Assessing and Optimising the Performance of Data Assimilation Algorithms

Noeleene Mallia-Parfitt

Department of Mathematics and Statistics

University of Reading

This thesis is submitted to the Department of Mathematics and Statistics in partial fulfilment of the requirement for the degree of Doctorate of Philosophy.

November 2016

Abstract

Data Assimilation means to find a trajectory of a dynamical model that matches a given set of observations. A problem of data assimilation experiments is that there is no possibility of replication. This is due to the fact that truly 'out-of-sample' observations from the same underlying flow pattern but with independent errors are usually not available. A direct evaluation against the available observations is likely to yield optimistic results since the observations were already used to find the solution.

A possible remedy is presented which simply consists of estimating the optimism, giving a more realistic picture of the out-of-sample performance. The approach is simple when applied to data assimilation algorithms employing linear error feedback. Moreover, the simplicity of this method allows the optimism to be calculated in operational settings. In addition to providing a more accurate picture of performance, this approach provides a simple and efficient means to determine the optimal feedback gain matrix.

A key feature of data assimilation schemes which employ linear error feedback, is the feedback gain matrix used to couple the underlying dynamical system to the assimilating algorithm. A persistent problem in practice is to find a suitable feedback. Striking the right balance of coupling strength requires a reliable assessment of performance which is provided by our estimate of the out-of-sample error. Numerical and theoretical results demonstrate that in linear systems with gaussian perturbations, the feedback determined in this way will approach the optimal Kalman Gain in the limit of large observational windows.

Declaration

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

Signed

_____

Noeleene Mallia-Parfitt

Acknowledgements

I would like to thank my supervisors, Dr Jochen Bröcker and Professor Peter Jan van Leeuwen. Over the past three years Jochen has provided me with continuous support and was always there to answer my questions. His door was always open and it is truly appreciated. Discussions with Peter Jan were very helpful and provided constructive suggestions and ideas which motivated some work in this thesis.

I would like to thank the University of Reading and the Department of Mathematics and Statistics for financial support.

My heartfelt thanks goes to my fellow PhD students for our interesting discussions and well earned, enjoyable tea breaks. I would also like to thank other members of staff in the department who made my time studying for my PhD a happy experience. In particular I thank Mrs Peta-Ann King who went above and beyond her role to help, support and provide encouragement every time the occasion called for it.

I am eternally grateful to my husband Max who provided endless patience, support and understanding throughout this process. His belief in my potential has been instrumental in completing this PhD.

For my father, who is always remembered

Contents

List of Figures

Chapter 1

Introduction

Our daily weather forecasts start out as initial value problems on the national weather services supercomputers (Kalnay 2001). Numerical weather prediction (NWP) provides the basis for weather forecasting beyond the first few hours. These forecasts are performed by running computer models of the atmosphere that, given some observations, can simulate the evolution of the atmosphere. The integration in time of an atmospheric model is an initial value problem. In order to achieve a good forecast, it is necessary that the computer model be a realistic representation of the atmosphere and that the initial conditions be known accurately. The process which we call *data assimilation*, uses both observations of the atmosphere and short range forecasts to estimate the initial conditions.

Formally, data assimilation involves the incorporation of observational data into a numerical model to produce a model state that accurately describes the observed reality, Le Dimet & Talagrand (1986). An illustration of data assimilation is shown in figure (1.1). The problem is as follows: Given some observations (black line, left panel) and a dynamical model (right panel), find a trajectory of the model (red line, right panel) which, when mapped into observation space, follows the observations (red line, left panel).

The data assimilation algorithms must produce a trajectory that is close to the observations up to a certain degree of accuracy and must verify dynamical and/or statistical

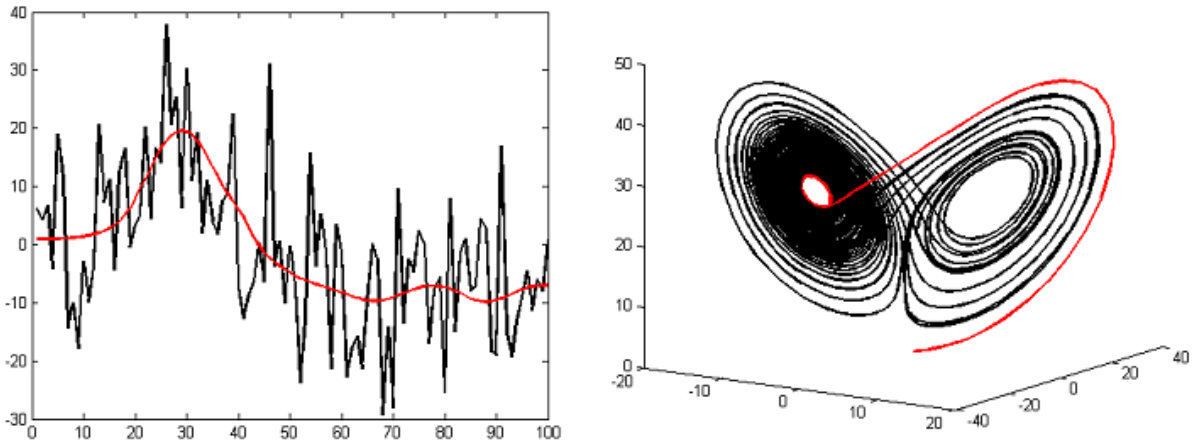


Figure 1.1: The problem of data assimilation: Given some observations (black line, left panel) and a dynamical model (right panel), find a trajectory of the model (red line, right panel) which, when mapped into observation space, follows the observations (red line, left panel).

relationships which are known to be satisfied by the model. Some initial information about the dynamical and/or statistical properties of the reality should be introduced into the analysis of the assimilation algorithm. The trajectory of the dynamical system is then evaluated using the observations.

There are many different types of data assimilation algorithms that approach the problem in different ways; however ultimately their goals are very similar. As such most data assimilation schemes work in cycles over time. The initial information from the previous cycle, called the *background field*, is used at the start of every new cycle. Since any cycle uses observations available up to that point, the initial guess at time n only depends on observations up to $n-1$. Nonetheless, the background field is meant to be the first guess of the state at time n. The feedback gain matrix couples the model of the underlying state to the data assimilation scheme. It depends on the observations up to the previous cycle and may or may not depend on the time evolution. Determining this gain (or coupling) matrix has proven to be problematic as striking the right balance of coupling strength is difficult. Data assimilation schemes differ on how the background field and gain matrix are calculated.

Once the trajectory is obtained, it is desirable to know how good a trajectory it is. This is done by mapping the trajectory into observation space and comparing it with the measured observations. However, these observations were used to obtain the trajectory in the first place, therefore if the trajectory produced follows the observations well, it does not necessarily mean that it is a 'good' trajectory. It is possible that the algorithm is simply reproducing the observations without picking up any of the underlying dynamics. In other words comparing the observations with the output of the data assimilation algorithm may provide an overly optimistic picture of performance. Moreover, assessing the performance this way could easily be cheated. An example of such a case is taking the output of the scheme to be the observations themselves. This would result in zero error however the trajectory obtained is not a good one.

The problem is that correlations between the output and observations are not taken into account. These correlations are present because the observations we are comparing against have been used in the data scheme to obtain the output. An immediate and easy solution would be to use independent observations from the same period and region as the original data and compare the obtained trajectory with these independent observations. Such measurements however are hardly ever available and as such alternative methods must be found.

The problem outlined above appears frequently in statistics. It is known as *overfitting* and there are many ways to deal with this problem. One way is to consider the *out-of-sample* performance of the model. This concept is used to measure how well a process generalises to unseen data and is used in many different applications; see for example Bishop (1995) where it is used in neural networks and Efron (1986) where it is used in statistical learning.

To implement the out-of-sample error in data assimilation, we assume that the observations are corrupted by additive random noise as done in Mallia-Parfitt & Bröcker (2016). If these observations are then assimilated into a dynamical model, the results should be close

to hypothetical observations from the same underlying flow patterns but with independent errors. If the results are not close to these hypothetical observations, then the scheme will not be reproducing the underlying dynamics of the model. The out-of-sample error simply gives us an assessment of how close the results are to these theoretical observations. On average, we can think of the out-of-sample error as the error with respect to the true observations plus a constant; the variance of the observations (Mallia-Parfitt & Bröcker 2016).

Calculating the out-of-sample error can be easily done in the case of data assimilation schemes that employ linear error feedback. The expression derived to determine the out-of-sample error is similar to Mallows' C_p statistic used in model selection in statistical learning (Hastie et al. 2009, Efron 2004). It will be shown that for schemes employing linear error feedback, the out-of-sample error is easily calculated even operationally.

A key feature of data assimilation schemes which employ linear error feedback is the gain or coupling matrix used to couple the underlying system to the algorithm. A persistent problem in practice is to find a suitable feedback gain matrix. If the coupling is too weak the stability of the system cannot be guaranteed while if the coupling is too strong, results deteriorate because the noise will be overly attenuated. Striking the right balance requires a reliable assessment of the performance which is provided by our estimate of the out-of-sample error.

In the case of linear systems with gaussian perturbations, the optimal gain matrix is the Kalman Gain (Anderson & Moore 1979). This is a particular form of the feedback gain matrix that minimises the mean-squared error and provides the best linear unbiased estimate. Computing the theoretically optimal Kalman Gain requires knowledge of the dynamical noise which is not usually available in practice. However, our experiments suggest that choosing the feedback gain matrix by assessing the out-of-sample performance produces a gain matrix which has the same asymptotic behaviour as the Kalman Gain. An advantage of this is that the gain chosen in this way does not require the explicit

knowledge of the dynamical model or dynamical noise. Our experiments demonstrate that the technique can be used in situations where the feedback gain matrix is completely unspecified and also in situations where it has a pre-determined structure but contains unknown parameters.

This suggestion, that the gain matrix minimising the out-of-sample error, converges to the asymptotic Kalman Gain in the limit of large observational windows is intriguing and as such is investigated further.

We first consider constant feedback gain matrices that minimise the expected out-of-sample error. We consider such matrices as we believe they will lead to simpler filters since they would not need to the updated at every time step. Given the fact that the Kalman gain converges to a known limit, called the asymptotic Kalman gain, we prove that a constant gain matrix that minimises the expected out-of-sample error converges to this same limit.

In practice however, we cannot calculate this expected error. Instead it is only possible to estimate the error by for example the empirical mean. This leads to estimates of the gain. The question then is does the same results hold true for the estimate of the minimising gain? To answer this question, we think of the problem in a slightly different way. An alternative way of looking at the problem outlined above, is to think of it as an estimation problem. By this we mean that we look for the feedback gain matrix that minimises a given criterion function.

Suppose that we are interested in a parameter θ attached to the distributions of some observations $X_1, \ldots, X_n$ and let the sample space be denoted by χ. A popular method for finding an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$, is to maximise (or minimise) a criterion function of the type

$$\theta \mapsto M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m_\theta(X_i) \tag{1.1}$$

where $m_\theta : \chi \mapsto \mathbb{R}$ are known functions. In our data assimilation setting the parameter θ represents the feedback gain matrix and the known functions m_θ represent the out-of-sample

error.

An estimator maximising (or minimising) $M_n(\theta)$ over some parameter space Θ, is called an *M-estimator* (Van der Vaart 2000). We are interested in the asymptotic behaviour of sequences of such estimators. The ultimate goal is to establish that the sequence of estimators is *consistent*. This means that the sequence of $\hat{\theta}_n$ converges in probability to θ, where in our setting θ represents the asymptotic Kalman Gain. This non-trivial fact is shown to be true in the case of linear systems with gaussian perturbations employing a data assimilation scheme which employs linear error feedback. The proof presented to establish this result however is missing a small piece. There is one small fact that we were unable to prove completely. It comes down to a very specific result that ensures all minimising feedback gains are stabilising. This fact was rigorously proven in the deterministic version of the proof (Chapter 5), however we were unable to do the same for the stochastic case (Chapter 6). Full details and an intuitive argument are given in the relevant sections.

Some further numerical experiments are also presented. These concern non-linear systems with linear observations as the out-of-sample error theory developed for linear systems is applicable to such systems. We present numerical results for two non-linear systems, one in Lur'e form and one fully non-linear dynamical system. In this setting, it is not so straightforward to establish the convergence of the gain matrix. Non-linear systems are more complicated in that without dynamical noise, it cannot be said that the feedback gain matrix converges in a meaningful way. Nonetheless, the numerical experiments show some interesting results.

1.1 Some Notation

1. The symbol 'D' is used to represent the total derivative of a function $f : \mathbb{R}^m \to \mathbb{R}^n$, so $\mathrm{D}f(x)$ is a linear mapping from $\mathbb{R}^m \to \mathbb{R}^n$.

2. We use D_x and D_y to represent partial derivatives, i.e. for a function $f : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to$

$\mathbb{R}^n$, $\mathrm{D}_x f(x,y)$ is linear $\mathbb{R}^{m_1} \to \mathbb{R}^n$ and $\mathrm{D}_y f(x,y)$ is linear $\mathbb{R}^{m_2} \to \mathbb{R}^n$.

3. We denote by $\mathcal{O}(\mathbf{A}, \mathbf{H})$ the observability matrix of the pair $(\mathbf{A}, \mathbf{H})$. See Section 4.2 for definition.

4. For symmetric matrices $\mathbf{F}$, $\mathbf{G}$ we have,

   (a) $\mathbf{F} \geq 0$ means $\mathbf{F}$ has nonnegative eigenvalues $\Leftrightarrow x^T \mathbf{F} x \geq 0 \quad \forall x \neq 0$.

   (b) $\mathbf{F} \gneq 0$ means $\mathbf{F} \geq 0$ but $\mathbf{F} \neq 0$ which means that $\mathbf{F}$ has nonnegative eigenvalues, not all of them zero.

   (c) $\mathbf{F} > 0$ means that all eigenvalues are positive $\Leftrightarrow x^T \mathbf{F} x \geq 0 \quad \forall x \neq 0$. If $\mathbf{F}$ is nonsingular then $\mathbf{F} \geq 0$ is equivalent to $\mathbf{F} > 0$.

5. Stochastic o and O symbols. The notation $o_P(1)$ denotes a sequence of random vectors that converges to zero in probability. The expression $O_P(1)$ is short for a sequence that is bounded in probability (Van der Vaart 2000).

1.2 Chapter Overview

In Chapter 2 we review different data assimilation algorithms and introduce the assimilation schemes which employ linear error feedback. Mallows' C_p statistic is introduced as it is the motivation behind the work presented in Chapter 3 and we consider current linear data assimilation diagnostics.

In Chapter 3, having chosen an assimilation scheme and using Mallows' C_p statistic as motivation, we investigate the concept of the out-of-sample error for linear dynamical systems and present numerical experiments (Mallia-Parfitt & Bröcker 2016).

Chapter 4 gives details of the Kalman Filter, its asymptotic properties and an in-depth discussion on the notions of Observability and Controllability. These concepts play a crucial role in the following chapters and thus are given the attention required.

Chapter 5 is interested in how the constant feedback gain matrix chosen to minimise the expected out-of-sample error compares to the Kalman Gain for linear systems. In Chapter 6 we consider the asymptotic behaviour of the gain minimising the empirical mean of the out-of-sample error, as this is the error we can calculate in practice. To do this we treat the problem as an estimation problem and prove that the sequence of estimators is consistent.

In Chapter 7 we consider non-linear systems and use algorithms which employ linear error feedback to test the concept of the out-of-sample error for two different non-linear systems with linear observations (Mallia-Parfitt & Bröcker 2016). Numerical experiments are presented.

Concluding remarks follow in Chapter 8.

Chapter 2

Data Assimilation, Diagnostic Methods and Ridge Regression

2.1 Review of Data Assimilation Algorithms

There are many different algorithms used to achieve the goals set out by the data assimilation problem. These algorithms fall into different classes, with each class varying in the approach taken to achieve the required results, Le Dimet & Talagrand (1986). Recall that data assimilation algorithms must produce a trajectory that satisfies two requirements. The trajectory must be close to the measured observations up to some degree of accuracy and secondly, it must satisfy dynamical and/or statistical relationships satisfied by the reality.

In early data assimilation experiments, interpolations of the measured observations to grid points were done by hand, Kalnay (2001). These fields of initial conditions were then manually digitized and due to the time consuming nature of the task, it soon became evident that an automatic objective analysis was required, Charney (1951). This led to the development of spatial interpolation methods (Panofsky 1949, Gilchrist & Cressman 1954, Barnes 1964).

However, spatial interpolation of observations to gridded fields is not the only problem.

The fact that the data available (i.e the measured observations) are not enough to initialise the model, is a far greater problem that needs to be addressed. Therefore it became apparent that some additional information needs to be added into the algorithm to prepare the initial conditions for forecasts. This additional information is called the *background or a priori information*, Le Dimet & Talagrand (1986). Initially climatology was used as this first guess however eventually, a short range forecast was chosen, Kalnay (2001).

Data Assimilation algorithms work in cycles over time. In a cycle for a global model, the background field is a model forecast, x^b. To obtain the *a priori* information, the background field is interpolated to the location of the observation, and if they are different, converted from model variables to observed variables, η. Therefore, the initial guess of the observations is $h(x^b)$, where $h(\cdot)$ is the observation operator that maps model variables into observation space. The difference $\eta - h(x^b)$ are called *innovations* and the *analysis*, x^a, is obtained by

$$x^a = x^b + \mathbf{K}[\eta - h(x^b)] \tag{2.1}$$

where we simply add the innovations to the background with weights $\mathbf{K}$, determined based on the statistical error covariance of the forecast and observations.

The different classes of algorithms are based on (2.1); they differ only by the approach taken to calculate the background and the weights to produce the analysis. The early methods such as Successive Correction Method (SCM),(Cressman 1959, Barnes 1964), calculate the weights empirically in which they are a function of distance between the observation and the grid point.

In Optimal Interpolation methods, (Gandin 1965, Lahoz et al. 2010, Lorenc 1981), the matrix of weights is determined by minimizing the analysis errors at each grid points. Such methods are essentially linear regression algorithms and thus are statistical in nature.

A third class of algorithms are variational methods. These methods produce results which minimise a given measure of the distance to the observations, while also satisfying an explicit dynamical constraint, Le Dimet & Talagrand (1986). In variational approaches,

one defines a cost function proportional to the square of the distance between the analysis and both the background and the observations, Kalnay (2001). The method which uses the cost function

$$J(x) = \frac{1}{2}(x - x^b)^T \mathbf{B}^{-1}(x - x^b) + \frac{1}{2}(h(x) - \eta)^T \mathbf{R}_n^{-1}(h(x) - \eta) \tag{2.2}$$

where $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix, is known as 3D-VAR (Sasaki 1958, 1970).

The background term of the cost function is important for many reasons. Observations are not regularly distributed in time or space and not all areas in the assimilation window are observed. The covariance matrix $\mathbf{B}$ will determine how information is extrapolated from observed regions to unobserved areas. Mathematically, the problem would be underdetermined in those regions without the background term, Tremolet (2006).

The observation term in the cost function describes the discrepancy between recorded observations and their equivalent obtained from the estimated state x. The cost function J is a weighted measure of those discrepancies. This gives data a weight inversely proportional to the variance of the errors affecting them, giving more weight to accurate information, Lawless (2012).

Lorenc (1986) showed that if the cost function in (2.2) is used, the Optimal Interpolation method and 3D-VAR approach are in fact equivalent. The minimum of the cost function is obtained for $x = x^a$ (i.e the analysis) and the solution obtained by minimising (2.2) is the same as in (2.1) if the weight matrix is defined by

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{2.3}$$

The difference between Optimal Interpolation and the 3D-VAR approach is in the method used to obtain the solution. In Optimal Interpolation, the weights are determined for each grid point while in 3D-VAR the minimisation of the cost function is performed

directly, thus allowing global use of the data Kalnay (2001). The resulting solution is called the Best Linear Unbiased Estimate (BLUE), Greene (1997).

The variational approach has been extended to four dimensions (4D-VAR) by including within the cost function the distance to the observations over a time window. Formally, the problem of 4D-VAR is to find the initial state that minimises the weighted least squares distance to the background while minimising the weighted least squares distance of the model trajectory to the observation over the time interval $[t_0, t_N]$, Lawless (2012). Mathematically, we write this as an optimization problem:

Find the analysis state x_0^a at time t_0 that minimizes the function

$$J(x_0) = \frac{1}{2}(x_0 - x^b)^T \mathbf{B}^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{n=0}^{N}(h(x_n) - \eta_n)^T \mathbf{R}_n^{-1}(h(x_n) - \eta_n) \tag{2.4}$$

subject to the states x_n satisfying a specified non-linear dynamical system. In the case $N = 0$, there is no model evolution and the scheme reverts to being three-dimensional variational data assimilation (3D-VAR).

As previously stated, the BLUE analysis is equivalently obtained as a solution to the variational optimisation problem and through statistical interpolation methods. Equation (2.1) with weight matrix (2.3) is the mathematical expression of the fact that we want the analysis to depend linearly on the innovations. We also want the analysis state to be as close as possible to the true state in the sense that we want it to be a minimum variance estimate. In the case of Gaussian errors (which we assume here), the minimum variance estimate is equivalent to the maximum likelihood estimate in the probabilistic approach to understanding the data assimilation problem. See Appendix A for details.

The Kalman Filter The Kalman Filter is a sequential method used to assimilate observations over time. A more in depth analysis of the Kalman Filter can be found in Chapter 4. It is an extension to the BLUE concept described earlier in which the background is provided by a forecast that starts from the previous analysis. Whereas

4D-Var assimilates all the observations at once in the assimilation time window, the Kalman Filter steps through the observations sequentially, producing the optimal analysis each time. A feature of the Kalman Filter is the forecast of the covariance matrices, which we denote by Γ_n for the analysis error covariance matrix at time t_n and Σ_n for the forecast error covariance matrix at time t_n, Jazwinski (1970).

Suppose we have the following linear system

$$x_{n+1} = \mathbf{A}_n x_n + q_n \tag{2.5}$$

where q_n is an unbiased gaussian error with covariance matrix $\mathbf{Q}_n$ with linear observations

$$\eta_n = \mathbf{H} x_n + r_n \tag{2.6}$$

where r_n is unbiased gaussian error with covariance matrix $\mathbf{R}_n$. Then the Kalman Filter algorithm is as follows:

State Forecast	$\hat{z}_n = \mathbf{A}_{n-1} z_{n-1}$
Error Covariance Forecast	$\Sigma_n = \mathbf{A}_{n-1} \Gamma_{n-1} \mathbf{A}_{n-1} + \mathbf{Q}_{n-1}$
Kalman Gain	$\mathbf{K}_n = \Sigma_n \mathbf{H}_n^T (\mathbf{H}_n \Sigma_n \mathbf{H}_n^T + \mathbf{R}_n)^{-1}$
State Analysis	$z_n = \hat{z}_n + \mathbf{K}_n (\eta_n - \mathbf{H}_n \hat{z}_n)$
Analysis Error Covariance	$\Gamma_n = (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \Sigma_n$

(2.7)

This can be generalised to have non-linear model and observation operators in which case it is called the Extended Kalman Filter, Anderson & Moore (1979). There are many similarities between 4D-Var and the Kalman Filter however it is important to understand the differences between them. 4D-Var is cheaper computationally and it is more optimal inside the time interval for optimisation since it uses all the observations at once. However, 4D-Var assumes that the model is perfect (i.e. $\mathbf{Q} = 0$) and it can only be run for a finite

time interval while the Kalman Filter can, in principle, be run indefinitely. The Kalman Filter also provides an estimate of any uncertainty in the final analysis whereas 4D-Var does not.

All the schemes mentioned above fall into the category of algorithms that employ linear error feedback. The term 'linear error feedback' refers to the fact that the analysis depends linearly on the innovations.

Regardless of which scheme is used, the performance of the algorithm needs to be evaluated and this is done by mapping the obtained trajectory (the analysis) into observation space to compare it with the observations which were already used to obtain the trajectory in the first place. The trouble is that just because the trajectory produced follows the observations well, it does not mean that it is a good trajectory. It is possible that the algorithm is simply reproducing the observations without picking up any of the underlying dynamics.

The easiest solution would be to use independent observations from the same period and region as the original data and compare the obtained trajectory with these independent observations. Such measurements however are hardly ever available. The aim of this thesis is to find a way to analyse the true performance of data assimilation algorithms which employ linear error feedback. We first investigate current tools and diagnostics available to assess the performance of data assimilation algorithms.

2.2 Linear DA Diagnostics

Since most operational assimilation schemes are based on the variational formalism (Courtier & Talagrand 1987), the tools available to evaluate the performance of the algorithms rely on this formalism. However, as we have seen, the variational approach to the problem is similar to the other methods.

Variational algorithms rely on the theory of least-variance linear statistical estimation

(Talagrand 1997). The pieces of information used in these schemes are given by observation and background estimates of the state. The analysis systems are thusly dependent on appropriate statistics for both the observation and background errors. One source of information on these errors is contained in the statistics of the innovations.

For linear data assimilation there exist two innovation based diagnostics we can use to determine an optimal linear analysis. These are the χ^2 test (Ménard & Chang 2000*b*) and Desroziers diagnostic (Desroziers et al. 2005).

2.2.1 χ^2 Diagnostic

The χ^2 diagnostic is a measure of consistency between the variances of random variables. In data assimilation, the random variable is an innovation, i.e. the difference between the observations and the model equivalent, at the same time and location. This diagnostic for data assimilation has been studied by many including, Ménard & Chang (2000*a*) and Bennett & Thorburn (1992).

For data assimilation the χ^2 test is defined as

$$\chi^2 = \mathbf{d}^T \Gamma^{-1} \mathbf{d}, \quad \mathbf{d} = \eta - \mathbf{H}\hat{z} \tag{2.8}$$

where $\mathbf{d}$ is the innovation vector. The innovation covariance, which we denote by Ξ, is defined by

$$\Xi = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} \tag{2.9}$$

is the innovation covariance, $\mathbf{B}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix and $\mathbf{H}$ is the observation operator.

The expected value of χ^2 is given by

$$\mathbb{E}(\chi^2) = \mathbb{E}(\mathbf{d}^T \Xi^{-1} \mathbf{d}) = \mathrm{tr}(\Xi^{-1}\overline{\Xi}) \tag{2.10}$$

where $\overline{\Xi} = \mathbb{E}(\mathbf{d}\mathbf{d}^T)$ is the sample covariance of the innovations. If $\Xi = \overline{\Xi}$ then the expected value becomes

$$\mathbb{E}(\chi^2) = d \tag{2.11}$$

where d is the dimension of the observation space. Note that the condition $\Xi = \overline{\Xi}$ is a necessary but not sufficient condition for (2.11) to hold. Equation (2.11) must hold in an optimal linear analysis.

2.2.2 Desroziers Diagnostic

As an extension to the χ^2 diagnostic Desroziers & Ivanov (2001) developed a method to tune observation error and background error parameters. Desroziers et al. (2005) proposed a more direct approach to estimate observation and background error parameters. It involves four consistency checks: Consistency diagnostic on innovations, background error, observation errors and analysis errors.

Denote the following relations:

$$\begin{aligned} \mathbf{d}_b^o &= \eta - \mathbf{H}\hat{z} \\ \mathbf{d}_b^a &= \mathbf{HK}(\eta - \mathbf{H}\hat{z}) \\ \mathbf{d}_a^o &= (1 - \mathbf{HK})(\eta - \mathbf{H}\hat{z}). \end{aligned} \tag{2.12}$$

Then the four diagnostics are given by

$$\begin{aligned} &\mathbb{E}[\mathbf{d}_b^o(\mathbf{d}_b^o)^T] = \mathbf{HBH}^T + \mathbf{R}, \qquad &&\mathbb{E}[\mathbf{d}_b^a(\mathbf{d}_b^o)^T] = \mathbf{HBH}^T \\ &\mathbb{E}[\mathbf{d}_a^o(\mathbf{d}_b^o)^T] = \mathbf{R}, &&\mathbb{E}[\mathbf{d}_b^a(\mathbf{d}_a^o)^T] = \mathbf{H}\Gamma\mathbf{H}^T \end{aligned} \tag{2.13}$$

where Γ is the analysis error covariance matrix defined by $\Gamma = \mathbf{B} - \mathbf{KHB}$. The above conditions should be fulfilled in an optimal linear analysis. The four conditions above require

$$\mathbb{E}[\eta - \mathbf{H}\hat{z}][\eta - \mathbf{H}\hat{z}]^T = \mathbf{HBH}^T + \mathbf{R} \tag{2.14}$$

to be satisfied; which is the same as χ^2 diagnostic requirement. This condition is necessary but not sufficient for an optimal linear analysis.

Diagnostic Methods provide tools to answer questions of the form: How much would the analysis change if one single influential observation were removed? How much information is extracted from the available data? How large is the influence of the latest data on the analysis? How much influence is due to the background?

In an attempt to answer such questions, Cardinali et al. (2004) derive statistical concepts of ordinary least squares regression to corresponding statistical data assimilation schemes. They show that the observation and background influences complement each other. For any observations, either very large or small influence could be sign of inadequacy in the assimilation. This is similar to the discussion on Ridge Regression given next, in Section 2.3, which leads on to a discussion on Mallows' C_p statistic and the out-of-sample performance of processes.

2.3 Linear Ridge Regression

As we have already stated, the purpose of this thesis is to assess the performance of data assimilation algorithms keeping in mind that even though the obtained trajectory might follow the observations, that does not mean the algorithm is working as expected. There is always the possibility that the algorithm is simply reproducing the results without picking up any of the underlying dynamics.

This problem however, appears everywhere in statistics and there are various ways to deal with it; BIC, AIC and Mallows' C_p Statistic are three examples (Hastie et al. 2009). In particular we shall focus on the C_p statistic and hence give after a brief discussion on this topic next. In order to understand the C_p statistic in more detail we first give a quick overview of regularised linear regression.

Given a vector of inputs $\mathbf{X}^T = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p)$, we want to predict the output Y (a

univariate random variable) via the linear regression model

$$Y = X\beta + \epsilon \tag{2.15}$$

where β are the regression coefficients. We want to fit the linear model to a set of training data and we do this by implementing a regularised least squares approach. Regularized regression finds the β that minimises the penalized residual sum of squares given by

$$\mathrm{RSS}(\beta) = \sum_{i=1}^{N}(y_i - x_i^T\beta)^2 + \lambda \left\|\beta\right\|^2 \tag{2.16}$$

which we write in matrix form as

$$\mathrm{RSS}(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta \tag{2.17}$$

where $\lambda \geq 0$ is a complexity (or ridge) parameter, $\mathbf{X}$ is an $N \times p$ matrix and $\mathbf{y}$ is an N-vector of the outputs of the training set. Differentiating with respect to β and setting equal to zero yields the unique solution $\hat{\beta}^{ridge}$ given by

$$\hat{\beta}^{ridge} = (\mathbf{X}^T\mathbf{X} - \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.18}$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. The predicted values are then defined by

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}. \tag{2.19}$$

Ridge regression shrinks the coefficients by imposing a penalty on their size. The complexity parameter λ controls the amount of shrinkage; the larger the value of λ, the greater the shrinkage. Notice that with the choice of the quadratic penalty $\beta^T\beta$ in (2.18), the ridge regression solution is a linear function in $\mathbf{y}$.

The motivation for ridge regression is that even if the input matrix $\mathbf{X}$ is not of full

rank, the problem is non-singular. This is because the solution adds a positive constant to the diagonal of $\mathbf{X}^T\mathbf{X}$ before inversion. The Singular Value Decomposition of the input matrix $\mathbf{X}$ gives us some further insight into the nature of ridge regression. See Appendix B for details.

