}_{\widetilde{\mathbf{S}}_p}(V_{max}) = \\ \frac{q}{N(n+1)}\frac{1}{\sigma_o^2}\left(\sigma_b^2\lambda_{max}(C_B)\gamma_k + \sigma_q^2\lambda_{max}(C_Q)\omega_k + \sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\phi_k\right), \tag{5.113}$$

which by the bounds of the Rayleigh quotient, Theorem 3.4.7 gives,

$$\mathcal{R}_{\widetilde{\mathbf{S}}_p}(V_{max}) \geq 1 + \frac{q}{N(n+1)}\frac{1}{\sigma_o^2}(\sigma_b^2\lambda_{max}(C_B)\gamma_{min} + \sigma_q^2\lambda_{max}(C_Q)\omega_{min} \\ + \sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\phi_{min}), \tag{5.114}$$

establishing the lower bound.

For the upper bound we recognise that for a circulant matrix $C \in \mathbb{R}^{N\times N}$ as in Definition (3.3.7), the following is always true:

$$||C||_\infty = ||C||_1. \tag{5.115}$$

The upper bound in Theorem (5.2.2) becomes

$$\kappa(\hat{\mathbf{S}}_p) \leq 1 + \frac{max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}\left(\sum_{k=0}^{n}||M||_\infty^k\right)^2, \tag{5.116}$$

which completes the proof, as required. ■

In both the upper and lower bounds the contribution of the eigenvalues and norm of M are influential. Thus the effect of the assimilation window length still exists in both bounds, and the lower bound has a further dependency on the eigenvalues of M. The lower bound operators γ, ω and ψ all depend on the assimilation window length and the size of the smallest eigenvalue of M, all multiplied by either the largest eigenvalue of the background or model error covariance matrices. The upper bound is much clearer in that it quadratically depends on the infinity-norm of M. Therefore the only definitive message we can deduce here is that bounds suggest that the assimilation window length will increase the condition number. The bounds also suggest that the condition number of $\mathbf{D}$ no longer affects $\kappa(\mathbf{S}_p)$. We instead have the ratio $\frac{max\left\{\sigma_b^2\lambda_{max}(C_B),\sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}$ now influencing both bounds.

In a bid to extract more meaningful information we now deduce bounds using the 1D advection model.

Theorem 5.2.4 *In addition to the assumptions in Theorem 5.2.2, we assume the model operator $M_{i,i-1} \in \mathbb{R}^{N\times N}$ represents the matrix presented by the discretisation of the advection equation using the upwind scheme, (3.71) with $M_{i,i} = I$. Then for Courant number $\mu \in (-1, 0)$ the following bounds on the condition number of $\hat{\mathbf{S}}_p$ therefore hold:*

$$1 + \frac{q}{N(n+1)}\frac{1}{\sigma_o^2}\left(\sigma_b^2\lambda_{max}(C_B)\gamma_{min}^{adv} + \sigma_q^2\lambda_{max}(C_Q)\omega_{min}^{adv} + \sigma_b\sigma_q\sqrt{\lambda_{max}(C_B)\lambda_{max}(C_Q)}\phi_{min}^{adv}\right)$$
$$\leq \kappa(\hat{\mathbf{S}}_p) \leq 1 + \frac{max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}(n+1)^2 \quad (5.117)$$

where

$$\gamma_{min}^{adv} = \frac{1-|1+2\mu|^{2(n+1)}}{1-|1+2\mu|^2}, \quad (5.118)$$

$$\phi_{min}^{adv} = \sum_{i=1}^{n}(2(1+2\mu)^i).\left(\frac{1-|1+2\mu|^{2(n-i+1)}}{1-|1+2\mu|^2}\right), \quad (5.119)$$

$$\omega_{min}^{adv} = \left(\frac{1-|1+2\mu|^{2n}}{1-|1+2\mu|^2}\right) + \sum_{i=2}^{n}(2(1+2\mu)^i).\left(\frac{1-|1+2\mu|^{2(n-i+1)}}{1-|1+2\mu|^2}\right). \quad (5.120)$$

Proof: The absolute row and column sums of a circulant matrix are all equal. For the advection equation we know

$$||M||_\infty = (1+\mu) + (-\mu) = 1. \quad (5.121)$$

We compute the geometric series in Theorem (5.2.3) for the advection equation,

$$\sum_{i=0}^{n}||M||_\infty^i = \sum_{i=0}^{n}(1)^i = n+1, \quad (5.122)$$

substituting this into Theorem (5.2.3), we establish the upper bound

$$\kappa(\hat{\mathbf{S}}_p) \leq 1 + \frac{max\left\{\sigma_b^2\lambda_{max}(C_B), \sigma_q^2\lambda_{max}(C_Q)\right\}}{\sigma_o^2}(n+1)^2. \quad (5.123)$$

For the lower bound we have

$$\lambda_{min}(M)\begin{cases} = 1+2\mu & \text{(for } N \text{ even)} \\ \geq 1+2\mu\frac{1-|1+2\mu|^{2(n+1)}}{1-|1+2\mu|^2} & \text{(for } N \text{ odd)}\end{cases} \quad (5.124)$$

We substitute $\lambda_{min}(M) = 1 + 2\mu$, into the lower bound expression presented in Theorem 5.2.3 and compute the values of γ_{min}, ϕ_{min} and ω_{min}:

$$\gamma_{min}^{adv} = \sum_{i=0}^{n} |1 + 2\mu|^{2i} = \frac{1 - |1 + 2\mu|^{2(n+1)}}{1 - |1 + 2\mu|^2}, \tag{5.125}$$

and

$$\phi_{min}^{adv} = \sum_{i=1}^{n} \sum_{j=0}^{n-i} |1 + 2\mu|^{2j}.(2(1 + 2\mu)^i) = \sum_{i=1}^{n} (2(1 + 2\mu)^i). \left(\frac{1 - |1 + 2\mu|^{2(n-i+1)}}{1 - |1 + 2\mu|^2} \right), \tag{5.126}$$

and

$$\omega_{min}^{adv} = \sum_{l=1}^{2} \sum_{i=k}^{n} \sum_{j=0}^{n-i} |1 + 2\mu|^{2j}.(2Re(1 + 2\mu)^{(l-1)i} - 1), \tag{5.127}$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{n-i} |1 + 2\mu|^{2j} + \sum_{i=2}^{n} \sum_{j=0}^{n-i} |1 + 2\mu|^{2j}.(2(1 + 2\mu)^i - 1), \tag{5.128}$$

$$= \sum_{i=1}^{n} \left(\frac{1 - |1 + 2\mu|^{2(n-i+1)}}{1 - |1 + 2\mu|^2} \right) + \sum_{i=2}^{n} \left(\frac{1 - |1 + 2\mu|^{2(n-i+1)}}{1 - |1 + 2\mu|^2} \right).(2(1 + 2\mu)^i - 1), \tag{5.129}$$

$$= \left(\frac{1 - |1 + 2\mu|^{2n}}{1 - |1 + 2\mu|^2} \right) + \sum_{i=2}^{n} \left(\frac{1 - |1 + 2\mu|^{2(n-i+1)}}{1 - |1 + 2\mu|^2} \right).(2(1 + 2\mu)^i), \tag{5.130}$$

which completes the proof. ■

We see here that the lower bound is similar to that of the unpreconditioned Hessian in Theorem 5.1.3, in that the parameters γ, ω and ψ all involve the Courant number and the length of the assimilation window governs the number of terms in the sum. These parameters can be amplified or otherwise by the largest eigenvalue of B and Q.

The upper bound is also similar to the upper bound in Theorem 5.1.3, showing the quadratic influence of the assimilation window length which can be amplified or otherwise by the size of the ratio $\frac{max\left\{\sigma_b^2 \lambda_{max}(C_B), \sigma_q^2 \lambda_{max}(C_Q)\right\}}{\sigma_o^2}$. We also see that $\kappa(\mathbf{D})$ is absent from both bounds as expected.

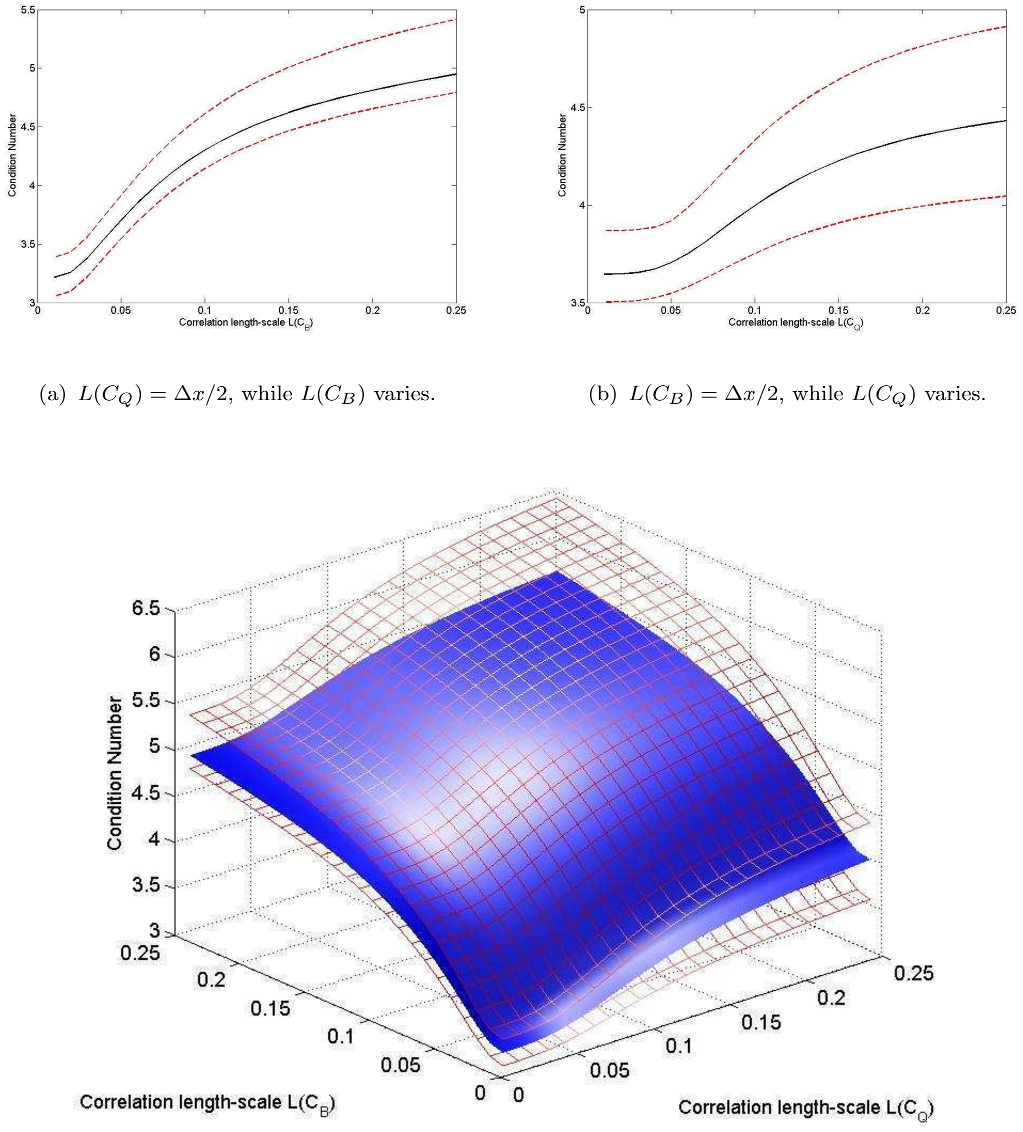
We now demonstrate the bounds through numerical experiments on the condition number of $\hat{\mathbf{S}}_p$.

5.2.1 Numerical Results

The parameter settings for the experiments in this section are identical to the settings in Section 5.1.3 unless stated otherwise. We show the alleviation of the sensitivities previously exhibited by $\mathbf{S}_p$ in the preconditioned Hessian $\hat{\mathbf{S}}_p$, while also demonstrating the quality of the theoretical bounds obtained in the previous section.

5.2.1.1 Experiment 1: Correlation Length-Scales

We begin by showing the sensitivity of the condition number of $\hat{\mathbf{S}}_p$ to increasing the correlation length-scales of the matrices composing $\mathbf{D}$.



(a) $L(C_Q) = \Delta x/2$, while $L(C_B)$ varies.

(b) $L(C_B) = \Delta x/2$, while $L(C_Q)$ varies.

(c) $L(C_B)$ and $L(C_Q)$ varying.

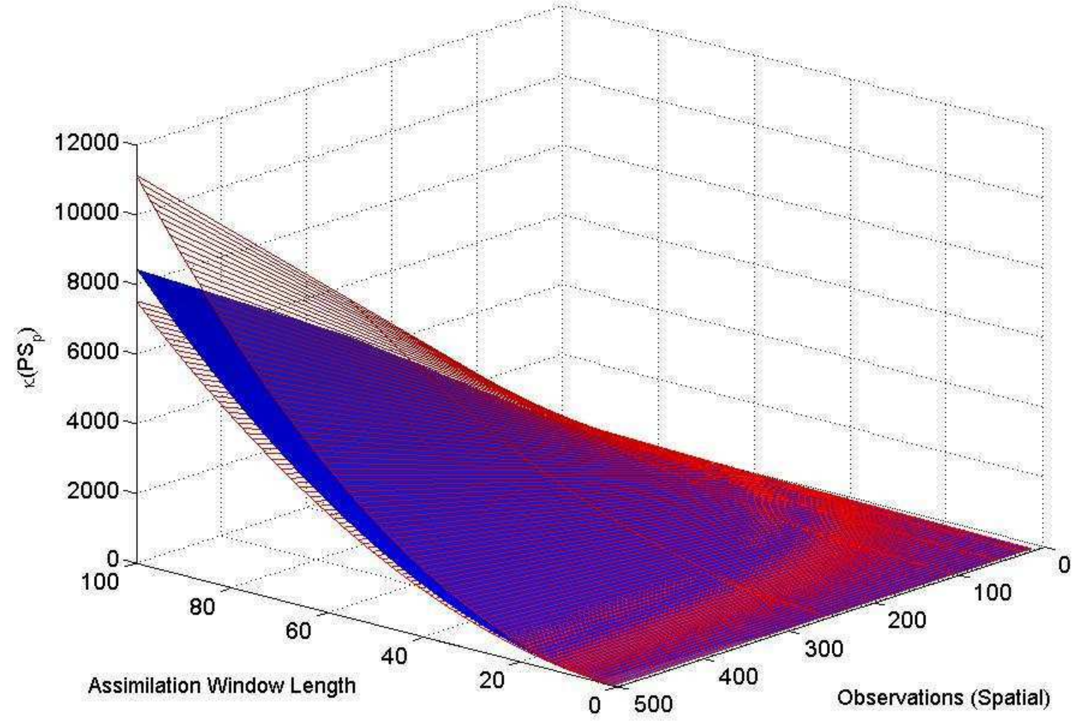
Figure 5.6: Graph (a) and (b) $\kappa(\hat{\mathbf{S}}_p)$ (black line) and theoretical bounds (red dotted lines) plotted against $L(C_B)$ (a) and $L(C_Q)$ (b). Graph (c) is a 3D representation of (a) and (b), $\kappa(\hat{\mathbf{S}}_p)$ (blue surface) with theoretical bounds (red-mesh surfaces), against $L(C_B)$ and $L(C_Q)$.

We state the correlation length-scales in terms of the grid spacing of the model, so for example $\Delta x = 0.1$. In Figure 5.6(a) the condition number rises more rapidly in comparison to 5.6(b) since $C_B = C_{SOAR}$ is known to be more ill-conditioned than $C_Q = C_{LAP}$, [41]. The main message from Figures 5.6(a) and 5.6(b) is that the rise in correlation length-scale of the correlation matrices composing $\mathbf{D}$ now has a greatly reduced effect on $\kappa(\hat{\mathbf{S}}_p)$ compared to $\kappa(\mathbf{S}_p)$ as shown in Section 5.1.3,

Figure 5.1. We also see that the bounds for $\kappa(\hat{\mathbf{S}}_p)$ are a good estimate of the condition number.

5.2.1.2 Experiment 2: Assimilation Window Length and Observation Density

We now examine the effects of varying observation density and assimilation window length on the condition number of $\hat{\mathbf{S}}_p$.



(a)

Figure 5.7: $\kappa(\hat{\mathbf{S}}_p)$ (blue surface) and theoretical bounds (red-mesh surfaces) with assimilation window length, n, and number of spatial observations, q.

Figure 5.7 shows that the condition number of $\hat{\mathbf{S}}_p$ grows as the assimilation window length increases and as the number of spatial observations at every assimilation step is increased. This is not dissimilar from the unpreconditioned problem as shown in Section 5.1.3, Figure 5.5. The bounds in Theorem 5.2.3, show a dependence on the assimilation window length, the upper bound shows a potential quadratic influence on the assimilation window length, which becomes much clearer in the upper bound of Theorem 5.2.4. Examining Figure 5.7 further, we see a quadratic increase of the actual condition number of $\kappa(\hat{\mathbf{S}}_p)$ for example with 500 observed points at $n = 50$, $\kappa(\hat{\mathbf{S}}_p) = 2026$, and at $n = 100$ $\kappa(\hat{\mathbf{S}}_p) = 8056$.

In this section we have demonstrated the bounds derived in Section 5.2 of the preconditioned Hessian $\hat{\mathbf{S}}_p$.

5.2.2 Summary

We have shown through numerical experiments that the original exhibited sensitivity of the unconditioned Hessian $\mathbf{S}_p$ to $\mathbf{D}$ has been greatly reduced. The absence of $\kappa(\mathbf{D})$ can be seen in Theorems 5.2.1, 5.2.2 and 5.2.3 when compared to the bounds derived for the unconditioned Hessian in Section 5.1. The numerical experiments in Figure 5.6 compared to Figure 5.1 also confirm the alleviation of the sensitivity of $\kappa(\mathbf{S}_p)$ to $\kappa(\mathbf{D})$, since the rise in correlation length-scale increases $\kappa(\mathbf{D})$ (shown in Figure 5.1).

The preconditioner chosen in this thesis does not address any ill-conditioning which could arise from the second term of $\mathbf{S}_p$. We see that $\mathbf{S}_p$ and $\hat{\mathbf{S}}_p$ both exhibit sensitivities to the length of the assimilation window and the spatial observation density through the theory (Theorems 5.2.2 and 5.1.2) and in Figures 5.7 and Figure 5.5 in Section 5.1.3. This is an inherent trait of $\mathbf{S}_p$ as well as the preconditioned Hessian $\hat{\mathbf{S}}_p$.

We also notice in the experiments that the lower bound is usually poorer than the upper bound. The Rayleigh quotient was used to obtain the lower bound, while the Courant Fischer theorem (Theorem 3.4.2) was used to obtain the upper bound. Although the Rayleigh quotient yields expressions that have aided in our analysis, it has proven to be a poorer estimator than the Courant Fisher theorem.

We now show results of the effect of the condition number sensitivities found in this chapter on the minimisation of $\mathcal{J}(\mathbf{p})$ and its preconditioned counter-part.

5.3 Convergence Results: Model Error Formulation vs Preconditioned Model Error formulation

We begin by designing numerical experiments for both the unpreconditioned problem $\mathcal{J}(\mathbf{p})$ and the preconditioned problem $\hat{J}(\delta\mathbf{z})$. We perform data assimilation experiments which focus on the minimisation problems $\mathcal{J}(\mathbf{p})$ and $\hat{J}(\delta\mathbf{z})$, rather than experiments on the Hessian themselves. We now discuss the experimental design for our experiments.

5.3.1 Experimental Design

The model is the 1-dimensional linear advection equation discretised using the upwind scheme, yielding a matrix M as in (3.71). The spatial domain is size $N = 50$ with a spatial resolution of $\Delta x = 0.01$. We use time-intervals of $\Delta t = 0.01$ and a wave speed of $a = -1$, thus giving us a Courant number of $\mu = -1$.

We choose the the background error, $B_0 = \sigma_b^2 C_{SOAR}$, such that the correlation length-scale $L = \Delta x = 0.01$ and $\sigma_b = 1$. The model error, $Q_i = \sigma_q^2 C_{LAP}$ is such that the correlation length-scale $L = \Delta x = 0.01$ and $\sigma_q = 1$. The observation error is such that $R_i = \sigma_o^2 I$, where $\sigma_o = 1$. We take observations every $\Delta q = 3$ model time-steps, $n = 60$ in total, with 5 equally spaced observed grid-points out of $N = 50$ grid-points per assimilation step.

