# AN INTERVAL KALMAN FILTER, INTERVAL EM ALGORITHM WITH APPLICATION TO WEATHER PREDICTION

BY

**OBAID JEFAIN ALGAHTANI**

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 OBAID JEFAIN ALGAHTANI 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. Mohamed A. El-Gebeily
Dissertation Committee Chairman

Dr. Jaafar Al-Mutawa
Co-Chairman

R.P. Agarwal
Prof. Ravi P. Agarwal
Member

Dr. Hattan Tawfiq
Department Chairman

Prof. Ashfaque H. Bokhari
Member

Dr. Salam Zummo
Dean of Graduate Studies

Dr. Kassem Mustapha
Member

1/6/11
Date

To my parents, my wife, , my children, my brothers,

and my teachers.

## ACKNOWLEDGEMENTS

Praise be to Allah, Lord of the worlds, the Almighty, with whose gracious help it was possible to accomplish this work.

I am grateful to Professor Mohamed A. El-Gebeily for all his help that has made my Ph.D. study a fulfilling experience. I am grateful for his dedication, the direction, and patience in helping me, learn and guiding this research to a completion. Working with him in a motivating environment was really a learning experience.

I extend my deepest gratitude to my thesis co-advisor Dr. Jaafar Al-Mutawa, thesis committee members Prof. R. P. Agarwal, Prof. A. H. Bokhari and Dr. K. Mustapha for their support, encouragement and suggestions.

It is my sense of duty to thank all faculty members of the Department of Mathematics and Statistics at the King Fahd University of Petroleum and Minerals for providing a genial research environment and a wonderful research working place.

I would also like to thank my parents and my wife for their kindness, pray and being the source for my strength. To my brothers, specially Arnan, as well as family and friends whose names were not mentioned for their love, friendship, and encouragement.

## TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES viii

ABSTRACT(ENGLISH) ix

ABSTRACT(ARABIC) x

1. INTRODUCTION **1**

1.1 Motivation ..... 1

1.2 Objectives ..... 4

1.3 Outline of Thesis ..... 5

2. PRELIMINARIES **6**

2.1 Introduction ..... 6

2.2 Matrices ..... 7

2.2.1 Operations and Functions ..... 7

2.2.2 Derivatives ..... 11

2.3 Probability ..... 16

2.3.1 The Probability Distribution Function ..... 16

2.3.2 The Expectation and Variance of Random Variables ..... 17

2.3.3 Joint Probability Distribution Function ..... 19

2.3.4 Conditional Probability ..... 21

2.4 Interval Analysis ..... 23

2.4.1 Concepts and Properties of Intervals ..... 24

2.4.2 Basics of Interval Arithmetic ..... 25

2.4.3 Interval Functions ..... 27

2.4.4 Interval Matrices ..... 30

2.5 Least-Squares Estimation ..... 31

2.6 The EM-Algorithm ..... 35

3. THE KALMAN FILTER..... **38**

3.1 Introduction ..... 38

3.2 Gaussian State Space Models ..... 39

3.3 Deriving the Kalman Filter from Least Squares Estimation ..... 41

3.4 Summary of the Kalman Filter and the Kalman Smoother ..... 47

3.5 Ensemble Kalman Filter ..... 51

3.6 Identification of Linear State Space Models ..... 52

4. CONVEXITY INTERVAL ANALYSIS ..... **61**

4.1 Introduction ..... 61

4.2 Convexity Interval Arithmetic ..... 62

4.3 Interval Matrices, Interval Linear Systems ..... 65

4.3.1 Interval Matrices ..... 65

4.3.2 Interval Linear System ..... 68

4.4 Interval Random Variables ..... 72

4.4.1 Measurable Set Valued Maps ..... 72

4.4.2 Normally Distributed Interval Random Variables ..... 73

5. THE CONVEXITY INTERVAL KALMAN FILTER ..... **87**

5.1 Introduction ..... 87

5.2 Discussion ..... 88

5.3 An Interval State Space Model ..... 89

5.4 Deriving the Convexity Interval Kalman Filter ..... 90

5.5 Derivation of Convexity Interval Kalman Smoother ..... 94

5.6 Summary of the Convexity Interval Kalman Filter and Smoother ..... 97

5.7 Simulation ..... 99

5.8 Identification of Interval State Space Models ..... 102

5.8.1 Motivation ..... 102

5.8.2 Interval Parameters Identification ..... 104

5.8.3 Simulation ..... 110

5.9 Application to Weather Prediction ..... 114

BIBLIOGRAPHY **125**

VITAE **134**

## LIST OF FIGURES

3.1 Kalman Filter (n=50).....50

3.2 Kalman Filter (n=100).....50

4.1 Exact Solution Region for Interval Linear System.....70

5.1 Interval Kalman Filter (n=50) ..... 101

5.2 Interval Kalman Filter (n=100)..... 101

5.3 Identification of Interval Parameters (n=20)..... 113

5.4 Identification of Interval Parameters (n=50)..... 113

5.5 Case 1 of Interval Prediction..... 117

5.6 Case 2 of Interval Prediction..... 118

5.7 Case 3 of Interval Prediction..... 120

5.8 Case 4 of Interval Prediction..... 122

5.9 Case 5 of Interval Prediction..... 123

5.10 Comparison between dimensions..... 124

## ABSTRACT

**Name:** Obaid Jefain Julaighim Algahtani.

**Title:** An Interval Kalman Filter, Interval EM Algorithm with Application to Weather Prediction.

**Degree:** Doctor of Philosophy.

M**ajor Field:** Mathematics.

**Date of degree:** May, 2011.

The Kalman filter (KF) gives the optimal estimates of the unknown state vector in time series linear stochastic state space model (SSM). If we have observed data of the state space model, we can identify the unknown parameters using system identification techniques. One way to do this is called Expectation Maximization (EM).

In the system certain elements such as the coefficient matrices are not precisely known or gradually change with time. One way to take these uncertainties into account is to allow interval state space models and extend the statistical concepts to interval setting.

The traditional Kalman filter technique can not be used directly when the system parameters are not precisely known or change with time. So, it is important to introduce an interval Kalman filter(IKF) to handle the current situation. Also, the interval parameters could be identified from a given record of interval measurements.

# ملخص الرسالة

**الاســـم** : عبيد جفين آل جليغم القحطاني

**عنوان الرســالة** : فلتر كالمان للفترات وخوارزمية تعظيم التوقع للفترات مع تطبيقات على تنبؤ الطقس

**الدرجــة العلمية:** دكتوراه

**التخـصــص** : الرياضيات

**تــاريخ التخرج** : مايو ٢٠١١

فلتر كالمان يعطي أفضل التقديرات لمتجه الحالة في نموذج فضاء الحالة الخطي والسلاسل الزمنية. إذا كان لدينا بيانات مرصودة لنموذج فضاء الحالة فانه يمكننا تحديد الوسائط المجهولة لهذه النماذج الرياضية باستخدام تقنية نظام التحديد وإحدى الطرق هي خوارزمية تعظيم التوقع.

بعض العناصر في تلك الأنظمة مثل مصفوفات المعاملات قد تكون غير معلومة بدقة أو تتغير تدريجياً بالنسبة للزمن. إحدى الطرق لتمثيل عدم التحديد حسابياً هو تعميم نموذج فضاء الحالة إلى نموذج فضاء الحالة للفترات . هذا التعميم يتطلب تعميم المفاهيم الإحصائية بحيث تطبق على مجموعة الفترات.

فلتر كالمان لايمكن تطبيقه مباشرة على نموذج فضاء الحالة للفترات لذا نحتاج إلى تقديم نموذج فلتر كالمان للفترات بحيث يتعامل مع هذا الوضع. أيضاً يمكننا تعميم خوارزمية تعظيم التوقع إلى خوارزمية تعظيم توقع للفترات لتحديد معاملات نموذج فضاء الحالة للفترات بمعلومية مجموعة من القياسات التي تكون على شكل فترات.

# Chapter 1

# INTRODUCTION

## 1.1 Motivation

A physical model is said to be in state space form if it is completely specified by two basic equations. These two equations are known as the measurement and transition equations [11,12,31,58]. The state space model is a linear when the observations in the measurement equation are a linear function of the state vector and, in the transition equation, the state vector is itself a linear function of the state vector in the previous time period.

Kalman filter was first proposed by Rudolf Kalman in the year 1960 as optimal esti-

mation filter for the linear state space model [32]. To estimate the state, the Kalman filter (KF) has access to measurements. Those measurements are linearly related to the state and are corrupted by noise. The KF processes all available measurements to estimate the state. It uses knowledge of the system and sensor dynamics, probabistic description of the system and measurement noise, and any available data about the initial values of the state. Some authors have called the discovery of the KF one of the greatest discoveries in the twentieth century [3,4,31,58].

The KF equations could be derived from many approachs. In [33], the conventional KF was derived by minimizing a quadratic cost function. This is intimately related to the least square estimation which is widely studied in control and optimization theories [31,65]. Kalman described his filter using state space technique which enables the filter to be used as either a smoother, a filter or a predictor [23,31,32,58,65].

The KF is applied to a wide range of tracking, manufacturing processes, aircraft, ships and navigation problems [4,8,58].

The Kalman smoother is an efficient algorithm for **E**-step in the Expectation-Maximization (**EM**) algorithm for linear Gaussian state space models. The **EM** algorithm is an iterative technique for obtaining the maximum likelihood estimation and consists of two steps: the **E**-step and the **M**-step [10,26,58,59]. In the **E**-step, we compute the conditional expectation of the log maximum likelihood function and in the **M**-step, we maximize the expected likelihood function with respect to unknown parameters in

the state space model (see Sec. 2.6).

Interval arithmetic is an arithmetic defined on sets of intervals rather than sets of real numbers. Modern development of interval arithmetic began with R. E. Moore's book [45]. Hansen and Smith [21] started the use of interval arithmetic in matrix computations. After this motivation and inspiration, several authors such as Alefeld and Herzberger [1], Hansen et al ([20]), Jaulin et al [24], Neumaier [48] and Rohn [54], have studied interval matrices.

In general interval analysis, some algebra properties do not hold. For example, the distributive law is not true. So, the inverse of an interval matrix is not well defined. J. Rohn defined the inverse of an interval matrix as the narrowest interval matrix containing the set of all inverse matrices included in the original interval matrix [54]. In the existing literature, no method is available to find the exact solution for the linear system of interval equations but there are methods available for computing the smallest box containing the exact solution of the system of interval linear equations [50].

In the state space model, certain elements, such as the coefficient matrices, are not precisely known or gradually change with time. One way to take these uncertainties into account, is to allow interval state space model presentation. The interval state space model requires an extension of some statistical concepts in interval setting. The traditional Kalman filter technique could not be used directly to handle the interval

state space model. G. Chen, J. Wang and L. Shieh wrote a paper about the interval Kalman filter (IKF) in 1997 in which most of the concepts from the ordinary case were extended in a strait forward manner to the interval case [18]. As well, this definition for IKF was introduced in the 4th edition of a published book titled "Kalman Filtering with Real Time Applications" (2009)[8].

## 1.2 Objectives

We extend the state space model concept to the interval state space model. In order to do this, we need to extend the statistical concepts to interval settings, e.g., interval random variables, interval expectation, interval variance, interval covariance, interval conditional expectation, interval conditional variance and investigate the algebraic properties of interval random variables.

Moreover, we introduce a definition of the determinant and inverse of interval matrices using a new convexity approach. This definition will enable us to obtain the computation of an interval inverse matrix.

In addition, we rigorously derive the interval Kalman filter using the definitions and statistical properties. Also, we identify interval parameters of interval state space model from a given record of interval measurements using a generalization of the **EM**

algorithm. Finally, we present a simulation for the identification of interval state space model and the weather prediction experiment.

## 1.3 Outline of Thesis

The organization of the thesis is as follows: in the second chapter a brief description of matrix algebra, probability theory, interval analysis, least square estimation and Expectation-Maximization(**EM**) algorithm are given. In chapter three, we introduce Gaussian state space model, deriving the Kalman filter and the identification of parameters of linear state space model. In chapter four, we present convexity interval analysis, definitions, interval linear systems and interval random variables. In chapter five, we introduce interval state space model, the convexity interval Kalman filter, convexity interval Kalman smoother, identification of the interval parameters of interval state space model and a simulation of interval parameter estimation for interval linear state space model. We also come up with an interval prediction model for weather.

# Chapter 2

# PRELIMINARIES

## 2.1 Inroduction

In this chapter, we will introduce some fundamental concepts that are necessary for proper understanding of interval Kalman filter and identification of parameters of interval state space model discussed in this thesis. We begin with an introduction to matrix algebra which is necessary to explain how to derive the Kalman filter equations. Then, we discuss brief background of certain basic concepts in probability theory. We will extend these concepts to interval setting in chapter 4. In addition, we review some preliminary results on interval analysis are needed throughout our research. Most

interval analysis preliminary results could be found in [1]. Since the Kalman filter could be derived from linear least square estimation (LSE), we introduce a summary of ordinary least square estimation. For more details of LSE, see [31,33,65]. Finally, we present the maximum likelihood parameter estimation and the **EM** algorithm. We will use the EM algorithm to identify the parameters of linear state space model in section 3.6.

## 2.2 Matrices

In this section we introduce some operations, formulas, inequalities and derivatives in matrix algebra which will be helpful in our research. More details about matrices, could be found in books on linear algebra and matrix theory [16].

### 2.2.1 Operations and Functions

**Definition 1** ***(Positive Definite and Semi-definite Matrices):*** *An $n \times n$ real matrix $A$ is positive definite if*

$$x^T Ax > 0, \forall x \neq 0 \in \mathbb{R}^n. \tag{2.1}$$

*A is positive semi-definite if*

$$x^T Ax \geq 0, \forall x \neq 0 \in \mathbb{R}^n. \tag{2.2}$$

**Remark** If $A, B \in \mathbb{R}^{n \times n}$, we will use the notation

$$A > B$$

when the matrix $A - B$ is positive definite and

$$A \geq B$$

when the matrix $A - B$ is positive semi-definite.

**Definition 2** ***(The trace of a matrix):*** *The trace of* $A = [a_{ij}]_{n \times n}$*, denoted by* $\mathrm{tr}A$*, is defined as the sum of its diagonal elements, namely:*

$$\mathrm{tr}A = \sum_{i=1}^{n} a_{ii}.$$

**Lemma 3** *The trace of matrices have these properties*

$$\begin{aligned} \mathrm{tr}A &= \sum_i \lambda_i, \lambda_i \in \mathrm{eig}(A) && (2.3) \\ \mathrm{tr}A^T A &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 && (2.4) \\ E[\mathrm{tr}A] &= \mathrm{tr}E[A] && (2.5) \end{aligned}$$

*where* $\mathrm{eig}(A)$ *is the eigenvalues of* $A$ *and* $E(A)$ *is the expectation of* $A$*.*

It follows from (2.3) that if $A$ is positive definite $\mathrm{tr}A > 0$.

**Definition 4** ***(Inverses of Matrix)****: The inverse of a matrix $A \in \mathbb{R}^{(n\times n)}$ is a matrix $A^{-1} \in \mathbb{R}^{(n\times n)}$ such that*

$$AA^{-1} = A^{-1}A = I_n.$$

The inverse matrix can be constructed, using the adjoint matrix $\mathrm{Adj}(A)$, by

$$A^{-1} = \frac{1}{\det(A)}\mathrm{Adj}(A), \det(A) \neq 0,$$

where $\det(A)$ is the determinate of $A$ and

$$\mathrm{Adj}(A) = (\mathrm{cof}(A))^T$$

and the matrix of cofactors is defined by

$$(\mathrm{cof}(A, i, j)) = (-1)^{i+j}\det([A]_{ij}),$$

where the submatrix $[A]_{ij}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the $ith$ row and the $jth$ column of $A$.

**Lemma 5** ***(Matrix Schwarz Inequality)*** *Let $P$ and $Q$ be $(m \times n)$ and $(m \times k)$ matrices, respectively, such that $P^TP$ is nonsingular, then*

$$Q^TQ \geq (P^TQ)^T(P^TP)^{-1}(P^TQ). \tag{2.6}$$

**Proof.** Let $(Q-PS)^T(Q-PS) \geq 0$ and choose $S = (P^TP)^{-1}(P^TQ)$. So, we get

$$\begin{aligned}
(Q-PS)^T(Q-PS) &\geq 0 \\
&\Rightarrow Q^TQ \geq S^TP^TQ + (P^TQ)^TS - S^T(P^TP)S \\
&= S^TP^TQ + (P^TQ)^T(P^TP)^{-1}(P^TQ) - [(P^TP)^{-1}(P^TQ)] \\
(P^TP)(P^TP)^{-1})P^TQ & \\
&= S^TP^TQ + (P^TQ)^T(P^TP)^{-1}(P^TQ) \\
-[(P^TP)^{-1}(P^TQ)]^TP^TQ & \\
&= (P^TQ)^T(P^TP)^{-1}(P^TQ) + (P^TQ)^T(P^TP)^{-1}(P^TQ) \\
-(P^TQ)^T(P^TP)^{-1}P^TQ & \\
&= (P^TQ)^T(P^TP)^{-1}(P^TQ)
\end{aligned}$$

■

**Lemma 6** *(**Completing the Square**) Let $X, A, B, C \in \mathbb{R}^{n\times n}$, with $A$ invertable then*

$$XAX^T - BX^T - (BX^T)^T + C = (X - BA^{-1})A(X - BA^{-1})^T + C - BA^{-1}B^T. \quad (2.7)$$

**Definition 7** ***Characteristic Values**: For any variable $\lambda$, the polynomial*

$$P_A(\lambda) = \det[A - \lambda I] = \Sigma_{i=0}^n a_i\lambda^i$$

*is called the characteristic polynomial of A. The roots of $P_A(\lambda)$ are called the characteristic values (or eigenvalues) of A and denoted by eig(A).*

**Characteristic Vectors**: For each real characteristic value $\lambda_i$ of a real symmetric $A$, there is a corresponds a characteristic vector $e_i \neq 0$ and $Ae_i = \lambda_i e_i$.