2.3.1 Mallows' C_p Statistic

Suppose we have the model

$$Y = f(X) + \epsilon \tag{2.20}$$

where $\mathbb{E}\epsilon = 0$ and $\mathbb{E}\epsilon^2 = \sigma^2$. We can derive an expression for the expected prediction error of the regression fit $\hat{f}(X)$ at some input point x_0. Using squared error loss we see that

$$\begin{aligned} \text{Err}(x_0) &= \mathbb{E}[(Y - \hat{f}(x_0))^2 \,|X = x_0] \\ &= \mathbb{E}[(Y - f(x_0))^2] + \mathbb{E}[(f(x_0) - \mathbb{E}\hat{f}(x_0))^2] + \mathbb{E}[(\hat{f}(x_0) - \mathbb{E}\hat{f}(x_0))^2] \\ &= \sigma^2 + \text{Bias}^2(\hat{f}(x_0)) + \text{Var}(\hat{f}(x_0)) \end{aligned} \tag{2.21}$$

This expression suggests that there will be a trade-off between the bias and variance. For a linear model fit,

$$\hat{f}(X) = X\hat{\beta} \tag{2.22}$$

where $\hat{\beta}$ is the parameter vector fitted by least squares, we have

$$\text{Err} = \mathbb{E}[Y - \hat{f}(X)] = \sigma^2 + \{(I - \mathbf{H})f(X)\}^2 + \text{tr}(\mathbf{H})\sigma^2 \tag{2.23}$$

where $\mathbf{H}$ is the hat matrix defined by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T. \tag{2.24}$$

Equation (2.23) is called the *expected prediction error or test error* which is the expected

error over an independent test sample. The *training error*, which is the error over the test sample, is expressed as

$$\text{err} = \sigma^2 + \{(I - \mathbf{H})f(X)\}^2 - \text{tr}(\mathbf{H})\sigma^2, \tag{2.25}$$

which, together with (2.23), gives us a relationship between the test error and the training error. The degrees of freedom is defined by

$$\text{df} = \text{tr}(\mathbf{H}) = d \tag{2.26}$$

where d is the dimension (see Appendix B for details) and so, if we let $\text{Err}_{\text{IN}} = \frac{1}{N}\sum \text{Err}$ and $\overline{\text{err}} = \frac{1}{N}\sum \text{err}$, we have,

$$\overline{\text{err}} - \text{Err}_{\text{IN}} = -\frac{2\text{tr}(\mathbf{H})\sigma^2}{N} = -\frac{2d\sigma^2}{N}. \tag{2.27}$$

Once we have an estimate $\hat{\sigma}^2$ to σ^2, the noise variance, we write

$$C_p = \overline{\text{err}} + \frac{2d\hat{\sigma}^2}{N} \tag{2.28}$$

which is a version of the C_p statistic. The C_p statistic can be expressed in a different way as in James et al. (2013), however we shall use this formulation as it illustrates the point we are trying to make very well.

Using this criterion we adjust the training error by a factor proportional to the number of basis function used (i.e the number of degrees of freedom, d). Typically the training error, $\overline{\text{err}}$, will be less than the prediction error, Err, because the same data is being used to fit the method and assess its error. A fitting method typically adapts to the training data and hence the training error will be an optimistic estimate of the test error.

2.3.2 Numerical Simulations

We present some numerical simulations using the theory described above. We implement the ridge regression method on a given set of data with known coefficient vector so that we will be able to see the idea of the Bias-Variance Trade-off.

Recall that in ridge regression we want to minimise the residual sum of squares defined by,

$$\mathrm{RSS}(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^T\beta \tag{2.29}$$

where $\lambda \geq 0$ is a complexity (or ridge) parameter, $\mathbf{X}$ is an $N \times p$ matrix and $\mathbf{y}$ is an N-vector of the outputs of the training set. Differentiating with respect to β and setting equal to zero yields the unique solution $\hat{\beta}$ given by

$$\hat{\beta}^{ridge} = (\mathbf{X}^T\mathbf{X} - \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.30}$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. Note however that the intercept β_0 is left out of the penalty term. So, more accurately, the function we want to minimise is given by

$$\mathrm{RSS}(\beta) = \sum_{i=0}^{N}(y_i - x_i^T\beta) + \lambda\sum_{i=1}^{p}\beta^2 + \beta_0^2. \tag{2.31}$$

Figure (2.1(a)) shows the coefficient paths as lambda is increased. Notice that the intercept is not affected by the penalty and while the other coefficients get shrunk to zero as lambda increases, it settles down to a non-zero constant.

There are two errors we are interested in. The first is the error between our targets, $\mathbf{y}$, and the fitted values, $\hat{\mathbf{y}}$ and the second is between the true coefficient values and our estimated coefficient values. We consider first the error between our targets and the fitted values. As explained above, we consider this error by considering both the training error and the test error (red diamonds). A plot of these is shown in figure (2.1(b)). It is evident that, as expected, the training error (blue circles) is smaller than the test error and that it

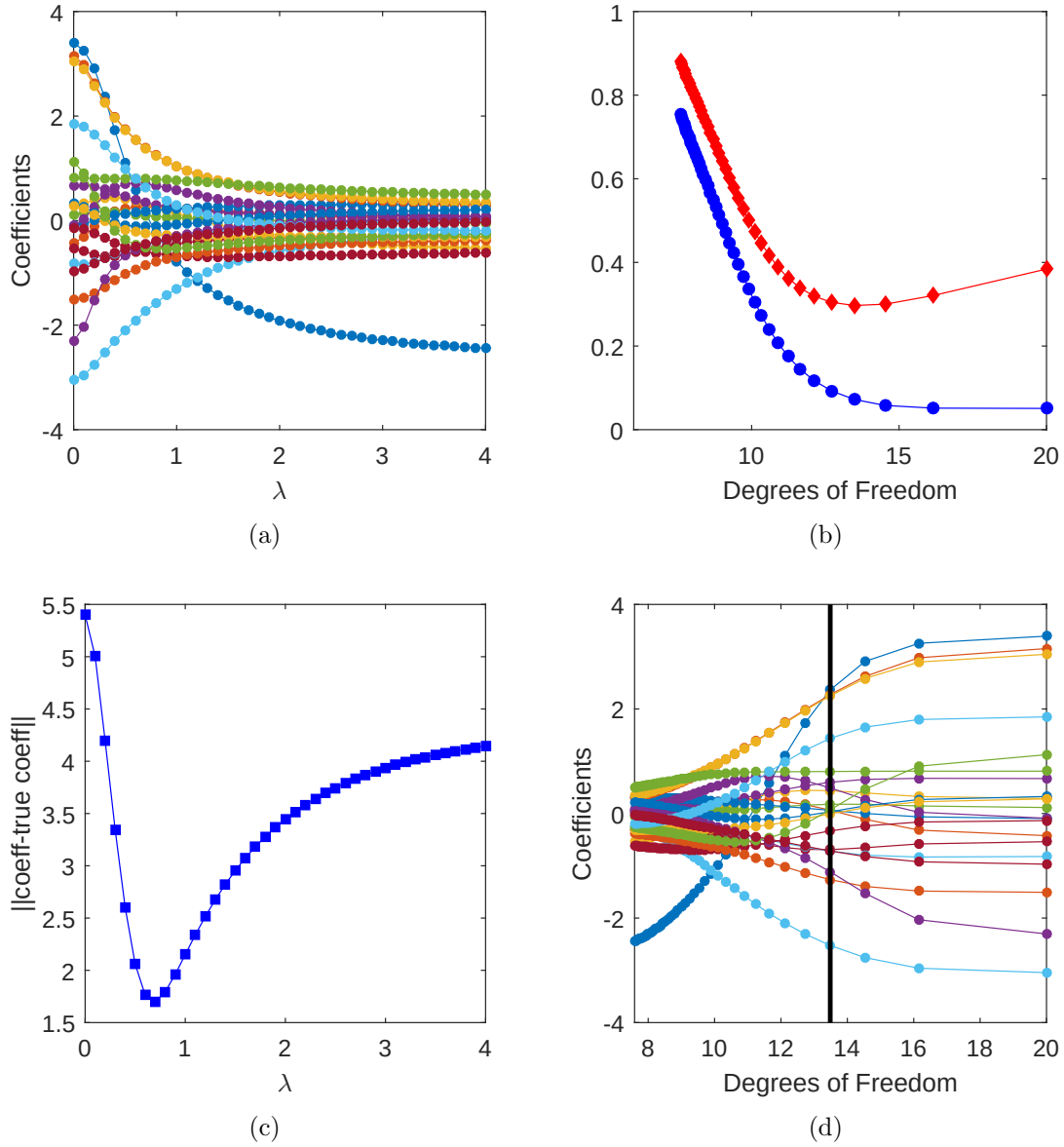


Figure 2.1: Plots produced in the ridge regression numerical experiments. The regression coefficient paths are plotted against the ridge parameter λ in fig. 2.1(a) and against the degrees of freedom in fig. 2.1(d). The vertical line draws attention to the optimal value of the degree of freedom. The test and prediction errors are shown in fig. 2.1(b) in blue circles, red diamonds respectively. The error between the true coefficient and the estimated coefficient is shown in fig. 2.1(c) in blue squares. This latter plot illustrates the trade-off between the bias and the variance. The minimum of the curve provides the optimal value of the degree of freedom that minimises the test error.

gets smaller as the model complexity is increased. The test error on the other hand begins to increase again. This increase in the test error is due to the increase in the variance. There is some intermediate model complexity that gives minimum expected test error. In this example, this minimum is achieved when df = degrees of freedom ≈ 13.5 which corresponds to a value of lambda = 0.3.

Figure (2.1(d)) shows the coefficient paths plotted against the degrees of freedom with a vertical line drawn at df = 13.5. For the purposes of model selection we should take the model with lambda value = 0.3 as this gives us minimum prediction test error. This method involves in-sample prediction error which is achieved by estimating $2d\sigma^2/N$ and adding it to the training error $\overline{\text{err}}$. Note that this only works for estimates that are linear in their parameters. In this case we have used the C_p statistic (2.28) where we adjust the training error by a factor proportional to the number of basis functions used.

Since in this example we have the true coefficient values we can determine the error between the true vector and our estimated vector of coefficients. Figure (2.1(c)) shows this error as a function of lambda. We can see that there is an initial decrease which corresponds to the decrease in variance and then an increase in the error, which corresponds to an increase in bias. We can see here, there is a trade-off between bias and variance. As the model becomes more complex, it uses training data more and thus is able to adapt to more complicated underlying structures. Hence there is a decrease in bias but an increase in variance.

In the next chapter we use the C_p statistic as inspiration to assess the performance of data assimilation algorithms by evaluating the so called out-of-sample error, analogous to the test error. We hope that evaluating this error will give us a more realistic idea of the model performance. The trade off between the bias and the variance here will be replaced, when used in the setting of data assimilation, by a trade off in the strength of the coupling introduced between the underlying model of the reality and the assimilation scheme to obtain an optimal analysis. In the above, we see that for model selection purposes there is

an optimal λ which minimises the expected prediction error. Similarly for data assimilation algorithms we expect to find an optimal weight or coupling matrix to achieve optimal performance.

Chapter Summary In this chapter we have seen a brief review of the concept of data assimilation. We also saw some examples of the different data assimilation algorithms available. These algorithms differ in how the background field and gain matrix are calculated.

We briefly discussed current diagnostics used in the setting of data assimilation to determine how good the algorithms truly are. We considered two diagnostics that enable us to determine if the analysis determined is optimal.

An explanation of ridge regression and Mallows' C_p statistic was also presented. The work in this section is the motivation behind the work presented in Chapter 3 and Mallia-Parfitt & Bröcker (2016) where we adapt these concepts to be used for data assimilation algorithms.

Chapter 3

The Out-of-Sample Error for Data Assimilation

In this chapter we shall investigate the out-of-sample error for data assimilation algorithms which employ linear error feedback. To implement the out-of-sample error in data assimilation, we assume that the observations obtained are corrupted by random noise as done in Mallia-Parfitt & Bröcker (2016). If these observations are then assimilated into a dynamical model, the results should be close to theoretical observations with independent errors. These theoretical observations must be from the same underlying flow patterns but with independent errors. If the results are not close to these hypothetical observations, then the model will not be reproducing the underlying dynamics of the model.

The out-of-sample error simply gives us an assessment of how close the results are to theoretical observations. The tracking error, which is the error with respect to the measured observations, is not a good estimate of the out-of-sample error. This is because the measured observations have already been used to find the solution and so the tracking error tends to misestimate the true out-of-sample performance. On average, we can think of the out-of-sample error as the error with respect to the true observations plus a constant; the variance of the observations (Mallia-Parfitt & Bröcker 2016).

The expression we develop to calculate the out-of-sample error can be estimated using terms that are readily available. Specifically we show that the out-of-sample error is the sum of the tracking error and a term which we call the optimism. This optimism gives us a representation of how the model and observations depend on each other and it quantifies how much the tracking error misestimates the out-of-sample error. The derived expression is very similar to the C_p statistic used in model selection in statistical learning, see Chapter 2 and Hastie et al. (2009), Efron (2004). The optimism takes a very simple form if we assume that the model employs a linear error feedback.

Numerical experiments are presented to validate the expression for the out-of-sample error and the optimism. Further numerical results illustrate the convergence of the gain matrix that minimises the out-of-sample error, to the asymptotic Kalman Gain in the limit of large observational windows. The experiments show that the technique can be used in situations where the feedback gain matrix is completely unspecified and also in situations where it has a pre-determined structure as done in Mallia-Parfitt & Bröcker (2016).

3.1 Estimating the Optimism

Suppose we have observations, $\eta_n \in \mathbb{R}^d$ which are given by

$$\eta_n = \zeta_n + \sigma r_n \tag{3.1}$$

where the desired signal, ζ_n, is made up of non random, unknown parameters which we try to estimate. The observation errors, r_n are assumed to be serially independent errors with mean $\mathbb{E}r_n = 0$ and variance $\mathbb{E}r_n r_n^T = \mathbb{1}$.

Data assimilation is the procedure by which trajectories $\{z_n \in \mathbb{R}^D\}$ are computed with the help of a dynamical model and observations, η_n. These trajectories should reproduce the observations up to some degree of accuracy. We express this latter part of the procedure formally as: There exists a function $h : \mathbb{R}^D \to \mathbb{R}^d$ so that the output $y_n = h(z_n)$ is close to

the observations η_n up to some degree of accuracy. The exact structure of the model is not important at this stage.

We measure the deviation of the output from the observations by means of the tracking error,

$$E_T = \mathbb{E}[y_n - \eta_n]^2. \tag{3.2}$$

To define the out-of-sample error we assume that we have another set of observations, η'_n, which are given by

$$\eta'_n = \zeta_n + \sigma r'_n \tag{3.3}$$

where r_n has the same stochastic properties as r_n but is independent from r_n, i.e $\mathbb{E}r_n r'_n = 0$. The desired signal, ζ_n, is the same as in η_n. Therefore we can define the out-of-sample error as

$$E_S = \mathbb{E}[y_n - \eta'_n]^2 = \mathbb{E}[y_n - \zeta_n]^2 + \sigma^2, \tag{3.4}$$

where the second equation is obtained by substituting (3.3) into $\mathbb{E}[y_n - \eta'_n]^2$ and noting that r'_n is uncorrelated with both y_n and ζ_n. The output error (first term on the right hand side of (3.4)), is ultimately the error we want to minimise with our choice of parameters. However, since the observations are corrupted, the output error is a difficult quantity to determine.

The tracking error is a bad estimate of the output error and can easily be cheated. It is not difficult to design an algorithm that produces zero tracking error by simply using the observations themselves as the output. That is any data assimilation scheme which satisfies $y_n = \eta_n$, $n = 1, \ldots, N$ achieves optimal performance with respect to the tracking error as a performance measure.

Using this idea of out-of-sample error it is possible to get a handle on the output error as it is evident that the out-of-sample error is simply the output error added to the variance of the observational noise. The relationship between the tracking and out-of-sample errors

is given by

$$\mathbb{E}[y_n - \eta_n']^2 = \mathbb{E}[y_n - \eta_n]^2 + 2\sigma\mathbb{E}[y_n^T r_n]. \tag{3.5}$$

This is seen by substituting (3.1) into (3.2) and noting that $\mathbb{E}[(y_n - \zeta_n)r_n] = \mathbb{E}[y_n r_n]$ since ζ_n is not a random variable. The term $2\sigma\mathbb{E}[y_n r_n]$ is called the *optimism*. The optimism should be understood as a correlation between r_n filtered through y_n and r_n itself. It is a measure of how much the tracking error misestimates the out-of-sample error.

3.2 Data Assimilation through Synchronisation

Synchronisation between dynamical systems has been studied for some time, see for example Pikovsky et al. (2001); Huijberts et al. (1999); Boccaletti et al. (2002). Synchronisation in the setting of data assimilation has also been studied, see Bröcker & Szendro (2012); Szendro et al. (2009); Yang et al. (2006).

As motivation suppose that the reality is given by the non linear dynamical system

$$\begin{aligned} x_{n+1} &= \tilde{f}(x_n) \\ \zeta_n &= \tilde{h}(x_n) \end{aligned} \tag{3.6}$$

where $x_n \in \mathbb{R}^D$ is referred to as the state and $\zeta_n \in \mathbb{R}^d$ are the true observations. For this non linear dynamical system we construct a sequential scheme

$$\begin{aligned} \hat{z}_{n+1} &= f(z_n) \\ z_{n+1} &= \hat{z}_{n+1} - \mathbf{K}_n(h(\hat{z}_{n+1}) - \eta_{n+1}) \\ y_n &= h(z_n) \end{aligned} \tag{3.7}$$

where $\mathbf{K}_n$ is a $D \times d$ coupling matrix which may depend on the observations $\eta_1, \ldots \eta_{n-1}$ but not on η_n and y_n is the model output where we hope that $y_n \cong \zeta_n$. Here f and h are approximations to the functions $\tilde{f}$ and $\tilde{h}$, respectively. The function $f(z_n)$ describes the model dynamics and is thought of as capturing our *a priori* knowledge of the observations.

The coupling introduced in this scheme creates a linear feedback, in the sense that the error between $y_n = h(\hat{z}_n)$ and the observations η_n, i.e the innovation, is fed back into the model.

Synchronisation refers to a situation in which, due to coupling, the error $y_n - \eta_n$ becomes small asymptotically irrespective of the initial conditions for the model (Pikovsky et al. 2001). Often a control theoretic approach is taken to determine conditions which guarantee the model output, $y_n = h(z_n)$, converging to the observations, η_n or even $z_n \to x_n$ which ultimately, is what we want to achieve.

Consider now the optimism as in (3.5). In order to calculate the optimism, assume that the function $h(x_n)$ is linear so that $h(x_n) = \mathbf{H}x_n$, where $\mathbf{H}$ is a $d \times D$ matrix. Then we can re-write the system (3.7) as

$$\begin{aligned} z_{n+1} &= f(z_n) - \mathbf{K}_n(h(f(z_n)) - \eta_{n+1}) \\ &= (\mathbb{1} - \mathbf{K}_n\mathbf{H})f(z_n) + \mathbf{K}_n(\zeta_{n+1} + \sigma r_{n+1}). \end{aligned} \tag{3.8}$$

We have seen in (3.5) that the tracking and out-of-sample errors are related by the optimism, $2\sigma\mathbb{E}[y_n r_n]$. For this particular system (3.7) the explicit expression for the optimism is given by

$$2\sigma\mathbb{E}[y_n^T r_n] = 2\sigma\mathbb{E}[(\mathbf{H}z_n)^T r_n] \tag{3.9}$$

$$= 2\sigma\mathbb{E}[\{\mathbf{H}(\mathbb{1} - \mathbf{K}_n\mathbf{H})f(z_{n-1}) + \mathbf{H}\mathbf{K}_n(\zeta_n + \sigma r_n)\}^T r_n] \tag{3.10}$$

$$\begin{aligned} &= 2\sigma\mathbb{E}[(\mathbf{H}(\mathbb{1} - \mathbf{K}_n\mathbf{H})f(z_{n-1}))^T r_n] \\ &\quad + 2\sigma\mathbb{E}[(\mathbf{H}\mathbf{K}_n\zeta_n)^T r_n] + 2\sigma^2\mathbb{E}[(\mathbf{H}\mathbf{K}_n r_n)^T r_n] \end{aligned} \tag{3.11}$$

$$= 2\sigma^2\mathbb{E}[r_n^T\mathbf{K}_n^T\mathbf{H}^T r_n] \tag{3.12}$$

$$= 2\sigma^2\mathrm{tr}(\overline{\mathbf{K}}_n^T\mathbf{H}^T\mathbb{E}[r_n r_n^T]) \tag{3.13}$$

where $\overline{\mathbf{K}}_n = \mathbb{E}[\mathbf{K}_n]$. The first two equalities, (3.9) and (3.10), are obtained by substi-

tuting the relevant information while (3.11) is obtained by simply expanding the previous equation. The derivation from (3.11) to (3.12) requires some explanation. Notice first that only the third term of (3.11) survives. The first term is equal to zero because $f(z_{n-1})$ and $\mathbf{K}_n$ are uncorrelated with r_n since they only depend on observations up to $n-1$. The second term is also equal to zero because ζ_n is not a random variable and because the coupling matrix $\mathbf{K}_n$ is uncorrelated with r_n. Therefore, we are only left with the third term of (3.11) in (3.12). Since $\mathbb{E}(r_n r_n^T) = \mathbb{1}$, (3.13) implies that

$$\mathbb{E}[y_n - \eta_n]^2 = \mathbb{E}[y_n - \eta'_n]^2 - 2\sigma^2 \mathrm{tr}(\overline{\mathbf{K}}_n^T \mathbf{H}^T). \tag{3.14}$$

In the case when $d = 1$, which is the case we consider in the numerical experiment later, this reduces to

$$\mathbb{E}[y_n - \eta_n]^2 = \mathbb{E}[y_n - \eta'_n]^2 - 2\mathbf{H}\overline{\mathbf{K}}_n \sigma^2. \tag{3.15}$$

Equation (3.15) has this simple form because of the linearity assumption in the observation operator. It tells us that to estimate the out-of-sample error, we need to estimate the optimism and then add it to the tracking error. This means that, in theory, it is possible to approximate the out-of-sample error using information that is readily available.

This is particularly useful as it is not necessary to know anything about the dynamical noise in the model. The terms required to calculate the out-of-sample error are all needed in the scheme itself; these include the gain matrix, observational noise variance and the system matrices. This is advantageous and as such can be applied operationally as no information about the underlying dynamical noise is required.

3.3 Numerical Experiment I: Linear Map

In this first numerical example the following experimental setup was used: The reality is given by

$$x_{n+1} = \underbrace{\begin{bmatrix} -1 & 10 \\ 0 & 0.5 \end{bmatrix}}_{\mathbf{A}} x_n + \rho q_{n+1} \tag{3.16}$$

with corresponding observations

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{3.17}$$

where $\mathbf{H} = [1 \quad 0]$, and $\zeta_n = \mathbf{H}x_n$. We assume that the dynamical model and observations are corrupted by random noise. For these experiments we have $x_n \in \mathbb{R}^2$ and $\eta_n \in \mathbb{R}$. The model and observation errors, q_n and r_n respectively, are assumed to be independent gaussian errors with mean 0 and variance 1. The notation $\rho \in \mathbb{R}^{d\times d}$ and $\sigma \in \mathbb{R}^{d\times d}$ represent the standard deviation of the model and observational noise respectively. Their values are taken to be between 0 and 1 (both not included).

Here we consider data assimilation by means of synchronisation so we set up an observer analogous to our sequential scheme (3.7),

$$z_{n+1} = \hat{z}_{n+1} + \mathbf{K}_n(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}), \qquad y_n = \mathbf{H}z_n \tag{3.18}$$

where

$$\hat{z}_{n+1} = \underbrace{\begin{bmatrix} -1 & 10 \\ 0 & 0.5 \end{bmatrix}}_{\mathbf{A}} z_n. \tag{3.19}$$

In this case the model is coupled to the observations through a linear coupling term which is dependent on the difference between the actual output and the output value expected based on the next estimate of the state. For these experiments we will take the

coupling matrix $\mathbf{K}_n$ to be constant, so from here on in we write $\mathbf{K}_n = \mathbf{K}$.

We need to choose the matrix $\mathbf{K}$ appropriately so that we can vary the coupling strength. If the coupling is too strong the observations will be tracked too closely and if the coupling is too weak the observations are tracked badly or not at all.

The error dynamics in this example are given by

$$\begin{aligned} e_{n+1} &= x_{n+1} - z_{n+1} \\ &= (\mathbf{A} - \mathbf{KHA})e_n + \mathbf{K}r_{n+1} - (\mathbb{1} - \mathbf{KH})q_{n+1}. \end{aligned} \tag{3.20}$$

Since the noisy part of the error dynamics is stationary, synchronisation can be guaranteed if the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ all lie within the unit circle. In order to synchronise the model and observer we use a result from control theory, for which we need a few definitions. Let $\mathbf{HA} = \mathbf{C}$ so that the error dynamics are given by $e_{n+1} = (\mathbf{A} - \mathbf{KC})e_n$ plus the stationary terms. A pair of matrices $(\mathbf{A}, \mathbf{C})$ is called *observable* when the observability matrix

$$\mathcal{O} = [\mathbf{C} \quad \mathbf{CA} \quad \mathbf{CA}^2 \quad \dots \quad \mathbf{CA}^{D-1}]^T \tag{3.21}$$

has full rank. If this condition holds then the poles of the matrix $(\mathbf{A} - \mathbf{KC})$ can be placed anywhere by proper selection of $\mathbf{K}$. In particular they can be placed within the unit circle where they are stable and so the error e_n will tend to zero asymptotically (Dorf & Bishop 2005).

In our example, $x_n \in \mathbb{R}^2$ so our observability matrix is

$$\mathcal{O} = [\mathbf{HA} \quad \mathbf{HA}^2]^T. \tag{3.22}$$

It is straightforward to check that the linear system we are working with here is

observable even though $\mathbf{A}$ itself is not stable. Since

$$\mathbf{H} = [1 \quad 0] \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} -1 & 10 \\ 0 & 0.5 \end{bmatrix} \tag{3.23}$$

it follows that

$$\mathbf{HA} = [-1 \quad 10] \quad \text{and} \quad \mathbf{HA}^2 = [1 \quad -5] \tag{3.24}$$

and hence the observability matrix defined by (3.22), in this case, has full rank.

The appropriate $\mathbf{K}$ for a desired characteristic polynomial, $q(\lambda)$, of the matrix $(\mathbf{A} - \mathbf{KHA})$ follows from Ackermann's Formula (Dorf & Bishop 2005) which is given by

$$\mathbf{K} = q(A)\mathcal{O}^{-1}[0 \dots 1]^T. \tag{3.25}$$

where $\mathcal{O}$ is the observability matrix. Suppose that the desired characteristic equation is given by

$$q(\lambda) = (\lambda + \alpha)(\lambda - \alpha) \tag{3.26}$$

so that $\lambda_1 = -\lambda_2$ and $|\lambda_1| = |\lambda_2| = \alpha$. Then Ackermann's formula yields

$$\mathbf{K} = \begin{bmatrix} 1 - 2\alpha^2 \\ 0.05 - 0.2\alpha^2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{HK} = 1 - 2\alpha^2. \tag{3.27}$$

From (3.27) we see that as $\alpha \to 0$, $\mathbf{HK} \to 1$. Thus,

$$y_n = \mathbf{H}z_n = (\mathbb{1} - \mathbf{HK})\mathbf{H}\hat{z}_n + \mathbf{HK}\eta_n \to \eta_n, \tag{3.28}$$

meaning that the data assimilation scheme simply replaces y_n with η_n, implying that the tracking error tends to zero. However this does not imply perfect data assimilation, by which we mean that the tracking tending to zero does not imply that the out-of-sample

error is also small.

From (3.15) and (3.27) we know that

$$\mathbb{E}[y_n - \eta'_n]^2 - \mathbb{E}[y_n - \eta_n]^2 = 2\sigma^2 \left(1 - 2\alpha^2\right). \tag{3.29}$$

To calculate the errors in the numerical simulation we approximate the expected value of a random variable, $\mathbb{E}[X]$, by the empirical mean squared error. Thus, (3.29) becomes

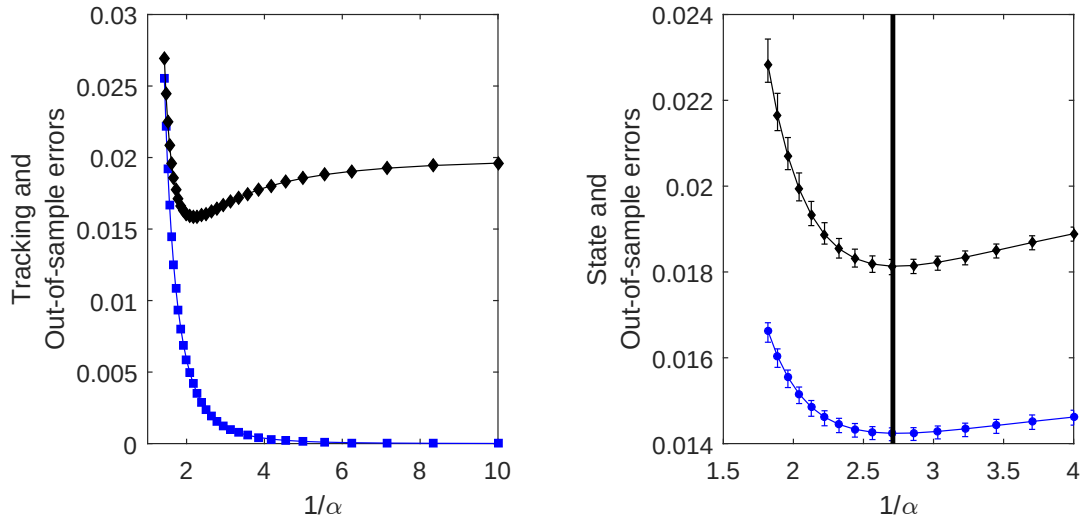
$$\frac{1}{N}\sum_{n=1}^{N}(y_n - \eta'_n)^2 - \frac{1}{N}\sum_{n=1}^{N}(y_n - \eta_n)^2 = 2\sigma^2 \left(1 - 2\alpha^2\right). \tag{3.30}$$

Any uncertainty in the calculation of the optimism will be assessed by running the experiment many times, each time changing the observational noise r_n so that the sample estimate is different every time. We then construct confidence intervals as a measure of accuracy.

The results obtained from our numerical experiment to test the theory described above are shown in Figure 3.1 and Mallia-Parfitt & Bröcker (2016). Figure 3.1(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. It is clear that the tracking error tends to zero with decreasing α. This is what we expected and is confirmed by using our analytical expression for the optimism.

It is evident from Figure 3.1(a) that while the tracking error tends to zero, the out-of-sample error initially decreases and then increases resulting in a well-defined minimum. This is because as the coupling strength increases, the observations are tracked too closely and thus the output adapts too closely to the observations resulting in an increase of the out-of-sample error; however the tracking error continues to decrease to zero. On the other hand when α is large and the coupling strength is weak, the observations are tracked poorly resulting in large tracking and out-of-sample errors.

The well defined minimum of the out-of-sample error can also be seen in Figure 3.1(b). Figure 3.1(b) shows the out-of-sample error (black diamonds) for the range of α where



(a) (b)

Figure 3.1: Figure 3.1(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. The errors are plotted against the inverse of α for $\sigma = 0.1$ and $\rho = 0.01$. Figure 3.1(b) shows a plot of the state error in blue circles and the out-of-sample error (black diamonds) for 100 realisations of the observational noise r_n with $\sigma = 0.1$. It is displayed for the range of α where the minimum occurs. The error bars represent 90% confidence intervals. The black vertical line draws attention to the minimum of the out-of-sample error.

the minimum occurs. The figure shows the out-of-sample error for 100 realisations of the observation noise r_n with $\sigma = 0.1$. The error bars represent 90% confidence intervals for each value of α with the lower bound for the errorbars plotted at the fifth percentile and the upper bound plotted at the 95th percentile.

When running data assimilation algorithms, the state error, defined by

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - x_i)^2, \tag{3.31}$$

is what we ultimately want to be minimal. However, we only have access to the observed error namely

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \eta_i)^2. \tag{3.32}$$

Due to this we consider whether minimising the out-of-sample error is equivalent to

minimising the state error. Figure 3.1(b) also shows the state error (blue circles) for $\sigma = 0.1$ and $\rho = 0.01$. Again 90% confidence intervals are plotted for every α. The black vertical line draws attention to the minimum of the out-of-sample error which coincides with the minimum of the state error. It is evident, at least in this example, that the minimising gain is the same for both errors.

