A BILINEAR KALMAN FILTER, A BILINEAR KALMAN SMOOTHER, AND THE BILINEAR EM ALGORITHM, WITH APPLICATIONS TO LOTKA-VOLTERRA MODEL

BY

ABDULLAH EQAL AL-MAZROOEI

A Dissertation Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In

MATHEMATICS

MAY 2011

KING FAHD UNIVERSITY OF PETROLEUM & MINIRALS
DHAHRAN, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This dissertation, written by ABDULLAH EQAL AL-MAZROOEI under the direction of his thesis advisors and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY IN MATHEMATICS.

Dissertation Committee

Prof. Muhammad El-Gebiely
Dissertation Committee Chairman

Dr. Jaafar Al-Mutwaa
Co-Chairman

Prof. R. P. Agarwal
Member

Dr. Hattan Tawfiq
Department Chairman

Dr. Ibrahim Hotait
Member

Dr. Salam Zummo
Dean of Graduate Studies

Dr. Mustapha Kassem
Member

1/6/11
Date

To

my parents,

my wife,

and to

Nura, Mohammed, Jna, and Jwana.

ACKNOWLEDGEMENTS

First, praise be to ALLAH who helped and guided me to accomplish this work. I ask him to inspire me and grant me a successful future.

I would like to thank my supervisor Professor Muhammed El-Gebiely, for his advice and prompt replies whenever I needed help. I greatly appreciate his honest comments on my work. Professor El-Gebiely has been a great mentor and a true friend throughout the years. I ask God to protect him and I wish him great success in his life. Many thanks to my co-supervisor Dr. Jaafar Almutawa for his suggestions and support throughout this thesis. His kind encouragement is much appreciated. I am also deeply grateful to the committee members, Prof. R. P. Agarwal, Dr. K. Mustapha and Dr. I. Hotait (from KAUST).

I extend my thanks to my brothers, in particular Ra'ed, for their help, support, and encouragement during my study.

Most importantly, I thank my parents and my wife from the bottom of my heart for their unconditional love and support throughout my life. They are the main factor behind the completion of this work. They supported me with their sincere prayers, love and concern. My deep thanks and gratitude to them.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Name: Abdullah Eqal Al-Mazrooei.

Title: A bilinear Kalman filter, a bilinear Kalman smoother, and a bilinear EM algorithm, with applications to Lotka-Volterra model.

Major Field: Mathematics.

Date of degree: May, 2011

The Kalman filter (KF) is an effective estimator of linear dynamical systems. However, most processes in real life are nonlinear. Bilinear systems are special kinds of nonlinear systems capable of representing a variety of important physical processes. Additionally, bilinear systems can be used in approximation or alternate representations for a range of other nonlinear systems. They are also used to model nonlinear processes in signal, images and communication systems. In particular, they arise in areas of engineering, chemistry, socioeconomics and biology. Bilinear systems can be represented as state space models.

In this dissertation, we develop a new bilinear model, in state space form. The evolution of this model not only depends on the state vectors, but also on the product of the state space by itself. Using this technique, we generalize many famous models such as: the well-known Lorenz 96 model and the Lotka-Volterra model which have many applications in real life. Since this bilinear model does not work with the traditional Kalman filter, we derive a new filter and a new smoother for our model. We refer to

them as the bilinear Kalman filter and the bilinear Kalman smoother, respectively.

Due to the widespread use of bilinear models, there is a strong motivation to develop identification algorithms for such systems. In this dissertation, we present the identification of our bilinear model to estimate the parameters which are used in the model under some statistical assumptions. We use a generalized technique known as bilinear Expectation Maximization algorithm which identifies the parameters, for certain classes of systems through Maximum Likelihood estimation.

To illustrate numerically the effectiveness of our proposed theoretical contribution (nonlinear filtering problem and the new technique for the identification process), we simulate our results by applying them to the well-known Lotka-Volterra model.

الخلاصة

الاسم: عبدالله بن عقال المزورعي

عنوان الأطروحة: فلتر كالمان ثنائي الخطية، منعم كالمان ثنائي الخطية، وخوارزمية تعظيم التوقع ثنائية الخطية، مع تطبيقات على نظام لوتكا- فولتيرا غير الخطي

التخصص الرئيسي: رياضيات

تاريخ نيل الدرجة: مايو ٢٠١١

فلتر (مصفي) كالمان هو مقدر فعال وناجح للأنظمة الديناميكية الخطية. غير أنَّ معظم الأنظمة في الحياة الواقعية غير خطية. تشكل الأنظمة ثنائية الخطية نوعاً خاصاً من الأنظمة غير الخطية، لها القدرة على تمثيل أنواع مختلفة وهامة من الأنظمة الفيزيائية غير الخطية. إضافة إلى ذلك، الأنظمة ثنائية الخطية يمكن أن تستخدم كتقريب أو تمثيل بديل لمجموعة من الأنظمة غيرالخطية الأخرى. وهي تستخدم أيضا لنمذجة العمليات غير الخطية في معالجة الصور والإشارات وأنظمة الاتصالات. على وجه الخصوص، ظهرت هذه الأنظمة في مجالات الهندسة، والكيمياء، والاقتصاد، والأحياء وغيرها. الأنظمة ثنائية الخطية يمكن تمثيلها على صورة ما يعرف بنموذج فضاء الحالة.

سنطور في هذه الأطروحة نموذجاً جديداً ثنائي الخطية في صورة نموذج فضاء حالة. تطوير هذا النموذج الجديد لا يعتمد فقط على متجه الحالة، بل يعتمد أيضاً على ضرب هذا المتجه بنفسه. بهذا الأسلوب الجديد في تطوير النموذج نعمم عدداً من النماذج الشهيرة، مثل نموذج لورنز٩٦ ونموذج لوتكا- فولتيرا والتي لها تطبيقات عملية. ولأنَّ هذا النموذج ثنائي الخطية لا يصلح للتعامل مع فلتر كالمان الخطي فإننا نشتق فلتراً جديداً ومنعماً جديداً تعمل مع النظام ثنائي الخطية الجديد. سوف نسميهما فلتر كالمان ثنائي الخطية ومنعم كالمان ثنائي الخطية، على الترتيب.

ونظراً للاستعمال الواسع للأنظمة ثنائية الخطية، فهناك محفز قوي لتطوير خوارزميات وطرق تشخيص وسائط تلك الأنظمة. في هذه الأطروحة سوف نقدم تشخيص نظام ثنائي الخطية الجديد لتقدير الوسائط المستخدمة في النظام الجديد تحت بعض الفرضيات الإحصائية. أسلوبنا الجديد في تشخيص النظام ثنائي الخطية الجديد يعتمد على تعميم ما يعرف بخوارزمية تعظيم التوقع. حيث نستخدم نظريات الأمثلة غير الخطية.

ونقدم أيضا في هده الأطروحة، محاكاة لنتائجنا الجديدة التي توصلنا إليها وأثبتناها. حيث نطبق هذه النتائج على ما يعرف بنظام لوتكا- فولتيرا غير الخطي، وذلك لشرح وتوضيح مدى فعالية وفائدة نتائجنا الجديدة التي توصلنا إليها على مسائل التقدير غير الخطية وعلى الأسلوب الجديد في عملية التشخيص.

Chapter 1

INTRODUCTION

1.1 Overview

Estimation theory is a branch of statistics and signal processing that deals with estimating the values of parameters based on measured/empirical data that has a random component [38]. The parameters describe an underlying physical setting in such a way that the value of the parameters affects the distribution of the measured data. An estimator attempts to approximate the unknown parameters using the measurements. The problem of estimating the values of a random process given measurements related to a random process is encountered in many fields of science and engineering such as

signal processing and control systems. Estimation theory has a rich history and its formative stages can be attributed to illustrious investigators such as Gauss, Legendre and others [31]. Gauss and Legendre made the first attempts in estimation theory. The first method for forming an optimal estimate from noisy data is the method of least squares. Its introduction is generally attributed to Gauss in 1795 while trying to estimate the positions of planets and comets using telescopic measurements. Then, the development in estimation theory appeared in the 1940s by Wiener and Kolmogorov when they introduced the filtering work.

In particular, Wiener introduced the importance of modeling not just noise, but noise and signal as a random process. His thought-provoking was released for open publication in 1949 and now is available in paperback form under the title "Time Series analysis". For very worthwhile background reading, see [16,31].

Between 1940 and 1950, many researches tried to relax the assumptions of the Wiener-Kolmogorov filtering theory and extended it. The breakthrough came with the Kalman filter which was introduced by Kalman in 1960. The Kalman filter extended the Wiener-Kolmogorov filter by changing the conventional formulation of the estimation problem and put it in a new direction completely different from the theory of stationary stochastic processes. Kalman made this change by introducing state-space theory where the important relationship between the estimation theory and the state-space

theory was established in [7,16,32,33].

The Kalman filter is an estimator used to estimate the state of a linear dynamic system perturbed by Gaussian white noise using measurements that are linear functions of the system state but corrupted by additive Gaussian white noise. The mathematical model used in the derivation of the Kalman filter is a reasonable representation for many practical problems, including control problems as well as estimation problems. The resulting estimator is statistically optimal with respect to certain quadratic function of estimation error. The model which is used with Kalman filter is called the state-space model. The state-space model is a mathematical model of a dynamical process that uses differential equations (or deference equations) to represent both deterministic and stochastic phenomena. The state variables of this model are the variables of interest. Random processes are characterized in terms of their statistical properties in the time domain, rather than in the frequency domain.

The Kalman filter was derived as the solution of the wiener filtering problem using the state space model for a dynamical and random process. The result is easier to derive and use than other filter. Practically, the Kalman filter is certainly one of the greater discoveries in the history of estimation theory and possibly the greatest discovery in the twentieth century. It has enabled humans to do many things that could not have been done without it [16]. Therefore, when the state space model is linear, then the

Kalman filter is the optimal estimator.

Unfortunately, most of the phenomena in the real life are nonlinear, so the Kalman filter is not suitable to use, and an appropriate extension of the Kalman filter has to be sought. There are different sorts of nonlinear systems depending on the variety in complexity and structures. Bilinear systems are a special type of nonlinear systems capable of representing a variety of important physical processes. They can be used in many applications in real life such as chemistry, biology, robotics, manufacturing, engineering and economics [37,51] that cannot be effectively modeled under the assumption of linearity. Bilinear systems have been extensively studied in recent years for three main reasons. Firstly, it has been shown that the bilinear systems are feasible mathematical models for large classes of problems of practical importance. Secondly, bilinear systems provide more flexible approximations to nonlinear systems than do linear systems. Thirdly, bilinear systems have rich geometric and algebraic structures that promise a fruitful field for researches [51]. Bilinear models were first introduced in the control theory literature in 1960s [54].

Due to the widespread use of bilinear models, there is a strong motivation to identify and develop identification algorithms for such models. There is a lot of work which presents methods of parameter estimation of bilinear systems. Many methods are used for the identification such as, least square estimation and maximum likelihood estimation. The maximum likelihood estimation can be computed by using the well-known

Expectation-Maximization algorithm (EM) [9]. The Expectation-Maximization (EM) algorithm is a broadly applicable approach to the iterative computation of Maximum Likelihood (ML) estimates, useful in a variety of incomplete data problems, where algorithms such as the Newton-Raphson method may turn out to be more complicated. On each iteration of the EM algorithm, there are two steps-called the Expectation step or the E-step and the Maximization step or the M-step. Because of this, the algorithm is called the EM algorithm [47]. This name was given by Dempster, Laird, and Rubin (1977) in their fundamental paper [9]. The identification of bilinear systems through the EM algorithm can be accomplished by using appropriate nonlinear filtering and nonlinear smoothing to compute the expectation step. However, optimization theory plays such an important role in establishing the maximization step that it will be considered as a nonlinear optimization problem.

1.2 Objectives of this dissertation

In this dissertation, we focus on studying linear and nonlinear estimation methods for specific linear and nonlinear mathematical models, by investigating how to derive filtering and smoothing for such models in a mathematical manner. We also study how the identification process can be accomplished for a linear mathematical model in

order to generalize to a nonlinear model of the bilinear class. Furthermore, we study how to apply and use optimization theory to achieve our objectives. Our objectives can be summarized as follows:

One: To study and understand the well-known Kalman filter and Kalman smoother for a linear and Gaussian discrete model in the state space form. We discuss the basic ideas and introduce the derivation of the Kalman filter using two techniques; (i) from a perspective of least square estimation and (ii) from a perspective of the Baysian estimation. We also introduce the derivation of the Kalman smoother.

Two: To explain the parameters identification of linear, Gaussian discrete state space models by using the expectation maximization algorithm.

Three: To introduce a new mathematical model. This new model is a bilinear, Gaussian, in a state space system, where, the dynamical system is bilinear and its evolution depends on state vector as well as the product of the state vector by itself. The measurements system of our model is linear. This bilinear model is a generalization of some other famous models, such as the Lorenz-96 model [41,42,43] and the Lotka-Volterra model [3,20,61] which are widely used in the study of atmospheric dynamics, ecological systems, chemistry, economic, neural network and control systems.

Four: Since our new model is not linear, the traditional Kalman filter and Kalman smoother are applied to such a model. So, we will derive a new filtering and a new smoothing procedure for our bilinear model. We will call the new filter "the bilinear

Kalman filter", and the new smoother "the bilinear Kalman smoother".

Five: To identify the parameters of such a bilinear model. The identification process will be achieved by using the expectation maximization algorithm (EM), where, in the expectation step, we use new approaches completely and totally different, to find the bilinear Kalman filter and the bilinear Kalman smoother to compute the estimation in the process of the nonlinear terms. This approach is new. In the maximization (M) step, we apply some theorems, lemmas and results from linear algebra and linear and nonlinear optimization theory.

Six: Since the applications are an important procedure to illustrate the effectiveness of the theorems and the results, we will apply our new theorems and our new results to a famous nonlinear model. This model in known by the Lotka-Volterra model which has applications in many fields in real life.

1.3 Outline of the dissertation

This dissertation is organized as follows: In Chapter 2, we introduce some basic tools which are to be used throughout this dissertation, as well as, basic concepts in statistics. In addition, methods for estimation and identification such as the least squares estimation, maximum likelihood estimation and the expectation maximization

algorithm will be introduced. We also present the definitions of the state-space model, type of estimators and Kronecker product. Definitions and theorems from linear algebra and the optimization theory which are used to prove and derive our results and theorems in this dissertation are introduced in this chapter.

In Chapter 3, we focus on the study of the Kalman filter and the Kalman smoother for a discrete linear, Gaussian model in the state space form. For this reason, we introduce the derivation of the equations of the Kalman filter by using two different approaches. First, by using least squares estimation and then, by using Baysian estimation to derive the same result. Since, the Kalman smoother gives more accurate results for the estimation of the same problem, we present the derivation of the Kalman smoother by using a famous technique [4].

Since the identification of parameters of such linear, Gaussian state space models is an important ingredient in estimation theory, we present, in Chapter 4, the identification process by using the well-known expectation maximization algorithm (EM), where we apply the Kalman filter and the Kalman smoother within the procedure of parameters estimation.

In Chapter 5, we present a new nonlinear, Gaussian mathematical state space model of bilinear class. This new model generalizes other corresponding famous models. Since our model is nonlinear, we present a new technique to derive what we call the "bilinear Kalman filter" and the "bilinear Kalman smoother".

In Chapter 6, we identify the parameters of our bilinear, Gaussian state space model using the expectation maximization algorithm, and we explain the methodology to be used. Furthermore, we apply the bilinear Kalman filter and the bilinear Kalman smoother to estimate the parameters in a new approach.

In Chapter 7, we introduce the applications of our results and theorems, where we apply the bilinear Kalman filter and the bilinear Kalman smoother to the well-known nonlinear Lotka-Volterra model. We also apply the new identification process to estimate the parameters of the nonlinear lotka-Volterra model.

Chapter 2

PRELIMINARIES

In this chapter, we briefly present the main background material needed throughout this dissertation. Some proofs of the known results are omitted, since they are available in standard books. In Section 2.1, we introduce some basic definitions and facts of statistical concepts. In section 2.2, we state the notion of the state space model. In section 2.3, we present the method of least square estimation for the linear system. In Section 2.4, we present the maximum likelihood identification and its algorithm. In Section 2.5, we introduce the expectation maximization (EM) algorithm. In Section 2.6, the common types of the estimators will be defined. In Section 2.7, we introduce the definition and properties of Kronecker product. In Section 2.8, we introduce

the differentiation rules which are used to derive some results and theorems in this dissertation.

2.1 Statistical concepts

Here, we introduce some statistical concepts which are used throughout this dissertation. These concepts can found in many statistics and time series analysis books, such as [23,27,30,36,45,52,58].

2.1.1 Random variables

A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment. There are two types of random variables, discrete and continuous.

A discrete random variable can only take a countable number of values. However, a continuous random variable is a random variable with continuous set of real numbers for its range.

2.1.2 Distribution functions

The distribution function F of a random variable X is defined by

$$F(x) = P[X \leq x],$$

for the real number x. Most of the commonly encountered distribution functions F can be expressed either as

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{2.1}$$

or

$$F(x) = \sum_{x_j \leq x} p(x_j) \tag{2.2}$$

where $\{x_0, x_1, \cdots\}$ is a finite or countably infinite set of real numbers.

In the case (2.1) we shall say that the random variable X is continuous, the function f is called the probability density function (pdf) of X, and can be found from the relation

$$f(x) = \frac{d}{dx}F(x).$$

While in the second case (2.2), the possible values of X are restricted to the set $\{x_0, x_1, \cdots\}$, and we shall say that the random variable X is discrete. The function f is called the probability mass function (pmf) of X.

The distribution functions have the following properties:

1. F is nondecreasing, i.e. $F(x) \leq F(y)$ if $x \leq y$

2. $F(x) = 1$ and $F(y) = 0$ for $x \to \infty$ and $y \to -\infty$, respectively.

2.1.3 Expectation, mean and variance

The expectation of a function g of a random variable X is defined by

$$E(g(X)) = \int g(x)dF(x)$$

where

$$\int g(x)dF(x) = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{in the continuous case} \\ \sum_{j=0}^{\infty} g(x_j)p(x_j), & \text{in the discrete case} \end{cases} \tag{2.3}$$

and g is any function such that $E(|g(X)|) < \infty$. The mean of X is defined as

$$\mu = E(X),$$

which is evaluated by setting $g(x) = x$ in the definition of $E(g(X))$. The variance of X is defined as

$$\sigma^2 = E(X - \mu)^2$$

which is evaluated by setting $g(x) = (x - \mu)^2$ in the definition of $E(g(X))$. It is clear from the definition that expectation has the linearity property

$$E(aX + b) = aE(X) + b$$

for any real constants a and b (provided that $E(X) < \infty$).

2.1.4 Random vectors

An n-dimensional random vector is a column vector $\mathbf{X} = (X_1, \cdots, X_n)^T$ each of whose components is a random variable. The distribution function F of $\mathbf{X}$, also called the **joint distribution** of $X_1, \cdots, X_n$, is defined by

$$F(x_1, \cdots, x_n) = P[X_1 \leq x_1, \cdots, X_n \leq x_n] \tag{2.4}$$

for all real numbers $x_1, \cdots, x_n$. This can be expressed in a more compact form as

$$F(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}], \quad \mathbf{x} = (x_1, \cdots, x_n)^T$$

for all real vector $\mathbf{x} = (x_1, \cdots, x_n)^T$. A random vector with distribution function F is continuous if F has a density function, i.e.

$$F(x_1, ..., x_n) = \int_{-\infty}^{x_n} ... \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(y_1, ..., y_n) dy_1 .... dy_n. \tag{2.5}$$

The probability density of $\mathbf{X}$ is then found from

$$f(x_1, \cdots, x_n) = \frac{\partial^n F(x_1, \cdots, x_n)}{\partial x_1, \cdots \partial x_n}.$$

The random vector $\mathbf{X}$ is said to be discrete if there exist real- valued vectors $\mathbf{x_0}, \mathbf{x_1}, \cdots$ and a probability mass function $p(\mathbf{x}_j) = P[\mathbf{X} = \mathbf{x}_j]$ such that

$$\sum_{j=0}^{\infty} p(\mathbf{x}_j) = 1.$$

The expectation of a function g of a random vector $\mathbf{X}$ is defined by

$$E(g(\mathbf{X})) = \int g(\mathbf{x})dF(\mathbf{x}) = \int g(x_1, \cdots, x_n)dF(x_1, \cdots, x_n),$$

where

$$\begin{aligned} & \int g(x_1, \cdots, x_n)dF(x_1, \cdots, x_n) \qquad (2.6) \\ = & \begin{cases} \int \cdots \int g(x_1, \cdots, x_n)f(x_1, \cdots, x_n)dx_1 \cdots dx_n, & \text{in the continuous case,} \\ \sum_{j_1} \cdots \sum_{j_n} g(x_{j_1}, \cdots, x_{j_n})p(x_{j_1}, \cdots, x_{j_n}), & \text{in the discrete case.} \end{cases} \end{aligned}$$

and g is any function such that $E|g(\mathbf{X})| < \infty$.

Definition 2.1.4.1. The random variables $X_1, \cdots, X_n$ are said to be independent if

$$P[X_1 \leq x_1, \cdots, X_n \leq x_n] = P[X_1 \leq x_1] \cdots P[X_n \leq x_n],$$

i.e.,

$$F(x_1, \cdots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Definition 2.1.4.2. (Conditional Expectation) For two random vectors $\mathbf{X} = (X_1, \cdots, X_n)^T$ and $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$ with joint density function $f_{\mathbf{X},\mathbf{Y}}$, the conditional density of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ is

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{Y}) = \begin{cases} \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}, & f_{\mathbf{Y}}(\mathbf{y}) > 0 \\ f_{\mathbf{X}}(\mathbf{x}), & f_{\mathbf{Y}}(\mathbf{y}) = 0. \end{cases} \qquad (2.7)$$

Since,

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})\mathbf{f}_{\mathbf{X}}(\mathbf{x}),$$

we have

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{Y}) = \frac{\mathbf{f}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})\mathbf{f}_{\mathbf{X}}(\mathbf{x})}{\mathbf{f}_{\mathbf{Y}}(\mathbf{y})}. \tag{2.8}$$

Equation (2.8) is called Bayes' rule.