### 2.2.2 Derivatives

**Derivatives:** Let $F : X \to Y$, where $X, Y$ are finite dimensional spaces with norm $\|.\|$. The derivative of $F$ is a linear operator $F'$ such that

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

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

**Computational procedure** : To compute $F'(x)$ we extract the linear part of $F(x+h)$, or equivalently, compute $\frac{\partial F(x+th)}{\partial t}|_{t=0}$.

**Chain Rule**: Let $G : X \to Y, F : Y \to Z$, we define $F(G(x)) = F \circ G : X \to Z$. $F(G(x))' \in \mathcal{L}(X, Z)$ and

$$F(G(x))'h = F'(G(x))G'(x)h, \forall h \in X$$

where $F'(Y) \in \mathcal{L}(Y, Z)$, $G'(x) \in \mathcal{L}(X, Z)$.

**Lemma 8** *Let $A$ and $B$, be square matrices, then*

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

**Proof.** : Let $C = AB$, then $C_{ii} = \sum_{j=1}^{n} a_{ij}b_{ji}$. Therefore

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

■

**Lemma 9** *Let $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{(n\times n)}$, then*

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

**Proof.**

$$x^T Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j$$

$$= \sum_{i,j=1}^{n} (xx^T)_{ij} a_{ij}$$

by Lemma 8

$$x^T Ax = \mathrm{tr}(xx^T A^T).$$

■

**Formulas for derivatives**

**Lemma 10** *Let $A$ be a linear operator, then*

$$\frac{d}{dx}(Ax)h = Ah, \forall h \in \mathbb{R}^n.$$

**Lemma 11**

$$(\frac{d}{dA}A^{-1})M = -A^{-1}MA^{-1}, \forall M \in \mathbb{R}^{n\times n}$$

**Proof.**

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

Using Neumann formula gives

$$(A+tM)^{-1} = A^{-1} - tA^{-1}MA^{-1} + O(t^2) \Rightarrow \frac{d}{dt}(A+tM)^{-1}|_{t=0} = -A^{-1}MA^{-1}.$$

■

**Theorem 12** *Let $|A|$ denotes the determinate of $A \in \mathbb{R}^{(n\times n)}$: $|.| : \mathbb{R}^{(n\times n)} \to \mathbb{R}$, then*

$$\frac{d}{dA}|A|M = \mathrm{tr}(\mathrm{adj}AM^T), M \in \mathbb{R}^{n\times n} \tag{2.8}$$

**Proof.** We will use the fact that $|A|$ is an $n$-linear form when regarded as a function on 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.

$$|A + tM| = |a_1 + tM_1, a_2 + tM_2, ..., a_n + tM_n| \tag{2.9}$$
$$= |a_1, a_2, ..., a_n| + t \sum_{i=1}^{n} |a_1, ..., M_i, a_{i+1}, ..., a_n| + O(t^2)$$
$$= |A| + t \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} A_{ij} + O(t^2),$$

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

Therefore by lemma 8, we have

$$\frac{d}{dt}|A + tM|_{t=0} = \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} A_{ij} = \mathrm{tr}(\mathrm{adj} A M^T) \tag{2.10}$$

■

**Corollary 13**

$$\frac{d}{dt} Log|A|M = \frac{1}{|A|} \mathrm{tr}(\mathrm{adj} A M^T) = \mathrm{tr}(A^{-1} M^T). \tag{2.11}$$

**Lemma 14**

$$\frac{d}{dA}(x^T A x) M = x^T M x, \forall M \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n.$$

**Lemma 15** *Let $A, M \in \mathbb{R}^{(n \times n)}$, then*

$$\frac{d}{dt}(\mathrm{tr}(A)) M = \mathrm{tr} M. \tag{2.12}$$

**Proof.** Notice that $\text{tr} : \mathbb{R}^{(n\times n)} \to \mathbb{R}$ is a linear operator. So, use Lemma 10 directly.

■

Combining the above results we can show that

$$\frac{d}{dA}(A^{-1}B)M = -A^{-1}MA^{-1}B \tag{2.13}$$

$$\frac{d}{dA}\text{tr}(A^{-1}B)M = \text{tr}(-A^{-1}MA^{-1}B) = -\text{tr}(A^{-1}MA^{-1}B) \tag{2.14}$$

$$\frac{d}{dA}(A^{-1}BA)M = A^{-1}BM - A^{-1}MA^{-1}BA \tag{2.15}$$

**Computation of critical points**

• Let

$$F(x) = A + (x - x_0)'B(x - x_0),$$

where $B$ is invertible.

Then the critical points of $F(x)$ are computed as

$$\frac{d}{dx}(F(x))v = 2v^TB(x - x_0) \tag{2.16}$$

$$2v^TB(x - x_0) = 0, \forall v \in \mathbb{R}^n$$

$$\Rightarrow B(x - x_0) = 0 \Rightarrow x - x_0 \in \text{ker}B,$$

$$\therefore x - x_0 = 0; x = x_0.$$

• Let

$$F(A) = \log|A| + \text{tr}(AP) + x^TAx,$$

where $A$ is invertible. Then,

$$\frac{d}{dA}F(A)M = \mathrm{tr}(A^{-1}M^T) + \mathrm{tr}(MP) + x^T M x \tag{2.17}$$

$$= \mathrm{tr}(A^{-1}M^T + P^T M^T + xx^T M^T)$$

$$= \mathrm{tr}([A^{-1} + P^T + xx^T]M^T)$$

$$= 0, \forall M \in \mathbb{R}^{(n\times n)} iff A^{-1} + P^T + xx^T = 0$$

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

## 2.3 Probability

### 2.3.1 The probability Distribution Function

**Definition 16** *let $S$ be a sample space and $X : S \to \mathbb{R}$ be a random variable. For each measurable set $A \subset \mathbb{R}$, define $P$ : events $\to [0, 1]$, where each event is a set $\{s \in S : X(s) \in A \subset \mathbb{R}\}$ or, briefly, $\{X \in A\}$, subject to the following conditions:*

*(1)$P(X \in A) \geq 0$ for any measurable set $A \subset \mathbb{R}$,*

*(2)$P(X \in \mathbb{R}) = 1$, and*

*(3) for any countable sequence of pairwise disjoint measurable sets $A_i$ in $\mathbb{R}$,*

$$P(X \in \cup A_i) = \sum_{i=1}^{\infty} P(X \in A_i).$$

*$P$ is called the probability distribution function of the random variable $X$.*

If there exists an integrable function $f$ such that

$$P(X \in A) = \int_A f(x)dx \tag{2.18}$$

for all measurable sets $A$, we say that $P$ is a continuous (as opposed to discrete) probability distribution and $f$ is called the probability density function of the random variable $X$.

**Definition 17** *The probability density function $f$ given by*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\{\frac{-(x-\mu_x)^2}{2\sigma_x^2}\}, \sigma_x > 0, \mu_x \in \mathbb{R} \tag{2.19}$$

*is called the Gaussian probability density function, and $P$ is called Gaussian distribution of the random variable $X$.*

It is completely determined by $\mu_x$ and $\sigma_x$. Hence, we use the notation: $X \sim N(\mu_x, \sigma_x^2)$.

### 2.3.2 The Expectation and Variance of Random Variables

**Definition 18** *Let $X$ be an $m$-dimensional random variable. The expectation of $X$ indicates the mean of the values of $X$, and is defined by*

$$E\{X\} = \int_{-\infty}^{\infty} xf(x)dx. \tag{2.20}$$

Note that $E(X)$ is an $m$-vector for any $m$-dimensional random variable $X$ with probability density function $f$. For the normal distribution, using the substitution $y = (x - \mu_x)/(\sqrt{2}\sigma_x)$, we have

$$\begin{aligned} E\{X\} &= \textstyle\int_{-\infty}^{\infty} x f(x) dx \\ &= \textstyle\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma_x} \exp\{\frac{-(x-\mu_x)^2}{2\sigma_x^2}\} dx \\ &= \mu_x \frac{1}{\sqrt{\pi}} \textstyle\int_{-\infty}^{\infty} e^{-} y^2 dy \\ &= \mu_x. \end{aligned} \tag{2.21}$$

**Definition 19** *The variance of $X$ is defined by*

$$\mathrm{Var}\{X\} = E\{X - E\{X\}\}^2 = \textstyle\int_{-\infty}^{\infty} (x - E\{X\})^2 f(x) dx. \tag{2.22}$$

For the Gaussian distribution, using the substitution $y = (x - \mu_x)/\sqrt{2}\sigma_x)$, we have

$$\begin{aligned} \mathrm{Var}\{X\} &= \textstyle\int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} \textstyle\int_{-\infty}^{\infty} (x - \mu_x)^2 exp\{\frac{-(x-\mu_x)^2}{2\sigma_x^2}\} dx \\ &= 2\sigma_x^2 \frac{1}{\sqrt{\pi}} \textstyle\int_{-\infty}^{\infty} y^2 e^{-y^2} dy \\ &= \sigma_x^2, \end{aligned} \tag{2.23}$$

where we have used the equality $\int_{-\infty}^{\infty} y^2 e^{-y^2} dy = \frac{\sqrt{\pi}}{2}$.

### 2.3.3 Joint Probability Distribution Function

Let $X = [X_1, X_2, ..., X_n]$ be a random vector whose components are random variables where $X_i(s) \in \mathbb{R}, s \in S$.

**Definition 20** *Let $P$ be a continuous probability distribution function of $X$. That is,*

$$P(X_1 \in A_1, ..., X_n \in A_n) = \int_{A_1} ... \int_{A_n} f(x_1, ..., x_n) dx_1 ... dx_n, \tag{2.24}$$

*where $A_1, ..., A_n$ are measurable sets in $\mathbb{R}$ and $f$ an integrable function. $f$ is called a joint probability density function of $X$ and $P$ is called a joint probability distribution function of $X_1, X_2, ..., X_n$.*

**Definition 21** *Let $X$ and $Y$ be random $n-$and $m-$vectors, respectively.*

*The covariance of $X$ and $Y$ is defined by the $n \times m$ matrix*

$$\text{Cov}(X, Y) = E[(X - E\{X\})(Y - E\{Y\})^T]. \tag{2.25}$$

When $Y = X$, we have the variance matrix, which is sometimes called a covariance matrix of $X$, Var($X$) =Cov($X, X$).

**Lemma 22** *The expectation, variance, and covariance have the following properties:*

$$E(AX + BY) = AE(X) + BE(Y) \tag{2.26}$$

$$E((AX)(BY)^T) = A(E(XY^T))B^T \tag{2.27}$$

$$\mathrm{Cov}(X, Y) = (\mathrm{Cov}(Y, X))^T \tag{2.28}$$

$$\mathrm{Cov}(X, Y) = E(XY^T) - E(X)(E(Y))^T \tag{2.29}$$

$$\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z), \tag{2.30}$$

*where $A$ and $B$ are constant matrices of appropriate dimensions.*

If

$$f(x) = \frac{1}{(2\pi)^{n/2}(detR)^{1/2}} \exp\{-\frac{1}{2}(x - \mu_x)^T R^{-1}(x - \mu_x)\} \tag{2.31}$$

where $\mu_x$ is a constant $n-$vector and $R$ is a covariance symmetric positive definite matrix, we say that $f(x)$ is a Gaussian probability density function of $X$.

We can show by using a substitution as that used for the scalar case of equations (2.21) and (2,23), that

$$E\{X\} = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} x_1 x_2 ... x_n f(X) dx_1 dx_2 ... dx_n = \mu_x, \tag{2.32}$$

and

$$\mathrm{Var}\{X\} = E\{X - \mu_x)(X - \mu_x)^T = R. \tag{2.33}$$

### 2.3.4 Conditional Probability

**Definition 23** *The conditional probability of $X_1 \in A_1$ given $X_2 \in A_2$ is defined by*

$$P(X_1 \in A_1 | X_2 \in A_2) = \frac{P(X_1 \in A_1, X_2 \in A_2)}{P(X_2 \in A_2)}. \tag{2.34}$$

Let $f(x_1|x_2)$ denote the probability density function corresponding to $P(X_1 \in A_1|X_2 \in A_2)$. $f(x_1|x_2)$ is called the conditional probability density function corresponding to the conditional probability distribution function $P(X_1 \in A_1|X_2 \in A_2)$.

Bayes' formula states that

$$f(x_1, x_2) = f(x_1|x_2)f_2(x_2) = f(x_2|x_1)f(x_1). \tag{2.35}$$

This formula also holds for random vectors $X_1$ and $X_2$.

$X$ and $Y$ are said to be independent if $f(\mathbf{x}|\mathbf{y}) = f_1(\mathbf{x})$ and $f(\mathbf{y}|\mathbf{x}) = f_2(\mathbf{y})$, and $X$ and $Y$ are said to be uncorrelated if $\mathrm{Cov}(X, Y) = 0$. It is easy to see that if $X$ and $Y$ are independent then they are uncorrelated. Indeed, if $X$ and $Y$ are independent then $f(\mathbf{x}, \mathbf{y}) = f_1(\mathbf{x})f_2(\mathbf{y})$. Indeed, by property (2.29) we find that $\mathrm{Cov}(X, Y) = 0$ .

Similar to the definitions of expectation and variance, the conditional expectation of $X$ under the condition that $Y = y$ is defined to be

$$E(X|Y = y) = \int_{-\infty}^{\infty} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x} \tag{2.36}$$

and the conditional variance of $X$, under the condition that $Y = y$ to be

$$\mathrm{Var}(X|Y=\mathbf{y}) = \int_{-\infty}^{\infty}[\mathbf{x} - E(X|Y=\mathbf{y})][\mathbf{x} - E(X|Y=\mathbf{y})]^T f(\mathbf{x}|\mathbf{y})d\mathbf{x}. \qquad (2.37)$$

Suppose that

$$E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

and

$$\mathrm{Var}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}.$$

From (2.31), we have

$$f(\mathbf{x},\mathbf{y}) = f\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\right)$$

$$= \frac{1}{(2\pi)^{n/2}\left(\det\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}\right)^{1/2}}$$

$$.\exp\left\{-1/2\left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)^T \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}^{-1} \left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)\right\}.$$

We find that[8]

$$
\begin{aligned}
& f(\mathbf{x}|\mathbf{y}) = \tfrac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} \\
& = \tfrac{1}{(2\pi)^{n/2}(det\tilde{R})^{1/2}} \exp\left\{-\tfrac{1}{2}(\mathbf{x}-\tilde{\mu}^T)\tilde{R}^{-1}(\mathbf{x}-\tilde{\mu})\right\}, \text{where} \\
& \tilde{\mu} = \mu_x + R_{xy}R_{yy}^{-1}(\mathbf{y}-\mu_y), \\
& \tilde{R} = R_{xx} - R_{xy}R_{yy}^{-1}R_{yx}.
\end{aligned} \tag{2.38}
$$

So, by rewriting $\tilde{\mu_x}$ and $\tilde{R}$, we get

$$E(X|Y=\mathbf{y}) = E(X) + \mathrm{Cov}(X,Y)\mathrm{Var}^{-1}(Y)(\mathbf{y}-E(Y)) \tag{2.39}$$

and

$$\mathrm{Var}(X|Y=\mathbf{y}) = \mathrm{Var}(X) + \mathrm{Cov}(X,Y)\mathrm{Var}^{-1}(Y)\mathrm{Cov}(Y,X). \tag{2.40}$$

## 2.4 Interval Analysis:

In this section, we provide some preliminary results on interval arithmetic and interval analysis that are needed throughout this thesis.

The birth of modern interval arithmetic was marked by the appearance of the book "Interval Analysis" by Ramon E. Moore in 1966 [45]. Hansen and Smith [21] started the use of interval arithmetic in matrix computations. We will follow the notations and definitions from [1](Alferld).

### 2.4.1 Concepts and Properties of Intervals

**Definition 24** *An interval* $[x]$ *is a closed and connected subset of* $\mathbb{R}$*; it is characterized by its lower and upper bounds* $\underline{x}$ *and* $\overline{x}$ *as* $[\underline{x}, \overline{x}]$.

Firstly, we introduce some useful concepts and properties of intervals:

**Equality**: Two intervals, $[\underline{x_1}, \overline{x_1}]$ and $[\underline{x_2}, \overline{x_2}]$, are said to be equal if and only if $\underline{x_1} = \underline{x_2}$ and $\overline{x_1} = \overline{x_2}$.

**Intersection:** The intersection of two intervals, $[\underline{x_1}, \overline{x_1}]$ and $[\underline{x_2}, \overline{x_2}]$ is defined by

$$[\underline{x_1}, \overline{x_1}] \cap [\underline{x_2}, \overline{x_2}] = [\max\{\underline{x_1}, \underline{x_2}\}, \min\{\overline{x_1}, \overline{x_2}\}].$$

**Union**: The union of two non-disjoint intervals, $[x_1]$ and $[x_2]$

$$[\underline{x_1}, \overline{x_1}] \cup [\underline{x_2}, \overline{x_2}] = [\min\{\underline{x_1}, \underline{x_2}\}, \max\{\overline{x_1}, \overline{x_2}\}].$$

**Inequality:** The interval $[x_1]$ is said to be less than (similarly, greater than) the interval $[x_2]$, iff $\overline{x_1} < \underline{x_2}$; otherwise, they cannot be compared. The relations $\leq$ and $\geq$: are not defined for intervals.

