T H E U N I V E R S I T Y O F R E A D I N G

Variational Data Assimilation: A Study

by

M.A. Wlasak

Numerical Analysis Report 10/97

D E P A R T M E N T O F M A T H E M A T I C S

THE UNIVERSITY OF READING
DEPARTMENT OF MATHEMATICS

Variational Data Assimilation: a Study

by M.A.Wlasak

Abstract

This dissertation is a study on variational data assimilation. Motivation for the subject is given in the Introduction. The mathematical background needed is given in Chapter 2. This followed by an examination of the minimisation procedure. Chapter 4 considers data assimilation on a discretisation of the Lorenz equations. It was shown that the time period over which data assimilation is successful may be extended by applying time-varying weights to discrete system.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

This dissertation is primarily about the application of data assimilation to various models, including the Lorenz equations. Data assimilation, as its name suggests, involves using observed data to find a solution to some model. The model usually consists of a set of differential equations whose general solution portrays the prevalent dynamics of the system being studied. The data is used to identify initial values that define a particular solution, in which comparable solution values and observed data are in some sense consistent to one another.

From diverse fields in engineering to problems in meteorology and oceanography, the practical importance of developing data assimilation techniques cannot be overemphasised. This is particularly true in areas such as oil recovery, meteorology and oceanography where there is insufficient data to find the the exact true behavior of a given model. The data assimilation problem has created new and interesting challenges for the applied mathematician.

A few of thse challenges can be identified by considering the applications and difficulties of the data assimilation in meteorology and oceanography. In meteorology the most pressing concern is to forecast - to predict future events from data/observations in the past. The model differential equations which describe the dynamics of the atmosphere are adequately representative of the dynamics that occur. The aim is to use past data in such a way as to cover the inadequacies in the differential model equations caused by having insuficient data to acccurately specify the initial conditions.

For oceanography, the general emphasis in usage is different. Forecasting though as ultimate aim is only presently being used to predict seasonal and interannual time scales in the tropical ocean and for short range forecasting for navies, fisheries and off-shore drilling. Unfortunately at present the forecasts are quite inaccurate unlike those in meteorology. The model equations do not adequately represent the true dynamics of the system. Thus data assimilation is used as a techinque to determine the specific inadequacies of the model equations so that they can be rectified in the future by the consideration of additional parameters.

The techniques used in meteorology and oceanography will be different.(Ghil et al.1991) This is due to the inherent differences in the problems considered. There is far less data availiable in oceanography, and the data by itself is insufficient to model the complex changes that occur. The current aim is to optimise the usage of the data so that the model equations can be improved upon. Added complexity arises from the fact that many of the variables demanded by the model equations can only be indirectly measured. For example in oceanography velocity components are needed. These have to be calculated from other variables such as surface height or wind stress. In many cases the relationship between direct and indirectly observed variables is not precisely known. Thus different techiniques need to be used in both meteorology and oceanography and techniques used in meteorology will need to be adapted and reinterpreted to be used in oceanography.

The aim of this chapter is to describe the data assimilation problem and briefly show how the method used to solve this problem in this dissertation relates to other approaches in the field. This will not only reveal the variety of approaches that exist but also the general terminology used within the subject area. This is followed by a brief synopsis of what is being covered in this dissertation.

1.1 The Data Assimilation Problem

1.1.1 The continuous problem

The continuous data assimilation problem is defined as

Problem 1 *Minimise cost functional*

$$\mathcal{J} = \int_{t_0}^{t_f} \frac{1}{2}(\tilde{\boldsymbol{x}} - C\boldsymbol{x})^T D(\tilde{\boldsymbol{x}} - C\boldsymbol{x})dt, \tag{1.1}$$

subject to the constraints

$$\boldsymbol{g}(\boldsymbol{x}, t) - \dot{\boldsymbol{x}}(t) = 0, \tag{1.2}$$

throughout the time interval

$$t \in [t_0, t_f] \tag{1.3}$$

where $\tilde{\boldsymbol{x}}(t)$ *is a* $(n \times 1)$ *real* C^2 *vector function of observations and* $\boldsymbol{x}(t)$ *is a* $(m \times 1)$ *real* C^2 *vector function satisfying the model equations (1.2).*

The $(n \times m)$ matrix C relates observations to the solutions of the model equation. The variables used to observe the data may well be different to those described by the model equations.

The $(n \times n)$ matrix D is assumed to be symmetric positive definite. The symmetric positive definiteness guarantees that the cost functional $\mathcal{J}$ has a unique value. It is used as a weighting function so that more credulence can be placed on some observational parameters than on others.

Problem (1) can be considered equivalent to finding an appropriate value of α so that the initial value $\boldsymbol{x}(t_0) = \alpha$ determines a solution to the model equations (1.2) that minimises the cost functional $\mathcal{J}$ in (1.1) throughout the time interval $t \in [t_0, t_f]$.

1.1.2 The discrete problem

The discrete data assimilation problem is

Problem 2 *Minimise*

$$\mathcal{J} = \sum_{i=0}^{i=N-1} \frac{1}{2}(\tilde{\boldsymbol{x}}_i - C\boldsymbol{x}_i)^T D(\tilde{\boldsymbol{x}}_i - C\boldsymbol{x}_i)\Delta t \tag{1.4}$$

subject to the constraints

$$\boldsymbol{g}(\boldsymbol{x}_i) - \boldsymbol{x}_{i+1} = 0 \tag{1.5}$$

for $i = 0, 1...N-1$

where $\boldsymbol{x}_i$ *is an* $(m \times 1)$ *real vector representing the model solution at* t_i *and* $\boldsymbol{g}$ *is a nonlinear vector function which gives the value of the model solutions at the next time step,* $\boldsymbol{x}_{i+1}$ *at* t_{i+1}.

The terms used in the discrete problem $\tilde{\boldsymbol{x}}, \boldsymbol{x}, C, D$ are similar to those in the continuous one. There is also a measure of similarity between the two problems. As in the continuous problem, the discrete problem can be considered in a different way. It is equivalent to finding an appropriate value of α so that the initial value $\boldsymbol{x}_0 = \alpha$ determines a solution to the difference equations (1.5) that minimises the objective function $\mathcal{J}$ in (1.4). In Chapter 2 further similarities between the discrete and continuous problems will be shown.

A variational approach will be used thoughout this dissertation. This involves minimising an objective function $\mathcal{J}$ subject to satifying the model equations exactly. The objective function/cost functional is a measure of the discrepancy between the data and the solution considered. The precise details of how this are done is left to Chapter 2.

Satisfying the model equations exactly while performing the minimisation on $\mathcal{J}$ is called the strong constraint approach (Sasaki, 1969). One can incorporate error terms in the model equations (Griffith and Nichols, 1996). This allows greater freedom to be introduced into the solution. This is consequently called the weak constraint approach.

1.2 Other Approaches

Although a variational approach is being used, one can also rephrase the problem in a number of ways to accomodate differing subject areas such as probability theory and control theory. This all adds to the richness of the topic.

(i). Other than the variational approach there are two general approaches which have been popular. The simplest of these is to insert observational data values into the model and then use some smoothing method to obtain a solution (analysis) possibly through the use of polynomial splines. Combining some prior estimate with actual observational data in some way that is consistent with the model equations to form a new estimate is called 'the analysis' (Cressman, 1959). This is the main stage in the data assimilation process.

A purely statistical approach can be used. One such technique is called Optimal Interpolation (Lorenc, 1986), (Ghil et al., 1991). It presumes that the prior estimate values and the observational data are both inaccurate to certain tolerances. The tolerances can be described by a measure called the variance which accounts for how the probability varies around the mean data value. Through the use of Baysean statistics the 'likeliest result' is found which is consistent with both data and model values but also with their variances. In other words the analysis values are those which are consistent with model and data values and have the smallest variance possible.

The problem with the former approach is that it is too simplistic and may place too much weight on the observed data. The later approach is a good one. However it does presume that the variances for each variable are known which is not always true.

(ii). In this dissertation the minimisation between model and data values occurs simultaneously across the whole assimilation time period. The consequence is that all the data within the time period iteracts with itself making the analysis at any time, dependent upon both past and future information within the assimilation time interval. Any scheme which exhibits this property is considered to be 4-D (Talagrand, 1991). A 4-D scheme which is both variational and statistical in nature is at present being assembled by the Meteorological Office to produce improved forecasts.

In contrast, schemes in which comparisons between the model and data values occur independently at specific points in time are called 3-D. Whether a 4-D or a 3-D scheme is used depends on the problem considered.

1.3 Summary of Thesis

Chapter 2 covers the necessary mathematical background needed. It details the necessary conditions for a minimum. The solution of adjoint equations are shown to satisfy the necessary conditions for a miminum under a constraint. This is applied to the case where the observational data and model equations are continuous as well as discrete. The conditions are written in such a form that they readily adapt to systems of equations.

Chapter 3 explains the minimisation procedure . This is followed by a number of numerical experiments applied to a scalar model and a linear system of two coupled first order differential equations. In both cases the numerical procedure used to solve the differential equations is the modified Euler scheme . A few insights are given into the properties of the minimisation procedure.

Chapter 4 gives a brief discussion of the Lorenz equations. Data assimilation is performed upon this system using the same numerical scheme as before. Amongst other results, the application of time-varying weights to the system allows an increase in the size of the time interval over which data assimilation is successful.

Chapter 2

MATHEMATICAL BACKGROUND

2.1 Overview of Chapter

The aim of this chapter is to develop sufficient mathematical theory to be able to adequately derive the necessary conditions needed to solve the data assimilation problems described in Chapter 1.

It is instructive to start from a broader perspective and to consider minima in general. Minimisation subject to satisfying some model equations exactly can be done through the well known method of Lagrange multipliers. This effectively transforms the problem from one that is explicitly constrained by a set of model equations to an associated easier problem in which the model equations are intrinsically incorporated into the objective function which is unconstrained. From this associated problem, the necessary conditions needed for a minimum can be obtained for both the general discrete and continuous problems in the consequent derivation of the Euler-Lagrange Equations. In presenting both the discrete and continuous problems together, the similarities between the two problems will be revealed.

The general results are readily applied to the two data assimilation problems. Three examples are presented: a scalar continuous example, a scalar discrete example and a second order discrete example. The first two examples show that discretising the necessary condition known as the adjoint equation in the continuous example is not in general the same as the equivalent condition in the discrete example. However as the size of timestep in both disretisations tend to zero, the two discretisations of the adjoint converge to the same result. The discrete examples are used in an examination of the numerical optimisation procedure presented in Chapter 3.