The conditional expectation of $g(\mathbf{X})$ given $\mathbf{Y} = \mathbf{y}$ is then

$$E\{g(\mathbf{X})|\mathbf{Y} = \mathbf{y}\} = \int_{-\infty}^{\infty} g(\mathbf{x}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) d\mathbf{x}.$$

If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $f_{\mathbf{y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X}) = f_{\mathbf{y}}(y)$ and so the conditional expectation of $g(\mathbf{Y})$ given $\mathbf{X} = \mathbf{x}$ is

$$E(g(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = E(g(\mathbf{Y})),$$

which is as expected, does not depend on $\mathbf{x}$. The same ideas hold in the discrete case with the probability mass function assuming the role of the density function.

Definition 2.1.4.3. If $E|X_i| < \infty$ for each i, then we define the mean or expected value of a random vector $\mathbf{X} = (X_1, \cdots, X_n)^T$ to be the column vector

$$\mu_{\mathbf{X}} = E\mathbf{X} = (EX_1, \cdots, EX_n)^T.$$

In the same way we define the expected value of any array whose elements are random variables (e.g., a matrix of random variables) to be the same array with each random variable replaced by its expected value (if the expectation exists).

Definition 2.1.4.4. If $\mathbf{X} = (X_1, \cdots, X_n)^T$ and $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$ are random vectors such that each X_i and Y_j has a finite variance, then the **covariance matrix** of $\mathbf{X}$ and $\mathbf{Y}$ is defined to be the matrix

$$\begin{aligned} \Sigma_{\mathbf{XY}} &= \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = E[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T] \\ &= E(\mathbf{XY}) - (E\mathbf{X})(E\mathbf{Y})^T. \end{aligned}$$

The (i, j) element of $\Sigma_{\mathbf{XY}}$ is the covariance

$$\mathrm{Cov}(X_i, Y_j) = E(X_i, Y_j) = E(X_i Y_j) - E(X_i)E(Y_j).$$

In the special case where $\mathbf{Y} = \mathbf{X}$, $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})$ reduces to the covariance matrix of the random vector $\mathbf{X}$.

Lemma 2.1.4.5. Suppose that $\mathbf{Y}$ and $\mathbf{X}$ are linearly related through the equation

$$\mathbf{Y} = a + B\mathbf{X},$$

where a is an m-dimensional column vector and B is an $m \times n$ matrix. Then $\mathbf{Y}$ has mean

$$E\mathbf{Y} = a + BE\mathbf{X} \tag{2.9}$$

and covariance matrix

$$\Sigma_{\mathbf{YY}} = B\Sigma_{\mathbf{XX}}B^T. \tag{2.10}$$

Proposition 2.1.4.6. The covariance matrix $\Sigma_{\mathbf{XX}}$ of a random vector $\mathbf{X}$ is symmetric and nonnegative definite, that is, $\mathbf{b}^T\Sigma_{\mathbf{XX}}\mathbf{b} \geq 0$ for all vectors $\mathbf{b} = (b_1, \cdots, b_n)^T$ with real components.

Definition 2.1.4.7. The correlation between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined by

$$\text{Corr}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}\mathbf{Y}^T).$$

If $\text{Corr}(\mathbf{X}, \mathbf{Y}) = 0$, then, we say $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated.

2.1.5 Gaussian distribution

Gaussian distribution is the most common and useful distribution which is also referred to as the normal distribution. It is defined by its mean and variance and has a bell shaped or Gaussian form. It represents a family of distributions of the same general form, characterized by their mean μ and the variance σ^2. The standard normal distribution is a normal (Gaussian) distribution with a mean of zero and a variance of one.

The pdf of the normal (Gaussian) distribution is of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

2.1.6 Multivariate Gaussian distribution

Let $\mathbf{X} = (X_1, \cdots, X_n)^T$ be a random vector, then $\mathbf{X}$ has a multivariate Gaussian, or multivariate normal distribution if its pdf takes the form

$$f(\mathbf{X}) = \frac{\mathbf{1}}{(\mathbf{2}\pi)^{\mathbf{n}}|\mathbf{\Sigma}|} \exp[(-\mathbf{1/2})(\mathbf{X}-\mu)^{\mathbf{T}}\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{X}-\mu)],$$

where μ is the mean, Σ is the covariance, and $|\Sigma|$ is the determinant.

Next, we state an important lemma which is used to derive some results in this dissertation [27].

Lemma 2.1.6.1. If $\mathbf{Y}$ and $\mathbf{X}$ are jointly normally (Gaussian) distributed, then, the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ is also normally distributed with,

$$E(\mathbf{Y}|\mathbf{X}) = E\mathbf{Y} + \mathrm{Cov}(\mathbf{Y}, \mathbf{X})\mathrm{Var}^{-1}(\mathbf{X})(\mathbf{X} - \mathbf{E}(\mathbf{X}))$$

$$\mathrm{Var}(\mathbf{Y}|\mathbf{X}) = \mathrm{Var}(\mathbf{Y}) - \mathrm{Cov}(\mathbf{Y}, \mathbf{X})\mathrm{Var}^{-1}(\mathbf{X})\mathrm{Cov}(\mathbf{X}, \mathbf{Y})$$

2.1.7 Markov chain

A Markov chain is a collection of random variables $\{x_k\}, k = 0, 1, \cdots$, having the property

$$p(x_{k+1}|x_k, x_{k-1}, \cdots, x_1) = p(x_{k+1}|x_k)$$

Basically, if we know x_k, then, knowledge of $x_{k-1}, \cdots, x_1$ does not give any more information about x_{k+1}.

2.2 State-space model

The state space model is a mathematical model of a system as a set of input, output and state variables related by differential (difference) equations.

The state space model consists of two systems, the dynamical system, and the measurement system.

A dynamical system is a concept in mathematics that describes a process that includes a set of possible states, and a rule that determines the present state in terms of past states. The measurements system describes how the measurements of the dynamical system are related to the states.

In this dissertation, we consider a discrete state-space model of the form:

$$x_{k+1} = Ax_k + w_k \tag{2.11}$$

$$y_k = Cx_k + v_k, \tag{2.12}$$

for $k = 0, 1, 2, \cdots$, such that, Eq. (2.11) is the dynamical system, where, $x_k \in \mathbb{R}^n$ is the state vectors at time k,

$A \in \mathbb{R}^{n \times n}$ is the transition operator, which maps a state space into itself, and $w_k \in \mathbb{R}^n$ is the dynamical noise.

While, Eq. (2.12) is the measurement system, where $y_k \in \mathbb{R}^{p \times n}$ is the measurement vector at time k,

$C \in \mathbb{R}^{p \times 1}$ is the measurement operator, which maps the state space into a measurement space. $v_k \in \mathbb{R}^p$ is the measurements noise.

2.3 Least square estimation

The first method for forming an optimal estimate from noisy data is the method of least square. Its introduction is generally attributed to Gauss in 1795 [16,31].

Gauss discovered that if he wrote a system of equations in matrix form, as

$$\begin{pmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \tag{2.13}$$

or

$$Hx = z,$$

then, he could consider the problem of solving for that value of an estimate $\hat{x}$ that minimizes the estimation error $H\hat{x} - z$. He defined the estimation error in terms of the Euclidean vector norm

$$||H\hat{x} - z||,$$

or, its square,

$$\begin{aligned} e^2(\hat{x}) &= ||H\hat{x} - z||^2 \\ &= \sum_{i=1}^{m} \left[\sum_{j=1}^{n} h_{ij}\hat{x}_j - z_i \right]^2, \end{aligned}$$

which is a continuously differentiable function of the n unknowns $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$.

This function attains its minimum value when all its partial derivatives are zero. There are n such equations

$$0 = \frac{\partial e^2}{\partial \hat{x}_k} = 2\sum_{i=1}^{m} h_{i_k}\left[\sum_{j=1}^{n} h_{ij}\hat{x}_j - z_i\right], \quad k = 1, 2, \ldots$$

Thus, we have

$$0 = 2H^T[H\hat{x} - z] = 2H^T H\hat{x} - 2H^T z.$$

Therefore

$$H^T H\hat{x} = H^T z, \tag{2.14}$$

Eq. (2.14) is called the normal equation of the linear least square problem.

The solution of the normal equation is

$$\hat{x} = (H^T H)^{-1} H^T z,$$

provided that $H^T H$ is nonsingular. $H^T H$ is called the Gramain matrix. If $|HH^T| = 0$, then the column vectors are linearly dependent, and $\hat{x}$ can not be determined uniquely.

If X is a random variable [31], consider

$$y = Hx + v,$$

where, y is a given $n \times 1$ vector, H is a given $m \times n \quad (m \geq n)$, $\quad x$ is an unknown $n \times 1$ vector, and v is a $n \times 1$ vector. A least square solution $\hat{x}$ is one that minimizes

the length of the residual vector v; i.e., it is one with the property that

$$||y - H\hat{x}||^2 \leq ||y - Hx||^2,$$

for all $x \in \mathbb{R}^n$, where $||.||$ as before, is the Euclidean norm

$$||v||^2 = v^T v = \sum_{i=1}^{n} |v(i)|^2.$$

Here $v(i)$ stands for the i-th entry of the vector v. Now, let the cost function $J(x)$ be defined as

$$\begin{aligned} J(x) &= ||y - Hx||^2 \\ &= (y - Hx)^T (y - Hx) \\ &= y^T y - y^T Hx - x^T H^T y + x^T H^T Hx. \end{aligned}$$

Now,

$$\begin{aligned} 0 = \frac{\partial}{\partial x} J(x)|_{x=\hat{x}} &= \frac{\partial}{\partial x} [y^T y - y^T Hx - x^T H^T y + x^T H^T Hx]|_{x=\hat{x}} \\ &= -y^T H + \hat{x}^T H^T H, \end{aligned}$$

which shows that every solution $\hat{x}$ must satisfy the normal equations

$$H^T H\hat{x} = H^T y.$$

We can see that $\hat{x}$ minimizes the cost function $J(x)$ by noting that

$$\frac{\partial^2 ||y - Hx||^2}{\partial x^T \partial x} = H^T H \geq 0,$$

since $H^T H$ is positive-semi-definite. The value of J at $\hat{x}$ can be expressed as

$$\begin{aligned} J(\hat{x}) &= ||y - H\hat{x})|| \\ &= (y - H\hat{x})^T (y - H\hat{x}) \\ &= y^T (y - H\hat{x}) - \hat{x}^T H^T (y - H\hat{x}). \end{aligned}$$

since $H^T(y - H\hat{x}) = 0$, by the normal equation, we get,

$$\begin{aligned} J(\hat{x}) &= y^T (y - H\hat{x}) \\ &= y^T y - y^T H\hat{x} \\ &= ||y||^2 - (H^T y)^T \hat{x} \\ &= ||y||^2 - (H^T H\hat{x})^T \hat{x} \\ &= ||y||^2 - ||H\hat{x}||^2. \end{aligned}$$

When the matrix H has full rank n, the matrix $H^T H$ will be nonsingular, and there is a unique solution $\hat{x}$ given by

$$\hat{x} = (H^T H)^{-1} H^T y.$$

2.4 Maximum likelihood estimation

Maximum likelihood estimation is a popular statistical method used for fitting a set of mathematical models to measured data, and providing estimates for the model pa-

rameters (identification) [23,47,52,58].

Generally, the method of maximum likelihood identifies values of the parameters that produce the distribution most likely to have resulted in the observed measurements. That is, the parameters that maximize the likelihood function. The maximum likelihood methodology is described as follows:

Suppose $X = (x_1, x_2, ..., x_n)$ is a vector of independent random variables with probability density function:-

$$p(X, \theta_1, \theta_2, \ldots, \theta_k),$$

where $\theta_1, \theta_2, \ldots, \theta_k$ are k unknown parameters which need to be estimated.

Then, the likelihood function of the parameters is given by the following product:

$$L(x_1, \ldots, x_n | \theta_1, \ldots, \theta_k) = L = \prod_{i=1}^{n} p(x_i; \theta_1, \ldots, \theta_k).$$

Now, taking the logarithm of the likelihood function,

$$\ln L = \sum_{i=1}^{n} \ln p(x_i; \theta_1, \ldots, \theta_k i).$$

The maximum likelihood estimators of $\theta_1, \theta_2, \ldots, \theta_k$ are obtained by maximizing L or $\ln L$. But, maximizing $\ln L$ is easier to work with than L. So, the maximum likelihood estimators of the parameters $\theta_1, \ldots, \theta_k$ are the simultaneous solutions of k equations

$$\frac{\partial \ln L}{\partial \theta_m} = 0, \quad m = 1, \ldots, k.$$

2.5 The Expectation maximization (EM) algorithm

The Expectation-Maximization (EM) algorithm is an iterative procedure for computing the maximum likelihood estimator when only a subset of the complete data set is available [47,52]. Dempster, Laird and Rubin [9] demonstrated the wide applicability of the EM algorithm and are largely responsible for popularizing this method in statistics.

In the usual formulation of the EM algorithm, the "complete" data vector $\mathbf{W}$ is made up of "observed" data $\mathbf{Y}$ (sometimes called incomplete data) and " unobserved" data $\mathbf{X}$. In many applications, $\mathbf{X}$ consists of values of a " latent" or unobserved process occurring in the specification of the model. For example, in the state-space model

$$x_{k+1} = Ax_k + w_k$$

$$y_k = Cx_k + v_k,$$

for $k = 0, 1, 2, \cdots$. Thus, $\mathbf{Y}$ could consist of the observed vectors $\{y_1, y_2, \cdots, y_t\}$ and $\mathbf{X}$ of the unobserved state vectors $\{x_1, x_2, \cdots, x_t\}$. The parameters of the model are A, C, Q, R, and V, where, Q and R are the covariance of the noises w_k and v_k respectively, and μ and V are the mean and the variance, respectively, of the initial state x_0. The EM algorithm provides an iterative procedure for computing the maximum likelihood estimator based only on the observed data $\mathbf{Y}$. Each iteration of the EM

algorithm consists of two steps. If $\theta(i)$ denotes the estimated value of the parameter vector

$$\theta = \{A, C, Q, R, \mu, V\}$$

after i iterations, then the two steps in the $(i+1)$th iteration are

E-step. Calculate $\psi(\theta, \theta(i)) = E_{\theta^{(i)}}[\log L(\theta; X, Y|Y)]$ where, $L(\theta; X, Y|Y)$ is the maximum likelihood function, and

M-step. Maximize $\psi(\theta, \theta(i))$ with respect to θ.

2.6 Filter, prediction and smoother

Generally, there are three types of estimators [16,58]:-

- Predictors: The measurements used are strictly prior to the time that the state of the dynamical system is to be estimated i.e.,

$$k_{\text{meas.}} < k_{\text{est.}}.$$

- Filters: In this type of estimators, the measurements used are up to and including the time that the state of the dynamical system is to be estimated. i.e.,

$$k_{\text{meas.}} \leq k_{\text{est.}}.$$

- Smoothers: This type uses the measurements beyond the time that the state of the dynamical system is to be estimated. i.e.,

$$k_{\text{meas.}} > k_{\text{est.}}$$

2.7 Kronecker product

In mathematics, the Kronecker product, denoted by $\otimes$, is an operation on two matrices of arbitrary size resulting in a block matrix. The Kronecker product should not be confused with the usual matrix multiplication, which is an entirely different operation. It is named after German mathematician Kronecker [21]. In this dissertation, we will use the notion of the Kronecker product to introduce a new mathematical model, where the utility of the Kronecker product makes our new model more general due the bilinearity property.

In the following definition we define the Kronecker product.

Definition 2.7.1. Let $A \in R^{m\times n}, B \in R^{p\times q}$. The Kronecker product (or tensor

product) of A and B is defined as the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \vdots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}$$

2.8 Differentiation rules

In this section, we introduce useful definitions, lemmas, theorems and properties about the differentiation of matrix expressions. These results are used throughout this dissertation to derive and prove some theorems.

Definition 2.8.1. Let $F : X \to Y$, where X, Y finite dimensional spaces with norm $\|\|$. Let $x \in X$. F is differentiable at x if there is a bounded linear operator , denoted $F'(x)$ or $\frac{dF(x)}{dx}$, such that,

$$F(x+h) - F(x) = F'(x)h + O(\|h\|)$$

for all $h \in X$. In other words $F'(x) \in \mathcal{L}(X, Y)$.

Remark 2.8.2. To compute $F'(x)$ we extract the linear part of $F'(x+h)$, or equivalently, compute

$$\frac{\partial F(x+th)}{\partial t}|_{t=0} \tag{2.15}$$

Theorem 2.8.3. (Cain Rule) Let $G : X \to Y$, be differentiable at x, and $F : Y \to Z$, be differentiable at $y = G(x)$. Then, $F(G(x)) = F \circ G : X \to Z$, is differentiable at x, and,

$$\frac{d}{dx}F \circ G(x) = \frac{d}{dy}F(y)\frac{d}{dx}G(x). \tag{2.16}$$

We sometimes write Eq.(2.16) as

$$F(G(x))' = F'(G(x))G'(x).$$

Note that, $F' \circ G \in \mathcal{L}(Y, Z)$ and $G' \in \mathcal{L}(X, Z)$.

In the following lemma, we present results from Linear Algebra.

Lemma 2.8.4. Let A and B be two $n \times n$ matrices.

1) $\mathrm{Tr}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}$.

2) $x^T Ax = \mathrm{Tr}(xx^T A^T)$

Proof. 1) Let $C = AB$, then

$$C_{ii} = \sum_{j=1}^{n} a_{ij}b_{ji}.$$

Therefore

$$\mathrm{Tr(AB)} = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}.$$

2) The second part can be proved as follows.

$$x^T Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j = \sum_{i,j=1}^{n} (xx^T)_{ij}a_{ij} = \mathrm{Tr}(xx^T A^T).$$

2.8.1 Derivatives with respect to vectors and operators

Here, we state a theorem which is used to compute the derivatives with respect to vectors and operators.

Theorem 2.8.1.1. Let x, h be vectors, and A, M be matrices.

1) $\frac{d}{dx}(Ax)h = Ah$.

2) $\frac{d}{dA^{-1}}A^{-1}M = -A^{-1}MA^{-1}$

3) $\frac{d}{dA}|A|M = \mathrm{Tr}(\mathrm{adj}(A)M^T), \quad M \in \mathbb{R}^{n\times n}$

4) $\frac{d}{dA}(x^TAx)M = x^TMx$.

5) $\frac{d}{dA}(\mathrm{Tr}(A))M = \mathrm{Tr}M$.

Proof. 1) We compute that,

$$A(x+th) = Ax + Ath,$$

then,

$$\frac{d}{dt}A(x+th)|_{t=0} = Ah. \tag{2.17}$$

2) We know that,

$$(A+tM)^{-1} = [A(I+tA^{-1}M)]^{-1} = (I+tA^{-1}M)^{-1}A^{-1}.$$

Using Neumann formula gives

$$(A+tM)^{-1} = A^{-1} - tA^{-1}MA^{-1} + O(t^2).$$

Therefore,

$$\frac{d}{dt}(A+tM)^{-1}|_{t=0} = -A^{-1}MA^{-1}.$$

3) We will use the fact that $|A|$ is an n-linear form when regarded as a function of the columns of A. For $A, M \in \mathbb{R}^{n\times n}$, denote by $a_i, M_i, 1 \leq i \leq n$ the columns of A, M, respectively.

$$\begin{aligned} |A+tM| &= |a_1 + tM_1, a_2 + tM_2, ..., a_n + tM_n| \\ &= |a_1, a_2, ..., a_n| + t\sum_{i=1}^{n} |a_1, ..., M_i, a_{i+1}, ..., a_n| + O(t^2) \\ &= |A| + t\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}A_{ij} + O(t^2), \end{aligned}$$

where A_{ij} is the cofactor of a_{ij} in A. Therefore,

$$\frac{d}{dt}|A+tM|_{t=0} = \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}A_{ij} = \mathrm{Tr}(\mathrm{adj}AM^T)$$

4) First, we compute

$$(x^T(A+tM)x) = (x^TAx + tx^TMx)$$

Then, we get

$$\frac{d}{dt}(x^T(A+tM)x)|_{t=0} = (x^TMx).$$

5) Notice that $Tr : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$ is a linear operator. Thus,

$$\mathrm{Tr}(A + th) = \mathrm{Tr}A + t\mathrm{Tr}h.$$

So,

$$\frac{d}{dt}\mathrm{Tr}(A + th)|_{t=0} = \mathrm{Tr}h.$$

Corollary 2.8.1.2.

$$\frac{d}{dt}\log|A|M = \frac{1}{|A|}\mathrm{Tr}(\mathrm{adj}AM^T) = \mathrm{Tr}(A^{-1}M^T).$$

2.8.2 Derivatives of bilinear operators

Here, we present an important result about derivatives of bilinear operators. First, we define the bilinear operator as follows.

Definition 2.8.2.1. A bilinear operator is a mapping $\mathcal{F} : X \times Y \longrightarrow Z$ such that

$$\mathcal{F}(ax, y) = a\mathcal{F}(x, y),$$

$$\mathcal{F}(x, by) = b\mathcal{F}(x, y),$$

$$\mathcal{F}(x + y, z) = \mathcal{F}(x, z) + \mathcal{F}(y, z)$$

$$\mathcal{F}(x, y + z) = \mathcal{F}(x, y) + \mathcal{F}(x, z).$$

For example, the Kronecker product of two $n \times n$ matrices is a bilinear operator with $Y = X = \mathbb{R}^{n \times n}$ and $Z = \mathbb{R}^{n^2 \times n^2}$.

In the following lemma we introduce an expression for the derivative of bilinear operators.