**Inclusion:** The interval $[x_1]$ is said to be included in $[x_2]$ iff $\underline{x_2} \leq \underline{x_1}$ and $\overline{x_1} \leq \overline{x_2}$.

### 2.4.2 Basics of Interval Arithmetic

Let $[x] = [\underline{x}, \overline{x}]$ and $[y] = [\underline{y}, \overline{y}]$ be real compact intervals and $o$ one of the basic operations 'addition', 'subtraction', 'multiplication' and 'division', respectively, for real numbers, that is $o \in \{+, -, *, /\}$. Then we define the corresponding operations for intervals $[x]$ and $[y]$ by

$$[x]o[y] = \{xoy | x \in [x], y \in [y]\},$$

where we assume $0 \notin [y]$ in case of division. $[x]o[y]$ can be represented by using only the bounds of $[x]$ and $[y]$ as the following :

$$\begin{aligned} &a.[x] + [y] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}] \\ &b.[x] - [y] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}] \\ &c.[x] * [y] = [\min(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \bar{x}\bar{y}), \max(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \bar{x}\bar{y})] \\ &d.\tfrac{[x]}{[y]} = [\min(\tfrac{\underline{x}}{\underline{y}}, \tfrac{\underline{x}}{\overline{y}}, \tfrac{\overline{x}}{\underline{y}}, \tfrac{\overline{x}}{\overline{y}}, \max(\tfrac{\underline{x}}{\underline{y}}, \tfrac{\underline{x}}{\overline{y}}, \tfrac{\overline{x}}{\underline{y}}, \tfrac{\overline{x}}{\overline{y}})], 0 \notin [y]. \end{aligned} \tag{2.41}$$

Equations (2.41) show that the set $\mathbb{IR}$ of real compact intervals is closed with respect to the operations $\{+, -, *, /\}$.

**Definition 25** *We define the the center and the width of interval* $[x]$ *as the following*

$$\begin{aligned} c([x]) &= \tfrac{(\underline{x}+\overline{x})}{2} \\ w([x]) &= \overline{x} - \underline{x}. \end{aligned} \tag{2.42}$$

If $[x]$ consists only of the element $x$, then we identify the real number $x$ with the degenerate interval $[x, x]$ keeping the real notation, i.e., $x = [x, x]$. In this way one recovers at once the real numbers $\mathbb{R}$ and the corresponding real arithmetic when restricting $\mathbb{IR}$ to the set of degenerate real intervals equipped with the arithmetic defined in (2.41). Unfortunately, $(\mathbb{IR}, +, *)$ is neither a field nor a ring. The structures $(\mathbb{IR}, +)$ and $(\mathbb{IR}/\{0\}, *)$ are commutative semigroups with the neutral elements 0 and 1, respectively, but they are not groups. A nondegenerate interval $[x]$ has no inverse with respect to addition or multiplication. Even the distributive law has to be replaced by the so-called subdistributivity

$$[x] * ([y] + [z]) \subseteq [x] * [y] + [x] * [z]. \tag{2.43}$$

For example, let $[x] = [-2, 2]$, $[y] = 1$ and $[z] = [-1]$,then

$$[x]*([y]+[z]) = [-2, 2](1-1) = 0 \subset [-4, 4] = [-2, 2]*1+[-2, 2]*(-1) = [x]*[y]+[x]*[z].$$

Also,

$$[-2, 2] + (-[-2, 2]) = [-2, 2] + [-2, 2] = [-4, 4] \neq 0,$$

which means that $-[-2, 2]$ not the inverse of $[-2, 2]$ with respect to addition.

Equality holds in equation (2.43) in some important particular cases, for instance if $[x]$ is degenerate or if $[y]$ and $[z]$ lie on the same side with respect to 0.

From (2.41) it follows that the operations for intervals are inclusion monotone in the

following sense:

$$[x] \subseteq [z], [y] \subseteq [w] \Rightarrow [x] \circ [y] \subseteq [z] \circ [w]. \tag{2.44}$$

### 2.4.3 Interval Functions

**Definition 26** *If $f$ is a continuous real valued function, then we define the standard interval function $W(f)$ as*

$$W(f, [x]) = \{f(x), x \in [x]\} = [\min_{x\in[x]} f(x), \max_{x\in[x]} f(x)] \tag{2.45}$$

*which are extensions of the corresponding real functions.*

The standard interval functions are inclusion monotone, i.e., they satisfy

$$[x] \subseteq [y] \Rightarrow W(f, [x]) \subseteq W(f, [y]). \tag{2.46}$$

**Definition 27** *Let $f : D \subset \mathbb{R} \to \mathbb{R}$ be given by a mathematical expression $f(x)$ which is composed of finitely many elementary operations $+, -, *, /$. If one replaces the variable $x$ by an interval $[x] \subset D$ and if one can evaluate the resulting interval expression following the rules in (2.41) and (2.44) then one gets again an interval. It is denoted by $f([x])$ and is called an interval arithmetic evaluation of $f$ over $[x]$.*

From (2.43) and (2.45) the interval arithmetic evaluation turns out to be inclusion monotone, i.e.,

$$[x] \subseteq [y] \Rightarrow f([x]) \subseteq f([y]). \tag{2.47}$$

From (2.46) we obtain

$$x \in [x] \Rightarrow f(x) \in f([x]), \tag{2.48}$$

hence,

$$W(f, [x]) \subseteq f([x]). \tag{2.49}$$

This formula is called the Fundamental Theorem of Interval Arithmetic.

**Example:** Suppose

$$f(x) = \frac{x}{1-x}, x \notin \{0, 1\},$$

and the interval $[x] = [2, 3]$. We could also write $f(x)$ as

$$f(x) = \frac{1}{1/x - 1}, x \notin \{0, 1\}.$$

In either case

$$W(f, [x]) = W(f, [2, 3]) = [-2, -\frac{3}{2}].$$

Denoting the first expression for $f$ by $f^{(1)}$ and the second by $f^{(2)}$ we get

$$f^{(1)}([x]) = f^{(1)}([2, 3]) = [-3, -1]$$

$$f^{(2)}([x]) = f^{(2)}([2, 3]) = [-2, -\frac{3}{2}] = W(f, [x]).$$

This example confirms (2.49) and shows that the quality of the inclusion of $f([x])$ in $W(f,[x])$ is strongly dependent on how the expression for $f(x)$ is written.

**Definition 28** *The distance between two intervals is defined by*

$$q([x],[y]) = \max\{|\underline{x} - \underline{y}|, |\overline{x} - \overline{y}|\}. \tag{2.50}$$

**Definition 29** *The absolute value of an interval* $[x]$ *is defined by*

$$|[x]| = \max\{|x| : x \in [x]\} = \max\{|\underline{x}|, |\overline{x}|\}. \tag{2.51}$$

The map $q$ defines a metric in $\mathbb{IR}$ .

**Theorem 30** *Suppose an expression* $\widetilde{f}\left(x^{(1)},\ldots,x^{(n)}\right)$ *is formed from the real function* $f$ *of the real variable* $x$ *by replacing every occurrence of* $x$ *with a new variable* $x_i$, $1 \leq i \leq n$. *Suppose the expression* $\widetilde{f}$ *satisfies the Lipschitz condition*

$$\left|\widetilde{f}\left(x^{(1)},\ldots,y^{(i)},\ldots,x^{(n)}\right) - \widetilde{f}\left(x^{(1)},\ldots,z^{(i)},\ldots,x^{(n)}\right)\right| \leq l_i \left|y^{(i)} - z^{(i)}\right| \tag{2.52}$$

*for each* $i$ *with* $1 \leq i \leq n$. *Then*

$$q\left(f\left(X\right), f\left(Y\right)\right) \leq \left(\sum_{i=1}^{n} l_i\right) q\left(X,Y\right). \tag{2.53}$$

### 2.4.4 Interval Matrices

We introduce $m \times n$ interval matrices $[\mathbf{A}] = ([a_{ij}])$ with interval entries $[a_{ij}], i = 1, 2, ..., m, j = 1, 2, ..., n,$ and interval vectors $[x] = ([x_j])$ with $n$ components $[x]$. We denote the corresponding sets by $\mathbb{IR}^{(m\times n)}$ and $\mathbb{IR}^n$, respectively. Trivially, $[\mathbf{A}]$ coincides with the interval matrix $[\underline{A}, \overline{A}] = \{B \in \mathbb{R}^{(m\times n)} : \underline{A} \leq B \leq \overline{A}\}$, where $\underline{A} = (\underline{a_{ij}}), \overline{A} = (\overline{a_{ij}}) \in \mathbb{R}^{(m\times n)}$. Since interval vectors can be identified with $n \times 1$ matrices, a similar property holds for them. The null matrix $\mathbf{0}$ and the identity matrix $\mathbf{I}$ have the usual meaning, $e$ denotes the vector $e = (1, 1, ..., 1)^T \in \mathbb{R}^n$.

Operations between interval matrices and between interval vectors are defined in the usual manner. They satisfy an analogue of (2.46)-(2.48). For example,

$$\{Ax : A \in [A], x \in [x]\} \subseteq [A][x] = (\textstyle\sum_{j=1}^{n} [a_{ij}][x_j]) \in \mathbb{IR}^m \tag{2.54}$$

if $[\mathbf{A}] \in \mathbb{IR}^{(m\times n)}$ and $[\mathbf{x}] \in \mathbb{IR}^n$.

An interval matrix $[\mathbf{A}] \in \mathbb{IR}^{(n\times n)}$ is called nonsingular if it contains no singular real $n\times n$ matrix. The distance, the center, the width and the absolute value in (2.42), (2.50) and (2.51) can be generalized to interval matrices and interval vectors, respectively. Note that the results are real matrices and vectors, respectively, as can be seen, e.g., for

$$q([\mathbf{A}], [\mathbf{B}]) = [q([\mathbf{a}_{ij}], [\mathbf{b}_{ij}])] \in \mathbb{R}^{(m\times n)}$$

if $[\mathbf{A}], [\mathbf{B}] \in \mathbb{IR}^{(m\times n)}$.

**Theorem 31** ***(Triangular Inequality)****: Let* $X = [x_{il}, x_{iu}], Y = [y_{il}, y_{iu}]$ *be two intervals, then*

$$|X + Y| \leq |X| + |Y|. \tag{2.55}$$

**Proof.**

$$\begin{aligned} |X+Y| &= \max\{|x_{il}+y_{il}|, |x_{iu}+y_{iu}|\} \\ &\leq \max\{|x_{il}|+|y_{il}|, |x_{iu}|+|y_{iu}|\} \\ &= \max\{|x_{il}|, |x_{iu}|\} + \max\{|y_{il}|, |y_{iu}|\} \\ &= |X| + |Y|. \end{aligned}$$

■

## 2.5 Least-Squares Estimation:

In our thesis, we use the least square estimation approach to derive the Kalman filter equations. In this section we mainly follow [8] to present the ordinary least square estimation (LSE).

Consider the observation equation

$$y_t = Hx_t + v_t \tag{2.56}$$

where $x_t$ is the state vector, $y_t$ the data vector, $v_t$ is a random vector of zero mean Gaussian white noise, namely: $E(v_t) = 0$ and $E(v_t v_j^T) = R_t \delta_{tj}$, where $R_t \delta_{tj} = \text{Cov}(v_t, v_j)$ and $\delta_{tj} = 1$ if $t = j$ and 0 if $t \neq j$ with $R_t$ being symmetric and positive definite.

Let $H$ be given. Our goal is to obtain an optimal estimate $\hat{u}_t$ of the state vector $x_t$ from the information $y_t$. When the data is contaminated with noise, we will minimize the quantity:

$$F(u_t, W_t) = E((y_t - Hu_t)^T W_t (y_t - Hu_t)) \tag{2.57}$$

over all $n-$ vectors $u_t$ where $W_t$ is a positive definite and symmetric weight matrix. That is, we wish to find a $\hat{u}_t = \hat{u}_t(W_t)$ such that

$$F(\hat{u}_t, W_t) = \min_{u_t} F(u_t, W_t). \tag{2.58}$$

Also, we wish to determine the optimal weight $\hat{W}_t$ in a sense to be explained later. To find $\hat{u}_t = \hat{u}_t(W_t)$, assuming that $(H_t^T W_t H_t)$ is nonsingular, using equation (2.7) we rewrite

$$\begin{aligned} F(u_t, W_t) &= E((y_t - Hu_t)^T W_t (y_t - Hu_t)) \qquad (2.59) \\ &= E[(H^T W_t H)u_t - H^T W_t y_t]^T (H^T W_t H)^{-1} [(H^T W_t H)u_t - H^T W_t y_t] \\ &+ E(y_t^T [I - W_t H (H^T W_t H)^{-1} H^T] W_t y_t), \end{aligned}$$

where the first term on the right hand side is non-negative definite. To minimize $F(u_t, W_t)$, the first term on the right must vanish, so that

$$\hat{u_t} = (H^T W_t H)^{-1} H^T W_t y_t. \tag{2.60}$$

To find the optimal weight $\hat{W}_t$, let us consider

$$F(\hat{u_t}, W_t) = E((y_t - H\hat{u_t})^T W_t (y_t - H\hat{u_t})) \tag{2.61}$$

as a function of $W_t$.

It is clear that this quantity does not attain a minimum value at a positive definite weight $W_t$ since such a minimum would result from $W_t = 0$. Hence, we need another measurement to determine an optimal $\hat{W}_t$. Noting that the original problem is to estimate the state vector $x_t$ by $\hat{u_t}(W_t)$, it is natural to consider a measurement of the error $(x_t - \hat{u_t}(W_t))$. But since not much about $x_t$ is known and only the noisy data can be measured, this measurement should be determined by the variance of the error. That is, we will minimize $\mathrm{Var}(x_t - \hat{u_t}(W_t))$ over all positive definite symmetric matrices $W_t$. So, we have

$$\begin{aligned}
&\mathrm{Var}(x_t - \hat{u_t}) = \mathrm{Var}[(H^T W_t H)^{-1}(H^T W_t H)x_t - (H^T W_t H)^{-1} H^T W_t y_t] \\
&= \mathrm{Var}[(H^T W_t H)^{-1} H^T W_t (H x_t - y_t)] \\
&= \mathrm{Var}[-(H^T W_t H)^{-1} H^T W_t v_t].
\end{aligned}$$

By the linearity of the expectation, we have

$$\begin{aligned} \mathrm{Var}(x_t - \hat{u_t}) &= (H^T W_t H)^{-1} H^T W_t E(v_t v_t^T) W_t H (H^T W_t H)^{-1} \\ &= (H^T W_t H)^{-1} H^T W_t R_t W_t H (H^T W_t H)^{-1}. \end{aligned} \tag{2.62}$$

This is the quantity to be minimized. To write this as a perfect square, we need the positive square root of the positive definite symmetric matrix $R_t$. That is $(R_t^{1/2})(R_t^{1/2}) = R_t$. It follows that $\mathrm{Var}(x_t - \hat{u_t}) = Q^T Q$ where $Q = (R_t^{1/2}) W_t H (H^T W_t H)^{-1}$. By the matrix Schwarz inequality (2.6) and the assumption that $P$ is a matrix with nonsingular $P^T P$, we have

$$Q^T Q \geq (P^T Q)^T (P^T P)^{-1} (P^T Q). \tag{2.63}$$

We may choose $P = (R_t^{1/2})^{-1} H$, so that $P^T P = H^T ((R_t^{1/2})^T)^{-1} (R_t^{1/2}) H$ $= H^T R_t^{-1} H$ is nonsingular, where $(H^T R_t H)$ is nonsingular, and

$$\begin{aligned} &(P^T Q)^T (P^T P)^{-1} (P^T Q) \\ &= [H^T ((R_t^{1/2})^{-1})^T (R_t^{1/2})^T W_t H (H^T W_t H)^{-1}]^T (H^T R_t^{-1} H)^{-1} \\ &[H^T ((R_t^{1/2})^{-1})^T (R_t^{1/2})^T W_t H (H^T W_t H)^{-1}] \\ &= (H^T R_t^{-1} H)^{-1} \\ &= \mathrm{Var}(x_t - \hat{u_t}(R_t^{-1})). \end{aligned} \tag{2.64}$$

Hence, $\mathrm{Var}(x_t - \hat{u_t}(W_t)) \geq \mathrm{Var}(x_t - \hat{u_t}(R_t^{-1}))$ for all positive definite symmetric weight matrices $W_t$. Therefore, the optimal weight matrix is $\hat{W}_t = R_t^{-1}$, and the optimal

estimate of $x_t$ using this optimal weight is

$$\hat{x_t} = \hat{u_t}(R_t^{-1}) = (H^T R_t^{-1} H)^{-1} H^T R_t^{-1} y_t. \tag{2.65}$$

We call $\hat{x}_t$ the least-squares optimal estimate of $x_t$.

## 2.6 The EM-Algorithm

In this section we describe the maximum-likelihood parameter estimation problem and how the Expectation- Maximization (**EM**) algorithm can be used for its solution. This method will be used for the identification of linear state space model as in Sec. 3.5.

Let $P(\mathbf{x}, \Theta)$ be a density function governed by the set of parameters $\Theta$. We also have a data set of size $N$, $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$. That is, we assume that these data vectors are independent and identically distributed with distribution $P$. Therefore, the resulting density for the samples is

$$P(X|\Theta) = \prod_{i=1}^{N} P(\mathbf{x}_i, \Theta) = L(\Theta|X).$$

This function $L(\Theta, X)$ is called the likelihood function. In the maximum likelihood problem, our goal is to find the $\Theta$ that maximizes $L$. We maximize $\mathrm{Log}L(\Theta, X)$ instead because it is analytically easier [26].