2.2 Preliminaries

The data assimilation problem as described in Chapter 1 has been formulated as a variational problem. General continuous and discrete variational problems are stated below.

A general continuous problem can be described as

Problem 3 *Minimise the cost functional*

$$\mathcal{K} = \int_{t_0}^{t_f} \mathcal{F}(\boldsymbol{x}, t) \tag{2.1}$$

subject to the constraint equations

$$\dot{\boldsymbol{x}} = \boldsymbol{g}(\boldsymbol{x}, t) \tag{2.2}$$

where $\boldsymbol{g}$ *is a vector of* n *continously differentiable real functions from* $\Re^n \times [t_0, t_f]$ *to* $\Re$ *assuming* $\nabla_{\boldsymbol{x}} g(\boldsymbol{x}, t)$ *is linearly independent for all* $(\boldsymbol{x}, t)$ *belonging to* $\Re^n \times [t_0, t_f]$.

Similarly, a general discrete problem can be described as

Problem 4 *Minimise the objective function*

$$\mathcal{K} = \sum_{i=0}^{i=N-1} \mathcal{F}(\boldsymbol{x}_i, t_i) \tag{2.3}$$

subject to the constraints

$$\boldsymbol{x}_{i+1} = \boldsymbol{g}(\boldsymbol{x}_i, t_i) \tag{2.4}$$

It makes sense to now describe the two general variational problems and the associated terminology.

The definition for a local optimum of $\mathcal{K}$ folllows.

Definition 1 *The cost functional/objective function $\mathcal{K}$ has a local optimum at u^* if there exists $\epsilon > 0$ such that $\mathcal{K}(u) - \mathcal{K}(u^*)$ is of just one sign for all u such that $||u - u^*|| < \epsilon$.*

The *first variation* of $\mathcal{K}$ at u^*, written as $\delta\mathcal{K}(u^*, \delta u)$ can be expressed as

$$\delta\mathcal{K} = \mathcal{K}(u^* + \delta u) - \mathcal{K}(u^*) \tag{2.5}$$

where δu is some small perturbation in direction δu with magnitude $||\delta u||$.

For the general nonlinear function $\mathcal{K}$ to have a local optimum at u^*, the necessary condition is that $\delta\mathcal{K} = 0 \;\forall\; \delta u$ such that $||\delta u|| < \epsilon$. A curve which satisfies the necessary conditions in the continuous case is called an *extremal*. The same term is given to a sequence of values from $i = 0$ to $i = N - 1$ which satisfy the necessary conditions in the discrete case. To guarantee that a extremal gives a local minimum, the Hessian of $\mathcal{K}$ needs to be positive definite for all pertubations δu) in the local neighbourhood of u^*. If the Hessian is positive semidefinite no conclusions can be reached, while if it is indefinite, the minimum does not exist at u^*.

2.2.1 Minimisation with/without constraints

Problems 3, 4 are constrained minimisation problems and can be solved by converting them into unconstrained problems. This can be done using the method of Lagrangian multipliers. As mentioned in the introduction to this chapter, this involves dealing with another problem where the constraints are a part of the cost functional/objective function $\mathcal{L}$. As the constraints are no longer explicitly defined, the new problems are unconstrained and consequently are far easier to solve, numerically and analytically.

The alternate general unconstrained continuous problem corresponding to Problem 3 is

Problem 5 *Find an initial value $\boldsymbol{x}(0)$ to produce a solution to the model equation (2.2) and a continuous function $\boldsymbol{l}(t)$ that minimises*

$$\mathcal{L}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{l}(t), t) = \mathcal{K} + \int_{t_0}^{t_f} \boldsymbol{l}(t)^T(\boldsymbol{g}(\boldsymbol{x}(t), t) - \dot{\boldsymbol{x}}(t))\, dt \tag{2.6}$$

$$= \int_{t_0}^{t_f} \mathcal{H}(\boldsymbol{x}, \dot{\boldsymbol{x}}(t), \boldsymbol{l}(t), t), \tag{2.7}$$

$\boldsymbol{l}$ is a vector belonging to $\Re^n$ whose components are Lagrangian multipliers l_k, $k = 1, \ldots, n$.

Similarly, the general discrete problem correponding to Problem 4 is

Problem 6 *Find an initial value $\boldsymbol{x}_0$ to produce a solution to the model equations (2.4) and a set of parameters $\boldsymbol{l}_i$ that minimise the cost functional*

$$\mathcal{L}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \boldsymbol{l}_{i+1}, t_i) = \mathcal{K} + \sum_{i=0}^{i=N-1} \boldsymbol{l}_{i+1}^T(\boldsymbol{g}(\boldsymbol{x}_i, t_i) - \boldsymbol{x}_{i+1}) \tag{2.8}$$

$$= \sum_{i=0}^{i=N-1} \mathcal{H}(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \boldsymbol{l}_{i+1}, t_i), \tag{2.9}$$

$\boldsymbol{l}_{i+1}$ being a vector whose components are Lagrangian multipliers $l_{k(i+1)}$, $k = 1, \ldots, n$.

It is possible to define Problems 5, 6 in the way described above, because it is the initial value $\boldsymbol{x}_0$ which determines the solution to the model equations and as such is the only degree of freedom available.

2.3 Necessary Conditions for a Minimum

With the problem transformed into a more acceptable form, it is now possible to find the necessary conditions for a minimum. These can be obtained by finding the appropriate Euler-Lagrange equations for the continuous and discrete problems.

2.3.1 The continuous problem

With $\mathcal{L}$ described as in (2.7) the first variation of $\mathcal{L}$ is

$$\delta\mathcal{L} = \int_{t_0}^{t_f} \frac{\partial\mathcal{H}}{\partial\boldsymbol{x}}\delta\boldsymbol{x} + \frac{\partial\mathcal{H}}{\partial\dot{\boldsymbol{x}}}\delta\dot{\boldsymbol{x}}\; dt \tag{2.10}$$

On application of integration by parts to the second term $\delta\mathcal{L}$ becomes

$$\delta\mathcal{L} = \int_{t_0}^{t_f} \frac{\partial\mathcal{H}}{\partial\boldsymbol{x}}\delta\boldsymbol{x}\; dt + \left[\frac{\partial\mathcal{H}}{\partial\dot{\boldsymbol{x}}}\delta\boldsymbol{x}\right]_{t_0}^{t_f} - \int_{t_0}^{t_f} \frac{d}{dt}\left(\frac{\partial\mathcal{H}}{\partial\dot{\boldsymbol{x}}}\right)\delta\boldsymbol{x}\; dt \tag{2.11}$$

Thus applying the Fundamental Lemma (Pars, 1962) shows that the necessary conditions for $\delta\mathcal{L}= 0$ are

(i).

$$\left[\frac{\partial\mathcal{H}}{\partial\dot{\boldsymbol{x}}}\delta\boldsymbol{x}\right]_{t_0}^{t_f} = 0; \tag{2.12}$$

(ii).

$$\frac{\partial\mathcal{H}}{\partial\boldsymbol{x}} - \frac{d}{dt}\left(\frac{\partial\mathcal{H}}{\partial\dot{\boldsymbol{x}}}\right) = 0. \tag{2.13}$$

Applying these conditions to (2.6) gives (i).

$$\left[\boldsymbol{l}^T \delta \boldsymbol{x} \right]_{t_0}^{t_f} = 0; \tag{2.14}$$

(ii).

$$\frac{\partial \mathcal{K}}{\partial \boldsymbol{x}} + \boldsymbol{l}^T \frac{\partial g}{\partial \boldsymbol{x}} + \left(\frac{d\boldsymbol{l}}{dt} \right)^T = 0. \tag{2.15}$$

REMARKS

(i). If the end points are fixed i.e $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ and $\boldsymbol{x}(t_f) = \boldsymbol{x}_f$ then $\delta \boldsymbol{x}(t_0) = \delta \boldsymbol{x}(t_f) = 0$, satisfies (2.12). Where a fixed end point is not present, a transversality condition is used. i.e.

$$\frac{\partial \mathcal{H}}{\partial \dot{\boldsymbol{x}}} = 0; \tag{2.16}$$

(ii). (2.15) is known as the adjoint equation. The adjoint together with the model equation form the Euler Lagrange equations.

2.3.2 The discrete problem

Using $\mathcal{L}$ as described in problem (2.7), the first variation is

$$\delta \mathcal{L} = \sum_{i=0}^{i=N-1} \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_i} \right)_i \delta \boldsymbol{x}_i + \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_{i+1}} \right)_{i+1} \delta \boldsymbol{x}_{i+1}. \tag{2.17}$$

Rearranging this equation gives

$$\delta \mathcal{L} = \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_0} \right)_0 \delta \boldsymbol{x}_0 + \sum_{i=1}^{i=N-1} \left[\left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_i} \right)_i \delta \boldsymbol{x}_i + \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_{i+1}} \right)_i \delta \boldsymbol{x}_i \right] + \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_N} \right)_N \delta \boldsymbol{x}_N. \tag{2.18}$$

Hence, a set of necessary conditions for $\delta \mathcal{L} = 0$ are:

(i).

$$\left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_0} \right)_0 \delta \boldsymbol{x}_0 = 0; \tag{2.19}$$

(ii).

$$\left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_N} \right)_N \delta \boldsymbol{x}_N = 0; \tag{2.20}$$

(iii). for $i = 1, \dots N-1$,

$$\left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_i}\right)_i + \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}_{i+1}}\right)_i = 0. \tag{2.21}$$

Using (2.7) these necessary conditions can be equivalently expressed as

(i).

$$\left(\left(\frac{\partial \mathcal{K}}{\partial \boldsymbol{x}_0}\right)_0 + \boldsymbol{l}_1^T \left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}_0}\right)_0\right) \boldsymbol{\delta x}_0 = 0, \tag{2.22}$$

(ii).

$$\boldsymbol{l}_N^T \boldsymbol{\delta x}_N = 0, \tag{2.23}$$

(iii). and for $i = N-1, \dots 1$,

$$\left(\frac{\partial \mathcal{K}}{\partial \boldsymbol{x}_i}\right)_i + \boldsymbol{l}_{i+1}^T \left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}_i}\right)_i - \boldsymbol{l}_i^T = 0. \tag{2.24}$$

2.4 Application to Data Assimilation Problems

2.4.1 Continuous data assimilation problem

Application of the theory for the general problem, Problem 3, to the data assimilation problem, Problem 1, produces

$$\mathcal{K}(\boldsymbol{x}(\boldsymbol{t}), t) = \frac{1}{2} \int_{t_0}^{t_f} (\tilde{\boldsymbol{x}} - C\boldsymbol{x})^T D (\tilde{\boldsymbol{x}} - C\boldsymbol{x}) \; dt. \tag{2.25}$$