We use the linear CG method as described in Section 3.2.1 to minimise $\mathcal{J}(\mathbf{x})$, with a iterative minimisation tolerance (as described in Chapter 3, Section 3.2.4) of $\tau = 10^{-10}$ throughout this section. The solution relative errors is calculated in the same way as shown in Chapter 4 Section 4.1.6.

5.3.2 Experimental Results 1: Correlation Length-Scales

We now examine the effect of varying correlation length-scales of the background and model error covariance matrices composing $\mathbf{D}$ on minimisation problems $\mathcal{J}(\mathbf{p})$ and $\hat{J}(\delta\mathbf{z})$.

The model error length-scale remains at $L(C_Q) = \Delta x/2$, while we vary the background correlation length-scale $L(C_B)$ to understand the impact it has on the minimisation process. From the insight gained through the bounds in Sections 5.1 and 5.2, we expect the rise in correlation length-scale to increase the condition number and increase the number of iterations required for convergence of the unpreconditioned problem but not for the preconditioned problem.

Correlation length-scale	No. of iterations		Solution relative error		Condition number		
$L(C_B)$	$\mathcal{J}(\mathbf{p})$	$\hat{J}(\delta\mathbf{z})$	$\mathcal{J}(\mathbf{p})$	$\hat{J}(\delta\mathbf{z})$	$\mathbf{S}_p$	$\hat{\mathbf{S}}_p$	$\mathbf{D}$
0.01	47	22	0.12	0.12	294	18	58
0.02	85	24	0.13	0.13	2047	32	837
0.03	116	26	0.13	0.13	6967	102	4323
0.04	138	26	0.14	0.14	17558	236	13889
0.05	155	27	0.13	0.13	37483	455	33665
0.06	189	28	0.14	0.14	70892	783	67961
0.07	204	29	0.14	0.14	121743	1239	121022
0.08	214	29	0.14	0.14	193774	1846	196977
0.09	231	29	0.13	0.13	290579	2626	299839
0.10	246	29	0.14	0.14	415651	3598	433526

Table 5.1: Convergence Figures: Varying correlation length-scales

We see in Table 5.1 that as the correlation length-scale of the background matrix increases to $L(C_B) = 10\Delta x$, the condition numbers of the unconditioned Hessian, the preconditioned Hessian and $\mathbf{D}$ all increase. The condition number of $\hat{\mathbf{S}}_p$ is $\mathcal{O}(10^3)$ smaller than the other condition numbers as early as $L(C_B) = 4\Delta x$. The number of iterations for $\hat{J}(\delta\mathbf{z})$ are of order $\mathcal{O}(10)$ less than $\mathcal{J}(\mathbf{p})$. We also see the solution accuracies are not effected since we are solving to the same solution accuracy.

We can conclude that as the condition number of $\mathbf{D}$ increases, the condition numbers of $\mathbf{S}_p$ and $\hat{\mathbf{S}}_p$ and the number of iterations to minimise $\mathcal{J}(\mathbf{p})$ and $\hat{J}(\delta\mathbf{z})$ also increase respectively. The preconditioned Hessian condition number increases at a much reduced rate and the number of iterations of the preconditioned problem barely increase at all.

5.3.3 Experimental Results 2: Assimilation Window Length

We now show the effect of the length of the assimilation window on the minimisation problem. From our results on the condition number both theoretically and numerically, we know that the length of the assimilation window increases the condition number of $\mathbf{S}_p$ and $\hat{\mathbf{S}}_p$. We also expect that this will increase the number of iterations required for convergence.

The experiment parameters are identical to the previous experiment with $L(C_B) = L(C_Q) = \Delta x$.

Assimilation window length	No. of iterations		Solution relative error		Condition number	
n	$\mathcal{J}(\mathbf{p})$	$\hat{J}(\delta\mathbf{z})$	$\mathcal{J}(\mathbf{p})$	$\hat{J}(\delta\mathbf{z})$	$\mathbf{S}_p$	$\hat{\mathbf{S}}_p$
1	43	7	0.35	0.35	58	6
10	46	18	0.20	0.20	135	8
20	48	23	0.11	0.11	317	19
30	54	27	0.07	0.07	611	37
40	57	30	0.05	0.05	1016	63
50	59	32	0.04	0.04	1529	96
60	63	35	0.03	0.03	2150	135
70	66	39	0.03	0.03	2879	182
80	71	42	0.02	0.02	3717	236
90	70	43	0.02	0.02	4663	296

Table 5.2: Convergence figures: Varying assimilation window length

We see from Table 5.2 that as the assimilation window length increases so do the number of iterations to minimise $\mathcal{J}(\mathbf{p})$ and $\hat{J}(\delta\mathbf{z})$. We also see that the numerical condition numbers of both $\mathbf{S}_p$ and $\hat{\mathbf{S}}_p$ increase. The rate of increases in the condition numbers of $\mathbf{S}_p$ and $\hat{\mathbf{S}}_p$ differ in that $\mathbf{S}_p$ increases much more rapidly, as we expected. However, the rate of increase in the number of iterations required for convergence of $\mathcal{J}(\mathbf{p})$ and $\hat{J}(\delta\mathbf{z})$ are very similar. We also see a decrease in the overall relative solution error as the length of the assimilation window increases, by an order of magnitude for both quantities.

We conclude that as the assimilation window is lengthened the condition numbers of both the unconditioned and preconditioned problems both increase, as do the number of iterations to solve both problems to the same iterative tolerance. We also conclude that as the assimilation window increases then the relative solution error decreases, as observed in Chapter 4, Experiment 3. The reason for the decrease in relative solution error however, is because the algorithms are permitted to iterate until the tolerance is reached without stopping it prematurely.

We now summarise this chapter.

5.4 Summary

The aim of this chapter was to explore the sensitivities of the problem $\mathcal{J}(\mathbf{p})$ by bounding the condition number of the Hessian matrix (2.38). Since the eigenvalues and eigenvectors of the Hessian matrices of these large non-linear least squares problems are not explicitly known, we bounded the condition number as an estimator.

We derived bounds for the unconditioned Hessian $\mathbf{S}_p$. We then chose a route of preconditioning that alleviates the evidently strong dependance of $\mathbf{S}_p$ on the matrix $\mathbf{D}$. We also derived bounds on the resultant preconditioned Hessian $\hat{\mathbf{S}}_p$. The sensitivities exposed by the bounds were demonstrated through numerical experiments using the 1D advection equation. Through the bounds, we demonstrated the following sensitivities both theoretically and numerically:

1. Error variance ratios.
2. Correlation length-scales.
3. Assimilation window length.

More specifically we showed:

1. As the error variance ratio of the background and model error σ_b/σ_q increases, so does the condition number of $\mathbf{S}_p$.
2. As the observation error variance σ_o decreases, the condition number of $\mathbf{S}_p$ increases. This is because decreasing σ_o increases the size of the ratio
$$\frac{\max/\min\left\{\sigma_b^2\lambda_{min}(C_B), \sigma_q^2\lambda_{min}(C_Q)\right\}}{\sigma_o}.$$
This sensitivity also holds for the Hessian of the preconditioned problem $\hat{\mathbf{S}}_p$.
3. Increasing the correlation length-scale of the background error covariance matrix increases the condition number of $\mathbf{D}$ and hence $\mathbf{S}_p$.
4. Increasing the correlation length-scale of the model error covariance matrix increases the condition number of $\mathbf{D}$ and hence $\mathbf{S}_p$.
5. Increasing the length of the assimilation window increases the condition number of $\mathbf{S}_p$.
6. Preconditioning with $\mathbf{D}$ improves condition number sensitivity to an ill-conditioned $\mathbf{D}$ matrix. The condition number of $\hat{\mathbf{S}}_p$ was shown to have greatly reduced sensitivity to increasing correlation length-scales in the background and model error covariance matrices compared with $\mathbf{S}_p$.

In addition to analysing the condition number of the unconditioned and preconditioned Hessians, we showed that the convergence of the preconditioned problem $\hat{J}(\delta\mathbf{z})$ is no longer sensitive to the increase in correlation length-scale of the matrices inside $\mathbf{D}$, and hence the condition number of $\mathbf{D}$. The convergence rate is much improved for the preconditioned problem $\hat{J}(\delta\mathbf{z})$ over the original problem $\mathcal{J}(\mathbf{p})$. We also showed that the condition number of the preconditioned problem is still sensitive to the length of the assimilation window and spatial observation density, which in turn was shown to affect the number of iterations required to converge.

This concludes the analysis of the Hessian condition number and convergence rates of the wc4DVAR $\mathcal{J}(\mathbf{p})$ formulation and its preconditioned counter-part *complement*. It is important to realise that these results are illustrative examples of the behaviour we expected to see from the theory we have derived. We now consider the alternative formulation $\mathcal{J}(\mathbf{x})$, (2.33).

Chapter 6

Conditioning of the State Formulation: $\mathcal{J}(\mathbf{x})$

The previous chapter was dedicated to the conditioning of the Hessian $\mathbf{S}_p$. We bounded the condition number of $\mathbf{S}_p$ and uncovered the parameters exhibiting the largest sensitivities with respect to the Hessian condition number. We found the Hessian $\mathbf{S}_p$ to be sensitive to the $\mathbf{D}$ matrix, containing the background and model error correlations. We then preconditioned the Hessian using the symmetric square root of $\mathbf{D}$ which improved the condition number sensitivity characteristics with respect to the condition number of $\mathbf{D}$. We then demonstrated the sensitivities obtained from the bounds through numerical experiments on the condition number. We further demonstrated the effect of some of these sensitivities on the number of iterations required for convergence.

In this chapter we bound the condition number of

$$\mathbf{S}_x = \begin{pmatrix} B_0^{-1}+M_1^T Q_1^{-1} M_1 & -M_1^T Q_1^{-1} & & & \\ -Q_1^{-1} M_1 & Q_1^{-1}+M_2^T Q_2^{-1} M_2 & -M_2^T Q_2^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & & \\ & & -Q_{n-1}^{-1} M_{n-1} & Q_{n-1}^{-1}+M_n^T Q_n^{-1} M_n & -M_n^T Q_n^{-1} \\ & & & -Q_n^{-1} M_n & Q_n^{-1} \end{pmatrix} +$$

$$\begin{pmatrix} H_0^T R_0^{-1} H_0 & & & \\ & H_1^T R_1^{-1} H_1 & & \\ & & \ddots & \\ & & & H_n^T R_n^{-1} H_n \end{pmatrix}. \tag{6.1}$$

Through bounding the condition number of $\mathbf{S}_x$ we uncover the parameter sensitivities and demonstrate these through numerical experiments on the condition number. We then show that these sensitivities can also effect the minimisation of $\mathcal{J}(\mathbf{x})$ by examining their effect on the number of iterations required for convergence and solution accuracy.

We begin by deriving new bounds on the condition number of $\mathbf{S}_x$.

6.1 Theoretical Results: Bounding the Condition Number of $\mathbf{S}_x$

The following theorem bounds the spectral condition number of $\mathbf{S}_x$,

Theorem 6.1.1 *Let $\mathbf{D} \in \mathbb{R}^{N(n+1)\times N(n+1)}$ be our background and model error covariance matrix. Suppose we take $q < N$ observations at each time interval t_i for $i = 0, ..., n$, with observation error covariance $R_i \in \mathbb{R}^{q\times q}$. Let $H_i \in \mathbb{R}^{q\times N}$ for $i = 0, .., n$, be the observation operator. Finally, let $M_i \in \mathbb{R}^{N\times N}$ for each time step $i = 1, .., n$ represent the model equations. Then the following bounds are satisfied by the spectral condition number of $\mathbf{S}_x$:*

$$\frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})} \leq \kappa(\mathbf{S}_x) \leq \frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}, \tag{6.2}$$

Proof: We begin by bounding $\lambda_{min}(\mathbf{S}_x)$ and $\lambda_{max}(\mathbf{S}_x)$ using Theorem 3.4.2, yielding

$$\begin{aligned}\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{min}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \leq & \lambda_{min}(\mathbf{S}_x) \\ & \leq \lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}^{-1}),\end{aligned} \tag{6.3}$$

and

$$\begin{aligned}\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{min}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \leq & \lambda_{max}(\mathbf{S}_x) \\ & \leq \lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}).\end{aligned} \tag{6.4}$$

We take the upper bound of $\lambda_{max}(\mathbf{S}_x)$ and lower bound of $\lambda_{min}(\mathbf{S}_x)$ to give us an upper bound on the condition number of $\mathbf{S}_x$

$$\kappa(\mathbf{S}_x) \leq \frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{min}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}, \tag{6.5}$$

similarly for the lower bound on $\kappa(\mathbf{S}_x)$, we take the lower bound of $\lambda_{max}(\mathbf{S}_x)$ and upper bound of $\lambda_{min}(\mathbf{S}_x)$ yielding

$$\kappa(\mathbf{S}_x) \geq \frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{min}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}. \tag{6.6}$$

Our assumption on the observation operator was that we take less observations than the state vector, $q < N$. So $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ will be rank deficient, a singular matrix with possibly more than one zero eigenvalue, thus $\lambda_{min}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) = 0$. Therefore,

$$\frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})} \leq \kappa(\mathbf{S}_x) \leq \frac{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) + \lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}, \tag{6.7}$$

as required. ■

Comparing these bounds to the bounds in Theorem 5.1.1, we notice the emphasis here is on the extreme eigenvalues of the term $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$, where in the $\mathbf{S}_p$ bounds the emphasis was clearly on $\kappa(\mathbf{D})$. The bounds in Theorem 6.1.1 can be expressed such that

$$\frac{\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}{1 + \frac{\lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}} \leq \kappa(\mathbf{S}_x) \leq \kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) \left(1 + \frac{\lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})}{\lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}\right), \tag{6.8}$$

which shows the influence of $\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$, instead of just $\kappa(\mathbf{D})$ when compared to Theorem 5.1.1. We also see that as $\lambda_{max}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}) \to 0$ both bounds tend to the condition number of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$. Therefore the bounds in Theorem 5.1.1 show that the condition number of $\mathbf{S}_x$ depends heavily on $\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$.

We now make more specific assumptions to obtain a more meaningful estimate on the condition number of $\mathbf{S}_x$

Theorem 6.1.2 *Let $B_0 \in \mathbb{R}^{N\times N}$ be the background error covariance matrix such that $B_0 = \sigma_b^2 C_B$, where C_B is a symmetric, positive-definite circulant matrix and $\sigma_b^2 > 0$ is the background error variance. Let $Q_i \in \mathbb{R}^{N\times N}$ be the model error covariance matrix such that $Q_i = \sigma_q^2 C_Q$, for $i = 1, ..., n$, where C_Q is a symmetric, positive-definite circulant matrix and $\sigma_q^2 > 0$ is the model error variance. Assume $q < N$ observations are taken with the same uncorrelated error variance at each time interval such that $R_i \in \mathbb{R}^{q\times q}$, $R_i = \sigma_o^2 I_q$ for $i = 0, ..., n$, where I_q is a $q \times q$ identity matrix and $q < N$. Assume that observations of the parameter are made at the same grid points at each time interval such that $H_i^T H_i = H^T H \in \mathbb{R}^{N\times N}$, so $H^T H$ is a diagonal matrix with unit entries at observed points and zeros otherwise. Finally, we assume that $M_i = M \in \mathbb{R}^{N\times N}$ for $i = 1, .., n$ and $M_0 = I_N$ where M is a circulant matrix. Then the following bounds hold on the spectral condition number of $\mathbf{S}_x$*

$$\frac{\frac{\sigma_q^2}{\sigma_o^2}\frac{q(n+1)}{N}+n\Big(\lambda_{max}(C_Q^{-1})+\lambda_{min}(M^T C_Q^{-1}M)-2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))\Big)+2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{max}(C_B^{-1})}{\frac{\sigma_q^2}{\sigma_o^2}\frac{q(n+1)}{N}+n\Big(\lambda_{min}(C_Q^{-1})+\lambda_{max}(M^T C_Q^{-1}M)-2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))\Big)+2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{min}(C_B^{-1})} \leq$$

$$\kappa(\mathbf{S}_x) \leq \frac{\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) + 2\sigma_q^{-2}\lambda_{max}(C_Q^{-1})\lambda_{max}(M)}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})} \quad (6.9)$$

where $\mathbf{S}_{x(i,i)}$ for $i = 1, ..., n+1$ refers to the main block diagonal entries of $\mathbf{S}_x$.

Proof: We begin by applying Theorem 3.4.9 with the intent of improving the upper bound on the spectral condition number of $\mathbf{S}_x$. Let G_i be the set of Gershgorin circles such that

$$G_i : ||(\mathbf{S}_{x(i,i)} - \lambda I)^{-1}||^{-1} \leq \sum_{\substack{i\neq j\\ j=1}}^{n} ||\mathbf{S}_{x(i,j)}||, \quad (6.10)$$

where $\lambda \in [\lambda_{min}(\mathbf{S}_x), \lambda_{max}(\mathbf{S}_x)]$ and $\mathbf{S}_{x(i,j)}$ refers to the block matrix on the i^{th} block row and j^{th} block column. The left hand side of (6.10) for $\mathbf{S}_x$ yields

$$||(\mathbf{S}_{x(i,i)} - \lambda I)^{-1}||_2^{-1} = \sqrt{\lambda_{min}\left((\mathbf{S}_{x(i,i)} - \lambda I)^H(\mathbf{S}_{x(i,i)} - \lambda I)\right)}. \tag{6.11}$$

To obtain an expression for the minimum eigenvalue we define eigenvectors x_k^i with corresponding eigenvalues $\mu_k^{(i,i)}$ of $\mathbf{S}_{x(i,i)}$ for $i = 1, ..., n+1$ and $k = 1, ..., N$. It follows that

$$\left(\mathbf{S}_{x(i,i)} - \lambda I\right) x_k^i = (\mu_k^{(i,i)} - \lambda) x_k^i, \tag{6.12}$$

$$(x_k^i)^H \left(\mathbf{S}_{x(i,i)} - \lambda I\right)^H = (\bar{\mu}_k^{(i,i)} - \bar{\lambda})(x_k^i)^H, \tag{6.13}$$

and therefore

$$\lambda_{min}\left((\mathbf{S}_{x(i,i)} - \lambda I)^H(\mathbf{S}_{x(i,i)} - \lambda I)\right) = \min_{i,k} \sqrt{\frac{(\mu_k^{(i,i)} - \lambda)(\bar{\mu}_k^{(i,i)} - \bar{\lambda})(x_k^i)^H x_k^i}{(x_k^i)^H x_k^i}}, \tag{6.14}$$

$$= \min_{i,k} |\mu_k^{(i,i)} - \lambda|. \tag{6.15}$$

To further illustrate the meaning of (6.15), we list the constituents of the set of Geršgorin circles G_i,

$$G_i : min\Big\{|\mu_1^{(1,1)} - \lambda|, ..., |\mu_N^{(1,1)} - \lambda|, |\mu_1^{(2,2)} - \lambda|, ..., |\mu_N^{(2,2)} - \lambda|, ...,$$

$$..., |\mu_1^{(n+1,n+1)} - \lambda|, ..., |\mu_N^{(n+1,n+1)} - \lambda|\Big\} \leq \sum_{\substack{i \neq j \\ j=1}}^{n} ||\mathbf{S}_{x(i,j)}||_2 \tag{6.16}$$

where $\mu_{1,...,N}^{(i,i)} \in [\lambda_{min}(\mathbf{S}_{x(i,i)}), \lambda_{max}(\mathbf{S}_{x(i,i)})]$.