The **EM**-algorithm (Expectation-Maximization algorithm) is an iterative procedure for computing the maximum likelihood estimator from a given data set when the data is incomplete or has missing values. The first proper theoretical study of the algorithm was done by Dempster, Laird, and Rubin (1977) [10]. The EM algorithm is extensively used throughout the statistics literature and has already become a multipurpose tool for building a method of statistical analysis based on likelihood and other substitution methods [26]. We will give an overview on how it works.

We assume that data $X$ is observed and is generated by some distribution. We call $X$ the incomplete data. We assume that a complete data set exists $Z = (X; Y)$ and also assume a joint density function:

$$P(\mathbf{z}|\Theta) = P(\mathbf{x}, \mathbf{y}|\Theta) = P(\mathbf{y}|\mathbf{x}, \Theta)P(\mathbf{x}|\Theta).$$

With this new density function, we can define a new likelihood function, $L(\Theta|Z) = L(\Theta|X, Y) = P(X, Y|\Theta)$, called the complete-data likelihood.

The **EM** algorithm first finds the expected value of the complete-data log-likelihood $\mathrm{Log}P(X, Y|\Theta)$ with respect to the unknown data $Y$ given the observed data $X$ and the current parameter estimates. That is, we define:

$$G(\Theta, \Theta^{(i-1)}) = E[\mathrm{Log}P(X, Y|\Theta)|X, \Theta^{(i-1)}], \tag{2.66}$$

where $\Theta^{(i-1)}$ are the current parameters estimates that we used to evaluate the expectation and $\Theta$ are the new parameters that optimize $G$. The evaluation of this

expectation is called the **E**-step of the algorithm.

The second step (the **M**-step) of the **EM** algorithm is to maximize the expectation we computed in the first step with respect to $\Theta$.

These two steps are repeated as necessary. Each iteration is guaranteed to increase the log likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function.

# Chapter 3

# THE KALMAN FILTER

## 3.1 Introduction

Kalman filter was first proposed by Rodolf Kalman in the year 1960 as optimal estimation filter for linear state space model [32]. In this chapter, we derive the Kalman filter algorithm based on least square estimation [65]. One can see other derivations of Kalman filter in [58,65]. In addition, we give an application of the Kalman filter to linear state space model. Kalman described his filter using state space techniques which enable the filter to be used as either a smoother, a filter or a predictor. So, we can find Kalman smoother equations which is required for the **E**-step in the Expectation-

Maximization (**EM**) algorithm for linear Gaussian state space models. We use the (**EM**) algorithm to identify the unknown parameters in state space model (SSM) (3.1) and (3.2) [59]. The maximum likelihood estimation procedure is to maximize an innovation form of the likelihood function where all values used are obtained in filtering and smoothing equations which are explained in this chapter. For more details about the Maximum Likelihood Method, see [10,26,58].

## 3.2 Gaussian State Space Models

A linear Gaussian state space model is characterized by an unobserved series of vectors $x_1, x_2, ..., x_n$ (called states), that are associated with a series of observations $y_1, y_2, ..., y_n$. The relation between the states and the observations is specified through the observation and the state equations as folloes

$$x_{t+1} = Ax_t + w_t, \tag{3.1}$$

$$y_t = Hx_t + v_t \tag{3.2}$$

$$w_t \sim N(0, Q), \tag{3.3}$$

$$v_t \sim N(0, R), t = 1, 2, ..., n \tag{3.4}$$

$$E(w_j w_t^T) = \delta_{j,t} Q_t, \tag{3.5}$$

$$E(v_j v_t^T) = \delta_{j,t} R_t, \tag{3.6}$$

$$E(w_i v_j^T) = 0, \forall i, j. \tag{3.7}$$

where $y_t$ is a $p\times1$ vector of observations, $x_t$ is a $k\times1$ vector of states. The equation (3.2) is called the observation equation, which has the structure of a linear regression relating the state vector to the observed time series. Eq. (3.1) is called the state equation, which describes the dynamics of the states. The relation between the state vector and the observed values is characterized by the matrix $H$ of size $p\times k$. The dependence of the current state on the past is determined by the transition matrix, $A$.

Eqs. (3.3) and (3.4) imply that the dynamical noise $w_t$ and the observation noise $v_t$ are uncorrelated, white, and Gaussian with zero means. In particular, Eqs. (3.5) and (3.6) indicate that $w_t$ and $v_t$ are white Gaussian processes with the covariances at time $t$ being $Q_t$ and $R_t$, respectively. The symbol $\sim$ in Eqs.(3.3) and (3.4) means "following the distribution". The notation $N(\mu, \Sigma)$ represents a Gaussian distribution with x being the random variable, whose mean and covariance are $\mu$ and $\Sigma$, respectively. Finally, $\delta_{j,t}$ denotes the Kronecker delta function, i.e.,

$$\delta_{i,j} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

## 3.3 Deriving the Kalman filter from least squares estimation

In this section we will derive the Kalman filter equations from point of view of least squares estimation. We mainly follow [65] in our derivation.

If $\Theta = \{A, H, R, Q\}$ is known in our linear Gaussian state space model (3.1) and (3.2), we can estimate the unknown state vector by the Kalman filter procedure.

We will define some notations for later use. $x_t^s = E\{x_t|Y_s\}$, where $Y_s = \{y_1, y_2, ..., y_s\}$ denotes the vectors up to time $s$. For $s = t-1$ the expectation is a forecast whereas for $s = t$, the expectation is the Kalman filtered value. For $s = n$, the expectation is conditional on the entire data and is the Kalman smoother. The conditional covariances $P_t^s = E\{(x_t - x_t^s)(x_t - x_t^s)^T|Y_s\}$, and $P_{t,u}^s = E\{(x_t - x_t^s)(x_u - x_u^s)^T|Y_s\}$, are interpreted in the same way.

Without loss of generality, we suppose that we know the initial state vector $x_0^0$ and the corresponding initial error covariance $P_0^0$ . We will divide the procedures of the Kalman filter into two steps: propagation (or prediction) and filtering [65].

**Propagation step**

The expectation of the state $x_t$ is given from Eq. (3.1) by

$$E(x_t) = AE(x_{t-1}), \tag{3.8}$$

which is normally used as the estimation of the background at instant $t$. The background $x_t^{t-1}$ is estimated as

$$\begin{aligned} x_t^{t-1} &= E(x_t|Y_{t-1}) \\ &= E(Ax_{t-1} + w_t|Y_{t-1}) \\ &= AE(x_{t-1}|Y_{t-1}) \\ &= Ax_{t-1}^{t-1}. \end{aligned} \tag{3.9}$$

The corresponding background error covariance $P_t^{t-1}$ is given by

$$\begin{aligned} P_t^{t-1} &= E((x_t - x_t^{t-1})(x_t - x_t^{t-1})^T|Y_{t-1}) \\ &= E((A(x_{t-1} - x_{t-1}^{t-1}) + w_t)(A(x_{t-1} - x_{t-1}^{t-1}) + w_t)^T|Y_{t-1}) \\ &= AE((x_{t-1} - x_{t-1}^{t-1})(x_{t-1} - x_{t-1}^{t-1})^T|Y_{t-1})A^T + E(w_t w_t^T|Y_{t-1}) \\ &= AP_{t-1}^{t-1}A^T + Q_t. \end{aligned} \tag{3.10}$$

Note that to derive Eq. (3.10), we have assumed that the analysis error $\epsilon_{t-1}^{t-1}$=$x_{t-1}^{t-1} - x_{t-1}$ is independent of the dynamical noise $w_t$.

**Filtering step**

After a new observation $y_t$ is available, one incorporates the new information so as to update the background $x_t^{t-1}$ to the analysis $x_t^t$ . To this end, one needs to find an

optimal weight matrix $K_t$ (Kalman gain), so that the analysis $x_t^t$, is updated according to the following rule

$$x_t^t = x_t^{t-1} + K_t(y_t - Hx_t^{t-1}). \tag{3.11}$$

Minimize the expectation of the energy (the cost function)

$$J_t = E(\|\epsilon_t^t\|^2) = E(\|x_t^t - x_t\|^2) \tag{3.12}$$

of the analysis error $\epsilon_t^t$=$x_t^t$−$x_t$. The reason to use Eq. (3.11) to update the background $x_t^{t-1}$ is because one would normally expect the background, the analysis, and the observation to be unbiased estimations, i.e.,

$$\begin{aligned} E\{\epsilon_t^{t-1}\} &= E(x_t^{t-1} - x_t) = 0, \\ E\{\epsilon_t^t\} &= E(x_t^t - x_t) = 0, \\ E\{v_t\} &= E(y_t - Hx_t) = 0, \end{aligned} \tag{3.13}$$

where $\epsilon_t^{t-1}$, $\epsilon_t^t$ denote the background and analysis errors, respectively. To see the rationale behind Eq. (3.11), one may first write the analysis as a linear combination of the background $x_t^{t-1}$ and the observation $y_t$. That is

$$x_t^t = Cx_t^{t-1} + Wy_t, \tag{3.14}$$

where $C$ and $W$ are $k \times k$ and $k \times p$ constant matrices, respectively. Because of the unbiasedness Eq. (3.13) one has

$$\begin{aligned} E(x_t^t - x_t) &= E(Cx_t^{t-1} + Wy_t - x_t) \\ &= E(Cx_t^{t-1} + W(Hx_t + v_t) - x_t) \\ &= (C + WH - I_k)E(x_t) \\ &= 0, \end{aligned} \tag{3.15}$$

So that $C = I_k - WH$. Substituting this identity for $C$ into Eq. (3.14), we find

$$\begin{aligned} x_t^t &= Cx_t^{t-1} + Wy_t \\ &= (I - WH)x_t^{t-1} + Wy_t \\ &= x_t^{t-1} + Wy_t - WHx_t^{t-1} \\ &= x_t^{t-1} + W(y_t - Hx_t^{t-1}). \end{aligned} \tag{3.16}$$

By replacing $W$ by $K_t$, we obtain Eq. (3.11).

On the other hand, the analysis error covariance

$$\begin{aligned} P_t^t &= E((\epsilon_t^t - E\epsilon_t^t)(\epsilon_t^t - E\epsilon_t^t)^T), \\ &= E(\epsilon_t^t(\epsilon_t^t)^T). \end{aligned} \tag{3.17}$$

Thus it is clear that the cost function in Eq.(3.12) is equivalent to the trace of the error covariance $P_t^t$, i.e.,

$$J_t = E\|\epsilon_t^t\|^2 = E((\epsilon_t^t)^T\epsilon_t^t) = \mathrm{tr}(P_t^t). \tag{3.18}$$

Consequently, the optimal state estimation problem in Eq.(3.1) now becomes an optimization problem whose objective is to minimize $\mathrm{tr}(P_t^t)$ over all possible weights $K_t$. Subtracting the truth $x_t$ from Eq. (3.11), we have

$$x_t^t - x_t = (x_t^{t-1} - x_t) + K_t((y_t - Hx_t) - (Hx_t^{t-1} - Hx_t)). \tag{3.19}$$

Thus Eq.(3.19) can be re-written as

$$\epsilon_t^t = \epsilon_t^{t-1} + K_t(v_t - H\epsilon_t^{t-1}) \tag{3.20}$$

$$= (I - K_tH)\epsilon_t^{t-1} + K_tv_t. \tag{3.21}$$

Therefore, one can obtain the analysis error covariance in terms of the background error covariance by noting that

$$\begin{aligned} P_t^t &= E(\epsilon_t^t(\epsilon_t^t)^T) \\ &= (I - K_tH)E(\epsilon_t^{t-1}(\epsilon_t^{t-1})^T)(I - K_tH)^T + K_tR_tK_t^T \\ &= (I - K_tH)P_t^{t-1}(I - K_tH)^T + K_tR_tK_t^T. \end{aligned} \tag{3.22}$$

Note that to obtain the above result, we have assumed that the background error $\epsilon_t^{t-1}$ and observation noise $v_t$ are independent, so that $E(\epsilon_t^{t-1}(v_t)^T) = E(v_t(\epsilon_t^{t-1})^T) = 0$. Also note that $P_t^{t-1} = E(\epsilon_t^{t-1}(\epsilon_t^{t-1})^T))$ is the background error covariance, and $R_t = E(v_t(v_t)^T)$ is the covariance of the observation noise. Thus one can re-write Eq.(3.22) as

$$P_t^t = P_t^{t-1} - P_t^{t-1}H^TK_t^T + K_tR_tK_t^T - K_tHP_t^{t-1} + K_tHP_t^{t-1}H^TK_t^T. \tag{3.23}$$

Therefore the trace of $P_t^t$ is given by

$$\mathrm{tr}(P_t^t) = \mathrm{tr}(P_t^{t-1}) + \mathrm{tr}(K_t R_t K_t^T) - 2\mathrm{tr}(K_t H P_t^{t-1}) + \mathrm{tr}(K_t H P_t^{t-1} H^T K_t^T). \quad (3.24)$$

Note that to derive Eq.(3.24), we have utilized the fact that $P_t^{t-1} H^T K_t^T$ is the transpose of $K_t H P_t^{t-1}$, hence their traces are the same. To minimize $\mathrm{tr}(P_t^t)$, a necessary condition for an optimal weight $K_t$ is that $\frac{d\mathrm{tr}(P_t^t)}{dK_t} = 0$. The differentiation is carried out using the rules given in subsection 2.2.2.

So, from Eq. (3.24), we get

$$\frac{d}{dK_t}\mathrm{tr}(P_t^t)M = \mathrm{tr}(K_t(R_t + R_t^T) - 2(HP_t^{t-1})^T + K_t(HP_t^{t-1}H^T + (HP_t^{t-1}H^T)^T)M, \forall M \in \mathbb{R}^{n\times n}.$$

Since the covariance matrices are symmetric, we have

$$\begin{aligned} &\tfrac{d}{dK_t}\mathrm{tr}(P_t^t)M = \mathrm{tr}([2K_tR_t - 2P_t^{t-1}H^T + 2K_tHP_t^{t-1}H^T]M) \quad (3.25)\\ &= 2\mathrm{tr}([K_tR_t - P_t^{t-1}H^T + KHP_t^{t-1}H^T]M). \end{aligned}$$

Equaling the above expression to zero and observing that equality must hold for all $M \in \mathbb{R}^{n\times n}$,

$$2[K_tR_t - P_t^{t-1}H^T + KHP_t^{t-1}H^T] = 0.$$

This gives

$$K_t = P_t^{t-1}H^T(HP_t^{t-1}H^T + R_t)^{-1}, \quad (3.26)$$

where $HP_t^{t-1}H^T + R_t$ is invertible because it is positive definite.

Substituting Eq.(3.26) into Eq.(3.23), it can be shown that

$$P_t^t = P_t^{t-1} - K_t H P_t^{t-1}. \tag{3.27}$$

## 3.4 Summary of the Kalman filter and the Kalman smoother:

We summarize for later reference the equations for the Kalman filter and the Kalman smoother as follows

**a.The Kalman Filter**

For the state space model specified in (3.1) and (3.2) with initial conditions $x_0^0 = \mu_0$ and $P_0^0 = \Sigma_0$, for $t = 1, 2, ..., n$

$$x_t^{t-1} = A x_{t-1}^{t-1} \tag{3.28}$$

$$P_t^{t-1} = A P_{t-1}^{t-1} A^T + Q_t \tag{3.29}$$

$$x_t^t = x_t^{t-1} + K_t(y_t - H x_t^{t-1}) \tag{3.30}$$

$$P_t^t = (I - K_t H) P_t^{t-1} \tag{3.31}$$

where the Kalman gain is

$$K_t = P_t^{t-1} H^T (H P_t^{t-1} H^T + R_t)^{-1}. \tag{3.32}$$

**b.The Kalman Smoother**

For the state space model specified in (3.1) and (3.2) with initial conditions $x_n^n = \mu_n$ and $P_n^n = \Sigma_n$ and for $t = n, n-1, ..., 1$,

$$x_{t-1}^n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - Ax_{t-1}^{t-1}) \tag{3.33}$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}^T, \tag{3.34}$$

$$J_{t-1} = P_{t-1}^{t-1}A^T(P_t^{t-1})^{-1}. \tag{3.35}$$

where $x_t^{t-1}$ and $P_t^{t-1}$ are given by (3.28)-(3.31).

**c. The Lag-One Covariance Smoother**

For the state space model specified in (3.1) and (3.2), with $K_t, J_t$, and $P_n^n$ and with initial condition

$$P_{n,n-1}^n = (I - K_nH_n)A_nP_{n-1}^{n-1}, \tag{3.36}$$

for $t = n, n-1, ..., 2$,

$$P_{t-1,t-2}^n = P_{t-1}^{t-1}J_{t-2}^T + J_{t-1}(P_{t,t-1}^n - A_tP_{t-1}^{t-1})J_{t-2}^T. \tag{3.37}$$

For more details about mathematical theory of Kalman filtering and Kalman smoothing, see [3,23,31,32,58,65].

**Example** (Applying Kalman Filter to State Space Model)

In this example, we introduce an application of Kalman filter equations (3.28)-(3.31)

to the linear state space model (3.1) and (3.2). Let

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & -2 & 1 \end{bmatrix}, H = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}.$$

We take the noises $w_t$ and $v_t$ as random noises. The covariance matrices $Q$ and $R$ are constant depending on $w_t$ and $v_t$, respectively. We start by the initial state vector $x_0 = 0_{3\times 1}$ with error covariance $P_0 = 0_3$. From the initial state $x_0$, we find true states $x_1, x_2, ..., x_n$ by the dynamical equation (3.1) and from these states we find the observations $y_1, y_2, ..., y_n$ by the measurement equation.