3.4 Numerical Experiment II : Gain Convergence for the Linear Map

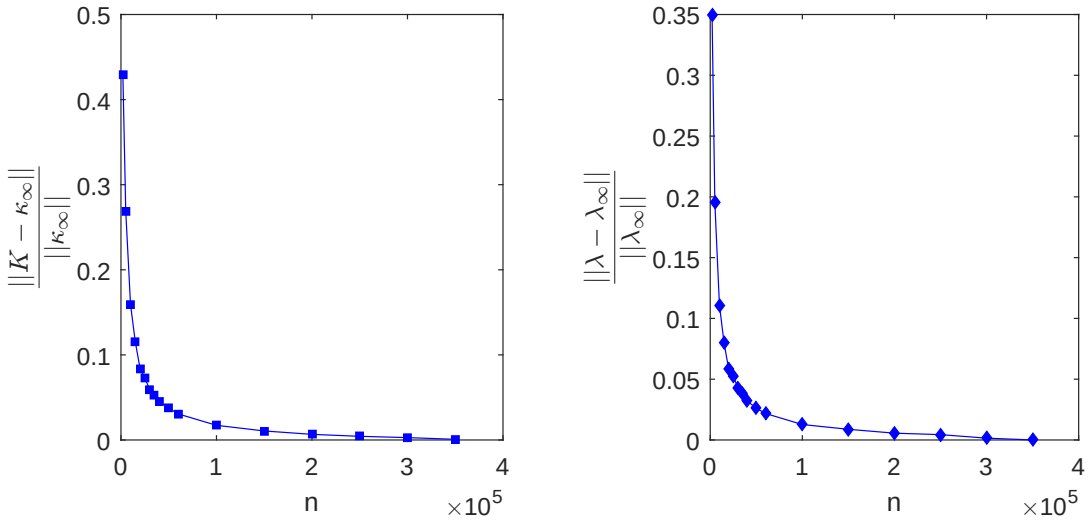
As a result of the process outlined above we are also able to determine the optimal coupling matrix, $\mathbf{K}$, to be used in the algorithm. The gain that minimises the out-of-sample error in the above experiments, is determined by arbitrarily choosing the parameter α. In order to analyse the asymptotic behaviour of this gain, we shall consider all possible gains that stabilise the system.

We ran some numerical experiments to test how the gain matrix that minimises the out-of-sample error behaves asymptotically. For the linear example in Section 3.3, the following experimental setup was used: The reality is given by the linear system (3.16) and (3.17) and the observer is set up in exactly the same way as in (3.18).

The results obtained in this experiment are shown in Figure 3.2 and Mallia-Parfitt & Bröcker (2016). The model noise is iid with $\mathbb{E}q_n = 0$, $\mathbb{E}q_n q_n^T = 1$ and $\rho = 0.01$ while for the observational noise, which was also iid with mean zero and variance one, we used $\sigma = 0.1$. The time evolution of the model which we denote by n was taken to vary between zero and 3.5×10^5. For each n the optimal gain was determined and recorded.

It is expected that the gain matrix will converge as n increases. A natural question that arises from this expectation is what the limit if that convergence is. Consider the equation

$$\Sigma_\infty = \mathbf{A}[\Sigma_\infty - \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \sigma^2)^{-1} \mathbf{H}\Sigma_\infty]\mathbf{A}^T + \mathbf{Q}. \tag{3.33}$$



(a) (b)

Figure 3.2: Figure 3.2(a) shows the convergence of the gain minimising the out-of-sample error to the asymptotic gain for increasing n. We plot the quantity $\|\mathbf{K} - \boldsymbol{\kappa}_\infty\| / \|\boldsymbol{\kappa}_\infty\|$ against n in blue squares. Figure 3.2(b) shows the quantity $\|\lambda - \lambda_\infty\| / \|\lambda_\infty\|$ against n in blue diamonds, where $\lambda = (\lambda_1, \lambda_2)$ represents the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$.

This equation describes the limit $n \to \infty$ of the covariance matrix Σ_n defined by $\Sigma_n = \mathbb{E}[(x_n - \hat{z}_n)(x_n - \hat{z}_n)^T]$. Equation (3.33) is called the Discrete Algebraic Riccati Equation (DARE). It is well known in Kalman Filter theory (see for example Anderson & Moore (1979)) that the optimal gain matrix for a linear filter is the Kalman Gain which is defined by

$$\boldsymbol{\kappa}_n = \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \sigma^2)^{-1} \tag{3.34}$$

where Σ_n is given by

$$\Sigma_n = \mathbf{A}(\Sigma_n - \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \sigma^2)^{-1} \mathbf{H}\Sigma_n)\mathbf{A}^T + \mathbf{Q}. \tag{3.35}$$

Kalman Filter theory states that for large n, the error covariance (3.35) converges to (3.33) which in turn implies that the Kalman Gain (3.34) converges to the asymptotic gain which is defined by

$$\boldsymbol{\kappa}_\infty = \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \sigma^2)^{-1} \tag{3.36}$$

The asymptotic gain, $\boldsymbol{\kappa}_\infty$, is obtained by solving the Discrete Algebraic Riccati Equation (DARE) given by (3.33) and using the solution to calculate (3.36). Using Maple's inbuilt DARE solver it is straightforward to determine the solution to this equation for the experimental setup described above. The Algebraic Riccati Equation is solved using the method described in Arnold III & Laub (1984). We expect that the constant gain matrix that minimises the out-of-sample error, also converges to the asymptotic gain.

The results obtained are shown in Figure 3.2 and Mallia-Parfitt & Bröcker (2016). Figure 3.2(a) shows a plot in blue squares of the relative error, $\|\mathbf{K} - \boldsymbol{\kappa}_\infty\| \,/\, \|\boldsymbol{\kappa}_\infty\|$ against n. It is clear that the constant gain matrix that minimises the out-of-sample (or output) error converges exponentially to the asymptotic gain. Moreover, it is illustrated in Figure 3.2(b) that the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ for each gain minimising the out-of-sample error, converge to the eigenvalues of the matrix $(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{HA})$. Figure 3.2(b) shows the quantity $\|\lambda - \lambda_\infty\| \,/\, \|\lambda_\infty\|$ plotted against n in blue diamonds, where $\lambda = (\lambda_1, \lambda_2)$ represents the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$. The convergence of the eigenvalues is also exponential. The values of these eigenvalues confirm that the minimising gains stabilise the system since all of then are within the unit circle.

It is worth noting that these eigenvalues are not symmetric. Therefore even though the control theoretic approach provided us with a minimising gain it wasn't the optimal one since we had constrained it by fixing the eigenvalues of the matrix in question. However, it provided us with a good motivation to investigate the convergence of the optimal gain.

Chapter Summary In this chapter we have defined the out-of-sample error and optimism in the context of data assimilation. Using data assimilation through synchronisation as our algorithm, we presented several numerical experiments for linear systems with linear observations. The results presented show that the out-of-sample error is indeed a good measure of performance and that it is easily calculated even in operational settings. This is because the observations are taken to be linear and the calculation of the out-of-sample error does not require explicit knowledge of the model error covariance.

These results also raise some interesting questions about the asymptotic behaviour of the errors and the gain matrices that minimise these errors. Further numerical simulations suggest that the gain matrix that minimises the out-of-sample error converges to the asymptotic Kalman Gain in the limit of large observational windows. Moreover, the results presented suggest that the gain matrix that minimises the out-of-sample error is the same as the gain that minimises the state error.

Chapter 4

Optimal Filtering

The numerical experiments in the previous chapter and in Mallia-Parfitt & Bröcker (2016) suggest that the feedback gain matrix minimising the out-of-sample error converges to the asymptotic gain in the limit of large observational windows. In Chapters 5 and 6 we shall prove this fact rigorously, however before we do so, we digress briefly to give a detailed introduction to the Kalman Filter and its asymptotic properties. We present in detail the discrete time Kalman Filter for linear systems with gaussian perturbations. In this setting, the Kalman Filter is the optimal linear filter. It is essential to understand these concepts prior to the main proof of this thesis as certain ideas are used and/or adapted in the next chapters.

Following this in-depth discussion regarding the Kalman Filter and its asymptotic properties, we consider in detail the notions of controllability and observability. We have already seen the importance of observability in the numerical experiments presented in Chapter 3 and in Mallia-Parfitt & Bröcker (2016). Both controllability and observability are crucial in the set up of the Kalman Filter equations, the asymptotic properties of the filter and eventually in the main work performed for this thesis.

4.1 The Discrete-Time Kalman Filter

Section 2.1 gave a brief overview of the Kalman Filter equations. Here, they are derived in detail and some further information and properties of the filtering problem are presented. In particular we give extra attention to its asymptotic properties. Suppose we have, for $n \geq 0$, the system defined by the following equations,

$$\begin{aligned} x_{n+1} &= \mathbf{A}_n x_n + q_n \\ \eta_n &= \mathbf{H}_n x_n + r_n \end{aligned} \tag{4.1}$$

where $\{q_n\}$, $\{r_n\}$ are independent, zero mean, gaussian white processes with

$$\mathbb{E}(q_n q_n^T) = \mathbf{Q}_n, \quad \mathbb{E}(r_n r_n^T) = \mathbf{R}_n. \tag{4.2}$$

The filtering problem, in broad terms, requires the deduction of information about x_n using measurements up until time n, Anderson & Moore (1979). However, in order to simplify the problem, we shall seek to deduce information about x_n using observations until time $n-1$ and then update the system to time n so that, in effect, we shall be considering a *one-step prediction problem*. This one-step prediction problem requires computations of the sequence $\mathbb{E}\{x_n|\eta_0, \ldots \eta_{n-1}\}$ for $n = 0, 1, \ldots$. We shall denote this quantity by $\hat{z}_n$.

Once we have this quantity, we want to know how good of an estimate it is. This is measured by the error covariance matrix Σ_n, which is defined by

$$\Sigma_n = \mathbb{E}\{(x_n - \hat{z}_n)(x_n - \hat{z}_n)^T|\eta_0, \ldots, \eta_{n-1}\}. \tag{4.3}$$

Once we have these estimates, we will want to compute the true filtered estimate, $\mathbb{E}\{x_n|\eta_0, \ldots, \eta_n\}$ which we shall denote by z_n. The associated error covariance, which we

are also interested in calculating, is denoted by Γ_n and defined by

$$\Gamma_n = \mathbb{E}\{(x_n - z_n)(x_n - z_n)^T | \eta_0, \ldots, \eta_n\}. \tag{4.4}$$

Due to the '$n-1$' notation we define the initial data for $n = 0$ to be $\hat{z}_0 = \mathbb{E}(x_0)$, $\Sigma_0 = \{(x_0 - z_0)(x_0 - z_0)^T\}$ given no measurements. Bringing all these ideas together, we state the Discrete-Time Kalman Filtering Problem formally as follows:

For the linear, finite-dimensional, discrete-time system of (4.1) defined for $n \geq 0$, suppose that $\{q_n\}$, $\{r_n\}$ are independent, zero mean gaussian processes with $\mathbb{E}r_n r_n^T = \mathbf{R}_n$, $\mathbb{E}q_n q_n^T = \mathbf{Q}_n$. Suppose further that the initial state x_0 is a gaussian random variable with mean $\hat{z}_0$ and covariance Σ_0 independent of $\{q_n\}$ and $\{r_n\}$. Determine the estimates

$$\hat{z}_n = \mathbb{E}\{x_n | \eta_0, \ldots, \eta_{n-1}\}, \quad z_n = \mathbb{E}\{x_n | \eta_0, \ldots, \eta_n\} \tag{4.5}$$

and the associated error covariances Σ_n and Γ_n as defined in (4.3) and (4.4) respectively.

The solution to the Kalman Filtering problem is given below. We omit the proof here however a full First Principles Derivation of the Kalman Filtering Equation can be found in Chapter 3 of Anderson & Moore (1979).

The Kalman Filter is described, for $n \geq 0$, by the equations

$$\hat{z}_n = \mathbf{A}_n z_{n-1}, \quad z_n = \hat{z}_n + \mathbf{K}_n(\eta_n - \mathbf{H}_n \hat{z}_n) \tag{4.6}$$

where $\mathbf{K}_n$ is the gain matrix and is determined from the error covariance matrix by

$$\mathbf{K}_n = \Sigma_n \mathbf{H}^T (\mathbf{H}_n \Sigma_n \mathbf{H}_n + \mathbf{R}_n)^{-1}. \tag{4.7}$$

We assume here that $\mathbf{H}_n\Sigma_n\mathbf{H}_n + \mathbf{R}_n$ is invertible. This normally holds and is in fact guaranteed if $\mathbf{R}_n$ is positive definite, Anderson & Moore (1979).

In order to relate the above equations with the discussions on data assimilation algorithms in Chapter 2, note that the term denoted by $\hat{z}_n$ is the *background* term and the term z_n is the *analysis*. The error covariance matrices Σ_n and Γ_n are the background and analysis covariance matrices respectively. The gain matrix $\mathbf{K}_n$ is the same weight matrix given in Chapter 2 and the structure of the gain matrix here is the same as in (2.3) since Σ_n is the background error covariance matrix.

The conditional error covariance matrices are given recursively by

$$\begin{aligned} \Sigma_n &= \mathbf{A}_{n-1}[\Sigma_{n-1} - \mathbf{K}_{n-1}\mathbf{H}_{n-1}\Sigma_{n-1}]\mathbf{A}_{n-1}^T + \mathbf{Q}_{n-1} \\ &= \mathbf{A}_{n-1}(\mathbb{1} - \mathbf{K}\mathbf{H})\Sigma_{n-1}\mathbf{A}_{n-1}^T + \mathbf{Q}_{n-1} \end{aligned} \tag{4.8}$$

and

$$\Gamma_n = (\mathbb{1} - \mathbf{K}_n\mathbf{H}_n)\Sigma_n. \tag{4.9}$$

The equations yielding z_n and Γ_n are sometimes called *measurement-update equations* and the equations yielding $\hat{z}_n$ and Σ_n are called *time-update equations*, Anderson & Moore (1979).

The Kalman Filter is a linear, discrete-time, finite-dimensional system. These are all desirable qualities making this filter rather nice to work with. Since Σ_n, $\mathbf{K}_n$ are independent of the measurement process, they can be calculated before the filter is actually run. This means that no one set of measured observations helps any more than any other to eliminate some uncertainty about x_n.

The Kalman Filter is the optimal filter of all linear filters (Anderson & Moore 1979) and the particular gain $\mathbf{K}_n$ as given in (4.7), which is called the *Kalman Gain*, minimises the error covariance Γ_n. This is straightforward to calculate by taking the derivative of Γ_n with respect to the gain and setting equal to zero. The resulting expression that must be

satisfied is given by

$$(\mathbf{K}_n\mathbf{H} - \mathbb{1})\Sigma_n\mathbf{H}^T + \mathbf{H}\Sigma_n(\mathbf{K}_n\mathbf{H} - \mathbb{1})^T + \mathbf{K}_n\mathbf{R} + \mathbf{R}\mathbf{K}_n^T = 0, \tag{4.10}$$

and it follows that the Kalman Gain defined by (4.7) is indeed the optimal solution.

It is possible to generalise the above and have one or more of the matrices $\mathbf{A}_n, \mathbf{H}_n, \mathbf{Q}_n, \mathbf{R}_n$ take values which depend on the measurement process. In this case some of the previous statements still hold true. For example the expressions for $\hat{z}_n$ and Σ_n are still valid but the gain matrix $\mathbf{K}_n$ and the error covariance Σ_n cannot be computed in advance as they now depend on $\{\eta_0, \ldots \eta_{n-1}\}$.

4.2 Time-Invariance and Asymptotic Stability of the Kalman Filter

In general, $\mathbf{A}_n$, $\mathbf{H}_n$ and $\mathbf{K}_n$ depend on n; that is the filter described in Section 4.1 is a time-varying filter. Time-invariant filters are those with $\mathbf{A}_n$, $\mathbf{H}_n$ and $\mathbf{K}_n$ independent of n. Clearly for the filter in Section 4.1 to be time invariant, the gain matrix $\mathbf{K}_n$ must be constant and unless there is some cancellation in the time variation of $\mathbf{A}_n$ and $\mathbf{K}_n\mathbf{H}_n$ to force $(\mathbf{A}_n - \mathbf{K}_n\mathbf{H}_n\mathbf{A}_n)$ to be constant, the matrices $\mathbf{A}_n$ and $\mathbf{H}_n$ must be constant too.

Certain assumptions applied to the underlying system do lead to the filter being time-invariant. These assumptions are time invariance of the system being filtered and stationarity of the random processes associated with the underlying system. It can be shown that these two conditions are in fact sufficient to guarantee time invariance of the filter, Anderson & Moore (1979).

As well as time-invariance of the filter, we are interested in the asymptotic stability of the filter; we shall only consider time invariant filters when investigating this concept. An equivalent task is to explain when the eigenvalues of the error matrix, $(\mathbf{A} - \mathbf{AKH})$ (if

we consider the background error covariance), or $(\mathbf{A} - \mathbf{KHA})$ (if we consider the analysis error covariance), lie inside the unit circle. We shall present precise conditions under which the filter is time-invariant and asymptotically stable.

In order to pin down the conditions which guarantee simultaneously that the optimal filter is both time-invariant (or asymptotically time-invariant) and asymptotically stable, we make the assumptions that the system is both completely controllable and observable. Denote by $\mathcal{O}(\mathbf{A}, \mathbf{H})$, the observability matrix which is defined by

$$\mathcal{O}(\mathbf{A}, \mathbf{H}) = [\mathbf{H} \quad \mathbf{HA} \quad \mathbf{HA}^2 \quad \dots \quad \mathbf{HA}^{n-1}]^T. \tag{4.11}$$

Then we have the following definitions.

Definition 4.2.1. *A linear dynamical system as in (4.1) is said to be observable if any of the following equivalent conditions hold.*

1. *The observability matrix, $\mathcal{O}(\mathbf{A}, \mathbf{H})$, has rank n.*
2. *ker $\mathbf{H}$ has no $\mathbf{A}$ invariant subspaces.*
3. *If $\mathbf{A}x = \lambda x$ then $\mathbf{H}x \neq 0$.*

Definition 4.2.2. *The linear system given by (4.1) is called controllable if the signal process noise is non-degenerate which means that*

$$\mathbf{A}^T x = \lambda x \quad \Rightarrow \quad x^T \mathbf{Q} x \neq 0. \tag{4.12}$$

Observability and controllability are two very important concept that will have a big impact on later work. An in-depth discussion of these notions is presented in Section 4.3.

Asymptotic time invariance of the filter arises when there is an asymptotically constant solution to the variance equation,

$$\Sigma_n = \mathbf{A}(\Sigma_n - \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}\Sigma_n)\mathbf{A}^T + \mathbf{Q}. \tag{4.13}$$

Denote by Σ_∞ the asymptotically constant solution to (4.13). The associated gain is called the *asymptotic gain*, denoted by $\boldsymbol{\kappa}_\infty$ and defined by

$$\boldsymbol{\kappa}_\infty = \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \mathbf{R})^{-1} \tag{4.14}$$

and the question arises as to whether the eigenvalues of $(\mathbf{A} - \mathbf{A}\boldsymbol{\kappa}_\infty \mathbf{H})$ all lie within the unit circle, ensuring asymptotic stability of the filter. Note that this is the limit of the convergence of the gain $\mathbf{K}$ minimising the out-of-sample error in the numerical examples of Chapter 3. The main conclusions are given in theorem 4.2.1 below and in Chapter 4 of Anderson & Moore (1979).

Theorem 4.2.1. *If the model is time invariant, observable, controllable and* $\mathbf{R}$ *is strictly positive definite, then*

1. *For any non negative symmetric initial condition we have*

$$\lim_{n\to\infty} \Sigma_n = \Sigma_\infty \tag{4.15}$$

with Σ_∞ *independent of the initial condition and satisfying a steady-state version of (4.13):*

$$\Sigma_\infty = \mathbf{A}[\Sigma_\infty - \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}\Sigma_\infty]\mathbf{A}^T + \mathbf{Q}. \tag{4.16}$$

This equation is called the Discrete Algebraic Riccati Equation (DARE).

2.

$$|\lambda_i(\mathbf{A} - \mathbf{A}\boldsymbol{\kappa}_\infty \mathbf{H})| < 1 \tag{4.17}$$

with $\boldsymbol{\kappa}_\infty$ *as in (4.14).*

3. Σ_∞ *is the unique non-negative definite solution to (4.16).*

The proof of this theorem is given in Chapter 4 of Anderson & Moore (1979) and

Appendix C. Complete controllability is required to establish asymptotic stability of the filter, and this is explicitly seen in the proof.

Complete observability on the other hand makes a more subtle appearance in the proof. Observability of the system is required to ensure the existence of Σ_∞. To see this suppose there is a mode that is not observed and not asymptotically stable, yet it is excited by the input. Since it is not observed, the best estimate of it is zero and the error variance will be the variance of the mode. Since it is not asymptotically stable the variance will be unbounded and a steady state value cannot exist. Therefore complete observability is needed to ensure the existence of Σ_∞.

4.3 Observability and Controllability

As was shown in Section 4.2, the assumptions that the system being analysed is both controllable and observable are crucial. They will also play a very important role in establishing the main result required of this thesis. Observability in particular will be essential. This is mainly because we are minimising the out-of-sample error which is an error in observation space. For this section we digress briefly to explore these concepts further and discuss their implications.

Recall that a pair of matrices $(\mathbf{A}, \mathbf{H})$, as given in the system (4.1), is said to be observable when the observability matrix

$$\mathcal{O}(\mathbf{A}, \mathbf{H}) = [\mathbf{H} \quad \mathbf{HA} \quad \mathbf{HA}^2 \quad \dots \quad \mathbf{HA}^{n-1}]^T \tag{4.18}$$

has full rank. There are other equivalent definitions which are omitted here but given in definition 4.2.1. If this condition holds then the poles of the error matrix can be placed anywhere by proper selection of $\mathbf{K}$. In particular they can be placed within the unit circle ensuring that the error dynamics are stable. In our situation the error dynamics in the

noise free case (i.e $r_n, q_n = 0$) are given by

$$x_{n+1} - z_{n+1} = (\mathbf{A} - \mathbf{KHA})(x_n - z_n). \tag{4.19}$$

Therefore, we require that the pair of matrices $(\mathbf{A}, \mathbf{HA})$ be observable.

Lemma 4.3.1. *Suppose* $\mathbf{A}$ *is invertible. Then* $(\mathbf{A}, \mathbf{H})$ *observable implies* $(\mathbf{A}, \mathbf{HA})$ *observable.*

Proof. Consider the observability matrix $\mathcal{O}(\mathbf{A}, \mathbf{HA})$,

$$\begin{aligned} \mathcal{O}(\mathbf{A}, \mathbf{HA}) &= [\mathbf{HA} \quad \mathbf{HA}^2 \quad \mathbf{HA}^3 \quad \dots \quad \mathbf{HA}^n]^T \\ &= \mathcal{O}(\mathbf{A}, \mathbf{H}) \cdot \mathbf{A}. \end{aligned} \tag{4.20}$$

Since $(\mathbf{A}, \mathbf{H})$ is an observable pair, the corresponding observability matrix has full rank. The matrix $\mathbf{A}$ is also of full rank as it is invertible so it follows that

$$\text{rank}(\mathcal{O}(\mathbf{A}, \mathbf{H}) \cdot \mathbf{A}) = \text{rank}(\mathcal{O}(\mathbf{A}, \mathbf{H})). \tag{4.21}$$

Hence, $\mathcal{O}(\mathbf{A}, \mathbf{HA})$ has full rank and $(\mathbf{A}, \mathbf{HA})$ is an observable pair. □

Lemma 4.3.2. *Suppose* $(\mathbf{A}, \mathbf{H})$ *is an observable pair and let* $\mathbf{K}$ *be an arbitrary feedback gain matrix with the appropriate dimensions. Then the pair* $(\mathbf{A} - \mathbf{KH}, \mathbf{H})$ *is also observable.*

Proof. The pair of matrices $(\mathbf{A}, \mathbf{H})$ being observable means

$$\mathbf{A}x = \lambda x, \quad (x \neq 0) \quad \Rightarrow \quad \mathbf{H}x \neq 0. \tag{4.22}$$

Suppose that for arbitrary $\mathbf{K}$, $(\mathbf{A} - \mathbf{KH}, \mathbf{H})$ is not an observable pair. Then there exists $x \neq 0$ such that $(\mathbf{A} - \mathbf{KH})x = \lambda x$ so that $\mathbf{H}x = 0$. However, this implies that $\mathbf{A}x = \lambda x$ which means that x is now an eigenvector of $\mathbf{A}$. But since $(\mathbf{A}, \mathbf{H})$ is observable, we cannot have $\mathbf{H}x = 0$. Thus we have a contradiction and so it follows that $(\mathbf{A} - \mathbf{KH}, \mathbf{H})$ is an observable pair. □

Lemma 4.3.3. *Suppose* $\mathbf{A}$ *is invertible and* $(\mathbf{A}, \mathbf{H})$ *is an observable pair. Then the pair* $(\mathbf{A} - \mathbf{KHA}, \mathbf{HA})$ *is also observable.*

Proof. From lemma 4.3.1 it follows that $(\mathbf{A}, \mathbf{HA})$ is an observable pair. Applying lemma 4.3.2 to the pair $(\mathbf{A}, \mathbf{HA})$ yields the required result, namely that $\mathcal{O}(\mathbf{A} - \mathbf{KHA}, \mathbf{HA})$ is an observable pair.

□

Consider now the pair $(\mathbf{A} - \mathbf{KHA}, \mathbf{H})$. To investigate whether or not this pair of matrices is observable, consider the corresponding the observability matrix:

$$\begin{aligned}\mathcal{O}(\mathbf{A} - \mathbf{KHA}, \mathbf{H}) &= \begin{bmatrix} \mathbf{H} \\ \mathbf{H}(\mathbf{A} - \mathbf{KHA}) \\ \vdots \\ \mathbf{H}(\mathbf{A} - \mathbf{KHA})^{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ (\mathbb{1} - \mathbf{HK})\mathbf{HA} \\ \vdots \\ (\mathbb{1} - \mathbf{HK})\mathbf{HA}(\mathbf{A} - \mathbf{KHA})^{n-2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H} \\ (\mathbb{1} - \mathbf{HK})\mathcal{O}(\mathbf{A} - \mathbf{KHA}, \mathbf{HA}) \end{bmatrix}.\end{aligned} \tag{4.23}$$

Since lemma 4.3.3 tells us that $\mathcal{O}(\mathbf{A} - \mathbf{KHA}, \mathbf{HA})$ is of full rank, it follows that provided $(\mathbb{1} - \mathbf{HK})$ is non-singular, the pair $(\mathbf{A} - \mathbf{KHA}, \mathbf{H})$ is observable. A natural question that arises then is: When is this matrix invertible?

Example: The Kalman Gain

If the feedback gain matrix $\mathbf{K}$ is given by the Kalman Gain,

$$\mathbf{K}_n = \Sigma_n \mathbf{H}^T (\mathbf{H} \Sigma_n \mathbf{H}^T + \mathbf{R})^{-1} \tag{4.24}$$

then the matrix $(\mathbb{1} - \mathbf{HK}_n)$ is non-singular. This follows from the fact that

$$\begin{aligned} \mathbf{HK}_n &= \mathbf{H}\Sigma_n\mathbf{H}^T(\mathbf{H}\Sigma_n\mathbf{H}^T + \mathbf{R})^{-1} \\ &= (\mathbf{H}\Sigma_n\mathbf{H}^T + \mathbf{R} - \mathbf{R})(\mathbf{H}\Sigma_n\mathbf{H}^T + \mathbf{R})^{-1} \\ &= \mathbb{1} - \mathbf{R}(\mathbf{H}\Sigma_n\mathbf{H}^T + \mathbf{R})^{-1} \\ \Rightarrow \mathbb{1} - \mathbf{HK}_n &= \mathbf{R}(\mathbf{H}\Sigma_n\mathbf{H}^T + \mathbf{R})^{-1} \end{aligned} \tag{4.25}$$

which is an invertible matrix.

This is just one example of when $(\mathbb{1} - \mathbf{HK})$ is non-singular. Since we are working with minimisers of the observed errors, it is extremely important that we have the above result and we shall see that this fact plays a very important role in proving that the minimising gain of either the out-of-sample or observed error, converges to the asymptotic Kalman gain $\boldsymbol{\kappa}_\infty$ as given in equation (4.14).

Controllability is also a very important concept. We have already seen the definition for controllability in definition 4.2.2. There are other equivalent definitions that can be given to define controllability, see for example Appendix C of Anderson & Moore (1979). What is of particular interest is the connection between observability and controllability.

It is not difficult to see that the pair $(\mathbf{A}, \mathbf{H})$ being observable is equivalent to the pair $(\mathbf{A}^T, \mathbf{H}^T)$ being controllable according to definition 4.2.2. For clarity purposes consider the following alternative definition of controllability,

Definition 4.3.1. *The pair of matrices* $(\mathbf{A}, \mathbf{H})$ *is said to be controllable if*

$$\mathbf{A}^T x = \lambda x \quad \Rightarrow \quad \mathbf{H}^T x \neq 0, \quad x, \lambda \neq 0. \tag{4.26}$$

The implication of this is as follows. Suppose the pair $(\mathbf{A}, \mathbf{H})$ is observable. Then by definition 4.2.1,

$$\mathbf{A}x = \lambda x \quad \Rightarrow \quad \mathbf{H}x \neq 0. \tag{4.27}$$

However, by definition 4.3.1, this is equivalent to $(\mathbf{A}^T, \mathbf{H}^T)$ being controllable. This in turn implies that when the system is assumed to be completely controllable and observable, equations (4.26) and (4.27) both hold at the same time. This duality property of observability and controllability will be used in later chapters to establish important facts about the out-of-sample error, its minimisers and their asymptotic behaviour.

Chapter Summary In this chapter we have presented an in-depth discussion on the discrete-time Kalman Filter for linear systems with gaussian perturbations. It was established that the Kalman Filter is the optimal linear filter for such systems.

We investigated its asymptotic properties and determined that it is both asymptotically stable and time-invariant. In order to prove asymptotic stability and time-invariance, it was necessary to assume the system was completely observable and controllable. As such, the concepts of observability and controllability were discussed in more detail and their duality property was introduced.

Chapter 5

Minimising the Error Covariance

The Kalman Filter, as described in Chapter 4, is the best linear filter available for linear systems with gaussian perturbations. Asymptotically, it is stable and time invariant; two desirable qualities. However, it can be computationally expensive to run even though some terms can be computed prior to running the filter itself. This is because a matrix inversion is required at every step to determine the optimal feedback gain. This feedback also depends on n and on the background error covariance matrix (which we denote by Σ_n) and this matrix is difficult to determine. If one studies the recursive equation for this covariance matrix, it becomes evident that knowledge of the model error covariance is essential. Unfortunately this information is often unavailable in practice and so certain assumptions and compromises must be made. In operational settings the model error covariance has to be estimated, resulting in further uncertainties.

The numerical experiments presented in Chapter 3 suggest that a constant feedback gain matrix that minimises the empirical out-of-sample error exists and that it is the same as the gain matrix that minimises the state error. Moreover, the numerical experiments suggest that the minimising feedback gain converges to the asymptotic Kalman gain in the limit of large observational windows. An advantage of determining the optimal feedback gain matrix in this way is that knowledge of the dynamical model error covariance is

unnecessary.

In this chapter, using the results in Section 4.2, we shall rigorously prove that the constant feedback gain matrix minimising the out-of-sample error exists, and with increasing observational windows converges to the asymptotic Kalman gain. Unfortunately however, this is not quite what can be done in practice. This is because we can only estimate the true error covariance as we do not have the access to the true values required. Therefore, in practical situations and in fact in our numerical experiments in Chapter 3, the errors are estimated by the empirical mean namely,

$$\frac{1}{n}\sum_{i=1}^{n}(z_i - x_i)^2 \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \eta_i')^2. \tag{5.1}$$

leading to estimators of the optimal gain. The question that arises then concerns the asymptotic behaviour of this estimator. In Chapter 6, we establish that this estimator has the same asymptotic behaviour as the Kalman gain.