Theorem 2.8.2.2. Let $\mathcal{F} : X \times X \longrightarrow Y$ be a bilinear operator, then,

$$\frac{d}{dx}\mathcal{F}(x,x)(a,b) = \mathcal{F}(a,x) + \mathcal{F}(x,b).$$

Proof. We compute,

$$\mathcal{F}((x,x) + t(a,b)) = \mathcal{F}(x + ta, x + tb) = \mathcal{F}(x,x) + t\mathcal{F}(a,x) + t\mathcal{F}(x,b) + t^2\mathcal{F}(a,b).$$

Now,

$$\frac{d}{dt}\mathcal{F}(x,x)(a,b)|_{t=0} = \mathcal{F}(a,x) + \mathcal{F}(x,b).$$

As a first corollary of this theorem we have the derivative of the Kronecker product.

Corollary 2.8.2.3.

$$\frac{d}{dx}(x \otimes x)v = (v \otimes x) + (x \otimes v).$$

Proof. In Theorem 2.8.2.2 put $\mathcal{F}(x,y) = x \otimes y$. Then,

$$\mathcal{F}'(x,x)(v,v) = \mathcal{F}(x,v) + \mathcal{F}(v,x) = (x \otimes v) + (v \otimes x).$$

As a second corollary we have,

Corollary 2.8.2.4. Let A be a symmetric matrix. Then,

$$\frac{d}{dx}(x^TAx)v = v^TAx + x^TAv = 2v^TAx. \tag{2.18}$$

Proof. In Theorem 2.8.2.2 put $\mathcal{F}(x,y) = x^TAy$. Then,

$$\begin{aligned}
\mathcal{F}'(x,x)(v,v) &= \mathcal{F}(x,v) + \mathcal{F}(v,x) \\
&= x^TAv + v^TAx \\
&= v^TA^Tx + v^TAx \\
&= v^T(A + A^T)x \\
&= 2v^TAx,
\end{aligned}$$

since A is symmetric.

The following results are stated without proof. The proof can be provided using the foregoing theorems and lemmas.

Lemma 2.8.2.5. Let $A, B, M \in \mathbb{R}^{n\times n}$, then

$$\frac{d}{dA}(A^{-1}B)M = -A^{-1}MA^{-1}B,$$

$$\frac{d}{dA}\mathrm{Tr}(A^{-1}B)M = \mathrm{Tr}(-A^{-1}MA^{-1}B) = -\mathrm{Tr}(A^{-1}MA^{-1}B),$$

$$\frac{d}{dA}(A^{-1}BA)M = A^{-1}BM - A^{-1}MA^{-1}BA$$

2.8.3 Computation of critical points

Here, we give some examples, which are used later as building blocks, for computing critical points.

Example 2.8.3.1. Let

$$F(x) = A + (x - M)^T B(x - M),$$

where B is symmetric. We can compute the critical point of F as follows: Since

$$\frac{d}{dx}(F(x))v = 2v^T B(x - M),$$

it follows that,

$$2v^T B(x - M) = 0, \quad \forall v \in \mathbb{R}^n.$$

Thus,

$$B(x - M) = 0,$$

this means that, $x - M \in \ker B$, if B is invertible, then, $x - M = 0$ or $x = M$.

Example 2.8.3.2. Let,

$$F(A) = \log |A| + \mathrm{Tr}(AP) + x^T Ax,$$

where, P is positive definite. Then, the procedure for computing the critical point of

this function is as follows,

$$\begin{aligned}\frac{d}{dA}F(A)M &= \mathrm{Tr}(A^{-1}M^T) + \mathrm{Tr}(MP) + x^T Mx \\ &= \mathrm{Tr}(A^{-1}M^T + P^T M^T + xx^T M^T) \\ &= \mathrm{Tr}([A^{-1} + P^T + xx^T]M^T) \\ &= 0, \quad \forall M \in \mathbb{R}^{n\times n},\end{aligned}$$

This would be true if and only if $A^{-1} + P^T + xx^T = 0$. Therefore,

$$A = -(P^T + xx^T)^{-1}.$$

(Observe that $P^T + xx^T$ is positive definite and therefore has an inverse.)

Example 2.8.3.3. Let,

$$F(K) = \mathrm{Tr}(A + BK + CK^T + KDK^T),$$

where, D is positive definite and symmetric. Then,

$$\frac{dF}{dK} = \mathrm{Tr}(M(B + C^T + 2DK^T)) = 0, \tag{2.19}$$

for all $M \in \mathbb{R}^{n\times n}$, where we use the facts that $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$, $\mathrm{Tr(AB)} = \mathrm{Tr(A)Tr(B)}$ and the linearity of the Tr(.) operator.

Eq.(2.19) gives

$$B + C^T + 2DK^T = 0$$

or

$$K = -\frac{1}{2}(B^T + C)D^{-1}.$$

Chapter 3

KALMAN FILTER AND KALMAN SMOOTHER

3.1 Introduction

Kalman filter is an estimator for so-called linear-quadratic problem, which is the problem of estimating the instantaneous "state" of a linear dynamic system perturbed by white noise using measurements linearly related to the state but also corrupted by white noise. It produces values that tend to be closer to the true values of the measurements and their associated calculated values by predicting an estimate of

uncertainty of the predicted via a weighted average of the predicted and measured values. Theoretically, the Kalman filter is an algorithm for efficiently doing exact inference in a linear state space model which has some statistical properties. The resulting estimator is statistically optimal with respect to some quadratic function of the estimation error [16].

Mathematically, the Kalman filter is a set of equations that provides an efficient recursive solution of the least square method. It provides estimates of the past, present, and future states and it can do so when the precise nature of the system model is unknown [16,52,58].

The Kalman filter was introduced by Kalman in 1960 [32]. It can be derived by many techniques [62]. In some works, it was derived by minimizing a quadratic cost function, which is related to the least squares estimation. The other technique for deriving Kalman filter is by using Bayesian statistics [16,27,31,52,58,62,64].

The second sort of the estimator is smoother. A smoother estimates the state of a system at time k, using measurements made before and after time k [16]. The accuracy of a smoother is generally superior to that of a filter, because it uses more measurements for its estimate. The Kalman smoother can be derived from the Kalman filter model. The general derivation methodology uses the Kalman filter for measurements up to (each) time k that state is to be estimated, combined with another algorithm derived

from Kalman filter for the measurements beyond that time [4,52,58]. This second algorithm of smoother can be derived by running the Kalman filter backward from the last measurement to just past k, then optimally combining the two independent estimates (forward and backward) of the state at time k based on the two independent sets of measurements.

In Section 3.2, we present the Kalman filter and its derivation by using two different techniques as discussed above. In Section 3.3, we introduce the notion of the Kalman smoother and the method of deriving it. We also introduce the well-known lag-one covariance smoother and its derivation.

3.2 Kalman filter

In this section, we study the Kalman filter for a linear-discrete Gaussian state space model. To begin with, we will introduce the nonlinear Kalman filter.
Consider a discrete-time linear state space model with dynamical system

$$x_{k+1} = Ax_k + w_k \tag{3.1}$$

and measurements

$$y_k = Cx_k + v_k, \tag{3.2}$$

where $x_k \in \mathbb{R}^n$ is the system state vector at time k,

$A \in \mathbb{R}^{n\times n}$ is the transition matrix,

$y_k \in \mathbb{R}^p$ is the corresponding measurement vector at time k,

$C \in \mathbb{R}^{p\times n}$ is the observation matrix,

$w_k \in \mathbb{R}^n$ is the dynamical noise at time k and

$v_k \in \mathbb{R}^p$ is the observation noise at time k.

Indeed, w_k and v_k are uncorrelated, white and Gaussian with zero mean and covariance Q and R, respectively, that is,

$$w_k \sim N(0, Q),$$

$$v_k \sim N(0, R),$$

$$E(w_k w_l^T) = \begin{cases} Q, \text{for} & \text{k=l} \\ 0, \text{for} & \text{k}\neq\text{l}, \end{cases}$$

$$E(v_k v_l^T) = \begin{cases} R, \text{for} & \text{k=l} \\ 0, \text{for} & \text{k}\neq\text{l} \end{cases}$$

and

$$E(w_i v_j^T) = 0, \quad \forall i, j.$$

3.2.1 Derivation of Kalman filter using least-squares estimation

There are many approaches to derive the Kalman filter [62]. We will introduce two derivations. The first is based on the least square estimation [27] while the second is based on Bayes' rule [52,62]. Next, we will use the following notations [19]:

x_k^a : is the estimation of x_k.

x_k^f : is the forecast state of x_k,

e_k^a : is the analysis state error defined by

$$e_k^a = x_k - x_k^a,$$

e_k^f : is the forecast state error defined by

$$e_k^f = x_k - x_k^f,$$

P_k^a : is the analysis state error covariance,

P_k^f : is the forecast state error covariance.

In the following theorem, we will present the Kalman filter which gives the filtering and forecasting equations [16,32,52,58].

Theorem 3.2.1. For the state space model defined by Eq.(2.11) and Eq.(2.12), with

initial conditions $x_0 = \mu_0$ and $P_0 = V_0$, we have

$$x_{k+1}^f = Ax_k^a \tag{3.3}$$

$$P_{k+1}^f = AP_k^a A^T + Q \tag{3.4}$$

for $k = 0, \cdots, t$, with

$$x_{k+1}^a = x_{k+1}^f + K_{k+1}(y_k - Ax_{k+1}^f) \tag{3.5}$$

$$P_{k+1}^a = [I - K_{k+1}C]P_{k+1}^f \tag{3.6}$$

where

$$K_{k+1} = P_{k+1}^f C^T [CP_{k+1}^f C^T + R]^{-1} \tag{3.7}$$

is called Kalman gain.

Proof. Taking expectation of the dynamic equation $x_{k+1} = Ax_k + w_k$, we obtain

$$\overline{x}_{k+1} = A\overline{x}_k,$$

where

$$\overline{x}_k = E(x_k).$$

Since $\overline{x}_k$ in unknown, we will use x_x^a as an estimation. Thus

$$x_{k+1}^f = Ax_k^a.$$

The corresponding forecast state error covariance $P^f_{k+1} \in \mathbb{R}^{n\times n}$ is defined by:

$$\begin{aligned}
P^f_{k+1} &= E\left[e^f_k (e^f_k)^T\right] \\
&= E\left[(x_{k+1} - x^f_{k+1})(x_{k+1} - x^f_{k+1})^T\right] \\
&= E\left[(Ax_k + w_k - Ax^a_k)(Ax_k + w_k - Ax^a_k)^T\right] \\
&= E\left[A(x_k - x^a_k) + w_k)(A(x_k - x^a_k) + w_k)^T\right] \\
&= AE\left[(x_k - x^a_k)(x_k - x^a_k)\right] A^T + E(w_k w^T_k) \\
&= AP^a_k A^T + Q,
\end{aligned}$$

where x^a_k is the estimate of the state x_k and $P^a_k \in \mathbb{R}^{n\times n}$ is the analysis state error covariance given by

$$P^a_k = E\left[e^a_k (e^a_k)^T\right] = E\left[(x_k - x^a_x)(x_k - x^a_k)^T\right].$$

In the analysis step, the objective is to estimate the state x_k by x^a_k using measurements y_k. So we have an optimization problem of the form,

$$\text{minimize} \quad \text{Tr}(E\left[e^a_k (e^a_k)^T\right] = \text{Tr}(P^a_k).$$

On the other hand, when a new measurement y_k is available, the new information will be used to update the prior x^f_k to the analysis x^a_k. So, we need to find an optimal $n \times p$ weight matrix K_k, (called "Kalman gain"), so the analysis x^a_k is updated by

$$x^a_k = x^f_k + K_k(y_k - Cx^f_k), \tag{3.8}$$

which minimizes the expectation of the energy "cost function",

$$J_k = E||e_k^a||^2 = E||x_k^a - x_k||^2.$$

We expect the forecast, the analysis and the measurements to be unbiased estimations, that is,

$$E(e_k^f) = E(x_k - x_k^f) = 0,$$

$$E(e_k^a) = E(x_k - x_k^a) = 0,$$

and

$$E(v_k) = E(y - Cx_k) = 0.$$

Now write x_k^a as a linear combination of x_k^f and y_k,

$$x_k^a = Fx_k^f + Hy_k \tag{3.9}$$

where $F \in \mathbb{R}^{n\times n}$ and $H \in \mathbb{R}^{n\times p}$ are constant matrices. By using one of unbiased equations, we get

$$E(x_k^a - x_k) = E[Fx_k^f + Hy_k - x_k] = 0\,.$$

Thus,

$$E[Fx_k + HCx_k - x_k] = 0$$

and so,

$$(F + HC - I_n)E(x_k) = 0,$$

where I_n is the $n \times n$ identity matrix. Hence,

$$F = I_n - HC.$$

By substituting the above relation in Eq(3.9), we get

$$\begin{aligned} x_k^a &= (I_n - HC)x_k^f + Hy_k \\ &= x_k^f - HCx_k^f + Hy_k \\ &= x_k^f + H(y_k - Cx_k^f). \end{aligned}$$

Replacing H with K_k, we get

$$x_k^a = x_k^f + K_k(y - Cx_k^f).$$

Now, subtracting x_k from Eq(3.8) gives

$$x_k^a - x_k = (x_k^f - x_k) + K_k \left[(y_k - Cx_k) - (Cx_k^f - Cx_k) \right].$$

Thus,

$$e_k^a = e_k^f + K_k(v_k - Ce_k^f),$$

that is, we obtain the analysis error covariance in terms of forecast error covariance,

$$\begin{aligned} P_k^a &= E(e_k^a(e_k^a)^T) \\ &= E\left[(e_k^f + K_k(v_k - Ce_k^f))(e_k^f + K_k(v_k - Ce_k^f))^T \right] \\ &= E\left[(e_k^f + K_k(v_k - Ce_k^f))((e_k^f)^T + (v_k^T - (e_k^f)^T C^T)K_k^T) \right] \end{aligned}$$

$$
\begin{aligned}
&= E\big[e_k^f(e_k^f)^T + e_k^f(v_k^T + (e_k^f)^T C^T)K_k^T + K_k(v_k - Ce_k^f)(e_k^f)^T \\
&\quad + K_k(v_x - Ce_k^f)(v_k^T - (e_k^f)^T C^T)K_k^T\big] \\
&= E\big[e_k^f(e_k^f)^T + e_k^f v_k^T - e_k^f(e_k^f)^T C^T K_k^T + K_k v_k(e_k^f)^T - K_k C e_k^f(e_k^f)^T \\
&\quad + K_k(v_k v_k^T)K_k^T - K_k(v_k(e_k^f)^T C^T - Ce_k^f v_k^T)K_k^T + K_k C(e_k^f(e_k^f)^T)C^T K_k^T\big].
\end{aligned}
$$

Since x_k^f and v_k are independent,

$$E(e_k^f(v_k)^T) = E(v_k(e_k^f)^T) = 0.$$

Therefore,

$$
\begin{aligned}
P_k^a &= E(e_k^f(e_k^f)^T) - E(e_k^f(e_k^f)^T)C^T K_k^T - K_k C E(e_k^f(e_k^f)^T) \\
&\quad K_k E(v_k v_k^T)K_k^T + K_k C E(e_k^f(e_k^f)^T)C^T K_k^T.
\end{aligned} \tag{3.10}
$$

Because by definition the forecast error covariance is $P_k^f = E[(e_k^f(e_k^f)^T]$ and the measurement noise covariance is $R = E(v_k v_k^T)$, then Eq (3.10) amounts to

$$P_k^a = P_k^f - P_{K_k}^f C^T K_k^T - K_k C P_k^f + K_k R K_k^T + K_k C P_k^f C^T K_k^T. \tag{3.11}$$

Thus,

$$\mathrm{Tr}(P_k^a) = \mathrm{Tr}(P_k^f) - \mathrm{Tr}(P_k^f C^T K_k^T) - \mathrm{Tr}(K_k C P_k^f) + \mathrm{Tr}(K_k R K_k^T) + \mathrm{Tr}(K_k C P_k^f C^T K_k^T), \tag{3.12}$$

and hence, using $(K_k C P_k^f)^T = P_k^f C^T K_k^T$ and $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$, we have

$$\mathrm{Tr}(P_k^a) = \mathrm{Tr}(P_k^f) - 2\mathrm{Tr}(K_k C P_k^f) + \mathrm{Tr}(K_k R K_k^T) + \mathrm{Tr}(K_k C P_k^f C^T K_k^T). \tag{3.13}$$

To minimize $\mathrm{Tr}(P_k^a)$ with respect to K_k, we must have

$$\frac{\partial \mathrm{Tr}(P_k^a)}{\partial K_k} = 0.$$

Now, using Example 2.8.3.3, with $A = P_k^f$, $B = -CP_k^f$, and $D = R + CP_k^f C^T$, we get

$$K_k = -\frac{1}{2}(-2P_k^f C^T)(R + CP_k^f C^T)^{-1}$$

and consequently,

$$K_k = P_k^f C^T (R + CP_k^f C^T)^{-1}, \tag{3.14}$$

provided that $(R + CP_k^f C^T)$ is non-singular. Noting that,

$$\frac{\partial^2}{\partial K_k^2}\mathrm{Tr}(P_k^a) = 2(CP_k^f C^T + R),$$

which is positive definite, we conclude that $\mathrm{Tr}P_k^a$ attains its minimum at K. Now, by inserting Eq. (3.14) in Eq. (3.11) and after simple manipulations, we can show that

$$P_k^a = P_k^f - K_k C P_k^f.$$

3.2.2 Derivation of Kalman filter from Beyesian estimation

Here, Bayesian statistics and the Markov property of system (3.1) and (3.2) are utilized to prove Theorem 3.2.2.1. We use the fact that the linear dynamical system and

measurements are Gaussian. The multivariate Gaussian distribution will be then used as the probability density function. We aim in this filtering problem to determine $P(x_k|\{y\}_1^k)$ and find the mean and the variance of this distribution. Here, we use the following notations:

$$E_k(x_k) = E(x_k|\{y\}_1^k) = x_k^a$$

$$E_{k-1}(x_k) = E(x_k|\{y\}_1^{k-1}) = x_k^f$$

$$\text{Cov}(x_k|\{y\}_1^k) = P_k^a$$

and

$$\text{Cov}(x_k|\{y\}_1^{k-1}) = P_k^f.$$

where the sequence $\{y\}_1^k$ is the set of measurements given by

$$\{y\}_1^k = \{y_1, y_2, \cdots, y_k\}.$$

Consider the linear state space defined by Eq.(3.1) and Eq.(3.2) where w_k, v_k are assumed (as before) to be uncorrelated, white, Guassian with zero mean and covariance Q and R respectively.

Here, we will use the above notation to prove Theorem 3.2.2.1, such that the statistical assumptions of the state-space model, Bayes' rule, and the Markov property are employed to derive the Kalman filter equations, [52,62].

For the state-space model defined by Eq.(3.1) and Eq.(3.2), the probability densities

are

$$p(x_{k+1}|x_k) = N(x_{k+1} : Ax_k, Q) \tag{3.15}$$

and

$$p(y_k|x_k) = N(y_k : Cx_k, R) \tag{3.16}$$

where $N(x : \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. Next, we will find the filtering and prediction steps.

By applying Beyes' theorem,

$$p(x_k|\{y\}_1^k) = \frac{p(y_k|x_k)p(x_k|\{y\}_1^{k-1})}{p(y_k|\{y\}_1^{k-1})} \tag{3.17}$$

and

$$p(x_{k+1}|\{y\}_1^k) = \int p(x_k|\{y\}_1^k)p(x_{k+1}|x_k)dx_k. \tag{3.18}$$

The Equations (3.17) and (3.18) are both Gaussian [52]. Thus

$$p(x_k|\{y\}_1^k) = N(x_k : x_k^a, P_k^a)$$

and

$$p(x_{k+1}|\{y\}_1^k) = N(x_{k+1}, x_k^f, P_k^f).$$

Inserting (3.16) in (3.18) yields

$$p(x_{k+1}|\{y\}_1^k) = \int N(x_{k+1} : Ax_k, Q)N(x_k : x_k^a, P_k^a)dx_k. \tag{3.19}$$

Now,

$$\begin{aligned} x_{k+1}^f = E_k(x_{k+1}) &= \int_{-\infty}^{\infty} x_{k+1} p(x_{k+1}|\{y\}_1^k dx_{k+1} \\ &= \int_{-\infty}^{\infty} x_{k+1} \int_{-\infty}^{\infty} p(x_k|\{y\}_1^k) p(x_{k+1}|x_k) dx_k dx_{k+1} \\ &= \int_{-\infty}^{\infty} p(x_k|\{y\}_1^k) [\int_{-\infty}^{\infty} x_{k+1} p(x_{k+1}|x_k) dx_{k+1}] dx_k \\ &= \int_{-\infty}^{\infty} A x_k p(x_k|\{y\}_1^k dx_k \\ &= A x_k^a. \end{aligned}$$

We can also show that

$$P_{k+1}^f = A P_k^a A^T + Q.$$

That is, $p(x_{k+1}|\{y\}_1^k)$ follows a Gaussian distribution $N(x_{k+1} : x_{k+1}^f, p_{k+1}^f)$. In filtering step; we will use Beyes's rule (Eq. (3.17)) to update $p(x_{k+1}|\{y\}_1^k)$ to $p(x_{k+1}|\{y\}_1^{k+1})$ and thus,

$$\begin{aligned} p(x_{k+1}|\{y\}_1^{k+1}) &= p(x_{k+1}|y_{k+1}, \{y\}_1^k) \\ &= \frac{p(y_{k+1}|x_{k+1}, \{y\}_1^k) p(x_{k+1}|\{y\}_1^k)}{p(y_{k+1}|\{y\}_1^k)} \\ &= \frac{p(y_{k+1}|x_{k+1}) p(x_{k+1}|\{y\}_1^k)}{\int p(y_{k+1}|x_{k+1}) p(x_{k+1}|\{y\}_1^k) dx_x}. \end{aligned} \tag{3.20}$$

Using

$$p(y_{k+1}|x_{k+1}) = N(y_{k+1} : C x_{k+1}, R) \tag{3.21}$$

in (3.20), we get

$$p(x_{k+1}|\{y\}_1^{k+1}) = \frac{N(y_{k+1} : C x_{k+1}, R) N(x_{k+1} : x_k^f, P_k^f)}{\int N(y_{k+1} : C x_{k+1}, R) N(x_{k+1} : x_k^f, P_k^f) dx_k}. \tag{3.22}$$

Using this, we obtain

$$\int N(y_{k+1} : Cx_{k+1}, R)N(x_{k+1} : x_k^f, P_k^f)dx_k = N(y_{k+1} : Cx_k^f, CP_k^f C^T + R)$$

and so, Eq. (3.22) can be reduced to

$$\begin{aligned} p(x_{k+1}|\{y\}_1^{k+1}) &= \frac{N(y_{k+1} : Cx_{k+1}, R)N(x_{k+1} : x_k^f, P_k^f)}{N(y_{k+1} : Cx_k^f, CP_k^f C^T + R)} \\ &= N(x_{k+1} : x_k^a, P_k^a) \end{aligned}$$

where

$$x_k^a = x_k^f + K_k(y_k - Cx_k^f)$$

$$P_k^a = P_k^f - K_k C P_k^f$$

$$K_k = P_k^f C^T (CP_k^f C^T + R)^{-1}.$$

For a complete proof of these identities, one may refer to [62].