Using the Kalman filter equations (3.28)-(3.31), we estimate the state vectors $\hat{x_1}, \hat{x_2}, ..., \hat{x_n}$. The figure shows the error between the true states its estimations at time $t = 1, 2, ..., n$. The X-axis is the time and Y-axis is the error.

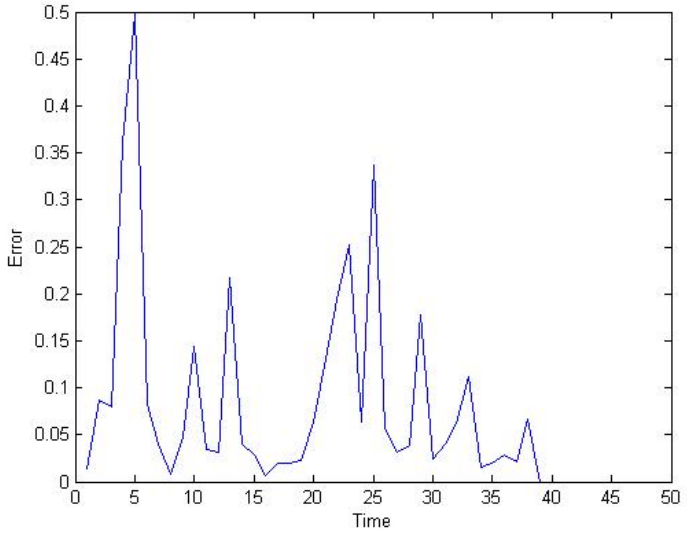


Figure 3.1: Kalman filter (n=50)

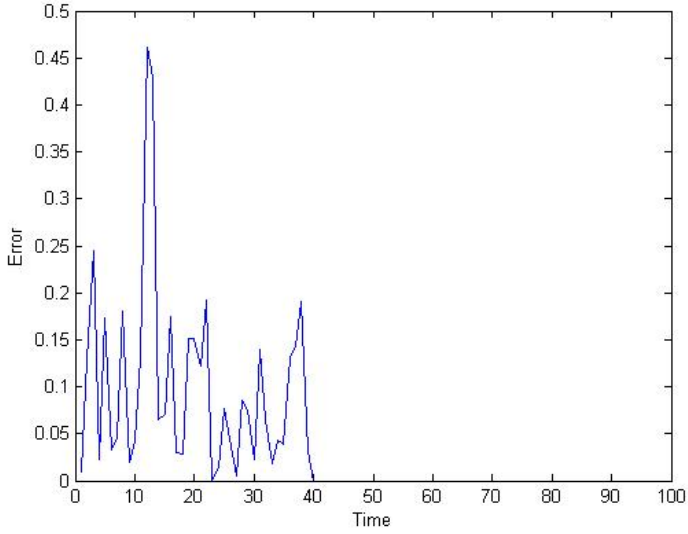


Figure 3.2: Kalman filter (n=100)

## 3.5 Ensemble Kalman Filter

In section 3.3, we derived the Kalman filter based on two fundamental assumptions, namely, the linearity of the state and observation equations and the Gaussianity of the dynamical and observation noise. Indeed, the assumption of linearity is not often valid. Thus the ensemble Kalman filter (EnKF), initially proposed by Evensen (1994), is designed to tackle this problem [13].

The EnKF is a sophisticated sequential data assimilation method. It applies an ensemble of model states to represent the error statistics of the model estimate, it applies ensemble integrations to predict the error statistics forward in time, and it uses an analysis scheme which operates directly on the ensemble of model states when observations are assimilated.

Suppose that, at the beginning of each assimilation cycle, one has an ensemble of the background (called background ensemble), usually obtained from the previous assimilation cycle. Then, with an incoming observation, one applies the Kalman filter equations (3.28)-(3.31) to update each individual member of the background ensemble. To do this, the mean and error covariance of the background are approximated by the sample mean and sample covariance of the background ensemble, so that one can apply KF Eqs. (3.28) and (3.29) to obtain an ensemble of the analysis. The analysis ensemble is then used to estimate the mean and covariance of the underlying system

states. By propagating the analysis ensemble forward through the dynamical system, one obtains a new background ensemble for the next assimilation cycle. In this way, by using only a small ensemble to evaluate the statistics (mean and covariance) at both the propagation and filtering steps, the computational cost of the filter can be reduced. A recent review and overview of the EnKF is given in [13], which provides detailed information on the formulation, interpretation and implementation of the EnKF.

## 3.6 Identification of linear state space models:

In this section, we will use the **EM** (Expectation-Maximization) algorithm to estimate the parameters $\Theta = \{A, H, Q, R\}$ in the state space model defined by Equations (3.1) and (3.2).

The **EM** algorithm for state space model requires the computation of the Kalman filter and the Kalman smoother for $x_t$ which was studied in the previous section.

Let $X_n = \{x_1, x_2, ..., x_n\}$ and $Y_n = \{y_1, y_2, ..., y_n\}$ be the states and observations, respectively. Under the Gaussian assumption, the probability density $P(x_0)$ is given by

$$P(x_0) = (2\pi)^{-k/2}|\Sigma_0|^{-1/2}\exp(-1/2(x_0-\mu_0)^T\Sigma_0^{-1}(x_0-\mu_0)). \tag{3.38}$$

Based on (3.1) and (3.2) we can write the conditional densities for the state and output

$$P(y_t|x_t) = (2\pi)^{-k/2}|R|^{-1/2}\exp(-1/2(y_t-Hx_t)^TR^{-1}(y_t-Hx_t)) \tag{3.39}$$

$$P(x_t|x_{t-1}) = (2\pi)^{-k/2}|Q|^{-1/2}\exp(-1/2(x_t-Ax_{t-1})^TQ^{-1}(x_t-Ax_{t-1})).$$

Since the dynamic systems are Markovian and the by the Beyes' formula (2.35), the joint likelihood for the complete data is given by

$$\begin{aligned} P(Y_n, X_n) &= P(y_1, y_2, ..., y_n, x_1, x_2, ..., x_n) \\ &= P(y_n, x_n|X_{n-1}, Y_{n-1})P(X_{n-1}, Y_{n-1}) \\ &= P(y_n|x_n)P(x_n|X_{n-1}, Y_{n-1})P(Y_{n-1}, X_{n-1}) \\ &= P(y_n|x_n)P(x_n|x_{n-1})P(Y_{n-1}, X_{n-1}) \\ &= ... \\ &= \prod_{t=1}^n P(y_t|x_t)\prod_{t=1}^n P(x_{t+1}|x_t)P(x_0). \end{aligned} \tag{3.40}$$

Taking the Log of Eq. (3.40), we have

$$\begin{aligned}
&\mathrm{Log}P(Y_n, X_n; \Theta) = \mathrm{Log}[\textstyle\prod_{t=1}^{n} P(y_t|x_t) \prod_{t=1}^{n} P(x_{t+1}|x_t)P(x_0)] \\
&= \textstyle\sum_{t=1}^{n} \mathrm{Log}[(2\pi)^{-k/2}|R|^{-1/2} \exp(-1/2(y_t - Hx_t)^T R^{-1}(y_t - Hx_t))] \\
&+ \textstyle\sum_{t=1}^{n} \mathrm{Log}[(2\pi)^{-k/2}|Q|^{-1/2} \exp(-1/2(x_t - Ax_{t-1})^T Q^{-1}(x_t - Ax_{t-1}))] \\
&+\mathrm{Log}[(2\pi)^{-k/2}|\Sigma_0|^{-1/2} \exp(-1/2(x_0 - \mu_0)^T \Sigma_0^{-1}(x_0 - \mu_0))] \\
&= \textstyle\sum_{t=1}^{n} \mathrm{Log}[|R|^{-1/2}] + \sum_{t=1}^{n}(-1/2(y_t - Hx_t)^T R^{-1}(y_t - Hx_t)) \\
&+ \textstyle\sum_{t=1}^{n} \mathrm{Log}[|Q|^{-1/2}] + \sum_{t=1}^{n}(-1/2(x_t - Ax_{t-1})^T Q^{-1}(x_t - Ax_{t-1})) \\
&+\mathrm{Log}[|\Sigma_0|^{-1/2} + (-1/2(x_0 - \mu_0)^T \Sigma_0^{-1}(x_0 - \mu_0))] + \mathrm{Const.}
\end{aligned} \tag{3.41}$$

Therefore, the joint log likelihood can be written as

$$\begin{aligned}
&\mathrm{Log}P(Y_n, X_n; \Theta) = -n/2\mathrm{Log}|R| - \textstyle\sum_{t=1}^{n}(y_t - Hx_t)^T R^{-1}(y_t - Hx_t) - n/2\mathrm{Log}|Q| \\
&- \textstyle\sum_{t=1}^{n}(x_t - Ax_{t-1})^T Q^{-1}(x_t - Ax_{t-1}) - 1/2\mathrm{Log}|\Sigma_0| - 1/2(x_0 - \mu_0)^T \Sigma_0^{-1}(x_0 - \mu_0) \\
&+\mathrm{Const.}
\end{aligned} \tag{3.42}$$

The EM algorithm tries to maximize $\mathrm{Log}P$ with respect to $\Theta$.

The two steps of the **EM** algorithm are (see Sec. 2.6):

1. The expectation step (**E**-step): we compute the conditional expectation of the joint log likelihood of the complete data at iteration $j, j = 1, 2, ...$

$$G(\Theta|\Theta^{(j-1)}) = E[\mathrm{Log}P(Y_n, X_n; \Theta)|Y_n, \Theta^{(j-1)}], \tag{3.43}$$

where $\Theta^{j-1}$ is the parameter vector obtained in the previous iteration.

2. The maximization step (**M**- step): we maximize the expected likelihood function

with respect our unknown parameters $\Theta$.

By using Lemma 9 and linearity of the trace and the conditional expectation, we can calculate Eq. (3.43) as.

$$G(\Theta|\Theta^{(j-1)}) = E[\mathrm{Log}P(Y_n, X_n; \Theta)|Y_n, \Theta^{(j-1)}]$$

$$= -n/2\mathrm{Log}|R| - \mathrm{tr}R^{-1}\sum_{t=1}^{n} E([(y_t - Hx_t)(y_t - Hx_t)^T]|Y_n)\}$$

$$-n/2\mathrm{Log}|Q| - 1/2\mathrm{tr}\{Q^{-1}\sum_{t=1}^{n} E([x_t - Ax_{t-1})(x_t - Ax_{t-1})^T]|Y_n)\} - 1/2\mathrm{Log}|\Sigma_0|$$

$$-1/2\mathrm{tr}|\{\Sigma_0^{-1}E((x_0 - \mu_0)(x_0 - \mu_0)^T|Y_n) + \mathrm{Const}..$$

Using this property of $\mathrm{Cov}(X, Y)$

$$\mathrm{Cov}(X, Y) = E(XY^T) - E(X)(E(Y))^T,$$

and recalling that

$$P_t^n = E[(x_t - x_t^n)(x_t - x_t^n)^T]$$

$$P_{t,t-1}^n = E[(x_{t-1} - x_{t-1}^n)(x_t - x_t^n)^T],$$

we have

$$G(\Theta|\Theta^{(j-1)}) = -n/2\mathrm{Log}|R| - \mathrm{tr}\{R^{-1}\sum_{t=1}^{n}[(y_t - Hx_t^n)(y_t - Hx_t^n)^T \tag{3.44}$$

$$+HP_t^nH^T]\} - n/2\mathrm{Log}|Q| - 1/2\mathrm{tr}\{Q^{-1}[\sum_{t=1}^{n}[x_t^n(x_t^n)^T + P_t^n]$$

$$-\sum_{t=1}^{n}[x_t^n(x_{t-1}^n)^T + P_{t,t-1}^n]A^T - A[\sum_{t=1}^{n}[x_t^n(x_{t-1}^n)^T + P_{t,t-1}^n]]^T$$

$$+A\sum_{t=1}^{n}[x_{t-1}^n(x_{t-1}^n)^T + P_{t-1}^n]A^T]\} - 1/2\mathrm{Log}|\Sigma_0|$$

$$-1/2\mathrm{tr}|\{\Sigma_0^{-1}[x_0^n - \mu_0)(x_0^n - \mu_0)^T + P_0^n]\} + \mathrm{Const}.$$

Putting

$$B = \sum_{t=1}^{n} [x_{t-1}^n (x_{t-1}^n)^T + P_{t-1}^n], \tag{3.45}$$

$$C = \sum_{t=1}^{n} [x_t^n (x_{t-1}^n)^T + P_{t,t-1}^n], \tag{3.46}$$

$$D = \sum_{t=1}^{n} [x_t^n (x_t^n)^T + P_t^n] \tag{3.47}$$

in Eq. (3.44), we get

$$\begin{aligned} G(\Theta|\Theta^{(j-1)}) &= -n/2\text{Log}|R| - \text{tr}\{R^{-1} \textstyle\sum_{t=1}^{n} [(y_t - Hx_t^n)(y_t - Hx_t^n)^T \\ &+ HP_t^n H^T]\} - n/2\text{Log}|Q| - 1/2\text{tr}\{Q^{-1}[D - CA^T - AC^T + ABA^T]\} - 1/2\text{Log}|\Sigma_0| \\ &- 1/2\text{tr}|\{\Sigma_0^{-1}[x_0^n - \mu_0)(x_0^n - \mu_0)T + P_0^n]\} + \text{Const.} \end{aligned} \tag{3.48}$$

In the above equations, the components $x_{t-1}^n, x_t^n, P_{t-1}^n, P_t^n$ and $P_{t,t-1}^n$ can be calculated by using the Kalman filter equations and the Kalman smoother equations (3.28)-(3.37). Calculation of equation (3.48) is the expectation step (E-step).

Now, we maximization of the function $G(\Theta|\Theta^{(j-1)})$ with respect to the unknown parameters set, $\Theta^{(j)} = \{H(j), A(j), Q(j), R(j), \mu_0(j)\}$ at iteration $j$ can be calculated from Eq. (3.48) as follows

• To find $A$, we will minimize the function $G(\Theta|\Theta^{(j-1)})$ with respect to $A$. It is enough to minimize the term

$$D - CA^T - AC^T + ABA^T \tag{3.49}$$

by the Lemma 6 of completing the squares, we have

$$D - CA^T - AC^T + ABA^T = D - CA^T - (CA^T)^T + ABA^T \tag{3.50}$$
$$= (A - CB^{-1})C(A - CB^{-1}) + D - CB^{-1}C^T.$$

So, this expression is minimum with respect to $A$ when

$$A = CB^{-1}. \tag{3.51}$$

• To find $Q$, we will take the partial derivative of the function $G(\Theta|\Theta^{(j-1)})$ with respect to $Q$ and equate to zero. So,

$$\tfrac{d}{dQ}G(\Theta|\Theta^{(j-1)})M = \tfrac{d}{dQ}[-n/2\text{Log}|Q| - 1/2\text{tr}\{Q^{-1}[D - CB^{-1}C^T]\}]M$$
$$= -n/2\text{tr}(Q^{-1}M) - 1/2\text{tr}(Q^{-1}MQ^{-1}(D - CA^T - AC^T + ABA^T)$$
$$= 1/2\text{tr}\{nQ^{-1}M - Q^{-1}MQ^{-1}(D - CA^T - AC^T + ABA^T)$$

equating to zero:

$$\text{tr}\{Q^{-1}M(nI - Q^{-1}(D - CA^T - AC^T + ABA^T)) = 0, \forall M \in \mathbb{R}^{k\times k} \tag{3.52}$$
$$\Rightarrow (nI - Q^{-1}(D - CA^T - AC^T + ABA^T)) = 0$$
$$\Rightarrow Q^{-1}(D - CA^T - AC^T + ABA^T) = nI$$
$$Q = 1/n(D - CA^T - AC^T + ABA^T).$$

Put Eq. (3.51) in Eq. (3.52), we have

$$Q = 1/n(D - CB^{-1}C^T). \tag{3.53}$$

• To find $H$, We will minimize Eq. (3.48) with respect to $H$. So ,we will minimize the terms which include $H$.

$$\begin{aligned}
&\mathrm{tr}(R^{-1}\textstyle\sum_{t=1}^{n}(y_t - Hx_t^n)(y_t - Hx_t^n)^T + HP_t^nH^T) \\
&= \mathrm{tr}(R^{-1}\textstyle\sum_{t=1}^{n} y_t y_t^T - Hx_t^n y_t^T - y_t(x_t^n)^T H^T + Hx_t^n(x_t^n)^T H^T + HP_t^nH^T) \\
&= \mathrm{tr}(R^{-1}\{(\textstyle\sum_{t=1}^{n} y_t y_t^T) - H(\sum_{t=1}^{n} x_t^n y_t^T) - (\sum_{t=1}^{n} y_t(x_t^n)^T)H^T \\
&+H(\textstyle\sum_{t-1}^{n} x_t^n(x_t^n)^T + P_t^n)H^T\}) \\
&= \mathrm{tr}(R^{-1}(S_1 - HS_2 - S_2^TH^T + HS_3H^T)).
\end{aligned}$$

where

$$\begin{aligned}
S_1 &= \textstyle\sum_{t=1}^{n} y_t y_t^T \\
S_2 &= \textstyle\sum_{t=1}^{n} x_t^n y_t^T \\
S_3 &= \textstyle\sum_{t=1}^{n}(x_t^n(x_t^n)^T + P_t^n).
\end{aligned}$$