The eigenvalues of $\mathbf{S}_x$ will always satisfy the inequality (6.16) above. This expression forms the set G_i which composes the well-known 'Geršgorin circles' in the complex plane. It is understood from the conventional scalar Geršgorin circle theorem that the eigenvalues will lie in the union of these regions G_i. The block analogue of the Geršgorin theorem as used here is no different. There will be $N(n+1)$ Geršgorin circles for $\mathbf{S}_x$ in total with centres $\mu_{1,...,N}^{(i,i)}$ and corresponding radiuses such that

$$r_j = \sum_{\substack{i \neq j \\ j=1}}^{n+1} ||\mathbf{S}_{x(i,j)}||_2. \tag{6.17}$$

We know the eigenvalues of $\mathbf{S}_x$ will lie on the positive real line since it is positive definite. Using (6.16) and recalling that $\mathbf{S}_x$ is block tri-diagonal, we have the following Geršgorin circles:

$$\begin{aligned}
|\mu_{1,...,N}^{(1,1)} - \lambda| &\leq ||\mathbf{S}_{x(1,2)}||_2, && (6.18)\\
|\mu_{1,...,N}^{(2,2)} - \lambda| &\leq ||\mathbf{S}_{x(2,1)}||_2 + ||\mathbf{S}_{x(2,3)}||_2, && (6.19)\\
&\vdots \\
|\mu_{1,...,N}^{(n,n)} - \lambda| &\leq ||\mathbf{S}_{x(n,n-1)}||_2 + ||\mathbf{S}_{x(n,n+1)}||_2, && (6.20)\\
|\mu_{1,...,N}^{(n+1,n+1)} - \lambda| &\leq ||\mathbf{S}_{x(n+1,n)}||_2, && (6.21)
\end{aligned}$$

all or some of which *could* contain a certain number of eigenvalues of $\mathbf{S}_x$, but the union of which will *definitely* contain *all* the eigenvalues of $\mathbf{S}_x$.

We now turn our attention to the radii. We know that for any $A \in \mathbb{R}^{m \times m}$, A^T shares the same determinant and the same eigenvalues albeit with different eigenvectors. It follows that the eigenvalues of $A^T A$ are the same as the eigenvalues of AA^T. Therefore the 2-norm of all the terms not on the main block diagonal are equal since,

$$|| - M^T Q^{-1}||_2 = || - Q^{-1} M||_2. \tag{6.22}$$

The two terms in (6.22) are the only two possible off-diagonal block terms in $\mathbf{S}_x$. The two possible radii for the main block diagonal terms are

$$\begin{aligned}
&||\mathbf{S}_{x(1,2)}||_2 = ||\mathbf{S}_{x(n+1,n)}||_2, && (6.23)\\
&||\mathbf{S}_{x(i,i-1)}||_2 + ||\mathbf{S}_{x(i,i+1)}||_2 \text{ (for } i = 2, ..., n). && (6.24)
\end{aligned}$$

The expression (6.23) refers to the smaller radii associated with blocks $\mathbf{S}_{x(1,1)}$ and $\mathbf{S}_{x(n+1,n+1)}$. The larger radius is associated with the remaining blocks $\mathbf{S}_{x(i,i)}$, for $(i = 2, ..., n)$.

To compute an explicit expression for the radii, we utilise the fact that our covariance and model matrices are circulant and have the Fourier eigendecomposition structure as in Theorem 3.3.10,

$$\begin{aligned}||\mathbf{S}_{x(1,2)}||_2 &= || - M^TQ^{-1}||_2 = | - \sigma_q^{-2}|.||M^TC_Q^{-1}||_2,\\ &= \sigma_q^{-2}||F\Lambda_M^H\Lambda_{C_Q}^{-1}F^H||_2,\\ &= \sigma_q^{-2}\sqrt{\lambda_{max}\left((F\Lambda_M^H\Lambda_{C_Q}^{-1}F^H)^H(F\Lambda_M^H\Lambda_{C_Q}^{-1}F^H)\right)},\\ &= \sigma_q^{-2}\sqrt{|\lambda_{max}(C_Q^{-1})|^2|\lambda_{max}(M)|^2},\\ &= \sigma_q^{-2}|\lambda_{max}(C_Q^{-1})||\lambda_{max}(M)|, \qquad (6.25)\end{aligned}$$

where Λ_M denotes the diagonal matrix containing the eigenvalues of M. We observe that the blocks $\mathbf{S}_{x(1,1)}$ and $\mathbf{S}_{x(n+1,n+1)}$ will yield the same term on the right-hand side of the block Geršgorin theorem. The blocks $\mathbf{S}_{x(i,i)}$ for $i = 2, ..., n$ will yield a term that is *exactly* twice as large.

The eigenvalue $\lambda_{max}(\mathbf{S}_x)$ is bounded above by the edge of the Geršgorin circle furthest from the origin on the positive real line. So the quantity we are interested in for the upper bound is

$$\begin{aligned}\lambda_{max}(\mathbf{S}_x) &\leq max\left\{||(\mathbf{S}_{x(i,i)} - \lambda I)^{-1}||_2^{-1} + \sum_{\substack{i\neq j\\ j=1}}^{n}||\mathbf{S}_{x(i,j)}||_2\right\}, \qquad (6.26)\\ &\leq max\left\{||(\mathbf{S}_{x(i,i)} - \lambda I)^{-1}||_2^{-1}\right\} + max\left\{\sum_{\substack{i\neq j\\ j=1}}^{n}||\mathbf{S}_{x(i,j)}||_2\right\}. \qquad (6.27)\end{aligned}$$

The Geršgorin circle furthest from the origin on the positive real line will be the largest eigenvalue of the main diagonal blocks of $\mathbf{S}_x$, denoted $\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right)$, plus its radius. Therefore,

$$\begin{aligned}\lambda_{max}(\mathbf{S}_x) &\leq \lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) + 2|| - M^TQ^{-1}||_2, \qquad (6.28)\\ &\leq \lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) + 2\sigma_q^{-2}|\lambda_{max}(C_Q^{-1})||\lambda_{max}(M)|. \qquad (6.29)\end{aligned}$$

where

$$\begin{aligned}\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) = max\big\{&\lambda_{max}(B^{-1} + M^TQ^{-1}M + H^TR^{-1}H),\\ &\lambda_{max}(Q^{-1} + M^TQ^{-1}M + H^TR^{-1}H),\\ &\lambda_{max}(Q^{-1} + H^TR^{-1}H)\big\}. \qquad (6.30)\end{aligned}$$

We now have a bound on the largest eigenvalue of $\mathbf{S}_x$. We combine this with the bound on $\lambda_{min}(\mathbf{S}_x)$ in (6.3) to obtain

$$\kappa(\mathbf{S}_x) \leq \frac{\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) + 2\sigma_q^{-2}|\lambda_{max}(C_Q^{-1})||\lambda_{max}(M)|}{\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}, \tag{6.31}$$

which establishes the upper bound.

For the lower bound we apply the Rayleigh quotient to $\mathbf{S}_x$. We choose a vector $\tilde{\mathbf{y}} \in \mathbb{R}^{N(n+1)}$ such that

$$\tilde{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{n+1} \end{pmatrix}. \tag{6.32}$$

The constituent vectors $\mathbf{y}_i \in \mathbb{R}^N$ are chosen such that $\mathbf{y}_1$ is the *orthonormal* eigenvector corresponding to $\lambda_{max}(B^{-1})$ and $\mathbf{y}_i$ for $i = 2, ..., n+1$ is the orthonormal eigenvector corresponding to $\lambda_{max}(Q^{-1})$. So we have

$$\mathcal{R}_{\mathbf{S}_x}(\tilde{\mathbf{y}}) = \frac{\tilde{\mathbf{y}}^H\left(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^H\tilde{\mathbf{y}}}, \tag{6.33}$$

yielding

$$\tilde{\mathbf{y}}^H\left(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\tilde{\mathbf{y}} = \tag{6.34}$$

$$\begin{array}{rll} & \left.\begin{array}{l} \mathbf{y}_1^H\left(B^{-1} + M^TQ^{-1}M + H^TR^{-1}H\right)\mathbf{y}_1 \\ + \ \sum_{i=2}^{n}\mathbf{y}_i^H\left(Q^{-1} + M^TQ^{-1}M + H^TR^{-1}H\right)\mathbf{y}_i \\ + \ \mathbf{y}_{n+1}^H\left(Q^{-1} + H^TR^{-1}H\right)\mathbf{y}_{n+1} \end{array}\right\} & f_1 \\ + & \sum_{i=1}^{n}\mathbf{y}_{i+1}^H(-Q^{-1}M)\mathbf{y}_i \quad \} & f_2 \\ + & \sum_{i=1}^{n}\mathbf{y}_i^H(-M^TQ^{-1})\mathbf{y}_{i+1} \quad \} & f_3\ , \end{array} \tag{6.35}$$

which we have segmented for clarity of calculation. The terms forming f_1 come from the main block diagonal, whereas f_2 and f_3 come from the sub-diagonal and super-diagonal terms of $\mathbf{S}_x$, respectively.

Taking each term in f_1, we have

$$\begin{aligned} &\mathbf{y}_1^H\left(B^{-1} + M^TQ^{-1}M + H^TR^{-1}H\right)\mathbf{y}_1 \\ &\qquad = \lambda_{max}(B^{-1}) + \lambda_a(M^TQ^{-1}M) + \sigma_o^{-2}\frac{q}{N}, \end{aligned} \tag{6.36}$$

and

$$\sum_{i=2}^{n} \mathbf{y}_i^H \left(Q^{-1} + M^T Q^{-1} M + H^T R^{-1} H\right) \mathbf{y}_i$$
$$= (n-1)\left(\lambda_{max}(Q^{-1}) + \lambda_b(M^T Q^{-1} M) + \sigma_o^{-2}\frac{q}{N}\right), \tag{6.37}$$

and finally

$$\mathbf{y}_{n+1}^H \left(Q^{-1} + H^T R^{-1} H\right) \mathbf{y}_{n+1} = \lambda_{max}(Q^{-1}) + \sigma_o^{-2}\frac{q}{N}, \tag{6.38}$$

where $\lambda_a, \lambda_b \in \mathbb{R}$ are some arbitrary eigenvalues of $M^T Q^{-1} M$. Therefore,

$$f_1 = \lambda_{max}(B^{-1}) + \lambda_a(M^T Q^{-1} M) + \lambda_{max}(Q^{-1}) + 2\sigma_o^{-2}\frac{q}{N}$$
$$+ (n-1)\left(\lambda_{max}(Q^{-1}) + \lambda_b(M^T Q^{-1} M) + \sigma_o^{-2}\frac{q}{N}\right). \tag{6.39}$$

We now compute f_2. Notice that due to our choice of $\tilde{\mathbf{y}}$, the first constituent of $\mathbf{y}$, namely $\mathbf{y}_1$ is the only vector that is different to the other $\mathbf{y}_i$ for $i = 2, ..., n+1$, so the first term in the sum f_2 is

$$\mathbf{y}_2^H(-Q^{-1}M)\mathbf{y}_1 = 0, \tag{6.40}$$

since we chose the vectors in $\mathbf{y}$ to be orthonormal. The remaining constituent vectors of $\mathbf{y}$ are all identical, and will therefore yield non-zero terms,

$$f_2 = \sum_{i=2}^{n} \mathbf{y}_{i+1}^H(-Q^{-1}M)\mathbf{y}_i = -\sum_{i=2}^{n} \left(\bar{\lambda}_{max}(Q^{-1})\lambda_c(M)\right) \mathbf{y}_{i+1}^H \mathbf{y}_i,$$
$$= -(n-1)(\lambda_{max}(Q^{-1})\lambda_c(M)), \tag{6.41}$$

where $\lambda_c(M) \in \mathbb{C}$ is some arbitrary eigenvalue of M and $\bar{\lambda}_{max}(Q^{-1}) = \lambda_{max}(Q^{-1})$ since Q is a symmetric positive-definite matrix. Similarly for f_3, we have

$$f_3 = \sum_{i=2}^{n} \mathbf{y}_{i+1}^H(-M^T Q^{-1})\mathbf{y}_i = -(n-1)(\lambda_{max}(Q^{-1})\bar{\lambda}_c(M)), \tag{6.42}$$

which when combined with f_2 gives us

$$f_2 + f_3 = -2(n-1)\lambda_{max}(Q^{-1})Re(\lambda_c(M)), \tag{6.43}$$

where $Re(\lambda_c(M))$ denotes the real part of $\lambda_c(M) \in \mathbb{C}$. Combining f_1, f_2 and f_3, we have the following expression for the Rayleigh quotient (6.33),

$$\mathcal{R}_{\mathbf{S}_x}(\tilde{\mathbf{y}}) = \frac{1}{n+1}\left[(n-1)\left(\lambda_{max}(Q^{-1})(1 - 2Re(\lambda_c(M))) + \lambda_b(M^T Q^{-1} M) + \sigma_o^{-2}\frac{q}{N}\right)\right.$$
$$\left. + \lambda_{max}(Q^{-1}) + \lambda_{max}(B^{-1}) + \lambda_a(M^T Q^{-1} M) + 2\sigma_o^{-2}\frac{q}{N}\right]. \tag{6.44}$$

To obtain a bound on $\lambda_{max}(\mathbf{S}_x)$ we recall the bounds of the Rayleigh quotient from Theorem 3.4.7,

$$\begin{aligned}\lambda_{max}(\mathbf{S}_x) \geq & \tfrac{1}{n+1}\left[(n-1)\left(\lambda_{max}(Q^{-1})(1-2Re(\lambda_c(M)))+\lambda_b(M^TQ^{-1}M)+\sigma_o^{-2}\tfrac{q}{N}\right)\right. \\ & \left. +\lambda_{max}(Q^{-1})+\lambda_{max}(B^{-1})+\lambda_a(M^TQ^{-1}M)+2\sigma_o^{-2}\frac{q}{N}\right], \\ \geq & \frac{1}{\sigma_o^2}\frac{q}{N}+\frac{1}{\sigma_q^2}\frac{1}{n+1}\left[n\Big(\lambda_{max}(C_Q^{-1})+\lambda_{min}(M^TC_Q^{-1}M)-2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))\Big)\right. \\ & \left. +2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{max}(C_B^{-1})\right]. \end{aligned} \tag{6.45}$$

We also do a similar calculation for $\lambda_{min}(\mathbf{S}_x)$ by choosing $\tilde{\mathbf{y}}$ in a similar fashion to (6.32). So $\mathbf{y}_i \in \mathbb{R}^N$ for each $i = 1, ..., n+1$ is chosen such that $\mathbf{y}_1$ is the *orthonormal* eigenvector corresponding to $\lambda_{min}(B^{-1})$ and $\mathbf{y}_i$ for $i = 2, ..., n+1$ is the orthonormal eigenvector corresponding to $\lambda_{min}(Q^{-1})$. This gives us

$$\begin{aligned}\lambda_{min}(\mathbf{S}_x) \leq & \frac{1}{\sigma_o^2}\frac{q}{N}+\frac{1}{\sigma_q^2}\frac{1}{n+1}\left[n\Big(\lambda_{min}(C_Q^{-1})+\lambda_{max}(M^TC_Q^{-1}M)-2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))\Big)\right. \\ & \left. +2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{min}(C_B^{-1})\right]. \end{aligned} \tag{6.46}$$

Combining the bounds on the lowest and largest eigenvalues of $\mathbf{S}_x$, we divide (6.45) by (6.46) to obtain the lower bound on the spectral condition number of $\mathbf{S}_x$

$$\kappa(\mathbf{S}_x) \geq \frac{\frac{\sigma_q^2}{\sigma_o^2}\frac{q(n+1)}{N}+n\Big(\lambda_{max}(C_Q^{-1})+\lambda_{min}(M^TC_Q^{-1}M)-2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))\Big)+2\lambda_{max}(C_Q^{-1})Re(\lambda_{min}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{max}(C_B^{-1})}{\frac{\sigma_q^2}{\sigma_o^2}\frac{q(n+1)}{N}+n\Big(\lambda_{min}(C_Q^{-1})+\lambda_{max}(M^TC_Q^{-1}M)-2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))\Big)+2\lambda_{min}(C_Q^{-1})Re(\lambda_{max}(M))+\frac{\sigma_q^2}{\sigma_b^2}\lambda_{min}(C_B^{-1})}, \tag{6.47}$$

which completes the proof. ■

The bounds obtained here are quite complex and require analysis before any definitive conclusions can be drawn about the nature of the sensitivities of the condition number of $\mathbf{S}_x$. We now analyse the $\mathbf{S}_x$ matrix and condition number bounds further and discuss interpretations of the bounds.

6.2 Discussion

We begin by highlighting some simple points by inspecting $\mathbf{S}_x$ under simplified assumptions. We make simplistic assumptions in addition to the assumptions made in Theorem 6.1.2: $M = I_N$, $B = \sigma_b^2 I_N$, $Q = \sigma_q^2 I_N$, $R = \sigma_o^2 I_N$ and $HH^T = I_q$, thus

$$\mathbf{S}_x = \frac{1}{\sigma_q^2} \begin{pmatrix} ((\frac{\sigma_q}{\sigma_b})^2+1)I & -I & & & \\ -I & 2I & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I & -I \\ & & & -I & I \end{pmatrix} + \frac{1}{\sigma_o^2} \begin{pmatrix} H^T H & & & & \\ & H^T H & & & \\ & & \ddots & & \\ & & & H^T H & \\ & & & & H^T H \end{pmatrix}. \qquad (6.48)$$

Examining (6.48) we can clearly see the parameters governing both the first and second term of the Hessian. The first term depends on the ratio of σ_b/σ_q, arising from $\mathbf{D}$. This is different from $\mathbf{S}_p$ since the first term of $\mathbf{S}_p$ is $\mathbf{D}^{-1}$ and the bounds in the previous chapter emphasise the dependence of the condition number of $\mathbf{S}_p$ on the condition number of $\mathbf{D}$. It is likely that $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ will be more ill-conditioned than $\mathbf{D}^{-1}$, hence the condition number of $\mathbf{S}_x$ may be more vulnerable to the condition number of $\mathbf{D}^{-1}$ than $\mathbf{S}_p$.

The constituents of the second term of $\mathbf{S}_x$ depend on:

1. the number of spatial observations per assimilation step;
2. the observation error variance σ_o^2.

Increasing the observation density means that $\sigma_o^{-2}\mathbf{H}^T\mathbf{H} \rightarrow \sigma_o^{-2}\mathbf{I}$, whereas decreasing it will increase the number of zero rows in $\sigma_o^{-2}\mathbf{H}^T\mathbf{H}$. We notice that decreasing observation accuracy (increasing σ_o^2) decreases the contribution of $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, which increases emphasis on $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$. We can show this more clearly using (6.8) from Theorem 6.1.1 with the simplistic assumptions we have made here:

$$\frac{\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}{1 + \frac{1}{\sigma_o^2 \lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}} \leq \kappa(\mathbf{S}_x) \leq \kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}) \left(1 + \frac{1}{\sigma_o^2 \lambda_{max}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})}\right). \qquad (6.49)$$

We can see that as $\sigma_o^2 \rightarrow \infty$, both bounds tend to $\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$. This shows that as we decrease the observation accuracy (increase σ_o^2), the condition number of $\mathbf{S}_x$ depends on $\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$, thus the presence of the second term of $\mathbf{S}_x$ is diminished. We also see that as $\sigma_o^2 \rightarrow 0$ the lower bound tends to zero, and the upper bound diverges, yielding no definitive information.

The lower bound in Theorem 6.1.2 shows that σ_o^2 is tied to the ratios σ_q/σ_o and σ_b/σ_o. We also see that changes in $\frac{\sigma_q^2}{\sigma_o^2}$ will not affect the overall size of the lower bound, since it is present in both the numerator and denominator of the lower bound with identical coefficients. Whereas if $\frac{\sigma_q^2}{\sigma_b^2}$ changes then the bound could increase if C_B is ill-conditioned, which is highly likely in an operational NWP context.

We now turn our attention to the upper bound of Theorem 6.1.2, where we have used a novel approach in an attempt to uncover the condition number sensitivities of $\mathbf{S}_x$. We see three separate things here:

1. the model error variance σ_q^2;
2. the largest eigenvalue of the main diagonal blocks of $\mathbf{S}_x$;
3. the denominator of the upper bound, the minimum eigenvalue of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$.

We see that as $\sigma_q \rightarrow \infty$ the upper bound will increase since $\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right)$ will increase. We also see that as $\sigma_q \rightarrow 0$, the upper bound will increase because of the term $2\sigma_q^{-2}\lambda_{max}(C_Q^{-1})\lambda_{max}(M)$. Therefore the upper bound shows that the condition number of $\mathbf{S}_x$ will increase as $\sigma_q \rightarrow 0, \infty$.