5.1 The Gain Minimising the Out-of-Sample Error

5.1.1 Design of an Observer

Suppose we have an initial state $x_0 \in \mathbb{R}^D$, with mean $\hat{z}_0$ and covariance Σ_0. Suppose also that we have a time invariant dynamical model given by

$$x_{n+1} = \mathbf{A}x_n + q_n \tag{5.2}$$

where $x_n \in \mathbb{R}^D$ is the state and q_n are iid random variables with zero mean and covariance $\mathbf{Q}$. The measured observations are given by

$$\eta_n = \mathbf{H}x_n + r_n \tag{5.3}$$

where η_n are observations in some space which we take to be $\mathbb{R}^d$ and r_n are iid random variables with zero mean and covariance $\mathbf{R}$. It is assumed that r_n and q_n are uncorrelated. Note also that this model is time invariant since $\mathbf{A}$, $\mathbf{H}$, $\mathbf{Q}$ and $\mathbf{R}$ are taken to be constant. When we refer to the model we shall mean the quadruple $(\mathbf{A}, \mathbf{H}, \mathbf{Q}, \mathbf{R})$ and we will say that the initial data is given by $(\hat{z}_0, \Sigma_0)$.

For the system defined by (5.2) and (5.3), we construct an observer of the form

$$\begin{aligned} \hat{z}_n &= \mathbf{A} z_{n-1} \\ z_n &= \hat{z}_n + \mathbf{K}_n(\eta_n - \mathbf{H}\hat{z}_n) \end{aligned} \tag{5.4}$$

where $\mathbf{K}_n$ is the gain matrix which may or may not depend on n. The feedback gain matrix depends on the observations up to time $n-1$ but does not depend on η_n. The coupling introduced by this gain matrix creates a linear feedback in the sense that the error between $\mathbf{H}\hat{z}_n$ and the observations (i.e. the innovations) is fed back into the model. The background term $\hat{z}_n$ is an estimate of x_n based on our *a priori* knowledge of the system, up to but not including time n. The trajectory obtained from this scheme, $z_n \in \mathbb{R}^D$ (i.e the analysis), lives in the state space $\mathbb{R}^D$.

Lemma 5.1.1.

$$\mathbb{E}\hat{z}_n = \mathbb{E}x_n \tag{5.5}$$

$$\Sigma_n = \mathbb{E}[(\hat{z}_n - x_n)(\hat{z}_n - x_n)^T] = \mathbf{A}\Gamma_{n-1}\mathbf{A}^T + \mathbf{Q} \tag{5.6}$$

$$\Gamma_n = \mathbb{E}[(z_n - x_n)(z_n - x_n)^T] = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\Sigma_n(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T \tag{5.7}$$

Proof. We prove the equality in (5.5) by induction. First note that

$$\mathbb{E}x_n = \mathbf{A}\mathbb{E}x_{n-1} = \mathbf{A}^n\hat{z}_0, \tag{5.8}$$

which follows by induction since the random variables q_n are iid with zero mean and

the initial state x_0 has mean $\hat{z}_0$. Then we have for the case $k = 1$,

$$\mathbb{E}\hat{z}_1 = \mathbf{A}\hat{z}_0 - \mathbf{A}\mathbf{K}_0\mathbf{H}\,(\hat{z}_0 - \mathbb{E}x_0) = \mathbf{A}\hat{z}_0 = \mathbb{E}x_1. \tag{5.9}$$

Assume this relation holds for the case $k = n-1$ and we want to show it also holds for the case $k = n$;

$$\mathbb{E}\hat{z}_n = \mathbf{A}\mathbb{E}\hat{z}_{n-1} - \mathbf{A}\mathbf{K}_{n-1}\mathbf{H}\,(\mathbb{E}\hat{z}_{n-1} - \mathbb{E}x_{n-1}) = \mathbf{A}\mathbb{E}\hat{z}_{n-1} = \mathbb{E}x_n. \tag{5.10}$$

Therefore, by induction, equation (5.5) holds.

We derive equations (5.6) and (5.7) together. Consider the error covariance matrix

$$\begin{aligned}
\Gamma_n &= \mathbb{E}[(z_n - x_n)(z_n - x_n)^T] \\
&= \mathbb{E}[((\mathbb{1} - \mathbf{K}_n\mathbf{H})(\hat{z}_n - x_n) + \mathbf{K}_n r_n)(\mathbb{1} - \mathbf{K}_n\mathbf{H})(\hat{z}_n - x_n) + \mathbf{K}_n r_n)^T] \\
&= (\mathbb{1} - \mathbf{K}_n\mathbf{H})\Sigma_n(\mathbb{1} - \mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T
\end{aligned} \tag{5.11}$$

where

$$\begin{aligned}
\Sigma_n &= \mathbb{E}[(\hat{z}_n - x_n)(\hat{z}_n - x_n)^T] \\
&= \mathbb{E}[(\mathbf{A}(z_{n-1} - x_{n-1}) - q_n)(\mathbf{A}(z_{n-1} - x_{n-1}) - q_n)^T] \\
&= \mathbf{A}\Gamma_{n-1}\mathbf{A}^T + \mathbf{Q}.
\end{aligned} \tag{5.12}$$

These relations are obtained by simply substituting the expressions for x_n and z_n into Γ_n and using the fact that the observational noise r_n is uncorrelated with the dynamical noise q_n.

See Chapter 3 of Anderson & Moore (1979) for an alternative derivation of these covariance matrices. □

The relations given in lemma 5.1.1 are in their most general form and hold for any feedback gain matrix $\mathbf{K}_n$. The Kalman gain (see Chapter 4) is a particular form of the

matrix $\mathbf{K}_n$, which we henceforth denote by $\boldsymbol{\kappa}_n$, that minimises the mean squared error given the model and initial data. The equation and properties of the Kalman Gain are given in the following lemma,

Lemma 5.1.2. *The Kalman Gain, $\boldsymbol{\kappa}_n$, minimises the mean squared error, Γ_n and it is defined by*

$$\boldsymbol{\kappa}_n = \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \mathbf{R})^{-1}. \tag{5.13}$$

Proof. For the Kalman gain to minimise the mean squared error covariance, it needs to satisfy

$$\mathrm{D}\Gamma_n(\boldsymbol{\kappa}_n).\Delta = 0 \tag{5.14}$$

for any perturbations Δ of $\boldsymbol{\kappa}_n$; that is any $\Delta \in \mathbb{R}^{D\times d}$. By differentiating (5.7) and using that $D\Gamma_n = 0$ and that Σ_n does not depend in $\boldsymbol{\kappa}_n$, it follows that $\boldsymbol{\kappa}_n$ must satisfy

$$(\mathbf{K}_n\mathbf{H} - \mathbb{1})\Sigma_n\mathbf{H}^T + \mathbf{H}\Sigma_n(\mathbf{K}_n\mathbf{H} - \mathbb{1})^T + \mathbf{K}_n\mathbf{R} + \mathbf{R}\mathbf{K}_n^T = 0 \tag{5.15}$$

from which it is straightforward to establish that $\boldsymbol{\kappa}_n$ is indeed given by (5.13).

To show that $\boldsymbol{\kappa}_n$ is a minimiser of the mean squared error covariance, we need to establish that

$$\Gamma_n(\boldsymbol{\kappa}_n + \Delta) \geq \Gamma_n(\boldsymbol{\kappa}). \tag{5.16}$$

Using a Taylor series expansion, it is straightforward to show that

$$\Gamma_n(\boldsymbol{\kappa}_n + \Delta) = \Gamma_n(\boldsymbol{\kappa}_n) + \mathrm{D}\Gamma_n(\boldsymbol{\kappa}_n).\Delta + \Delta\mathbf{H}\Sigma_n\mathbf{H}^T\Delta^T + \Delta\mathbf{R}\Delta^T \tag{5.17}$$

and since we know $\boldsymbol{\kappa}_n$ must satisfy (5.14) and that the last two terms in (5.17) are non-negative definite the required result is obtained. Even though no assumptions on uniqueness are being made here, it can be concluded that $\boldsymbol{\kappa}_n$ is unique if $\mathbf{R} > 0$, Anderson & Moore (1979). This condition, that the observation error covariance matrix is strictly

positive definite, is required anyway to guarantee the existence of the Kalman Gain. □

An observer given by (5.4) that uses the Kalman gain (5.13) as its feedback gain matrix is known as the Kalman Filter. In the class of linear filters which produce an estimate by minimising a mean squared error, the Kalman Filter is the optimal one (Anderson & Moore 1979); see Chapter 4.

We will now investigate the properties of the feedback gain matrix that minimises the out-of-sample error covariance. Recall that to define the out-of-sample error we assume that we have another set of observations, η'_n, which are given by

$$\eta'_n = \zeta_n + \sigma r'_n \tag{5.18}$$

where r'_n has the same stochastic properties as r_n but is independent from r_n, i.e $\mathbb{E}r_n r'_n = 0$. The out-of-sample error is then defined by

$$\begin{aligned} \mathbb{E}[(y_n - \eta'_n)^2] &= \mathbf{H}\mathbb{E}(z_n - x_n)(z_n - x_n)^T\mathbf{H}^T + \mathrm{tr}(\mathbf{R}') \\ &= \mathbf{H}\Gamma_n\mathbf{H}^T + \mathrm{tr}(\mathbf{R}'). \end{aligned} \tag{5.19}$$

where $\mathbf{R}'$ is the covariance of the iid noise r'_n. Note that $\mathbf{R}' = \mathbf{R}$ however for notational purposes we write $\mathbf{R}'$ in order to distinguish between the two. To minimise the out-of-sample error as defined above, we take the derivative with respect to $\mathbf{K}$ and set equal to zero. Doing so yields

$$\mathbf{H}\mathrm{D}_{\mathbf{K}}\Gamma_n(\mathbf{K}, \Gamma_0).\Delta\mathbf{H}^T = 0 \tag{5.20}$$

and we see from this, that minimising the out-of-sample error is equivalent to minimising the observed error, $\mathbf{H}\Gamma_n\mathbf{H}^T$, since the observation error covariance $\mathbf{R}$ (or $\mathbf{R}'$) has no dependence on the feedback gain matrix.

Lemma 5.1.3. *The Kalman Gain, $\boldsymbol{\kappa}_n$, defined by (5.13) minimises the observed error, $\mathbf{H}\Gamma_n\mathbf{H}^T$.*

Proof. To see this consider the following. Since $\boldsymbol{\kappa}_n$ is the optimal gain matrix in the sense that it minimises the state error covariance matrix Γ_n, we have

$$\Gamma_n(\boldsymbol{\kappa}_n + \Delta) \geq \Gamma_n(\boldsymbol{\kappa}_n). \tag{5.21}$$

This implies that

$$\mathbf{H}\Gamma_n(\boldsymbol{\kappa}_n + \Delta)\mathbf{H}^T \geq \mathbf{H}\Gamma_n(\boldsymbol{\kappa}_n)\mathbf{H}^T \tag{5.22}$$

and by the definition of " $\geq$ " for matrices (see Chapter 1), it follows that the Kalman Gain minimises the observed error. □

It is clear from the above, that the Kalman gain minimises the observed error; thus the Kalman Filter is optimal also in the sense that it minimises the observed or even the out-of-sample error. We have seen however that the Kalman Filter may be problematic as it requires knowledge of $\mathbf{Q}$. In order to calculate the empirical out-of-sample error, knowledge of $\mathbf{Q}$ is not required. Therefore, using this error as a measure of performance and to determine the optimal gain matrix, is advantageous. Therefore, we investigate the expected out-of-sample error, in particular for a constant gain matrix. We choose such a gain as we hope it will lead to a simpler filter as the feedback matrix will not need to be updated at every step.

5.1.2 Design of a Suboptimal Filter

Consider an observer of the form (5.4) for which we keep the feedback gain matrix constant so that we have $\mathbf{K}_n = \mathbf{K}$. Our aim is to choose this gain matrix so that it minimises the observation error, $\mathbf{H}\Gamma_n\mathbf{H}^T$, or the out-of-sample error over the whole assimilation window. For notational purposes we write, when the gains are all the same

$$\mathbf{H}\Gamma_n\mathbf{H}^T = \psi_n(\mathbf{K}, \Gamma_0) \tag{5.23}$$

to indicate the dependence on the initial condition, Γ_0.

Suppose that the system defined by (5.2) and (5.3) is completely observable and controllable as defined in definitions 4.2.1 and 4.2.2 respectively. Suppose also that the observation error covariance matrix is strictly positive definite (i.e. $\mathbf{R} > 0$) so that the results presented in Chapter 4 hold.

In order to establish that the feedback gain matrix that minimises the expected out-of-sample error converges to the asymptotic Kalman Gain, we use the following result. It is a deterministic version of theorem 5.7 in Van der Vaart (2000) which is stated and proven in Chapter 6, theorem 6.1.1.

Theorem 5.1.1. *Consider the continuous functions*

$$\psi : \mathcal{K} \to \mathbb{R}_{\geq 0}, \quad \psi_n : \mathbb{R}^D \to \mathbb{R}_{\geq 0} \tag{5.24}$$

with $\mathcal{K} \subset \mathbb{R}^D$ compact and assume that ψ_n has a minimiser, which we shall denote by $\mathbf{K}_n$. This minimiser is not necessarily unique. Assume further that

1. *$\mathbf{K}_n \in \mathcal{K}$ for $n \geq n_0$ for some n_0*

2. *$\psi_n \to \psi$ uniformly on $\mathcal{K}$*

3. *ψ has a unique minimiser $\boldsymbol{\kappa}_\infty$.*

Then $\mathbf{K}_n \to \boldsymbol{\kappa}_\infty$.

The n stated in the above theorem refers to the size of the observational window. That is, the minimising gain converges to the asymptotic Kalman gain as the size of the observational window increases.

Proof. Assume $n \geq n_0$. Then

$$\begin{aligned} 0 &\leq \psi(\mathbf{K}_n) - \psi(\boldsymbol{\kappa}_\infty) \\ &= \underbrace{\psi(\mathbf{K}_n) - \psi_n(\mathbf{K}_n)}_{A} + \underbrace{\psi_n(\mathbf{K}_n) - \psi_n(\boldsymbol{\kappa}_\infty)}_{B} + \underbrace{\psi_n(\boldsymbol{\kappa}_\infty) - \psi(\boldsymbol{\kappa}_\infty)}_{C}. \end{aligned} \tag{5.25}$$

Now, $A \to 0$ and $C \to 0$ by assumption (2) in the statement of the theorem and $B \leq 0$ because $\mathbf{K}_n$ minimises ψ_n by assumption and so $\psi_n(\boldsymbol{\kappa}_\infty) \geq \psi_n(\mathbf{K}_n)$. Hence $\psi(\mathbf{K}_n) \to \psi(\boldsymbol{\kappa}_\infty)$.

Since $\mathcal{K}$ is compact, we consider the sub-sub-sequence n_{l_k} such that $\mathbf{K}_{n_{l_k}}$ converges to some $\boldsymbol{\kappa}^*$. But since $\psi_n \to \psi$ uniformly and $\boldsymbol{\kappa}_\infty$ is the unique minimiser of ψ, $\boldsymbol{\kappa}^*$ must be equal to $\boldsymbol{\kappa}_\infty$. Repeating the argument for all converging subsequences yields the same conclusion, thus $\mathbf{K}_n \to \boldsymbol{\kappa}_\infty$. □

In order to prove that the gain matrix minimising the expected observed error or the expected out-of-sample error converges to the asymptotic Kalman gain, we need to check that each of the four items given in theorem 5.1.1 hold.

We first establish point (3) in theorem 5.1.1. Recall that the asymptotic Kalman gain is defined by

$$\boldsymbol{\kappa}_\infty = \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \mathbf{R})^{-1}. \tag{5.26}$$

The expression for ψ is given by

$$\begin{aligned} \psi = \mathbf{H}\Gamma_\infty \mathbf{H}^T \quad &= \mathbf{H}(\mathbf{A} - \mathbf{KHA})\Gamma_\infty (\mathbf{A} - \mathbf{KHA})^T \mathbf{H}^T \\ &+ \mathbf{H}(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T \mathbf{H}^T + \mathbf{HKRK}^T \mathbf{H}^T \end{aligned} \tag{5.27}$$

which is obtained by taking limits in equation (5.7). Then we have the following proposition.

Proposition 5.1.1. *$\boldsymbol{\kappa}_\infty$ minimises ψ uniquely.*

Proof. Taking the derivative, assuming it exists in an open neighbourhood of the gain

matrix, of the observed asymptotic error, ψ as given in (5.27) and setting this equal to zero tells us that we need the gain matrix to satisfy

$$0 = \mathrm{D}\mathbf{H}\Gamma_\infty(\mathbf{K})\mathbf{H}^T.\Delta \tag{5.28}$$

$$= \Delta\mathbf{H}\left[\mathbf{R}\mathbf{K}^T - \mathbf{H}\mathbf{Q}(\mathbb{1} - \mathbf{K}\mathbf{H})^T - \mathbf{H}\mathbf{A}\Gamma_\infty(\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A})^T\right]\mathbf{H}^T \tag{5.29}$$

$$+ \mathbf{H}\left[\mathbf{K}\mathbf{R} - (\mathbb{1} - \mathbf{K}\mathbf{H})\mathbf{Q}\mathbf{H}^T - (\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A})\Gamma_\infty\mathbf{A}^T\mathbf{H}^T\right]\mathbf{H}^T\Delta^T \tag{5.30}$$

and since this must hold for all Δ the minimising $\mathbf{K}$ must satisfy

$$\begin{aligned} 0 &= \mathbf{H}\left[\mathbf{K}\mathbf{R} - (\mathbb{1} - \mathbf{K}\mathbf{H})\mathbf{Q}\mathbf{H}^T - (\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A})\Gamma_\infty\mathbf{A}^T\mathbf{H}^T\right]\mathbf{H}^T \\ &= \mathbf{H}\left[\mathbf{K}(\mathbf{H}\Sigma_\infty\mathbf{H}^T + \mathbf{R}) - \Sigma_\infty\mathbf{H}^T\right]\mathbf{H}^T \end{aligned}$$

where $\Sigma_\infty = \mathbf{A}\Gamma_\infty\mathbf{A}^T + \mathbf{Q}$. So we can see that $\mathbf{K} = \boldsymbol{\kappa}_\infty$ is one solution to the problem, however we wish to show that it is unique. Notice that

$$\mathbf{H}\mathbf{K} = \mathbf{H}\Sigma_\infty\mathbf{H}^T(\mathbf{H}\Sigma_\infty\mathbf{H}^T + \mathbf{R})^{-1} \quad \Rightarrow \quad \mathbb{1} - \mathbf{H}\mathbf{K} = \mathbf{R}(\mathbf{H}\Sigma_\infty\mathbf{H}^T + \mathbf{R})^{-1}, \tag{5.31}$$

is a non-singular matrix so it follows from the results presented in Chapter 4.3 that the pair of matrices $(\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A}, \mathbf{H})$ is observable.

For any two symmetric matrices $\mathbf{M}_1, \mathbf{M}_2$, we write $\mathbf{M}_1 \geq \mathbf{M}_2$ if $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite but not zero. Let $\mathbf{K}_1, \mathbf{K}_2$ be two stabilising feedback gains so that $\Gamma(\mathbf{K}_1) \geq \Gamma(\mathbf{K}_2)$; that is $\mathbf{K}_2$ performs better than $\mathbf{K}_1$.

Bearing this in mind, suppose that there exists another stabilising feedback gain, $\mathbf{K}^*$, so that $\Gamma(\boldsymbol{\kappa}_\infty) \geq \Gamma(\mathbf{K}^*)$, i.e. $\boldsymbol{\kappa}_\infty$ performs worse than $\mathbf{K}^*$. Multiplying from the left and right by $\mathbf{H}$ preserves the inequality so

$$\mathbf{H}\Gamma(\boldsymbol{\kappa}_\infty)\mathbf{H}^T \geq \mathbf{H}\Gamma(\mathbf{K}^*)\mathbf{H}^T. \tag{5.32}$$

Assuming that $\mathbf{H}\Gamma(\boldsymbol{\kappa}_\infty)\mathbf{H}^T = \mathbf{H}\Gamma(\mathbf{K}^*)\mathbf{H}^T$ would then imply,

$$\begin{aligned} 0 &= \mathbf{H}(\Gamma_\infty(\boldsymbol{\kappa}_\infty) - \Gamma_\infty(\mathbf{K}^*))\mathbf{H}^T \\ &= \mathbf{H}(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{H}\mathbf{A})^n(\Gamma_\infty(\boldsymbol{\kappa}_\infty) - \Gamma_\infty(\mathbf{K}^*))\ (\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{H}\mathbf{A})^{n\,T}\,\mathbf{H}^T \end{aligned} \tag{5.33}$$

as all the other terms in the expression for Γ would cancel each other out. Let $\mathbf{M} = \Gamma_\infty(\boldsymbol{\kappa}_\infty) - \Gamma_\infty(\mathbf{K}^*)$, so it follows that

$$\mathbf{H}\mathbf{M}\mathbf{H}^T = 0 \quad \Rightarrow \quad \mathbf{H}\left\{(\mathbf{A} - \mathbf{K}^*\mathbf{H}\mathbf{A})^n\mathbf{M}\,(\mathbf{A} - \mathbf{K}^*\mathbf{H}\mathbf{A})^{n\,T}\right\}\mathbf{H}^T = 0. \tag{5.34}$$

Using the spectral decomposition of $\mathbf{M}$,

$$\mathbf{M} = \sum_{i=1}^{d} \lambda_i v_i v_i^T \tag{5.35}$$

where λ_i are the eigenvalues of the matrix $\mathbf{M}$ and v_i are the corresponding eigenvectors, we see that

$$0 = \mathbf{H}\mathbf{M}\mathbf{H}^T = \sum_{i=1}^{d} \lambda_i(\mathbf{H}(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{H}\mathbf{A})^n v_i)^2 \tag{5.36}$$

for all n. Since $\mathbf{M} \neq 0$ there is $\lambda_j > 0$ and hence $\mathbf{H}(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{H}\mathbf{A})^n v_j = 0$ for all n, which contradicts the observability of $(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{H}\mathbf{A}, \mathbf{H})$. Thus $\mathbf{M} = 0$, finishing the proof. Therefore $\boldsymbol{\kappa}_\infty$ is the unique minimiser. □

Take the parameter space $\mathcal{K}$ to be defined as $\mathcal{K} = \{\mathbf{K}; \sigma(\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A}) \leq 1 - \epsilon\} \cap \mathcal{K}_0$, where $\mathcal{K}_0$ is a compact region. We will find ϵ and $\mathcal{K}_0$ later, see 5.72.

Since $\mathbf{K}$ is a stabilising feedback gain in $\mathcal{K}$, ψ is well defined and given $\mathcal{K}$ as above we can confirm point (2) in theorem 5.1.1.

Lemma 5.1.4. *Let $\sigma(\mathbf{X})$ denote the spectral radius of $\mathbf{X}$. Then we have $\psi_n(\mathbf{K}, \Gamma_0) \to \psi_\infty(\mathbf{K}, \Gamma_0)$ uniformly if and only if $\sigma(\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A}) \leq 1 - \epsilon$ and $\mathbf{K} \in \mathcal{K}$ where $\mathcal{K} = \{\mathbf{K}; \sigma(\mathbf{A} - \mathbf{K}\mathbf{H}\mathbf{A}) \leq 1 - \epsilon\} \cap \mathcal{K}_0$ is compact.*

Proof. We need to show that $\left\|\mathbf{H}(\Gamma_{n+l} - \Gamma_n)\mathbf{H}^T\right\| \leq C\lambda^n$ for some $\lambda < 1$. Consider the trace of this matrix;

$$\text{trace}[\mathbf{H}(\Gamma_{n+l} - \Gamma_n)\mathbf{H}^T] \leq d' \left\|\mathbf{H}^T\mathbf{H}\right\| \|(\Gamma_{n+l} - \Gamma_l)\| \tag{5.37}$$

where the inequality follows by definition and d' is the dimension of the system. Note that

$$\Gamma_{n+l} - \Gamma_n = (\mathbf{A} - \mathbf{KHA})^n(\Gamma_l - \Gamma_0)\,(\mathbf{A} - \mathbf{KHA})^{n\,T}\,. \tag{5.38}$$

Let $\mathbf{W} = (\mathbf{A} - \mathbf{KHA})$. Since $\mathbf{K}$ is chosen to stabilise the system, as well as minimise the mean squared error, we know that the eigenvalues of the stability matrix must lie within the unit circle. Therefore, if $\sigma(\mathbf{X})$ denotes the spectral radius of a matrix $\mathbf{X}$, we have that $\sigma(\mathbf{W}) \leq 1 - \epsilon$. Now consider taking the trace of (5.38);

$$\begin{aligned}
\text{trace}\left[\mathbf{W}^n(\Gamma_l - \Gamma_0)\,\mathbf{W}^{n\,T}\right] &\leq d.\sigma\left[(\Gamma_l - \Gamma_0)\,\mathbf{W}^{n\,T}\,\mathbf{W}^n\right] && (5.39)\\
&= d\left\|(\Gamma_l - \Gamma_0)\,\mathbf{W}^{n\,T}\,\mathbf{W}^n\right\| && (5.40)\\
&\leq d\,\|(\Gamma_l - \Gamma_0)\|\,\|\mathbf{W}^q\|^{\frac{1}{q}2n} && (5.41)
\end{aligned}$$

where d is the dimension of the system. The inequality in (5.39) follows by definition and (5.40) is obtained as an equality because the matrix is symmetric; (5.41) follows from the definition of a norm.

Since $\sigma(\cdot)$ denotes the spectral radius, we can write

$$\sigma(\mathbf{W}) = \lim_{q\to\infty} \|\mathbf{W}^q\|^{\frac{1}{q}} \tag{5.42}$$

and since $\sigma(\mathbf{W}) \leq 1 - \epsilon$ it follows that

$$\|\mathbf{W}^q\|^{\frac{1}{q}} \leq 1 - \epsilon. \tag{5.43}$$

Therefore, using this fact in (5.41) we see that

$$\text{trace}\left[\mathbf{W}^n(\Gamma_l - \Gamma_0)\,\mathbf{W}^{n\,T}\right] \leq d\,\|(\Gamma_l - \Gamma_0)\|\,(1-\epsilon)^n. \tag{5.44}$$

Since we are working on the compact parameter space $\mathcal{K}$, we can establish that q and ϵ are independent of $\mathbf{K}$. The value q that satisfies the bound above depends on $\mathbf{K}$ but the same q is valid for an open neighbourhood of the matrix $\mathbf{W}$. Since $\mathcal{K}$ is compact there can only be finitely many q's that satisfy (5.43). Therefore we can choose the largest such q and corresponding ϵ to get a uniform upper bound. Therefore set $C = d\,\|(\Gamma_l - \Gamma_0)\|$ and $\lambda = 1 - \epsilon$ and it follows that

$$\text{trace}[\mathbf{H}(\Gamma_{n+l} - \Gamma_n)\mathbf{H}^T] \leq d'\left\|\mathbf{H}^T\mathbf{H}\right\| C\lambda^n = C'\lambda^n \tag{5.45}$$

where we simply let $C' = d'\left\|\mathbf{H}^T\mathbf{H}\right\| C$, to obtain the required result.

To prove the other direction assume $\psi_n(\mathbf{K}, \Gamma_0) \to \psi_\infty(\mathbf{K}, \Gamma_0)$. Then we have that

$$\begin{aligned}\psi_n(\mathbf{K}, \Gamma_0) \;&= \mathbf{H}(\mathbf{A} - \mathbf{KHA})\Gamma_{n-1}(\mathbf{K}, \Gamma_0)(\mathbf{A} - \mathbf{KHA})^T\mathbf{H}^T \\ &\quad + \mathbf{H}(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T\mathbf{H}^T + \mathbf{HKRK}^T\mathbf{H}^T\end{aligned} \tag{5.46}$$

which implies by continuity that

$$\begin{aligned}\psi_\infty(\mathbf{K}, \Gamma_0) \;&= \mathbf{H}(\mathbf{A} - \mathbf{KHA})\Gamma_\infty(\mathbf{K}, \Gamma_0)(\mathbf{A} - \mathbf{KHA})^T\mathbf{H}^T \\ &\quad + \mathbf{H}(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T\mathbf{H}^T + \mathbf{HKRK}^T\mathbf{H}^T\end{aligned} \tag{5.47}$$

from which it follows that $\mathbf{K}$ must stabilise the system. To see this suppose that it doesn't so that we have $(\mathbf{A} - \mathbf{KHA})^T\omega = \lambda\omega$ for some λ with $|\lambda| \geq 1$ and non-zero ω. Since $(\mathbf{A} - \mathbf{KHA}, \mathbf{H})$ is an observable pair, we only need to consider eigenvalues of the form $\omega = \mathbf{H}^T x$. This follows from the duality of controllability and observability (see Section 4.3). It follows then that $(\mathbf{A} - \mathbf{KHA})^T\mathbf{H}^T x = \lambda\mathbf{H}^T x$.

By rearranging equation (5.47) we get that

$$\begin{aligned} & \mathbf{H}\left\{\Gamma_\infty - (\mathbf{A}-\mathbf{KHA})\Gamma_\infty(\mathbf{A}-\mathbf{KHA})^T\right\}\mathbf{H}^T \\ = \quad & \mathbf{H}\left\{(\mathbb{1}-\mathbf{KH})\mathbf{Q}(\mathbb{1}-\mathbf{KH})^T + \mathbf{KRK}^T\right\}\mathbf{H}^T \end{aligned} \tag{5.48}$$

from which it follows that

$$(1-|\lambda|^2)x^T\mathbf{H}\Gamma_\infty\mathbf{H}^Tx = x^T\mathbf{H}\left\{(\mathbb{1}-\mathbf{KH})\mathbf{Q}(\mathbb{1}-\mathbf{KH})^T + \mathbf{KRK}^T\right\}\mathbf{H}^Tx. \tag{5.49}$$

The left hand side of the above equation is non-positive while the right hand side is clearly non-negative, therefore for the equation to make sense both sides must be equal to zero. This implies that $(\mathbf{HK})^Tx = 0$ and $x^T\mathbf{HQH}^Tx = 0$. By our assumption that $\mathbf{K}$ doesn't stabilise the system, $(\mathbf{HK})^Tx = 0$ implies that $\mathbf{A}^T\mathbf{H}^Tx = \lambda\mathbf{H}^Tx$ which together with $x^T\mathbf{HQH}^Tx = 0$ implies a lack of controllability since

$$\mathbf{A}^T\mathbf{H}^Tx = \lambda\mathbf{H}^Tx \quad \Rightarrow \quad x^T\mathbf{HQH}^Tx = 0 \tag{5.50}$$

follows from the definition of controllability given in definition 4.2.2. Therefore, $\mathbf{K}$ must be a stabilising gain. □

This then just leaves the assumption that ψ_n has a minimiser and point (2) in theorem 5.1.1 to be checked and ϵ and $\mathcal{K}_0$ to be determined. We shall prove these together.

Lemma 5.1.5. *There is $n_0 \in \mathbb{N}$, $\delta > 0$ so that for any $n \geq n_0$, if ψ_n has a minimiser $\mathbf{K}_n$, we must have $\psi_n(\mathbf{K}_n) \leq \mathbf{R}(\mathbb{1}-\delta)$.*

Proof. Consider first the following calculation

$$\begin{aligned}\mathbf{H}\Gamma_\infty(\boldsymbol{\kappa}_\infty)\mathbf{H}^T &= \mathbf{H}\Sigma_\infty\mathbf{H}^T - \mathbf{H}\Sigma_\infty\mathbf{H}^T(\mathbf{H}\Sigma_\infty\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\Sigma_\infty\mathbf{H}^T \\ &= \mathbf{H}\Sigma_\infty\mathbf{H}^T - (\mathbf{H}\Sigma_\infty\mathbf{H}^T+\mathbf{R}-\mathbf{R})(\mathbf{H}\Sigma_\infty\mathbf{H}^T+\mathbf{R})^{-1}(\mathbf{H}\Sigma_\infty\mathbf{H}^T+\mathbf{R}-\mathbf{R}) \\ &= \mathbf{R} - \mathbf{R}(\mathbf{H}\Sigma_\infty\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{R} \\ &< \mathbf{R}\end{aligned} \tag{5.51}$$

since $\mathbf{R} > 0$ and so we have $\psi(\boldsymbol{\kappa}_\infty) < \mathbf{R}$. Then we can say $\psi(\boldsymbol{\kappa}_\infty) \leq \mathbf{R}(\mathbb{1} - 2\delta)$. Since $\psi_n(\boldsymbol{\kappa}_\infty) \to \psi_\infty(\boldsymbol{\kappa}_\infty)$, as established in lemma 5.1.4, we can pick n_0 large enough so that

$$\psi_n(\boldsymbol{\kappa}_\infty) \leq \mathbf{R}(\mathbb{1} - \delta), \quad \forall n \geq n_0. \tag{5.52}$$

If ψ_n has a minimiser, $\mathbf{K}_n$, then this potential minimiser has to be better than $\boldsymbol{\kappa}_\infty$. Otherwise, we may as well use the asymptotic gain in the algorithm. Hence, using the bound in (5.52) we have the required result. □

Note that if $\mathbf{K}$ is such that $\psi_n(\mathbf{K}) \leq \mathbf{R}(\mathbb{1} - \delta)$ for all $n \geq n_0$ then it also true that

$$\psi(\mathbf{K}) \leq \mathbf{R}(\mathbb{1} - \delta). \tag{5.53}$$

Let $\mathcal{K}_0 := \{\mathbf{K}; \psi(\mathbf{K}) \leq \mathbf{R}(\mathbb{1} - \delta)\}$. Then it follows that this set is closed and will contain any potential minimiser of $\psi_n, n > n_0$.