On expanding, assuming D is symmetric, we find

$$\mathcal{K} = \frac{1}{2} \int_{t_0}^{t_f} \tilde{\boldsymbol{x}}^T D \tilde{\boldsymbol{x}} - 2\tilde{\boldsymbol{x}}^T D C \boldsymbol{x} + \boldsymbol{x}^T C^T D C \boldsymbol{x} \; dt. \tag{2.26}$$

Hence applying (2.15) to $\mathcal{K}$ and $\mathcal{L}$ gives the appropriate Euler-Lagrange Equations for Problem 1 as

$$-\frac{d\boldsymbol{l}^T}{dt} = -\tilde{\boldsymbol{x}}^T D C + \boldsymbol{x}^T C^T D C + \boldsymbol{l}^T \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}(\boldsymbol{x}(t), t), \tag{2.27}$$

or more clearly by taking the transpose,

$$-\frac{d\boldsymbol{l}}{dt} = -C^T D\tilde{\boldsymbol{x}} + C^T DC\boldsymbol{x} + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}^T(\boldsymbol{x}(t), t)\boldsymbol{l}. \tag{2.28}$$

Finally, the necessary conditions for an extremal to the data assimilation problem are;

(i). equation (2.28),

(ii).

$$\boldsymbol{l}(t_0) = \boldsymbol{l}(t_f) = 0, \tag{2.29}$$

from the transversality conditions (2.14), as our end points are not fixed;

(iii). the model equations

$$\boldsymbol{g}(\boldsymbol{x}(t), t) - \dot{\boldsymbol{x}} = 0. \tag{2.30}$$

2.4.2 Discrete data assimilation problem

Similar application of the theory for the general Problem 6 to the discrete data assimilation problem, Problem 2, gives

$$\mathcal{K}(\boldsymbol{x}_i, t_i) = \frac{1}{2}\sum_{i=0}^{i=N-1}(\tilde{\boldsymbol{x}}_i - C\boldsymbol{x}_i)^T D(\tilde{\boldsymbol{x}}_i - C\boldsymbol{x}_i). \tag{2.31}$$

On expanding,

$$\mathcal{K} = \sum_{i=0}^{i=N-1} \frac{1}{2}\tilde{\boldsymbol{x}}_i^T D\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_i^T DC\boldsymbol{x}_i + \frac{1}{2}\boldsymbol{x}_i^T C^T DC\boldsymbol{x}_i. \tag{2.32}$$

Substituting $\mathcal{K}$ in Problem 6 and using (2.24) gives the Euler-Lagrange Equations for the discrete data assimilation problem as

$$-\tilde{\boldsymbol{x}}_i^T DC + \boldsymbol{x}_i^T C^T DC + \boldsymbol{l}_{i+1}^T\left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}_i}\right)_i - \boldsymbol{l}_i^T = 0. \tag{2.33}$$

for $i = N-1, \ldots, 1$. Rearranging and taking the transpose gives

$$\boldsymbol{l}_i = -C^T D\tilde{\boldsymbol{x}}_i + C^T DC\boldsymbol{x}_i + \left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}_i}\right)_i^T \boldsymbol{l}_{i+1}, \tag{2.34}$$

for $i = N-1, \ldots, 1$. The respective boundary conditions with transversality conditions applied are

(i).

$$\boldsymbol{l}_0 \equiv C^T D \tilde{\boldsymbol{x}}_0 + C^T D C \boldsymbol{x}_0 + \left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}_0} \right)_0^T \boldsymbol{l}_1 = 0; \tag{2.35}$$

(ii).

$$\boldsymbol{l}_N = 0. \tag{2.36}$$

Thus, the necessary conditions for an extremal to the discrete data assimilation problem are

(i). 2.34;

(ii). the boundary conditions (2.35),(2.36);

(iii). and the model difference equations

$$\boldsymbol{g}(\boldsymbol{x}_i, t) - \boldsymbol{x}_{i+1} = 0. \tag{2.37}$$

2.5 Examples

2.5.1 Continuous scalar example

Consider the first order nonlinear equation

$$\dot{x}(t) = x^2(t) \tag{2.38}$$

$$x(t_0) = \alpha, \tag{2.39}$$

where $x \in \Re$, $t \in [t_0, t_f]$ and $\tilde{x}(t)$ is some continuous function, with $\tilde{x} \in \Re$, which we can consider to be the true solution. Let the matrices $C = 1$ and $D = 2$ and let $\mathcal{J}$ be given by

$$\mathcal{J} = \int_{t_0}^{t_f} (x(t) - \tilde{x}(t))^2 dt \tag{2.40}$$

Then the data assimilation problem reduces to finding the initial value α such that the model solution to (2.38) performs a least square fit to the observations $\tilde{x}$. Such a least squares fit minimises $\mathcal{J}$.

The associated augmented functional is

$$\mathcal{L} = \int_{t_0}^{t_f} (x(t) - \tilde{x}(t))^2 + l(t)(x^2(t) - \dot{x}(t))dt \tag{2.41}$$

The first variation is

$$\delta\mathcal{L} = \int_{t_0}^{t_f} (2(x(t) - \tilde{x}) + 2l(t)x(t))\delta x(t) - l(t)\delta\dot{x}(t)\ dt. \tag{2.42}$$

Integrating by parts, rearranging and setting $\delta\mathcal{L} = 0$ produces the correponding adjoint equation

$$-\dot{l} = 2(x(t) - \tilde{x}(t)) + 2l(t)x(t) \tag{2.43}$$

with boundary conditions

$$l(t_0) = l(t_f) = 0. \tag{2.44}$$

The above three equations (2.38), (2.43), (2.44) give the necessary conditions for the optimal.

2.5.2 Discrete scalar example

The constraint to be used in the subsequent minimisation is obtained by transforming the differential equation (2.38) into a difference equation using the modified Euler scheme.

The modified Euler scheme is a second order Runga-Kutta method used to solve the differential equation $\dot{x} = g(x, t)$. It is defined by:

For

$$t_i = t_0 + i\Delta t, \qquad i = 1, \ldots, N-1, \tag{2.45}$$

where Δt is the size of the timestep and

$$x_0 = x(t_0), \tag{2.46}$$

successive values x_i which approximate the values $x(t_i)$ are obtained from

$$\begin{aligned} k1 &= \Delta t g(x_i, t_i), \\ k2 &= \Delta t g(x_i + k1, t_i + \Delta t), \\ x_{i+1} &= x_i + \frac{1}{2}(k1 + k2). \end{aligned} \tag{2.47}$$

The model equation $x_{i+1} = g(x_i, t_i) = x_i^2$ produces the difference scheme.

$$x_{i+1} = x_i + \Delta t x_i^2 + (\Delta t)^2 x_i^3 + \frac{(\Delta t)^3 x_i^4}{2}. \tag{2.48}$$

Let the observed data points $\tilde{x}_j$ be known at points where model values are defined and let there be a direct comparison between the two. Matrix C then is just the identity matrix. Let matrix $D = 2I$.

Problem 7 *The control problem is to find $\alpha = x_0$ to minimise*

$$\mathcal{J} = \sum_{i=0}^{i=N-1} (x_i - \tilde{x}_i)^2 \Delta t \tag{2.49}$$

subject to the constraint 2.48.

The associated augmented function $\mathcal{L}$ is

$$\mathcal{L} = \sum_{i=0}^{i=N-1} (x_i - \tilde{x}_i)^2 \Delta t - l_{i+1}(x_{i+1} - x_i - \Delta t x_i^2 - (\Delta t)^2 x_i^3 - \frac{(\Delta t)^3 x_i^4}{2}) \tag{2.50}$$

giving the first variation $\delta\mathcal{L}$ as

$$\delta\mathcal{L} = \sum_{i=0}^{i=N-1} 2(x_i - \tilde{x}_i)\Delta t \delta x_i + l_{i+1}(1 + 2\Delta t x_i + 3(\Delta t)^2 x_i^2 + 2(\Delta t)^3 x_i^3)\delta x_i - l_{i+1}\delta x_{i+1} \tag{2.51}$$

which needs to be zero, that is,

$$
\begin{aligned}
0 &= (2(x_0 - \tilde{x}_0)\Delta t + l_1(1 + 2\Delta t x_0 + 3(\Delta t)^2 x_0^2 + 2(\Delta t)^3 x_0^3))\delta x_0 \\
&+ \sum_{i=1}^{i=N-1} 2(x_i - \tilde{x}_i)\Delta t \delta x_i + l_{i+1}(1 + 2\Delta t x_i + 3(\Delta t)^2 x_i^2 + 2(\Delta t)^3 x_i^3) - l_i)\delta x_i \\
&+ l_N \delta x_N
\end{aligned}
\tag{2.52}
$$

To satisfy (2.52) we enforce

(i). (2.48) for $i = 0 \ldots N-1$;

(ii). for $i = N-1, \ldots, 0$

$$
l_i = 2(x_i - \tilde{x}_i)\Delta t + l_{i+1}(1 + 2\Delta t x_i + 3(\Delta t)^2 x_i^2 + 2(\Delta t)^3 x_i^3); \tag{2.53}
$$

(iii). and

$$
l_0 = l_N = 0. \tag{2.54}
$$

to give the necessary conditions for an optimal solution.

REMARK

The adjoint equation derived from the discrete scalar example (2.53) is not the same as the corresponding adjoint equation derived by applying the Modified Euler scheme to the continuous scalar example. However as $\Delta t \to 0$ both difference schemes tend towards the same result.

2.5.3 Discrete second order system

Although the Euler-Lagrange equations have already been defined for the general discrete data assimilation problem, it is instructive to look at an example that is a system.

Applying the Modified Euler scheme to discretise the second order linearised system.

$$\begin{aligned}\dot{x} &= 2x + 5y \\ \dot{y} &= x - 2y, \qquad\qquad t \in [0, t_f],\end{aligned} \tag{2.55}$$

produces

$$x_{i+1} = x_i + \Delta t(2x_i + 5y_i) + \frac{9\Delta t^2 x_i}{2} \tag{2.56}$$

$$y_{i+1} = y_i + \Delta t(x_i - 2y_i) + \frac{9\Delta t^2 y_i}{2} \tag{2.57}$$

for $i = 0, \ldots, j-1$.