The largest eigenvalues of the main diagonal blocks depend on

$$\begin{aligned}\lambda_{max}\left(\mathbf{S}_{x(i,i)}\right) = max\big\{&\lambda_{max}(B^{-1} + M^T Q^{-1} M + H^T R^{-1} H),\\ &\lambda_{max}(Q^{-1} + M^T Q^{-1} M + H^T R^{-1} H),\\ &\lambda_{max}(Q^{-1} + H^T R^{-1} H)\big\}. \qquad (6.50)\end{aligned}$$

The parameters that will cause this term to increase in size are: σ_b, σ_q, which we have discussed, and σ_o. The largest eigenvalues of C_B and C_Q will also contribute, but we focus on the error variances in this discussion. As the background error variance $\sigma_b \rightarrow 0, \infty$, the ratio σ_b/σ_q will grow, thus causing the term (6.50) to increase. As the observation error variance σ_o decreases, it will cause (6.50) to increase but because σ_o is linked to the second term of the $\mathbf{S}_x$ matrix, we cannot determine anything definitive.

We now examine the eigenvalue spectrum of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ to understand the impact of $\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ on the upper-bound. Note that

$$\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} = \begin{pmatrix} B_0+M_{0,1}^TQ_1M_{1,0} & -M_{0,1}^TQ_1 & & & \\ -Q_1M_{1,0} & Q_1+M_{1,2}^TQ_2M_{2,1} & -M_{1,0}^TQ_1 & & \\ & -Q_2M_{2,1} & Q_2+M_{2,3}^TQ_3M_{3,2} & \ddots & \\ & & \ddots & \ddots & -M_{n-1,n}^TQ_n \\ & & & -Q_nM_{n,n-1} & Q_n \end{pmatrix}. \tag{6.51}$$

In addition to the assumptions at the beginning of this section, we assume $\sigma_o = 1$ and let $\sigma_b > \sigma_q$ since it is intuitive that the variance of the errors in the previous forecast will be larger than the variance of the model errors in a single time step. Therefore,

$$\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} = \sigma_q^{-2} \begin{pmatrix} ((\frac{\sigma_q}{\sigma_b})^2+1)I & -I & & & \\ -I & 2I & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I & -I \\ & & & -I & I \end{pmatrix}, \tag{6.52}$$

which is similar to the *discretised* 2^{nd} *derivative* matrix, which arises quite often in finite difference schemes solving the heat equation (See [40], page 50). If we further assume $\sigma_q = \sigma_b = 1$ and assume a one variable linear model $N = 1$, then matrix (6.52) becomes

$$P = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}. \tag{6.53}$$

We now analyse the matrix P to extract information about its condition number sensitivities in this simplified scenario. We now require a result on the eigenvalues of P to get an understanding of the sensitivities of the ratio of its extreme eigenvalues.

Theorem 6.2.1 *The eigenvalues of* $P \in \mathbb{R}^{n+1\times n+1}$ *are*

$$\lambda_k(P) = 4\sin^2\left(\pi\frac{k-\frac{1}{2}}{2n+1}\right), \tag{6.54}$$

for $k = 1, ..., n+1$.

Proof: We solve the eigenvalue equation

$$P\mathbf{v} = \lambda\mathbf{v}, \tag{6.55}$$

where $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector such that

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n+1} \end{pmatrix}, \tag{6.56}$$

with corresponding eigenvalue λ. We can rewrite the eigenvalue equation as a recurrence relation

$$-v_{k-1} + 2v_k - v_{k+1} = \lambda v_k, \tag{6.57}$$

where

$$v_0 = 0, \tag{6.58}$$

$$v_{n+2} = v_{n+1}. \tag{6.59}$$

We introduce the appropriate auxiliary equation

$$x^2 - (2 - \lambda)x + 1 = 0, \tag{6.60}$$

which has 2 *distinct* roots

$$x_{1,2} = \frac{(2 - \lambda) \pm \sqrt{\lambda^2 - 4\lambda}}{2}. \tag{6.61}$$

Since the roots x_1 and x_2 are distinct we can write the auxiliary equation such that

$$(x - x_1)(x - x_2) = x^2 - (x_1 + x_2) + x_1 x_2, \tag{6.62}$$

which implies

$$x_2 = \frac{1}{x_1}, \tag{6.63}$$

$$x_1 + x_2 = 2 - \lambda. \tag{6.64}$$

The solution to the original recurrence relation (6.57) is a linear combination of the distinct roots,

$$v_k = Ax_1^k + Bx_2^k, \tag{6.65}$$

for some constants A, B yet to be determined. Boundary condition (6.58) dictates that $A = -B$. Therefore we have

$$v_k = A(x_1^k - x_2^k). \tag{6.66}$$

Using boundary condition (6.59) on (6.66) yields

$$(x_1^{n+1} - x_2^{n+1})A = (x_1^{n+2} - x_2^{n+2})A, \tag{6.67}$$

which, with some manipulation becomes,

$$(1 - x_1)x_1^{n+1} = x_2^{n+1}(1 - x_2), \tag{6.68}$$

we then substitute (6.63) into the right hand side of (6.68) obtaining,

$$(1 - x_1)x_1^{n+1} = x_2^{n+1}(1 - \frac{1}{x_1}). \tag{6.69}$$

Since $1 - \frac{1}{x_1} = \frac{x_1 - 1}{x_1}$, we now have

$$x_1^{n+1} = -\frac{x_2^{n+1}}{x_1}, \tag{6.70}$$

which, upon using 6.63 again, yields the following solutions

$$x_1^{2n+2} = -1, \tag{6.71}$$

$$x_1 = 1. \tag{6.72}$$

We now solve for the non-trivial root (6.71),

$$e^{i\theta(2n+2)} = e^{i\pi(2k-1)}, \tag{6.73}$$

which implies,

$$\theta = \pi\frac{2k-1}{2n+2}, \tag{6.74}$$

for $k = 1, ..., n+1$. Using (6.64) we deduce

$$2 - \lambda_k = x_1 + x_2 = x_1 + \bar{x}_1 = 2Re(x_1) \tag{6.75}$$

and since $x_1 = e^{i\theta}$, and $e^{ix} = \cos(x) + i\sin(x)$, we obtain

$$\lambda_k = 2 - 2\cos\left(\pi\frac{2k-1}{2n+2}\right). \tag{6.76}$$

Therefore

$$\lambda_k(P) = 4\sin^2\left(\pi\frac{k-\frac{1}{2}}{2n+2}\right), \tag{6.77}$$

for $k = 1, ..., n+1$, as required. ■

Since the squared sine function is bounded between 0 and 1, the eigenvalues $\lambda_k(P)$ are bounded between 0 and 4 as the assimilation window length, n, grows. The extreme eigenvalues tend to their limits (0 and 4) at a rate of $4/n^2$. The possibility of a 0 eigenvalue as the assimilation window grows implies that $\kappa(P) \to \infty$ as n grows. The analysis in this simplified scenario shows that a major source of ill-conditioning of $\mathbf{S}_x$ can arise from the smallest eigenvalue of the $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ term as the assimilation window length, n, grows.

We now make the link between the sensitivity of $\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ and the previous analysis in (6.49). If the number of observations were to equal the number of states, the dependence of the condition number of $\mathbf{S}_x$ on $\kappa(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ term will no longer be an issue. This is because the second term of $\mathbf{S}_x$, $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, will be full rank and the condition number of $\mathbf{S}_x$ will not be vulnerable to the minimum eigenvalue of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$, since the lowest eigenvalue of $\mathbf{S}_x$ will be bounded by σ_o^{-2}. This also implies that if there were a full set of observations, long assimilation windows will not affect the conditioning of $\mathbf{S}_x$, since the minimum eigenvalue of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ is no longer an issue.

We now demonstrate the bounds and verify sensitivities of the condition number of $\mathbf{S}_x$ discussed here.

6.3 Numerical Results

The aim of this section is to numerically demonstrate the sensitivities of the condition number of $\mathbf{S}_x$. We organise this section as follows.

In the first part of this section we demonstrate the uses of the Geršgorin circle theorem both in scalar and block forms for estimating the condition number of $\mathbf{S}_x$, since this was used to obtain the upper bound in Theorem 6.1.2.

The second part is solely dedicated to the demonstration of the bounds on the condition number of $\mathbf{S}_x$, and the sensitivities obtained from the theoretical analysis in the previous sections. We demonstrate the following sensitivities of the condition number of $\mathbf{S}_x$, which were obtained from the theory in this chapter:

1. the model error variance σ_q^2;
2. correlation length-scales;
3. the length of the assimilation window with the number of spatial observations per assimilation step.

The third part further enforces the effect of these sensitivities on the problem of minimising $\mathcal{J}(\mathbf{x})$ in terms of the number of iterations required for convergence, again using the linear CG method.

6.3.1 Experimental Design

The model is the 1-dimensional advection equation discretised using the upwind scheme, yielding a matrix M as in (3.71). The spatial domain is size $N = 50$ with a spatial resolution of $\Delta x = 0.01$. We use time-intervals of $\Delta t = 0.01$ and a wave speed of $a = -0.3$, thus giving us a Courant number of $\mu = -0.3$.

The experiment settings are as follows unless otherwise stated. We choose the background error, $B_0 = \sigma_b^2 C_{SOAR}$, such that the correlation length-scale $L(C_B) = \Delta x = 0.01$ and $\sigma_b = 1$. The model error, $Q_i = \sigma_q^2 C_{LAP}$ is such that the correlation length-scale $L(C_Q) = 5\Delta x = 0.05$ and $\sigma_q = 1$. The observation error is such that $R_i = \sigma_o^2 I$, where $\sigma_o = 1$. We take observations every $\Delta q = 3$ model time-steps, $n = 60$ in total, with $q = 25$ equally spaced observed grid-points out of the $N = 50$ grid-points per assimilation step.

6.3.2 Experiment 1: Geršgorin's Circles

Firstly we illustrate some of the advantages of using the block Geršgorin circle theorem by applying it to $\mathbf{S}_x$.

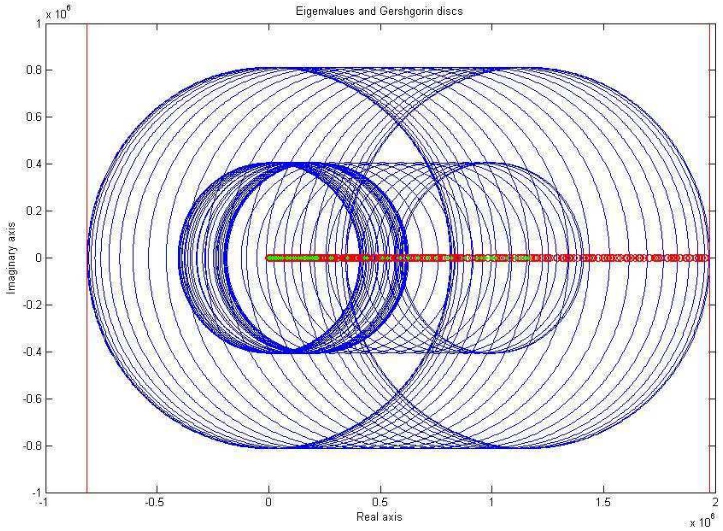


Figure 6.1: Block Geršgorin theorem applied to $\mathbf{S}_x$ where $\kappa(\mathbf{S}_x) = 3.912 \times 10^6$. Eigenvalues of $\mathbf{S}_x$ (small red circles). Eigenvalues of $\mathbf{S}_{x(i,i)}$ (green dots). Geršgorin discs (large blue circles) and estimated upper and lower bounds of the block Geršgorin Theorem (red vertical lines).

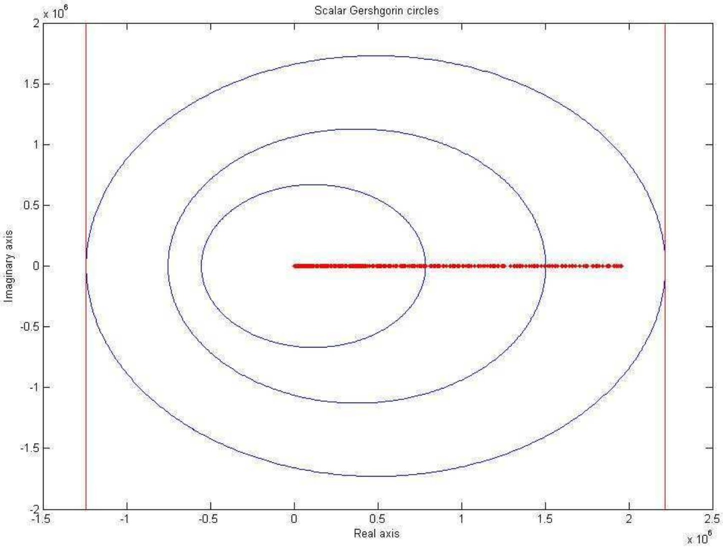


Figure 6.2: Scalar Geršgorin theorem applied to $\mathbf{S}_x$ where $\kappa(\mathbf{S}_x) = 3.912 \times 10^6$. Eigenvalues of $\mathbf{S}_x$ (small red circles). Geršgorin discs (large blue circles) and estimated upper and lower bounds of the scalar Geršgorin Theorem (red vertical lines).

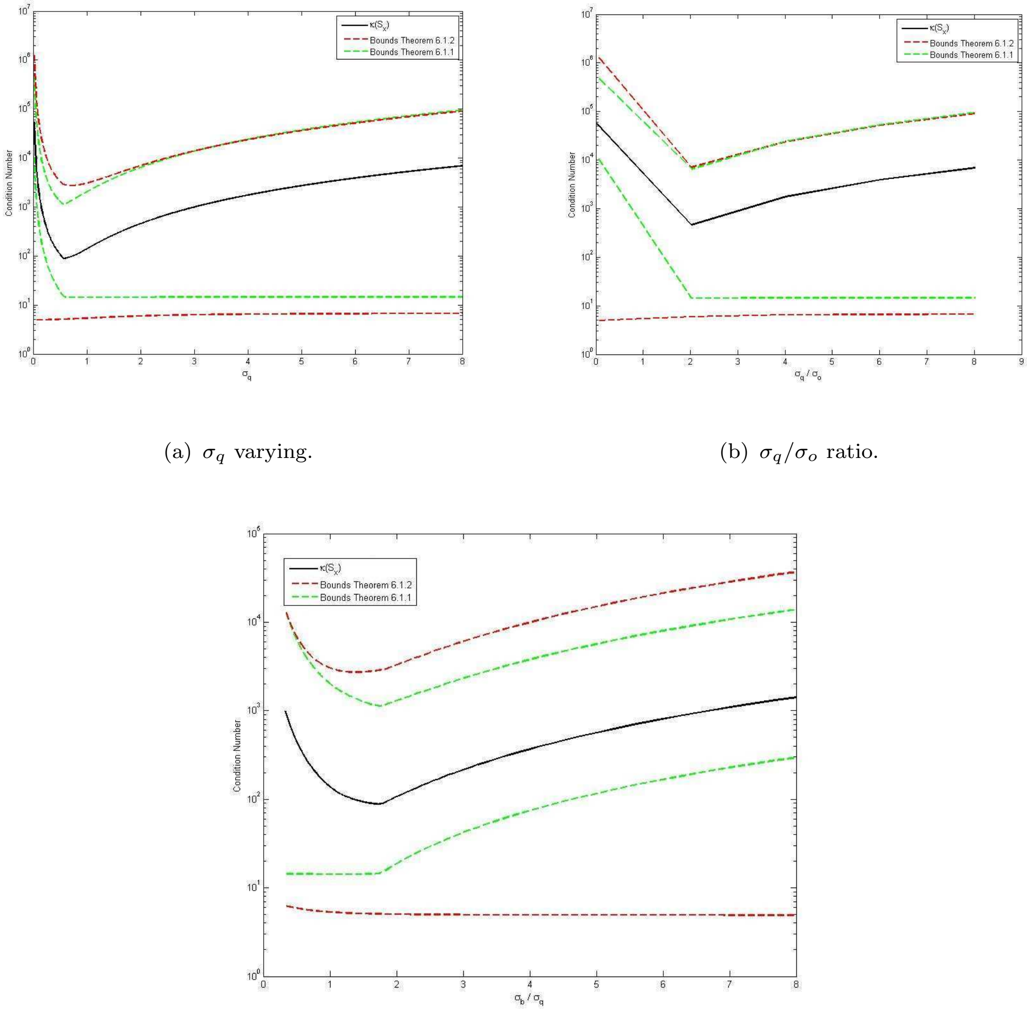
We note that we could not utilise either of the Geršgorin circle theorems for the lower bound, since $\mathbf{S}_x$ is positive definite, and the lower bounds shown in Figures 6.1 and 6.2 are negative. The condition number is relatively high due to the high correlation length-scale for the model error covariance matrix. This does not hinder the Geršgorin theorem from estimating the whereabouts of the eigenvalues. We can see from Figures 6.1 and 6.2 that the block Geršgorin circle theorem is at least as good as the Geršgorin circle theorem and that it gives a far better indication as to the whereabouts of the eigenvalues of $\mathbf{S}_x$ in this particular case. The same is also observed in [23], where the authors showed the block analogue of Geršgorin's theorem to be at least as good as the scalar Geršgorin circle theorem in general.

We also observe $\lambda_{max}(\mathbf{S}_x) = 1.956 \times 10^6$ and that the upper bound estimated by the block Geršgorin circle theorem is 1.976×10^6 compared with the upper bound scalar Geršgorin estimate of 2.218×10^6. We conclude that both bounds are good and the block Geršgorin circle theorem provides a tighter upper bound in this particular situation.

We now demonstrate the effects of the model error variance σ_q^2 on the condition number of $\mathbf{S}_x$

6.3.3 Experiment 2: Model Error Variance

The experiment parameters remain as stated in Section 6.3.1 with the exception of the following. The model error covariance matrix correlation length-scale is reduced to $L(C_Q) = \Delta x = 0.01$ and the observation standard deviation $\sigma_o = 0.5$. We also reduce the number of equally spaced spatial grid-points observed to 10 out of the $N = 50$ grid-points per assimilation step. These settings are arbitrarily chosen to ensure that the only source of ill-conditioning will be from σ_q.



(a) σ_q varying.

(b) σ_q/σ_o ratio.

(c) σ_b/σ_q ratio.

Figure 6.3: Log-scale graphs of $\kappa(\mathbf{S}_x)$ (black line) with bounds from Theorem 6.1.1 (green dotted lines) and Theorem 6.1.2 (red dotted lines) as a function of σ_q (a), σ_q/σ_o (b) and σ_b/σ_q (c).

As the parameter σ_q varies, so do the ratios σ_b/σ_q and σ_q/σ_o, prompting us to study the behaviour of the condition number of $\mathbf{S}_x$ with respect to these ratios as well as σ_q. Figure 6.3 demonstrates that the upper bounds of both Theorems 6.1.1 and 6.1.2 resemble the behaviour of the condition number of $\mathbf{S}_x$, whereas the lower bounds are uninformative. We also see that the quality of the bounds from Theorem 6.1.1 to Theorem 6.1.2 have deteriorated in accuracy. We were able to infer slightly more information from the bounds in Theorem 6.1.2 in comparison to Theorem 6.1.1, at the cost of being a worse estimator for the condition number of

$\mathbf{S}_x$. We see a minimum condition number value for $\sigma_b/\sigma_q = \sigma_q/\sigma_o = 2$ or $\sigma_q = 0.5$, but the condition number of $\mathbf{S}_x$ continues to rise as $\sigma_q \to 0, \infty$. This confirms the sensitivity of the condition number of $\mathbf{S}_x$ to the model error variance, which we obtained from the bounds in Theorem 6.1.1.

We have demonstrated the bounds and confirmed the sensitivity of $\mathbf{S}_x$ to σ_q.

6.3.4 Experiment 3: Correlation Length-Scales

In this section we discuss the effect of correlation length-scales on the condition number of $\mathbf{S}_x$. The parameters remain exactly the same as the experiment run in Section 5.1.3, Figure 5.2, to enable a comparison between the condition numbers of $\mathbf{S}_x$ and $\mathbf{S}_p$.