3.3 Kalman smoother

Here, we will consider the problem of obtaining x_k based on the data samples $y_1, \cdots, y_t$ where $k \leq t$; namely, x_k^t. These estimators are called smoothers because a time plot of the sequence $\{x_k^t : k = 1, \cdots, t\}$ is typically smoother than the forecasts $\{x_k^{k-1} : k = 1, \cdots, t\}$ or the filter $\{x_k^k : k = 1. \cdots, t\}$ [58]. Hence, smoothing implies

that each estimated value is a function of the past, present and future, whereas the filter estimator depends on the past and the present. The forecast depends only on the past.

3.3.1 Derivation of Kalman smoother

The Kalamn smoother can be derived in many ways. Here, we will follow the technique that was given originally by Ansley and Kohn [4].

The Kalman smoother algorithm is given in the following theorem.

Theorem 3.3.1.1. Consider the state space model specified in (3.1) and (3.2) with initial conditions x_k^a and P_k^a obtained from Theorem 3.2.1.1. For $k = t, t-1, \cdots, 1$, we have

$$x_k^t = x_k^a + J_k(x_{k+1}^t - x_{k+1}^f) \tag{3.23}$$

and

$$P_k^t = P_k^a + J_k(P_{k+1}^t - P_{k+1}^f)J_k^T \tag{3.24}$$

where

$$J_k = P_k^a A^T [P_{k+1}^f]^{-1}. \tag{3.25}$$

Proof. For $1 \leq k \leq t$, define

$$\{y\}_1^k = \{y_1, \cdots, y_k\}, \quad \text{with } \{y_0\} \text{ empty}$$

$$\eta_k = \{v_{k+1}, \cdots, v_t, w_{k+2}, \cdots, w_t\},$$

and let $q_k = E\{x_k | \{y\}_1^t, x_{k+1} - x_{k+1}^f, \eta_k\}$. Since, $\{y\}_1^k, \{x_{k+1} - x_k^f\}$ and η_k are matually independent and since x_k, η_k are independent, Lemma 2.1.6.1. gives

$$\begin{aligned} q_k &= E\{x_k | \{y\}_1^k, x_{k+1} - x_{k+1}^f, \eta_k\} \\ &= E_k\{x_k | x_{k+1} - x_{k+1}^f\} \\ &= x_k^a + \text{Cov}(x_k, x_{k+1} - x_{k+1}^f)[\text{Var}(x_{k+1} - x_{k+1}^f)]^{-1}(x_{k+1} - x_{k+1}^f). \end{aligned}$$

Now, compute $\text{Cov}(x_k, x_{k+1} - x_{k+1}^f)$ as follows:

$$\begin{aligned} \text{Cov}(x_k, x_{k+1} - x_{k+1}^f) &= \text{Cov}(x_k, Ax_k + w_k - x_{k+1}^f) \\ &= E[(x_k - x_k^a)(Ax_k + w_k - x_{k+1}^f - Ax_k^a + x_{k+1}^f)^T] \\ &= E[(x_k - x_k^a)(A(x_k - x_k^a))^T] \\ &= E[(x_k - x_k^a)(x_k - x_k^a)^T]A^T \\ &= P_k^a A^T. \end{aligned}$$

Similarly,

$$\text{Var}(x_{k+1} - x_{k+1}^f) = E[(x_{k+1} - x_{k+1}^f)(x_{k+1} - x_{k+1}^f)^T] = P_{k+1}^f.$$

Thus,

$$\begin{aligned} q_k &= x_k^a + P_k^a A^T [P_{k+1}^f]^{-1} (x_{k+1} - x_{k+1}^f) \\ &= x_k^a + J_k (x_{k+1} - x_{k+1}^f) \end{aligned}$$

where

$$J_k = P_k^a A^T [P_{k+1}^f]^{-1}.$$

Now, since $\{y\}_1^k, x_{k+1} - x_{k+1}^f$, and η_k generate $\{y\}_1^t = \{y_1, \cdots, y_t\}$ [58],

$$\begin{aligned} x_k^t &= E_t\{x_k\} \\ &= E_t\{q_k\} \\ &= x_k^a + J_k (x_{k+1}^t - x_{k+1}^f), \end{aligned}$$

which is Eq. (3.23).

Now to prove the identity given in Eq. (3.24), we use Eq. (3.3) and obtain

$$x_k - x_k^t = x_k - x_k^a - J_k (x_{k+1}^t - A x_k^a)$$

which is equivalents to

$$(x_k - x_k^t) + J_k x_{k+1}^t = (x_k - x_k^a) + J_k A x_k^a. \tag{3.26}$$

Multiplying both sides of Eq. (3.26) by the transpose of itself and taking the expectation, we get

$$P_k^t + J_k E(x_{k+1}^t x_{k+1}^{t^T}) J_k^T = P_k^a + J_k A E(x_k^a x_k^{a^T}) A^T J_k^T,$$

and because

$$\begin{aligned} E(x_{k+1}^t x_{k+1}^{t^T}) &= E(x_{k+1}x_{k+1}^T) - P_{k+1}^t \\ &= AE(x_k x_k^T)A^T + Q - P_{k+1}^t \end{aligned}$$

and

$$E(x_k^a x_k^{a^T}) = E(x_k x_k^T) - P_k^a,$$

we have

$$P_k^t + J_k[AP_k^a A^T + Q - P_{k+1}^t]J_k^T = P_k^a.$$

Hence,

$$P_k^t = P_k^a + J_k[P_{k+1}^t - P_{k+1}^f]J_k^T$$

which is Eq. (3.24).

3.3.2 The lag-one covariance smoother

The lag-one covariance smoother is a type of a smoother problem which is a set of recursions for obtaining $P_{k+1,k}^t$.

In the next theorem, we will state the well-known lag-one covariance smoother.

Theorem 3.3.2.1. For the state-space model specified in (3.1) and (3.2) with $K_k, J_k, (k = 1, 2, \cdots, t)$ and P_t^t are defined as in Theorems 3.2.1.1 and 3.3.1.1 and with initial con-

dition

$$P_{t+1,t}^{t+1} = (I - K_t C)AP_t^t \tag{3.27}$$

for, $t = k, k-1, \cdots, 1$. We have,

$$P_{k,k-1}^t = P_k^a J_{k-1}^T + J_k(P_{k+1,k}^t - AP_k^a)J_{k-1}^T. \tag{3.28}$$

Proof. To derive the initial condition (3.27), define

$$\tilde{x}_{k+1}^t = x_{k+1} - x_{k+1}^t,$$

where

$$x_{k+1}^t = E_t(x_{k+1}) = E\{x_{k+1}|\{y\}_1^t\}.$$

Thus,

$$\begin{aligned}
P_{k+1,k}^{k+1} &= E[\tilde{x}_{k+1}^{k+1}\tilde{x}_k^{k+1^T}] \\
&= E[(x_{k+1} - x_{k+1}^a)(x_k - x_k^{k+1})^T] \\
&= E\{[x_{k+1} - x_{k+1}^f - K_{k+1}(y_k - Cx_{k+1}^f)][x_k - x_k^a - J_k K_{k+1}(y_k - Cx_{k+1}^f)]^T\} \\
&= E\{[\tilde{x}_{k+1}^f - K_{k+1}(y_k - Cx_{k+1}^f)][\tilde{x}_k^a - J_k K_{k+1}(y_k - Cx_{k+1}^f)]^T\} \\
&= E\{[\tilde{x}_{k+1}^f - K_{k+1}(C(x_{k+1} - x_{k+1}^f) + v_k][\tilde{x}_{k+1}^a - J_k K_{k+1}(C(x_{k+1} - x_{k+1}^f) + v_k]^T\} \\
&= E\{[\tilde{x}_{k+1}^f - K_{k+1}(C\tilde{x}_{k+1}^f + v_k)][\tilde{x}_{k+1}^a - J_k K_{k+1}(C\tilde{x}_{k+1}^f + v_k)]^T\}
\end{aligned}$$

$$
\begin{aligned}
&= E\{\tilde{x}^f_{k+1}\tilde{x}^{a^T}_{k+1} - \tilde{x}^f_{k+1}\tilde{x}^{f^T}_{k+1}C^T K^T_{k+1}J^T_k + \tilde{x}^f_{k+1}v^T_k - K_{k+1}C\tilde{x}^f_{k+1}\tilde{x}^{a^T}_{k+1} \\
&\quad + K_{k+1}C\tilde{x}^f_{k+1}\tilde{x}^{f^T}_{k+1}C^T K^T_{k+1}J^T_k - K_{k+1}C\tilde{x}^f_{k+1}v^T_k - K_{k+1}v_k\tilde{x}_{k+1}a^T \\
&\quad - K_{k+1}v_k\tilde{x}^{f^T}_{k+1}CK^T_{k+1}J^T_k + K_{k+1}v_k v^T_k K^T_{k+1}J^T_k\} \\
&= P^k_{k+1,k} - P^f_{k+1}C^T K^T_{k+1}J_k - K_{k+1}CP^k_{k+1,k} + K_{k+1}CP^f_{k+1}C^T K^T_{k+1}J^T_k \\
&\quad + K_{k+1}RK^T_{k+1}J_k \\
&= P^k_{k+1,k} - P^f_{k+1}C^T K^T_{k+1}J_k - K_{k+1}CP^k_{k+1,k} + K_{k+1}[CP^f_{k+1}C^T + R]K^T_{k+1}J^T_k\,.
\end{aligned}
$$

Using $K_{k+1}[CP^f_{k+1}C^T + R] = P^f_{k+1}C^T$ and

$$
\begin{aligned}
P^k_{k+1,k} &= E\{(x_{k+1} - x^k_{k+1})(x_k - x^a_k)^T\} \\
&= E\{(Ax_k + w_k - Ax^a_k)(x_k - x^a_k)^T\} \\
&= AE\{(x_k - x^a_k)(x_k - x^a_k)^T\} \\
&= AP^a_k,
\end{aligned}
$$

we get

$$
\begin{aligned}
P^{k+1}_{k+1,k} &= AP^a_k - P^f_{k+1}C^T K^T_{k+1}J^T_k - K_{k+1}CAP^a_k + P^f_{k+1}C^T K^T_{k+1}J^T_k \\
&= (I - K_{k+1}C)AP^a_{k+1}.
\end{aligned}
$$

This relationship holds for any $t = 1, \cdots, n$. In particular, if we choose $k = t$, we obtain Eq. (3.27).

The task now is to derive Eq. (3.28). From Eq. (3.23), we have

$$x_k^t = x_k^a + J_k(x_{k+1}^t - x_{k+1}^f)$$

so,

$$x_k - x_k^t = x_k - x_k^a - J_k(x_{k+1}^t - x_{k+1}^f).$$

Thus,

$$\tilde{x}_x^t = \tilde{x}_k^a - J_k x_{k+1}^t + J_k A x_k^a \,.$$

That is,

$$\tilde{x}_k^t + J_k x_{k+1}^t = \tilde{x}_k^a + J_k A x_k^a, \tag{3.29}$$

and

$$\tilde{x}_{k-1}^t + J_{k-1} x_k^t = \tilde{x}_{k-1}^a + J_{k-1} A x_{k-1}^a \tag{3.30}$$

Multiplying the LHS of Eq. (3.29) by the transpose of the LHS of Eq. (3.30), and equate that with the corresponding result of the RHS of Eq. (3.29) and Eq. (3.30), we have

$$(\tilde{x}_k^t + J_k x_{k+1}^t)(\tilde{x}_{k-1}^t + J_{k-1} x_k^t)^T = (\tilde{x}_k^a + J_k A x_k^a)(\tilde{x}_{k-1}^a + J_{k-1} A x_{k-1}^a)^T$$

By taking the expectation of both sides we get

$$\begin{aligned} P_{k,k-1}^t \quad &+ \quad J_k E(x_{k+1}^t (x_k^t)^T) J_{k-1}^T \\ &= \quad P_{k,k-1}^{k-1} - K_k C P_{k,k-1}^{k-1} + J_k A K_k C P_{k,k-1}^{k-1} + J_k A E(x_k^a (x_{k-1}^a)^T) A^T J_{k-1}^T, \end{aligned}$$

where

$$\begin{aligned} E(\tilde{x}_k^a(\tilde{x}_{k-1}^a)^T) &= E(\tilde{x}_k^f - K_k(y_k - Cx_k^f)(\tilde{x}_{k-1}^a)^T) \\ &= P_{k,k-1}^{k-1} - K_k E((C(x_k - x_k^f + v_k))(\tilde{x}_{k-1}^a)^T) \\ &= P_{k,k-1}^{k-1} - K_k C E((\tilde{x}_k^f)(\tilde{x}_{k-1}^a)^T) \\ &= P_{k,k-1}^{k-1} - K_k C P_{k,k-1}^{k-1}. \end{aligned}$$

Also

$$\begin{aligned} E(x_k^a(\tilde{x}_{k-1}^a)^T) &= E\{(x_k^f + K_k(C\tilde{x}_k^f + v_k))((\tilde{x}_{k-1}^a)^T)\} \\ &= K_k C P_{k,k-1}^{k-1}. \end{aligned}$$

Using

$$E(x_{k+1}^t(x_k^t)^T) = E(x_{k+1}x_k^T) - P_{k+1,k}^t = AE(x_k x_{k-1}^T)A^T + AQ - P_{k+1,k}^t$$

and

$$\begin{aligned} E(x_k^a(x_{k-1}^a)^T) &= E\{(x_k^f + K_k(y_k + Cx_k^f))(x_{k-1}^a)^T\} \\ &= E(x_k^f(x_{k-1}^a)^T) \\ &= E(x_k x_{k-1}^T) - P_{k,k-1}^{k-1}, \end{aligned}$$

and after straightforward simplifications, we obtain

$$\begin{aligned}
P^t_{k,k-1} &= P^{k-1}_{k,k-1} - K_k C P^{k-1}_{k,k-1} + J_k A K_k C P^{k-1}_{k,k-1} - J_k A P^{k-1}_{k,k-1} A^T J^T_{k-1} - J_k A Q J^T_{k-1} \\
&\quad + J_k P^t_{k+1,k} J^T_{k-1} \\
&= (I - K_k C) P^{k-1}_{k,k-1} + J_k A (K_k C P^{k-1}_{k,k-1} - P^{k-1}_{k,k-1} A^T - Q) J^T_{k-1} \\
&\quad + J_K P^t_{k+1,k} J^T_{k-1} \\
&= P^a_k J^T_{k-1} - J_k A P^a_k J^T_{k-1} + J_k P^t_{k+1,k} J^T_{k-1} \\
&= P^a_k J^T_{k-1} + J_k (P^t_{k+1,k} - A P^a_k) J^T_{k-1}.
\end{aligned}$$

Chapter 4

IDENTIFICATION OF THE LINEAR GAUSSIAN STATE-SPACE MODEL

4.1 Introduction

System identification is a general term to describe the mathematical tools and algorithms that build dynamical models from measurement data [23,40,60]. System identification plays an important role in uncertain dynamical systems.

The Expectation-Maximization (EM) algorithm is a broadly applicable approach to the iterative computation of the maximum likelihood (ML) estimates, useful in a variety of incomplete data problems. On each iteration of the EM algorithm, there are two steps; the Expectation step (E-step) and the Maximization step (M-step). Thus, the algorithm is called the EM algorithm. This name was given by Dempster, Laird, and Rubin (1977) in their fundamental paper [9]. The EM algorithm is a technique that can be use to estimate the parameters after filling in the initial values for the missing data. The latter are then updated by their predicted values using these initial parameter estimates. The parameters are then re-estimated, and so on, proceeding iteratively until convergence. Indeed, this technique is so intuitive and natural. The EM have been formulated and applied to a variety of problems even before the work of Dempster, Laird, and Rubin (1977)[9]. But it was in this the work that the ideas were synthesized, a general formulation of the EM algorithm was established, its properties investigated, and a host of traditional and non-traditional applications indicated [47]. The formulation of the EM algorithm is given in Section 2.5.

The identification of a linear Gaussian state-space model was introduced by Shumway and Stoffer [59], when they identified the linear state-space model and estimated the parameters of the state space model using the expectation maximization algorithm (EM). Shumway and Stoffer used a Kalman filter and a Kalman smoother to estimate the parameters. The Kalman filter and the Kalman smoother are the basic tools to

calculate the expectation in the E-step. The aim of this chapter is to study the identification of the linear Gaussian state-space model by using the EM algorithm, where the Kalman filter and the Kalman smoother are applied in computing the E-step. In the M-step, we apply the rule of differentiation which was introduced in Section 2.8. This chapter is organized as follows; In the next section, we recall the linear state-space model which was introduced earlier. In Section 4.3, we present the technique of the identification by using the expectation maximization (EM) algorithm to estimate the parameters of the state-space model. We summarize the procedure of the identification in Section 4.4 .

4.2 The state-space model

Recall the linear Gaussian state-space model with dynamical system

$$x_{k+1} = Ax_k + w_k \tag{4.1}$$

and measurements system

$$y_k = Cx_k + v_k, \tag{4.2}$$

where, as before, A is the transition matrix ($n \times n$ matrix). C is the measurements matrix ($p \times 1$ matrix), w_k and v_k are the system noise and measurement noise respectively, where $w_k \sim N(0, Q)$ and $v_k \sim N(0, R)$, and Gaussian, white, and uncorrelated.

The initial state vector x_0 is assumed to be a Gaussian random vector with mean μ and covariance V, i.e., $x_0 \sim N(\mu, V)$.

4.3 Maximum likelihood estimation by the EM algorithm

To obtain the maximum Likelihood estimator of the parameters in the model, we use the EM algorithm [9,47]. Shumway and Stouffer [59] introduced the EM algorithm for the linear Gaussian state-space model.

Let $\{x\}_1^t = \{x_1, \cdots, x_t\}$ be a sequence of state variables, and $\{y\}_1^t = \{y_1, \cdots, y_t\}$ be a sequence of measurements. The joint likelihood function of the measurements and state variables can be computed by using Bayes' rule as follows,

$$\begin{aligned}
p(y_1, \cdots, y_t, x_1, \cdots, x_t) &= p(y_t|x_t, \{x\}_1^{t-1}, \{y\}_1^{t-1}) p(\{x\}_1^{t-1}, \{y\}_1^{t-1}) \\
&= p(y_t|x_t) p(x_t|\{x\}_1^{t-1}, \{y\}_1^{t-1}) p(\{y\}_1^{t-1}, \{x\}_1^{t-1}) \\
&= p(y_t|x_t) p(x_t|x_{t-1}) p(\{x\}^{t-1}, \{x\}_1^{t-1}) \\
&= \cdots \\
&= \prod_{k=0}^{t-1} p(y_k|x_k)) (\prod_{k=0}^{t-1} p(x_{k+1}|x_k) p(x_0).
\end{aligned}$$

Thus

$$p(\{x\}_1^t, \{y\}_1^t : \alpha) = p(x_0) \prod_{k=0}^{t-1} p(x_{k+1}, x_k) \prod_{k=0}^{t-1} p(y_k|x_k)$$

where $\alpha = \{A, C, Q, R, \mu\}$ is the parameter vector in the model. V is assumed to be known. The probability density functions $p(x_0), p(x_{k+1}|x_k)$ and $p(y_k|x_k)$ are given by

$$p(x_0) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\{-1/2(x_0 - \mu)^T V^{-1}(x_0 - \mu)\}$$

$$p(x_{k+1}|x_k) = \frac{1}{(2\pi)^{n/2}|Q|^{1/2}} \exp\{-1/2(x_{k+1} - Ax_k)^T Q^{-1}(x_{k+1} - Ax_k)\}$$

$$p(y_k|x_k) = \frac{1}{(2\pi)^{p/2}|R|^{1/2}} \exp\{-1/2(y_k - Cx_k)^T R^{-1}(y_k - Cx_k)\}.$$

Now, by taking the logarithm of the joint likelihood function we get

$$\begin{aligned}
\log P(\{x\}_1^t, \{y\}_1^t; \bar{\alpha}) &= -1/2 \log|V| - 1/2(x_0 - \mu)^T V^{-1}(x_0 - \mu) \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\sum_{k=0}^{t-1}(x_{k+1} - Ax_k)^T Q^{-1}(x_{k+1} - Ax_k) \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\sum_{k=0}^{t-1}(y_k - Cx_k)^T R^{-1}(y_k - Cx_k) \\
&\quad +const.
\end{aligned}$$

In the EM algorithm, to estimate the maximum likelihood parameter vector $\hat{\alpha}$, the

conditional expectation of the joint log-likelihood

$$
\begin{aligned}
q(\alpha|\bar{\alpha}) &= E(\log P(\{x_k\}_1^t, \{y\}_1^t, \alpha|\{y\}_1^t, \bar{\alpha}) \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}(V^{-1}E_t\{(x_0-\mu)(x_0-\mu)^T\} \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}\sum_{k=0}^{t-1}(E_t(x_{k+1}-Ax_k)(x_{k+1}-Ax_k)^T)\} \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=0}^{t-1}(E_t(y_k-Cx_k)(y_k-Cx_k)^T)\} \\
&\quad +const. \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}E_t(x_0x_0^T - x_0\mu^T - \mu x_0^T - \mu\mu^T)\} \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}\sum_{k=0}^{t-1}E_t(x_{k+1}x_{k+1}^T - x_{k+1}x_k^TA^T - Ax_kx_{k+1}^T \\
&\quad +Ax_kx_k^TA^T)\} \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=0}^{t-1}(E_t(y_ky_k^T - y_kx_k^TC^T - Cx_ky_k^T + Cx_kx_k^TC^T)\} \\
&\quad +const.
\end{aligned}
$$

Now taking conditional expectation on $\{y\}_1^t$ we get,

$$
\begin{aligned}
q(\alpha|\alpha_i) &= E[\log p(\{x\}_1^t, \{y\}_1^t, |\alpha)|\{y\}_1^t, \theta_i] \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}(p_0^t + (x_0^t-\mu)(x_0^t-\mu)^T] \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\sigma - \gamma A^T - A\gamma^T + A\beta A^T)] \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=0}^{t-1}[\xi - \varepsilon C^T - C\varepsilon^T + C\beta C^T] + const,
\end{aligned}
$$

where

$$P_0^t = E(x_0 x_0^T)$$

$$\sigma = \sum_{k=0}^{t-1} E_t(x_{k+1} x_{k+1}^T)$$

$$\beta = \sum_{k=0}^{t-1} E_t(x_k x_k^T)$$

$$\gamma = \sum_{k=0}^{t-1} E_t(x_{k+1} x_k^T)$$

$$\xi = \sum_{k=0}^{t-1} E_t(y_k y_k^T)$$

$$\varepsilon = \sum_{k=0}^{t-1} E_t(y_k x_k^T).$$

The quantities σ, β and γ can be computed using the Kalman smoother where

$$\beta = \sum_{k=0}^{t-1} (P_k^t + x_k^t (x_k^t)^T), \tag{4.3}$$

$$\gamma = \sum_{k=0}^{t-1} (P_{k+1,k}^t + x_{k+1}^t (x_k^t)^T), \tag{4.4}$$

$$\sigma = \sum_{k=0}^{t-1} (P_k^t + x_k^t x_k^{t^T}), \tag{4.5}$$

such that x_k^t, P_k^t and $P_{k+1,k}^t$, can be found using the Kalman filter and the Kalman smoother. More precisely,

$$x_k^t = E_t(x_k) = E(x_k | \{y\}_1^t)$$

$$P_k^t = \mathrm{Cov}(x_k | \{y\}_1^t)$$

and

$$P^t_{k,k-1} = \mathrm{Cov}(x_k, x_{k-1}|\{y\}^t_1),$$

• **The M-step**

The parameters of the system A, C, Q, R and μ, are collected in the vector $\alpha = \{A, C, Q, R, \mu\}$. To estimate these parameters, we take the corresponding partial derivatives of the expected log-likelihood function, and set to zero, then solve. So that, α_i is updated to α_{i+1} .