Lemma 5.1.6. *The set $\mathcal{K}_0$ as defined above is compact.*

Proof. Suppose we are on $\mathcal{K}_0$ and $n \geq n_0$. We begin by writing, for each $k = 1, \ldots, n$,

$\Gamma_k(\mathbf{K}, \Gamma_0)$ as

$$\begin{aligned}\Gamma_k(\mathbf{K}, \Gamma_0) &= \mathbf{W}^k \Gamma_0 \, {\mathbf{W}^k}^T + \sum_{i=0}^{k-1} \mathbf{W}^i (\mathbb{1} - \mathbf{KH}) \mathbf{Q} (\mathbb{1} - \mathbf{KH})^T \, {\mathbf{W}^i}^T \\ &\quad + \sum_{i=0}^{k-1} \mathbf{W}^i \mathbf{KRK}^T \, {\mathbf{W}^i}^T\end{aligned} \tag{5.54}$$

where $\mathbf{W} = (\mathbf{A} - \mathbf{KHA})$. Since Γ_n is a covariance matrix it is non-negative definite and will remain non-negative definite if we multiply from the left and right by $\mathbf{H}$. The individual terms on the right hand side of (5.54) are all non-negative definite, so we have that

$$\mathbf{R}(\mathbb{1} - \delta) \geq \psi_n \geq \mathbf{H} \Gamma_n \mathbf{H}^T \geq \sum_{k=0}^{n-1} \mathbf{H} \mathbf{W}^k \mathbf{KRK}^T \, {\mathbf{W}^k}^T \mathbf{H}^T. \tag{5.55}$$

Note that since the terms in (5.54) are all non-negative definite, they satisfy the bound individually.

When $n = 1$ we have that $\mathbf{R}(\mathbb{1} - \delta) \geq (\mathbf{HK})^2 \mathbf{R}$ which implies that $\mathbb{1} - \delta \geq (\mathbf{HK})^2$. This means that $\mathbf{HK}$ is bounded below and it follows that the pair $(\mathbf{H}, \mathbf{A} - \mathbf{KHA})$ is always observable on $\mathcal{K}$. This is because, as was explained in Section 4.3, the pair $(\mathbf{H}, \mathbf{A} - \mathbf{KHA})$ is observable when $\mathbf{HK} \neq \mathbb{1}$. Thus since $\mathbf{HK}$ is bounded below by $\mathbb{1} - \delta$, the pair $(\mathbf{H}, \mathbf{A} - \mathbf{KHA})$ is observable.

When $n = 2$ we get, after performing a similar calculation, that

$$\mathbb{1} - \delta > (\mathbf{HK})^2 + ((\mathbb{1} - \mathbf{HK})\mathbf{HAK})^2, \tag{5.56}$$

which implies that $\mathbf{HAK}$ must also be bounded. Repeating the argument for increasing powers of n up to $n - 1$ yields the implication that $\mathbf{HK}, \mathbf{HAK}, \ldots, \mathbf{HA}^{n-1}\mathbf{K}$, all be bounded below since we get the expression

$$\mathbb{1} - \delta \geq \mathbf{HK} + (\mathbb{1} - \mathbf{HK})\mathbf{HAK} + \ldots + (\mathbb{1} - \mathbf{HK})\mathbf{HA}^{n-1}\mathbf{K}. \tag{5.57}$$

However notice that this is simply the matrix $(\mathbb{1} - \mathbf{HK})\mathcal{O}(\mathbf{A}, \mathbf{H})$ applied to $\mathbf{K}$. So we have that $\mathcal{O}(\mathbf{A}, \mathbf{H}) \cdot \mathbf{K}$ must be bounded. But since we have assumed complete observability, $\mathcal{O}(\mathbf{A}, \mathbf{H})$ is invertible. Thus $\mathcal{K}_0$ is bounded. We have already established that $\mathcal{K}_0$ is closed, therefore it is compact. □

So far then, we have established that ψ_n has a minimiser on $\mathcal{K}_0$ because $\mathcal{K}_0$ is compact and ψ_n is continuous. We need to ensure that this minimiser exists on $\mathcal{K}$, that is, in addition the minimising gain must satisfy $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$, i.e. $\mathbf{K}$ must stabilise the error dynamics.

Lemma 5.1.7. *There is a constant C so that if $n \geq n_0$, then*

$$\psi_n(\mathbf{K}) \geq C\frac{1 - (1 - \epsilon)^{2n}}{1 - (1 - \epsilon)^2} \tag{5.58}$$

for all $\mathbf{K} \in \mathcal{K}_0$ with $\sigma(\mathbf{A} - \mathbf{KHA}) \geq 1 - \epsilon$.

Before we prove this lemma, we prove the following results as they will be needed in the proof.

Lemma 5.1.8. *There exists $c > 0$ so that for all $v \in \mathbb{C}^D$, $\|v\| = 1$ and $\lambda \in \mathbb{C}^D$ with $v^T(\mathbf{A} - \mathbf{KHA}) = \lambda v^T$ so that*

$$\forall \epsilon > 0 \exists \delta : if \left\|v^T\mathbf{K}\right\| < \delta \Rightarrow v^T\mathbf{Q}v \geq c - \epsilon. \tag{5.59}$$

Proof. By controllability, if $\omega^T\mathbf{A} = \lambda\omega^T \Rightarrow \omega^T\mathbf{Q}\omega > 0$. Take $c = \min\{\omega^T\mathbf{Q}\omega; \omega^T\mathbf{A} = \lambda\omega^T, \|\omega\| = 1\}$. Since $\mathbf{Q}$ is non-degenerate on every eigenspace of $\mathbf{A}$ and there are finitely many distinct eigenvalues we have $c > 0$.

Suppose the claim is not true. Then there exist sequences $\mathbf{K}_n, v_n$ with $\left\|v^T\mathbf{K}\right\| \to 0$ but $v_n^T\mathbf{Q}v_n \leq c - \epsilon$, satisfying

$$v_n^T(\mathbf{A} - \mathbf{K}_n\mathbf{HA}) = \lambda_n v_n^T \tag{5.60}$$

Since $\|v_n\| = 1$, we take subsequences so that $v_n \to v$. Taking the limit yields $v^T\mathbf{A} = \lambda v^T$. Then $\lambda_n \to \lambda$ as all other terms in (5.60) converge. But this means $v^T\mathbf{Q}v \geq c$. □

Corollary 5.1.1. *There exists $\alpha > 0$ so that for all v, $\|v\| = 1$ and $v^T(\mathbf{A} - \mathbf{KHA}) = \lambda v^T$,*

$$v^T\mathbf{KRK}^Tv + v^T(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T \geq \alpha. \tag{5.61}$$

Proof. Since $\mathbf{R} > 0$, there exists $r > 0$ such that $v^T\mathbf{KRK}^Tv \geq r\left\|v^T\mathbf{K}\right\|^2$. Further, $v^T(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^Tv = v^T\mathbf{Q}v + f(v^T\mathbf{K})$, $f(0) = 0$.

Let $\epsilon > 0$ so that $c - 2\epsilon$, c as in the above lemma. Now pick δ so small that if $\left\|v^T\mathbf{K}\right\| \leq \delta$, $v^T\mathbf{Q}v \geq c - \epsilon$, $|f(v^T\mathbf{K})| \leq \epsilon$ by the above lemma and the continuity of f.

Hence $v^T\mathbf{KRK}^Tv + v^T(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T \geq C - 2\epsilon$. If $\left\|v^T\mathbf{K}\right\| \geq \delta$, then $r\left\|v^T\mathbf{K}\right\|^2 \geq r\delta^2$. We can pick $\alpha = \min\{c - 2\epsilon, r\delta^2\} > 0$. □

We can now prove lemma 5.1.7.

Proof of Lemma 5.1.7. Using the results presented in lemma 5.1.8 and corollary 5.1.1, for $v \in \mathbb{C}^D$ with $\|v\| = 1$ and $|\lambda| \geq 1 - \epsilon$ we get that

$$v^T\Gamma_n v \geq \frac{1 - |\lambda|^{2n}}{1 - |\lambda|^2} \cdot \alpha \tag{5.62}$$

where α is as given in corollary 5.1.1. This equation is obtained by writing out in full the expression for $v^T\Gamma_n v$, and noting that $v^T(\mathbf{A} - \mathbf{KHA}) = \lambda v^T$. Using this together together with corollary 5.1.1 the result follows.

By the properties of the trace of a matrix it follows that

$$\mathrm{tr}(\Gamma_n) \geq \frac{1 - |\lambda|^{2n}}{1 - |\lambda|^2} \cdot \alpha. \tag{5.63}$$

Define $U(\mathbf{K}) := \mathcal{O}(\mathbf{A} - \mathbf{KHA}, \mathbf{H})$ and note that since the pair $(\mathbf{H}, \mathbf{A} - \mathbf{KHA})$ is observable, as explained in the proof of lemma 5.1.6, $U(\mathbf{K})$ is invertible on $\mathcal{K}_0$.

Now for any two non-negative definite matrices $\mathbf{X}$ and $\mathbf{Y}$ we have by the Cauchy-Schwartz inequality (see for example Hunter & Nachtergaele (2001)), that

$$\mathrm{tr}(\mathbf{XY}) \leq \sqrt{\mathrm{tr}(\mathbf{X}^2)\mathrm{tr}(\mathbf{Y}^2)}. \tag{5.64}$$

Since $\mathbf{X}$ and $\mathbf{Y}$ are non-negative definite in our case it follows that $\mathrm{tr}(\mathbf{X}^2) \leq \mathrm{tr}(\mathbf{X})^2$ so that (5.64) becomes

$$\mathrm{tr}(\mathbf{XY}) \leq \mathrm{tr}(\mathbf{X})\mathrm{tr}(\mathbf{Y}). \tag{5.65}$$

Bearing this in mind consider

$$\begin{aligned} \mathrm{tr}(\Gamma_n) &= \mathrm{tr}\left([U(\mathbf{K})^T U(\mathbf{K})]^{-1}[U(\mathbf{K})^T U(\mathbf{K})]\Gamma_n\right) \\ &\leq \mathrm{tr}\left([U(\mathbf{K})^T U(\mathbf{K})]^{-1}\right) \cdot \mathrm{tr}\left(U(\mathbf{K})\Gamma_n U(\mathbf{K})^T\right). \end{aligned} \tag{5.66}$$

The first term on the right hand side of the above equation is bounded by some constant C' since $\mathbf{K} \in \mathcal{K}_0$. As for the second term on the right hand side of (5.66) consider,

$$\begin{aligned} \mathrm{tr}\left(U(\mathbf{K})\Gamma_n U(\mathbf{K})^T\right) &= \sum_{\substack{i,j \\ d-1}} U_{ij}\Gamma_n U_{jk} \\ &= \sum_{k=0} \mathbf{H}(\mathbf{A} - \mathbf{KHA})^k \Gamma_n \left(\mathbf{A} - \mathbf{KHA}\right)^{k\,T} \mathbf{H}^T \\ &\leq \mathbf{H}\Gamma_n\mathbf{H}^T + \mathbf{H}\Gamma_{n+1}\mathbf{H}^T + \ldots + \mathbf{H}\Gamma_{d+n-1}\mathbf{H}^T \end{aligned} \tag{5.67}$$

These inequalities follow from the explicit expression for Γ given in (5.7). Since all the terms individually on the right hand side of the above are non-negative definite, it follows that one of them is bounded below by the left hand side divided by the dimension, in this case d. As it holds for all the terms individually, the bound holds for $\mathbf{H}\Gamma_n\mathbf{H}^T$ in particular, so that we have

$$\mathbf{H}\Gamma_n\mathbf{H}^T \geq \frac{\mathrm{tr}\left(U(\mathbf{K})\Gamma_n U(\mathbf{K})^T\right)}{d}. \tag{5.68}$$

By substituting the above into (5.66) and using the bound given in (5.63), it follows

that

$$\frac{1-|\lambda|^{2n}}{1-|\lambda|^2}\cdot\frac{\alpha}{dC'}\leq \mathbf{H}\Gamma_n\mathbf{H}^T \tag{5.69}$$

which with $\lambda = 1-\epsilon$, is the required result. □

Now take ϵ_0 so small that $\sigma(\mathbf{A}-\boldsymbol{\kappa}_\infty\mathbf{HA})\leq 1-\epsilon_0$. We have to find $\epsilon<\epsilon_0$ and $n_1\geq n_0$ so that

$$\frac{1-(1-\epsilon)^{2n_1}}{1-(1-\epsilon)^2}\geq \mathbf{R}. \tag{5.70}$$

Take $n_1=\max(S,n_0)$, then by de L'Hôpital's Rule

$$\frac{1-(1-\epsilon)^{2n_1}}{1-(1-\epsilon)^2}\overset{\epsilon\to 0}{\to} S \tag{5.71}$$

where S will be defined shortly. Take ϵ small so that $1-(1-\epsilon)^{2n_1}/1-(1-\epsilon)^2 > S/2$, where we take S so that

$$\frac{\alpha}{dC'}\frac{S}{2}\geq \mathbf{R}. \tag{5.72}$$

Therefore by lemma 5.1.7, if $n\geq n_1$, $\mathbf{K}\in\mathcal{K}$ with $\sigma(\mathbf{A}-\mathbf{KHA})\geq 1-\epsilon$, then $\psi_n(\mathbf{K})\geq\mathbf{R}$. This means that such a $\mathbf{K}$ cannot be a minimiser of ψ_n as soon as $n\geq n_1$, proving the remaining facts in theorem 5.1.1. Thus we have the following final result.

Theorem 5.1.2. *The feedback gain matrix* $\mathbf{K}$ *that minimises the out-of-sample error,* $\psi_n(\mathbf{K})$*, over the compact set* $\mathcal{K}$ *and stabilises the system, converges to the asymptotic Kalman Gain* $\boldsymbol{\kappa}_\infty$ *in the limit of large observational windows.*

Proof. It has been established above that the four points given in theorem 5.1.1 are satisfied by our problem. The claim then follows. □

Minimising the State Error The numerical experiments in Chapter 3 suggested that for linear systems, the out-of-sample error is equivalent (in a certain sense) to the asymptotic state error covariance. In this context equivalent means that minimising the out-of-sample

error covariance is equivalent to minimising the state error covariance. This can be easily seen as follows.

It was established in proposition 5.1.1 that the asymptotic Kalman Gain $\boldsymbol{\kappa}_\infty$, uniquely minimises the asymptotic out-of-sample error covariance, $\mathbf{H}\Gamma_\infty\mathbf{H}^T + \mathbf{R}'$. Consider the fixed point equation for the asymptotic state error covariance,

$$\Gamma_\infty = (\mathbf{A} - \mathbf{KHA})\Gamma_\infty(\mathbf{A} - \mathbf{KHA})^T + (\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T + \mathbf{KRK}^T. \tag{5.73}$$

Taking the derivative (assuming it exists) of this with respect to $\mathbf{K}$ and setting equal to zero yields

$$0 = \mathrm{D}\Gamma_\infty(\mathbf{K}).\Delta = \Delta\left[\mathbf{RK}^T - \mathbf{HQ}(\mathbb{1} - \mathbf{KH})^T - \mathbf{HA}\Gamma_\infty(\mathbf{A} - \mathbf{KHA})^T\right] \tag{5.74}$$

$$+ \left[\mathbf{KR} - (\mathbb{1} - \mathbf{KH})\mathbf{QH}^T - (\mathbf{A} - \mathbf{KHA})\Gamma_\infty\mathbf{A}^T\mathbf{H}^T\right]\Delta^T \tag{5.75}$$

and since this must hold for all Δ the minimising $\mathbf{K}$ must satisfy

$$\begin{aligned} 0 &= \mathbf{KR} - (\mathbb{1} - \mathbf{KH})\mathbf{QH}^T - (\mathbf{A} - \mathbf{KHA})\Gamma_\infty\mathbf{A}^T\mathbf{H}^T \\ &= \mathbf{K}(\mathbf{H}\Sigma_\infty\mathbf{H}^T + \mathbf{R}) - \Sigma_\infty\mathbf{H}^T \\ &\Rightarrow \mathbf{K} = \Sigma_\infty\mathbf{H}^T(\mathbf{H}\Sigma_\infty\mathbf{H}^T + \mathbf{R})^{-1} = \boldsymbol{\kappa}_\infty \end{aligned}$$

where $\Sigma_\infty = \mathbf{A}\Gamma_\infty\mathbf{A}^T + \mathbf{Q}$.

It follows that this is the unique optimal gain matrix. Therefore, both the state and out-of-sample error covariances are minimised uniquely by the same feedback gain matrix. Hence, in this sense they are equivalent.

Chapter Summary In this chapter we have presented the proof that the constant gain matrix that minimises the expected out-of-sample error exists. We considered constant gain matrices as they lead to simpler filters as the optimal gain matrix does not need

to be updated at every step, avoiding the matrix inversion required for the traditional Kalman Filter. Further to this it was established that such a gain matrix converges to the asymptotic gain in the limit of large observational windows. The asymptotic limit mentioned here is the limit of the Kalman Gain defined in Chapter 4.

This fact was established by first constructing the compact space in which we are working. This then led to the fact that the out-of-sample error was minimised by a feedback gain that always entered the region in which it stabilised the error dynamics. Using the fact that the asymptotic gain is the unique minimiser of the asymptotic observed error, the conclusion that the minimiser converges to the asymptotic Kalman Gain, $\boldsymbol{\kappa}_\infty$, was obtained. Some comments on the equivalence of minimising the state and out-of-sample error covariances were also made.

Chapter 6

Minimising the Empirical Mean of the Error

It has been established that the minimiser, $\mathbf{K}_n$, of the expected observed or out-of-sample error converges to the asymptotic Kalman gain, $\boldsymbol{\kappa}_\infty$ in the limit of large observational windows. However in practice, only an estimate of this minimiser is available. In practical situations and in fact in our numerical experiments in Chapter 3, the errors are estimated by the empirical mean namely,

$$\frac{1}{n}\sum_{i=1}^{n}(z_i - x_i)^2 \quad \text{or} \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - \eta_i')^2. \tag{6.1}$$

leading to estimators of the optimal gain. Therefore it is desirable to establish that the estimator $\hat{\boldsymbol{\kappa}}_n$, used to estimate $\mathbf{K}_n$, converges in probability to the asymptotic Kalman gain.

The problem then is as follows. Given that the estimator $\hat{\boldsymbol{\kappa}}_n$, minimises the function ϕ_n where ϕ_n is the empirical mean that estimates the out-of-sample error and that the asymptotic Kalman gain minimises the asymptotic out-of-sample error uniquely, is it true that $\hat{\boldsymbol{\kappa}}_n \to \boldsymbol{\kappa}_\infty$ as $n \to \infty$ in probability? Numerical evidence presented in Chapter 3 and in Mallia-Parfitt & Bröcker (2016) suggests that this is the true for linear dynamical

systems with gaussian perturbations and linear observations. We shall now prove this.

This is achieved by observing that the estimator considered essentially minimises a sum of functions of observed data and as such can be thought of as an *M-estimator* (see Chapter 1). Using this approach and results which exist to prove consistency of such estimators, we shall endeavor to prove that the minimising gain of the empirical mean of the out-of-sample error, converges to the asymptotic gain in the limit of large observational windows.

The proof used in this chapter is similar to the one presented in Chapter 5 for the deterministic case. Unfortunately, however there is one piece of the proof that is missing in this stochastic case. As part of the proof, we require that the probability for a minimiser to be stabilising goes to 1 for large n. We cannot however state this for certain as we cannot say that all potential minimisers stabilise the error dynamics. A full explanation of the problem is given in detail at the end of the chapter.

The theory of M-estimators and their properties, such as consistency and asymptotic normality, is covered in detail in Van der Vaart (2000), Ferguson (1996). A brief of overview of the main results in the theory of M-estimators is presented here.

6.1 Theory of M-Estimators

The work presented in this section is the general theory that motivated our approach. We present the theory of M-estimators and conditions for which such estimators are asymptotically consistent in a general context. The results given here are not applicable without compactness of the parameter space in which we are working. The content is obtained from Van der Vaart (2000).

Suppose we are interested in a parameter θ attached to the distribution of some observations, X_i for $i = 1, \ldots, n$. A popular method for finding an estimator $\hat{\theta}_n$ is to

maximise (or minimise) the criterion function of the type

$$\phi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f_\theta(X_i). \tag{6.2}$$

An estimator maximising ϕ_n over a set Θ, is called an M-estimator and we are interested in the asymptotic behaviour of sequences of M-estimators.

Usually the maximising (or minimising) value is found by setting a derivative equal to zero. Thus the term M-estimator is also used for estimators satisfying systems of equations of the form

$$\Psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i) = 0. \tag{6.3}$$

Such equations that define an estimator, are known collectively as *estimating equations* and when it corresponds to a maximisation problem, it is called a *Z-estimator*, however the name M-estimators is widely used. An example of an M-estimator is the maximum likelihood estimator, Van der Vaart (2000). To see this suppose $X_1, \ldots, X_n$ have a common density p_θ. Then the maximum likelihood estimator maximises the log likelihood

$$\theta \mapsto \sum_{i=1}^{n} \log p_\theta(X_i). \tag{6.4}$$

Thus, a maximum likelihood estimator is an M-estimator as in (6.2) with $f_\theta = \log p_\theta$. If the density is partially differentiable with respect to θ for each fixed x, then the maximum likelihood estimator also solves an equation of type (6.3), with m_θ equal to the vector of partial derivatives.

A note of interest is that the definition (6.2) of an M-estimator may apply in cases where (6.3) does not. For example, if $X_1, \ldots, X_n$ are iid according to the uniform distribution on $[0, \theta]$, then it makes sense (by defining $\log 0 = -\infty$) to maximise the log likelihood

$$\theta \mapsto \sum_{i=1}^{n} \left(\log \mathbb{1}_{[0,\theta]}(X_i) - \log \theta\right). \tag{6.5}$$

However, this function is not smooth in θ and there exists no natural version of (6.3). Thus, in this example the definition as the location of a maximum is more fundamental than the definition as a zero.

6.1.1 Consistency of M-Estimators

Since the estimator $\hat{\theta}_n$ is used to estimate the parameter θ, it would be ideal if the sequence converges in probability to θ. If this is the case for every possible parameter value, then the sequence of estimators is *consistent*, Van der Vaart (2000). For example the sample mean, $\overline{X}_n$ is asymptotically consistent for the population mean, $\mathbb{E}X$, provided it exists. This follows from the law of large numbers. This naturally extends to other sample characteristics, such as the sample median which is consistent for the population median. The question that follows then is what can be said about M-estimators in general?

Suppose that the M-estimator $\hat{\theta}_n$ maximises $\phi_n(\theta)$. The asymptotic value of $\hat{\theta}_n$ depends on the asymptotic value of ϕ_n. By the Law of Large Numbers we may have that

$$\phi_n(\theta) \xrightarrow{P} \phi(\theta) = \mathbb{E}f_\theta \tag{6.6}$$

for every θ, provided the expectation exists. The letter P above the arrow indicates convergence in probability. Convergence as given in (6.6) is not quite enough. Uniform convergence is needed.

It seems reasonable to expect that the maximiser $\hat{\theta}_n$ of ϕ_n converges to the maximising value θ_0 of ϕ. The main result that proves this is given in the following theorem, (Van der Vaart 2000). The theorem statement and proof are reproductions of theorem 5.7 in Van der Vaart (2000).

Theorem 6.1.1. *Let ϕ_n be random functions and let ϕ be a fixed function of θ such that*

$$\sup_{\theta\in\Theta} |\phi_n(\theta) - \phi(\theta)| \xrightarrow{P} 0, \tag{6.7}$$

and for all $\epsilon > 0$

$$\sup_{\theta: d(\theta,\theta_0)\geq\epsilon} \phi_n(\theta) < \phi(\theta_0). \tag{6.8}$$

Then any sequence of estimators, $\hat{\theta}_n$ *with* $\phi_n(\hat{\theta}_n) \geq \phi_n(\theta_0) - o_P(1)$ *converges in probability to* θ_0*.*

Proof. We have that $\phi_n(\hat{\theta}_n) \geq \phi_n(\theta_0) - o_P(1)$. We know that $\phi_n(\theta_0) \xrightarrow{P} \phi(\theta_0)$, therefore $\phi_n(\hat{\theta}_n) \geq \phi(\theta_0) - o_P(1)$ and hence

$$0 \leq \phi(\theta_0) - \phi(\hat{\theta}_n) \leq \phi_n(\hat{\theta}_n) - \phi(\hat{\theta}_n) + o_P(1) \xrightarrow{P} 0 \tag{6.9}$$

since

$$\sup_{\theta} |\phi_n(\theta) - \phi(\theta)| + o_P(1) \xrightarrow{P} 0 \tag{6.10}$$

by assumption. By the second part of the assumption, there exists for every $\epsilon > 0$ a number $\eta > 0$ such that $\phi(\theta) < \phi(\theta_0) - \eta$ for every θ with $d(\theta, \theta_0) \geq \epsilon$. The event $\{d(\hat{\theta}_n, \theta_0) \geq \epsilon\}$ is contained in the event $\{\phi_n(\hat{\theta}_n) < \phi(\theta_0) - \eta\}$ and the probability of this converges to zero. In the above, we can replace $o_p(1)$ with ϵ as there is no sign specified. □

The conditions of the theorem contain a stochastic and a deterministic part. The deterministic part, equation (6.8), ensures that the maximum θ_0 is a unique maximiser and also that it is a well-separated point of maximum of ϕ. This means that only parameters close to θ_0 may yield a value of $\phi(\theta)$ close to the maximum value $\phi(\theta_0)$. The stochastic condition, equation (6.7), requires uniform convergence of ϕ_n.

6.1.2 Conditions for Consistency of M-Estimators

The above approach to prove consistency has two requirements; one deterministic and one stochastic. We shall discuss these requirements separately and determine a set of conditions which guarantee these.

First we shall discuss conditions for which the maximiser θ_0 is a well-separated point of maximum. A sufficient set of conditions is given in lemma 6.2.1, see problem 5.27 in Van der Vaart (2000). This result tells us that uniqueness of the minimiser for continuous functions on a compact space are the conditions required to establish that the minimiser is a well-separated (or isolated) point of minimum. The stochastic condition in theorem 6.1.1 requires uniform convergence of ϕ_n. In our situation, the asymptotic gain is the expected value of the gain minimising the actual error, not its empirical mean. Therefore, we are interested in more generic random functions and we need a method to prove uniform convergence in probability.

In this spirit, let $G_n(\theta)$ be a generic sequence of random functions, that we consider to be given by

$$G_n(\theta) = \phi_n(\theta) - \phi(\theta). \tag{6.11}$$

Then we have the following theorem for generic Uniform Convergence, Newey (1991). This theorem uses the concept of stochastic equicontinuity (which we discuss in more detail later) and pointwise convergence to characterise uniform convergence on a compact set. It is a stochastic generalisation of the continuous result with the same goal, see Rudin (1964). It is motivated by its relationship to well known results on weak convergence of stochastic processes, e.g Billingsley (1968).

Theorem 6.1.2. *If Θ is a compact space, $G_n(\theta) \xrightarrow{P} 0$, $\forall\theta \in \Theta$ and $\{G_n(\theta) : n \geq 1\}$ is stochastically equicontinuous, then*

$$\sup_{\theta\in\Theta} |G_n(\theta)| \xrightarrow{P} 0. \tag{6.12}$$

Before we prove this theorem, we make some remarks on the conditions required to achieve the result. In particular, we are interested in discussing the concept of stochastic equicontinuity. The formal definition of stochastic equicontinuity is given below and in Andrews (1994).

Definition 6.1.1. $\{G_n(\theta) : n \geq 1\}$ *is stochastically equicontinuous on* Θ *if* $\forall \epsilon > 0$, $\exists \delta > 0$ *such that*

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta} \sup_{\theta'\in B(0,\delta)} |G_n(\theta) - G_n(\theta')| > \epsilon \right) < \epsilon. \tag{6.13}$$

The following two lemmas give equivalent definitions of stochastic equicontinuity. We omit the proof of this here however see Section 2 of Andrews (1994) for the details.

Lemma 6.1.1. $\{G_n(\theta) : n \geq 1\}$ *is stochastically equicontinuous on* Θ *if for any random sequences* $\{\theta_n \in \Theta\}_{n\geq 1}$ *and* $\{\theta_n^* \in \Theta\}_{n\geq 1}$ *such that* $\|\theta_n - \theta_n^*\| \xrightarrow{P} 0$, $\|G_n(\theta_n) - G_n(\theta_n^*)\| \xrightarrow{P} 0$.

Lemma 6.1.2. *The sequence of random functions* $\{G_n(\theta) : n \geq 1\}$ *is stochastically equicontinuous if and only if for every sequence of constants* $\{\delta_n : n \geq 1\} \subseteq \mathbb{R}_+$ *with* $\delta_n \to 0$, *we have*

$$\sup_{\theta,\theta^*\in\Theta, d(\theta,\theta^*)\leq\delta_n} \|G_n(\theta) - G_n(\theta^*)\| \xrightarrow{P} 0.$$

Using these definitions we can prove the stochastic generalisation of uniform convergence given in theorem 6.1.2 and Newey (1991).

Proof of Theorem 6.1.2. Since Θ is compact, for any $\delta > 0$, there exists a finite subset $\{\theta_k : k = 1, \ldots, K\}$ of Θ such that $B(\theta_k, \delta : k = 1, \ldots, K)$ cover Θ. Let $\epsilon > 0$, arbitrary and δ be the positive number such that (6.13) holds. Then,

$$\begin{aligned}
\mathbb{P}\left(\sup_{\theta\in\Theta} |G_n(\theta)| > 2\epsilon \right) &= \mathbb{P}\left(\max_k \sup_{\theta\in B(\theta_k,\delta)} |G_n(\theta) - G_n(\theta_k) + G_n(\theta_k)| > 2\epsilon \right) \\
&\leq \mathbb{P}\left(\max_k \sup_{\theta\in B(\theta_k,\delta)} |G_n(\theta) - G_n(\theta_k)| + \max_k |G_n(\theta_k)| > 2\epsilon \right) \\
&\leq \mathbb{P}\left(\max_k \sup_{\theta\in B(\theta_k,\delta)} |G_n(\theta) - G_n(\theta_k)| > \epsilon \right) + \mathbb{P}\left(\max_k |G_n(\theta_k)| > \epsilon \right) \\
&\leq \mathbb{P}\left(\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} |G_n(\theta) - G_n(\theta_k)| > \epsilon \right) + \mathbb{P}\left(\max_k |G_n(\theta_k)| > \epsilon \right).
\end{aligned} \tag{6.14}$$

Thus,

$$\limsup_{n\to 0} P(\sup_{\theta\in\Theta} |G_n(\theta)| > 2\epsilon) \leq \epsilon + 0 = \epsilon \tag{6.15}$$

which implies that

$$\sup_{\theta\in\Theta} |G_n(\theta)| \xrightarrow{P} 0. \tag{6.16}$$

□

In order to establish that the random functions we are considering are stochastic equicontinuous we need to find a set of conditions that guarantee this fact. Results which give us conditions with which to prove stochastic equicontinuity as in Andrews (1994),Newey (1991) will be presented next. However first we have a short discussion on the notion of tightness for random vectors (Van der Vaart 2000, Newey 1991).