These two coupled difference equations give the constraints

$$\boldsymbol{g}(\boldsymbol{x}_i) = \begin{pmatrix} g_1(\boldsymbol{x}_i) \\ g_2(\boldsymbol{x}_i) \end{pmatrix} \tag{2.58}$$

for the objective function

$$\mathcal{J} = \sum_{i=0}^{i=N-1} ((x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2)\Delta t, \tag{2.59}$$

which needs to be minimised. The associated Lagrange function is

$$\begin{aligned}\mathcal{L} = & \sum_{i=0}^{i=N-1} ((x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2)\Delta t + \\ & lx_{i+1}(g_1(\boldsymbol{x}_i) - x_{i+1}) + ly_{i+1}(g_2(\boldsymbol{x}_i) - y_{i+1})\end{aligned} \tag{2.60}$$

where lx_i and ly_i are both Lagrange multipliers applied to the i^{th} point in the series. Taking the first variation, setting it to zero and rearranging gives a coupled system of adjoint equations,

$$\begin{aligned}lx_i &= 2\Delta t(x_i - \tilde{x}_i) + lx_{i+1}(1 + 2\Delta t + 4.5\Delta t^2) + ly_{i+1}\Delta t \\ ly_i &= 2\Delta t(y_i - \tilde{y}_i) + ly_{i+1}(1 - 2\Delta t + 4.5\Delta t^2) + 5lx_{i+1}\Delta t \\ & \quad for\ i = N-1, \ldots, 0;\end{aligned} \tag{2.61}$$

which with the boundary conditions,

$$lx_0 = ly_0 = lx_N = ly_N = 0 \tag{2.62}$$

and the model equations (2.56), (2.57) give the necessary conditions for a minimum.

Now that the necessary background has been covered, the numerical minimisation procedure can be explained using the discrete scalar example and the discrete second order system as examples, as shown in Chapter 3.

Chapter 3

NUMERICAL EXPERIMENTS

3.1 Outline

The purpose of this chapter is twofold: to describe the general numerical minimisation procedure used in the dissertation and to examine aspects of the data assimialtion process. The general numerical minimisation procedure is restricted to dealing with the discrete case. Two examples from Chapter 2 will be looked at: the discrete scalar example and a second order linear system. Validation tests will show that both discretised model equations and minimisation process converge. The accuracy of a number of measures of performance are compared as well as conditons where the procedure fails.

3.2 Numerical Minimisation Procedure

The overall aim of the minimisation procedure is for all the necessary conditions for an optimal solution to the augmented functional to be satisfied. This needs the calculation of the adjoint equations as shown in Chapter 2, the calculation of model values, as well as enforcing the boundary conditions of the adjoint. An algorithm is considered in which successive approximations to the initial value describing the model solution are produced such that corresponding successive approximations to $\boldsymbol{l}_0$ tend to zero.

The procedure can be described as follows:

(i). Start with an initial guess to the initial value to the model difference equation.

(ii). Use the model difference equation to obtain the rest of the model values.

(iii). These model values are incorporated into the adjoint difference equations. The adjoint values are gathered starting with $\boldsymbol{l}_N = 0$, by iterating backwards until a value for $\boldsymbol{l}_0$ is obtained. $\boldsymbol{l}_0$ is the gradient of cost functional $\mathcal{J}$.

(iv). The gradient $\boldsymbol{l}_0$ is used in a minimisation procedure to produce a better estimate for the 'true' initial value to the model equation.

(v). The procedure is repeated using the new estimated initial value until some stopping criteria is met.

3.2.1 Description of minimisation procedure used

The minimisation procedure is the steepest descent method. It is a local method. It is an iteration procedure of the form

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k - s\nabla(\boldsymbol{z}^k) / \left\| \nabla(\boldsymbol{z}^k) \right\| \tag{3.1}$$

where

(i). $\boldsymbol{z}^k$ is the n-th iterate of the initial conditions;

(ii). the negative gradient $-\boldsymbol{l}_0^k \equiv -\nabla(\boldsymbol{z}^k)$ is the descent direction;

(iii). s is the step length.

The gradient is obtained from the initial value of the adjoint equation. This gradient is normalised under the L_2 norm to have unit length. This is done because the length of this vector has no relation to the discrepancy between the computed k-th iterate of the initial conditions and its optimal value.

In most cases the starting value for the step length s in (3.1) was 0.5. This value is used provided that the new estimate z^{k+1} under some measure of performance is better than previous estimates. Otherwise the new estimate is ignored. Instead a new value for z^{k+1} is obtained from (3.1) by halving the steplength . A measure of performance for this value is gathered, and the new value of s is used so long as the new estimate is better than z^k. Otherwise the steplength s is repeatedly halved until a better estimate is obtained. In subsequent iterations this new value for s is used.

3.2.2 Measures of performance, stopping criteria

There are a number of measures of performance. These can be used as stopping criteria for the minimisation procedure. A list of all the types of stopping criteria used in this dissertation are given below. They are;

(i).

$$\left\| l_0^k \right\| < tol1, \tag{3.2}$$

(ii).

$$\left\| l_0^k - l_0^{k-1} \right\| < tol2, \tag{3.3}$$

(iii).

$$\mathcal{J}^k < tol3, \tag{3.4}$$

(iv).

$$\frac{\mathcal{J}^{k+1} - \mathcal{J}^k}{\mathcal{J}^k} < tol4, \tag{3.5}$$

(v).

$$\mathcal{J}^{k+1} - \mathcal{J}^{k} < tol5, \tag{3.6}$$

(vi).

$$\left\| \boldsymbol{x}_0 - \boldsymbol{x}_0^k \right\| < tol6, \tag{3.7}$$

(vii).

$$\left\| \boldsymbol{x}_0^{k+1} - \boldsymbol{x}_0^k \right\| < tol7, \tag{3.8}$$

3.2.3 Variants on the procedure

A number of variants to the general steepest descent method were used. These primarily varied in how the steplength was obtained at each iteration step in the minimisation procedure.

In dealing with second/third order systems, the steplength s can become so small, as to make improvements in $\boldsymbol{z}^{k+1}$ negligible even when $\boldsymbol{z}^{k+1}$ is quite a distance away from the optimum. One way to get around this is to let s increase by a factor (say 3.0) at every iterate, provided improvement is made. Another way is to find an optimum steplength at each minimisation step. By starting each time with a steplength of 0.5 and repeatedly halving this value until an improved estimate to initial value is found.

The first suggestion uses less computational effort at each step in the iteration, but may require more iterations. The second suggestion may find the least number of minimisation steps needed but needs a lot of effort in doing each step. It is likely that the optimum step size s at the $k+1^{th}$ iteration is going to be of the same order of magnitude as at the k^{th} iteration, where linear systems are used. In these cases the first suggestion is used. Otherwise the second one is applied.

3.3 The Scalar Model

The first order ODE, as mentioned in the overview, is the same as in Chapter 2, *i.e.*

$$\dot{x}(t) = x^2(t), \qquad t \in [1, t_f]. \tag{3.9}$$

This equation is discretised using the Modified Euler Scheme

$$x_{i+1} = x_i + \Delta t x_i^2 + \Delta t^2 x_i^3 + \frac{\Delta t^3 x_i^4}{2}, \tag{3.10}$$

for $i = 0 \ldots j-1$

As explained previously, the aim of the numerical minimisation procedure is to achieve the necessary conditions needed for a minimum to the discrete data assimilation problem. An iterative method is used where all the necessary conditions are satisified except for, $l_0 = 0$. If the observations are produced by (3.10) and with the same time-step as the solution to the data assimialtion problem, then provided the value of l_0 converges to zero as the iteration proceeds, a good approximation to the minimum will be guaranteed.

Thus we satisfy;

(i). (3.10);

(ii). the respective adjoint equations (2.53);

(iii).

$$l_N = 0, \tag{3.11}$$

stepping (3.10) forward in time and stepping with the adjoint equation (2.53) backward in time from the end condition (3.11) in order to eventually get an approximation for $l(0)$. This value of $l(0)$ is used to provide a better guess for the initial value to the model equation.

error in modeuler for $\dot{x} = x^2$ at t=1.5 (I.C x(1)=1)	Number of timesteps to reach t=1.5	order of magnitude improvement in reduction of error
1.8568527058660D-03	16	
4.7651153062178D-04	32	3.90
1.2061845894151D-04	64	3.95
3.0337398903013D-05	128	3.97
7.6069564751702D-06	256	3.99

Table 3.1: showing (3.9) is second order accurate

3.3.1 Validation tests

We need to show that the Modified Euler scheme applied to this first order nonlinear differential equation is second order accurate. This can be shown by comparing the global error that arises when a varied number of timesteps are used to cover a given time window.

In Table 3.1 the initial value of the difference scheme starts at $t = 1$. The error at $t = 1.5$ is measured for differing numbers of timesteps within the time interval $t \in [1, 1.5]$.

Doubling the the number of timesteps has the effect of diminishing the error by four, especially where the number of timesteps is large. This verifies that the Modified Euler scheme applied to (3.9) is second order accurate.

Ideally, before the data asssimialtion techinque is completely valided the optimisation procedure needs to be checked. The observations were obtained using (3.10), starting from a 'true' initial value with the same size of time-step as in the data assimilation process. Where the solution to the data assimilation process can perfectly match the observations, the initial value of the adjoint $l(0)$ has to become zero for a minimum to be present. Thus subscribing $l(0) < tol1$ can act as a good measure for the accuracy prescribed by the minimisation procedure.

The tolerance placed on the stopping criterion $l(0)$ does provide a rough measure of the final accuracy in the initial conditions. The results are generally accurate to the same number of decimal places. Decreasing the tolerance has the effect of similarly decreasing the error, validating the minimisation procedure. This can seen in Table 3.2 . The Table 3.2 compares the error in the initial conditions as a result of the minimisation procedure applied to the nonlinear first order system for different tolerances on $l(0)$ and for different starting guesses for the initial conditions.

A few additional remarks can be made about Table 3.2.

(i). The minimisation procedure has been verified using $l(0)$ as a measure of performance. Reducing the value on $tol1$ has the effect of increasing the accuracy of the final computed initial value.

(ii). The amount of error in the final initial condition is in all cases within the $tol1$. The error in the final initial condition is in effect the sixth measure of performance, $\left\|x_0 - x_0^k\right\| < tol6$. These two measures are thus directly related.

(iii). There are a number of entries for which no result could be obtained. The reason for this lies in the character of the model equation. The solution of the model equation is

$$x(t) = \frac{x(1)}{1 + x(1) - tx(1)} \tag{3.12}$$

where x(1) is the initial value at $t = 1$.