Thus,

$$\alpha_{i+1} = \{A(i+1), C(i+1), Q(i+1), R(i+1), \mu(i+1)\},$$

is obtained by maximizing the log likelihood function with respect to each parameter. The value of μ is

$$\mu = x^t_0 = E_t(x_0). \tag{4.6}$$

Now, by completing the square for the log likelihood function with respect to A, we get,

$$\begin{aligned} q(\alpha|\alpha_i) &= E[\log p(\{x\}^t_1, \{y\}^t_1, |\alpha)|\{y\}^t_1, \theta_i] \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}(p^t_0 + (x^t_0 - \mu)(x^t_0 - \mu)^T] \\ &\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\sigma - (A - \gamma\beta^{-1})\beta(A - \gamma\beta^{-1})^T)\} \\ &\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=0}^{t-1}[\xi - \varepsilon C^T - C\varepsilon^T + C\beta C^T] + const. \end{aligned}$$

Hence,

$$A(i+1) = \gamma\beta^{-1}. \tag{4.7}$$

Substituting the value of A, and taking the partial derivative with respect to Q, and then equating the result with zero we get,

$$-\frac{t}{2}Q^{-1} - \frac{1}{2}Q^{-1}\{(\sigma - \gamma\beta^{-1}\gamma^T)\}Q^{-1} = 0,$$

hence,

$$Q(i+1) = t^{-1}(\sigma - \gamma\beta^{-1}\gamma^T). \tag{4.8}$$

By the same argument, we find the value of C to be

$$C(i+1) = \varepsilon\beta^{-1}. \tag{4.9}$$

Similarly, substituting the value of C, taking the partial derivative with respect to R, and equating the result with zero we get,

$$R(i+1) = t^{-1}(\xi - \varepsilon\beta^{-1}\varepsilon^T). \tag{4.10}$$

4.4 Summary

The procedure of the identification of the parameters of the state-space model (4.1) and (4.2) using the EM algorithm to obtain the maximum likelihood estimation vector

$\hat{\alpha}$ is summarized as follows:

1. Select initial values of $\mu(0), A(0), C(0), Q(0), R(0)$, and compute x_k^t, P_k^t and $P_{k,k-1}^t$ using the Kalman smoother.

2. Calculate the conditional expectation of the log Likelihood function.

3. Calculate Equations (4.6)-(4.10) and find the next iterative estimated parameters that maximize the conditional expectation of the log Likelihood.

4. Insert the new parameters in the state-space model, and compute the Kalman smoother.

5. Repeat the steps 2,3, and 4 until the log likelihood converge.

Chapter 5

BILINEAR FILTERING PROBLEM

5.1 Introduction

When the dynamical system and measurements system of the state-space model are linear, the Kalman filter is the optimal estimator. However, in most processes of interest the dynamical and/or the measurements systems are nonlinear and so a suitable extension of the Kalman filter is needed. In this chapter, we introduce a new nonlinear model of bilinear type and describe a new approach to generalize the

Kalman filter. Indeed, the well-known Lorenz-96 model [41,42] and Lotka-Volterra model [3,20,61] which are used for atmospheric dynamics, biology, chemistry, and control systems are special cases of our bilinear model. More precisely, the Lorenz-96 the Lotka-Volterra models can be obtained from our model when the coefficient matrices are chosen properly. The model consists of a nonlinear state and a linear measurements equation. The specific nonlinearity is of the bilinear form depending on the system dynamics. Previously, bilinear models consisted of bilinear function of the input and state dynamics [71,51,54].

In the next section, we introduce a new nonlinear Gaussian discrete state space model and present a new bilinear Kalman filter and a bilinear Kalman smoother which is an extension of the Kalman filter and Kalman smoother occurring in the linear case. Section 5.3 is devoted to present a derivation of the new bilinear filtering algorithm which we refer to as the bilinear Kalman filter. In Section 5.4, the derivation of the new bilinear smoother algorithm is introduced and we will name it the bilinear Kalman smoother.

5.2 A Bilinear state-space model

In this section, we present the bilinear Gaussian discrete state-space model. This model is described by

$$x_{k+1} = Ax_k + B(x_k \otimes x_k) + w_k; \tag{5.1}$$

$$y_k = Cx_k + v_k; \tag{5.2}$$

where;

$x_k \in \mathbb{R}^n$ is the state vector,

$y_k \in \mathbb{R}^p$ is the measurements vector,

while the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times \frac{n(n+1)}{2}}$ and $C \in \mathbb{R}^{p\times n}$ are the parameters of the model.

The noise corruption signals $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^p$ are white, uncorrelated and Gaussian with zero mean and covariances Q and R respectively. That is,

$$w_k \sim N(0, Q),$$

$$v_k \sim N(0, R).$$

Also,

$$E(w_k w_l^T) = \begin{cases} Q, \text{for} \quad \text{k=l} \\ 0, \text{for} \quad \text{k}\neq\text{l}, \end{cases}$$

$$E(v_k v_l^T) = \begin{cases} R, \text{for} & \text{k=l} \\ 0, \text{for} & \text{k}\neq\text{l}, \end{cases}$$

and

$$E(w_k v_k^T) = 0.$$

Here, $x_k \otimes x_k$ denotes the Kronecker product of the state x_k with itself. The definition of the Kronecker product is given in Section 2.7.

The Lorenz-96 model [41,42] is given by

$$\frac{dx_k}{dt} = (x_{k+1} - x_{k-2})x_{k-1} - x_k + F, \tag{5.3}$$

for $k = 1, ..., n$. Where F is the forcing constant which is independent of k. In this model, the definition of x_k is extended in such a way that they form cyclic boundary conditions; that is, for all values of k,

$$x_{k-n} = x_k \text{ and } x_{k+n} = x_k.$$

For example, when $n = 3$, we have the following system

$$\frac{dx_1}{dt} = -x_1$$

$$\frac{dx_2}{dt} = -x_2$$

$$\frac{dx_3}{dt} = -x_3 + (x_3x_2 - x_1x_2).$$

This system can be reformulated as,

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_1x_3 \\ x_2^2 \\ x_2x_3 \\ x_3^2 \end{pmatrix}$$

Comparing the last expression with our bilinear model (5.1), we conclude that;

$$A = -I_3$$

and

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Thus, the well-known Lorenz 96 model is a special case of our new bilinear model. Dealing with the nonlinear term creates a lot of technicality as well as mathematical computations. For this reason we introduce the following approximation of the nonlinear term. Using Taylor expansion we get,

$$z_k = z_k^j + f'(x_k^j)(x_k - x_k^j) + \frac{1}{2}\mathcal{H}(x_k, x_k^j)(x_k - x_k^j),$$

where,

$$z_k^j = E_j(z_k)$$

$$f'(x) = [\frac{\partial x_i x_j}{\partial x_l}]_{i,j,l=1,2,\ldots,m},$$

$$\mathcal{H}(x_{k+1}, x_{k+1}^j) = \begin{pmatrix} (x_{k+1} - x_{k+1}^j)^T D_1 \\ (x_{k+1} - x_{k+1}^j)^T D_2 \\ \vdots \\ (x_{k+1} - x_{k+1}^j)^T D_m \end{pmatrix}$$

and, $\mathcal{D} = \begin{pmatrix} D_1 \\ \vdots \\ D_m \end{pmatrix}$ is the $m^2 \times m$ matrix of second derivatives.

- **Reasonable approximations for z_k^j**:

All approximations are based on replacing the term $(x_k - x_k^j)^T$ in the last expression by a reasonable approximation.

- In the case of prediction, the interest is in computations involving $(x_k - x_k^{k-1})$. We take,

$$(x_k - x_k^{k-1}) \approx (x_k^{k-2} - x_k^{k-1}),$$

with $x_1^{-1} = 0$. That is,

$$z_k = z_k^{k-1} + f'(x_k^{k-1})(x_k - x_k^{k-1}) + \frac{1}{2}\mathcal{H}(x_k^{k-2}, x_k^{k-1})(x_k - x_k^{k-1}),$$

- In the case of filtering, the interest is in computations involving $(x_k - x_k^k)$.

  We take,

  $$(x_k - x_k^k) \approx (x_k^{k-1} - x_k^k).$$

  This means that,

  $$z_k = z_k^k + f'(x_k^k)(x_k - x_k^k) + \frac{1}{2}\mathcal{H}(x_k^{k-1}, x_k^k)(x_k - x_k^k),$$

- In the case of smoothing, the interest is in computations involving $(x_k - x_k^{k+1})$.

  We take,

  $$(x_k - x_k^{k+1}) \approx (x_k^{k+2} - x_k^{k+1}),$$

  with $x_{N-1}^{N+1} = 0$. This means that,

  $$z_k = z_k^{k+1} + f'(x_k^{k+1})(x_k - x_k^{k+1}) + \frac{1}{2}\mathcal{H}(x_k^{k+2}, x_k^{k+1})(x_k - x_k^{k+1}),$$

Thus, we can summarize all previous cases in the following equation;

$$\begin{aligned} z_k &= z_k^j + f'(x_k^j)(x_k - x_k^j) + \frac{1}{2}\mathcal{H}(x_k^{j\pm1}, x_k^j)(x_k - x_k^j) \qquad (5.4) \\ &= z_k^j + V_k^j(x_k - x_k^j), \end{aligned}$$

where,

$$V_k^j = f'(x_k^j) + \frac{1}{2}\mathcal{H}(x_k^{j\pm1}, x_k^j). \qquad (5.5)$$

The nonlinearity presents challenge as far as the great sensitivity in the structure and the components of these systems, such as the parameters of the system.

For example, in the nonlinear Lotka-Volterra model which is of the form

$$x_{k+1} = (I + hA)x_k + rhBz_k,$$

where, $z_k = x_k \otimes x_k$ and $B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix}$. For different choices of the parameter A and considering different values of h and r, the obtained results are completely different, although we deal with the same nonlinear model. Results of many cases are described in Figures 5.1, 5.2 and 5.3, as follows:

Figure 5.1: When $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $h = 0.1$ and $r = 0.1$.

Figure 5.2: When $A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$, $h = 0.1$ and $r = -1$.

Figure 5.3: When $A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$, $h = 0.1$ and $r = 1$.

Figure 5.1. The Lotka-Volterra Model with r=0.1, h=0.1 and A=[0 1;-1 0]

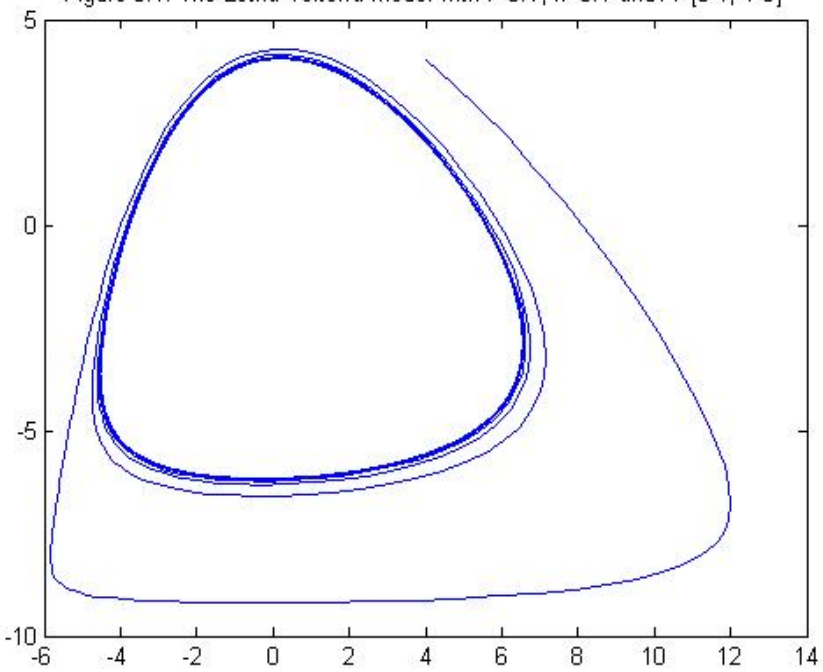
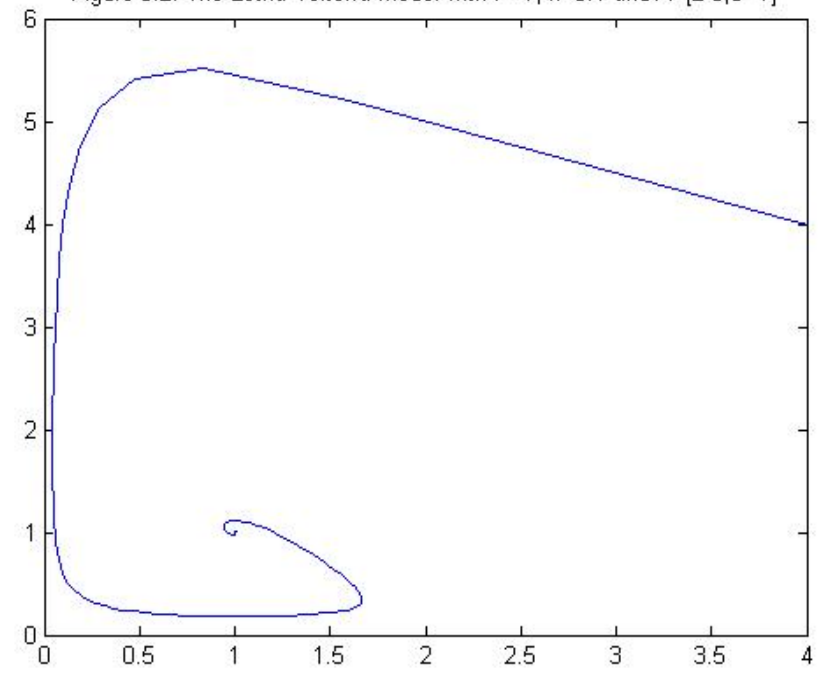
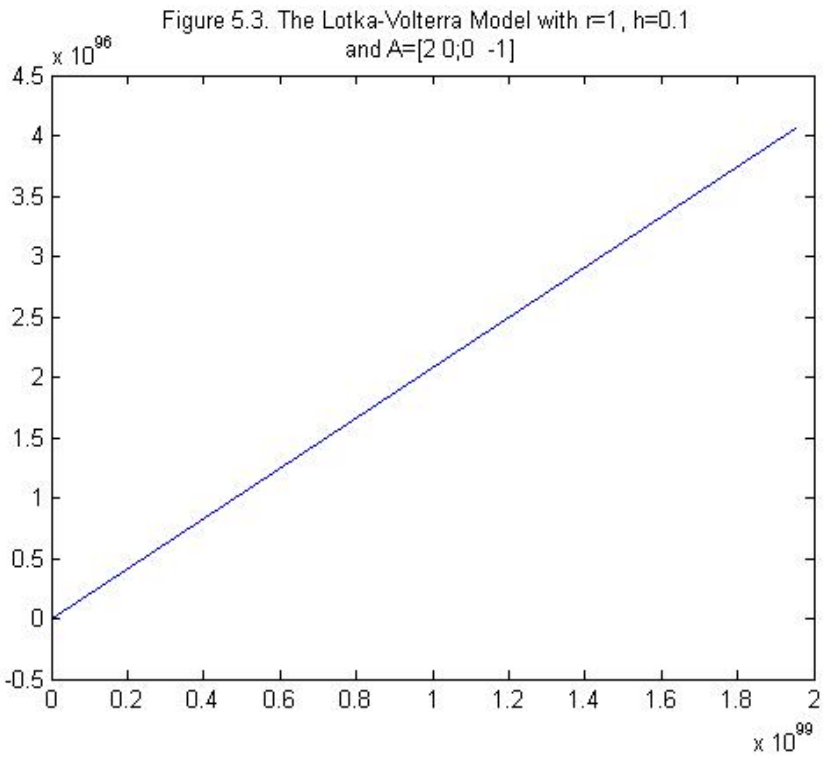


Figure 5.2. The Lotka-Volterra Model with r=-1, h=0.1 and A=[2 0;0 -1]





The other challenges of the nonlinearity are the identification and the filtering. Typically approaches concentrated on linearization, such as the extended Kalman filter and/or replacing the nonlinear system with a set of systems, such as the ensemble Kalman filter. These challenges will be discussed in the next chapter.

5.3 A Bilinear Kalman filter

In recent years, the state space representation and the associated Kalman recursions have had a profound impact on the time series analysis and many related areas. The techniques were originally developed in connection with control of linear systems [7,16,52].

Here, we have a nonlinear system and thus, using the linear Kalman filter is not possible in this case for two reasons; the first is self-evident; the linear Kalman filter is appropriate only for linear systems. Secondly, we do not want to approximate our nonlinear system by a linear one as in the case of the extended Kalman filter [16,19]. Therefore, we need to derive a nonlinear filter suitable for the nonlinear system.

In this section, we derive a bilinear filter associated with the bilinear state space model defined by (5.1) and (5.2). We call it the bilinear Kalman filter.

We will adopt the following notations to derive the bilinear Kalman filter:

$$z_k = x_k \otimes x_k$$

$$x_k^t = E\{x_k | \{y\}_1^t\} = E_t(x_t)$$

$$P_k^t = E_t\{(x_k - x_k^t)(x_k - x_k^t)^T\}$$

$$z_k^t = E\{z_k | \{y\}_1^t\} = E_t(z_t)$$

$$\dot{P}_k^t = E_t\{(z_k - z_k^t)(z_k - z_k^t)^T\}$$

$$\ddot{P}_k^t = E_t\{(x_k - x_k^t)(z_k - z_k^t)^T\},$$

where

$$1 \leq k \leq t$$

$$1 \leq t \leq n$$

and $\{y\}_1^t$ is the measurements sequence

$$\{y\}_1^t = \{y_1, ..., y_t\}.$$

In the following theorem, we state the bilinear Kalman filter.

Theorem 5.3.1. For the bilinear state-space model defined by (5.1) and (5.2), we have

$$x_{k+1}^k = Ax_k^k + Bz_k^k \tag{5.6}$$

$$P_{k+1}^k = AP_k^k A^T + A\ddot{P}_k^k B^T + B(\ddot{P}_k^k)^T A^T + B\dot{P}_k^k B^T + Q. \tag{5.7}$$

with

$$x_{k+1}^{k+1} = x_{k+1}^k + K_{k+1}[y_k - Cx_{k+1}^k] \tag{5.8}$$

$$P_{k+1}^{k+1} = [I - K_{k+1}C]P_{k+1}^k \tag{5.9}$$

$$\ddot{P}_{k+1}^{k+1} = P_{k+1}^{k+1}[V_{k+1}^{k+1}]^T \tag{5.10}$$

$$\dot{P}_{k+1}^{k+1} = (V_{k+1}^{k+1})\ddot{P}_{k+1}^{k+1} \tag{5.11}$$

where

$$K_{k+1} = P_{k+1}^k C^T [C P_{k+1}^k C^T + R]^{-1} \tag{5.12}$$

and

$$V_{k+1}^{k+1} = f'(x_{k+1}^{k+1}) + \frac{1}{2}\mathcal{H}(x_{k+1}^k, x_{k+1}^{k+1}), \tag{5.13}$$

for $k = 0, ..., t$.