A random vector X is *tight* if for every $\epsilon > 0$ there exists a constant M such that $\mathbb{P}(\|X\| > M) < \epsilon$. A set of random vectors $\{X_\alpha : \alpha \in A\}$ is called *uniformly tight* if M can be chosen the same for every X_α. That is, for every $\epsilon > 0$ there exists a constant M such that

$$\sup_{\alpha} \mathbb{P}(\|X_\alpha\| > M) < \epsilon. \tag{6.17}$$

This mean that there exists a compact set to which all X_α give probability almost one. Another name for uniformly tight is *bounded in probability*, (Van der Vaart 2000), and we shall use the notation $X_n = \mathcal{O}_p(1)$.

Every weakly converging sequence X_n is uniformly tight. According to Prohorov's theorem, (Prohorov 1956), the converse is also true: Every uniformly tight sequence contains a weakly converging subsequence.

The following theorem as given in Newey (1991) characterises the connection between tightness and stochastic equicontinuity.

Theorem 6.1.3. *Suppose there exists $N \in \mathbb{N}$ such that almost surely*

$$|G_n(\theta) - G_n(\theta^*)| \leq B_n h(d(\theta, \theta^*))$$

holds for all $\theta, \theta^* \in \Theta$ and $n \geq N$, where h is a deterministic function and $h(x) \to 0$ as

$x \to 0$ *and* $B_n = \mathcal{O}_p(1)$. *Then* $\{G_n(\theta) : n \geq 1\}$ *is stochastically equicontinuous.*

Proof. Let $\delta_n \to 0$. Then for n sufficiently large,

$$\sup_{\theta,\theta^* \in \Theta, d(\theta,\theta^*) < \delta_n} |G_n(\theta) - G_n(\theta^*)| \leq B_n h(\delta_n) = \mathcal{O}_p(1)o(1) = o_p(1). \tag{6.18}$$

Hence by definition of stochastic equicontinuity we conclude that $\{G_n(\theta) : n \geq 1\}$ is stochastically equicontinuous. □

Extending the tools designed above, we shall prove that the estimator $\hat{\boldsymbol{\kappa}}_n$ of the optimal gain $\mathbf{K}_n$ converges to the asymptotic Kalman gain for linear systems. The theory discussed and developed here reduces our task to proving the assumptions given in theorem 6.1.1. This theorem is very similar to theorem 5.1.1. The difference here is that the conditions are now stochastic in nature. The main difficulty we have is we do not have compactness for our problem but this is a requirement for the theory above to hold.

6.2 Minimising the Out-of-Sample Error

The set up of the problem is the same as in Chapter 5. However, we shall recall the details for clarity before we prove the main result. Suppose we have an initial state x_0, with mean $\hat{z}_0$ and covariance Σ_0. Suppose also that our model is given by

$$x_{n+1} = \mathbf{A}x_n + q_n \tag{6.19}$$

where x is the state and q_n are iid random variables with zero mean and covariance $\mathbf{Q}$ and we have observations given by

$$\eta_n = \mathbf{H}x_n + r_n \tag{6.20}$$

where r_n are iid with zero mean and covariance $\mathbf{R}$ and $\mathbf{R}$ is strictly positive definite.

Assume that r_n and q_n are uncorrelated and note that this model is time invariant since $\mathbf{A}$, $\mathbf{H}$, $\mathbf{R}$ and $\mathbf{Q}$ are taken to be constant.

For the system defined by (6.19) and (6.20), we construct an observer of the form

$$\begin{aligned} \hat{z}_n &= \mathbf{A} z_{n-1} \\ z_n &= \hat{z}_n + \mathbf{K}(\eta_n - \mathbf{H}\hat{z}_n) \end{aligned} \tag{6.21}$$

where $\mathbf{K}$ is the gain matrix that is kept constant. The coupling introduced by this gain matrix creates a linear feedback in the sense that the error between $\mathbf{H}\hat{z}_n$ and the observations is fed back into the model. The $\hat{z}_n$ is an estimate of x_n based on our *a priori* knowledge of the system, up to but not including time n.

We also assume that the system is completely observable and controllable. Recall the definitions for observability and controllability are given in Chapter 3

It was established in Chapter 5 that the gain matrix, $\boldsymbol{\kappa}_n$, that minimises the expected out-of-sample error,

$$\mathbb{E}[\mathbf{H}e_n + r'_n]^2 = \mathbf{H}\Gamma_n\mathbf{H}^T + \mathbf{R}' \tag{6.22}$$

converges to the asymptotic Kalman gain, $\boldsymbol{\kappa}_\infty$. We now want to show that the sequence $\hat{\boldsymbol{\kappa}}_n$ that minimises the empirical mean of the out-of-sample error, and estimates $\mathbf{K}_n$ also converges to the asymptotic gain. Here $e_n = z_n - x_n$ and r'_n is iid noise with covariance matrix $\mathbf{R}'$, which is independent of the observation noise r_n but comes from the same underlying flow pattern. The empirical mean is the quantity which we can calculate in practical situations and as such are interested in its asymptotic properties. Since we want to minimise the mean of the out-of-sample error, we think of the estimator $\hat{\boldsymbol{\kappa}}_n$, as an M-estimator and thus the problem becomes one of proving consistency of the M-estimator.

Naturally, the asymptotic value of $\hat{\boldsymbol{\kappa}}_n$ depends on the asymptotic value of ϕ_n. The deterministic function $\psi(\mathbf{K})$ in this case is defined by the asymptotic error covariance

$\mathbf{H}\Gamma_\infty\mathbf{H}^T + \mathbf{R}'$, which is defined by

$$\begin{aligned}\psi(\mathbf{K}) &= \mathbf{H}(\mathbf{A} - \mathbf{KHA})\Gamma_\infty(\mathbf{A} - \mathbf{KHA})^T\mathbf{H}^T \\ &\quad + \mathbf{H}(\mathbb{1} - \mathbf{KH})\mathbf{Q}(\mathbb{1} - \mathbf{KH})^T\mathbf{H}^T + \mathbf{HKRK}^T\mathbf{H}^T + \mathbf{R}'.\end{aligned} \tag{6.23}$$

This is a fixed point equation that characterises the asymptotic behaviour of the error covariance.

6.2.1 Consistency of the Estimator

The requirements for consistency as characterised by theorem (6.1.1) are that $\boldsymbol{\kappa}_\infty$ is a well-separated point of minimum and that the sample average that describes the empirical mean of the out-of-sample error converges uniformly to the asymptotic error in (6.23). Establishing these facts reduces to proving the assumptions in the following theorem. First, let $e_n = z_n - x_n$, then the out-of-sample error is defined by

$$\phi_n(\mathbf{K}) = \frac{1}{n}\sum_{i=0}^{n-1}(\mathbf{H}e_i - r_i')^2 - 2\mathrm{tr}\,(\mathbf{HKR}) \tag{6.24}$$

Theorem 6.2.1. *Consider the continuous functions ϕ_n and ψ as defined in (6.24) and (6.23) respectively, with $\mathcal{K} \subset \mathbb{R}^D$ compact. Let $\hat{\boldsymbol{\kappa}}_n$ be the minimiser of ϕ_n. Assume*

1. $\mathbb{P}(\phi_n \quad \text{has} \quad \text{no} \quad \text{minimiser}) \to 0$
2. $\mathbb{P}(\hat{\boldsymbol{\kappa}}_n \notin \mathcal{K}) \to 0$
3. $\sup_{\mathbf{K}\in\mathcal{K}} |\phi_n(\mathbf{K}) - \psi(\mathbf{K})| \xrightarrow{P} 0$
4. *ψ has a unique minimiser $\boldsymbol{\kappa}_\infty$.*

Then $\hat{\boldsymbol{\kappa}}_n \xrightarrow{P} \boldsymbol{\kappa}_\infty$.

Notice that this theorem is very similar to theorem 5.1.1. The difference here is that the conditions are now stochastic in nature.

Before we prove this theorem, we establish a small detail that will be required in the proof. This detail concerns the asymptotic Kalman gain, $\boldsymbol{\kappa}_\infty$. Point (4) in the theorem explains that $\boldsymbol{\kappa}_\infty$ must be the unique minimiser of ψ. In fact more is true. Coupled with the fact that $\mathcal{K}$ is compact and ψ is continuous, it follows that $\boldsymbol{\kappa}_\infty$ is a well-separated (or isolated) point of minimum. This means that only parameters close to $\boldsymbol{\kappa}_\infty$ may yield a value of $\psi(\mathbf{K})$ close to the minimum value $\psi(\boldsymbol{\kappa}_\infty)$. The following results establish this fact. See problem 5.27 in Van der Vaart (2000).

Lemma 6.2.1. *For a compact set Θ and continuous function ϕ, uniqueness of θ_0 as a maximiser implies that θ_0 is a well-separated point of maximum.*

Proof. Let $G \subset \Theta$ be open. Then G^c is a closed subset of Θ and is compact. Since ϕ is continuous it achieves its maximum on a compact set. Hence there exists some $\theta^* \in G^c$ such that

$$\phi(\theta^*) = \sup_{\theta \in G^c} \phi(\theta). \tag{6.25}$$

Since θ_0 is the unique maximiser of ϕ we have that

$$\phi(\theta_0) > \phi(\theta^*) = \sup \phi(\theta) \tag{6.26}$$

which is the required result. □

Lemma 6.2.2. *Under the same conditions of theorem 5.1.1, $\boldsymbol{\kappa}_\infty$ is a well-separated point of minimum of $\psi(\mathbf{K})$.*

Proof. Our parameter space is given by $\mathcal{K}$ which we assume is compact as in theorem 5.1.1. This will be proved later. Proposition (5.1.1) gives us uniqueness of the minimiser $\boldsymbol{\kappa}_\infty$ and so from lemma (6.2.1) it follows that $\boldsymbol{\kappa}_\infty$ is a well separated point of minimum of $\psi(\mathbf{K})$. □

Proof of Theorem 6.2.1. Let $\mathcal{G}_n := \{\phi_n \quad \text{has} \quad \text{a} \quad \text{minimiser} \quad \} \cap \{\boldsymbol{\kappa}_n \in \mathcal{K}\}$. If $\mathcal{G}_n$ happens then the proof of this theorem goes through in exactly the same way as the proof of theorem

6.1.1.

If not however, it follows that

$$\mathbb{P}(\mathcal{G}^C) \leq \mathbb{P}(\psi_n \quad \text{has} \quad \text{no} \quad \text{minimiser}) + \mathbb{P}(\mathbf{K}_n \notin \mathcal{K}). \tag{6.27}$$

By assumption both of the terms on the right hand side of the above equation converge to zero in probability, so the result stays the same. □

Notice that compactness is a necessary condition in the result. However, in our specific case, the isolation of the point is true regardless of compactness. This can be seen by considering the space $\mathcal{K}$ established in Chapter 5 as this will be the space used here also. By design, it ensures that for any gain that is not a minimiser, the error $\psi(\mathbf{K})$ grows very large. Therefore, there cannot be another gain that gives a value close to $\psi(\mathbf{K})$. Hence, $\boldsymbol{\kappa}_\infty$ is a well-separated point of minimum. The compactness of the space however, is needed for other elements of the proof and as such still needs to be included.

As we did in Chapter 5, we need to check that each of the four assumptions made in theorem 6.2.1 hold. If this is true then we will have proven that the estimator $\hat{\boldsymbol{\kappa}_n}$ that minimises ϕ_n and estimates $\boldsymbol{\kappa}_n$ converges to the asymptotic gain in probability.

Assumption (4) in the statement of thereom 6.2.1 is identical to its counterpart in theorem 5.1.1. This is because the limit and its minimser are the same in both cases. Thus point (4) in the above has already been established in proposition 5.1.1.

The parameter space $\mathcal{K}$ is defined again by $\mathcal{K} = \{\mathbf{K}; \sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon\} \cap \mathcal{K}_0$, where $\mathcal{K}_0$ is a compact region. We will determine ϵ and $\mathcal{K}_0$ later.

Since $\mathbf{K}$ is a stabilising feedback gain in $\mathcal{K}$, ψ is well defined and given $\mathcal{K}$ as above we will now confirm point (3) in theorem 6.2.1. It is simply the statement that ϕ_n converges to ϕ uniformly in probability. We shall prove stochastic uniform convergence using the results presented in Section 6.1.2.

Let $e_n = z_n - x_n$, then as we have already seen, the out-of-sample error is defined by

$$\phi_n(\mathbf{K}) = \frac{1}{n}\sum_{i=0}^{n-1}(\mathbf{H}e_i - r_i')^2 - 2\mathrm{tr}\,(\mathbf{HKR}) \tag{6.28}$$

which depends on, in particular, the term $\frac{1}{n}\sum e_i e_i^T$. We will now prove that $|e_i - \epsilon_i| \to 0$ where ϵ_i is a stationary process with $\mathrm{Cov}(\epsilon_i) = \Gamma_\infty$.

Consider the error e_n which is defined by

$$e_n = (\mathbf{A} - \mathbf{KHA})e_{n-1} + \mathbf{K}r_n - (\mathbb{1} - \mathbf{KH})q_n \tag{6.29}$$

and by induction it follows that

$$\begin{aligned} e_{n+m} &= (\mathbf{A} - \mathbf{KHA})^m e_0 \\ &\quad + \sum_{i=0}^{m-1}(\mathbf{A} - \mathbf{KHA})^i\{\mathbf{K}r_{n+m-i} - (\mathbb{1} - \mathbf{KH})q_{n+m-i}\}. \end{aligned} \tag{6.30}$$

Now define

$$e_0^{(m)} = \sum_{i=0}^{m-1}(\mathbf{A} - \mathbf{KHA})^i\{\mathbf{K}r_{-i} - (\mathbb{1} - \mathbf{KH})q_{-i}\} \tag{6.31}$$

where we assume r_i, q_i are extended to the past. This can be done since the noise terms are all iid.

Lemma 6.2.3. *The error, $e_0^{(m)}$, is a Cauchy Sequence if $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$.*

Proof. Let $\mathbf{W} = \mathbf{A} - \mathbf{KHA}$ and consider

$$e_0^{(m+l)} - e_0^{(m)} = \sum_{i=m}^{m+l-1} \mathbf{W}^i \left\{\mathbf{K}r_{-i} - (\mathbb{1} - \mathbf{KH})q_{-i}\right\}. \tag{6.32}$$

Then taking the expected value of the above yields,

$$\begin{aligned}\mathbb{E}\left[(e_0^{(m+l)} - e_0^{(m)})^2\right] &= \sum_{i=m}^{m+l-1} \mathbf{W}^i \left\{\mathbf{K}r_{-i} - (\mathbb{1} - \mathbf{KH})q_{-i}\right\} \mathbf{W}^{i\,T} \\ &\leq \sum_{i=m}^{\infty} \mathbf{W}^i \left\{\mathbf{K}r_{-i} - (\mathbb{1} - \mathbf{KH})q_{-i}\right\} \mathbf{W}^{i\,T} \\ &= \mathbf{W}^m \left(\sum_{i=0}^{\infty} \mathbf{W}^i \left\{\mathbf{K}r_{-i} - (\mathbb{1} - \mathbf{KH})q_{-i}\right\} \mathbf{W}^{i\,T}\right) \mathbf{W}^{m\,T}\end{aligned} \tag{6.33}$$

which converges in L_2 as required since $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$. □

Consider now the L_2 limit given by

$$\epsilon_0 = \lim_{m\to\infty} e_0^{(m)} = \sum_{i=0}^{\infty} (\mathbf{A} - \mathbf{KHA})^i \{\mathbf{K}r_{-i} + -(\mathbb{1} - \mathbf{KH})q_{-i}\}. \tag{6.34}$$

If we now use the random variable ϵ_0 as the initial condition, the process given by

$$\epsilon_n = (\mathbf{A} - \mathbf{KHA})^n \epsilon_0 + \sum_{i=0}^{n-1} (\mathbf{A} - \mathbf{KHA})^i \{\mathbf{K}r_{n-i} - (\mathbb{1} - \mathbf{KH})q_{n-i}\} \tag{6.35}$$

is stationary. Comparing this to the error obtained when we use the initial condition $e_0 = z_0 - x_0$ we see that

$$e_n - \epsilon_n = (\mathbf{A} - \mathbf{KHA})^n (e_0 - \epsilon_0) \overset{n\to\infty}{\to} 0 \tag{6.36}$$

since $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$. From now on we assume e_n is stationary. Equipped with the above information we can now prove stochastic uniform convergence of the sample error as explained in theorem 6.1.2. Define $G_n(\mathbf{K}) := \phi_n(\mathbf{K}) - \psi(\mathbf{K})$ then pointwise convergence is given in the following proposition.

Proposition 6.2.1. *$G_n(\mathbf{K}) \overset{P}{\to} 0$ for all $\mathbf{K} \in \mathcal{K}$.*

Proof. The out-of-sample error, $\phi_n(\mathbf{K})$, as defined in (6.28) converges to $\psi(\mathbf{K})$ by the ergodic theorem (see for example Collet & Eckmann (2007)). Thus the pointwise convergence in

probability is established.

□

Stochastic equicontinuity can also be established using theroem 6.1.3.

Lemma 6.2.4. *If* $\mathbf{K}$ *is in the compact set* $\mathcal{K}$ *and* $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$ *then* $G_n(\mathbf{K})$ *is stochastically equicontinuous.*

Proof. We establish stochastic equicontinuity by showing that

$$G_n(\mathbf{K}) - G_n(\mathbf{K}') \leq B_n h(|\mathbf{K} - \mathbf{K}'|) \tag{6.37}$$

for $\mathbf{K}, \mathbf{K}' \in \mathcal{K}$, $B_n = \mathcal{O}_p(1)$ and h a deterministic, continuous function. Since G_n is defined by $G_n(\mathbf{K}) := \phi_n(\mathbf{K}) - (\mathbf{H}\Gamma_\infty(\mathbf{K})\mathbf{H}^T + \mathbf{R}')$ we have that

$$|G_n(\mathbf{K}) - G_n(\mathbf{K}')| \leq \frac{1}{n}\sum_i |\upsilon_i^2(\mathbf{K}) - \upsilon_i^2(\mathbf{K}')| + |\mathbf{H}(\Gamma_\infty(\mathbf{K}) - \Gamma_\infty(\mathbf{K}'))\mathbf{H}^T| \tag{6.38}$$

where $\upsilon_i = \mathbf{H}e_i - r_i$. The first term on the right hand side of the above equation can be expressed in the following way

$$\upsilon_i^2(\mathbf{K}) - \upsilon_i^2(\mathbf{K}') = (\upsilon_i(\mathbf{K}) - \upsilon_i(\mathbf{K}'))(\upsilon_i(\mathbf{K}) + \upsilon_i(\mathbf{K}')) \tag{6.39}$$

and

$$\upsilon_i(\mathbf{K}) - \upsilon_i(\mathbf{K}') = \mathbf{H}(e_i(\mathbf{K}) - e_i(\mathbf{K}')); \quad \upsilon_i(\mathbf{K}) + \upsilon_i(\mathbf{K}') = \mathbf{H}(e_i(\mathbf{K}) + e_i(\mathbf{K}')) + 2r_i. \tag{6.40}$$

From the structure of these errors since we have assumed complete observability and r_i' are iid and therefore tight, we simply need to establish stochastic equicontinuity for $(e_i(\mathbf{K}) - e_i(\mathbf{K}'))$, $(e_i(\mathbf{K}) + e_i(\mathbf{K}'))$ and $(\Gamma_\infty(\mathbf{K}) - \Gamma_\infty(\mathbf{K}'))$. Proving stochastic equicontinuity for the term $\mathbf{HKR}$ as it appears in (6.28) is trivial.

By using the explicit expression for the error, e_i, we obtain

$$\begin{aligned} e_i(\mathbf{K}) - e_i(\mathbf{K}') &= (\mathbf{A} - \mathbf{KHA})(e_{i-1}(\mathbf{K}) - e_{i-1}(\mathbf{K}')) + (\mathbf{K} - \mathbf{K}')(r_i + \mathbf{H}q_i) \\ &\quad + ((\mathbf{A} - \mathbf{KHA}) - (\mathbf{A} - \mathbf{K}'\mathbf{HA}))\, e_{i-1}(\mathbf{K}'). \end{aligned} \tag{6.41}$$

The second and third terms on the right hand side of (6.41) can be expressed in the following way:

$$\begin{aligned} &((\mathbf{A} - \mathbf{KHA}) - (\mathbf{A} - \mathbf{K}'\mathbf{HA}))\, e_i(\mathbf{K}') + (\mathbf{K} - \mathbf{K}')(r_i + \mathbf{H}q_i) \\ &= (\mathbf{K} - \mathbf{K}')(r_i + \mathbf{H}q_i - \mathbf{HA}e_{i-1}(\mathbf{K}')) \\ &= b_i(\mathbf{K} - \mathbf{K}') \end{aligned} \tag{6.42}$$

and since $(r_i + \mathbf{H}q_i)$ are iid, tightness follows while since $e_i(\mathbf{K}')$ converges in distribution it too is tight; so that $b_i = \mathcal{O}_p(1)$. The convergence in distribution of e_i follows from the earlier discussion when we considered the stationary random variable ϵ_0. This is because asymptotically, e_i and ϵ_0 have the same distribution.

For the first term on the right hand side of (6.41) we use induction. Suppose we start with the same initial condition, i.e $e_0(\mathbf{K}) = e_0(\mathbf{K}')$, then for $i = 1$ we have

$$e_1(\mathbf{K}) - e_1(\mathbf{K}') = (\mathbf{K} - \mathbf{K}')(r_1 + \mathbf{H}q_1 - \mathbf{HA}e_0) = B_1 h(|\mathbf{K} - \mathbf{K}'|) \tag{6.43}$$

where $B_1 = \mathcal{O}_p(1)$ and $h(x) = x$. Assume this is true for $i = l$ and consider

$$\begin{aligned} e_{l+1}(\mathbf{K}) - e_{l+1}(\mathbf{K}') &= (\mathbf{A} - \mathbf{KHA})(e_l(\mathbf{K}) - e_l(\mathbf{K}')) \\ &\quad + (\mathbf{K} - \mathbf{K}')(r_{l+1} + \mathbf{H}q_{l+1} - \mathbf{HA}e_l(\mathbf{K}')) \\ &\leq (\mathbf{A} - \mathbf{KHA})B_l h(|\mathbf{K} - \mathbf{K}'|) + b_{l+1}(\mathbf{K} - \mathbf{K}') \end{aligned} \tag{6.44}$$

and since $\sigma(\mathbf{A} - \mathbf{KHA}) \leq 1 - \epsilon$ we have

$$|e_i(\mathbf{K}) - e_i(\mathbf{K}')| \leq \tilde{B}_{i-1} h(|\mathbf{K} - \mathbf{K}'|) + b_i |\mathbf{K} - \mathbf{K}'| \tag{6.45}$$

with $\tilde{B}_i = \mathcal{O}_p(1)$.

Performing the same calculation and argument on $(e_i(\mathbf{K}) + e_i(\mathbf{K}'))$ yields

$$|e_i(\mathbf{K}) + e_i(\mathbf{K}')| \leq C_{i-1} g(|\mathbf{K} - \mathbf{K}'|) + c_i |\mathbf{K} + \mathbf{K}'| \tag{6.46}$$

with $C_i, c_i = \mathcal{O}_p(1)$ and so it follows that

$$|e_i(\mathbf{K}) - e_i(\mathbf{K}')||e_i(\mathbf{K}) + e_i(\mathbf{K}')| \leq L_i f(|\mathbf{K} - \mathbf{K}'|) \tag{6.47}$$

where $L_i = \mathcal{O}_p(1)$ is a combination of B_i, C_i, c_i, b_i and f is a continuous function.

Therefore we have that

$$\frac{1}{n} \sum_i |e_i^2(\mathbf{K}) - e_i^2(\mathbf{K}')| \leq \frac{1}{n} \sum_i L_i f(|\mathbf{K} - \mathbf{K}'|). \tag{6.48}$$

Now consider the final term to prove is stochastic equicontinous, $(\Gamma_\infty(\mathbf{K}) - \Gamma_\infty(\mathbf{K}'))$. Note that

$$|\Gamma_\infty(\mathbf{K}') - \Gamma_\infty(\mathbf{K})| \leq \mathbb{E}|e_0(\mathbf{K}') - e_0(\mathbf{K})||e_0(\mathbf{K}') + e_0(\mathbf{K})|^T \tag{6.49}$$

where $e_0 = (\mathbf{A} - \mathbf{KHA})e_0 + \mathbf{K}r_0 - (\mathbb{1} - \mathbf{KH})q_0$ and $\Gamma_\infty = \mathbb{E}e_0 e_0^T$. it is then straightforward to see that

$$\begin{aligned} |\Gamma_\infty(\mathbf{K}') - \Gamma_\infty(\mathbf{K})| \quad &\leq \mathbb{E}|(\mathbf{K}' - \mathbf{K})(\mathbf{HA}e_0 - r_0 - \mathbf{H}q_0)||(\mathbf{K}' + \mathbf{K})(\mathbf{HA}e_0 - r_0 - \mathbf{H}q_0) \\ &\quad +2(e_0 + q_0)| \\ &\leq Q_0 v(|\mathbf{K}' - \mathbf{K}|) \end{aligned} \tag{6.50}$$

where $Q_0 = \mathcal{O}_p(1)$ and v is a continuous function.

Since all the expression are stochastic equicontinuous, it follows that

$$|G_n(\mathbf{K}) - G_n(\mathbf{K}')| \leq L_n f(|\mathbf{K} - \mathbf{K}'|) + Q_0 v(|\mathbf{K}' - \mathbf{K}|) \tag{6.51}$$

which proves the required result. □

Therefore, using these results we can establish uniform convergence in probability.

Lemma 6.2.5. *$G_n(\mathbf{K})$ converges uniformly in probability, i.e* $\sup_{\mathbf{K}\in\mathcal{K}} |G_n(\mathbf{K})| \xrightarrow{P} 0$.

Proof. By theorem 6.1.2 and lemma 6.2.4 with the addition outlined above, it follows that

$$\sup_{\mathbf{K}\in\mathcal{K}} |G_n(\mathbf{K})| \xrightarrow{P} 0. \tag{6.52}$$

□

This then just leaves points (1) and (2) in theorem 6.2.1 to be checked and ϵ and $\mathcal{K}_0$ to be determined. We shall prove these together.

The results proven in lemmas 5.1.5 and 5.1.6, still hold and apply here. The former result states that for any minimiser $\hat{\boldsymbol{\kappa}}_n$, given that it exists, satisfies $\phi_n(\hat{\boldsymbol{\kappa}}_n) \leq \mathbf{R}(\mathbb{1} - \delta)$. This is still true now except that it is true in probability. The proof is the same; the conclusion is as follows,

$$\mathbb{P}(\phi_n(\hat{\boldsymbol{\kappa}}_n) > \mathbf{R}(\mathbb{1} - \delta)) \to 0. \tag{6.53}$$

The second lemma mentioned above proves that the set $\mathcal{K}_0$ defined by

$$\mathcal{K}_0 := \{\mathbf{K}; \psi(\mathbf{K}) \leq \mathbf{R}(\mathbb{1} - \delta)\}, \tag{6.54}$$

is compact. This remains identical in this case as the limit ψ is the same in both the deterministic and stochastic cases.

Therefore we have established that ϕ_n has a minimiser on $\mathcal{K}_0$ because $\mathcal{K}_0$ is compact and ϕ_n is continuous. We need to ensure that this minimiser exists on $\mathcal{K}$, that is, in addition the minimising gain must satisfy $\sigma(\mathbf{A}-\mathbf{KHA}) \leq 1-\epsilon$, i.e. $\mathbf{K}$ must stabilise the error dynamics.

In the deterministic case presented in Chapter 5, using the result established in lemma 5.1.8, we determined that $\psi_n(\mathbf{K}) \geq \mathbf{R}$, which excluded this gain and any others like it from being minimisers. In other words, we proved that any minimising gain of the expected out-of-sample error must stabilise the error dynamics. We did this by finding a uniform bound outside of the space $\mathcal{K}$.

Proving this in the stochastic case has been difficult and as such has not been completed. This result is formulated in the following conjecture.

Conjecture 6.2.1. *Let $\nu_k = \mathbf{H}e_k$, where $e_k = x_k - z_k$. There exists an ϵ such that for all δ*

$$\mathbb{P}\left(\inf_{\mathbf{K}:\sigma(\mathbf{A}-\mathbf{KHA})>1-\epsilon} \frac{1}{n}\sum_{k=1}^{n} \nu_k \nu_k^T \leq \mathbf{R} - \delta\right) \to 0. \tag{6.55}$$

For an intuitive discussion about this conjecture, consider for a moment the state error, $e_n = x_n - z_n$, given by

$$\begin{aligned} e_{n+1} &= (\mathbf{A}-\mathbf{KHA})e_n - (\mathbb{1}-\mathbf{KH})q_{n+1} + \mathbf{K}r_{n+1} \\ &= (\mathbf{A}-\mathbf{KHA})e_n + s_n. \end{aligned} \tag{6.56}$$

By the results in lemma 5.1.8 and corollary 5.1.1 we know that $\mathbb{E}s_n^2 \geq \alpha > 0$. We can also write

$$v^T e_n = \sum_{l=0}^{n-1} \lambda^l s_{n-l} \tag{6.57}$$

for $v^T(\mathbf{A}-\mathbf{KHA}) = \lambda v^T$. Now by results in the theory of random polynomials (see for example Erdos & Turán (1950), Hughes & Nikeghbali (2008)), the zeros of $v^T e_n$ cluster on the unit circle. Since s_n are iid, any meaningful cancellation between in $v^T e_n$ can only

happen if $\lambda = \mathcal{O}(1)$. But the out-of-sample error is an average over such polynomials squared. Intuitively, (6.55) can only happen if the zeros are very different to $v^T e_n$.

Once the Conjecture is established the following final result then follows immediately.

Theorem 6.2.2. *The gain matrix* $\mathbf{K}$ *that minimises* $\phi_n(\mathbf{K})$ *over the compact set* $\mathcal{K}$ *and stabilises the system such that* $\sigma(\mathbf{A} - \mathbf{KHA}) < 1$*, converges in probability to the asymptotic gain* $\boldsymbol{\kappa}_\infty$*.*

Chapter Summary When calculating errors in practical situations, it is only possible to determine an estimate of the errors. In the numerical experiments presented in Chapter 3 we calculated the out-of-sample error by means of the empirical mean. As such it was necessary to determine whether the minimiser of this empirical mean converges to the asymptotic Kalman gain as suggested in Chapter 3 and Mallia-Parfitt & Bröcker (2016). In this chapter we have presented the proof that the constant gain matrix that minimises the empirical mean of the out-of-sample error and estimates the optimal gain, converges to the asymptotic Kalman gain in the limit of large observational windows.

This was accomplished by treating the estimator $\hat{\boldsymbol{\kappa}}_n$ as an M-estimator and used the concept of stochastic uniform convergence to prove that it is a consistent estimator. The conditions required to prove this non-trivial fact are that $\boldsymbol{\kappa}_\infty$ is a well-separated point of minimum, the error uniformly converges in probability to the asymptotic error and that the parameter space is compact. Unfortunately, we were unable to complete the result as Conjecture 6.2.1 has not been proven.

Chapter 7

The Out-of-Sample Error for Non-Linear Systems

By considering data assimilation schemes which employ linear error feedback, it has been established in Chapters 5 and 6 that the feedback gain matrix minimising the out-of-sample error, or even the empirical mean of the out-of-sample error (which is what can be calculated in practice), converges to the asymptotic Kalman gain in the limit of large observational windows. We now wish to consider non-linear systems.

We define the out-of-sample error, optimism and tracking error for non-linear systems and determine, numerically, that the theory developed in Chapter 3 for linear systems, applies in the non-linear setting. The theory is applicable to non-linear systems with linear observations since calculation of the out-of-sample error only depends on the structure of the observations not on the underlying dynamical system. Knowledge of the dynamical system enters the calculation through the assimilation algorithm.