Now, we will calculate the partial derivative of this expression w.r.t $H$ and equate to zero.

$$\tfrac{d}{dH}(\mathrm{tr}(R^{-1}(S_1 - HS_2 - S_2^TH^T + HS_3H^T)))M = \mathrm{tr}(R^{-1}(-MS_2 - S_2^TM^T + MS_3H^T + HS_3M^T)) \tag{3.54}$$

But

$$\begin{aligned}
\mathrm{tr}(R^{-1}S_2^TM^T) &= \mathrm{tr}(S_2^TM^TR^{-1}) \\
&= \mathrm{tr}(R^{-1}MS_2).
\end{aligned}$$

So, Eq. (3.54) becomes

$$\tfrac{d}{dH}(\mathrm{tr}(R^{-1}(S_1 - HS_2 - S_2^TH^T + HS_3H^T)))M = \mathrm{tr}(R^{-1}M(-2S_2 + 2S_3H^T)$$

Equating to zero and observing that equality must hold for all $M \in \mathbb{R}^{k\times k}$ gives

$$S_3 H^T = S_2 \tag{3.55}$$
$$\Rightarrow H^T = S_3^{-1} S_2$$
$$\Rightarrow H = S_2^T S_3^{-1} = (\textstyle\sum_{t=1}^{n}(\mathbf{y}_t(\mathbf{x}_t^n)^T))(\sum_{t=1}^{n}(\mathbf{x}_t^n(\mathbf{x}_t^n)^T + \mathbf{P}_t^n))^{-1}$$

• To find $R$, it is enough to minimize

$$-n/2\text{Log}|R| - \text{tr}\{R^{-1}\textstyle\sum_{t=1}^{n}[(y_t - Hx_t^n)(y_t - Hx_t^n)^T + HP_t^n H^T]\}. \tag{3.56}$$

This expression can be written using the same steps in Eqn (3.54) as

$$-n/2\text{Log}|R| - \text{tr}(R^{-1}(S_1 - HS_2 - S_2^T H^T + HS_3 H^T)). \tag{3.57}$$

Now, we use the same steps used in finding $Q$. We differentiate with respect to $R$ and equate to zero

$$\tfrac{d}{dR}(-n/2\text{Log}|R| - \text{tr}(R^{-1}(S_1 - HS_2 - S_2^T H^T + HS_3 H^T))M = 0 \tag{3.58}$$
$$\Rightarrow -1/2\text{tr}(nR^{-1}M - R^{-1}MR^{-1}(S_1 - HS_2 - S_2^T H^T + HS_3 H^T) = 0$$
$$\Rightarrow -1/2\text{tr}R^{-1}M(nI - R^{-1}(S_1 - HS_2 - S_2^T H^T + HS_3 H^T)) = 0$$
$$\Rightarrow nI - R^{-1}(S_1 - HS_2 - S_2^T H^T + HS_3 H^T)) = 0$$
$$\Rightarrow R = 1/n(S_1 - HS_2 - S_2^T H^T + HS_3 H^T)).$$

Putting the values of $S_1$, $S_2$, $S_3$ and $H$, we have

$$R = 1/n\{\textstyle\sum_{t=1}^{n} y_t y_t^T - (\sum_{t=1}^{n} y_t(x_t^n)^T)(\sum_{t=1}^{n} x_t^n(x_t^n)^T + P_t^n)^{-1}(\sum_{t=1}^{n} x_t^n y_t^T)\}. \tag{3.59}$$

Finally, we can find $\mu_0$ as

$$\mu_0 = \mathbf{x}_0^n \tag{3.60}$$

We summarize the procedure of the **EM** computations as following:

1. Initialize the procedure by selecting starting values for the parameters $\Theta^{(0)} = \{A^{(0)}, H^{(0)}, R^{(0)}, Q^{(0)}\}$ and estimate $\mu_0$.

2. (**E**-step) Use Eqns (3.28)-(3.37) to estimate the smoothed values $x_t^n, P_t^n$ and $P_{t,t-1}^n$, for $t = 1, 2, ..., n$, with parameters$\Theta^{(j-1)}, (j = 1, 2, ...)$.

3. (**M**-step) Update the estimates,$A, H, R$ and $Q$ using the Eqns.(3.51), (3.53), (3.55) and (3.59), to obtain $\Theta^{(j)}$.

4. Repeat 2 and 3 above until the estimates and the log likelihood function (3.42) are stable.

# Chapter 4

# CONVEXITY INTERVAL ANALYSIS

## 4.1 Introduction

In the ordinary interval analysis, some algebra properties of real number are not satisfied in interval setting (see Sec2.4). For example, the distribution law for intervals does not hold in general. So, the inverse of an interval matrix is not well defined because a determinate is not well defined. Also, the multiplication of interval matrices is not associative and hence we are not able to find the powers of interval matrices.

In this chapter, we introduce modified arithmetic operations on intervals and interval matrices introduced in section (2.4) by using the convexity which make some algebraic properties hold. In addition, we introduce definitions for interval random variables, interval expectation, interval variance, interval covariance, interval conditional expectation, interval conditional variance. We investigate some algebraic properties of interval random variables. All of these are important when we extend the Kalman filter equations to interval Kalman filter and interval Kalman smoother which will be studied in the next chapter.

## 4.2 Convexity Interval Arithmetic

Given an interval $I = [a, b]$ and $t \in [0, 1]$, we let $x(t)$ denote the convex combination

$$x(t) = (1 - t)a + tb.$$

Thus we may write

$$I = \{x(t) : t \in [0, 1]\}.$$

In the next subsection we define some basic interval operations that will be used for subsequent developments. We should note that the "special" operations given below produce subintervals of the more general interval operations given in [1]. Our special operations will be denoted $\oplus, \ominus, \otimes, \oslash$ as opposed to the more general operators

$+, -, \times, \div$

**Basic Operations**

**Addition**: For $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, we let

$$\begin{aligned} I_1 \oplus I_2 &= \{x(t) + y(t), x(t) \in I_1, y(t) \in I_2\} \\ &= \{(1-t)a_1 + tb_1 + (1-t)a_2 + tb_2\} \\ &= \{a_1 - a_1 t + tb_1 + a_2 - ta_2 + tb_2\} \\ &= \{a_1 + a_2 - t(a_1 + a_2) + t(b_1 + b_2)\} \\ &= [a_1 + a_2, b_1 + b_2] = I_1 + I_2. \end{aligned} \tag{4.1}$$

**Negation**: For $I = [a, b]$, we let

$$\begin{aligned} \ominus I_1 &= \ominus [a, b] \\ &= -\{(1-t)a + tb, t \in [0, 1]\} \\ &= \{t(a-b) - a\} \\ &= [-b, -a] \\ &= -I. \end{aligned} \tag{4.2}$$

**Subtraction**: For $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, we let

$$\begin{aligned} I_1 \ominus I_2 &= I_1 \oplus (\ominus(I_2)) \\ &= [a_1, b_1] \oplus [-b_2, -a_2] \\ &= [a_1 - b_2, b_1 - a_2] = I_1 - I_2. \end{aligned} \tag{4.3}$$

**Multiplication**: For $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, we let

$$\begin{aligned} I_1 \otimes I_2 &= \{x(t)y(t), x(t) \in I_1, y(t) \in I_2\} \\ &= \{((1-t)a_1 + tb_1)((1-t)a_2 + tb_2)\} \\ &\subseteq I_1 \times I_2. \end{aligned} \tag{4.4}$$

**Division**: For $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, we let

$$\begin{aligned} I_1 \oslash I_2 &= \{x(t)/y(t), x(t) \in I_1, y(t) \in I_2, 0 \notin I_2\} \\ &= \{((1-t)a_1 + tb_1)/((1-t)a_2 + tb_2)\} \\ &\subseteq I_1 \div I_2. \end{aligned} \tag{4.5}$$

The above operations have the same algebraic properties (i.e. closed, commutative, associative and identity) of the more general interval operations.. In addition, distributivity holds

$$I_1 \otimes (I_2 \oplus I_3) = (I_1 \otimes I_2) \oplus (I_1 \otimes I_3). \tag{4.6}$$

The proof is easily shown using equations (4.1) and (4.4). Distributivity allows us to define regously such notion as determinants of interval matrices.

## 4.3 Interval Matrices, Interval Linear Systems

### 4.3.1 Interval Matrices

The definition of interval matrices, interval vectors and their operations are introduced in subsection 2.4.4. Let $\mathbf{A}$ be an $n \times n$ interval matrix and $\mathbf{b}$ an $n \times 1$ interval vector. The determinant of an interval matrix is not well defined because the distributive law is not true under the more general interval operators.

**Example** Let

$$\mathbf{A} = \begin{bmatrix} [-2,2] & [0] & [-1,1] \\ [-2,0] & [1] & [1] \\ [-1] & [-1,1] & [-1] \end{bmatrix},$$

then we find det$\mathbf{A}$ using the first row

$$|\mathbf{A}| = [-2,2]([1][-1] - [-1,1][1]) - [0]([-2,0][-1] - [-1][1]) + [-1,1]([-2,0][-1,1] - [-1][1])$$
$$= [-7,7]. \tag{4.7}$$

The determinate using the second row

$$|\mathbf{A}| = -[-2,0](0 - [-1,1][-1,1]) + [1]([-2,2][-1] - [-1][-1,1]) - [1]([-2,2][-1,1] - 0)$$
$$= [-7,9]. \tag{4.8}$$

which is not equal to the determinate in Eq. (4.7).

**Definition 32** *The interval matrix* $\mathbf{A}$ *is said to be regular if every* $A \in \mathbf{A}$ *has an inverse.*

**Definition 33** *The inverse interval matrix* $\mathbf{A}^{-1}$ *is defined by [54]*

$$\mathbf{A}^{-1} = [\{A^{-1} : A \in \mathbf{A}\}]. \tag{4.9}$$

Here $[S]$ is the smallest interval vector (matrix) containing $S \subset \mathbb{R}^n$ ($S \subset \mathbb{R}^{n\times n}$).

Under "special" operations, we will introduce the definition of interval matrix, determinant, inverse and some properties.

**Definition 34** *Let* $\boldsymbol{A} = [\underline{A}, \overline{A}]$ *be an interval matrix, we define an interval matrix* $\boldsymbol{A}$ *as*

$$\boldsymbol{A} = \{A_\alpha, \alpha \in [0,1]\} = \{(1-\alpha)\underline{A} + \alpha\overline{A}, \alpha \in [0,1]\},$$

*where* $\underline{A}$ *is the lower bound of* $\boldsymbol{A}$ *and* $\overline{A}$ *is the upper bound.*

**Definition 35** *The determinant of a square interval matrix* $\boldsymbol{A}$ *is defined by*

$$\det(\boldsymbol{A}) = |\boldsymbol{A}| = \{\det A_\alpha, \alpha \in [0,1]\}. \tag{4.10}$$

*which is an interval.*

**Definition 36** *The adjoint interval matrix of a square interval matrix* $\boldsymbol{A}$ *is defined by*

$$\mathrm{adj}\boldsymbol{A} = \{\mathrm{adj}A_\alpha, \alpha \in [0,1]\}. \tag{4.11}$$

**Definition 37** *The inverse of a square interval matrix* $\boldsymbol{A}$ *is defined by*

$$\boldsymbol{A}^{-1} = \{A_\alpha^{-1}, \alpha \in [0,1]\} = \{\frac{\mathrm{adj}A_\alpha}{\det A_\alpha}, \alpha \in [0,1]\}, \tag{4.12}$$

*where* $0 \notin |\boldsymbol{A}|$.

It is clear that $\mathbf{A}^{-1}$ is an interval matrix.

Since,

$$\{A_\alpha^{-1}, \alpha \in [0,1]\} \subseteq \{A^{-1}, A \in \mathbf{A}\}$$

the inverse of an interval matrix in our sense is a subset of the inverse of an interval matrix in the general sense.

**Example** Let

$$\mathbf{A} = \begin{bmatrix} [2] & [-1,1] \\ [-1,1] & [2] \end{bmatrix},$$

then the inverse of $\mathbf{A}$ in general is

$$\mathbf{A}^{-1} = \begin{bmatrix} \left[\frac{1}{3}, \frac{2}{3}\right] & \left[\frac{-1}{3}, \frac{1}{3}\right] \\ \left[\frac{2}{5}, \frac{2}{3}\right] & \left[\frac{-1}{3}, \frac{1}{3}\right] \end{bmatrix},$$

and the inverse in our sense is

$$\mathbf{A}^{-1} = \begin{bmatrix} \left[\frac{1}{2}, \frac{2}{3}\right] & \left[0, \frac{1}{3}\right] \\ \left[0, \frac{1}{3}\right] & \left[\frac{1}{2}, \frac{2}{3}\right] \end{bmatrix},$$

### 4.3.2 Interval Linear Systems

The intuitive idea of an interval linear system

$$\mathbf{AX} = \mathbf{b} \tag{4.13}$$

is that it consists of all linear systems

$$AX = b \tag{4.14}$$

with $A \in \mathbf{A}$ and $b \in \mathbf{b}$.

**Definition 38** *The solution set $\mathcal{S}$ of (4.13) is defined by*

$$\mathcal{S} = \{X \in \mathbb{R}^n : AX = b, A \in \mathbf{A}, b \in \mathbf{b}\}. \tag{4.15}$$

**Proposition 39** *If $\mathbf{A}$ is regular, then $[\mathcal{S}] \subseteq \mathbf{A}^{-1}\mathbf{b}$.*

**Proof.** Let $X \in \mathcal{S}$. Then there is an $A \in \mathbf{A}$ and $b \in \mathbf{b}$ such that $AX = b$. Then $X = A^{-1}b \in \mathbf{A}^{-1}b \subset \mathbf{A}^{-1}\mathbf{b}$. Hence, $\mathcal{S} \subset \mathbf{A}^{-1}\mathbf{b}$. The minimality of $[\mathcal{S}]$ then implies

that $[\mathcal{S}] \subset \mathbf{A}^{-1}\mathbf{b}$. ■

The next example shows that equality may not occur in the previous proposition.

**Example:** For $\mathbf{A} = \begin{bmatrix} [2] & [-1,0] \\ [-1,0] & [2] \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} [1.2] \\ [-1.2] \end{bmatrix}$

$$\begin{aligned} \mathcal{S} &= \left\{ \begin{bmatrix} \frac{1.2(2-\alpha)}{4-\alpha\beta} \\ \frac{1.2(\beta-2)}{4-\alpha\beta} \end{bmatrix} : \alpha, \beta \in [0,1] \right\}, \\ [\mathcal{S}] &= \begin{bmatrix} [.3,.6] \\ [-.6,-.3] \end{bmatrix}. \end{aligned}$$

However,

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} \left[\frac{1}{2},\frac{2}{3}\right] & \left[0,\frac{1}{3}\right] \\ \left[0,\frac{1}{3}\right] & \left[\frac{1}{2},\frac{2}{3}\right] \end{bmatrix}, \\ \mathbf{A}^{-1}\mathbf{b} &= \begin{bmatrix} [.2,.8] \\ [-.8,.2] \end{bmatrix}. \end{aligned}$$

Hence, $[\mathcal{S}] \subsetneq \mathbf{A}^{-1}\mathbf{b}$.

The set $\mathcal{S}$ defined in (4.15) is not, in general, an interval vector. In fact it may not even be convex and may have a complicated structure. In the above example, the exact solution region indicate in Figure 4.1 which is not interval vector.

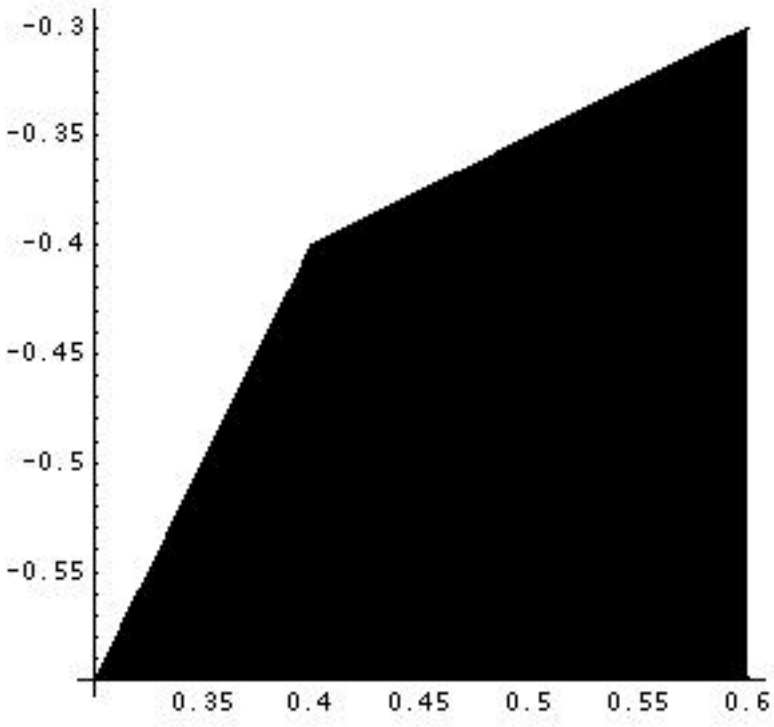


Figure 4.1: Exact solution region for the interval linear system

**Proposition 40** *Let* $\boldsymbol{A} = [\underline{A}, \overline{A}] \in \mathbb{IR}^{(n\times n)}$. *If* $\underline{A}$ *and* $\overline{A}$ *are regular and* $\underline{A}^{-1}, \overline{A}^{-1} \geq 0$, *then* $\boldsymbol{A}$ *is regular and*

$$\boldsymbol{A}^{-1} = [\overline{A}^{-1}, \underline{A}^{-1}] \geq 0.$$

Now, we define the solution set of interval linear system in the convexity sense.