A discontinuity will exist within the time window $t \in [1, 1.5]$ if there exists a value of t within this interval for which $t = 1 + 1/x(1)$. The modified Euler scheme is unable to deal with discontinuities causing the numerical procedure to fail. The results even show that if $x(1)$ lies just outside this interval it may still fail unless the size of the timestep is reduced to a sufficient extent to cover the difficulty. Also the continuous data assimilation problem will not be

Size of timestep h	$l(0) < tol1$	true initial value	original guess of initial value	Number of minimsation steps	error in final initial condition
0.1	0.01	-1.0	-5.0	11	5.4E-3
0.01				15	7.03E-3
0.001				15	9.3E-3
0.1	0.01	-1.0	-0.1	6	-6.03E-3
0.01				7	-7.04E-3
0.001				7	-7.63E-3
0.001	0.01	-1.0	-1000	N.A.	inf
0.0005			-1000	789	6.98E-3
0.1			1	9	6.71E-3
0.1			2.5	N.A.	inf
0.1			1000	N.A.	inf
0.1	0.0001	-1.0	-5.0	17	-3.11E-5
0.1			-0.1	12	-4.5E-5
0.001			-1000	N.A	inf
0.0005			-1000	796	8.29E-5

Table 3.2: over time interval [1,1.5]

defined, as a discontinuity contravenes the assumption that the solution needs to be a C^2 function.

The value $l(0)$ in effect gives the gradient direction $\delta\mathcal{J}$, the descent direction needed for an improved estimate for the initial value which will reduce the value of the objective function $\mathcal{J}$. The relative change in the objective function $\frac{\delta\mathcal{J}}{\mathcal{J}}$ is given by the fourth measure of performance $\frac{\mathcal{J}^{k+1}-\mathcal{J}^k}{\mathcal{J}^k} < tol4$. Just as with $l(0)$, a tolerance restriction $tol2$ on $\frac{\delta\mathcal{J}}{\mathcal{J}}$ can be used as a stopping condition. A few additional conditions need to be included.

(i). Once the value of $\mathcal{J}$ is within machine precision the iteration needs to be stopped. The true solution has then been reached and $\mathcal{J}$ has been been minimised to machine accuracy.

(ii). If there is negligible change in the value of $\frac{\delta\mathcal{J}}{\mathcal{J}}$ from one iteration of the minimisation procedure to the next, then the procedure needs to be stopped. Little improvement will result from letting the procedure run.

The results in Table 3.3, compare the error in the initial conditions for different tolerances upon the approximations to $\frac{\delta\mathcal{J}}{\mathcal{J}}$ for various starting initial values.

An explanation of the Table 3.3 follows,

(i). Three stopping conditions were used: condition 1 (3.5), condition 2 (3.6) and condition 3 (3.4). $tol3$ and $tol5$ were both set at 1E-15. $tol4$ was varied (as shown in the second column.) The results do verify the convergence of the minimisation procedure. Reducing $tol4$ from 1 to 0.01 did have the effect of improving accuracy, so much so, that it resulted in condition 2 being used as a stopping criterion.

(ii). A relative small change in $tol4$ can have a large effect on the final error in the initial value.

Size of timestep h	$\frac{\mathcal{J}^{k+1}-\mathcal{J}^{k}}{\mathcal{J}^{k}} < tol4$	true initial value	original guess of initial value	Number of minimsation steps	stop-ping crit-erion	final l(0)	error in final initial value
0.1	0.01	-1.0	-5.0	25	2	-2E-8	3.6E-8
0.01				32	2	-3E-8	-6.0E-8
0.001				33	2	-3E-8	5.4E-8
0.1	0.01	-1.0	-0.1	20	2	-2E-8	-2.9E-8
0.01				24	2	-3E-8	-5.9E-8
0.001				25	2	-2E-8	-4.4E-8
0.001	0.01	-1.0	-1000	2	1	-2.0	998
0.0005			-1000	2	1	-1.7	997
0.1			1	24	2	-2E-8	3.3E-8
0.1			2.5	2	1	N.A.	inf
0.1			1000	2	1	N.A.	inf
0.1	1.0	-1.0	-5.0	2	1	-0.9	2.76
0.1			-0.1	2	1	1E-1	-1.7E-1
0.001			-1000	2	1	-2	998
0.0005	0.003		-1000	2	1	-1.7	997
0.0005	0.002	-1.0	-1000	807	2	-2E-8	4.1E-8

Table 3.3: validation of minimisation procedure over time interval [1,1.5], with stopping criteria directly related to the objective function $\mathcal{J}$

(iii). Where a singularity in (3.12) is just outside the time interval as with a starting initial value of $x_0 = -1000$, not only does the time-step have to be very small for the integration method to recognise that the singularity is not within the time period but also the bound on relative change in the objective function $\mathcal{J}$ ($tol4$), needs to be very small to obtain a correct result. Once $tol4$ was reduced to 0.002 the iteration proceeded normally.

(iv). It is interesting to note that the final result in Table 3.3 was produced by the the numerical procedure stopping due to condition 2. This shows how critical it is to find the correct tolerances for the above stopping criteria.

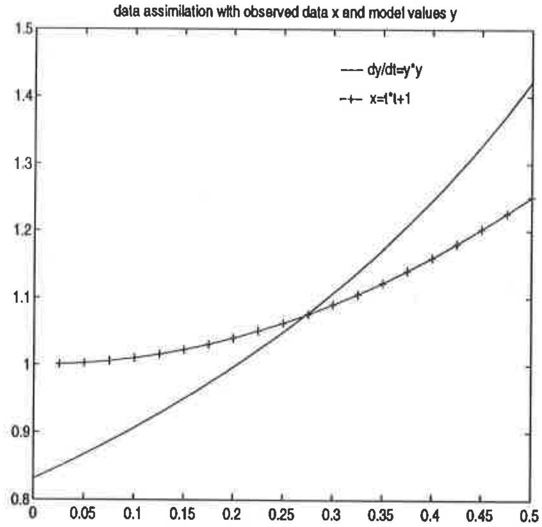
As the 'true' initial value is known throughout these experiments and defines the observed values, the error in minimising $\mathcal{J}$ is directly related to the error between the computed initial value and the 'true' initial value. This leads to another stopping criterion (3.7). Results have been already shown for this measure, disguised as the error in the final computed initial value.

In most of the numerical experiments carried out the 'true solution' is a solution which is obtainable. For interest, it might serve to examine the case where this is not so. One such example would be to find the best fit of a solution to the first order nonlinear equation to a parabola $x = t^2 + 1$. The results are shown in the Figure 3.1..

As seen in Figure 3.1 there is a discrepancy between the parabola and the solution. This shows that the best fit is restricted to the possible solutions to the model equations. If none of these possible solutions can exactly fit the observed data then the final solution will result in the minimised objective function $\mathcal{J}$ being a positive value.

An important question to consider is what effect there is in diminishing the timestep in a minimisation procedure. An obvious restriction on the maximum size of the timestep is the region of absolute stability. Reducing the stepsize past a

Figure 3.1: Best fit of model solution to a parabola



certain limit will have a detrimental effect on calculation of the descent direction in aggregation of roundoff error. It is thus possible that there may be a limit in certain cases to how small a tolerance can be applied to a stopping criterion that is achievable by the computer at hand.

3.4 Second Order System

A linear second order system is used (2.55) to which the Modified Euler scheme is applied, giving (2.56) and (2.57). The respective adjoint equations to (2.56) and (2.57) were calulated in Chapter 2 and are presented in (2.61)

The discretised model equations (2.56) and (2.57) are second order accurate. This is shown in Table 3.4.

3.4.1 Properties of the continuous system

It is instructive to look at the simple dynamical properties of the dynamical system that we are modeling. The second order linear system has eigenvalues $+3, -3$. There is only one equilibrium point and it is a saddle point located at the origin. There

norm of error in modeuler for (2.56) at t=1 (I.C$\boldsymbol{x}(0) = (1,1)$)	Number of timesteps to reach t=1.5	order of magnitude improvement in reduction of error
0.51853264050297	16	
0.13970712472975	32	3.71
3.6222666423452D-02	64	3.86
9.2192807410619D-03	128	3.93
2.3253487538627D-03	256	3.96

Table 3.4: showing that (2.56) and (2.57) are second order accurate

are two axes on the phase plane; one $y = 0.2x$ for which x and y diverge from the origin and the other $y = -x$ on which x and y converge on the origin with increasing time.

The solution can be expressed as the set of equations

$$\begin{aligned} x(t) &= \left(\frac{5}{6}e^{3t} + \frac{1}{6}e^{-3t}\right) x(0) + \left(\frac{5}{6}e^{3t} - \frac{5}{6}e^{-3t}\right) y(0) \\ y(t) &= \left(\frac{1}{6}e^{3t} - \frac{1}{6}e^{-3t}\right) x(0) + \left(\frac{1}{6}e^{3t} + \frac{5}{6}e^{-3t}\right) y(0) \end{aligned} \quad (3.13)$$

Information on the continuous system is essential for any form of prediction as to the success or failure of the assimilation technique on the discretised problem. As before the descent direction is found. Now it has two components: one in the x direction lx_0 and the other in the y direction ly_0. These are obtained from the discretised adjoint equations.

A few predictions can be made on account of the properties of the respective continuous model system equations.

(i). We expect the minimisation procedure to be verified to converge to the true solution.

(ii). There may be difficulties when the 'true' solution is located on one of the axes of the phase plane.

In Table 3.5 a single stopping criterion was used and that was $l(0) < tol1$. The size of the timestep, the tolerance on the stopping criterion, the 'true' initial value and the original guess of the initial values were varied. The error in the final initial condition was measured as the the square root of the sum of the squares of the respective errors in the x and y components.

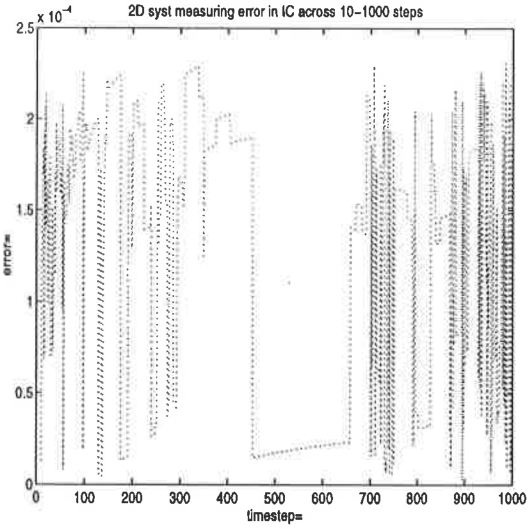
One of the most salient points in Table 3.5 is the large differences in final error in the initial condition resulting from different original guesses to the initial value. The most extreme results occur where the guess of the initial value and the true initial value lie both on either $y = 0.2x$ or $y = -x$, the two axes in the bifurcation diagram. On the stable axis $y = -x$ there are no problems. Extremely low tolerances on the stopping criterion can be easily achieved even with a relatively large timestep 0.1. However, where the true and starting initial values lie on the unstable axis, $y = 0.2x$, accuracy in the final computed result is is approximately fifty times worse. Reducing the tolerance on the stopping criterion does generally reduce the final error. The problem lies in forcing the minimising procedure to achieve such tolerances. Smaller and smaller sizes of timesteps are needed for smaller tolerances on the stopping criterion to be reached. Though this condition is a standard one, what is unusual is the relative amount of reduction in timestep needed to achieve a subscribed reduced tolerance. The tolerance can easily be reduced to a value which cannot be attained where the optimal accuracy of the numerical scheme is simply not good enough.