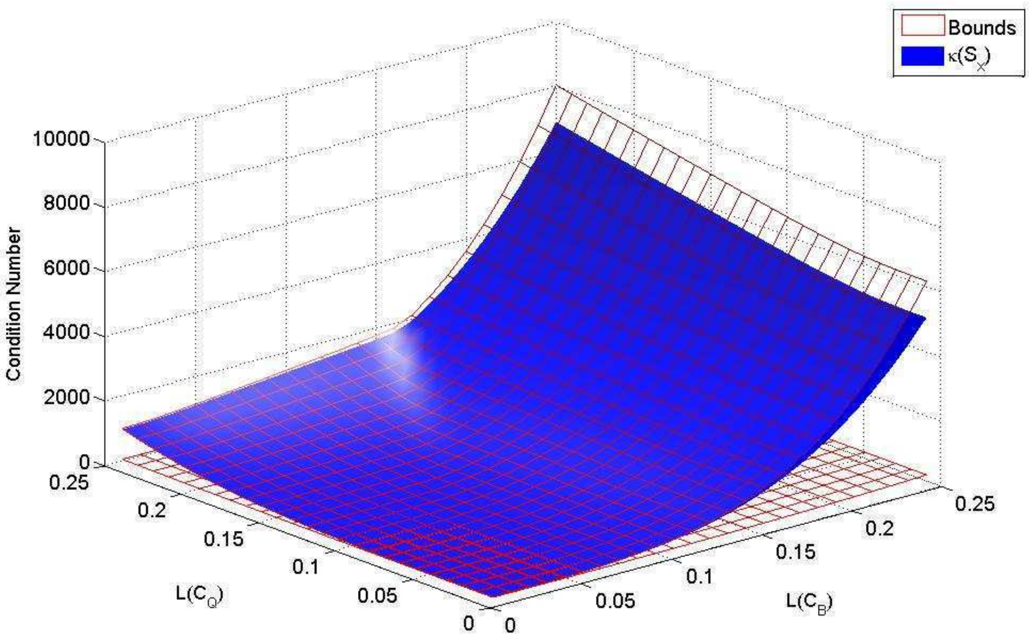


Figure 6.4: Surface plot of $\kappa(\mathbf{S}_x)$ (blue surface) and bounds (red mesh). Horizontal axes measure background error correlation length-scale $L(C_B)$ and model error correlation length-scale $L(C_Q)$. Vertical axis measures condition number.

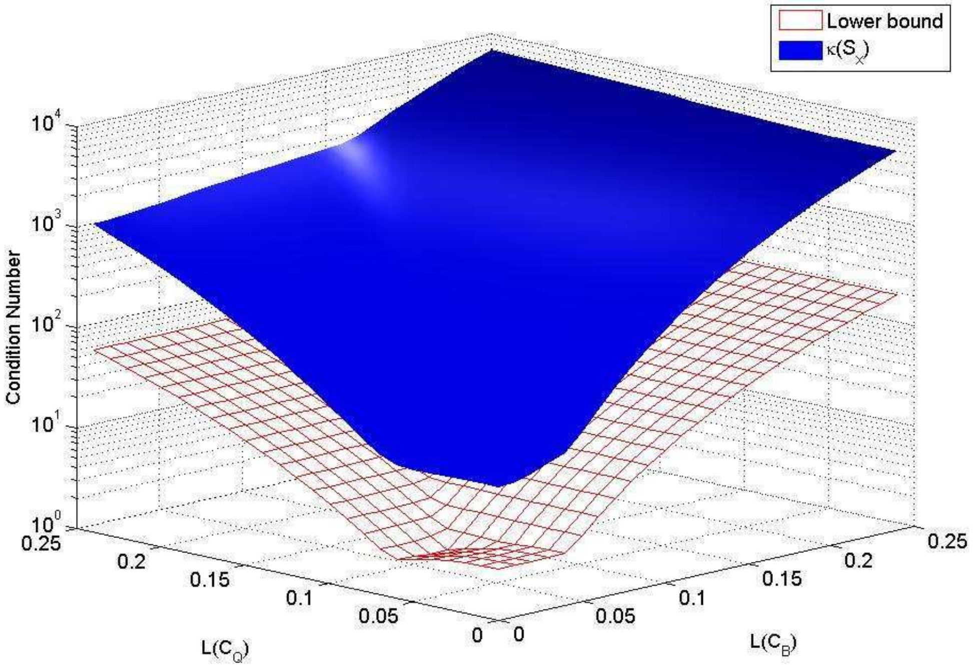


Figure 6.5: Log-scale surface plot of $\kappa(\mathbf{S}_x)$ (blue surface) and lower bound (red mesh). Horizontal axes are the background error correlation length-scale $L(C_B)$ and model error correlation length-scale $L(C_Q)$. Vertical axis measures condition number on a log scale.

Figures 6.4 and 6.5 demonstrate the sensitivity of the condition number of $\mathbf{S}_x$ to correlation length-scales in the background and model error covariance matrices. We see the upper bound is a good estimate of the condition number of $\mathbf{S}_x$ in Figure 6.4, while the lower bound is uninformative. However, Figure 6.5 shows that the behaviour of the lower bound is similar to the behaviour of the condition number of $\mathbf{S}_x$ on a log-scale.

Comparing this to the behaviour shown in the previous chapter [Section 5.1.3, Figure 5.2], the condition number of $\mathbf{S}_x$ is far more sensitive than $\kappa(\mathbf{S}_p)$ to changes in the correlation length-scales of C_B and C_Q and hence $\kappa(\mathbf{D})$, rising to a condition number range of $5000-7000$ for $L(C_B) = 0.25 = 2.5\Delta x$ compared to the maximum condition number of $\kappa(\mathbf{S}_p) = 1800$ in the same scenario. We also should keep in mind that correlation matrix C_B is a SOAR matrix which is more sensitive to correlation length-scale then C_Q which is a Laplacian. We see in Figure 6.4, when the value of $L(C_Q) = 0.25$, that $\kappa(\mathbf{S}_x)$ rises to 1400, compared to Section 5.1.3, Figure 5.2, where $\kappa(\mathbf{S}_p)$ rises to only 400 for the same value of $L(C_Q) = 0.25$.

We have demonstrated the sensitivity of the condition number of $\mathbf{S}_x$ to correlation length-scales in the background and model error covariance matrices, along with the bounds. We now investigate the sensitivity of the condition number of $\mathbf{S}_x$ to observation density and assimilation window length.

6.3.5 Experiment 4: Assimilation Window Length and Spatial Observation Density

The length of the assimilation window and the distribution of observations are important aspects of the data assimilation problem. In the previous chapter we found that the condition number of $\mathbf{S}_p$ and even the preconditioned Hessian suffered from ill-conditioning as the number of observations and assimilation window length both increased. In this experiment the parameters are identical to those used in Experiment 2 (Section 6.3.3) and Section 5.1.3, to enable a comparison between $\mathbf{S}_p$ and $\mathbf{S}_x$.

We investigate the dependence of the condition number of $\mathbf{S}_x$ on the condition number of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$, which is ill-conditioned due to its minimum eigenvalue. If the second term of the Hessian, $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, were to be full rank, then this would remedy the issue with the minimum eigenvalue, as we see below

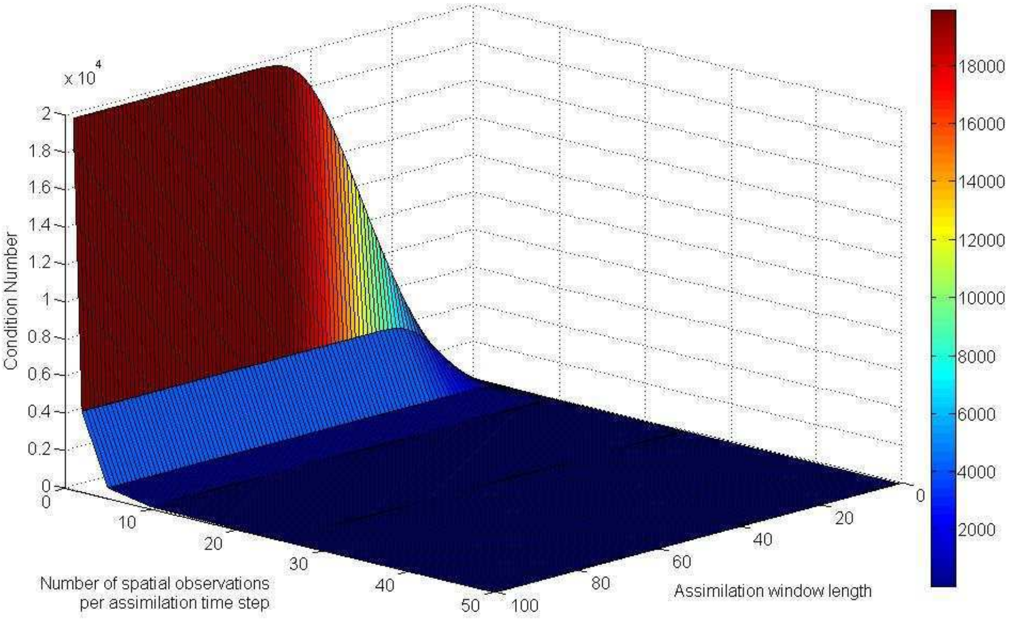


Figure 6.6: Surface plot of $\kappa(\mathbf{S}_x)$. Vertical axis measures condition number. The non-vertical axes measure spatial observation density q and assimilation window length, n.

Figure 6.6 demonstrates the sensitivity of the condition number of $\mathbf{S}_x$ to increasing assimilation window length as the number of spatial observations per assimilation step decreases below $q = N/5$. Interestingly, we see that the rise in assimilation window length has no effect on the condition number of $\mathbf{S}_x$ if there are a good number of spatial observations, more than $q = N/2$. This confirms our findings in the discussion in Section 6.2, that as the term $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ approaches full rank, the condition number of $\mathbf{S}_x$ becomes less dependent on the condition number of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$.

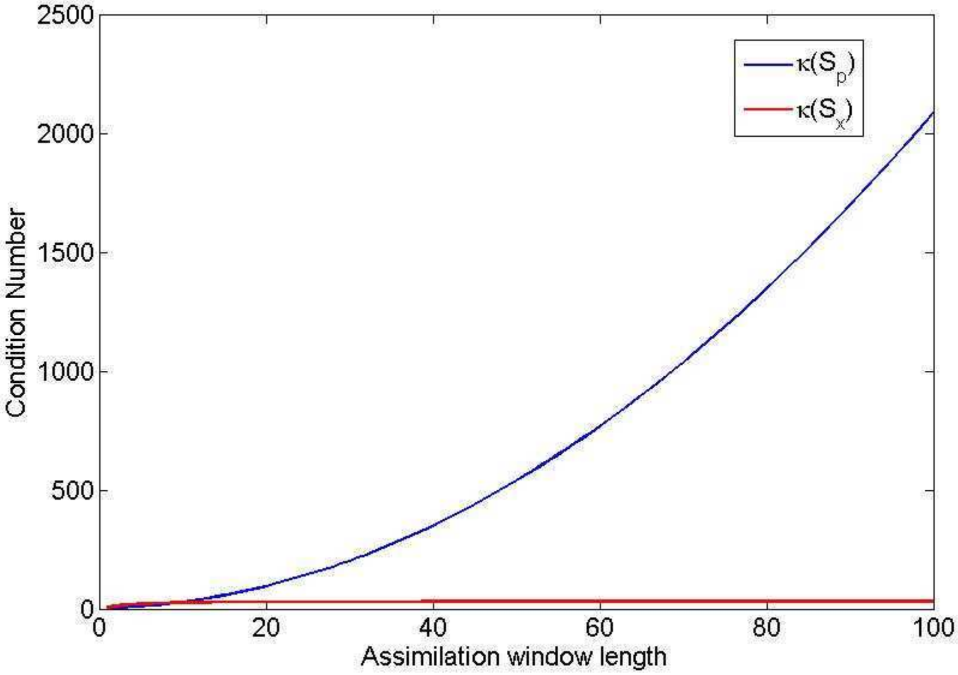


Figure 6.7: Condition numbers of $\mathbf{S}_p$ (blue line) and $\mathbf{S}_x$ (red line) as a function of assimmilation window length, n.

Figure 6.7 shows the condition numbers of $\mathbf{S}_x$ and $\mathbf{S}_p$ in the case where the domain is fully observed. Although this is unrealistic in an operational setting, it does show an inherent difference between both Hessians. This experiment shows that with a fully observed domain, the condition number of $\mathbf{S}_x$ is immune to increasing assimilation window length, whereas $\mathbf{S}_p$ is affected by increasing assimilation window length regardless of the number of observations.

This concludes our numerical demonstration of the theoretical findings from the bounds in this chapter. We now investigate the effect of the sensitivities discussed in this chapter on the rate of convergence using the linear CG method to minimise $\mathcal{J}(\mathbf{x})$ using the linear advection model.

6.4 Convergence Results

In this section we complement the findings in this chapter on the sensitivities of the condition number of $\mathbf{S}_x$. We investigate the effect of the sensitivities numerically demonstrated in the previous section on the minimisation convergence characteristics of $\mathcal{J}(\mathbf{x})$.

The data assimilation parameters are identical to those in described in Section 6.3.1, with the exception of $L(C_Q) = \Delta x = 0.01$ and the number of spatial observations is $q = 10$ out of the $N = 50$ grid-points per assimilation step. The parameters used to generate the truth are identical to the parameters used in the assimilation.

We use the linear CG method as described in Section 3.2.1 to minimise $\mathcal{J}(\mathbf{x})$, with a iterative minimisation tolerance of $\tau = 10^{-5}$, as described in Chapter 3, Section 3.2.4. The solution relative error is calculated in the same way as shown in Chapter 4, Section 4.1.6.

6.4.1 Experiment 1: Model Error Variance

The first sensitivity we investigate is the model error variance σ_q^2. Varying this parameter alters the values of the ratios σ_b/σ_q and σ_q/σ_o simultaneously. We use settings identical to those in Section 6.3.3, Experiment 2, since these settings were used arbitrarily to illustrate this sensitivity. The table below shows the effect that changes in this parameter have on the numerical condition number, solution accuracy and number of iterations to convergence.

σ_q	σ_q/σ_o	σ_b/σ_q	No. of iterations	Condition number	Solution relative error
0.11	0.11	9.09	220	4824	0.29
0.21	0.21	4.76	139	1351	0.30
0.31	0.31	3.23	115	641	0.29
0.41	0.41	2.44	98	385	0.28
1.81	1.81	0.55	101	367	0.23
2.81	2.81	0.36	126	882	0.25
3.81	3.81	0.26	151	1619	0.25
5.81	5.81	0.17	183	3762	0.29
7.81	7.81	0.13	208	6796	0.28

Table 6.1: Standard deviation ratios, number of iterations to convergence and the solution relative error of $\mathcal{J}(\mathbf{x})$, and the condition number of $\mathbf{S}_x$. Standard deviations $\sigma_b = \sigma_o = 1$.

We see here that when σ_q tends to zero or increases from 2, the condition number of $\mathbf{S}_x$, the number of iterations to convergence and the solution relative error all increase. The other ratios involving σ_q are the underlying reason for the changes seen in the minimisation characteristics in Table (6.1). As the ratios move away from ~ 2, the condition number of $\mathbf{S}_x$, number of iterations and relative solution error all increase.

6.4.2 Experiment 2: Correlation Length-Scales

We now investigate to the sensitivity of the minimisation problem to correlation length-scales. We preserve the settings from the previous experiment, Section 6.4.1 and we vary the correlation length-scales of C_B and C_Q, remembering that $\kappa(C_{LAP})$ is less sensitive than $\kappa(C_{SOAR})$ to identical changes in the correlation length-scales.

$L(C_B)$	No. of iterations	Condition number
0.01	86	134
0.05	608	43,637
0.10	978	492,394
0.15	1301	2,203,292
0.20	1687	6,537,759

$L(C_Q)$	No. of iterations	Condition number
0.01	86	134
0.05	361	65,670
0.10	491	560,326
0.15	572	1,924,374
0.20	596	4,563,487

Table 6.2: Tables of convergence and condition number values with varying correlation length-scales. Table on the left $L(C_B) = \Delta x$, while $L(C_Q)$ varies. Similarly the right table $L(C_Q) = \Delta x$, while $L(C_B)$ varies.

Table 6.2 shows the effects of correlation length-scale on the minimisation problem presented by $\mathcal{J}(\mathbf{x})$. Both tables confirm the sensitivity of the condition number of $\mathbf{S}_x$ to the correlation length-scales of C_B and C_Q, also shown in Section 6.3.4 Experiment 3. We have also shown the adverse affect this has on the number of iterates.

We now examine the effect of observation density and assimilation window length.

6.4.3 Experiment 3: Assimilation Window Length and Observation Density

In this experiment we examine the sensitivity of the minimisation problem presented by $\mathcal{J}(\mathbf{x})$ to the length of the assimilation window and the observation density simultaneously. We will discuss three tables in this section; number of iterates, solution accuracy and condition numbers.

We aim to show that increasing assimilation window length renders $\mathbf{S}_x$ ill-conditioned, as discussed in Section 6.2 for low observation densities. We also show that as we increase the number of spatial observations per assimilation step the condition number of $\mathbf{S}_x$ becomes less effected by the rise in assimilation window length. This due to the second term of the Hessian $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ approaching full rank as the observation density increases.

		No. of spatial observations					
		50	25	10	5	2	1
Assimilation window length	1	19	24	32	38	48	50
	11	20	26	49	83	165	193
	21	19	26	51	93	215	271
	31	19	25	51	97	230	349
	41	18	25	50	97	230	420
	51	18	24	49	98	241	460
	61	17	24	49	98	240	429
	71	17	24	49	97	241	459
	81	17	23	49	96	239	460
	91	16	23	49	94	241	465

Table 6.3: No. of iterations

		No. of spatial observations					
		50	25	10	5	2	1
Assimilation window length	1	0.26	0.27	0.30	0.30	0.27	0.35
	11	0.09	0.09	0.12	0.22	0.51	0.58
	21	0.05	0.05	0.06	0.11	0.40	0.64
	31	0.04	0.04	0.04	0.07	0.26	0.57
	41	0.03	0.03	0.03	0.05	0.17	0.48
	51	0.02	0.02	0.02	0.03	0.12	0.37
	61	0.02	0.02	0.02	0.03	0.09	0.28
	71	0.02	0.02	0.02	0.02	0.08	0.24
	81	0.01	0.01	0.02	0.02	0.07	0.19
	91	0.01	0.01	0.01	0.02	0.05	0.16

Table 6.4: Solution relative error

Table 6.3 shows that the main contributor to the rise in the number of iterates is the lack of spatial observations per assimilation time step. In comparison to minimising the preconditioned version of $\mathscr{J}(\mathbf{p})$, Section 5.2.1 Figure 5.7, the increased number of observations and assimilation window length both *decrease* the number of iterations required for the minimisation problem of $\mathscr{J}(\mathbf{x})$ to converge. We also see a sharp rise in iterates in the cases where there are not many observations, $q = 5, 2, 1$, settling quickly at $n = 41$.

Table 6.4 shows the error in the solution increases as the number of observations decreases for all lengths of assimilation window. We also see as the length of the assimilation window increases the solution errors generally decrease, allowing the algorithm more freedom to fit the data, which is a known feature of wc4DVAR. The solution relative error falls as the assimilation window length grows when more than half of the state is observed.

		No. of spatial observations					
		50	25	10	5	2	1
Assimilation window length	1	8	20	78	97	103	103
	11	9	78	100	147	221	232
	21	9	90	133	233	608	762
	31	9	94	144	268	816	1541
	41	9	95	148	284	892	2402
	51	9	96	150	292	950	2989
	61	9	96	151	297	974	3177
	71	9	96	152	300	986	3246
	81	9	97	153	302	996	3291
	91	9	97	153	303	1000	3332

Table 6.5: Condition number values of $\mathbf{S}_x$.

Table 6.5 shows the condition number values for varying assimilation window lengths and observation densities. The condition number behaviour complements the trends shown in Tables 6.3 and 6.4, which was expected.

This concludes our analysis and investigations into the effects of the sensitivities on the minimisation characteristics of the $\mathcal{J}(\mathbf{x})$ formulation minimised using the linear CG method with the linear advection equation as the model.

We now summarise this chapter.