**Definition 41** *The solution of an interval linear system*

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \tag{4.16}$$

*is defined by*

$$S = \{X \in \mathbb{R}^n : A_\alpha X_\alpha = b_\alpha, \alpha \in [0,1]\} \tag{4.17}$$

*which is an interval vector.*

Notice that,

$$\{X \in \mathbb{R}^n : A_\alpha X_\alpha = b_\alpha, \alpha \in [0,1]\} \subseteq \{X \in \mathbb{R}^n : AX = b, A \in \mathbf{A}, b \in \mathbf{b}\}.$$

## 4.4 Interval Random Variables

### 4.4.1 Measurable Set Valued Maps

We begin by discussing the measurability of set valued maps and then introduce the definition of an interval random variable. The basic definitions and more details can be found in [25].

A measurable space $(\Omega, \mathcal{A})$ consists of a basic set $\Omega$ together with a $\sigma$-algebra $\mathcal{A}$ of subsets of $\Omega$ called measurable sets. Here we consider closed convex valued set valued maps $F : \Omega \Rightarrow \mathbb{R}^k$, i.e., $F(\omega)$ is a closed convex subset of $\mathbb{R}^k$ for each $\omega \in \Omega$. This is the case when $F$ is interval valued. The latter notion means that, for each $\omega \in \Omega$, the components of $F(\omega)$ are closed intervals in $\mathbb{R}$.

We first define what it means for a set valued map to be measurable. Recall that the inverse image of a set $S \subset \mathbb{R}^k$ under the set valued map $F$ is defined by

$$F^{-1}(S) = \{\omega \in \Omega : F(\omega) \cap S \neq \emptyset\}$$

and that the graph of $F$ (Denoted by $G_F$) is defined by

$$G_F = \{(\omega, y) : \omega \in \Omega, y \in F(\omega)\}.$$

**Definition 42** *Let $(\Omega, \mathcal{A})$ be a measurable space and $F : \Omega \to \mathbb{R}^k$ be a set valued*

*map. $F$ is called measurable if the inverse image of each open set is a measurable set: if $O \subset \mathbb{R}^k$ is open, then $F^{-1}(O) \in \mathcal{A}$.*

For the rest of this thesis, we will use the special interval operations $\oplus, \ominus, \otimes$ and $\oslash$ defined in section(4.2). However, we will not use the above special symbols. Instead, we will use the usual symbols $+, -, \times, \div$.

We are now in a position to introduce the definition of interval random variables.

**Definition 43** *Let $(\Omega, \mathcal{S}, P)$ be a probability space. A measurable interval valued map $X : \Omega \to \mathbb{R}^k$ is called an interval random variable. A stochastic interval process is an indexed set of interval random variables.*

### 4.4.2 Normally Distributed Interval Random Variables

In order to arrive at a definition of normally distributed interval random variables and their expectations and variances, which are anticipated to be interval valued themselves, we need to discuss first the integral of set valued maps and interval valued maps. The discussion begins with the notion of measurable selections.

**Definition 44** *Let $(\Omega, \mathcal{A})$ be a measurable space and $F : \Omega \rightrightarrows \mathbb{R}^k$ be a measurable set valued map. A measurable selection of $F$ is a measurable map $f : \Omega \to \mathbb{R}^k$ satisfying*

$f(\omega) \in F(\omega)$ *for each* $\omega \in \Omega$.

It is well known that every measurable set valued map has at least one measurable selection. Furthermore, we have the following equivalence [25].

**Theorem 45** *Let* $(\Omega, \mathcal{A})$ *be a measurable space and denote by* $\mathcal{B}$ *the* $\sigma$*-algebra of Borel sets in* $\mathbb{R}^k$. *Let* $F : \Omega \rightrightarrows \mathbb{R}^k$ *be a set valued map. The following are equivalent:*

1. $F$ *is measurable.*
2. $F^{-1}(B) \in \mathcal{A}$ *for every* $B \in \mathcal{B}$.
3. *The map* $d(y, F(\cdot))$ *is measurable for each* $y \in \mathbb{R}^k$.
4. *There exists a sequence of measurable selections* $\{f_n\}_{n=1}^{\infty}$ *of* $F$ *such that*

$$F(\omega) = \overline{\cup_{n \geq 1} f_n(\omega)}$$

*for each* $\omega \in \Omega$.

A countable family of measurable selections satisfying the last property is called dense. Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be an interval valued map. We define the two special functions $l_F$ and $r_F : \Omega \to \mathbb{R}^k$ by $l_F(\omega) = a(\omega)$ and $r_F(\omega) = b(\omega)$, where $F(\omega) = [a(\omega), b(\omega)]$ for each $\omega \in \Omega$. The next lemma shows that $l_F$ and $r_F$ are measurable selections of $F$ when the latter is measurable.

**Lemma 46** *Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be a measurable interval valued map. Then the single valued functions $l_F$ and $r_F$ are measurable selections of $F$.*

**Proof.** Choose a sequence of measurable selections $\{f_n\}_{n=1}^{\infty}$ of $F$ such that

$$F(\omega) = \overline{\cup_{n\geq 1} f_n(\omega)}.$$

Then $l_F(\omega) = \inf_{n\geq 1} f_n(\omega)$ and $r_F(\omega) = \sup_{n\geq 1} f_n(\omega)$ (here the inf and sup operations are taken componentwise). Since the inf and the sup operators preserve measurability, we see that the functions $l_F$ and $r_F$ are measurable selections of $F$. ■

**Example:** Let $\Omega = [1, \infty)$ and define $F : \Omega \rightrightarrows \mathbb{R}$ by

$$F(t) = \left[t, t + \frac{1}{t}\right].$$

Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the rational numbers in the interval $[0, 1]$ and let's assume that $r_1 = 1$, $r_2 = 0$. Define $f_n : [1, \infty) \to \mathbb{R}$ by

$$f_n(t) = r_n t + (1 - r_n)\left(t + \frac{1}{t}\right).$$

Thus $l_F(t) = t = f_1(t)$ and $r_F(t) = \left(t + \frac{1}{t}\right) = f_2(t)$. For every $t \in [1, \infty)$, the set $\left\{r_n t + (1 - r_n)\left(t + \frac{1}{t}\right)\right\}_{n=1}^{\infty}$ is dense in the interval $\left[t, t + \frac{1}{t}\right]$.

Now suppose that $(\Omega, \mathcal{A}, \mu)$ is a measure space and $F : \Omega \rightrightarrows \mathbb{R}^k$ is a set valued map. A measurable selection $f$ of $F$ is an integrable selection if $f$ is integrable with respect to the measure $\mu$. The set of all integrable selections of $F$ will be denoted by $\mathcal{F}$. The map

$F$ is called integrally bounded if there exists a $\mu$-integrable function $g \in L^1(\Omega; \mathbb{R}, \mu)$ such that $F(\omega) \subset g(\omega)\mathbf{B}$ for $\mu$-almost every $\omega \in \Omega$. Here $\mathbf{B}$ denotes the unit ball in $\mathbb{R}^k$. In this case every measurable selection $f$ of $F$ is also an integrable selection since $f(\omega) \in F(\omega) \subset g(\omega)\mathbf{B}$ implies that $\|f(\omega)\| \leq |g(\omega)|$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$.

**Definition 47** *The integral of a set valued map $F$ is defined to be the set of integrals of integrable selections of $F$. That is,*

$$\int_\Omega F d\mu = \left\{ \int_\Omega f d\mu : f \in F \right\}. \tag{4.18}$$

We shall say that $F$ is integrable if every measurable selection is integrable.

The following two properties [25] will be responsible for many important results in what follows,

$$\int_\Omega \lambda F d\mu = \lambda \int_\Omega F d\mu, \tag{4.19}$$

$$\int_\Omega (F_1 + F_2) d\mu = \int_\Omega F_1 d\mu + \int_\Omega F_2 d\mu \tag{4.20}$$

**Lemma 48** *Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be an interval valued map. If $l_F$ and $r_F$ are integrable then $F$ is integrable and*

$$\begin{aligned} \int_\Omega F d\mu &= \left[ \int_\Omega l_F d\mu, \int_\Omega r_F d\mu \right] \\ &= \left\{ \int_\Omega f_\alpha d\mu : f_\alpha = \alpha l_F + (1-\alpha) r_F, \alpha \in [0,1] \right\}. \end{aligned}$$

**Proof.** The first equality is shown as follows. Since for every $\omega \in \Omega$ and every measurable selection $f$ of $F$ we have $l_F(\omega) \leq f(\omega) \leq r_F(\omega)$, $f$ is integrable and

$$\int_\Omega l_F(\omega)\, d\mu \leq \int_\Omega f(\omega)\, d\mu \leq \int_\Omega r_F(\omega)\, d\mu.$$

Therefore,

$$\int_\Omega F d\mu \subseteq \left[\int_\Omega l_F d\mu, \int_\Omega r_F d\mu\right].$$

On the other hand, let $\theta \in \left[\int_\Omega l_F d\mu, \int_\Omega r_F d\mu\right]$. We may write $\theta = \alpha \int_\Omega l_F d\mu + (1-\alpha) \int_\Omega r_F d\mu$ for some $\alpha \in [0,1]$. Then

$$\begin{aligned} \theta &= \int_\Omega (\alpha l_F + (1-\alpha) r_F)\, d\mu \\ &= \int_\Omega f_\alpha d\mu, \end{aligned}$$

where $f_\alpha = \alpha l_F + (1-\alpha) r_F$. Hence, $\theta \in \int_\Omega F d\mu$.

The second equality is an immediate consequence of this. ■

It will always be assumed that both $l_F$ and $r_F$ are integrable.

**Example:** Let $\Omega$ and $F$ be defined as in the previous example. Let $\mu$ be the measure defined by

$$d\mu = \frac{1}{t^3} dt.$$

Then

$$\int_\Omega F d\mu = \left[\int_1^\infty l_F(t)\, d\mu, \int_1^\infty r_F(t)\, d\mu\right] = \left[1, \frac{4}{3}\right].$$

In view of (4.20) we have the following corollary.

**Corollary 49** *Let $F_1, F_2 : \Omega \rightrightarrows \mathbb{R}^k$ be integrable interval valued maps. Then*

$$\begin{aligned}\int_\Omega (F_1 + F_2)\, d\mu &= \int_\Omega F_1 d\mu + \int_\Omega F_2 d\mu \\ &= \left[\int_\Omega l_{F_1} d\mu, \int_\Omega r_{F_1} d\mu\right] + \left[\int_\Omega l_{F_2} d\mu, \int_\Omega r_{F_2} d\mu\right] \\ &= \left[\int_\Omega (l_{F_1} + l_{F_2})\, d\mu, \int_\Omega (r_{F_1} + r_{F_2})\, d\mu\right].\end{aligned}$$

Let $(\Omega, \mathcal{S}, P)$ be a probability space and let $Z : \Omega \Rightarrow \mathbb{R}^k$ be an interval random variable. We have

$$Z(\omega) = [l_Z(\omega), r_Z(\omega)] = \{z_\alpha := \alpha l_Z(\omega) + (1-\alpha) r_Z(\omega) : \alpha \in [0,1]\}.$$

An interval random variable $Z$ will be called Gaussian if every $z \in Z$ is Gaussian. An interval stochastic process $\{Z_t\}_{t\in T}$ will be called Gaussian if each $t \in T$, $Z_t$ is a Gaussian interval random variable. From this point on, all interval random variables and interval stochastic processes will be allowed Gaussian.

An interval random variable will be called zero-mean if $\mu = [0]$. This happens if and only if each $z_\alpha$ is zero-mean in the usual sense. Observe that the latter statement is true if and only if $l_Z, r_Z$ are zero-mean.

**The interval expectation** of the interval random variable $Z$ can then be defined as

follows

$$\begin{aligned} E(Z) &= \{E(z_\alpha) : \alpha \in [0,1]\} \\ &= \{E(\alpha l_Z + (1-\alpha) r_Z) : \alpha \in [0,1]\} \\ &= \{\alpha E(l_Z) + (1-\alpha) E(r_Z) : \alpha \in [0,1]\} \\ &= [E(l_Z), E(r_Z)], \end{aligned}$$

where in the last equation we used the monotonicity of the expectation ( $y \leq z$) a.e. implies $E(y) \leq E(z)$). It should also be noted here that the expectation of a vector random variable is the vector of expectations of its components. It follows from equations (4.19) and (4.20) that

$$\begin{aligned} E(\lambda Z) &= \lambda E(Z) \\ E(Z_1 + Z_2) &= E(Z_1) + E(Z_2) \end{aligned}$$

Also, if $I = [a, b]$ then

$$\begin{aligned} E(IZ) &= \{E(tZ) : t \in [a,b]\} \\ &= \{tE(Z) : t \in [a,b]\} \\ &= IE(Z). \end{aligned}$$

Also, If $\mathbf{A}\,(k \times k)$ is an interval matrix then

$$\begin{aligned} E\,(\mathbf{A}Z) &= E\left(\sum_{j=1}^{k} \mathbf{A}_{ij} Z_j\right)_{i=1}^{k} \\ &= \left(\sum_{j=1}^{k} E\,(\mathbf{A}_{ij} Z_j)\right)_{i=1}^{k} = \left(\sum_{j=1}^{k} \mathbf{A}_{ij} E\,(Z_j)\right)_{i=1}^{k} \\ &= \mathbf{A}E\,(Z)\,. \end{aligned}$$

**The interval variance** has the form

$$\begin{aligned} &\mathrm{Var}(Z) = \{\mathrm{Var}(z_\alpha) : \alpha \in [0,1]\} \qquad (4.21) \\ &= \{\mathrm{Var}((1-\alpha)l_z + \alpha r_z), \alpha \in [0,1]\} \\ &= \{(\alpha-1)^2 \mathrm{Var}(l_z) + 2\alpha(1-\alpha)\mathrm{Cov}(l_z, r_z) + \alpha^2 \mathrm{Var}(r_z)\} \end{aligned}$$

which is clearly an interval. Observe that, if $Z$ is one-dimensional, then the above interval reduces to $\mathrm{int}(\mathrm{Var}(l_z), \mathrm{Var}(r_z))$, where

$$\mathrm{int}(x, y) = [\min\{x, y\}, \max\{x, y\}].$$

To see this, we note that

$$2\mathrm{Cov}(l_z, r_z) \leq 2\mathrm{Var}(l_z)^{1/2}\mathrm{Var}(r_z)^{1/2} \leq \mathrm{Var}(l_z) + \mathrm{Var}(r_z)$$

then

$$\begin{aligned} &(1-\alpha)^2 \mathrm{Var}(l_z) + 2\alpha(1-\alpha)\mathrm{Cov}(l_z, r_z) + \alpha^2 \mathrm{Var}(r_z) \\ &\leq (1-\alpha)^2 \mathrm{Var}(l_z) + \alpha(1-\alpha)(\mathrm{Var}(l_z) + \mathrm{Var}(r_z)) + \alpha^2 \mathrm{Var}(r_z) \\ &\leq (1-\alpha)^2 \mathrm{Var}(l_z) + \alpha \mathrm{Var}(r_z). \end{aligned}$$

**The interval covariance** has the form

$$\begin{aligned}
&\mathrm{Cov}(Y,Z) = \{\mathrm{Cov}(y_\alpha, z_\beta), \alpha, \beta \in [0,1]\} \qquad (4.22)\\
&= \{\mathrm{Cov}(((1-\alpha)l_y + \alpha r_y), ((1-\beta)l_z + \beta r_z)) : \alpha, \beta \in [0,1]\}\\
&= \{(1-\alpha)(1-\beta)\mathrm{Cov}(l_y, l_z) + \beta(1-\alpha)\mathrm{Cov}(l_y, r_z) + \alpha(1-\beta)\mathrm{Cov}(r_y, l_z)\\
&+\alpha\beta(r_y, r_z) : \alpha, \beta \in [0,1]\}.
\end{aligned}$$

Since the last equality can be regarded as a continuous image of $[0,1] \times [0,1]$, and since the projection operator onto the $i, j$ component of a matrix is also continuous, $\mathrm{Cov}(Y,Z)$ is an interval matrix.

The two interval random variables $Y, Z$ will be called uncorrelated if for each $y \in Y$, $z \in Z$, $y, z$ are uncorrelated. Therefore, $Y, Z$ are uncorrelated if and only if $\mathrm{Cov}\,(Y,Z) = [0]$.

**Theorem 50** *Let $Y = [l_y, r_y], Z = [l_z, r_z], W = [l_w, r_w]$ be interval random variables, $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{IR}^{(k\times k)}$ ,$\lambda \in \mathbb{R}$ and $A \in \mathbb{R}^{(k\times k)}$, then*

1. $\mathrm{Cov}(\lambda Y, Z) = \lambda \mathrm{Cov}(Y,Z)$
2. $\mathrm{Cov}(Y+Z, W) = \mathrm{Cov}(Y,W) + \mathrm{Cov}(Z,W)$
3. $\mathrm{Cov}(AY, Z) = A\mathrm{Cov}(Y,Z)$
4. $\mathrm{Cov}(Y, BZ) = \mathrm{Cov}(Y,Z)B^T$

5. $\mathrm{Cov}(\boldsymbol{A}Y, Z) = \boldsymbol{A}\mathrm{Cov}(Y, Z)$

6. $\mathrm{Cov}(\boldsymbol{A}Y, \boldsymbol{B}Z) = \boldsymbol{A}\mathrm{Cov}(Y, Z)\boldsymbol{B}^T$.