Figure 3.2 illustrates the second order system. Data assimilation was applied within a time window between 0 and 0.5 for a 'true solution' with initial conditions $(1,1)$ and an initial guess at $(2,2)$. The stopping criterion was $\|\boldsymbol{l}\| < 0.0001$.The diagram compares the error in the resulting initial value over a range of timesteps (between 10 and 1000). What can be seen is that the error is always lower than 2.5×10^{-4} and that there is no improvement whether there are 50 timesteps or a

Size of timestep h	$l(0) < tol1$	true initial value	original guess of initial value	Number of minimsation steps	error in final initial condition
0.1	0.01	(1,1)	(2,2)	11	4.29E-4
0.1	0.001	(1,1)	(2,2)	13	4.38E-4
0.1	0.0001	(1,1)	(2,2)	N.A.	inf
0.001	0.0001	(1,1)	(2,2)	19	1.7E-4
0.1	0.01	(1,1)	(1,3)	15	8.50E-4
0.1	0.001	(1,1)	(1,3)	21	8.46E-4
0.001	0.0005	(1,1)	(1,3)	13	3.03E-4
0.1	0.01	(1,1)	(1,0)	15	8.5E-4
0.1	0.01	(1,1)	(1,-1)	15	8.5E-4
0.1	0.01	(5,1)	(10,2)	10	1.3E-3
0.1	0.01	(10,2)	(5,1)	10	1.3E-3
0.1	0.01	(1,-1)	(2,-2)	10	2.3E-5
0.1	0.01	(2,-2)	(1,-1)	10	2.5E-5
0.1	0.01	(0,0)	(1,0)	9	1.7E-3
0.1	0.01	(0,0)	(5,1)	10	1.3E-3
0.1	0.01	(0,0)	(1,-1)	12	2.5E-5
0.1	0.00001	(0,0)	(1,-1)	10	5.3E-9
0.1	0.00001	(0,0)	(5,1)	N.A.	inf

Table 3.5: testing data assimilation on the second order system across time period $t \in [0, 1]$

Figure 3.2: Error in initial value in second order system across a range of timesteps

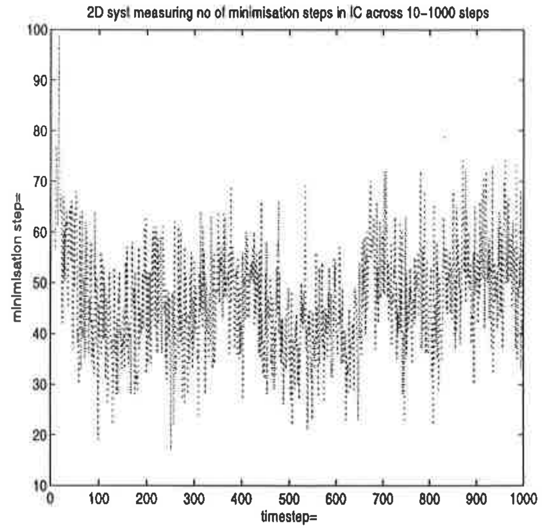


1000. The great fluctuations in error are because differing number of minimisation steps are used. There is a trough in the error between 450 and 650 timesteps. As the same number of minimisation steps are used throughout this interval it is suggested that the gradual increase in error is influenced by round off. The fact that there is no real improvement in reduction of the size of timestep indicates that the error in $\mathcal{J}$ dominates.

If the error in $\mathcal{J}$ dominates there should be no steady decrease in number of minimisation steps with decreasing the timestep. This is what happens. Data assimilation was applied within a time window between 0 and 0.5 between a 'true solution' with initial conditions $(1, 1)$ and an initial guess at $(2, 2)$. The experiment was repeated with number of timesteps successively increased from 10 to 1000 for the same time window. For every size of time step the number of minimisation steps needed to get a tolerance of $l < 1 \times 10{-}7$ was recorded. The results are shown in the Figure 3.3.

For comparison some other measures of performance were used. As with the scalar model conditions 1,2,3 were used. $tol4$ was set at 0.1. $tol3$ and $tol5$ were both set at 1E-15. Then the final error in initial values $x(0)$ and $y(0)$ with respect

Figure 3.3: How number of minimisation steps vary with size of timestep



to the number of timesteps within the time interval $(0, 0.5)$ was compared. This is represented in the graphs shown in Figure 3.4.

(i). There does not seem to be any noticeable reduction in error by increasing the number of timesteps. This is similar to what occured when $\boldsymbol{l}_0$ was used as a stopping criterion.

(ii). On graphs in Figure 3.4 there are gaps in the plotted graphs. At these points the objective function $\mathcal{J}$ itself becomes zero, making the approximation of $\frac{\delta \mathcal{J}}{\mathcal{J}}$ unintelligable to the computer.

(iii). There are a number of peaks in the results. The value at these peaks were examined and do satisfy the stopping criteria. It seems that the largest peak indicates the tolerance in the error. It seems that at other points the error is almost completely eliminated.

(iv). The error in x(0) seems to be mirrored roughly in magnitude in y(0) with a reversal in sign. This can be seen on Figure 3.5 where the errors are compared.

Figure 3.4: Error in x(0), y(0) across various timesteps

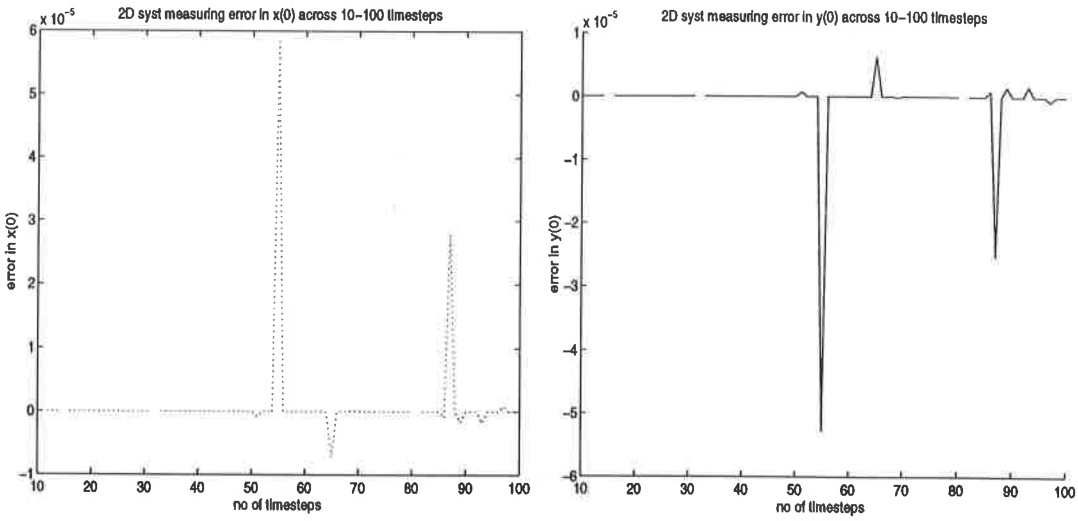
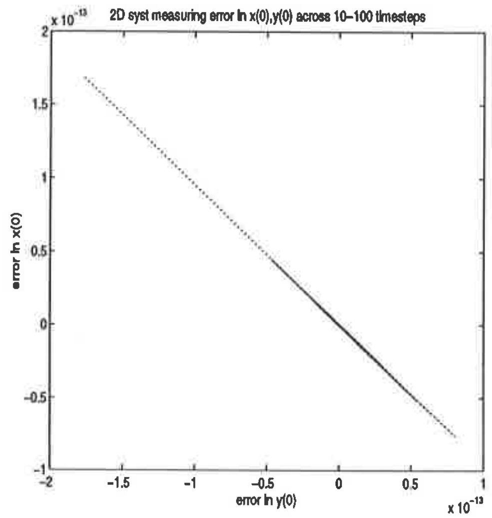


Figure 3.5: Error in x(0), y(0) with stricter stopping condition



(v). Reducing the tolerance on the approximation to $\frac{\delta \mathcal{J}}{\mathcal{J}}$ does have the effect of reducing the overall error in the initial conditions. This can be seen in Figure 3.5.

Where the stopping criterion of $\mathcal{J}$ is 0.00001 the error in either initial condiiton is no greater than 2×10^{-13}. In comparison when the stopping criterion $tol4$ is 0.1 the corresponding error is no greater than 6×10^{-5}. The minimisation process is convergent.

Now that the minimisation process has been examined, the data assimilation process can be studied on a more complex discrete system - a discrete system derived from the Lorenz equations. This is the subject of the next chapter.

Chapter 4

THE LORENZ EQUATIONS

4.1 Introduction

The Lorenz equations constitute a third order nonlinear system described by

$$\begin{aligned} \dot{x} &= a(y - x) \\ \dot{y} &= rx - y - xz \\ \dot{z} &= xy - bz \end{aligned} \tag{4.1}$$

where a, r, b are fixed parameters and x, y, z are continuous functions of time.

4.2 Properties of the Lorenz Equations

4.2.1 Stationary points

The number of the stationary points vary depending on the values given to the parameters a, r, b. The origin is a stationary point for all parameter values and is stable and globally attracting for values of r between 0 and 1. A simple bifurcation occurs at $r = 1$, such that for $r > 1$ two further stationary points C_1, C_2 exist at points $(\pm\sqrt{b(r-1)}, \pm\sqrt{b(r-1)}, r-1)$ with C_2 in the region $x > 0$. For $r > 1$ the origin is non-stable. The linearised flow around this point has two negative, and one positive eigenvalue, making it a saddle point.

The character of C_1 and C_2 also varies by changing the value of r. If $1 < r < 1.346$ with $(a = 10, b = 8/3)$, C_1 and C_2 are stable, both having a negative real part. Where $1.346 < r < 24.74$, the two stationary points are also stable, each having a pair of complex conjugate eigenvalues. Where $r > 24.74$, C_1 and C_2 are both non-stable. Both stationary points have a negative eigenvalue and a complex conjugate pair of eigenvalues whose real part is positive.