6.5 Summary

The aim of this chapter was to explore the sensitivities of the problem $\mathcal{J}(\mathbf{x})$ by bounding the condition number of $\mathbf{S}_x$ in a similar fashion to the exercise carried out in Chapter 5. We used the block analogue of the Geršgorin circle theorem as shown in [23] to demonstrate the theoretical result in the paper, where the block Geršgorin theorem is *at least as good as* the scalar Geršgorin circle theorem. This was shown on the Hessian matrix $\mathbf{S}_x$ through a simple example.

The bounds derived for the Hessian $\mathbf{S}_x$ were demonstrated through numerical experiments using the 1D advection equation. Through the bounds, we showed the sensitivities of the condition number of $\mathbf{S}_x$ to the following:

1. the model error variance σ_q^2;
2. correlation length-scales in the background and model error covariance matrices;
3. assimilation window length and observation density.

More specifically we showed

1. The condition number of $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}$ heavily influences the condition number of $\mathbf{S}_x$, shown in Theorem 6.1.1. We highlight this sensitivity further through the condition number of the background and model error covariance matrix, $\mathbf{D}$, which is sensitive to correlation length-scales and the σ_b/σ_q ratio. The theory suggests that $\mathbf{S}_x$ is potentially more vulnerable to the condition number of $\mathbf{D}$ than $\mathbf{S}_p$. This was shown theoretically in Section 6.2 and also demonstrated numerically in Section 6.3.4, Experiment 3.

2. The sensitivity of the condition number of $\mathbf{S}_x$ to assimilation window length. This is different to $\mathbf{S}_p$, which sees an increase in its condition number (as shown in Chapter 5) as the observation density increases *and* as the assimilation window increases.

   (a) The minimum eigenvalue of the first term of the $\mathbf{S}_x$ Hessian has the potential to converge to 0 as the assimilation window grows. The upper bound in Theorem 6.1.2 shows that as the assimilation window increases, $\lambda_{min}(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L})$ decreases and therefore increasing $\kappa(\mathbf{S}_x)$. We showed this through examination of the first term of $\mathbf{S}_x$ when reduced to the P matrix (as discussed in Section 6.2).

   (b) As the observation density decreases the condition number of $\mathbf{S}_x$ grows at a faster rate as the length of the assimilation window increases. While if we have a full rank observation term, the condition number of $\mathbf{S}_x$

becomes *immune* to increasing assimilation window length (as discussed in Section 6.2).

(c) Decreasing observation accuracy (increasing σ_o) reduces the contribution of the second term of $\mathbf{S}_x$ and puts greater emphasis on the first term of $\mathbf{S}_x$, which is sensitive to assimilation window length and the condition number of $\mathbf{D}$. This is shown through the analysis of the bounds in Theorem 6.1.1 in the discussion in Section 6.2, equation (6.49).

These sensitivities were shown through theoretical analysis of the bounds and numerical demonstrations of the theory on the condition number of $\mathbf{S}_x$. We showed further that these sensitivities also reflect in the minimisation characteristics, which we characterised by the number of iterations to converge to a required tolerance and the solution accuracy post-convergence.

This concludes the analysis of the condition number of the first-order Hessian $\mathbf{S}_x$.

Chapter 7

Weak-Constraint 4DVAR: Lorenz95 Model

In this chapter, we show an example where it is possible for the theory established in the previous chapters to provide valuable insight for applications in a wider context. We explore the application of the wc4DVAR algorithms discussed in this thesis on the non-linear chaotic model known as Lorenz 95, described in Chapter 3, Section 3.5.2. This model possesses error growth characteristics similar to that of weather prediction models. It is also one of the models used by the ECMWF in OOPS (Object-Oriented Programming System), which they use as a testing ground before operational implementation.

The theory derived in this thesis assumes linear time invariant models or models that present a circulant matrix, with periodic domain and appropriate covariance structures. The aim of this chapter is to demonstrate the potential scope of the condition number sensitivities found in Chapters 5 and 6 on the non-linear chaotic Lorenz 95 model.

7.1 Lorenz 95 Model Example

The purpose of this chapter is to put the theory in the previous chapters into wider context. We do this by testing if the parameters, which were found to be responsible for ill-conditioning in the theory on linear models, also have the same effect the solution process of wc4DVAR when applied to a non-linear model. The specific sensitivities we investigate are:

1. the observation density and assimilation window length;
2. the correlation length-scales in the background and model error covariance matrices.

The theory showed that as the observation density and assimilation window length increase, the condition number of $\mathbf{S}_p$ and hence the number of iterations for the model error formulation also increase. The theory also showed that as the number of observations *decreases* and the assimilation window length increases the condition number of $\mathbf{S}_x$ and the number of iterations of the state formulation to converge, also increase. We also found a particular special case where if the state domain was fully observed, the increase in assimilation window length *no longer affected* the condition number of $\mathbf{S}_x$ or the number of iterations required for convergence. We also saw that as the correlation length-scales grow $\mathbf{S}_p$ and $\mathbf{S}_x$ become more ill-conditioned, where $\mathbf{S}_x$ showed potential of being more sensitive to this than $\mathbf{S}_p$.

Both wc4DVAR algorithms implemented on the Lorenz 95 model have been tested and verified in the same manner as for the implementation of the wc4DVAR algorithms for the advection equation in Chapter 4. The adjoints and objective function gradients were all successfully coded and tested. We do not discuss the implementation details of the Lorenz 95 system in this chapter as it has already been done in Chapter 4. We now discuss the experimental design before discussing our experimental results.

7.1.1 Experimental Design

The model parameters used for the Lorenz 95 are explained in Chapter 3, Section 3.5.2, but we restate the parameter settings here for clarity. The variables are treated as points on a latitude circle, therefore the spacing between each of the $N = 40$ variables is $\Delta x = 1/N = 0.025$. Throughout this chapter we use a time-step of $\Delta t = 0.025$, which is equivalent to 3 hours. We use the Polak-Ribiere non-linear conjugate gradient technique as described in Chapter 3, Section 3.2.3, to minimise the objective functions. The iterative minimisation stopping criterion used is described in Chapter 3, Section 3.2.4, where we set the tolerance to $\tau = 10^{-3}$ for all experiments unless otherwise stated. The solution errors and relative errors are all calculated as in previous chapters, as shown in Chapter 4, Section 4.1.6. The model parameters chosen here remain unchanged throughout our experiments.

The assimilation parameters are as follows. The background covariance matrix is such that $B = \sigma_b^2 C_{SOAR}$ with $\sigma_b = 0.1$ and $L(C_B) = 0.005 = \Delta x/5$. The model error covariance matrix is such that $Q = \sigma_q^2 C_{LAP}$ with $\sigma_q = 0.05$ and $L(C_Q) = 0.005 = \Delta x/5$. The observation error covariance matrix is $R = \sigma_o^2 I$ with $\sigma_o = 0.05$. It is important to note that for all our experiments, the data assimilation parameters used to generate the truth are identical to the assimilation parameters.

We use the Polak-Ribiere code used is as described in Secion 3.2.3, written by C.E. Rasmussen, to minimise the objective functionals. The Polak-Ribiere code is written such that it requires the code for the procedure which evaluates $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$ and their respective gradients.

We now present our experimental results.

7.1.2 Experiment 1 (i): Assimilation Window Length and Observation Density

In this section we examine the sensitivity of the model error and state formulations to the length of the assimilation window and the observation density simultaneously. We now present the number of iterations needed for both formulations to achieve the minimisation tolerance τ.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	6	7	6	7	7	7	6	5
	6	19	34	24	18	23	17	12	11
	12	50	48	62	46	41	27	46	16
	18	84	93	99	63	66	44	34	37
	24	122	78	58	58	71	47	66	17
	30	189	90	89	50	50	123	32	21
	36	158	106	191	70	118	68	23	23
	42	213	188	80	73	68	44	42	36
	48	224	179	102	58	61	292	107	17

Table 7.1: Number of iterations to minimise $\mathcal{J}(\mathbf{p})$.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	5	12	13	11	10	11	13	13
	6	8	37	46	54	54	58	61	57
	12	8	39	71	85	111	123	135	168
	18	8	42	85	110	166	213	252	229
	24	8	66	99	121	191	245	336	352
	30	8	69	176	262	207	345	261	286
	36	8	59	175	282	235	379	258	303
	42	8	67	109	210	213	290	220	416
	48	8	57	129	185	239	237	278	540

Table 7.2: Number of iterations to minimise $\mathcal{J}(\mathbf{x})$.

$n = 48$ is equivalent to 6 days.

Table 7.1 shows the iteration counts for the model error formulation with assimilation window length and observation density. This table shows that as the number of observations and assimilation window length increase, the number of iterations for convergence also increases. The results in Table 7.1 are in line with our initial findings in Chapter 4, Section 4.2.3, Experiment 3, where we observed an increase in the number of iterations for the model error formulation as assimilation window length increased. These results also agree with theoretical evidence derived from the upper bound of Theorem 5.1.3, Section 5.1.1 for the advection equation, which was demonstrated on the condition number of $\mathbf{S}_p$ in Experiment 3, Section 5.1.3.3 and on the number of iterates of the model error formulation in Chapter 5, Section 5.3.3.

Table 7.2 shows that as the number of observations *decreases*, the number of iterations required for $\mathcal{J}(\mathbf{x})$ to converge increases, where this effect is amplified by the increasing length of the assimilation window. The results in Table 7.2 agree with findings in Chapter 6, Section 6.3.5, Experiment 4. We can also see the special case in Table 7.2, where the state is fully observed (first column, where observations $q = 40$), agreeing with Chapter 6, Section 6.3.5, Experiment 4.

Comparing the number of iterations of the two formulations, we see that the model error formulation generally performs better than the state formulation, unless the state is half ($q = 20$) or fully ($q = 40$) observed. The assimilation runs in Tables 7.1 and 7.2 show that with enough observations, the state formulation out-performs the model error formulation and has the unique property of not being affected by the assimilation window length with a fully observed state. This agrees with findings in Chapter 6.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	0.004	0.008	0.008	0.009	0.008	0.009	0.008	0.008
	6	0.005	0.009	0.011	0.009	0.011	0.011	0.012	0.012
	12	0.006	0.013	0.019	0.021	0.022	0.028	0.028	0.029
	18	0.006	0.013	0.028	0.037	0.041	0.046	0.061	0.052
	24	0.007	0.015	0.032	0.032	0.049	0.051	0.065	0.081
	30	0.007	0.022	0.033	0.040	0.099	0.090	0.082	0.106
	36	0.007	0.020	0.048	0.034	0.415	0.066	0.172	0.110
	42	0.009	0.023	0.041	0.058	0.079	0.093	0.109	0.146
	48	0.009	0.026	0.052	0.077	0.058	0.135	0.121	0.241

Table 7.3: Total solution relative error, $\mathcal{J}(\mathbf{p})$.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	0.004	0.009	0.008	0.009	0.009	0.01	0.008	0.009
	6	0.004	0.008	0.012	0.013	0.023	0.016	0.017	0.020
	12	0.006	0.025	0.035	0.041	0.054	0.046	0.046	0.075
	18	0.006	0.013	0.054	0.664	0.824	0.078	0.171	0.169
	24	0.007	0.025	0.048	0.415	0.722	0.146	0.619	0.272
	30	0.007	0.028	0.598	1.094	1.031	0.638	0.553	0.725
	36	0.007	0.022	0.828	0.772	0.616	0.699	0.817	1.235
	42	0.008	0.036	0.753	0.935	1.071	0.487	0.902	0.737
	48	0.009	0.047	0.974	0.936	0.583	0.897	1.335	1.245

Table 7.4: Total solution relative error, $\mathcal{J}(\mathbf{x})$.

Table 7.3 shows that the accuracy of the model error formulation solution increases as the number of observations increase at the cost of more iterations, which was to be expected. The solution relative errors also increase as the assimilation window length increases and requires more iterations to solve. While the increase in the number of iterations agrees with findings in Chapter 5, Section 5.2.1.2, Experiment 2, the increase in relative errors does not agree with previous findings. The reason for the increase in solution relative errors as the assimilation window increases is that the objective functions $\mathcal{J}(\mathbf{p})$ and $\mathcal{J}(\mathbf{x})$ become increasingly non-linear and thus the higher order terms in the Taylor expansion of these functions become larger. We showed in Section 3.1, equation 3.7, that the condition number acts only as an indicator of solution accuracy, *with* a second order approximation of the Taylor series of the non-linear objective functionals. Therefore the condition number alone may not be responsible for the increase in iterations and solution relative error.

In Table 7.4 we see a clear trend of increased relative errors in the solution with the increase of assimilation window length. We also see that as the observation density decreases, the relative errors in the solution increase, which is consistent with the increase in iterations shown in Table 7.2. The increase of relative errors with assimilation window length is evident for all numbers of observations except when the state is fully observed, $N = 40$. We also see many cases of divergence of the solution, where the solution relative error of the state formulation is ~ 1. This emphasises the sensitivity of the state formulation to observation density, where if there are not enough observations, the solution relative error can be $\sim \mathcal{O}(10^2)$ larger than errors in the model error formulation solution.

Tables 7.3 and 7.4 show that the accuracy of the model error formulation is clearly superior to the state formulation. The increased non-linearity of $\mathcal{J}(\mathbf{x})$ over $\mathcal{J}(\mathbf{p})$ could be the reason for the difference in solution relative errors. In Table 7.4 for $n = 48$ and $q \leq 10$, we see an example where the state formulation solution has diverged. This may be due to the increase in non-linearity of the state formulation objective function or the inadequacy of the stopping criterion.

We now examine the condition numbers.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	1.49E+05	5.44E+07	4.96E+07	4.36E+07	3.56E+07	4.45E+07	1.26E+07	2.73E+07
	6	1.07E+16	2.47E+19	4.04E+20	3.52E+20	2.24E+19	1.12E+20	4.23E+19	2.38E+15
	12	8.30E+30	1.16E+34	9.20E+33	4.41E+32	1.23E+34	2.93E+31	1.34E+32	6.30E+33
	18	4.37E+44	8.37E+47	1.10E+47	1.05E+47	2.06E+46	1.29E+48	1.10E+46	2.42E+43
	24	7.46E+55	2.54E+61	3.28E+61	1.48E+59	2.90E+57	3.45E+58	2.48E+54	1.57E+51
	30	2.16E+71	5.08E+73	6.02E+73	1.74E+71	1.10E+74	1.02E+71	1.58E+71	9.80E+63
	36	1.54E+79	1.40E+81	2.86E+82	2.76E+84	2.40E+80	1.33E+82	2.22E+77	4.27E+72
	42	6.35E+91	5.87E+91	3.42E+91	4.11E+91	6.61E+87	1.32E+89	9.99E+91	6.81E+77
	48	8.22E+101	9.31E+105	5.83E+105	1.74E+106	1.80E+101	6.29E+106	2.82E+101	1.53E+88

Table 7.5: Condition number values of $\mathbf{S}_p$. $n = 48$ is equivalent to 6 days.

The condition numbers in Table 7.5 are incredibly high, however we do see the general trend that the condition number of $\mathbf{S}_p$ increases with the length of the assimilation window for any number of observations. We also see that the condition number of $\mathbf{S}_p$ increases as the number of observations increase, which is in agreement with the iterations in Table 7.1 and the trend of solution relative errors in Table 7.3. This also agrees with our findings in Chapter 5.

		No. of spatial observations							
		40	20	10	8	5	4	2	1
Assimilation window length	1	1.00	2.83E+06	1.13E+07	1.22E+07	1.52E+07	1.43E+07	1.42E+07	1.57E+07
	6	1.00	3.36E+07	1.50E+14	4.20E+16	2.40E+17	5.23E+17	7.14E+17	4.08E+18
	12	1.00	8.04E+07	4.33E+14	2.31E+16	2.31E+21	3.25E+21	3.45E+21	1.06E+28
	18	1.00	1.28E+09	7.68E+12	2.65E+17	7.92E+20	2.14E+21	5.08E+21	2.76E+26
	24	1.00	9.44E+07	2.25E+14	2.22E+15	5.70E+20	1.46E+21	3.00E+21	5.33E+21
	30	1.00	6.27E+07	7.19E+11	5.83E+13	8.73E+20	1.52E+21	2.77E+21	1.44E+22
	36	1.00	2.86E+07	1.29E+13	9.88E+15	3.80E+21	3.65E+21	1.31E+22	1.99E+22
	42	1.00	4.90E+07	1.44E+13	2.20E+16	7.21E+21	1.28E+22	6.27E+21	4.27E+21
	48	1.00	2.68E+07	1.36E+12	2.46E+15	2.97E+22	7.96E+20	1.91E+21	7.70E+21

Table 7.6: Condition number values of $\mathbf{S}_x$. $n = 48$ is equivalent to 6 days.

The condition numbers for $\mathbf{S}_x$ in Table 7.6 show that if the state is fully observed, the condition number of $\mathbf{S}_x$ is consistently 1, which agrees with the lower number of iterations in the same column in Table 7.2 and also the low relative error in Table 7.4. We also see the as the number of observations decreases, the condition numbers of $\mathbf{S}_x$ rise very rapidly, reaching a plateau at around 5 observations.

As mentioned previously, the condition number is not the only influential factor for the poor solution accuracy of the state formulation as seen in Table 7.4, the increasing non-linearity of $\mathcal{J}(\mathbf{x})$ may also be a contributor. Evidences of increasing non-linearity of $\mathcal{J}(\mathbf{x})$ can be seen in the large number of iterations, poor solution relative errors (to the extent that it looks to have diverged in some cases) and very low condition numbers in comparison to the condition number of the Hessian of $\mathcal{J}(\mathbf{p})$. Another possibility is that the iterative minimisation stopping criterion used (as described in Section 3.2.4) is not suitable for this particular application.

We now show a further experiment to emphasise the strength of the model state formulation when the state is fully observed.

7.1.3 Experiment 1 (ii): Assimilation Window Length Special Case

In this experiment we aim to show the special case where if the state is fully observed, then the number of iterations of $\mathcal{J}(\mathbf{x})$ does not increase for a long assimilation window. The experiment settings are as in Section 7.1.1 with the exception of the following. We set the background related parameters $\sigma_b = 0.1$ and $L(C_B) = 0.025 = \Delta x$, the model error related parameters $\sigma_q = 0.05$ and $L(C_Q) = 0.01 = \Delta x/5$ and the observation error variance $\sigma_o = 0.01$. We set a long assimilation window, equivalent to 6 days $n = 48$, and the state is fully observed $q = N = 40$ at every assimilation step.

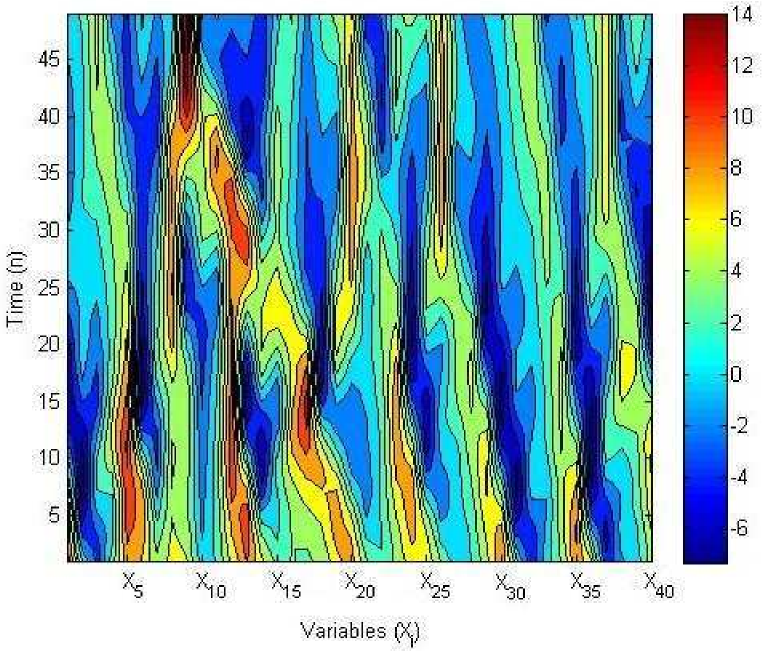
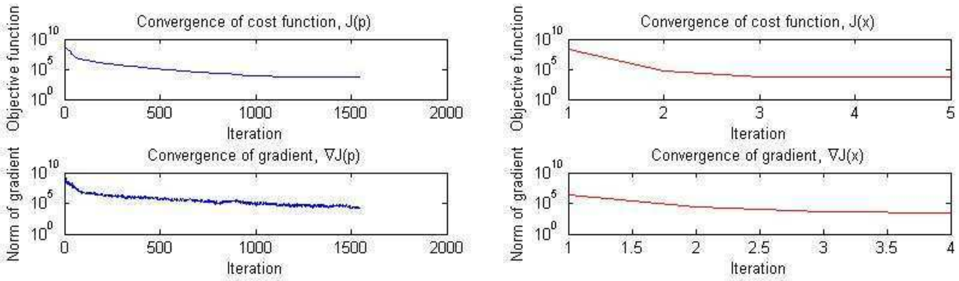


Figure 7.1: Contour plot of the time evolution (vertical axis) of the $N = 40$ variables (horizontal axis). Colour bar represents atmospheric quantity value.