In the case of non-linear dynamical systems we cannot as easily calculate an explicit expression for the asymptotic gain neither can we be certain that the optimal gain will converge in a meaningful way. This is because the asymptotic behaviour of the optimal gain depends heavily on the presence of dynamical noise and cannot be expected to converge

in a significant way without the presence of model noise. However, we present numerical experiments for two non-linear systems with linear observations as done in Mallia-Parfitt & Bröcker (2016).

7.1 Non-Linear System

Consider non-linear dynamical systems of the form

$$\begin{aligned} \tilde{x}_{n+1} &= f(\tilde{x}_n) \\ \eta_n &= h(\tilde{x}_n) \end{aligned} \tag{7.1}$$

where f and h are non-linear functions. As for the linear case, the construction of an observer requires some properties of observability. When a linear system is observable it is observable regardless of the noise input. For non-linear systems, this is no longer true as, in general, they have singular inputs that make them unobservable.

For such a system (7.1) define the observability map $\mathcal{O}$ by

$$\mathcal{O}(x) = \begin{bmatrix} h(x) \\ h \circ f(x) \\ \vdots \\ h \circ f^{n-1}(x) \end{bmatrix}, \tag{7.2}$$

where $h \circ f(x) = h(f(x))$, $f^1 = f$, $f^j = f \circ f^{j-1}$. The system in question is called observable around a point x_0 if the Jacobian $(\partial\mathcal{O}/\partial x)(x_0)$ is invertible. Observability is always required; however there are several approaches designed to construct an appropriate observer.

The design of state estimation for non-linear systems has been studied thoroughly, with different approaches being taken to achieve the required results. Krener & Isidori

(1983) and Krener & Respondek (1985) presented a contribution to this observer theory for systems in which the dynamics of the observation error is linear. However the conditions to achieve this are rather restrictive. Another algorithm was proposed by Zeitz (1987), in which time derivatives of the input were used; unfortunately convergence of the observer cannot be guaranteed in this case.

An approach based on high gain cancellation of the non-linearity was proposed by Tornambe (1989). However, this approach does not guarantee the asymptotic convergence of the estimated state to the true state. Ciccarella et al. (1993) construct High Gain Observers that can be extended to multiple input-multiple output non-linear systems. They show that the required asymptotic or sometimes exponential convergence can be achieved for a large enough gain.

A complete contribution to this theory is explained in Gauthier et al. (1992). They establish that a with non-linear change of coordinates, the state of a non-linear system can be globally asymptotically tracked by means of an observer whose gain is determined via a solution of a Lyapunov-like equation. This approach requires the existence of a global diffeomorphism and is the approach considered here.

Formally, we write this as the following problem. Consider the system (7.1), with no noise input and with scalar output. Assume that $f(0) = 0$, $h(0) = 0$. The problem is to find conditions ensuring existence of an invertible coordinate change $x = T(\tilde{x})$ such that the original non-linear system is equivalent to

$$\begin{aligned} x_{n+1} &= \mathbf{A}x_n + \xi(\eta_n) \\ \eta_n &= \mathbf{H}x_n \end{aligned} \tag{7.3}$$

where the pair $(\mathbf{A}, \mathbf{H})$ is observable in the traditional definition 4.2.1. The following result gives a solution to the problem, (Lin & Byrnes 1995, Huijberts et al. 1999). It is the discrete analogue of theorem 5.1.3 presented in Nijmeijer & van der Schaft (1990).

Theorem 7.1.1. *A discrete-time system (7.1) with single output is locally equivalent to a system (7.3) with observable pair* $(\mathbf{A}, \mathbf{H})$ *via a coordinate change* $x = T(\tilde{x})$ *if and only if*

i the pair $(\partial h(0)/\partial\tilde{x}, \partial f(0)/\partial\tilde{x})$ *is observable*

ii the Hessian matrix of the function $h \circ f^n \circ \mathcal{O}^{-1}(s)$ *is diagonal.*

Condition (i) means that the Jacobian $(\partial\mathcal{O}/\partial\tilde{x})(0)$ is invertible. Condition (ii) can be interpreted in the following way. If condition (i) holds, the transformation $s = \mathcal{O}(\tilde{x})$ is a local diffeomorphism and so s forms a new set of local coordinates for the dynamics (7.1) around the origin. In these new coordinates, the system (7.1) takes the form

$$s_{n+1} = \begin{bmatrix} s_n^{(2)} \\ s_n^{(3)} \\ \vdots \\ s_n^{(k)} \\ f_s(s_n) \end{bmatrix}, \quad \eta_n = s_n^{(1)} \tag{7.4}$$

where $f_s(s) = h \circ f^n \circ \mathcal{O}^{-1}(s)$ and $s^{(i)} := h \circ f^{i-1}$. Equation (7.4) is called the observable form of the system (7.1), (Huijberts et al. 1998). Condition (ii) is then equivalent to the local existence of functions $\phi_1, \ldots, \phi_n : \mathbb{R} \to \mathbb{R}$ such that

$$f_s(s) = \xi_1(s^{(1)}) + \xi_2(s^{(2)}) + \ldots + \xi_n(s^{(n)}). \tag{7.5}$$

With these functions known, the transformation

$$x_i = s_{n+1-i} - \sum_{k=i+1}^{n} \xi_k(s^{(k-i)}) \tag{7.6}$$

for $i = 1, \ldots, n$ then transforms the observable form (7.4) into the required form,

$$\begin{cases} x_{n+1}^{(1)} &= \xi_1(y_n) \\ x_{n+1}^{(2)} &= x_n^{(1)} + \xi_2(\eta_n) \\ &\vdots \\ x_{n+1}^{(k)} &= x_n^{(k-1)} + \xi_k(\eta_n) \\ \eta_n &= x_n^{(k)}. \end{cases} \tag{7.7}$$

Therefore, we shall be considering systems of the form

$$\begin{aligned} x_{n+1} &= \mathbf{A}x_n + \xi(\mathbf{H}x_n) \\ \eta_n &= \mathbf{H}x_n. \end{aligned} \tag{7.8}$$

Systems of this form are known as systems in Lur'e form. The observer is set up in a similar way to Chapter 3 so that our sequential scheme is given by

$$\begin{aligned} \hat{z}_{n+1} &= \mathbf{A}z_n + \xi(\eta_{n+1}) \\ z_{n+1} &= \hat{z}_{n+1} - \mathbf{K}_n(\mathbf{H}\hat{z}_{n+1} - \eta_{n+1}) \\ y_n &= \mathbf{H}z_n \end{aligned} \tag{7.9}$$

where $\mathbf{K}_n$ is the feedback gain matrix which may depend on the observations $\eta_1, \ldots \eta_{n-1}$ but not on η_n and y_n is the model output. Once again we shall be considering data assimilation through synchronisation.

Due to the linearity in the observation operator, the calculations for the out-of-sample error, tracking error and optimism are the same as in the linear case. The statistic we use to calculate the out-of-sample error is also identical to the linear case and recall that is given by

$$\mathbb{E}[y_n - \eta_n]^2 = \mathbb{E}[y_n - \eta_n']^2 - 2\sigma^2 \mathrm{tr}(\overline{\mathbf{K}}_n^T \mathbf{H}^T). \tag{7.10}$$

The linearity in the observation operator allows for simple calculation of the optimism and hence the out-of-sample error. Evidently, even in this non-linear case, calculating the out-of-sample error is straightforward and we do not need any information about the model error. The only information required is the feedback gain matrix, observation operator, and observational error covariance matrix

7.2 Numerical Experiment I: Hénon Map

We carried out numerical experiments to test the methodology described above and in Chapter 3 as done in Mallia-Parfitt & Bröcker (2016). The following experimental setup was used: The reality is given by

$$x_{n+1} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{\mathbf{A}} x_n + c \begin{bmatrix} (\mathbf{H}x_n)^2 \\ 0 \end{bmatrix} + d \tag{7.11}$$

which for the values $a = 0$, $b = 0.3$, c$=-1.4$, $d = [1 \quad 0]^T$ is the chaotic Henon Map with corresponding observations

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{7.12}$$

where $\mathbf{H} = [1 \quad 0]$, and $\zeta_n = \mathbf{H}x_n$. The model describing the reality is completely deterministic and we assume that the observations are corrupted by random noise. For these experiments we have $x_n \in \mathbb{R}^2$ and $\eta_n \in \mathbb{R}$.

Here we consider data assimilation by means of synchronisation so we set up an observer roughly analogous to our sequential scheme (7.9),

$$z_{n+1} = \hat{z}_{n+1} + \mathbf{K}_n(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}), \qquad y_n = \mathbf{H}z_n \tag{7.13}$$

where

$$\hat{z}_{n+1} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{\mathrm{A}} z_n + c \begin{bmatrix} \eta_n^2 \\ 0 \end{bmatrix} + d \tag{7.14}$$

where a, b, c, d are the same as for the reality. In this case the model is coupled to the observations through a linear coupling term which is dependent on the difference between the actual output and the output value expected based on the next estimate of the state. However there is also a non linear coupling introduced here by the presence of η_n^2 in the background term. Note that (7.10) is still valid nonetheless because $\hat{z}_{n+1}$ is still uncorrelated with r_{n+1}. For these experiments we will take the coupling matrix $\mathbf{K}_n$ to be constant so from here on in we write $\mathbf{K}_n = \mathbf{K}$.

We need to choose the matrix $\mathbf{K}$ appropriately so that we can vary the coupling strength. If the coupling is too strong the observations will be tracked too closely and if the coupling is too weak the observations are tracked badly or not at all. We first consider the noise-free situation so that $\eta_n = \mathbf{H}x_n$. The error dynamics in this case are given by

$$\begin{aligned} e_{n+1} &= x_{n+1} - z_{n+1} \\ &= x_{n+1} - \hat{z}_{n+1} - \mathbf{KH}(x_{n+1} - \hat{z}_{n+1}) \\ &= (\mathbb{1} - \mathbf{KH})(x_{n+1} - \hat{z}_{n+1}) \\ &= (\mathbf{A} - \mathbf{KHA})(x_n - z_n) \\ &= (\mathbf{A} - \mathbf{KHA})e_n. \end{aligned} \tag{7.15}$$

The matrix $(\mathbf{A} - \mathbf{KHA})$ is stable even if $\mathbf{K} = \mathbf{0}$. This means that synchronisation occurs even if there is no linear coupling between the model output and observations because of the non linear coupling introduced in the model (7.14). The eigenvalues for such a case are $\lambda_{1,2} = \pm\sqrt{b}$, where b is as in the matrix $\mathbf{A}$. However, it might be that with noise, the

out-of-sample error is not optimal for this coupling and can be improved with some other linear coupling.

To investigate this possibility we once again use results from control theory and thus need observability of the system to be satisfied. In our example, $x_n \in \mathbb{R}^2$ so our observability matrix is

$$\mathcal{O} = [\mathbf{HA} \quad \mathbf{HA}^2]^T. \tag{7.16}$$

It is straightforward to check that the system we are working with here is observable provided that $b \neq 0$. Since

$$\mathbf{H} = [1 \quad 0] \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 0 & 0.3 \\ 1 & 0 \end{bmatrix} \tag{7.17}$$

it follows that

$$\mathbf{HA} = [0 \quad 0.3] \quad \text{and} \quad \mathbf{HA}^2 = [0.3 \quad 0] \tag{7.18}$$

and hence the observability matrix in this case, has full rank.

The appropriate $\mathbf{K}$ for a desired characteristic polynomial, $q(\lambda)$ of the matrix $(\mathbf{A}-\mathbf{KHA})$ follows from Ackermann's Formula (Dorf & Bishop 2005) which is given by

$$\mathbf{K} = q(A)\mathcal{O}^{-1}[0 \ldots 1]^T. \tag{7.19}$$

Suppose that the desired characteristic equation is given by

$$q(\lambda) = (\lambda + \alpha)(\lambda - \alpha) \tag{7.20}$$

so that $\lambda_1 = -\lambda_2$ and $|\lambda_1| = |\lambda_2| = \alpha$. Then by Ackermann's formula we get

$$\mathbf{K} = \begin{bmatrix} 1 - \alpha^2/b \\ \\ a\alpha^2/b^2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{HK} = 1 - \frac{\alpha^2}{b} \tag{7.21}$$

where $a = 0$ and $b = 0.3$ as in the matrix $\mathbf{A}$. From (7.21) we see that as $\alpha \to 0$, $\mathbf{HK} \to 1$. Thus,

$$y_n = \mathbf{H}z_n = (\mathbb{1} - \mathbf{HK})\mathbf{H}\hat{z}_n + \mathbf{HK}\eta_n \to \eta_n, \tag{7.22}$$

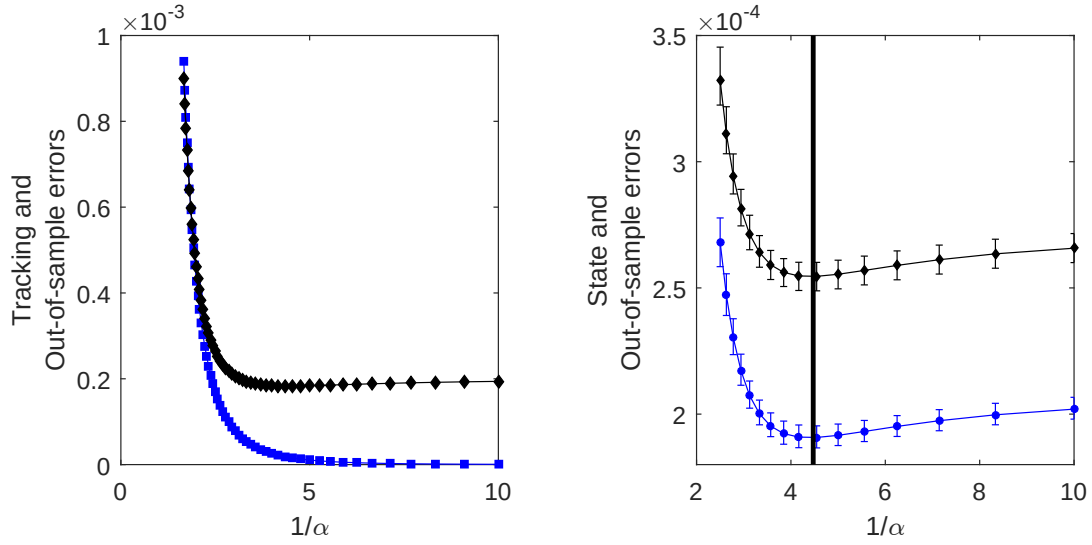
meaning that our data assimilation scheme simply replaces y_n with η_n, implying that the tracking error is zero. However this does not imply perfect data assimilation, by which we mean that the tracking tending to zero does not imply that the out-of-sample error is also small.

From (7.10) we know that

$$\mathbb{E}[y_n - \eta_n']^2 - \mathbb{E}[y_n - \eta_n]^2 = 2\sigma^2 \left(1 - \frac{\alpha^2}{b}\right). \tag{7.23}$$

Recall that the aim of this work is to find a way to estimate the out-of-sample error to get a more realistic picture of model performance. We have already determined that when there is no linear coupling (i.e. $\mathbf{K} = \mathbf{0}$) the system is stable and synchronisation occurs. We can see from (7.23) that this happens when $\alpha = \pm\sqrt{b}$. There are two further cases to consider. When $\alpha^2 > b$ the feedback, due to the linear coupling, is negative. Therefore, in this case we will not be able to improve the out-of-sample error. However as α tends to zero the optimism will increase and be bounded by $2\sigma^2$. Therefore when $\alpha^2 < b$ it may be possible to improve the out-of-sample error and determine a coupling matrix $\mathbf{K} \neq \mathbf{0}$ to use in the model.

To calculate the errors in the numerical simulation we approximate the expected value



(a) (b)

Figure 7.1: Figure 7.1(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. The errors are plotted against the inverse of α for $\sigma = 0.01$. Figure 7.1(b) shows a plot of the out-of-sample error in black diamonds for 100 realisations of the observational noise r_n with $\sigma = 0.01$. It is displayed for the range of α where the minimum occurs. The error bars represent 90% confidence intervals. The state error is show in blue circles also for 100 realisations of the observation noise with 90% confidence intervals. The vertical line draws attention to the minimum of both curves.

of a random variable, $\mathbb{E}[X]$, by the empirical mean squared error. Thus, (7.23) becomes

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - \eta'_n)^2 - \frac{1}{N}\sum_{n=1}^{N}(y_n - \eta_n)^2 = 2\sigma^2\left(1 - \frac{\alpha^2}{b}\right). \tag{7.24}$$

Any uncertainty in the calculation of the optimism will be assessed through bootstrapping. This is a statistical method used to assign measures of accuracy to sample estimates. In our case we run the experiment many times, each time changing the noise r_n so that the sample estimate is different every time. We then construct confidence intervals as a measure of accuracy.

The results obtained from our numerical experiment to test the theory described above are shown in Figure 7.1 and Mallia-Parfitt & Bröcker (2016). For these experiments we used $\sigma = 0.01$, $n = 10000$ and varied the parameter α between 0 and 1. Figure 7.1(a) shows the tracking error in blue squares and the out-of-sample error in black diamonds

plotted against the inverse of α. We can see that the tracking error tends to zero with decreasing α. This is what we expected and is confirmed by using our analytical expression for the optimism.

The tracking and out-of-sample errors meet when $\alpha^2 = b$. To the left of this, when $\alpha^2 > b$, the tracking error is greater than the out-of-sample error. To the right, when $\alpha^2 < b$, the tracking error is smaller than the out-of-sample error. In fact the tracking error tends to zero while the out-of-sample error decreases and then starts to increase again resulting in a well defined minimum. This is because as the coupling strength increases, the observations are tracked too closely and thus the model output adapts too closely to the observations resulting in an increase in the out-of-sample error, much like we saw in the linear case. On the other hand, when α is large and the coupling strength is weak, the observations are tracked badly resulting in large tracking and out-of-sample errors.

The well defined minimum of the out-of-sample error is shown more clearly in Figure 7.1(b). Figure 7.1(b) shows the out-of-sample error (black diamonds) for the range of α where the minimum occurs. The figure shows the out-of-sample error for 100 realisations of the noise r_n with $\sigma = 0.01$. The error bars represent 90% confidence intervals for each α where the lower limit of the errorbars is plotted at the fifth percentile while the upper limit is plotted at the 95th.

Figure 7.1(b) also shows the sate error (blue circles) for 100 realisations of the noise r_n with $\sigma = 0.01$ and again with 90% confidence intervals for each α. The state error which we recall is defined by

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - x_i)^2. \tag{7.25}$$

The black, vertical line draws attention to the minimum of the out-of-sample error. However, we can see that the minimum is actually the same for both errors. When running data assimilation schemes, the state error is the error we are interested in minimising, however we only have access to the error in observation space. Even though this is the case, we have shown numerically that the minimising gain is the same for both errors.

What is particularly of interest here is that even though the dynamical system included a non linear term, the methodology still applies, provided that the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ are $< 1 - \epsilon$. If we consider the error dynamics for the noisy case we see that

$$e_{n+1} = (\mathbf{A} - \mathbf{KHA})e_n + \mathbf{K}r_{n+1} - (\mathbb{1} - \mathbf{KH})(q_{n+1} + \xi(\mathbf{H}x_n) - \xi(\eta_n)) \tag{7.26}$$

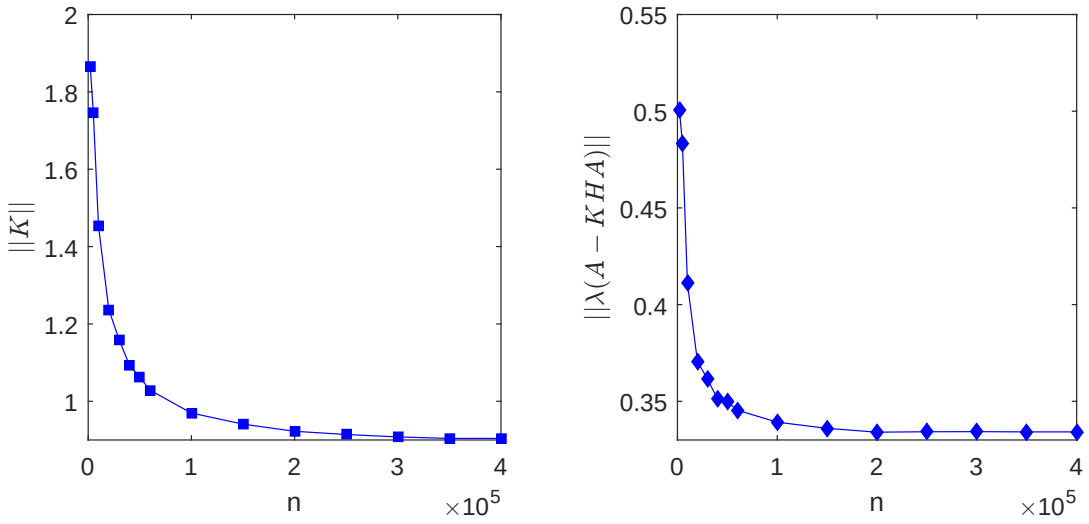
where $\xi(\cdot)$ represents the nonlinearity in the dynamical system. These error dynamics contain a linear part and a non linear part. This experiment suggests that the eigenvalues of the linear part of the error dynamics have to be $< 1 - \epsilon$ for the theory described above and in chapter 3 to hold.

7.3 Numerical Experiment II : Gain Convergence for Hénon Map

As a result of the process outlined above we are also able to determine the optimal coupling matrix, $\mathbf{K}$, to be used in the algorithm. The gain that minimises the out-of-sample error in the above experiments, is determined by arbitrarily choosing the parameter α. In order to analyse the asymptotic behaviour of this gain, we need to consider all possible gains that stabilise the system, much like we did for the linear systems in Chapter 3.

We ran some numerical experiments to test how the asymptotic behaviour of the gain matrix that minimises the out-of-sample error behaves asymptotically as in Mallia-Parfitt & Bröcker (2016). For this non-linear numerical experiment the following experimental setup was used: The reality is given by

$$x_{n+1} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{\mathbf{A}} x_n + c \begin{bmatrix} (\mathbf{H}x_n)^2 \\ 0 \end{bmatrix} + d + \rho q_{n+1} \tag{7.27}$$



(a) (b)

Figure 7.2: Figure 7.2(a) shows the convergence of the gain minimising the out-of-sample error by plotting the norm of the gain matrix $\mathbf{K}$ as n increases for $\sigma = 0.01$. Figure 7.2(b) is a plot of the norm of the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ for each gain minimising the out-of-sample error and we see that the eigenvalues too converge exponentially.

which for the values $a = 0$, $b = 0.3$, c=-1.4, $d = [1 \quad 0]^T$ is the chaotic Henon Map with corresponding observations

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{7.28}$$

where $\mathbf{H} = [1 \quad 0]$. The observer is set up in exactly the same way as in (7.13). The observational noise r_n is iid with mean zero and variance one and notice that we have added dynamical noise to the model in equation (7.27). This dynamical noise is also iid with $\mathbb{E}q_n = 0$ and $\mathbb{E}q_n q_n^T = 1$. If this noise wasn't present in the underlying system, then we could not expect the gain matrix to converge in a meaningful way as the gain may not be well defined. Even without coupling, it is possible that the observer and model will synchronise due to the presence of the η_n^2 term in the background term. However this does not mean the appropriate gain matrix in this case is the optimal one.

The results obtained in this experiment are shown in Figure 7.2. The observational noise is iid with $\mathbb{E}r_n = 0$, $\mathbb{E}r_n r_n^T = 1$ with $\sigma = 0.001$. The dynamical noise is also iid with mean zero and variance one with $\rho = 0.004$. The true evolution of the model which we

denote by n was taken to vary between 0 and 4×10^5. For each n the optimal gain was determined and recorded. We also calculated the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ for each minimising gain. It is expected that the gain matrix will converge as n increases, however in this non-linear case, determining the exact structure of this limit is not so straightforward.

When we considered the linear system in Chapter 3, we had the optimal linear filter (i.e the Kalman Filter) to compare the results with. However here, even though the observer is linear, the Kalman Filter is not optimal and thus the asymptotic gain as defined previously is not the limit in the convergence. Further to this, such a limit is difficult to determine. This is because we have little information on the correlation between the non-linear term and the other terms in the error dynamics. That being said we can still deduce some information from these numerical experiments.

The results are shown in Figure 7.2. Figure 7.2(a) shows a plot in blue squares of $||\mathbf{K}||$ against n. It is evident from the figure that the constant gain matrix that minimises the out-of-sample error converges exponentially. This is further confirmed in Figure 7.2(b) in which it is clear that the eigenvalues of $(\mathbf{A} - \mathbf{KHA})$ for each gain also converge. This second figure shows a plot in blue diamonds of $||\lambda||$ against n and we can see that the convergence here is also exponential.

7.4 Numerical Experiment III: Lorenz '96

For this third numerical experiment (as presented in Mallia-Parfitt & Bröcker (2016)), the reality is given by the Lorenz'96 model which is governed by the following equations

$$\dot{x_i} = -x_{i-1}(x_{i-2} - x_{i+1}) - x_i + F \tag{7.29}$$

and exhibits chaotic behaviour for $F = 8$. By solving the above differential equation we obtain a discrete model for our reality which we denote by

$$x_{n+1} = \Phi(x_n). \tag{7.30}$$

We take corresponding observations of the form

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{7.31}$$

where $\mathbf{H}$ is the observation operator and r_n is iid noise. We shall take the state dimension to be $D = 12$, the observation space to be $d = 4$ and we define the observation operator so that we observe every third element of the state. The system we construct here is fully non-linear with linear observations.

The assimilating model will use the Lorenz'96 model coupled to the observations through a simple linear coupling term, as done in the the previous numerical experiments. We set the coupling matrix $\mathbf{K}$, to be defined by

$$\mathbf{K} = \kappa\mathbf{H}^T \tag{7.32}$$

where κ is a coupling parameter taken to be between 0 and 1. With this information, the assimilating model is defined by the following equations

$$\hat{z}_{n+1} = \Phi(z_n); \quad z_{n+1} = \hat{z}_{n+1} + \kappa\mathbf{H}^T(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}). \tag{7.33}$$

Once again we will vary the coupling strength in the observer by adjusting the coupling parameter κ. If the coupling is too strong, the observations will be tracked too rigorously and so the observational noise will not be filtered out. If the coupling is too weak the observations are tracked poorly; so once again we expect the out-of-sample error to take a minimum at some non-trivial value of κ.

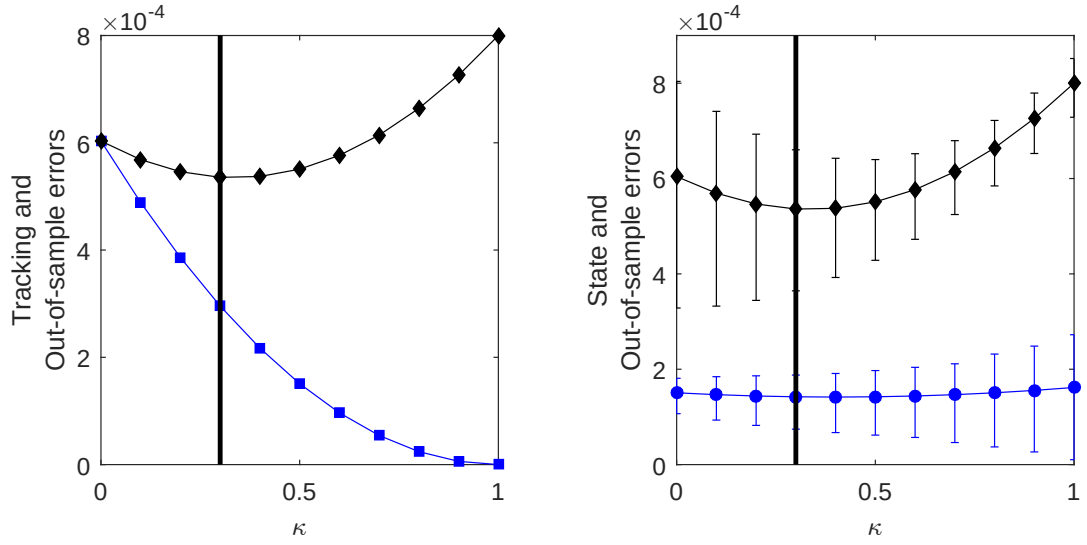


Figure 7.3: Figure 7.3(a) presents the out-of-sample error (black diamonds) and the tracking error (blue squares). Figure 7.3(b) illustrates the out-of-sample error (black diamonds) and the state error (blue circles) with the error bars representing 90% confidence intervals. The black vertical line draws attention to the minimum of the out-of-sample error.

As always we are interested in the behaviour of the state error and, ultimately, this is the error we want to be minimal. We saw in Section 7.2 that the minimiser for the out-of-sample error was the same as for the state error. This shall be investigated here also.

The results obtained are shown in Figure 7.3 and in Mallia-Parfitt & Bröcker (2016). The model was integrated with a time step $\delta = 1.5 \times 10^{-2}$. Once again the observational noise is iid with $\mathbb{E}r_n = 0$, $\mathbb{E}r_n r_n^T = 1$ and $\sigma = 0.01$. Since the gain is given by equation (7.32), the optimism reduces to $8\sigma^2\kappa$. To calculate the the errors, a transient time was ignored to give the system time to synchronise. In Figure 7.3(a) the out-of-sample error (black diamonds) is presented together with the tracking error (blue squares). The black vertical line draws the eye to the minimum of the out-of-sample error. As in the previous experiments, the tracking error reduces to zero while the out-of-sample error increases due to the change in coupling strength.

Figure 7.3(b) presents the out-of-sample error (black diamonds) and the state error

(blue circles). The figure shows the errors for 100 realisations of the observational noise, r_n. The error bars represent 90% confidence intervals for each value of κ with the lower limit of the error bars taken at the fifth percentile and the upper limit taken at the 95th. Again, the black line draws attention to the minimum of the out-of-sample error and we once again see that the minima of the state and out-of-sample errors coincide. It is evident that these results support the results determined previously in the numerical experiments.

The minimisers of the out-of-sample error and state error coincide just as shown in Section 7.3. The tracking error, with increasing coupling strength, decreases and converges to zero while the out-of-sample error increases and thus has a well defined minimum. The optimism monotonously increases with increasing coupling strength.

Thus it is clear here that even for fully non-linear systems with linear observations, the theory of out-of-sample error holds. This is because the calculation of the optimism only depends on the observation operator and does not depend on the structure of the background term; neither does it depend on how the background term is obtained.

The asymptotic behaviour of the optimal gain depends heavily on the presence of dynamical noise. As we have seen previously, the gain cannot be expected to converge in a significant way without the presence of model noise. For example it is possible that the dynamics may enter a region of stability resulting in a reduction of the error. In this case it would make sense to reduce or eliminate the coupling; however such gain matrices are not being considered here.

Ideally we would like to rigorously prove that for non-linear systems, the sequence of minimising feedback gains converges to some asymptotic gain. This task is not an impossible one; in fact the same linear proof presented in Chapter 6 can be adapted (for the most part) to work for non-linear systems. Unfortunately, there is one fundamental

hurdle that cannot be so easily overcome. To see this consider a dynamical model given by

$$\begin{aligned} x_{n+1} &= \mathbf{A}x_n + \xi(\mathbf{H}x_n) + q_{n+1} \\ \eta_n &= \mathbf{H}x_n + r_n \end{aligned} \tag{7.34}$$

where x is the state, η are the observations and q_n and r_n are iid model and observation errors with covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ respectively. Assume that q_n and r_n are uncorrelated and that ξ is Lipschitz continuous. Assume also that the observation error covariance matrix is strictly positive definite.