In this study the parameters are fixed at $a = 10, b = 8/3, r = 28$. Then three non-stable stationary points exist at $(0,0,0)$, $(6\sqrt{2}, 6\sqrt{2}, 27)$ and $(-6\sqrt{2}, -6\sqrt{2}, 27)$.

4.2.2 Other properties

Three other properties that are important are given below.

(i). The Lorenz equations (4.1) have a natural symmetry with $(x, y, z) \rightarrow (-x, -y, z)$

(ii). The z-axis is time invariant such that any trajectory that starts on the z-axis will remain on it and will approach the the origin as time progresses.

(iii). There is a bounded ellipsoid E in $\Re^3$ into which all trajectories eventually enter (Lorenz, 1963).

Further properies of the Lorenz equations can be obtained from (Sparrow,1982).

4.3 Application of Data Assimilation to the Lorenz Equations

4.3.1 A review of literature

Applying advanced data assimilation techniques to the Lorenz Equations has been a popular topic in recent years. The reason for this is that the Lorenz equations like many systems in oceanography, are dissipative in nature and volume reducing, tending asympotically to a set of zero volume.

Gauthier,(1992) applied variational data assimilation to the Lorenz equations. An adjoint method was used on a discretisation of the associated linear tangent model of the system. (Two good references to the linear tangent model are (Talagrand and Courtier, 1987) and (Courtier and Talagrand, 1990)). Observations were considered from both predictable and unpredictable settings. The predictable situation involved observations around just one strange attractor; the unpredictable setting used observations which revolved around both. In the unpredictable case, the final solution was found to be far more sensitive to initial conditions. There were more secondary minima present, making the choice for first guess to the initial value crucial.

In the paper by Miller et al.,(1994), the study of Gauthier,(1992) was extended. Advanced data assimilation techniques such as the Extended Kalman Filter and methods derived from the weak constraint approach were used to study the Lorenz equations. The results were comparable to those of Gauthier,(1992), but they clearly show that for weak constraint methods, the number of local minima with respect to the cost functional $\mathcal{J}$ increases as the time period is increased over which the minimisation takes place.

4.3.2 Present study

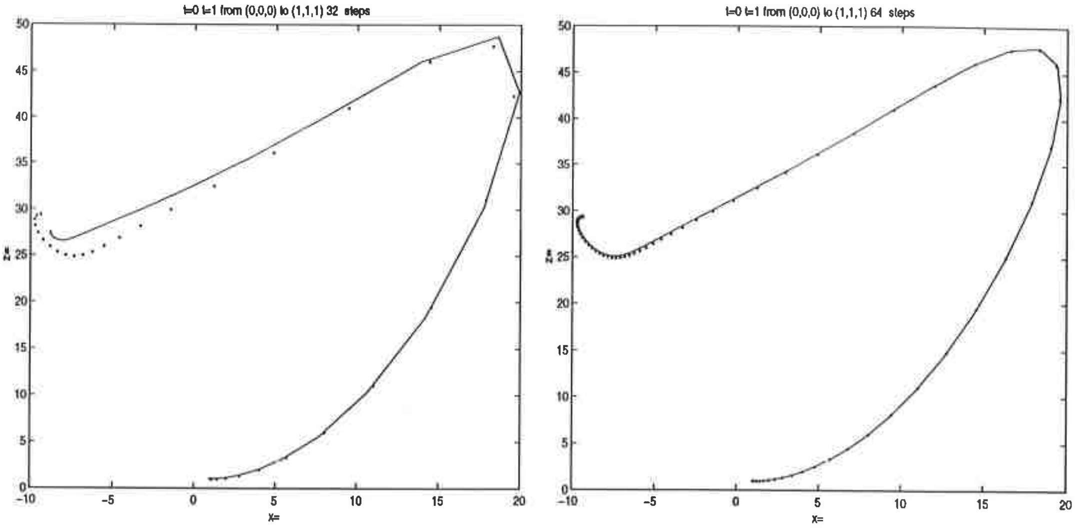
In this dissertation we shall not be looking for the true solution to the Lorenz equations. As the system is time varying, there is no analytic solution which can be used as a comparision. Moreover the system is very sensitive to the initial conditions. For any numerical scheme used, the error will increase exponentially with time. Thus instead of trying to find the true solution of the Lorenz equations, we consider just the properties of a discretisation of the Lorenz system. Any numerical approximation to the Lorenz equations is a discrete system which will display some similar characteristics in behavior. If no round-off error occurs, for a sufficiently small interval and a sufficiently small time step the approximation will be quantitatively similar to (4.1). As round-off error exists, this cannot be guaranteed.

Since we are not looking for the true solution to the Lorenz equations, a low order scheme can be used. As in Chapter 3 the Modified Euler scheme is used producing the discretised system

$$\begin{aligned}
x_{i+1} &= x_i + \frac{1}{2}a\Delta t(2(y_i - x_i) - a\Delta t(y_i - x_i) + \Delta t(rx_i - y_i - x_i z_i)) \\
y_{i+1} &= y_i + \frac{1}{2}\Delta t(2(rx_i - y_i - x_i z_i) + \Delta t(ra(y_i - x_i) - (rx_i - y_i - x_i z_i) - \\
&\quad x_i(x_i y_i - bz_i) - az_i(y_i - xi) - \Delta ta(y_i - x_i))(x_i y_i - bz_i))) \\
z_{i+1} &= z_i + \frac{1}{2}\Delta t(2(x_i y_i - bz_i) + \Delta t(ay_i(y_i - x_i) + x_i(rx_i - y_i - x_i z_i) \\
&\quad + \Delta ta(y_i - x_i)(rx_i - y_i - x_i z_i) - b(x_i y_i - bz_i))) \qquad (4.2)
\end{aligned}$$

The same numerical minimisation procedure as in Chapter 3 is adopted in applying data assimilation to the process. The two pictures in Figure 4.1 compare a 'true' solution (bold line) to a 'best fit' solution (dotted line) derived from the data assimilation process. The observations are derived by applying (4.2) to the time interval $0 < t < 1$, using a stepsize an eighth to that used in the data assimilation process and started from some 'true' initial value. In this example, the 'true' initial

Figure 4.1: Comparison of model error



condition is $(1, 1, 1)$ The observations comprise every eighth data point in the 'true' solution such that both observations and data points in the model equations occur at the same point in time. The final 'best fit' solution is obtained with a starting initial value of $(0, 0, 0)$, a stopping criterion of $||\boldsymbol{l}|| < 0.01$ and a time interval $0 < t < 1$.

As one can see from the pictures in Figure 4.1, Figure 4.2, the model error introduced by using different stepsizes is reduced by increasing the number of the timesteps in both the 'true' solution and in the data assimilation process. The general behavior is also consistent to what is expected from the Lorenz equations.

The main aim is to capture the general properties of the 'true' solution through the data assimilation process. However, with the Lorenz equations, as one increases the time interval, inaccuracies accumulate until one gets qualitatively the wrong pattern. This can be seen in Figure 4.3 (left) where the time interval is increased from $0 < t < 1$ to $0 < t < 14$.

We wish to consider the effect of applying a different set of weights to differing points in time. By placing a stronger weight on earlier terms in the system's progression, it should be less possible for the problem in Figure 4.3 to occur. The original initial value is presumed to be reasonably close to the true one. No such assumption

Figure 4.2: Comparison of model error

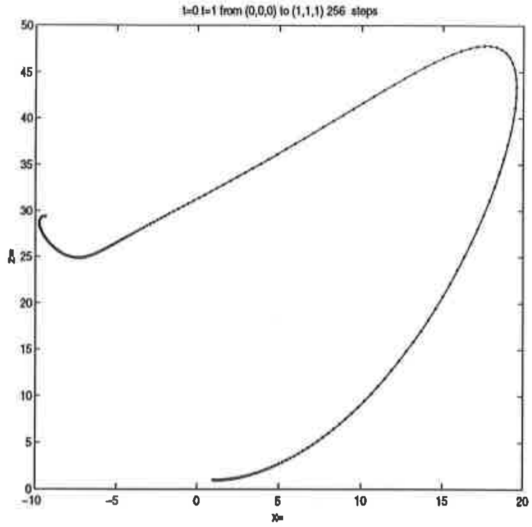
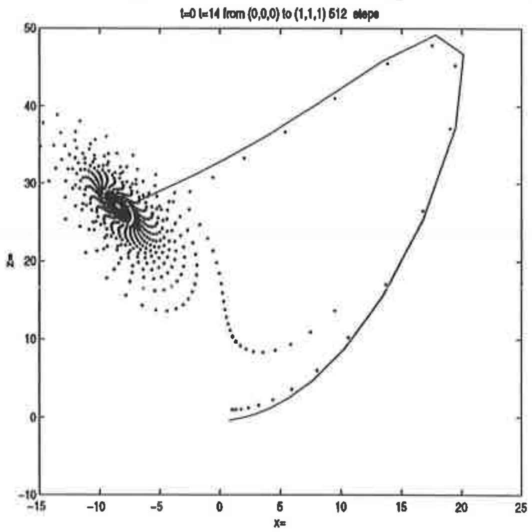


Figure 4.3: The problem with a longer time interval



can be made, however, for corresponding values at the end of the time interval. It is hoped that placing a stronger weight on the earlier terms will allow a more accurate solution to be obtained, making it possible to extend the time interval on which the data assimilation process is carried out.

As described in Chapter 2, weights can be applied to different model values by varying the values in the diagonal matrix D. The weights were set such that

$$d_j = \frac{N+1}{\sum_{i=0}^{i=N} exp(-10i\Delta t)} exp(-10j\Delta t), \tag{4.3}$$

where d_j is the j^{th} diagonal element corresponding to the j^{th} model value, with $j = 0, \ldots N$.

Two stopping criteria are used. The first is for $||\boldsymbol{l}_0||$ to be smaller than 0.0001. The second was for $||\boldsymbol{l}_k - \boldsymbol{l}_{k-1}||$ to be smaller than machine accuracy. The observations were obtained by starting the discretsed system of (4.2) from a 'true' initial value with the same step size as in the data assimilation process. Thus, unlike the previous experiment, no model error was present.