Proof. First, we derive the forecast steps which are given in Eq.(5.6) and Eq.(5.7). That is, we consider the case $t < k$ in the previous notations. We derive Eq.(5.6) as follows

$$\begin{aligned} x_{k+1}^k &= E_k(x_{k+1}) \\ &= E_k(Ax_k + Bz_k + w_k) \\ &= AE_k(x_k) + BE_k(z_k) + E_k(w_k) \\ &= Ax_k^k + Bz_k^k. \end{aligned}$$

Also,

$$\begin{aligned} P_{k+1}^k &= E_k\{(x_{k+1} - x_{k+1}^k)(x_{k+1} - x_{k+1}^k)^T\} \\ &= E_k\{(Ax_k + Bz_k + w_k - Ax_k^k - Bz_k^k)(Ax_k + Bz_k + w_k - Ax_k^k - Bz_k^k)^T\} \\ &= E_k\{[A(x_k - x_k^k) + B(z_k - z_k^k) + w_k][A(x_k - x_k^k) + B(z_k - z_k^k) + w_k]^T\} \end{aligned}$$

$$
\begin{aligned}
&= E_k\{[A(x_k - x_k^k) + B(z_k - z_k^k) + w_k][(x_k - x_k^k)^T A^T + (z_k - z_k^k)^T B^T + w_k^T]\} \\
&= E_k\{A(x_k - x_k^k)(x_k - x_k^k)^T A^T + A(x_k - x_k^k)(z_k - z_k^k)^T B^T + B(z_k - z_k^k)(x_k - x_k^k)^T A^T \\
&\quad + B(z_k - z_k^k)(z_k - z_k^k)^T B^T + w_k w_k^T\} \\
&= AP_k^k A^T + A\ddot{P}_k^k B^T + B(\ddot{P}_k^k)^T A^T + B\dot{P}_k^k B^T + Q
\end{aligned}
$$

which is Eq.(5.7).

Now, when $t = k$, we derive the filtering steps. Let

$$
\begin{aligned}
\rho_k &= y_k - E_{k-1}(y_k) \\
&= y_k - E_{k-1}(Cx_k - v_k) \\
&= y_k - Cx_k^{k-1} \\
&= Cx_k - Cx_k^{k-1} + v_k \\
&= C(x_k - x_k^{k-1}) + v_k,
\end{aligned}
$$

for $k = 1, ..., t$. Thus, we notice that,

$$
E_{k-1}(\rho_k) = \rho_k^{k-1} = 0 \tag{5.14}
$$

and

$$\begin{aligned}\Sigma_{k+1} &= \mathrm{Var}(\rho_{k+1}) \\ &= E\{[C(x_{k+1}-x_{k+1}^k)+v_k][C(x_{k+1}-x_{k+1}^k)+v_k]^T\} \\ &= CE\{[x_{k+1}-x_{k+1}^k][x_{k+1}-x_{k+1}^k]^T\}C^T + E(v_k v_k^T) \\ &= CP_{k+1}^k C^T + R.\end{aligned}$$

Also that,

$$E_k(\rho_{k+1}y_k^T) = E_k((y_{k+1}-y_{k+1}^k)y_k^T) = 0, \tag{5.15}$$

which means that the innovations are independent of the past measurements. On the other hand, the conditional covariance between x_{k+1} and ρ_{k+1} is computed as follows

$$\begin{aligned}\mathrm{Cov}(x_{k+1},\rho_{k+1}) &= \mathrm{Cov}(x_{k+1}-x_{k+1}^k, C(x_{k+1}-x_{k+1}^k)+v_k) \\ &= E\{[(x_{k+1}-x_{k+1}^k)-E(x_{k+1}-x_{k+1}^k)] \\ &\quad [C(x_{k+1}-x_{k+1}^k)+v_k-CE(x_{k+1}-x_{k+1}^k)]^T\} \\ &= P_{k+1}^k C^T.\end{aligned}$$

From these results, we conclude that x_{k+1} and ρ_{k+1} have a Gaussian joint conditional distribution on $\{y\}_1^k$. That is,

$$\{\begin{pmatrix} x_{k+1} \\ \rho_{k+1}\end{pmatrix} | \{y\}_1^k\} \sim N\{\begin{pmatrix} x_{k+1}^k \\ 0\end{pmatrix}, \begin{pmatrix} P_{k+1}^k & P_{k+1}^k C^T \\ CP_{k+1}^k & \Sigma_{k+1}\end{pmatrix}\}. \tag{5.16}$$

Now, applying Lemma 2.1.6.1., we have

$$\begin{aligned} x_{k+1}^{k+1} &= E_{k+1}(x_{k+1}) \\ &= E_k\{x_{k+1}|\rho_{k+1}\} \\ &= E_k(x_{k+1}) + \mathrm{Cov}(x_{k+1}, \rho_{k+1})\Sigma_{k+1}^{-1}\rho_{k+1} \\ &= x_{k+1}^k + P_{k+1}^k C^T[CP_{k+1}^k C^T + R]^{-1}\rho_{k+1} \\ &= x_{k+1}^k + K_{k+1}[y_{k+1} - Cx_{k+1}^k]; \end{aligned}$$

where

$$K_{k+1} = P_{k+1}^k C^T[CP_{k+1}^k C^T + R]^{-1}$$

represents the Kalman gain.

To derive Eq.(5.9), we use Eq.(5.16) and apply Lemma 2.1.6.1 to get

$$\begin{aligned} P_{k+1}^{k+1} &= \mathrm{Cov}(x_{k+1}, \rho_{k+1}) \\ &= \mathrm{Cov}(x_{k+1}) - \mathrm{Cov}(x_{k+1}, \rho_{k+1})\Sigma_{k+1}^{-1}\mathrm{Cov}(\rho_{k+1}, x_{k+1}) \\ &= P_{k+1}^k - P_{k+1}^k C^T[CP_{k+1}^k C^T + R]^{-1}CP_{k+1}^k \\ &= P_{k+1}^k - K_{k+1}CP_{k+1}^k \\ &= [I - K_{k+1}C]P_{k+1}^k\,. \end{aligned}$$

To derive Eq.(5.10), we use Eq.(5.4) as follows:

$$\begin{aligned}
\ddot{P}_{k+1}^{k+1} &= E((x_{k+1} - x_{k+1}^{k+1})(z_{k+1} - z_{k+1}^{k+1})^T) \\
&= E(((x_{k+1} - x_{k+1}^{k+1})(x_{k+1} - x_{k+1}^{k+1})^T)[V_{k+1}^{k+1}]^T \\
&= P_{k+1}^{k+1}[V_{k+1}^{k+1}]^T .
\end{aligned}$$

By employing similar arguments for deriving Eq(5.11), we obtain

$$\begin{aligned}
\dot{P}_{k+1}^{k+1} &= E((z_{k+1} - z_{k+1}^{k+1})(z_{k+1} - z_{k+1}^{k+1})^T) \\
&= (V_{k+1}^{k+1})E((x_{k+1} - x_{k+1}^{k+1})(x_{k+1} - x_{k+1}^{k+1})^T)[V_{k+1}^{k+1}]^T \\
&= (V_{k+1}^{k+1})P_{k+1}^{k+1}[V_{k+1}^{k+1}]^T \\
&= (V_{k+1}^{k+1})\ddot{P}_{k+1}^{k+1} .
\end{aligned}$$

5.4 A Bilinear Kalman smoother

In this section, we will consider the problem when $t > k+1$. As mentioned earlier, this kind of estimators is called a smoother. Here, we derive a bilinear Kalman smoother associated with the bilinear state-space model (5.1) and (5.2).

For convenience, we introduce the following notations:

$$P_{k_1,k_2}^t = E\{(x_{k_1} - x_{k_1}^t)(x_{k_2} - x_{k_2}^t)^T\}$$

$$\dot{P}^t_{k_1,k_2} = E\{(z_{k_1} - z^t_{k_1})(z_{k_2} - z^t_{k_2})^T\}$$

$$\ddot{P}^t_{k_1,k_2} = E\{(x_{k_1} - x^t_{k_1})(z_{k_2} - z^t_{k_2})^T\}.$$

The next lemma is needed in the derivation of the bilinear Kalman smoother.

Lemma 5.4.1. Let,

$$\epsilon_{k+1} = \{v_{k+1}, ..., v_t, w_{k+2}, ..., w_t\}.$$

Then, for $1 \leq k \leq t-1$, $\{y\}^k_1$,$\{x_{k+1} - x^k_{k+1}\}$ and ϵ_{k+1} build $\{y\}^t_1$.

Proof. Recall that

$$z_{k+1} = z^j_k + V^j_{k+1}(x_{k+1} - x^j_{k+1}),$$

that is,

$$(z_{k+1} - z^j_k) \in \{(x_{k+1} - x^j_{k+1})\}.$$

Knowing that,

$$y_{k+1} = Cx_{k+1} + v_{k+1},$$

and

$$y^k_{k+1} = Cx^k_{k+1},$$

yield

$$y_{k+1} - y^k_{k+1} = C(x_{k+1} - x^k_{k+1}) + v_{k+1}.$$

Thus,

$$\begin{aligned}\{y\}_1^{k+1} &= \{\{y\}_1^k, y_{k+1}\} \\ &= \{\{y\}_1^k, y_{k+1} - y_{k+1}^k\} \\ &= \{\{y\}_1^k, x_{k+1} - x_{k+1}^k, v_{k+1}\}.\end{aligned}$$

Similarly,

$$y_{k+2} - y_{k+2}^k = CA(x_{k+1} - x_{k+1}^k) + CB(z_{k+1} - z_{k+1}^k) + Cw_{k+1} + v_{k+2},$$

and hence,

$$\begin{aligned}\{y\}_1^{k+2} &= \{\{y\}_1^{k+1}, y_{k+2}\} \\ &= \{\{y\}_1^{k+1}, y_{k+2} - y_{k+2}^k\} \\ &= \{\{y\}_1^k, x_{k+1} - x_{k+1}^k, v_{k+1}, z_{k+1} - z_{k+1}^k, w_{k+1}, v_{k+2}\} \\ &= \{\{y\}_1^k, x_{k+1} - x_{k+1}^k, v_{k+1}, w_{k+1}, v_{k+2}\}.\end{aligned}$$

Continuing, we see that $\{y\}_1^t$ is generated by $y_k, x_{k+1} - x_{k+1}^k$ and ϵ_{k+1}.

We state the bilinear Kalman smoother in the following theorem.

Theorem 5.4.2. Consider the bilinear state-space model (5.1) and (5.2),with x_t^t and P_t^t as given in Theorem 5.3.1. Then for $k = t-1, ..., 1$, we have

$$x_k^t = x_k^k + J_k(x_{k+1}^t - x_{k+1}^k) \tag{5.17}$$

$$P_k^t = P_k^k + J_k[P_{k+1}^t - P_{k+1}^k]J_k^T; \tag{5.18}$$

where,

$$J_k = [P_k^k A^T + \ddot{P}_k^k B^T][P_{k+1}^k]^{-1}. \tag{5.19}$$

Proof. For $1 \leq k \leq t-1$, we define

$$\epsilon_{k+1} = \{v_{k+1}, ..., v_t, w_{k+2}, ..., w_t\}.$$

Let

$$\tau_{k+1} = E_k\{x_k | x_{k+1} - x_{k+1}^k, \epsilon_{k+1}\}.$$

We notice that, x_k and ϵ_{k+1} are independent, and the sequence of measurements $\{y\}_1^k$,$\{x_{k+1} - x_{k+1}^k\}$ and ϵ_{k+1} are mutually independent. So that, by applying Lemma 2.1.6.1., we obtain,

$$\begin{aligned}
\tau_{k+1} &= E_k\{x_k | x_{k+1} - x_{k+1}^k, \epsilon_{k+1}\} \\
&= x_k^k + \mathrm{Cov}(x_k, x_{k+1} - x_{k+1}^k)[\mathrm{Var}(x_{k+1} - x_{k+1}^k)]^{-1}(x_{k+1} - x_{k+1}^k)
\end{aligned}$$

Now, we compute $\mathrm{Cov}(x_k, x_{k+1} - x_{k+1}^k)$ as follows

$$\begin{aligned}
\mathrm{Cov}(x_k, x_{k+1} - x_{k+1}^k) &= \mathrm{Cov}(x_k, Ax_k + Bz_k + w_k - x_{k+1}^k) \\
&= E_k\{[x_k - x_k^k][Ax_k + Bz_k + w_k - x_{k+1}^k - Ax_k^k - Bz_k^k + x_{k+1}^k]^T\} \\
&= E_k\{[x_k - x_k^k][A(x_k - x_k^k) + B(z_k - z_k^k) + w_k]^T\}
\end{aligned}$$

$$
\begin{aligned}
&= E_k\{[(x_k - x_k^k)(x_k - x_k^k)^T A^T + (x_k - x_k^k)(z_k - z_k^k)^T B^T]\} \\
&= P_k^k A^T + \ddot{P}_k^k B^T,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(x_{k+1} - x_{k+1}^k) &= E\{(x_{k+1} - x_{k+1}^k)(x_{k+1} - x_{k+1}^k)^T\} \\
&= P_{k+1}^k.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\tau_{k+1} &= x_k^k + [P_k^k A^T + \ddot{P}_k^k B^T][P_{k+1}^k]^{-1}[x_{k+1} - x_{k+1}^k] \\
&= x_k^k + J_k(x_{k+1} - x_{k+1}^k),
\end{aligned}
$$

where

$$
J_k = [P_k^k A^T + \ddot{P}_k^k B^T][P_{k+1}^k]^{-1}.
$$

By using Lemma 5.4.1, we conclude that,

$$
\begin{aligned}
x_k^t &= E_t(x_k) \\
&= E_t(\tau_{k+1}) \\
&= x_k^k + J_k(x_{k+1}^t - x_{k+1}^k),
\end{aligned}
$$

which proves Eq.(5.17). To derive Eq(5.18), we subtract Eq.(5.17) from x_k and use Eq.(5.6);

$$
\begin{aligned}
x_k - x_k^t &= x_k - x_k^k - J_k(x_{k+1}^t - x_{k+1}^k) \\
&= x_k - x_k^k - J_k(x_{k+1}^t - Ax_k^k - Bz_k^k)
\end{aligned}
$$

and hence,

$$(x_k - x_k^t) + J_k x_{k+1}^t = (x_k - x_k^k) + J_k A x_k^k + J_k B z_k^k. \tag{5.20}$$

Multiplying both sides of Eq.(5.20) by the transpose of itself and taking the expectation, we have

$$\begin{aligned} P_k^t + J_k E(x_{k+1}^t (x_{k+1}^t)^T) J_k^T &= P_k^k + J_k A E(x_k^k (x_k^k)^T) A^T J_k^T + J_k B E(z_k^k (z_k^k)^T) B^T J_k^T \\ &\quad + J_k A E(x_k^k (z_k^k)^T)) B^T J_k^T + J_k B E(z_k^k (x_k^k)^T)) A^T J_k^T. \end{aligned}$$

Also,

$$\begin{aligned} E(x_{k+1}^t (x_{k+1}^t)^T) &= E(x_{k+1} x_{k+1}^T) - P_{k+1}^t \\ &= A E(x_k x_k^T) A^T + A E(x_k (z_k)^T)) B^T + B E(z_k (x_k)^T)) A^T \\ &\quad + B E(z_k (z_k)^T) B^T + Q - P_{k+1}^t \end{aligned}$$

and

$$E(x_k^k (x_k^k)^T) = E(x_k x_k^T) - P_k^k,$$

$$E(x_k^k (z_k^k)^T) = E(x_k z_k^T) - \ddot{P}_k^k,$$

$$E(z_k^k (z_k^k)^T) = E(z_k z_k^T) - \dot{P}_k^k.$$

Therefore,

$$P_k^t + J_k [A P_k^k A^T + A \ddot{P}_k^k B^T + B (\ddot{P}_k^k)^T A^T + B \dot{P}_k^k B^T + Q - P_{k+1}^t] J_k = P_k^k,$$

and since,

$$P_{k+1}^k = AP_k^k A^T + A\ddot{P}_k^k B^T + B(\ddot{P}_k^k)^T A^T + B\dot{P}_k^k B^T + Q$$

we get,

$$P_k^t = P_k^k + J_k[P_{k+1}^t - P_{k+1}^k]J_k^T,$$

which is Eq. (5.18) and the proof is completed now.

In the next theorem, we introduce the bilinear lag-one covariance smoother to obtain $P_{k+1,k}^t$ and $\ddot{P}_{k+1,k}^t$.

Theorem 5.4.3. Consider the state-space model (5.1) and (5.2). Then,

$$P_{k+1,k}^t = AP_k^k + B(\ddot{P}_k^k)^T \tag{5.21}$$

and

$$\ddot{P}_{k+1,k}^t = P_{k+1,k}^t[V_k^t]^T \tag{5.22}$$

where P_k^k and V_k^t are defined as before.

Proof. First, by the previous notations, we get

$$\begin{aligned} P_{k+1,k}^t &= E_t((x_{k+1} - x_{k+1}^{k+1})(x_k - x_k^k)^T) \\ &= E_t((Ax_k - Bz_k + w_k - Ax_k^k - Bz_k^k)(x_k - x_k^k)^T) \\ &= E_t((A(x_k - x_k^k) + B(z_k - z_k^k) + w_k)(x_k - x_k^k)^T) \end{aligned}$$

$$
\begin{aligned}
&= E_t((A(x_k - x_k^k)(x_k - x_k^k)^T + B(z_k - z_k^k)(x_k - x_k^k)^T) \\
&= AE_t((x_k - x_k^k)(x_k - x_k^k)^T) + BE_t((z_k - z_k^k)(x_k - x_k^k)^T) \\
&= AP_k^k + B(\ddot{P}_k^k)^T
\end{aligned}
$$

and also,

$$
\begin{aligned}
\ddot{P}_{k+1,k}^t &= E_t((x_{k+1} - x_{k+1}^t)(z_k - z_k^k)^T) \\
&= E_t((x_{k+1} - x_{k+1}^t)(V_k^t(x_k - x_k^k))^T) \\
&= E_t((x_{k+1} - x_{k+1}^t)(x_k - x_k^k)^T[V_k^t]^T) \\
&= E_t((x_{k+1} - x_{k+1}^t)(x_k - x_k^k)^T)[V_k^t]^T \\
&= P_{k+1,k}^t[V_k^t]^T.
\end{aligned}
$$

Chapter 6

IDENTIFICATION OF A BILINEAR MODEL VIA THE BILINEAR EM ALGORITHM

6.1 Introduction

System identification is a general term to describe mathematical tools and algorithms that build dynamical models from measured data. It plays an important role in uncertain control systems. The identification of nonlinear systems has been a chal-

lenging research area since Wiener published his work on nonlinear random theory in 1958 [64].

One of the objectives of this dissertation is to identify the bilinear state space model. When the nonlinearity consists of the Kronecker product of the states with input, the identification step suffers no problem, because the system automatically reduce to a linear identification problem (see [17,51,54]).

In this chapter, our aim is to introduce the identification of the bilinear state space model defined by (5.1) and (5.2) in Section 5.2 where we use the expectation maximization algorithm. In the next section, we present the method of identifying the parameters of our nonlinear model by using the expectation maximization algorithm.

6.2 Identification of the bilinear state-space model

The expectation maximization (EM) algorithm is an iterative technique can be used for obtaining the maximum likelihood estimation [see Section 2.5]. According to the EM algorithm, the steps are:

1. The expectation (E) step: we compute

$$q(\theta, \theta(i)) = E\{\log p(\theta, X, Y)|Y\}$$

where $p(\theta, X, Y)$ is the maximum likelihood function.

2. The maximization (M) step: maximize $q(\theta, \theta(i))$ with respect to θ.

Here, the symbol θ refers to the parameters vector

$$\theta = \{A, B, C, Q, R, V, \mu\}$$

and

$$\begin{aligned} X &= \{x\}_1^t = \{x_1, ..., x_t\}, \\ Y &= \{y\}_1^t = \{y_1, ..., y_t\} \end{aligned} .$$

The following theorem accomplishes the expectation step.