For the system given by (7.34) we construct an observer of the form

$$\begin{aligned} \hat{z}_n &= \mathbf{A}z_{n-1} + \xi(\eta_{n-1}) \\ z_n &= \hat{z}_n + \mathbf{K}_n(\eta_n - \mathbf{H}\hat{z}_n) \end{aligned} \tag{7.35}$$

where $\mathbf{K}_n$ is the gain matrix which may or may not depend on n. Then the error dynamics are given by

$$e_{n+1} = (\mathbf{A} - \mathbf{KHA})e_n - (\mathbb{1} - \mathbf{KH})(q_{n+1} - \xi(\eta_n) + \xi(\mathbf{H}x_n)) + \mathbf{K}r_{n+1} \tag{7.36}$$

and the error covariance matrices can be calculated to obtain the following equations:

$$\Gamma_n = \mathbb{E}[(z_n - x_n)(z_n - x_n)^T] = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\Sigma_n(\mathbf{I} - \mathbf{K}_n\mathbf{H})^T + \mathbf{K}_n\mathbf{R}\mathbf{K}_n^T \tag{7.37}$$

$$\Sigma_n = \mathbb{E}[(\hat{z}_n - x_n)(\hat{z}_n - x_n)^T] = \mathbf{A}\Gamma_{n-1}\mathbf{A}^T + \mathbf{Q} + T_{n-1} \tag{7.38}$$

where

$$\begin{aligned} T_n &= \mathbb{E}[(\xi(\eta_n) - \xi(\mathbf{H}x_n))(\xi(\eta_n) - \xi(\mathbf{H}x_n))^T] \\ &\quad + \mathbb{E}[\mathbf{A}(z_n - x_n)(\xi(\eta_n) - \xi(\mathbf{H}x_n))^T] + \mathbb{E}[(\xi(\eta_n) - \xi(\mathbf{H}x_n))(z_n - x_n)^T\mathbf{A}^T]. \end{aligned} \tag{7.39}$$

The ideas using the theory of M-estimators to prove the result can still be applied however it is essential that we have further information about the correlation between the non-linear term in (7.36) and the remaining terms in the error dynamics. This is the main difficulty in constructing a rigorous proof. More restrictive assumptions will have to be made on the non-linear term and on the system itself.

In addition to this problem, we still have to find a candidate for the asymptotic gain in the non-linear case. It may be reasonable to take this candidate to be the asymptotic limit of the error covariance in (7.37), however this will include taking the limit of the extra term, T_n, which represents the correlation between the non-linearity and the error itself. If the gain matrix converges then T_n also converges; this is not difficult to determine. However, more information is still required as we have very little information about how this term behaves asymptotically.

Chapter Summary In this chapter we considered non-linear systems and the concept of the out-of-sample error for systems in Lur'e form and fully non-linear systems but with linear observations. We illustrated the theory working using the chaotic Hénon Map and Lorenz '96 system as the underlying dynamical models in the experiments. Numerical results show that the theory works in a very similar way to the linear case presented earlier in Chapter 3 and Mallia-Parfitt & Bröcker (2016). Establishing that the feedback gain matrix converges is slightly trickier in this setting however it has been established numerically for system in Lur'e form. Rigorously proving this fact is not so straightforward as we have little information about the non-linear term and its correlation with the other terms in the error dynamics.

Chapter 8

Conclusion

When considering data assimilation algorithms, it is essential that the performance of these schemes is analysed. Perhaps even more important than simply assessing the performance, the analysis must be done in such a way that it provides a true and honest assessment of the algorithm. The traditional way of determining how well an algorithm performs, is to compare the output with the measured observations. However, this can easily provide a false assessment of the performance since the measurements are already used to obtain the output in the first place. Using completely independent observations from the same time and region to assess the performance will be the ideal option, however practically this is not feasible as such observations hardly ever exist.

A possible remedy was suggested by considering the out-of-sample error which we recall is simply the error between the output and the true observation added to the variance of the observational noise. Numerical experiments utilising both linear and non-linear systems, suggested that this error provides a better assessment of performance. Where the tracking error (the error between the output and measured observations) approached zero with increasing coupling strength, the out-of-sample error increased again, resulting in a well-defined minimum and thus an optimal coupling parameter.

When running data assimilation schemes, the error we are ultimately interested in

reducing is the error between the underlying state and the trajectory obtained by the data assimilation algorithm. However, calculating this error is not possible in practice as we do not have access to the underlying state. Thus we must do the best we can using errors in observation space as the measured observations are the only real data that we have access to. Numerical experiments for both linear and non-linear systems, suggest however that the minimum of the out-of-sample error coincides with the minimum of the state error.

These numerical experiments were also used to determine the asymptotic behaviour of the optimal gain matrix that produced minimal out-of-sample error. Moreover, for linear systems, the limit of this convergence was shown to be the same as that for the Kalman Gain, the optimal gain used in the Kalman Filter. The challenge that followed was to rigorously prove that the optimal gain did indeed converge to the asymptotic Kalman Gain.

Proving this result for the expected error covariance was done first and it was established that the sequence of gain matrices that minimise the out-of-sample error does indeed converge to the asymptotic gain. However, in practice, it is the empirical mean of the error that is calculated not the error itself. Therefore, it was necessary to determine whether the minimiser of the empirical mean of the out-of-sample error converged to the minimiser of the asymptotic error, i.e the asymptotic Kalman Gain.

Using ideas from the theory of asymptotic statistics, in particular the theory of M-estimators, a detailed proof of the aforementioned result was presented. The proof boiled down to establishing that the estimator was asymptotically consistent, however a direct application of known theorems was not so straightforward in this specific setting. That being said, results using stochastic equicontinuity to establish uniform convergence were adapted and given certain assumptions the required result was established. There is still one outstanding result (Conjecture 6.2.1) that prevents us from completing the proof in its entirety. We cannot state for certain that all potential minimisers stabilise the error dynamics. This was achieved for the deterministic case (i.e. Chapter 5), however it is not so straightforward for the stochastic case.

Naturally, proving this result for non-linear system was considered. The numerical experiments for systems in Lur'e form and indeed for fully non-linear systems also, established that the theory of the out-of-sample error applies. The results presented show that provided certain conditions are met, an optimal gain matrix can be determined in the sense that the out-of-sample error is minimised. Moreover, once again it was shown numerically that the minimum of the out-of-sample error is the same as that for the state error.

When considering the convergence of the optimal gain matrix for non-linear systems, the presence of dynamical noise in the underlying system becomes extremely important. If there is no model noise present, then we cannot expect the gain matrix to converge in a meaningful way as this gain may not be well defined. For example it is possible that the dynamics may enter a region of stability, resulting in a reduction of the error. In this case it would make sense to reduce or completely eliminate the coupling parameter. This would need the coupling matrix to be adaptive in some way; a concept not considered here. However, if one does add model noise to the system, convergence of the optimal gain may occur. In this event it is desirable to prove that the same results that hold for linear systems, apply in the non-linear case. In the case of a Lur'e system, the added complication comes from the presence of the nonlinearity. In particular the problem is in the correlation between the non-linear term and the other terms in the error dynamics. In order to apply a similar proof to that of the linear system, a good understanding of this correlation will be required.

Appendix A

The Best Linear Unbiased Estimate Analysis

Recall that the problem of four-dimensional variational data assimilation (4D-Var) is to find the initial state that minimizes the weighted least squares distance to the background while minimizing the weighted least squares distance of the model trajectory to the observation over the time interval $[t_0, t_N]$, Lawless (2012). Mathematically, we write this as an optimization problem:

Find the analysis state x_0^a at time t_0 that minimizes the function

$$J(x_0) = \frac{1}{2}(x_0 - x^b)^T \mathbf{B}^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{n=0}^{N}(h(x_n) - \eta_n)^T \mathbf{R}_n^{-1}(h(x_n) - \eta_n) \tag{A.1}$$

subject to the states x_n satisfying a specified non-linear dynamical system. The minimization problem given by (A.1) can be interpreted in a statistical or deterministic sense. From Bayes' Theorem it can be shown that x_0^a gives the maximum likelihood estimate of the state under the assumptions that all errors considered are Gaussian. Alternatively, the term measuring the fit to the background state can be thought of as a form of regularisation in fitting the observations, Lawless (2012).

The Best Linear Unbiased Estimate (BLUE), obtained through least squares fitting is

given by

$$x^a = x^b + \mathbf{K}(\eta - h(x^b)), \quad \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{A.2}$$

where $\mathbf{K}$ is called the gain or weight matrix. This BLUE analysis is equivalently obtained as a solution to the variational optimisation problem. Euqation (A.2) is the mathematical expression of the fact that we want the analysis to depend linearly on the difference between the background and the observations. We also want the analysis state to be as close as possible to the true state in the sense that we want it to be a minimum variance estimate. In the case of Gaussian errors (which we assume here), the minimum variance estimate is equivalent to the maximum likelihood estimate.

To show that the BLUE analysis is equivalently obtained as a solution to the variational optimisation problem, consider the optimization problem for 3D-Var: Find the state x that minimises the cost function,

$$J(x) = (x - x^b)^T\mathbf{B}^{-1}(x - x^b) + (\eta - h(x))^T\mathbf{R}^{-1}(\eta - h(x)). \tag{A.3}$$

If we assume that x^a is the state that minimises $J(x)$ so that $\nabla J(x^a) = 0$ and that h is linear so that $h(x) - h(x^b) = \mathbf{H}(x - x^b)$ where $\mathbf{H} = \partial h/\partial x$ we have that

$$\begin{aligned} 0 &= \nabla J(x^a) \\ &= \mathbf{B}^{-1}(x^a - x^b) - \mathbf{H}^T\mathbf{R}^{-1}(\eta - h(x^a)) \\ &= \mathbf{B}^{-1}(x^a - x^b) - \mathbf{H}^T\mathbf{R}^{-1}(\eta - h(x^b) - \mathbf{H}(x^a - x^b)) \\ \Rightarrow x^a - x^b &= (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\eta - h(x^b)). \end{aligned} \tag{A.4}$$

By the Shermann-Morrison-Woodbury formula (Bartlett 1951) we see that

$$(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{A.5}$$

Hence, we have that

$$x^a = x^b + \mathbf{K}(\eta - h(x^b)), \quad \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{A.6}$$

as expected. We can also calculate the analysis error covariance matrix which is defined by

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T. \tag{A.7}$$

This expression is obtained directly from calculating $\mathbb{E}[(x^a - x^t)(x^a - x^t)^T]$. In the case where $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$, the optimal choice for $\mathbf{K}$, the analysis error covariance matrix reduces to

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}. \tag{A.8}$$

Calculating the BLUE directly can result in some difficulties as it requires the inversion of large matrices. We also require a definition of the $\mathbf{B}$ matrix which is not always possible in real world systems and it is difficult to use non-linear observation operators. Therefore, we need to consider minimising the cost function directly which is the aim of variational data assimilation.

The cost function can be minimised using iterative numerical methods such as conjugate gradient or quasi-Newton methods. On each iteration the value of the cost function and its gradient at the current iterate must be calculated. To do this we solve the discrete adjoint equations which we obtain through the method of Lagrange Multipliers. Define Lagrange multipliers λ and the Lagrangian

$$L = J(x_0) + \sum_{n=0}^{N} \lambda_{n+1}^T (x_{n+1} - f(x_n)). \tag{A.9}$$

The necessary conditions for a minimum are

$$\frac{\partial L}{\partial \lambda_n} = 0 \quad \text{and} \quad \frac{\partial L}{\partial x_n} = 0. \tag{A.10}$$

The first of these conditions yields $x_{n+1} = f(x_n)$, which is just our original constraint while the second yields the discrete adjoint equations given by

$$\lambda_n = \mathbf{F}_n^T \lambda_{n+1} - \mathbf{H}_n^T \mathbf{R}_n^T (h(x_n) - \eta_n), \quad n > 0 \tag{A.11}$$

with $\lambda_{n+1} = 0$. Here $\mathbf{H}_n$ and $\mathbf{F}_n$ are the Jacobians of the non-linear operators h_n and f_n with respect to the state variables x_n.These Jacobians are referred to as the *tangent linear operator* and the *tangent linear model* (TLM) respectively, Lawless (2012). The gradient of the cost function, $J(x)$, with respect to the initial state, x_0, is then

$$\nabla J(x_0) = \mathbf{B}^{-1}(x_0 - x^b) - \lambda_0 \tag{A.12}$$

where the operators $\mathbf{H}_n^T$ and $\mathbf{F}_n^T$ are the adjoints of the observation operator and the non-linear model. The adjoint is equal to the matrix transpose of the Jacobians these adjoints are usually taken with respect to the Euclidean inner product, Lawless (2012).

For numerical optimization methods, each iteration requires one run of the forward model to calculate the value of the cost function and one run of the adjoint model (A.11) to calculate the value of the gradient. This makes 4D-Var very expensive from a computational point of view. The possibility of implementing variational data assimilation in an operational setting came with the proposal of incremental variational data assimilation, Courtier et al. (1994).

Appendix B

Singular Value Decomposition

Singular Value Decomposition (SVD) is the method of choice for solving most linear least squares problems since it is considered to be a very stable method, Press et al. (1988). SVD methods are based on the following theorem.

Theorem B.0.1. *Any $m \times n$ matrix $\mathbf{X}$ whose number of rows m is greater than or equal to the number of columns n can be written as the product of an $m \times n$ column orthogonal matrix $\mathbf{U}$, an $n \times n$ diagonal matrix $\mathbf{D}$ with positive or zero elements and the transpose of an $n \times n$ orthogonal matrix $\mathbf{V}$:*

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{B.1}$$

The SVD of the matrix $\mathbf{X}$ has the form $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. Here $\mathbf{U}, \mathbf{D}, \mathbf{V}$ are as required in Theorem (B.0.1).

Using the singular value decomposition we can write the ridge regression fitted vector as

$$\begin{aligned} \mathbf{X}\hat{\beta}^{ridge} &= \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}_\lambda\mathbf{y} \\ &= \mathbf{U}\mathbf{D}(\mathbf{D}^2 + \lambda\mathbf{I})^{-1}\mathbf{D}\mathbf{U}^T\mathbf{y} \\ &= \sum_{i=1}^{p} \mathbf{u}_i \frac{d_i^2}{d_i^2 + \lambda}\mathbf{u}_i^T\mathbf{y} \end{aligned} \tag{B.2}$$

where $\mathbf{u}_i$ are the columns of $\mathbf{U}$ and d_i are the singular values of $\mathbf{X}$ and $\mathbf{H}_\lambda$ is called the *hat matrix*. Notice that $d_i^2/(d_i^2+\lambda) \leq 1$ since $\lambda \geq 0$. Ridge regression computes the coordinates of $\mathbf{y}$ with respect to the orthonormal basis $\mathbf{U}$; it then shrinks these coordinates by the factors $d_i^2/(d_i^2+\lambda)$. The greater amount of shrinkage is applied to the coordinates of basis vectors with smaller d_j^2, Hastie et al. (2009).

Using the singular value decomposition for ridge regression (B.2), we can get a closed form expression for the effective degrees of freedom. This quantity is given by

$$\begin{aligned} \mathrm{df}(\lambda) &= \mathrm{tr}(\mathbf{H}_\lambda) \\ &= \mathrm{tr}(\mathbf{X}(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I})^{-1}\mathbf{X}^T) \\ &= \sum_{i=1}^{p} \frac{d_i^2}{d_i^2+\lambda}. \end{aligned} \tag{B.3}$$

This is a monotone decreasing function of λ and it is known as the effective degrees of freedom of the ridge regression fit. Usually in a linear regression fit with p variables, the degrees of freedom of the fit is p, the number of free parameters; however in ridge regression they are fit in a restricted fashion controlled by λ although all p coefficients will be non-zero. Note that when $\lambda = 0$, $\mathrm{df}(\lambda) = p$ and $\mathrm{df}(\lambda) \to 0$ as $\lambda \to \infty$.

Appendix C

Asymptotic Properties of the Kalman Filter

The proof of Theorem 4.2.1 is done in the following four steps, as done in Anderson & Moore (1979):

- Σ_n is bounded for all n.
- For zero initial condition, Σ_n is monotone increasing with n and together with the bound in the first point, it establishes the existence of $\lim \Sigma_n$. Equation (4.16) will be recovered.
- The stability property is established.
- Allow for arbitrary non-negative symmetric initial condition, Σ_0.
- Establish uniqueness and positive definiteness of Σ_∞.

We shall prove these items in that order next.

Proof of Theorem 4.2.1.

Bound on the Error Covariance

We define a suboptimal filter whose error covariance must over bound Σ_n. By the assumption of observability of the pair $(\mathbf{A}, \mathbf{H})$, there exists a matrix $\mathbf{K}_e$ such that $|\lambda_i(\mathbf{A} - \mathbf{A}\mathbf{K}_e\mathbf{H})| < 1$.

Define a suboptimal, asymptotically stable filter by

$$z_{n+1}^e = \mathbf{A}z_n^e + \mathbf{K}_e[\eta_n - \mathbf{H}z_n^e], \quad z_0^e = 0. \tag{C.1}$$

The error covariance is given by

$$\begin{aligned} \Sigma_n^e &= \mathbb{E}(x_n - z_n^e)(x_n - z_n^e)^T \\ &= (\mathbf{A} - \mathbf{A}\mathbf{K}_e\mathbf{H})\Sigma_n^e(\mathbf{A} - \mathbf{A}\mathbf{K}_e\mathbf{H})^T + \mathbf{A}\mathbf{K}_e\mathbf{R}\mathbf{K}_e^T\mathbf{A}^T + \mathbf{Q}. \end{aligned} \tag{C.2}$$

If we are comparing (C.1) with an optimal filter initialised by Σ_0, the initial uncertainty in x_0 is Σ_0 and by (C.1), we have $\Sigma_0^e = \Sigma_0$. But (C.1) is a sub-optimal filter so in general, $\Sigma_n^e \geq \Sigma_n \geq 0$. Because of the stability of this suboptimal filter, Σ_n^e has a bounded solution for any initial condition. Thus we have obtained a bound on Σ_n.

Use of Zero Initial Condition

Suppose now that $\Sigma_0 = 0$. We shall show that Σ_n is increasing with n. The argument simply says that if two filtering problems are considered that are identical except that the initial uncertainty is greater for one than the other, then the ordering property in the errors in estimating the state at an arbitrary time will be preserved.

Consider the variance equation in (4.13) with two initial conditions $\Sigma_0^{(1)}$, $\Sigma_0^{(2)}$ with the property that

$$0 = \Sigma_0^{(1)} \leq \Sigma_0^{(2)}. \tag{C.3}$$

We shall use induction to establish that

$$\Sigma_n^{(1)} \leq \Sigma_n^{(2)} \tag{C.4}$$

which implies that the ordering property of the initial conditions is preserved. This ordering is true for $n = 0$ and now assume that it is also true for $n = 1, \ldots, i-1$. Then an optimal version of the variance equation yields,

$$\begin{aligned}
\Sigma_i^{(2)} &= \min_{\mathbf{K}}[(\mathbf{A} - \mathbf{AKH})\Sigma_{i-1}^{(2)}(\mathbf{A} - \mathbf{AKH})^T + \mathbf{Q} + \mathbf{AKRK}^T\mathbf{A}^T] \\
&= (\mathbf{A} - \mathbf{AK}^*\mathbf{H})\Sigma_{i-1}^{(1)}(\mathbf{A} - \mathbf{AK}^*\mathbf{H})^T + \mathbf{Q} + \mathbf{AK}^*\mathbf{R}\,\mathbf{K}^{*T}\,\mathbf{A}^T \\
&\geq (\mathbf{A} - \mathbf{AK}^*\mathbf{H})\Sigma_{i-1}^{(1)}(\mathbf{A} - \mathbf{AK}^*\mathbf{H})^T + \mathbf{Q} + \mathbf{AK}^*\mathbf{R}\,\mathbf{K}^{*T}\,\mathbf{A}^T \\
&\geq \min_{\mathbf{K}}[(\mathbf{A} - \mathbf{AKH})\Sigma_{i-1}^{(1)}(\mathbf{A} - \mathbf{AKH})^T + \mathbf{Q} + \mathbf{AKRK}^T\mathbf{A}^T] \\
&= \Sigma_i^{(1)}
\end{aligned} \tag{C.5}$$

where $\mathbf{K}^*$ is the minimising gain. The underlying time-invariance of all the quantities in the variance equation save for the covariance matrix itself and that the initial condition was zero, leads to the required result that Σ_n is monotone increasing.

In the previous section we showed that the error covariance was bounded above, so we know that the limit in (4.15) exists when $\Sigma_0 = 0$. Taking limits in (4.13) yields the DARE (4.16).

Asymptotic Stability of the Filter

Assume controllability as defined in definition 4.2.2. Then we can prove asymptotic stability of the filter by contradiction. In that spirit, suppose asymptotic stability does not hold so that $(\mathbf{A} - \mathbf{AKH})^T v = \lambda v$ for some λ with $|\lambda| \geq 1$, $v \neq 0$. Then,

$$\Sigma_\infty = (\mathbf{A} - \mathbf{AK}_\infty\mathbf{H})\Sigma_\infty(\mathbf{A} - \mathbf{AK}_\infty\mathbf{H})^T + \mathbf{AK}_\infty\mathbf{RK}_\infty^T\mathbf{A}^T + \mathbf{Q} \tag{C.6}$$

and by our assumption, we get

$$(1 - |\lambda|^2)v^T\Sigma_\infty v = v^T\mathbf{A}\mathbf{K}_\infty\mathbf{R}\mathbf{K}_\infty^T\mathbf{A}^T v + v^T\mathbf{Q}v. \tag{C.7}$$

The left hand side of this equation is non-positive since $|\lambda| \geq 1$ while the right hand side is clearly non-negative. Therefore, both side must equal zero. This implies that $(\mathbf{A}\mathbf{K}_\infty)^T v = 0$ and $v^T\mathbf{Q}v = 0$. But $(\mathbf{A}\mathbf{K}_\infty)^T v = 0$ implies that $\mathbf{A}^T v = \lambda v$ by our assumption that asymptotic stability does not hold, and combined with the fact that $v^T\mathbf{Q}v = 0$, we have a lack of controllability. Thus, we have a contradiction and so the filter is asymptotic stable.

Non-Zero Initial Covariance

In order to generalise the above for any non-zero initial condition, we use the squeeze theorem. Consider, as we have seen so far, the optimal filter initialised by a zero initial condition ($\Sigma_0 = 0$). Then consider the same optimal filter being initialised with a non-zero initial condition so that $0 \leq \Sigma_0$. Finally consider a suboptimal stable filter initialised at some non-zero $\Sigma_0^e = \Sigma_\infty$. Since this filter is a suboptimal one, the initial data satisfies the following relation,

$$0 \leq \Sigma_0 \leq \Sigma_0^e. \tag{C.8}$$

The error covariance for the optimal filter initialised at zero converges to Σ_∞; the error covariance for the suboptimal filter initialised at Σ_∞ will remain equal to the initial condition. Therefore, by the squeeze theorem it follows that the error covariance initialized by Σ_0 converges to Σ_∞ also.

An alternative proof of this generalisation can be found in Anderson & Moore (1979).

Σ_∞ is the unique positive definite solution to (4.16)

The steady state equation (4.16) is non linear. Therefore it can be generally expected to

have more than one solution. Only one however can be non-negative definite and symmetric. Suppose that $\hat{\Sigma} \neq \Sigma_\infty$ is such a solution. Then with an initial condition of $\hat{\Sigma}$, (4.16) yields by continuity, $\Sigma_n = \hat{\Sigma}$ for all n while (4.15) yields $\lim_{n\to\infty} \Sigma_n = \Sigma_\infty \neq \hat{\Sigma}$, which is a contradiction. This means that the asymptotic gain in (4.14) is uniquely obtained.

□

Bibliography

Anderson, B. & Moore, J. (1979), *Optimal Filtering*, Dover Publications Inc.

Andrews, D. W. (1994), 'Empirical process methods in econometrics', *Handbook of econometrics* **4**, 2247–2294.

Arnold III, W. F. & Laub, A. J. (1984), 'Generalized eigenproblem algorithms and software for algebraic riccati equations', *Proceedings of the IEEE* **72**(12), 1746–1754.

Barnes, S. L. (1964), 'A technique for maximizing details in numerical weather map analysis', *Journal of Applied Meteorology* **3**(4), 396–409.

Bartlett, M. S. (1951), 'An inverse matrix adjustment arising in discriminant analysis', *The Annals of Mathematical Statistics* **22**(1), 107–111.

Bennett, A. F. & Thorburn, M. A. (1992), 'The generalized inverse of a nonlinear quasi-geostrophic ocean circulation model'.

Billingsley, P. (1968), *Convergence of probability measures*, Wiley New York.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press Inc.

Boccaletti, S., Kurths, J., Osipov, G., Valladares, D. & Zhou, C. (2002), 'The synchronization of chaotic systems', *Physics Reports* **366**, 1–101.

Bröcker, J. & Szendro, I. G. (2012), 'Sensitivity and out-of-sample error in continuous time data assimilation', *Quarterly Journal of the Royal Meteorological Society* **138**(664), 1785–801.

Cardinali, C., Pezzulli, S. & Andersson, E. (2004), 'Influence-matrix diagnostic of a data assimilation system', *Quarterly Journal of the Royal Meteorological Society* **130**(603), 2767–2786.

Charney, J. (1951), 'Dynamical forecasting by numerical process', *Compendium of meteorology* .

Ciccarella, G., Dalla Mora, M. & Germani, A. (1993), 'A luenberger-like observer for nonlinear systems', *International Journal of Control* **57**(3), 537–556.

Collet, P. & Eckmann, J.-P. (2007), *Concepts and results in chaotic dynamics: a short course*, Springer Science & Business Media.

Courtier, P. & Talagrand, O. (1987), 'Variational assimilation of meteorological observations with the adjoint vorticity equation. ii: Numerical results', *Quarterly Journal of the Royal Meteorological Society* **113**(478), 1329–1347.

Courtier, P., Thepaut, J. & Hollingsworth, A. (1994), 'A strategy for operational implementation of 4D-Var, using an incremental approach', *Quarterly Journal of the Royal Meteorological Society* **120**, 1367–1387.

Cressman, G. P. (1959), 'An operational objective analysis system', *Monthly Weather Review* **87**(10), 367–374.

Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2005), 'Diagnosis of observation, background and analysis-error statistics in observation space', *Quarterly Journal of the Royal Meteorological Society* **131**(613), 3385–3396.

Desroziers, G. & Ivanov, S. (2001), 'Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation', *Quarterly Journal of the Royal Meteorological Society* **127**(574), 1433–1452.

Dorf, R. & Bishop, R. (2005), *Modern Control Systems Tenth Edition*, Pearson Education Inc.

Efron, B. (1986), 'How biased is the apparent error rate of a prediction rule?', *Journal of the American Statistical Association* **81**(394), 461–470.

Efron, B. (2004), 'The estimation of prediction error: Covariance penalties and cross-validation', *Journal of the American Statistical Association* **99**(467).

Erdos, P. & Turán, P. (1950), 'On the distribution of roots of polynomials', *Annals of mathematics* pp. 105–119.

Ferguson, T. S. (1996), *A course in large sample theory*, Vol. 49, Chapman & Hall London.

Gandin, L. S. (1965), *Objective analysis of meteorological fields*, Vol. 242, Israel program for scientific translations Jerusalem.

Gauthier, J. P., Hammouri, H. & Othman, S. (1992), 'A simple observer for nonlinear systems applications to bioreactors', *Automatic Control, IEEE Transactions on* **37**(6), 875–880.

Gilchrist, B. & Cressman, G. P. (1954), 'An experiment in objective analysis', *Tellus* **6**(4), 309–318.

Greene, W. (1997), *Econometric Analysis*, New York: Prentice Hall.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction (Second Edition)*, Springer-Verlag.

Hughes, C. P. & Nikeghbali, A. (2008), 'The zeros of random polynomials cluster uniformly near the unit circle', *Compositio Mathematica* **144**(03), 734–746.

Huijberts, H. J. C., Nijmeijer, H. & Pogromsky, A. Y. (1999), 'Discrete-time observers and synchronization', *Controlling chaos and bifurcations in engineering systems* pp. 439–455.

Huijberts, H., Lilge, T. & Nijmeijer, H. (1998), 'Synchronization and observers for nonlinear discrete time systems', *Eindhoven University of Technology, Department of Mathematics and Computing Science* .

Hunter, J. K. & Nachtergaele, B. (2001), *Applied analysis*, Vol. 472, World Scientific.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013), *An introduction to statistical learning*, Vol. 112, Springer.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory Volume 64*, Academic Press Inc.

Kalnay, E. (2001), *Atmospheric Modeling, Data Assimilation and Predictability*, first edn, Cambridge University Press.

Krener, A. J. & Isidori, A. (1983), 'Linearization by output injection and nonlinear observers', *Systems & Control Letters* **3**(1), 47–52.

Krener, A. J. & Respondek, W. (1985), 'Nonlinear observers with linearizable error dynamics', *SIAM Journal on Control and Optimization* **23**(2), 197–216.

Lahoz, W., Khattatov, B. & Menard, R. (2010), *Data Assimilation: Making Sense of Observations*, Springer-Verlag.

Lawless, A. (2012), 'Variational Data Assimilation for very Large Environmental Problems'.

Le Dimet, F.-X. & Talagrand, O. (1986), 'Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects', *Tellus A* **38**(2), 97–110.

Lin, W. & Byrnes, C. I. (1995), 'Remarks on linearization of discrete-time autonomous systems and nonlinear observer design', *Systems & Control Letters* **25**(1), 31–40.

Lorenc, A. (1981), 'A global three-dimensional multivariate statistical interpolation scheme', *Monthly Weather Review* **109**(4), 701–721.

Lorenc, A. C. (1986), 'Analysis methods for numerical weather prediction', *Quarterly Journal of the Royal Meteorological Society* **112**(474), 1177–1194.

Mallia-Parfitt, N. & Bröcker, J. (2016), 'Assessing the performance of data assimilation algorithms which employ linear error feedback', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **26**(10).

Ménard, R. & Chang, L. (2000*a*), 'Stratospheric assimilation of chemical tracer observations using a kalman filter. part ii: Chi-square-validated results and analysis of variance and correlation dynamics', *Mon. Wea. Rev* **128**, 2672–2686.

Ménard, R. & Chang, L.-P. (2000*b*), 'Assimilation of stratospheric chemical tracer observations using a kalman filter. part ii: χ 2-validated results and analysis of variance and correlation dynamics', *Monthly weather review* **128**(8), 2672–2686.

Newey, W. K. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica: Journal of the Econometric Society* pp. 1161–1167.

Nijmeijer, H. & van der Schaft, A. (1990), *Nonlinear Control Systems*, Springer.

Panofsky, R. (1949), 'Objective weather-map analysis', *Journal of Meteorology* **6**(6), 386–392.

Pikovsky, A., Rosenblum, M. & Kurths, J. (2001), *Synchronization: A Univeral Concept in Nonlinear Sciences*, Cambridge University Press.

Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. (1988), *Numerical Recipes in C*, Cambridge University Press.

Prohorov, Y. V. (1956), 'Convergence of random processes and limit theorems in probability theory', *Theory of Probability & Its Applications* **1**(2), 157–214.

Rudin, W. (1964), *Principles of mathematical analysis*, Vol. 3, McGraw-Hill New York.

Sasaki, Y. (1958), 'An objective analysis based on the variational method', *Journal of the Meteorological Society, Japan* **36**, 77–88.

Sasaki, Y. (1970), 'Some basic formalisms in numerical variational analysis', *Monthly Weather Review* **98**, 875–883.

Szendro, I. G., Rodrìguez, M. A. & Lopez, J. M. (2009), 'On the problem of data assimilation by means of synchronization', *Journal Of Geophysical Research* **114**, D20109.

Talagrand, O. (1997), 'Assimilation of observations, an introduction', *Journal-Meteorogical Society of Japan Series 2* **75**, 81–99.

Tornambe, A. (1989), Use of asymptotic observers having-high-gains in the state and parameter estimation, *in* 'Decision and Control, 1989., Proceedings of the 28th IEEE Conference on', IEEE, pp. 1791–1794.

Tremolet, Y. (2006), 'Accounting for an imperfect model in 4D-Var', *Quarterly Journal of the Royal Meteorological Society* **132**(621), 2483–2504.

Van der Vaart, A. W. (2000), *Asymptotic statistics*, Cambridge university press.

Yang, S.-C., Baker, D. & Li, H. (2006), 'Data assimilation as snchronization of truth and model: Experiments with the three-variable Lorenz system', *Journal of the Atmospheric Sciences* **63**, 2340–2354.

Zeitz, M. (1987), 'The extended luenberger observer for nonlinear systems', *Systems & Control Letters* **9**(2), 149–156.