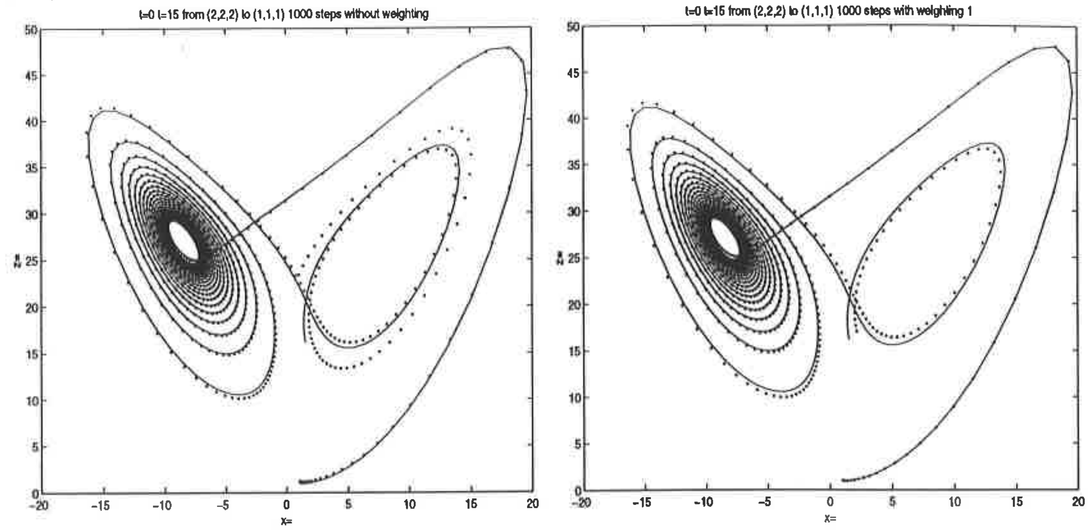
Although there is an improvement in the accuracy in the final initial value obtained through applying the weighting function (4.3) to the data assimilation process, there seems to be no real overall gain in accuracy over the whole time period; the graphs of the solution with/without weighting across the time period $0 < t < 1$ seem identical. This may be due to using too small a timestep as well as using too small a time period. Thus a comparison is made between using and not using the weighting function for a larger time period of $0 < t < 15$ and a larger timestep of 0.015. The results are shown in Figure 4.4 for a 'true' initial value at (1,1,1) with an original guess at (2,2,2).

As seen in the graphs of Figure 4.4, the application of weighting produced a better result overall, even with an improved estimate for the 'true' initial value. Over such a time period the data assimilation process without any weighting picks a solution which follows a different trajectory to that indicated by the observations. The values

	Original guess of initial value at (0,0,0)	Original guess of initial value at (2,2,2)
error in initial conditions without weight distribution (4.3)	4.48E-3	3.04E-3
final $\|\|\boldsymbol{l}_0\|\|$ value without weight distribution (4.3)	2.0E-4	2.0E-4
final number of minimisation steps without weight distribution (4.3)	137	188
error in initial conditions using a weight distribution (4.3)	2.72E-3	1.63E-3
final $\|\|\boldsymbol{l}_0\|\|$ value using weight distribution (4.3)	1.01E-3	5.94E-4
final number of minimisation steps using weight distribution (4.3)	101	103

Table 4.1: Comparison of the error in the final initial value when the weighting distribution is/is not used with a time-step of 0.0001, a 'true' initial value at (1,1,1) and a time period [0,1].

Figure 4.4: Comparision of applying weighting (left) to not applying weighting (right) at longer time interval



at the end of the time period weakly influence the overall data assimialtion process, allowing for the aggregation of error. This insight lead to choosing a less extreme weighting distribution as described in equation (4.4).

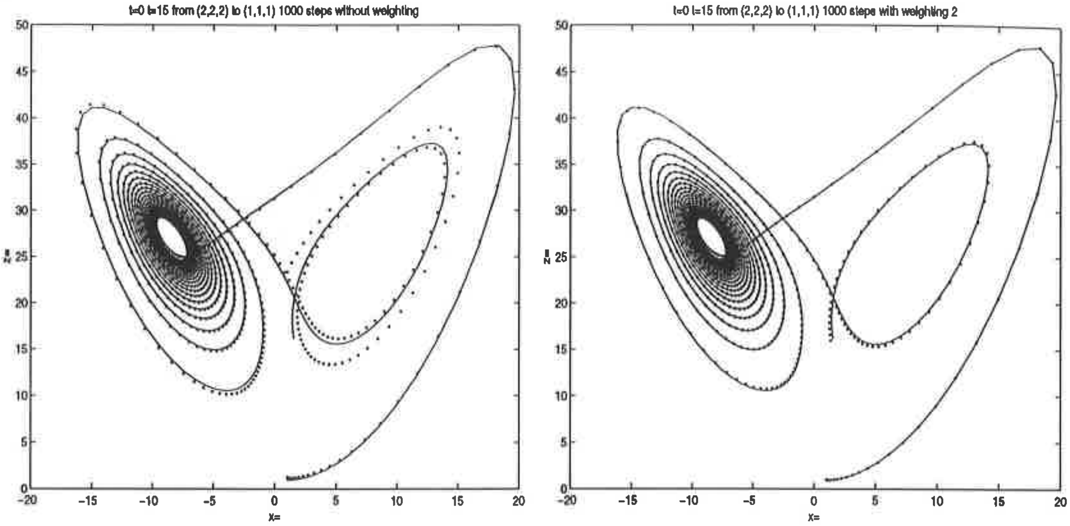
$$d_j = \frac{N+1}{\sum_{i=0}^{i=N} exp(-i\Delta t)} exp(-j\Delta t), \tag{4.4}$$

where d_j is the j^{th} diagonal element corresponding to the j^{th} model value, with $j = 0, \dots N$,

A comparison is made between using the weighting distribution of (4.4) and no weighting over a time period of $0 < t < 15$. The time-step is kept at 0.015 with a true initial value at (1,1,1) and a first guess at (2,2,2). When weighting (4.4) is applied, the trajectories obtained by the observations and data assimilation process match. The results are shown in Figure 4.5.

This shows that the weighting must be balanced so as not only to ensure a more accurate initial condition, but also to be such that the errors in model solution values at the end of the time period do not dominate over the gains in accuracy obtained

Figure 4.5: Comparision of applying weighting (left) to not applying weighting (right) over a time period $0 < t < 15$



at earlier moments in time.

Chapter 5

Conclusion

The aim of this thesis has been to investigate variational data assimilation. Chapter 2 gives the mathematical background to the problem. It shows that there are a number of ways to formulate a discretised adjoint to the system. A discretisation of the adjoint obtained from the continuous example is not the same as the respective discretised adjoint derived from the discrete model. Also Chapter 2 shows some of the similarities between the discrete data assimialtion problem and the continuous data assimilation problem. This is due to the fact that both problems share many properties when considered in terms of Hilbert Spaces.

Chapter 3 describes the numerical minimisation technique used throughout the rest of the thesis. In addition, a number of numerical experiments are carried out on the scalar discrete example and on the second order linear discrete example. In both cases the model equations were verified to be second order accurate. The minimisation procedure was also verified in both cases to converge to the true solution. The general emphasis rested on tests on the minimisation procedure and not on the the data assimialtion problem itself; the problems were too simple for anything to disrupt the data assimilation process. Any problems that occurred were a result of the model being studied. For instance, the data assimilation procedure for the discrete scalar example failed when a discontinuity in the solution entered the time period under consideration. Similarly, the speed of convergence of the minimisation process in the second order system varied in the position of the 'true' initial value as well as the position of the first guess to that value.

Chapter 4 considers the application of data assimilation to a discretised system of the Lorenz equations. A weighting distribution was applied to this system. Applying time-varying weights to the Lorenz system across the time interval has not previously been examined. However, the result of Figure 4.5 raises the question as to whether the weight distribution works well for other 'true' initial values and other first guesses to such values. A few other points have been considered and the preliminary findings indicate that the time period can be extended as in Figure 4.5. A more thorough investigation is needed.

The application of weights to discretised systems is of wide interest. The Meteorological Office in their forecast model use a background term which effectively acts as a weighting term on the initial value such that the value will not vary too much from what is expected. A problem, however, as shown in this study is finding the optimal strength of weights. Further research needs to be carried to extend this idea to variable weights.

Another pertinent line of interest concerns the effect of using fewer observations than model values, as occurs in the numerical weather prediction. A brief study where just one observation is applied to the discrete scalar example has been carried out. The results are similar to those in (Griffith and Nichols, 1994). The data assimilation process is only successful if the position in time where the observation point exists is not too far away from the beginning of the time period. Otherwise convergence to secondary minima arise. It was also found that the speed of the minimisation process was influenced not only by the time separation between the observation point and the start of the time period, but also by the size of the time step. The more time-steps that are present, the slower the process.

As mentioned in the introduction, the number of avenues for research in data assimilation are almost endless. In particular, the consideration of fewer observations as well as the study of applying weights to the data assimilation process are areas which will invoke interest for many years to come.

5.1 References

P.Cressman, An operational objective analysis system, *Monthly Weather Review*, 87, 367-374, 1959.

P.Courtier and O.Talagrand, Meterological observations with the direct and adjoint shallow water equations, *Tellus*, 42A, 531-549, 1990.

P. Gauthier, Chaos and quadri-dimensional data assimialtion: a study based on the Lorenz model, *Tellus*, 44A, 2-17, 1992.

M. Ghil and P. Malalnotte-Rissoli, Data assimilation in meterology and oceanography, *Adv. Geophys*, 33:141-266, 1991.

A.K.Griffith and N.K.Nichols, Data assimilation using optimal control theory. Numerical Analysis Report 10/94 The University of Reading, 1994.

A.K.Griffith and N.K.Nichols, Data Assimilation using observers. Numerical Analysis Report 11/94 The University of Reading, 1994.

A.K.Griffith and N.K.Nichols, Accounting for model error in data assimilation using adjoint methods. Numerical Analysis Report 2/96 The University of Reading , 1996.

A.C.Lorenc, Analysis methods for numerical weather prediction. *Quart.J.R.Met.Soc.*, 112, 1177-1194, 1986.

R.N.Miller,M.Ghil and F.Gauthiez, Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems, *Journal of Atmospheric Sciences*, vol 51(8), 1037-1056, 1994.

L.A.Pars, *Introduction to the Calculus of Variations.* Heinenmann, 1962.

C.Sparrow, *The Lorenz Equations : Bifurcations, Chaos, and Strange Attractors.* Springer-Verlag, 1982.

O.Talagrand. A study on the dynamics of four-dimensional data assimilation, *Tellus*, 33, 43-60, 1981.

O.Talagrand, On the mathematics of data assimilation. *Tellus*, 33, 321-339, 1981.

O.Talagrand and P.Courtier, Variational assimilation of the meterological observations with the adjoint vorticity equation. I: theory. *Q. J. R. Meterol.Soc.*, 113, 1311-1328, 1987.