Figure 7.1 shows a truth run of the Lorenz 95 model. The position of N sectors on a latitude circle at a given time are represented by the X_i variables on the horizontal axis. So imagine the latitude circle has been put onto a straight line.

The values of the variables X_i are represented by their colour. These variables can be any atmospheric quantity, for example, temperature [62]. The vertical axis represents time, thus the plot shows us the temporal evolution of these atmospheric quantities with respect to their position.



(a) $\mathcal{J}(\mathbf{p})$ (top) and $||\nabla\mathcal{J}(\mathbf{p})||$ (bottom). (b) $\mathcal{J}(\mathbf{x})$ (top) and $||\nabla\mathcal{J}(\mathbf{x})||$ (bottom).

Figure 7.2: Respective objective function and gradient norm values with the number of minimisation iterations.

We see here in Figures 7.2(a) and (b) that the model error formulation requires $\mathcal{O}(10^3)$ more iterations than the state formulation to converge to the same tolerance. We now examine the relative errors in the solutions.

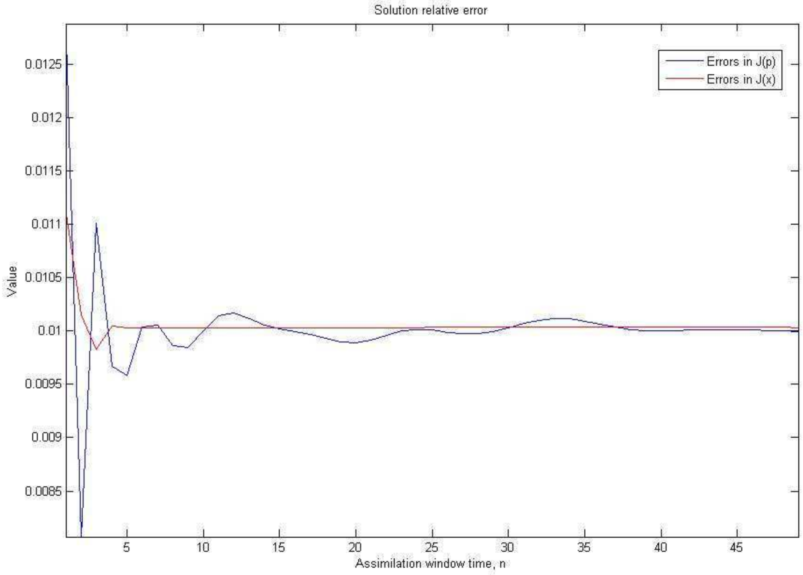


Figure 7.3: Solution relative errors throughout the assimilation window, $\mathcal{J}(\mathbf{p})$ (blue line) and $\mathcal{J}(\mathbf{x})$ (red line).

Figure 7.3 shows the errors are spread in a similar manner, with the range of errors exhibited by the solution to the $\mathcal{J}(\mathbf{p})$ problem being slightly larger than $\mathcal{J}(\mathbf{x})$. This is confirmed by the total solution relative error of both the model error and state formulations, which are 0.017 and 0.012 respectively.

The results in this experiment show that for long assimilation windows with plentiful observations, the model error formulation requires many iterations to converge, which agrees with findings in Chapter 5 Section 5.3.3. This experiment also demonstrates the special case for the state formulation, where if the state is fully observed, increasing the length of the assimilation window has no effect on the number of iterations or condition number of $\mathbf{S}_x$. This is consistent with our findings on the Hessian condition numbers in Chapter 6, Section 6.3.5, Experiment 4, Figure 6.7 and convergence iterates in Chapter 6, Section 6.4.3, Experiment 3.

7.1.4 Experiment 2: Background and Model Error Correlation Length-Scales

In this experiment we examine the sensitivity of the iterations, solution relative errors and Hessian condition numbers of the model error and state formulations to correlation length-scales of the matrices composing the $\mathbf{D}$ matrix. It is important to remember that the condition number of the background error covariance matrix will be more sensitive to its correlation length-scale in comparison to the condition number of the model error covariance matrix, since the SOAR covariance matrix is more sensitive to correlation length-scale than the Laplacian covariance matrix, (discussed in Chapter 4). We expect that increase in correlation length-scale of C_B and C_Q to increase the number of iterations, the solution relative errors and the Hessian condition numbers of both formulations, with the state formulation exhibiting an increased sensitivity over the model error formulation.

The experiment settings are the same as Section 7.1.2 except for the following. We reduce the assimilation window length to the equivalent of one day $n = 8$. We observe every 10^{th} variable such that $q = N/10 = 4$. The error variances are all equal $\sigma_b^2 = \sigma_q^2 = \sigma_o^2 = 1$ to ensure that the only source of ill-conditioning will arise from the correlation length-scales being varied.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	18	31	47	89	155	239
	0.03	25	46	76	79	146	214
	0.05	42	45	57	102	140	201
	0.07	49	56	60	100	170	202
	0.09	77	79	83	123	212	221
	0.11	102	110	100	135	162	277

Table 7.7: Number of iterations for $\mathcal{J}(\mathbf{p})$.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	61	176	370	685	891	1128
	0.03	174	191	420	695	967	1497
	0.05	348	508	385	742	1343	1108
	0.07	617	519	584	663	933	1182
	0.09	497	781	632	826	1739	933
	0.11	709	680	593	604	545	813

Table 7.8: Number of iterations for $\mathcal{J}(\mathbf{x})$.

$L(C) = 0.025$ is equivalent to Δx.

Tables 7.7 and 7.8 show the sensitivity of the iteration numbers of both formulations to the condition number of $\mathbf{D}$, which rises with correlation length-scales $L(C_B)$ and $L(C_Q)$. The number of iterations required for the state formulation to converge consistently exceeds the model error formulation. We also observe that the state formulation is much more sensitive to identical increases in the condition number of $\mathbf{D}$ than model error formulation. Taking the specific example where $L(C_B) == L(C_Q) = 0.09$, we see that the number of iterations for the state formulation exceeds the number of iterations for the model error formulation by nearly one order of magnitude. This agrees with findings in Chapter 5 and Chapter 6, where the state formulation is more sensitive to identical increases in the correlation length-scales of C_B and C_Q than the model error formulation.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	0.260	0.271	0.236	0.227	0.268	0.237
	0.03	0.372	0.441	0.353	0.293	0.314	0.351
	0.05	0.449	0.332	0.416	0.358	0.283	0.314
	0.07	0.346	0.486	0.220	0.284	0.347	0.246
	0.09	0.167	0.176	0.401	0.306	0.239	0.219
	0.11	0.184	0.274	0.098	0.277	0.187	0.274

Table 7.9: Total solution relative error, $\mathcal{J}(\mathbf{p})$.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	0.264	0.273	0.238	0.228	0.268	0.238
	0.03	0.383	0.447	0.376	0.296	0.316	0.355
	0.05	0.628	0.335	0.593	0.363	0.291	0.335
	0.07	0.365	0.511	0.413	0.309	0.417	0.254
	0.09	0.295	0.183	0.461	0.324	0.390	0.261
	0.11	0.274	0.428	0.190	0.397	0.237	0.513

Table 7.10: Total solution relative error, $\mathcal{J}(\mathbf{x})$.

Tables 7.9 and 7.10 show that the solution relative errors of the state formulation are consistently larger than the model error formulation. In a specific example where $L(C_B) = 0.05$ and $L(C_Q) = 0.11$, the solution relative error of the state formulation is almost one order of magnitude higher than the model error formulation. So it is clear that the model error formulation is less sensitive to correlation length-scale and provides consistently more accurate solution in comparison to the state formulation.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	6.11E+18	5.02E+18	7.16E+18	1.90E+19	3.74E+19	2.45E+20
	0.03	5.70E+18	9.42E+19	3.46E+19	5.70E+20	5.01E+20	6.41E+19
	0.05	3.39E+20	1.12E+19	1.15E+20	1.44E+20	5.55E+19	1.26E+20
	0.07	1.21E+19	2.53E+20	1.48E+20	8.32E+21	9.83E+20	8.40E+20
	0.09	2.73E+21	3.44E+21	2.86E+19	1.57E+20	2.48E+20	2.72E+20
	0.11	9.35E+18	1.56E+20	5.22E+20	3.53E+21	4.42E+20	2.34E+21

Table 7.11: Condition number values for $\mathbf{S}_p$.

		$L(C_B)$					
		0.01	0.03	0.05	0.07	0.09	0.11
$L(C_Q)$	0.01	2.18E+19	1.81E+19	6.50E+19	2.42E+19	7.28E+21	1.18E+19
	0.03	4.18E+20	1.41E+19	7.30E+18	8.48E+20	2.72E+18	1.72E+19
	0.05	8.08E+19	1.17E+19	9.92E+18	6.99E+18	1.95E+19	1.13E+19
	0.07	1.51E+19	5.39E+19	6.06E+19	1.67E+19	3.58E+20	6.93E+19
	0.09	3.08E+19	5.45E+19	1.17E+20	8.44E+20	1.38E+20	1.38E+19
	0.11	1.17E+20	7.82E+19	2.92E+20	1.44E+21	4.03E+22	2.28E+20

Table 7.12: Condition number values for $\mathbf{S}_x$.

The condition numbers in Tables 7.11 and 7.12 for both formulations are very similar which was not expected based on the results obtained in Chapters 5 and 6. However, Tables 7.11 and 7.12 show that as the correlation length-scales of B and Q increase, then so do the condition numbers of $\mathbf{S}_p$ and $\mathbf{S}_x$, which is compatible with the iteration results in Tables 7.7 and 7.8. These results do not complement the iteration number figures in Tables 7.7 and 7.8, which indicates that the higher order terms of the Taylor expansion of both objective functions may be large.

To summarise, we see that the results related to the number of iterations in

Tables 7.7 and 7.8 strongly agrees with our findings in Chapter 5, Section 5.2.1.1 and Section 5.1 and Chapter 6, Section 6.3.4 and Section 6.4.2. The number of iterations of both the model error and state formulations both rise as both correlation length-scales increase, with an increased sensitivity to $L(C_B)$ as we expected. The state formulation also exhibits a much more visible increase in iterations in comparison to the model error formulation, which was also to be expected. The relative solution errors in Tables 7.9 and 7.10 were also to be expected, since the experiments in Chapter 5 and Chapter 6 showed that the solution errors of both formulations did not rise with correlation length-scale. Finally, we would have expected to see differences in the condition numbers of both formulations, but Tables 7.11 and 7.12 do not reflect this.

We now summarise this chapter.

7.2 Summary

To summarise, we showed in Experiment 1 that the number of iterations, solution relative error and Hessian condition numbers of both formulations are sensitive to assimilation window length and observation density with the Lorenz 95 system as the model. More specifically, we showed that:

1. As the assimilation window increases, the condition number of $\mathbf{S}_p$ and the number of iterations for convergence of the model error formulation also increase. We also see that as the number of observations increase, the condition number of $\mathbf{S}_p$ and number of iterations of the model error formulation also increase. This agrees with findings in Chapter 5.

2. The state formulation solution errors and iteration count increase as the assimilation window length increases, for any number of observations. The exception to this is shown in Experiment 1(ii) where the state formulation out-performs the model error formulation when the state is *fully observed*. This coincides with findings in Chapter 6.

In Experiment 2 we showed that both formulations exhibit an increase in the number of iterations (Tables 7.7 and 7.8) and Hessian condition numbers (Tables 7.11 and 7.12) as the condition number of the background and model error matrix **D** increases. We increased the condition number of the background and model error matrix by increasing the correlation length-scales of the background and model errors. Additionally, the increased sensitivity of the state formulation over the model error formulation to the background and model error correlation length-scales was also seen in Table 7.7 and Table 7.8.

We now conclude the thesis.

Chapter 8

Conclusions

The weak-constraint 4DVAR problem is a variational data assimilation technique, which unlike the conventional sc4DVAR method, accounts for model error, [83]. The wc4DVAR technique has two known formulations, both of which have been employed in various applications in the literature, [84], [56], [19], [68], [66] and [67]. Obtaining a solution to the wc4DVAR problem requires the minimisation of the objective function and its gradient. The widely used method of choice for solving the variational problem is the gradient-based Gauss-Newton 'incremental' technique. As we showed in Section 3.1, the condition number of the Hessian is an appropriate measure of understanding the sensitivities of the solution to changes in the input data composing the data assimilation problem.

We now draw conclusions from the work in this thesis followed by our ideas for further research.

8.1 Conclusions

We intended to understand the differences between the model error and state formulations of the wc4DVAR problem. In Chapter 4 we showed that by changing a few data assimilation parameters, the iterative minimisation characteristics of both problems can change dramatically. We found that the formulations were both sensitive to observation density, error variances and the length of the assimilation window. We also found that even when using identical settings for the generated truth and assimilation, both wc4DVAR solutions consistently under-estimated the true model error variance slightly.

We then examined the model error formulation more closely in Chapter 5, by bounding the condition number of the first-order Hessian under simplified assumptions and examining the bound expressions for sensitivities of the solution to specific input parameters. We found that the model error formulation Hessian condition number was sensitive to the background and model error covariance matrix. This implied that the Hessian condition number is sensitive to both the correlation length-scales of the background and model errors, and the ratio of the background and model error variances. We also found that the Hessian condition number of the model error formulation to be sensitive to the observation accuracy, observation density and assimilation window length. We then examined the preconditioned model error formulation showing that the condition number and convergence rates are much improved.

An examination of the condition number of the first-order Hessian of the state formulation followed in Chapter 6. We found that, under simplified assumptions, the state formulation shared certain sensitivities with model error formulation. One of these was the sensitivity to the background and model error error covariance matrix, however this was more pronounced for the state formulation than for the model error formulation. We also found the state formulation to be sensitive to the observation density and assimilation window length, although there were some unique differences. The state formulation Hessian condition number becomes ill-conditioned as the observation density *decreases*, which also amplifies its sensitivity to the assimilation window length. If the state is fully observed, then the state formulation is no longer sensitive to the assimilation window length. This is an interesting advantage, however, a fully observed state is unrealistic in operational applications.

We finally explored the wider-scope application of the theoretical results on a non-linear, chaotic Lorenz 95 model in Chapter 7. We found that the sensitivities of both formulations also show in specific experiments for the observation density, assimilation window length and correlation length-scales.

The following points were covered in the thesis:

- In Chapter 4 we detailed the practical implementation of the wc4DVAR formulations on the 1-dimensional linear advection model, which highlighted clear differences in the minimisation characteristics of both formulations based on changes in experimental parameters. We also observed in several experiments that a general trait of both wc4DVAR formulations is that the model errors are under-estimated.
- The condition number of the Hessian of the sc4DVAR problem is bounded above by the condition number of the Hessian of the model error formulation, $\mathbf{S}_p$, shown in Appendix A.
- We identified and demonstrated the following sources of ill-conditioning of the Hessian of the model error formulation. We did this both theoretically and complemented it with numerical experiments to show similar effects on the rate of convergence in Chapter 5:
    - The condition number of the background and model error covariance matrix, $\mathbf{D}$.
        - As the ratio of the background and model error variance increases or decreases away from 1, $\mathbf{D}$ becomes ill-conditioned and therefore so does $\mathbf{S}_p$.
        - As the correlation length-scales of the background and model error covariance matrix increases, $\mathbf{S}_p$ becomes more ill-conditioned.
    - Increasing the assimilation window length increases the condition number of $\mathbf{S}_p$ at a potentially quadratic rate.

    - The ratio of the largest of the background and model error variance to the observation error variance also renders $\mathbf{S}_p$ ill-conditioned if it increases or decreases from 1. This means increasing *observation* accuracy (lower observation variance), *background* accuracy or even *model* accuracy can harm the conditioning of $\mathbf{S}_p$, if the ratios of these three error variances diverges away from 1.

- We also preconditioned the model error formulation with the symmetric square root of $\mathbf{D}$ and showed that the condition number of the preconditioned Hessian was much improved in comparison to that of $\mathbf{S}_p$, both theoretically and numerically. We also showed the convergence rate of the iterative solver used on the preconditioned objective function to be much improved as a result of preconditioning.

- We identified the following sources of ill-conditioning of the condition number of the Hessian of the state formulation, $\mathbf{S}_x$. We also demonstrated that these sources of ill-conditioning subsequently have an adverse effect on the iterative convergence rate of the state formulation in Chapter 6:

    - Assimilation window length and observation density. If we have a fully observed state then the condition number of $\mathbf{S}_x$ is no longer affected by the length of the assimilation window. We also see that as the observation density decreases the condition number of $\mathbf{S}_x$ becomes ill-conditioned. This was discussed in Section 6.2 and demonstrated numerically in Sections 6.3.5 and 6.4.3.

    - The sensitivity to the condition number of $\mathbf{D}$. As the correlation length-scales of the covariances matrices of the background and model errors increase, $\mathbf{S}_x$ becomes ill-conditioned and the iterative convergence rate suffers as a result. We also showed that the sensitivity of the condition number of $\mathbf{S}_x$ to the condition number of $\mathbf{D}$ is greater than $\mathbf{S}_p$.

- We examined the effect of assimilation window length, observation density and condition number of $\mathbf{D}$, via the correlation length-scales, on the minimisation characteristics of both the model error and state formulations applied to the non-linear chaotic Lorenz 95 model. We showed:

- Increasing the correlation length-scales of the matrices composing $\mathbf{D}$ increases the number of iterations required for the model error and state algorithms to converge, where the state formulation exhibits a larger increase in iterations than the model error formulation.
- An increase in the number of observations and assimilation window length increases the number of iterations for the model error algorithm to converge.
- Decreasing the number of observations for any length of assimilation window increases the number of iterations required for the state algorithm to converge.
- For a fully observed state, increasing the assimilation window length does not affect the number of iterates required for the state algorithm to converge.

From the research shown in this thesis we can draw a few general conclusions. The sensitivities shared by both formulations are: background and model error covariance matrix correlation length-scales, error variance ratios, observation density and assimilation window length. These sensitivities are shared but they have different effects on each wc4DVAR formulation, as we have discussed in this chapter. It is interesting and worth noting however that the state formulation is not affected by assimilation window length if the state is fully observed. Although a fully observed state is unrealistic, this suggests that there is a way of enabling the state formulation to be more stable. We also see throughout the thesis that the state formulation exhibits increased sensitivity in comparison to the parameters which influence its condition number. We conclude that the model error formulation is not as 'fragile' as the state formulation to its own sensitivities and therefore the model error formulation is the more stable of the two wc4DVAR algorithms to use until a suitable preconditioner for the state formulation is found.

We now discuss avenues for further work before bringing the thesis to a close.

8.2 Further Work

The work in this thesis establishes a theoretical basis for the conditioning of the model error and state estimation wc4DVAR problems. However, the theory established in this thesis is limited to the simple assumptions made to derive the theorems. We assumed that observations were taken of the state directly, which allows for a simple observation operator. In reality however, observations may be obtained from satellite radiances for example, which means that the observation operator would be some form of the radiative transfer equation. The radiative transfer equation has the potential of being highly non-linear and quite difficult to deal with, [65].

We could also relax the assumption of uncorrelated observation errors. Observation error spatial correlations are typically ignored in data assimilation while the error variances are over-inflated to compensate for the lack of information on correlations. While this assumption is not realistic, observation correlations are ignored because it makes the implementation of 4DVAR easier in general. Studies into the known sources of observation error have narrowed it down to four sources; measurement error, observation operator errors, quality control errors and representativity errors, [87]. The latter three sources of error are believed to be correlated in space, while it has been suggested that observation errors are potentially temporally correlated, [79]. Incorporating correlated observation errors has only begun to be operationally implemented by the Met Office, [90], while there are still problems with the conditioning of 4DVAR, [89].