Theorem 6.2.1. Consider the bilinear state-space model (5.1) and (5.2) with the following assumptions

$$x_0 \sim N(\mu, V)$$

$$w_k \sim N(0, Q)$$

$$v_k \sim N(0, R).$$

Then,

$$\begin{aligned} q(\theta, \theta(i)) &= E\{\log p(\theta, X, Y)|Y\} \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}(\Delta - \hat{x_0}\mu^T - \mu\hat{x_0}^T + \mu\mu^T)\} \\ &\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\Theta - \Psi A^T - \Pi B^T - A\Psi^T \\ &\quad +A\Phi A^T - B\Pi^T + B\Lambda B^T)\} \\ &\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}(\delta - \Omega C^T - C\Omega^T + C\Phi C^T)\} \\ &\quad +const., \end{aligned}$$

where

$$\Delta = E_t(x_0 x_0^T)$$

$$\hat{x_0} = E_t(x_0)$$

$$\Theta = \sum_{k=0}^{t-1} E_t(x_{k+1} x_{k+1}^T)$$

$$\Phi = \sum_{k=0}^{t-1} E_t(x_k x_k^T)$$

$$\Psi = \sum_{k=0}^{t-1} E_t(x_{k+1} x_k^T)$$

$$\Pi = \sum_{k=0}^{t-1} E_t(x_{k+1} z_k^T)$$

$$\Gamma = \sum_{k=0}^{t-1} E_t(x_k z_k^T)$$

$$\Lambda = \sum_{k=0}^{t-1} E_t(z_k z_k^T)$$

$$\Omega = \sum_{k=0}^{t-1} E_t(y_k x_k^T)$$

$$\delta = \sum_{k=0}^{t-1} (y_k y_k^T)$$

and

$$z_k = x_k \otimes x_k$$

Proof. We know that the dynamical systems are Markovian, and the joint likelihood function can be computed by using Beyes' rule as follows

$$\begin{aligned}
p(\theta, X, Y) &= p(y_1, ..., y_t, x_1, ..., x_t) \\
&= p(y_t, x_t | \{x\}_1^{t-1}, \{y\}_1^{t-1}) p(\{x\}_1^{t-1}, \{y\}_1^{t-1}) \\
&= p(y_t|x_t) p(x_t | \{x\}_1^{t-1}, \{y\}_1^{t-1}) p(\{y\}_1^{t-1}, \{x\}_1^{t-1}) \\
&= p(y_t|x_t) p(x_{t+1}|x_t) p(\{y\}_1^{t-1}, \{x\}_1^{t-1}) \\
&= ... \\
&= (\prod_{k=1}^{t} p(y_k|x_k)) p(\prod_{k=1}^{t} p(x_{k+1}|x_k)) p(x_0)\,.
\end{aligned}$$

By using the assumptions on x_0, w_k and v_k, the probability density functions $p(x_0)$, $p(x_{k+1}|x_k)$ and $p(y_k|x_x)$ are given by

$$p(x_0) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x_0 - \mu)^T V^{-1} (x_0 - \mu)\}$$

$$p(x_{k+1}|x_k) = \frac{1}{(2\pi)^{\frac{n}{2}}|Q|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x_{k+1} - Ax_k - Bz_k)^T Q^{-1}(x_{k+1} - Ax_k - Bz_k)\}$$

and

$$p(y_k|x_k) = \frac{1}{(2\pi)^{\frac{p}{2}}|R|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(y_k - Cx_k)^T R^{-1}(y_k - Cx_k)\}$$

where

$$z_k = x_k \otimes x_k \,.$$

Now, substituting these density functions in the log likelihood function and taking the logarithm of both sides we get

$$\begin{aligned} L &= \log p(y_1, ..., y_t, x_1, ..., x_t) \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}(x_0 - \mu)^T V^{-1}(x_0 - \mu) \\ &\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\sum_{k=0}^{t-1}(x_{k+1} - Ax_k - Bz_k)^T Q^{-1}(x_{k+1} - Ax_k - Bz_k) \\ &\quad -\frac{t}{2}\log|R| - \frac{1}{2}\sum_{k=0}^{t}(y_k - Cx_k)^T R^{-1}(y_k - Cx_k) + const \\ &= p(\theta, X, Y)\,. \end{aligned}$$

Taking the conditional expectation and applying the rule

$$E(X^T AX) = \mathrm{Tr}[AE(XX^T)], \tag{6.1}$$

we get

$$
\begin{aligned}
q(\theta,\theta(i)) &= E_t\{\log p(\theta,X,Y)\} \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}(V^{-1}E_t\{(x_0-\mu)(x_0-\mu)^T\} \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}\sum_{k=0}^{t-1}(E_t(x_{k+1}-Ax_k-Bz_k)(x_{k+1}-Ax_k-Bz_k)^T)\} \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=0}^{t}(E_t(y_k-Cx_k)(y_k-Cx_k)^T)\} \\
&\quad +const. \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}E_t(x_0x_0^T - x_0\mu^T - \mu x_0^T - \mu\mu^T)\} \\
&\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}\sum_{k=0}^{t-1}E_t(x_{k+1}x_{k+1}^T - x_{k+1}x_k^TA^T - x_{k+1}z_k^TB^T - Ax_kx_{k+1}^T \\
&\quad +Ax_kx_k^TA^T + Ax_kz_k^TB^T - Bz_kx_{k+1}^T + Bz_kx_k^TA^T + Bz_kz_k^TB^T)\} \\
&\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}\sum_{k=1}^{t}(E_t(y_ky_k^T - y_kx_k^TC^T - Cx_ky_k^T + Cx_kx_k^TC^T)\} \\
&\quad +const.
\end{aligned}
$$

Since odd moments of Gaussian random variables vanish,

$$E_t(x_kz_k^T) = 0. \tag{6.2}$$

Furthermore,

$$
\begin{aligned}
E_t(x_{k+1}z_k^T) &= E_t((Ax_k+Bz_k+w_k)z_k^T) \\
&= BE_t(z_kz_k^T).
\end{aligned}
$$

Put,

$$\Delta = E_t(x_0 x_0^T)$$

$$\hat{x_0} = E_t(x_0)$$

$$\Theta = \sum_{k=0}^{t-1} E_t(x_{k+1} x_{k+1}^T)$$

$$\Phi = \sum_{k=0}^{t-1} E_t(x_k x_k^T)$$

$$\Psi = \sum_{k=0}^{t-1} E_t(x_{k+1} x_k^T)$$

$$\Pi = \sum_{k=o}^{t-1} E_t(x_{k+1} z_k^T)$$

$$\Lambda = \sum_{k=0}^{t-1} E_t(z_k z_k^T)$$

$$\Omega = \sum_{k=0}^{t-1} E_t(y_k x_k^T)$$

$$\delta = \sum_{k=0}^{t-1} (y_k y_k^T)$$

Substituting these expressions, we get

$$\begin{aligned} q(\theta, \theta(i)) &= E\{\log p(\theta, X, Y)|Y\} \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}(\Delta - \hat{x_0}\mu^T - \mu\hat{x_0}^T + \mu\mu^T)\} \\ &\quad -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\Theta - \Psi A^T - \Pi B^T - A\Psi^T + A\Phi A^T - B\Pi^T + B\Lambda B^T)\} \\ &\quad -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}(\delta - \Omega C^T - C\Omega^T + C\Phi C^T)\} \\ &\quad +const. \end{aligned}$$

and therefore, the proof is completed.

Remark 6.2.2. From Theorem 6.2.1, we note that to compute $q(\theta, \theta(i))$, it is necessary to compute the quantities Θ, Φ, Ψ, Π and Λ. Indeed, Θ, Φ and Ψ may be calculated directly by using the bilinear Kalman filter and the bilinear Kalman smoother which have been introduced in Theorems 5.3.1. and 5.4.2., respectively. By using the notations introduced in Chapter 5, we have

$$\Theta = \sum_{k=0}^{t-1} E_t(x_{k+1}x_{k+1}^T) = \sum_{k=0}^{t-1}(x_{k+1}^t(x_{k+1}^t)^T + P_{k+1}^t).$$

$$\Phi = \sum_{k=0}^{t-1} E_t(x_k x_k^T) = \sum_{k=0}^{t-1}(x_k^t(x_k^t)^T + P_k^t).$$

and

$$\Psi = \sum_{k=0}^{t-1} E_t(x_{k+1}x_k^T) = \sum_{k=0}^{t-1}(x_{k+1}^t(x_k^t)^T + P_{k+1,k}^t)$$

where $P_{k+1,k}^t$ can be computed from Theorem 5.4.3.

We compute the quantity Π as follows,

$$\Pi = \sum_{k=0}^{t-1} E_t(x_{k+1}z_k^T) = \sum_{k=0}^{t-1}(x_{k+1}^t(z_k^t)^T + \ddot{P}_{k+1,k}^t)$$

where $\ddot{P}_{k+1,k}^t$ can be found by using Theorem 5.4.3. The term

$$\begin{aligned}\Lambda &= \sum_{k=0}^{t-1} E_t(z_k z_k^T) \\ &= \sum_{k=0}^{t-1}(z_k^t(z_k^t)^T + \dot{P}_k^t),\end{aligned}$$

where

$$\dot{P}_k^t = (V_k^t)\ddot{P}_k^t.$$

The next step of the EM algorithm is to maximize the function $q(\theta, \theta(i))$ with respect to the parameters vector θ.

Theorem 6.2.3. The function $q(\theta, \theta(i))$ which is given in Theorem 6.2.1. is maximized over θ provided that we choose

$$\mu = \hat{x_0}$$

$$V = \Delta$$

$$A = \Psi\Phi^{-1}$$

$$B = \Pi\Lambda^{-1}$$

$$Q = \tfrac{1}{t}(\Theta - \Psi^T\Phi^{-1}\Psi - \Pi\Lambda^{-1}\Pi^T)$$

$$C = \Omega\Phi^{-1}$$

and

$$R = t^{-1}\{\delta - \Omega\Phi^{-1}\Omega^T\}.$$

Proof. Define the following functions,

$$q_1(\mu, V) = -\frac{1}{2}\log|V| - \frac{1}{2}\mathrm{Tr}\{V^{-1}(\Delta - \hat{x_0}\mu^T - \mu\hat{x_0}^T + \mu\mu^T)\}$$

$$q_2(A,B,Q) = -\tfrac{t}{2}\log|Q| - \tfrac{1}{2}\mathrm{Tr}\{Q^{-1}(\Theta - \Psi A^T - \Pi B^T - A\Psi^T + A\Phi A^T$$
$$-B\Pi^T + B\Lambda B^T)\}$$

and

$$q_3(C,R) = -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}(\delta - \Omega C^T - C\Omega^T + C\Phi C^T)\}$$

Hence, the log likelihood function becomes

$$q(\theta,\theta(i)) = q_1(\mu,V) + q_2(A,B,Q) + q_3(C,R) + const.$$

Since $q_1(\mu,V)$ depends only on μ and V and since $q_2(A,B,Q)$ contains the parameters A, B and Q while $q_3(C,R)$ depends on the parameters C and R, the function $q(\theta,\theta(i))$ will be maximized by maximizing each of the subfunctions $q_1(\mu,V), q_2(A,B,Q)$ and $q_3(C,R)$ separately.

Now, from the function $q_1(\mu,V)$, we conclude that

$$\mu = \hat{x_0}. \tag{6.3}$$

Substituting this in $q_1(\mu,V)$, and then using the results of differentiation given in Section 2.8 (the partial derivative with the respect to V), and equating to zero, we get

$$\frac{\partial}{\partial V}q_1(\mu,V) = -\frac{1}{2}V^{-1} + \frac{1}{2}V^{-1}\Delta V^{-1} = 0$$

and thus,

$$V = \Delta.$$

By completing the square with respect to A, we obtain

$$\begin{aligned} q_2(A,B,Q) &= -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\Theta - (A-\Psi\Phi^{-1})\Phi(A-\Psi\Phi^{-1})^T \\ &\quad -\Pi B^T - B\Pi^T + B\Lambda B^T)\}, \end{aligned}$$

and so, we conclude that

$$A = \Psi\Phi^{-1}.$$

In a similar manner, we find the value of B as follows,

$$B = \Pi\Lambda^{-1}.$$

Substituting the values of A and B in $q_2(A,B,Q)$ we observe that

$$q_2(A,B,Q) = -\frac{t}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}\{Q^{-1}(\Theta - \Psi^T\Phi^{-1}\Psi - \Pi\Lambda^{-1}\Pi^T)\} \tag{6.4}$$

Applying the partial divertive of $q_2(A,B,Q)$ with respect to the parameter Q and applying the obtained result to zero yields

$$-\frac{t}{2}Q^{-1} + \frac{1}{2}Q^{-1}\{(\Theta - \Psi^T\Phi^{-1}\Psi - \Pi\Lambda^{-1}\Pi^T)\}Q^{-1} = 0$$

and hence,

$$Q = \frac{1}{t}(\Theta - \Psi^T\Phi^{-1}\Psi - \Pi\Lambda^{-1}\Pi^T).$$

Now, the function $q_3(C,R)$ can be reduced to

$$q_3(C,R) = -\frac{t}{2}\log|R| - \frac{1}{2}\mathrm{Tr}\{R^{-1}(\delta - (C-\Omega\Phi^{-1})\Phi(C-\Omega\Phi^{-1})^T)\}. \tag{6.5}$$

From the last equation, we find that

$$C = \Omega\Phi^{-1}. \tag{6.6}$$

Taking the partial derivative of $q_3(C, R)$ with respect to the variable R and then, equating it with zero, we have

$$-\frac{t}{2}R^{-1} + \frac{1}{2}R^{-1}[\delta - \Omega C^T - C\Omega^T + C\Phi C^T]R^{-1} = 0 \tag{6.7}$$

and therefore,

$$\begin{aligned} R &= t^{-1}[\delta - \Omega C^T - C\Omega^T + C\Phi C^T] \\ &= t^{-1}[\delta - \Omega\Phi^{-1}\Omega^T]. \end{aligned}$$

Chapter 7

SIMULATION AND APPLICATION TO THE LOTKA-VOLTERRA MODELS

7.1 Introduction

Applications are very important in this field of research as it illustrates the efficiency of the theoretical work and demonstrates its relevance to real life problems. Applications consist of transforming purely theoretical results to reality by applying them in

various applied fields. In this dissertation, we have developed a nonlinear state-space model of a bilinear class. The new development of our bilinear model generalizes some famous nonlinear models which had been applied in many fields of the real life, such as atmospheric studies, biology, chemistry, economics, engineering, neural networks and control systems [3,41,42,61]. We also derived new filter (the bilinear Kalman filter) and smoother (the bilinear Kalman smoother) algorithms that suit our bilinear model. Furthermore, we included a new technique to identify the parameters of the new bilinear model via another generalization of the expectation maximization (EM) algorithm, where the bilinear Kalman filter and the bilinear Kalman smoother are employed to estimate the parameters of the model.

In this chapter, we will apply our new results to the nonlinear estimation and, our method to the nonlinear identification of the Lotka-Volterra nonlinear model. This model has applications in various domains of life science, such as biology, chemistry, economic and neural networks [3,61]. In the next section, we introduce a general review and the mathematical framework for Lotka-Volterra model. In Section 7.3, the bilinear Kalman filter and the bilinear Kalman smoother are applied to simultaneously estimate states and parameters from noise data of a Lotka-Volterra system.

7.2 The nonlinear Lotka-Volterra model

The famous Lotka-Volterra equations play a fundamental role in the mathematical modeling of various ecological and chemical systems. The LotkaVolterra equations have been widely used in economic theory and neural networks. The LotkaVolterra predator-prey model was initially proposed by Alfred J. Lotka in the theory of auto-catalytic chemical reactions in 1910. Vito Volterra, who made a statistical analysis of fish catches in the Adriatic independently investigated the equations in 1926 [20]. There are two cases of Lotka-Volterra models; (i) competing species models and (ii) predator-prey models [3]. The Lotka-Volterra predator-prey equations, are a pair of first-order, non-linear, differential equations frequently used to describe the dynamics of biological systems in which two species interact; predator and its prey. It is of the form;

$$\frac{dx}{dt} = \lambda x - \beta xy \tag{7.1}$$

$$\frac{dy}{dt} = \beta xy - \gamma y \tag{7.2}$$

where y is the number of some predator and x is the number of corresponding prey; $\frac{dx}{dt}$ and $\frac{dy}{dt}$ represent the growth of the two populations against time; t represents the time; and λ, β and γ are parameters representing the interaction of the two species. The Lotka-Volterra competing species is a system of ordinary differential equations of

the form;

$$\frac{dx}{dt} = \lambda x(1-x) - \beta xy \tag{7.3}$$

$$\frac{dy}{dt} = \gamma y(1-y) - \beta xy. \tag{7.4}$$

Eq. (7.3) says that the population of species x grows according to a logistic law in the absence of species y (that is, when $y = 0$). In addition, the rate of growth of x is negatively proportional to y, representing competition between members of x and members of y. The larger the population y, the smaller the growth rate of x. Similarly, Eq. (7.4) describes the rate of growth for population y.

The Lotka-Volterra competing species model can be written in the form of our bilinear model (Eq. (5.1)) with

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \gamma \end{pmatrix},$$

and,

$$B = \begin{pmatrix} -\lambda & -\beta & 0 \\ 0 & -\beta & -\gamma \end{pmatrix}.$$

The same can be done for the Lotka-Volterra predator-prey model (7.1) and (7.2).

7.3 Simulation

In Chapter 5, we derived the bilinear Kalman filter and the bilinear Kalman smoother for an underlying nonlinear system of bilinear types. In Chapter 6 we presented the identification process for such systems via the EM algorithm. The nonlinear Lotka-Volterra model is a mathematical model which has wide applications in ecological and chemical systems. This model is a special case of our bilinear model. In this section we will simulate the bilinear Kalman filter and the bilinear Kalman smoother algorithms. This simulation demonstrates the utility and the usage of such algorithms to estimate the states of the system. On the other hand, we also explain the usage of the bilinear EM approach to maximum likelihood estimation which is proposed in this dissertation, where, the identification of the parameters for the nonlinear Lotka-Volterra model is done via the bilinear EM algorithm.

The simulations will enable us to see that the technique of the bilinear Kalman filter and the bilinear Kalman smoother used to estimate the states of the system and parameters for a nonlinear dynamical system relative to a linear measurement system, give a good approximation of the system states and also allow us to identify the parameters for such nonlinear dynamical systems. We conclude that the proposed algorithm and the present method for identification perform well. We will carry out the simu-

lation study in two steps using MATLAB codes. First, we simulate the performance of our bilinear Kalman filter and bilinear Kalman smoother (which were derived in Chapter 5). Secondly, we show the performance of using the bilinear EM approach in estimating the parameters of a nonlinear Lotka-Volterra model. The methodology of utilizing the bilinear EM algorithm with a nonlinear system was introduced in Chapter 6.

7.3.1 Simulation of the bilinear Kalman filter and the bilinear Kalman smoother

Here, we provide a simulation example in order to illustrate the utility of the bilinear Kalman filter and the bilinear Kalman smoother which were proposed in Chapter 5.

We consider the nonlinear Lotka-Voltera competition model in state-space form

$$x_{k+1} = Ax_k + Bz_k + w_k \tag{7.5}$$

$$y_{k+1} = Cx_k + v_k \tag{7.6}$$

where

$$A = \begin{pmatrix} 1 & 0.1 \\ -0.1 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} 0.01 & 0.01 & 0 \\ 0 & -0.01 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the bilinear term $z_k = x_k \otimes x_k$ (x_k is the state vector). The random noise w_k and v_k are uncorrelated with $w_k \sim (0, W)$ and $v_k \sim (0, V)$, where, $W = .0004I_2$ and $V = .0004I_2$. The initial state is $x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

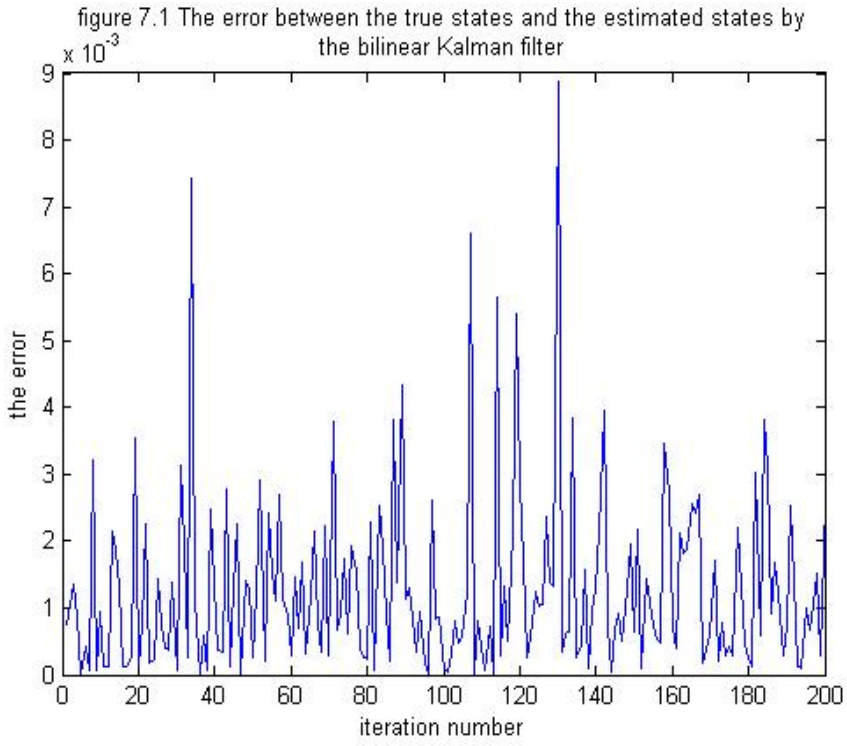
As usual, the error for the estimated quantities is required in order to state the reliability of the results. The error is provided by the covariance matrix to compute the difference between the true states and the estimated states. That is, if x_k is the true state, and x_k^t is the estimated state, then the estimation error can be computed by the relation

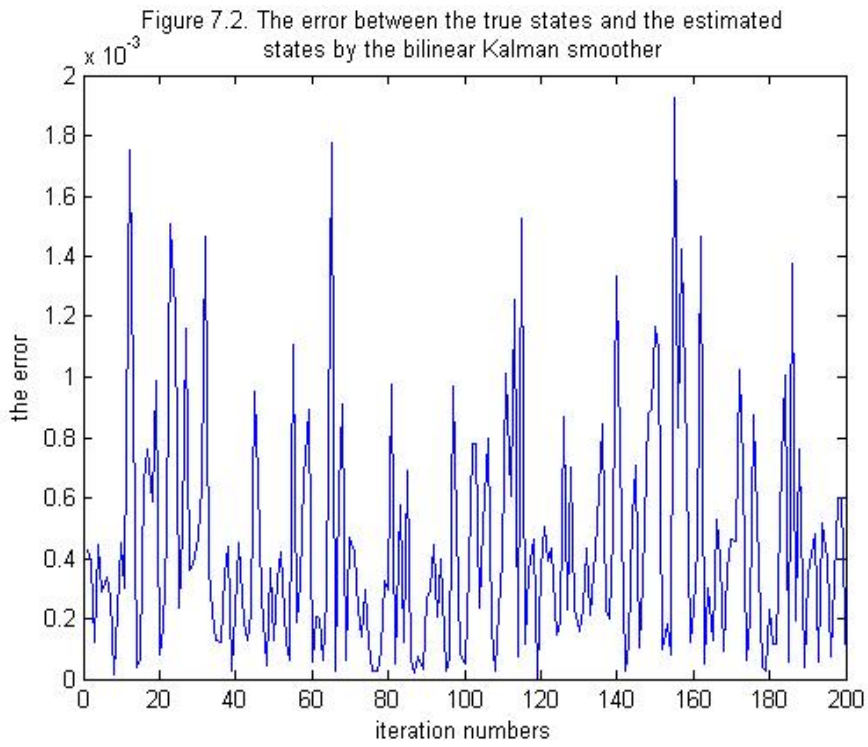
$$\epsilon = \| x_k - x_k^t \| = (x_k - x_k^t)^T (x_k - x_k^t). \tag{7.7}$$

Now, in the bilinear Kalman filter, the estimated state is x_{k+1}. The value of x_{k+1}^k can be obtained by applying the equations which are given in Theorem 5.3.1. After running the algorithm, we find that the error between the true states and the estimated states via the bilinear Kalman filter is very small. The error in the case of the bilinear

Kalman filter is given in Figures 7.1.