Another assumption we made to obtain the theory was that the background, model and observation errors were not time-correlated. It is common practice in NWP to ignore time correlations because it is simply too computationally expensive to deal with. However, there have been studies to show that, for example, model error can be correlated with time, [26], and also observation errors in remote sensing for example, are correlated in time, [80].

The work in Chapter 7 could have been complemented with using the Gauss-Newton 'incremental' wc4DVAR technique. We could also employ the preconditioned model error algorithm using both the incremental technique and the non-linear Polak-Ribiere conjugate gradient technique, to see if the preconditioning has similar effects to those shown in Chapter 5 on the linear advection equation. Comparing the differences in convergence rates and solution errors of the Polak-Ribiere and incremental approach would be interesting. We would expect the incremental approach to at least as good as the iterative minimisation performance of the Polak-Ribiere technique, if not better.

Another practical aspect worth considering would be to investigate the validity of the conditioning theory in this thesis on larger systems such as the ECMWF Object-Orientated Programming System (OOPS), or even the University of California's operational Regional Ocean Modeling System (ROMS). Testing the theory on bigger systems to investigate the sensitivity of both minimisation algorithms to the input parameters discovered to be sensitive in this thesis would be the next logical step.

In this thesis we preconditioned the model error formulation using the symmetric square root of **D**, which we showed to improve the conditioning and minimisation properties considerably. We could also consider the preconditioning of the state formulation, which was shown to be *very* sensitive to the condition number of **D**. As a first step we could precondition the state formulation using the symmetric square root of **D** to understand if it improves its stability. M. Fisher and S. Gürol have established an alternative saddle point formulation of the state formulation, which has the advantage of avoiding the need to invert **D**, [27], [25]. In [27] the authors identified that the Hessian of the state formulation can be preconditioned using an approximation of the wc4DVAR model propagator, **L**. However they also showed that the formulation is very sensitive to the approximation of **L**. We could also study the conditioning of the saddle-point formulation problem, where the

Hessian matrix is symmetric indefinite.

Another formulation that would be useful to consider is the weak-constraint equivalent of the dual formulation, [12]. The weak-constraint problem, which is considerably larger than the strong-constraint problem can be mapped into observation space to reduce the size of the problem and achieve an equivalent solution. The attractive prospect of this is that wc4DVAR is a much larger problem than sc4DVAR, so wc4DVAR would possibly benefit more from being solved in the lower dimensional observation space. Investigating the conditioning of the weak-constraint dual problem would also complement the work by A. El Akkraoui et al. [20], [21], but this has yet to be done.

This concludes the thesis *quod erat faciendum*.

Appendix A

General Upper Bound: The Strong-Constraint 4DVAR Hessian Condition Number

We write the sc4DVAR Hessian, $S \in \mathbb{R}^{N \times N}$ as

$$S = B_0^{-1} + \hat{H}^T \mathbf{R}^{-1} \hat{H}, \tag{A.1}$$

where

$$\hat{H} = \left[H_0^T, (H_1 M_{1,0})^T, (H_2 M_{2,0})^T, \ldots, (H_n M_{n,0})^T \right]^T, \tag{A.2}$$

notice that $\hat{H}$ is identical to the first block column of $\mathbf{HL}^{-1}$ in the weak constraint Hessian matrix (2.40).

We now present a general result, which shows the eigenvalue spectrum of the Hessian of sc4DVAR is bounded by the eigenvalue spectrum of the Hessian of wc4DVAR formulation (2.32).

Theorem A.0.1 *The condition number of the Hessian of the strong-constraint problem is bounded such that*

$$\kappa(S) \leq \kappa(\mathbf{S}_p). \tag{A.3}$$

Proof: We prove this by showing that the largest and smallest eigenvalues of S can be obtained by taking an appropriate Rayleigh Quotient of $\mathbf{S}_p$. To illustrate this we denote the spectrum of S by $[\lambda_N, \lambda_1]$, where λ_N is the smallest eigenvalue and λ_1 is the largest eigenvalue of S. Similarly we let the interval $[\sigma_{N(n+1)}, \sigma_1]$ denote the spectrum of $\mathbf{S}_p$. Since we know the bounds of the Rayleigh Quotient from Theorem 3.4.7 , we aim to show

$$\sigma_{N(n+1)} \leq \lambda_N \leq \lambda_1 \leq \sigma_1. \tag{A.4}$$

Note this does not mean that an eigenvalue of S is necessarily an eigenvalue of $\mathbf{S}_p$.

Consider the Rayleigh Quotient of $\mathbf{S}_p$

$$\mathcal{R}_{\mathbf{S}_p}(w) = w^T(\mathbf{D}^{-1} + \mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})w, \tag{A.5}$$

where $w \in \mathbb{R}^{N(n+1)}$ is such that

$$w = \begin{pmatrix} v_1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \tag{A.6}$$

where v_1 is an eigenvector of S corresponding to the largest eigenvalue.

We compute the first part of the Rayleigh Quotient of $\mathbf{S}_p$,

$$w^T\mathbf{D}^{-1}w = v_1^T B_0 v_1. \tag{A.7}$$

Computing the second part yields

$$\mathbf{H}\mathbf{L}^{-1}w = \hat{H}v_1. \tag{A.8}$$

The transpose of this statement is also true. Therefore the second term yields

$$w^T(\mathbf{L}^{-T}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1})w = v_1^T\hat{H}^T\mathbf{R}^{-1}\hat{H}v_1. \tag{A.9}$$

The Rayleigh Quotient of $\mathbf{S}_p$ is then

$$\mathcal{R}_{\mathbf{S}_p}(w) = v_1^T B_0 v_1 + v_1^T\hat{H}^T\mathbf{R}^{-1}\hat{H}v_1 = \mathcal{R}_S(v_1) = \lambda_1, \tag{A.10}$$

as required. The largest eigenvalue of the Hessian of the strong-constraint problem exists in the eigenvalue interval of the Hessian of the weak-constraint problem (2.32).

The same argument can be made for the smallest eigenvalue λ_N of S. If the largest and smallest eigenvalues of S both exist in the eigenvalue interval of $\mathbf{S}_p$, recalling the bounds of the Rayleigh Quotient from Theorem (3.4.7),

$$\lambda_N \leq \mathcal{R}_S(\mathbf{x}) \leq \lambda_1, \tag{A.11}$$

$$\sigma_{N(n+1)} \leq \mathcal{R}_{\mathbf{S}_p}(\mathbf{x}) \leq \sigma_1, \tag{A.12}$$

we have

$$\sigma_{N(n+1)} \leq \lambda_N \leq \lambda_1 \leq \sigma_1. \tag{A.13}$$

Finally, the condition number as defined in (3.9) is the ratio of the largest and smallest eigenvalue. So it follows that $\kappa(S)$ is less then or equal to $\kappa(\mathbf{S}_p)$.

This completes the proof. ■

The condition number of the Hessian of sc4DVAR being less then or equal to the condition number of the Hessian of wc4DVAR formulation (2.32) suggests that the iterative performance of wc4DVAR should not exceed the iterative performance of the sc4DVAR when solving identical data assimilation problems. Ideally, wc4DVAR will at least have the same convergence characteristics as sc4DVAR.

Bibliography

[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. *Quarterly Journal of the Royal Meteorological Society*, 126(565):1455–1472, 2000.

[2] E. Atkins, M. Morzfeld, and A. Chorin. Implicit particle methods and their connection with variational data assimilation. *arXiv preprint arXiv:1205.1830*, 2012.

[3] O. Axelsson. *Iterative solution methods.* Cambridge University Press, 1996.

[4] R.N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, 2008.

[5] M. Bell, M. Martin, and N. Nichols. Assimilation of data into an ocean model with systematic errors near the equator. *Quarterly Journal of the Royal Meteorological Society*, 130(598):873–893, 2004.

[6] A.F. Bennett, B.S. Chua, D. Harrison, and M.J McPhaden. Generalized inversion of tropical atmosphere-ocean data and a coupled model of the tropical pacific. *Journal of Climate*, 11(7):1768–1792, 1998.

[7] A.F. Bennett, B.S. Chua, D. Harrison, and M.J. McPhaden. Generalized inversion of tropical atmosphere-ocean (tao) data and a coupled model of the

tropical pacific. part ii: The 1995-96 la nina and 1997-98 el nino. *Journal of Climate*, 13(15):2770–2785, 2000.

[8] P. Bergthórsson and B. Döös. Numerical weather map analysis. *Tellus*, 7(3):329–340, 1955.

[9] M. Bocquet, C. Pires, and L. Wu. Beyond gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138(8):2997–3023, 2010.

[10] A. Clayton, A. Lorenc, and D. Barker. Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461, 2013.

[11] R. Courant and H. Robbins. *What is Mathematics? An elementary approach to ideas and methods*. Oxford University Press, 1996.

[12] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 123(544):2449–2461, 1997.

[13] P. Courtier, J. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[14] M. Cullen. Analysis of cycled 4d-var with model error. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1473–1480, 2013.

[15] R. Daley. *Atmospheric data analysis*, volume 2. Cambridge university press, 1993.

[16] D. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, 2005.

[17] D. Dee and S. Uppala. Variational bias correction of satellite radiance data in the era-interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 135(644):1830–1841, 2009.

[18] D.P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly weather review*, 123(4):1128–1145, 1995.

[19] G. Desroziers, J. Camino, and L. Berre. 4denvar: link with 4d state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2097–2110, 2014.

[20] A. El Akkraoui. *The primal and dual forms of variational data assimilation in the presence of model error.* PhD thesis, McGill Univeristy, Montreal, Canada, 2010.

[21] A. El Akkraoui and P. Gauthier. Convergence properties of the primal and dual forms of variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 136(646):107–115, 2010.

[22] D. Fairbairn, S. Pring, A. Lorenc, and I. Roulstone. A comparison of 4dvar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, 140(678):281–294, 2014.

[23] D. Feingold and R. Varga. Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem. *Pacific J. Math*, 12(4):1241–1250, 1962.

[24] E. Fertig, J. Harlim, and B. Hunt. A comparative study of 4d-var and a 4d ensemble kalman filter: Perfect model simulations with lorenz-96. *Tellus A*, 59(1):96–100, 2007.

[25] M. Fisher and S. Gürol. Parallelisation in the time dimension of four-dimensional variational data assimilation. 2014. submitted to Quarterly Journal of the Royal Meteorological Society.

[26] M. Fisher, M. Leutbecher, and G.A. Kelly. On the equivalence between kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3235–3246, 2005.

[27] M. Fisher, Y. Trémolet, H. Auvinen, D. Tan, and P. Poli. Weak-constraint and long-window 4d-var. Technical Report 621, European Centre for Medium-Range Weather Forecasts (ECMWF), 2011.

[28] R. Fletcher and C. Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154, 1964.

[29] L. Gandin and R. Hardin. *Objective analysis of meteorological fields*, volume 242. Israel program for scientific translations Jerusalem, 1965.

[30] G. Gaspari and S. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

[31] P. Gauthier, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau. Extension of 3dvar to 4dvar: Implementation of 4dvar at the meteorological service of canada. *Monthly weather review*, 135(6):2339–2354, 2007.

[32] S. Geršgorin. Uber die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 6(1):749–754, 1931.

[33] B. Gilchrist and G. Cressman. An experiment in objective analysis. *Tellus*, 6(4):309–318, 1954.

[34] P. Gill, W. Murray, and M. Wright. Practical optimization. 1981.

[35] G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3. Johns Hopkins Univ Pr, 1996.

[36] S. Gratton, S. Gürol, and P. Toint. Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear-least squares problems. *Computational Optimization and Applications*, 54(1):1–25, 2013.

[37] R.M. Gray. *Toeplitz and circulant matrices: A review.* Now Pub, 2006.

[38] A.K. Griffith. *Data assimilation for numerical weather prediction using control theory.* PhD thesis, University of Reading, 1997.

[39] A.K. Griffith and N.K. Nichols. Adjoint methods in data assimilation for estimating model error. *Flow, turbulence and combustion*, 65(3):469–488, 2000.

[40] D.F. Griffiths and A.R. Mitchell. *The finite difference method in partial differential equations*. John Wiley, 1980.

[41] S.A. Haben. *Conditioning and preconditioning of the minimisation problem in variational data assimilation*. PhD thesis, University of Reading, 2011.

[42] S.A. Haben, A.S. Lawless, and N.K. Nichols. Conditioning and preconditioning of the variational data assimilation problem. *Computers & Fluids*, 46(1):252–256, 2011.

[43] S.A. Haben, A.S. Lawless, and N.K. Nichols. Conditioning of incremental variational data assimilation, with application to the met office system. *Tellus A*, 63(4):782–792, 2011.

[44] M.R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*, volume 49. National Bureau of Standards Washington, DC, 1952.

[45] K. Ide, P. Courtier, M. Ghil, and A.C. Lorenc. Unified notation for data assimilation: operational, sequential and variational. *Practice*, 75(1B):181–189, 1997.

[46] B. Ingleby. The statistical structure of forecast errors and its representation in the met. office global 3-d variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127(571):209–231, 2001.

[47] T. Kadowaki. A 4-dimensional variational assimilation system for the jma global spectrum model. *CAS/JAC WGNE Research Activities in Atmospheric and Oceanic Modelling*, 34:1–17, 2005.

[48] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[49] E. Kalnay. *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press, 2003.

[50] E. Klinker, F. Rabier, G. Kelly, and J.F. Mahfouf. The ecmwf operational implementation of four-dimensional variational assimilation. iii: Experimental results and diagnostics with operational configuration. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1191–1215, 2000.

[51] A.S. Lawless. Variational data assimilation for very large environmental problems. *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences, Radon Series on Computational and Applied Mathematics*, 13:55–90, 2012.

[52] A.S. Lawless, S. Gratton, and N.K. Nichols. An investigation of incremental 4d-var using non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 131(606):459–476, 2005.

[53] A.S. Lawless and N.K. Nichols. Inner-loop stopping criteria for incremental four-dimensional variational data assimilation. *Monthly weather review*, 134(11):3425–3435, 2006.

[54] M-S. Lee and D-K. Lee. An application of a weakly constrained 4dvar to satellite data assimilation and heavy rainfall simulation. *Monthly weather review*, 131(9):2151–2176, 2003.

[55] P. Lewis, J. Gómez-Dans, T. Kaminski, J. Settle, T. Quaife, N. Gobron, J. Styles, and M. Berger. An earth observation land data assimilation system (eo-ldas). *Remote Sensing of Environment*, 120:219–235, 2012.

[56] M. Lindskog, D. Dee, Y. Trémolet, E. Andersson, G. Radnóti, and M. Fisher. A weak-constraint four-dimensional variational analysis system in the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, 135(640):695–706, 2009.

[57] A. Lorenc. Iterative analysis using covariance functions and filters. *Quarterly Journal of the Royal Meteorological Society*, 118(505):569–591, 1992.

[58] A. 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, et al. The met. office global

three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.

[59] A.C. Lorenc. The potential of the ensemble kalman filter for nwpa comparison with 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003.

[60] A.C. Lorenc, N.E. Bowler, A.M. Clayton, S.R. Pring, and D. Fairbairn. Comparison of hybrid-4denvar and hybrid-4dvar data assimilation methods for global nwp. *Monthly Weather Review*, 140(2014):281–294, 2014.

[61] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

[62] E.N. Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

[63] E.N. Lorenz. Designing chaotic models. *Journal of the atmospheric sciences*, 62(5), 2005.

[64] J-F. Mahfouf and F. Rabier. The ecmwf operational implementation of four-dimensional variational assimilation. ii: Experimental results with improved physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1171–1190, 2000.

[65] A.P. McNally, P.D. Watts, J.A. Smith, R. Engelen, G.A. Kelly, J.N. Thépaut, and M. Matricardi. The assimilation of airs radiance data at ecmwf. *Quarterly Journal of the Royal Meteorological Society*, 132(616):935–957, 2006.

[66] A.M. Moore, H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Foley, J.D. Doyle, D. Costa, and P. Robinson. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: part ii–performance and application to the california current system. *Progress in Oceanography*, 91(1):50–73, 2011.

[67] A.M. Moore, H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B. Powell, D. Foley, J.D. Doyle, D. Costa, and P. Robinson. The regional ocean modeling

system (roms) 4-dimensional variational data assimilation systems: Part iii–observation impact and observation sensitivity in the california current system. *Progress in Oceanography*, 91(1):74–94, 2011.

[68] A.M. Moore, H.G. Arango, G. Broquet, B.S. Powell, A.T. Weaver, and J. Zavala-Garay. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part i–system overview and formulation. *Progress in Oceanography*, 91(1):34–49, 2011.

[69] K.W. Morton, D.F. Mayers, and M. Cullen. *Numerical solution of partial differential equations*, volume 2. Cambridge university press Cambridge, 1994.

[70] H. Ngodock and M. Carrier. A 4dvar system for the navy coastal ocean model. part i: System description and assimilation of synthetic observations in monterey bay*. *Monthly Weather Review*, 142(6):2085–2107, 2014.

[71] H. Ngodock and M. Carrier. A 4dvar system for the navy coastal ocean model. part ii: Strong and weak constraint assimilation experiments with real observations in monterey bay*. *Monthly Weather Review*, 142(6):2108–2117, 2014.

[72] N.K. Nichols. Treating model error in 3-d and 4-d data assimilation. *Data assimilation for the earth system*, pages 127–135, 2003.

[73] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer verlag, 1999.

[74] F. Rabier, H. Järvinen, E. Klinker, J-F. Mahfouf, and A. Simmons. The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1143–1170, 2000.

[75] 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 office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133(623):347–362, 2007.

[76] Y. Sasaki. A fundamental study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan*, 33(6):262–275, 1955.

[77] Y. Sasaki. An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, 36(3):77–88, 1958.

[78] Y. Sasaki. Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98(12):875–883, 1970.

[79] L.M Stewart, S. Dance, and N.K. Nichols. Data assimilation with correlated observation errors: experiments with a 1-d shallow water model. *Tellus A*, 65, 2013.

[80] L.M. Stewart, S.L. Dance, and N.K. Nichols. Correlated observation errors in data assimilation. *International journal for numerical methods in fluids*, 56(8):1521–1527, 2008.

[81] E. Süli and D.F. Mayers. *An introduction to numerical analysis.* Cambridge University Press, 2003.

[82] O. Toeplitz. Zur theorie der quadratischen und bilinearen formen von unendlichvielen vernderlichen. i. teil: Theorie der l-formen. *Math. Annal.*, 70:351–376, 1911.

[83] Y. Trémolet. Accounting for an imperfect model in 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2483–2504, 2006.

[84] Y. Trémolet. Model-error estimation in 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 133(626):1267–1280, 2007.

[85] R. Varga. *Gershgorin and His Circles in Springer Series in Computational Mathematics, 36.* Springer, Berlin, 2004.

[86] P.A. Vidard, A. Piacentini, and F-X. Le Dimet. Variational data analysis with control of the forecast bias. *Tellus A*, 56(3):177–188, 2004.

[87] J.A. Waller, S.L. Dance, A.S. Lawless, N.K. Nichols, and J.R. Eyre. Representativity error for temperature and humidity using the met office

high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1189–1197, 2014.

[88] R.O. Weber and P. Talkner. Some remarks on spatial correlation function models. *MONTHLY WEATHER REVIEW-USA*, 121:2611–2611, 1993.

[89] P. Weston. Progress towards the implementation of correlated observation errors in 4d-var. *Met Office Forecasting Research Technical Report*, 560, 2011.

[90] P.P. Weston, W. Bell, and J.R. Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429, 2014.

[91] P. Wolfe. Convergence conditions for ascent methods. *SIAM review*, 11(2):226–235, 1969.

[92] A. Wood. When is a truncated covariance function on the line a covariance function on the circle? *Statistics & probability letters*, 24(2):157–164, 1995.

[93] L. Xu, T. Rosmond, J. Goerss, and B. Chua. Toward a weak constraint operational 4d-var system: application to the burgers' equation. *Meteorologische Zeitschrift*, 16(6):741–753, 2007.

[94] D. Zupanski. A general weak constraint applicable to operational 4dvar data assimilation systems. *Monthly Weather Review*, 125(9):2274–2292, 1997.