In the bilinear Kalman smoother, the estimated state is x_{k+1}^t, for $t > k+1$. The estimated state for the Lotka-Volterra model can be computed also by using the bilinear Kalman smoother which is given in Theorem 5.4.2. By applying the equations of this estimator to estimate the state x_k, we find that, the error between the true state and the estimated state is very small as illustrated in Figures 7.2. These results show that our new estimator (the bilinear Kalman filter and the bilinear Kalman smoother) work well since it is successfully applicable to a bilinear model due to the good results which we have obtained in the simulation.



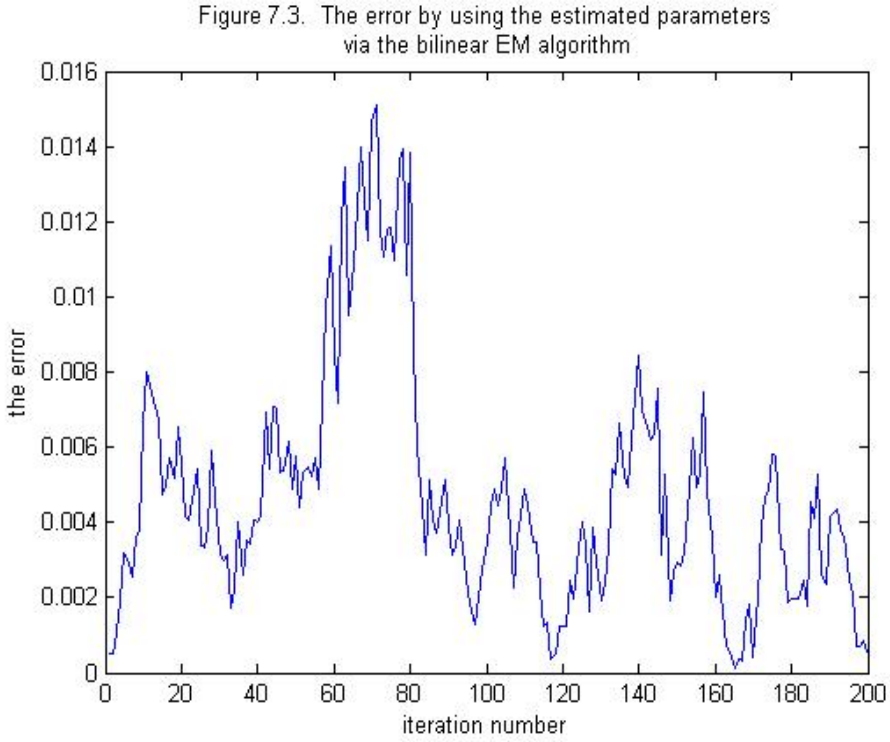


7.3.2 Simulation of the parameters estimation via the bilinear EM algorithm

We simulate the utility of the bilinear Kalman filter and the bilinear Kalman smoother for estimating the bilinear system parameters via the well-known bilinear EM approach. Here, we consider the Lotka-Volterra state-space model (7.5)–(7.6) where the parameters A, B, C, W and V are unknown. The initial value for the state

will be $x_0 = \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix}$, such that, $x_0 \sim N(\mu, P)$. The additive noises w_k and v_k are uncorrelated with $w_k \sim (0, W)$ and $v_k \sim (0, V)$, with initial estimates, W and V. The estimation for the system parameters of the model (7.5)–(7.6) is determined from this information about the model via the EM algorithm. We will start by initial guesses for such parameters, and they are updated recursively until convergence to the true system. That is, if the estimated parameters are very close to the true parameters, we will obtain a small error between the estimated state and the true state. In the E-step, we use Theorem 6.2.1. In the M-step we use Theorem 6.2.3.

In Figure 7.3, we display the error between the true state and the estimated state using the estimated values of the parameters which we have obtained from the simulation via the bilinear EM algorithm. The resulting small errors indicate that using the bilinear EM approach to estimate the exact values of the parameters is very reliable .



7.3.3 Comparison between the Bilinear Kalman filter and the ensemble Kalman filter (EnKF) and the extended Kalman filter (EKF)

The extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) are famous filters which are used with the nonlinear models. When we apply theses filters (EKF and EnKF) with the nonlinear Lotka-Volterra model, we observe good results. However, when we compare the results by applying the bilinear Kalman filter with

the corresponding results by applying the extended Kalman filter (EKF) and the corresponding results by applying the ensemble Kalman filter (EnKF), we find that the results of the error between the true states and the estimate states by applying the bilinear Kalman filter are smaller than the error from applying the EKF and the EnKF. That is, the bilinear Kalman filter gives a more accurate result than the EKF and the EnKF. These results are described in the following figures.

Figure 7.4. The error between the true states and the estimated states by the bilinear Kalman filter

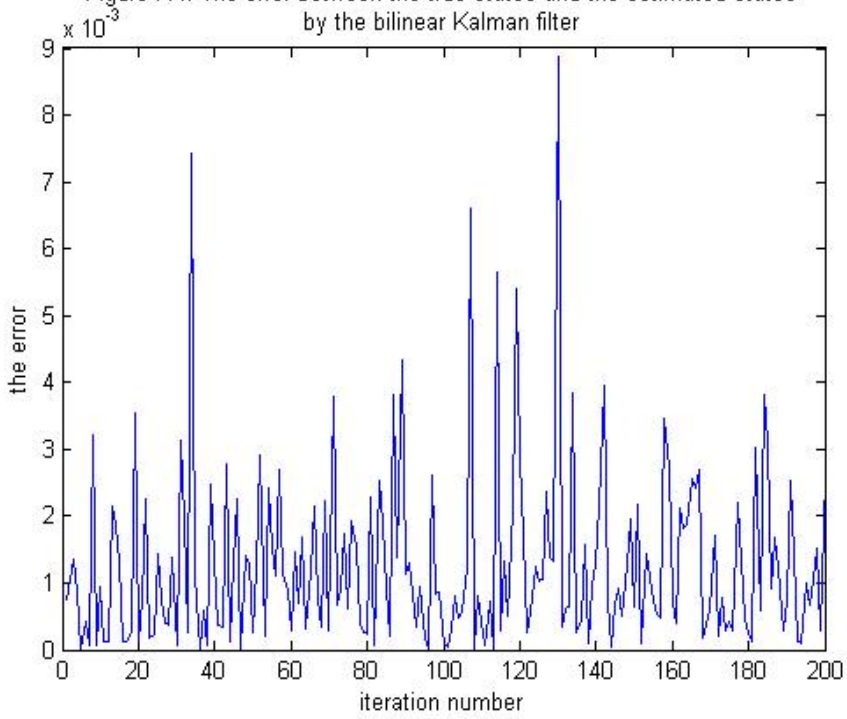


Figure 7.5 The error between the true states and the estimated states by the extended Kalman filter

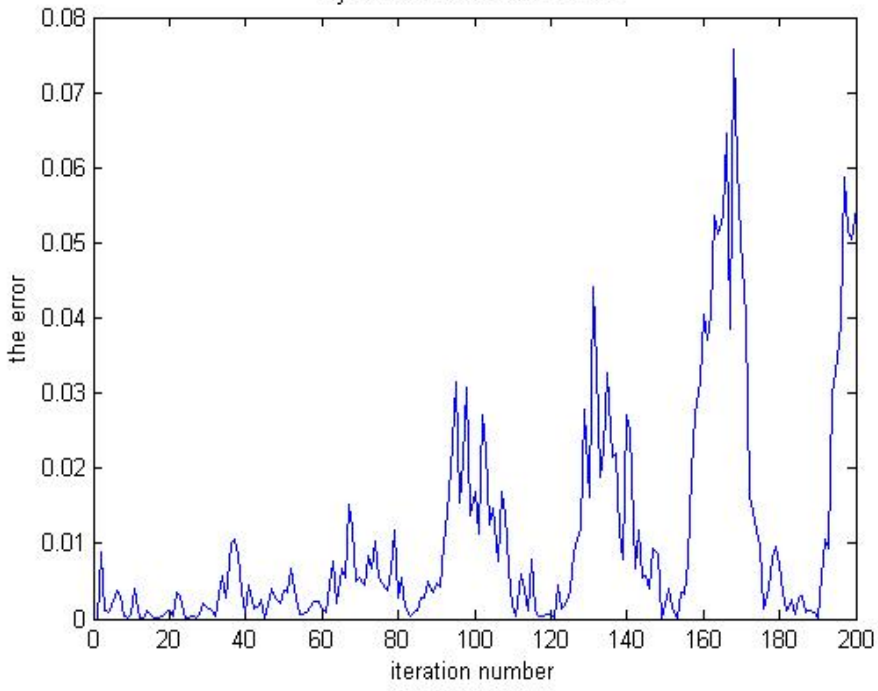
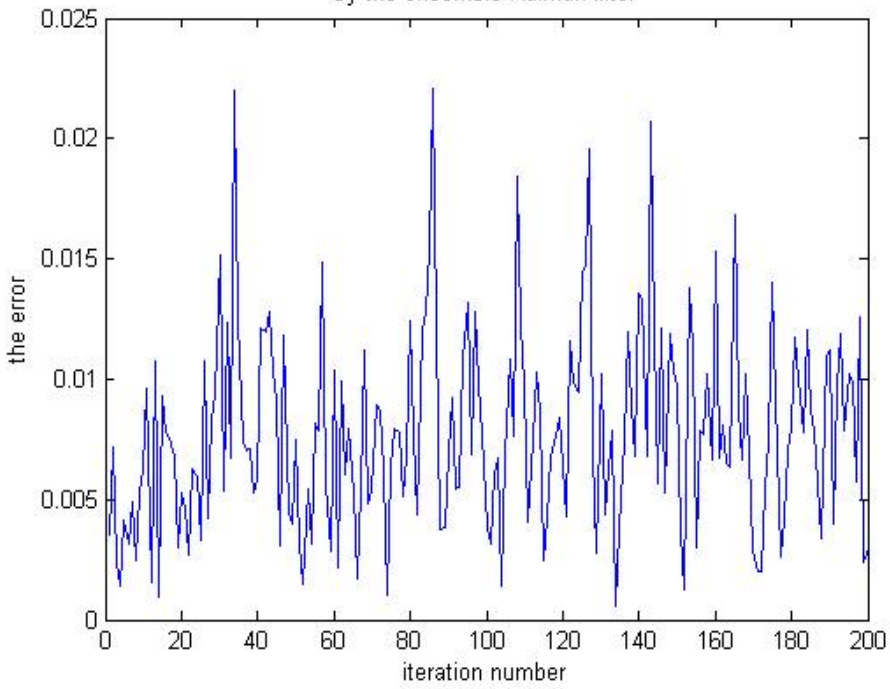


Figure 7.6. The error between the true states and the estimated states by the ensemble Kalman filter



Bibliography

[1] J. L. Anderson, S. L. Anderson, A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, Monthly Weather. Rev. 127, 2741-2758, 1999.

[2] B.D. Anderson and J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[3] K. T. Alligood, T. D. Sauer, J. A. Yorke, An Introduction to Dynamical Systems, Springer-Verlag New York, 1996.

[4] C. F. Ansley and R. Kohn, A geometrical derivation of the fixed interval smoothing algorithm, Biomelrilca, 69, 486-487, 1982.

[5] J. Bendat, Nonlinear System Analysis and Identification from Random Data, Wiley Interscience, 1990.

[6] F. Bouttier, P. Courtier, Data assimilation concepts and methods, Training Course Lecture Series, ECMWF, 1999.

[7] C. K. Chui and G. Chen, Kalman Filtering, New York: Springer, 1999.

[8] F. E. Daum, Exact Finite Dimensional Nonlinear Filters, IEEE Trans. Automatic Control, vol. AC-31, 7, 616-622, 1986.

[9] A. Dempster, N. Laird, and D. Rubin, Maximum Likelihood from Incomplete Data via The EM Algorithm, J. Royal Statist. Soc., ser. B, vol. 39, pp. 1-38, 1977.

[10] G. S. Duane, J. J. Tribbia, and J. B.Weiss, Synchronicity in Predictive Modeling: a New View of Data Assimilation, Nonlin. Processes Geophys., 13, 601-612, 2006.

[11] G. Evensen, Advanced Data Assimilation for Strongly Nonlinear Dynamics, Monthly Weather Review, vol. 125, 1342-1354, 1997.

[12] G. Evensen. Data assimilation: The Ensamble Klamn Filter, Springer, 2nd edition, 2009.

[13] R. Furrer and T. Bengtsson, Data Assimilation and The Ensemble Kalman Filter, Geophysical Stat. Proj. presented for NCAR, 2003.

[14] G. Galanis et. al., Applications of Kalman Filter Based on Nonlinear Functions to Numerical Weather Predictions, Annales Geophesicea, 24, 1-10, 2006.

[15] G. Galanis and M. anadranistakis, A One-Dimension Kalman Filter for Correction of Near Surface Temperature Forecasts, Meteorol. Appl. 9, 437-441, 2002.

[16] M. S. Gerwal and A. P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB, John Wily and Sons Inc., 2001.

[17] S. Gibson, A. Wills and B. Ninness, Maximum-Likelihood Parameter Estimation of Bilinear Systems, IEEE Transactions on Automatic Control 50, 1581-1596, 2005.

[18] S. Gibson and B. Ninness, Robust Maximum-Likelihood Estimation of Multivariable Dynamics Systems, Automatica 41, 1667-1682, 2005.

[19] S. Gillijns, et.al., What is The Ensemble Kalman Filter and How Well Does It Work?, Proceeding of the 2006 American Control System Conference, IEEE, 4448-4453, 2006.

[20] N. S. Goel, et al., On The Volterra and Other Nonlinear Models of Interacting Populations, Academic Press Inc., 1971.

[21] G. Golub and C. Van Loan, The Matrix Computations, Third edition, The Johns Hopkins University Press, 1996.

[22] Q. P. Ha, H. Trinh. State and Input Simultaneous Estimation for a Class of Nonlinear System, Automatica, 1779-1785, 2004.

[23] J. D. Hamilton, Time Series Analysis, Princeton university press, 1994.

[24] P. L. Houtekamer, H. L. Mitchell, A Sequential Ensemble Filter for Atmospheric Data Assimilation, Monthly Weather Rev., 129, 123-137, 2001.

[25] J Harlim and A J Majda, Filtering Nonlinear Dynamical Systems with Linear Stochastic Models, Nonlinearity 21, 1281-1306, 2008.

[26] A. C. Harvey, Forecasting Structural Time Series Models and The Kalman Filter, Cambridge University Press, New York, 1989.

[27] A. H. Jazwinski, Stochastic Processes and Filtering Theory, New York: Academic, 1970.

[28] S. J. Julier et. al. A New Approach for Filtering Nonlinear System, Proceeding of the 1995 American Control Conference, 3, 1628-1632, 1995.

[29] S. Juiler and J. Uhlamn, A New Extension of The Kalman Filter, Int Symp Aerospace Defense Sensing Simul and Controls, Volume: 3, Publisher: Citeseer, 1997.

[30] J. P. Kaipio and E. Somersalo, Statistical and Computational Inverse Problems, New York: Springer, 2005

[31] T. Kailath, A. sayed, and B. Hassibi, Linear Estimation, Prentice -Hall, Inc., 2000.

[32] R. Kalman, A New Approach to Linear Filtering and Prediction Problems, Trans. ASME, Ser. D, J. Basic Eng. ser. 82, 35-45, 1960.

[33] R. Kalman, R. Bucy, New Results in Linear Filtering and Prediction Theory, Trans. ASME, Ser. D, J. Basic Eng. ser. 83 , 95-108, 1961.

[34] V. Katari, S. Malireddi, S. Kanth, S.Bendapudi, G. Panda, Adaptive Nonlinear System Identification Using Comprehensive Learning PSO, IEEE, 434-439, 2008.

[35] G. Kerschen, Keith Worden, A. F. Vakakis, J. Golinval. Past, Present and Future of Nonlinear System Identification in Structural Dynamics, Mechanical Systems and Signal Processing 20, 505-592, 2006.

[36] T. Kollo and D. V. Rosen, Advanced Multivariate Statistics with Matrices, Springer, 2005.

[37] A. J. Krener, Bilinear and Nonlinear Realizations of Input-Output Maps, SIAM J. Control, 13, 827-834, 1975.

[38] E. Lehmann, Theory of Point Estimation, New York: Wiley, 1983.

[39] J. Liu and Eugenia Kalnay. Estimating Observation Impact Without Adjoint Model in an Ensemble Kalman Filter, Q. J. R. Meteorol. Soc. 134, 1327-1335, 2008.

[40] L. Ljung, System Identification: Theory for the User, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[41] E.N. Lorenz, Predictibility: A Problem Partly Solved, In Proc. Seminar on Predictability,1, 1-18, ECMWF, Reading, Berkshire, UK, 1996.

[42] E.N. Lorenz and K.E. Emanuel. Optimal Sites for Supplementary Weather Observations: Simulations with a small model, J. Atmos. Sci., 55, 399-414, 1998.

[43] E. N. Lorenz, Designing Chaotic Models. Jour. of the Atmo. Sci., 62, 1574-1588,2005

[44] X. Luo, Recursive Bayesian Filters for Data Assimilation, D. Phil. Thesis, University of Oxford, 2009.

[45] H. Lutkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2005.

[46] A. J. Majda and X. Wang, Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows, Cambridge University Press, 2006.

[47] G. McLachlan and T. Krishnan, " EM Algorithm and Extensions. New York: Wiley, 1996.

[48] V. Monbet, et. al., Survey of Stochastic Models for Wind and Sea State Time Series. Probabilistic Engineering Mechanics 22, 113-126, 2007.

[49] C. P. Mracek, J. R. Cloutier, and C. A. D'Souza, A New Technique for Nonlinear Estimation, Proc. IEEE Conf. Contr. App., 338-343,1996.

[50] M. Norgaard, N. K. Poulsen, O. Ravn, New Developments in State Estimation for Nonlinear Systems, Automatica 36 , 1627 - 1638, 2000.

[51] P. M. Pardalos, and V.Yatsenko, Optimization and Control of Bilinear Systems: Theory, Algorithms and Applications, Springer, 2008.

[52] J. Peter, A. Richard, A. Davis ,Introduction to Time Series and Forecasting, Second Edition, Springer, 2002.

[53] D. T. Pham , Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. Mon. Weather Rev 129, 1194-1207, 2001.

[54] M. B. Preistley, Non-Linear and Non-Stationary Time Series Analysis, Academic Press Inc., 1989.

[55] D. Roy , Z.E. Musielak, Generalized Lorenz Models and Their Routes to Chaos. III. Energy-Conserving Horizontal and Vertical Mode Truncations, Chaos, Solitons and Fractals 33, 1064-1070, 2007.

[56] D. Roy , Z.E. Musielak,, Generalized Lorenz Models and Their Routes to Chaos. II. Energy-Conserving Horizontal Mode Truncations, Chaos, Solitons and Fractals 31, 747-756, 2007.

[57] J. Sarangapani, Neural Network Control of Nonlinear Discrete-Time System, Taylor and Francis, 2006.

[58] R. Shumway and D. Stoffer, Time Series Analysis and Its Applications, New York: Springer-Verlag, 2006.

[59] R. Shumway and D. Stoffer, An Approach to Time Series Smoothing and Forecasting Using The EM Algorithm, J. Time Series Anal., 3, 253-264, 1982.

[60] T. Soderstrom, and P. Stoica (1989), System Identification, Prentice Hall. New York.

[61] S. H. Strgate, Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering, Presure Books l.l.c.,1994.

[62] H. Tanizaki, Nonlinear Filtering: Estimation and Applications, Springer-Verlag, 1996.

[63] E. A. Wan and R. van der Merwe, The Unscented Kalman Filter for Nonlinear Estimation, Proc. The IEEE AS-SPCC Symposium, 153-158, 2000.

[64] M. West and J. Harrisson, Bayesian Forecasting and Dynamic Model, 2nd edition, Springer-Verlag, New York, 1997.

[65] N. Wiener, Nonlinear Problems in Random Theory, Boston, MA: The M.I.T. Press, 1958.

[66] C. K. Wikle and L. M. Berliner, A Bayesian tutorial for data assimilation, Physica D 230, 1-16, 2007.

[67] R. Yamaguchi and T. Higuchi, State-Space Approache with The Maximum Likelihood Principle to Identify The System Generating Time-Course Gene Expression Data of Yeast, Int. J. Data Mining and Bioinformatics, 1, 77-87, 2006.

[68] S. Yang, D. Baker, H. Li, G. Nagpal, and E. Kalnay, Data Assimilation as Aynchronization of Truth and Model: Experiments with The Three-Variable Lorenz System. Journal of the Atmospheric Sciences; 63, 2340-2354, 2006.

VITA

ABDULLAH EQAL AL-MAZROOEI

- Born in Khulais, Saudi Arabia, on July, 1976.
- Faculty member, Mathematics Department, Faculty of Science, Taibah University, Al-Madinah Al-Munawwarah, since 2002.
- Joined, in 2007, Ph.D. program in the Department of Mathematics and statistics, King Fahd University of Petroleum and Minerals.
- Received M.S. degree in Mathematics from King Abdul Aziz University, Jeddah, in 2006.
- Received B.S. degree in Mathematics from King Abdul Aziz University, in 1998.
- **Contact Details:**

- Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, P. O. Box 8581, Dhahran 31261, Saudi Arabia.
- E-mail Address: abo.mhmd@hotmail.com

- **Permanent Address:** Department of Mathematics, Taibah University, P.O.Box 30002, Al-Madinah Al-Munawarah, Saudi Arabia.