**Proof.** To show 1:

$$\mathrm{Cov}(\lambda Y, Z) = \{(1-\alpha)(1-\beta)\mathrm{Cov}(\lambda l_y, l_z) + \beta(1-\alpha)\mathrm{Cov}(\lambda l_y, r_z) + \alpha(1-\beta)\mathrm{Cov}(\lambda r_y, l_z)$$

$$+\alpha\beta(\lambda r_y, r_z) : \alpha, \beta \in [0,1]\}$$

$$= \{\lambda(1-\alpha)(1-\beta)\mathrm{Cov}(l_y, l_z) + \lambda\beta(1-\alpha)\mathrm{Cov}(l_y, r_z) + \lambda\alpha(1-\beta)\mathrm{Cov}(r_y, l_z)$$

$$+\lambda\alpha\beta\mathrm{Cov}(ry,r_z) : \alpha, \beta \in [0,1]\}$$

$$= \{\lambda((1-\alpha)(1-\beta)\mathrm{Cov}(l_y, l_z) + \beta(1-\alpha)\mathrm{Cov}(l_y, r_z) + \alpha(1-\beta)\mathrm{Cov}(r_y, l_z)$$

$$+\alpha\beta\mathrm{Cov}(r_y, r_z)) : \alpha, \beta \in [0,1])\}$$

$$= \lambda\mathrm{Cov}(Y, Z).$$

To show 2:

$$\mathrm{Cov}(Y + Z, W) = \mathrm{Cov}([l_y + l_z, r_y + r_z], [l_w, r_w])$$

$= \{(1-\alpha)(1-\beta)\text{Cov}(l_y + l_z, l_w) + \beta(1-\alpha)\text{Cov}(l_y + l_z, r_w) + \alpha(1-\beta)\text{Cov}(r_y + r_z, l_w)$

$+\alpha\beta(r_y + r_z, r_w) : \alpha, \beta \in [0,1]\}$

$= \{(1-\alpha)(1-\beta)[\text{Cov}(l_y, l_w) + \text{Cov}(l_z, j_w)] + \beta(1-\alpha)[\text{Cov}(l_y, r_w) + \text{Cov}(l_z, r_w)]$

$+\alpha(1-\beta)[\text{Cov}(r_y, l_w) + \text{Cov}(r_z, l_w)] + \alpha\beta[\text{Cov}(r_y, r_w) + \text{Cov}(r_z, z_w)] : \alpha, \beta \in [0,1]\}$

$= \{(1-\alpha)(1-\beta)\text{Cov}(l_y, l_w) + (1-\alpha)(1-\beta)\text{Cov}(l_z, l_w)$

$+\beta(1-\alpha)\text{Cov}(l_y, r_w) + \beta(1-\alpha)\text{Cov}(l_z, r_w) + \alpha(1-\beta)\text{Cov}(r_y, l_w)$

$+\alpha(1-\beta)\text{Cov}(r_z, l_w) + \alpha\beta\text{Cov}(r_y, r_w) + \alpha\beta\text{Cov}(r_z, r_w) : \alpha, \beta \in [0,1]\}$

$= \{(1-\alpha)(1-\beta)\text{Cov}(l_y, l_w) + \beta(1-\alpha)\text{Cov}(l_y, r_w) + \alpha(1-\beta)\text{Cov}(r_y, l_w)$

$+\alpha\beta\text{Cov}(r_z, r_w) : \alpha, \beta \in [0,1]\}$

$+\{(1-\alpha)(1-\beta)\text{Cov}(l_z, l_w) + \beta(1-\alpha)\text{Cov}(l_z, r_w) + \alpha(1-\beta)\text{Cov}(r_z, l_w)$

$+\alpha\beta\text{Cov}(r_z, r_w)\alpha, \beta \in [0,1]\}$

$= \text{Cov}(Y, W) + \text{Cov}(Z, W).$

3. Let $Y, Z$ be interval random vector, then

$$\begin{aligned}
\text{Cov}\,(AY, Z) &= \text{Cov}(\left(\sum_{j=1}^{k} a_{ij} Y_j\right)_{i=1}^{k}, Z) \\
&= \left(\text{Cov}(\sum_{j=1}^{k} a_{ij} Y_j, Z)\right)_{i=1}^{k} \\
&= \left(\sum_{j=1}^{k} \text{Cov}(a_{ij} Y_j, Z)\right)_{i=1}^{k} \\
&= \left(\sum_{j=1}^{k} a_{ij} \text{Cov}(Y_j, Z)\right)_{i=1}^{k} \\
&= A\text{Cov}(Y, Z).
\end{aligned}$$

Before we prove part(5), we need to to prove that $\mathrm{Cov}(IY, Z) = I\mathrm{Cov}(Y, Z)$, where $I \in \mathbb{IR}$.

$$\begin{aligned}\mathrm{Cov}(IY, Z) &= \{\mathrm{Cov}(tY, Z) : t \in I\} \\ &= \{t\mathrm{Cov}(Y, Z) : t \in I\} = I\mathrm{Cov}(Y, Z).\end{aligned}$$

5. Let $Y, Z$ be interval random vector, then

$$\begin{aligned}\mathrm{Cov}\left(\mathbf{A}Y, Z\right) &= \mathrm{Cov}(\left(\sum_{j=1}^{k} A_{ij}Y_j\right)_{i=1}^{k}, Z) \\ &= \left(\mathrm{Cov}(\sum_{j=1}^{k} A_{ij}Y_j, Z)\right)_{i=1}^{k} \\ &= \left(\sum_{j=1}^{k} \mathrm{Cov}(A_{ij}Y_j, Z)\right)_{i=1}^{k} \\ &= \left(\sum_{j=1}^{k} A_{ij}\mathrm{Cov}(Y_j, Z)\right)_{i=1}^{k} \\ &= \mathbf{A}\mathrm{Cov}(Y, Z).\end{aligned}$$

To show 6:

$$\mathrm{Cov}\left(\mathbf{A}Y, \mathbf{B}Z\right) = \mathrm{Cov}(\left(\sum_{j=1}^{k} A_{ij}Y_j\right)_{i=1}^{k}, \left(\sum_{j=1}^{k} B_{ij}Z_j\right)_{i=1}^{k})$$

$$
\begin{aligned}
&= \mathrm{Cov}(\left(\sum_{j=1}^{k} A_{ij}Y_j, \sum_{j=1}^{k} B_{ij}Z_j\right)_{i=1}^{k}) \\
&= \left(\mathrm{Cov}(\sum_{j=1}^{k} A_{ij}Y_j, \sum_{j=1}^{k} B_{ij}Z_j)\right)_{i=1}^{k} \\
&= \left(\sum_{j=1}^{k} \mathrm{Cov}(A_{ij}Y_j, B_{ij}Z_j)\right)_{i=1}^{k} \\
&= \left(\sum_{j=1}^{k} A_{ij}\mathrm{Cov}(Y_j, Z_j)B_{ij}^T\right)_{i=1}^{k} \\
&= \mathbf{A}\mathrm{Cov}(Y, Z)\mathbf{B}^T.
\end{aligned}
$$

■

The interval conditional expectation is defined as

$$
\begin{aligned}
E\left(Z|Y\right) &= \{E(z_\alpha|y_\alpha) : \alpha \in [0,1]\} \\
&= \{(1-\alpha)E(l_z|y_\alpha) + \alpha E(r_z|y_\alpha) : \alpha \in [0,1]\}.
\end{aligned}
$$

Since the underling probabilities are continuous (Gaussian), $E(Z|Y)$ is an interval.

**Lemma 51** *For interval random variables $X, Y, Z$ and interval matrix $\boldsymbol{A}$,*

1. $E(X+Y|Z) = E(X|Z) + E(Y|Z)$

2. $E(\boldsymbol{A}Y|Z) = \boldsymbol{A}E(Y|Z)$.

**Proof.** To show 1:

$$\begin{aligned} E(X+Y|Z) &= \{E(x_\alpha + y_\alpha|z_\alpha) : \alpha \in [0,1]\} \\ &= \{E(x_\alpha|z_\alpha) + E(y_\alpha|z_\alpha) : \alpha \in [0,1]\} \\ &= \{E(x_\alpha|z_\alpha) : \alpha \in [0,1]\} + \{E(y_\alpha|z_\alpha) : \alpha \in [0,1]\} \\ &= E(X|Z) + E(Y|Z). \end{aligned}$$

To show 2:

$$\begin{aligned} E(\mathbf{A}Y|Z) &= E(\left(\sum_{j=1}^{k} \mathbf{A}_{ij} Y_j\right)_{i=1}^{k} |Z) \\ &= \left(\sum_{j=1}^{k} E\left(\mathbf{A}_{ij} Y_j|Z\right)\right)_{i=1}^{k} \\ &= \left(\sum_{j=1}^{k} \mathbf{A}_{ij} E\left(Y_j|Z\right)\right)_{i=1}^{k} \\ &= \mathbf{A}E\left(Y|Z\right). \end{aligned}$$

■

# Chapter 5

# A CONVEXITY INTERVAL KALMAN FILTER

## 5.1 Introduction

In this chapter, we extend the classical state space model (3.1) and (3.2) to interval state space model (5.1) and (5.2) with the same statistical assumptions.

In order to do this, we need to extend the statistical concepts to interval setting which we have already introduced in chapter 4.

Also, since the traditional Kalman filter technique could not be used directly with in-

terval state space model, we introduce interval Kalman filter to handle the situation. In addition, we rigorously derive the interval Kalman filter and the interval Kalman smoother using the sound definition and statistical properties given in the previous chapter.

Finally, we identify interval parameters of the interval state space model from a given record of interval measurements. We generalize the Expectation-Maximization method to interval setting using interval Kalman filter and interval Kalman smoother equations. Finally, we present a simulation of interval parameters estimation for interval linear state space model. We also give an interval prediction model for weather.

## 5.2 Discussion

In the existing literature, an optimal interval Kalman filter was attempted in [18]. That attempt suffered from lack of proper definitions and rigorous treatment. The simulation given there amounts to a way of dealing with interval systems. The idea in [18] paper was to replace the interval matrix inversion with its worst case inversion, while keeping everything else unchanged. Most of the concepts from the ordinary case were extended in a straightforward way, with not justification, to the interval case. Unfortunately, many concepts cannot be easily extended in this way. Some of

the equations they used do not even make sense in the interval case. For example, they wrote the equations for the IKF by replacing matrices and vectors by interval matrices and interval vectors without paying attention to any details. For example, the Kalman gain for the IKF uses the inverse of an interval matrix to be found. However, the concept of the inverse of an interval matrix as the quotient of an adjoint and determinant is not well defined because the determinant of an interval matrix is not well defined. On the other hand, there is no algorithm to invent an interval matrix in its classical definition.

In our work, we rigorously derive a suboptimal interval Kalman filter using our definitions and well defined statistical properties. Also, we check that the **EM** algorithm is applicable. Furthermore, numerically, we have a well defined procedure for choosing the ensembles in the suboptimal interval Kalman filter.

## 5.3 An Interval State Space Model

In a state space model, certain elements of the system parameters such as the coefficient matrices are not precisely known or gradually change with time. One way to take these uncertainties into account is to allow interval state space models. So, we

extend the state space model (3.1) and (3.2) to interval state space model

$$\mathbf{x}_{t+1}^I = A^I \mathbf{x}_t^I + \mathbf{w}_t, \tag{5.1}$$

$$\mathbf{y}_t^I = H^I \mathbf{x}_t^I + \mathbf{v}_t, \tag{5.2}$$

$$\mathbf{w}_t \sim N(0, Q^I), \tag{5.3}$$

$$\mathbf{v}_t \sim N(0, R^I), t = 1, 2, ..., n \tag{5.4}$$

$$E(\mathbf{w}_j \mathbf{w}_t^T) = \delta_{j,t} Q_t^I, \tag{5.5}$$

$$E(\mathbf{v}_j \mathbf{v}_t^T) = \delta_{j,t} R_t^I, \tag{5.6}$$

$$E(\mathbf{w}_i \mathbf{v}_j^T) = 0, \forall i, j. \tag{5.7}$$

where $\mathbf{y}_t^I$ is a $p \times 1$ interval vector of observations, $\mathbf{x}_t^I$ is a $k \times 1$ interval vector of states, $A^I$ is called interval transition matrix, $H^I$ is called the measurement interval matrix. The matrices $Q^I$ and $R^I$ are the interval covariances of $\mathbf{w}_t$ and $\mathbf{v}_t$, respectively. We will use the following notation in our work: $A^I = \mathbf{A}$, $H^I = \mathbf{H}$, $Q^I = \mathbf{Q}$ and $R^I = \mathbf{R}$

## 5.4 Deriving the Convexity Interval Kalman Filter

In this section, we introduce derivation of interval Kalman filter. The interval operations and the statistical concepts used here are those introduced in chapter 4. Let $\mathbf{X} = \mathbf{x}^I$, $\mathbf{Y} = \mathbf{y}^I$, and $\Upsilon_t = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_t\}$.

From equation (5.1) and Lemma 50, we find

$$\begin{aligned}
\mathbf{X}_t^{t-1} &= E(\mathbf{X}_t|\Upsilon_{t-1}), \\
&= E(\mathbf{A}\mathbf{X}_{t-1} + \mathbf{w}_t|\Upsilon_{t-1}), \\
&= E(\mathbf{A}\mathbf{X}_{t-1}|\Upsilon_{t-1}), \\
&= \mathbf{A}E(\mathbf{X}_{t-1}|\Upsilon_{t-1}), \\
&= \mathbf{A}\mathbf{X}_{t-1}^{t-1}
\end{aligned} \tag{5.8}$$

and

$$\begin{aligned}
\mathbf{P}_t^{t-1} &= E\{(\mathbf{X}_t - \mathbf{X}_t^{t-1})(\mathbf{X}_t - \mathbf{X}_t^{t-1})\}, \\
&= E\{(\mathbf{A}\mathbf{X}_t + \mathbf{w}_t - \mathbf{A}\mathbf{X}_{t-1}^{t-1})(\mathbf{A}\mathbf{X}_t + \mathbf{w}_t - \mathbf{A}\mathbf{X}_{t-1}^{t-1})^T\}, \\
&= E\{(\mathbf{A}(\mathbf{X}_t - \mathbf{X}_{t-1}^{t-1}) + \mathbf{w}_t)(\mathbf{A}(\mathbf{X}_t + \mathbf{X}_{t-1}^{t-1}) + \mathbf{w}_t)^T\} \\
&= E\{(\mathbf{A}(\mathbf{X}_t - \mathbf{X}_{t-1}^{t-1}) + \mathbf{w}_t)(\mathbf{X}_t - \mathbf{X}_{t-1}^{t-1})^T\mathbf{A}^T + \mathbf{w}_t^T\} \\
&= E\{(\mathbf{A}(\mathbf{X}_t - \mathbf{X}_{t-1}^{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1}^{t-1})^T\mathbf{A}^T\} + \mathbf{Q} \\
&= \mathbf{A}\mathbf{P}_{t-1}^{t-1}\mathbf{A}^T + \mathbf{Q}
\end{aligned} \tag{5.9}$$

Let us define the innovation as

$$\epsilon_t = \mathbf{Y}_t - E(\mathbf{Y}_t|\Upsilon_{t-1}) = \mathbf{Y}_t - \mathbf{H}\mathbf{X}_t^{t-1} \tag{5.10}$$

Note, $E(\epsilon_t) = [0]$ and

$$\begin{aligned}
\mathrm{Var}(\epsilon_t) &= \mathrm{Var}(\mathbf{H}(\mathbf{X}_t - \mathbf{X}_t^{t-1}) + \mathbf{v}_t), \\
&= \mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R}, \\
&= \Sigma_t.
\end{aligned} \tag{5.11}$$

The innovations are Gaussian and independent of the past observations. Now, we find the conditional covariance between $\mathbf{X}_t$ and $\epsilon_t$ given $\Upsilon_t$ which will be used in the next operations.

$$\begin{aligned}
&\mathrm{Cov}(\mathbf{X}_t, \epsilon_t|\Upsilon_t - 1) = \mathrm{Cov}(\mathbf{X}_t, \mathbf{Y}_t - \mathbf{H}\mathbf{X}_t^{t-1}|\Upsilon_{t-1}), \qquad (5.12)\\
&= \mathrm{Cov}(\mathbf{X}_t - \mathbf{X}_t^{t-1}, \mathbf{Y}_t - \mathbf{H}\mathbf{X}_t^{t-1}|\Upsilon_{t-1}),\\
&= \mathrm{Cov}(\mathbf{X}_t - \mathbf{X}_t^{t-1}, \mathbf{H}(\mathbf{X}_t - \mathbf{X}_t^{t-1}) + \mathbf{v}_t),\\
&= E\{(\mathbf{X}_t - \mathbf{X}_t^{t-1})(\mathbf{H}(\mathbf{X}_t - \mathbf{X}_t^{t-1}) + \mathbf{v}_t)^T\},\\
&= E\{(\mathbf{X}_t - \mathbf{X}_t^{t-1})(\mathbf{X}_t - \mathbf{X}_t^{t-1})\mathbf{H}^T\},\\
&= \mathbf{P}_t^{t-1}\mathbf{H}^T.
\end{aligned}$$

Therefore, the joint conditional distribution of $\mathbf{X}_t$ and $\epsilon_t$ given $\Upsilon_t$ is Gaussian

$$\begin{pmatrix} \mathbf{X}_t \\ \epsilon_t \end{pmatrix} \mid \Upsilon_{t-1} \sim N\left(\begin{bmatrix} \mathbf{X}_t^{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t^{t-1} & \mathbf{P}_t^{t-1}\mathbf{H}^T \\ \mathbf{H}\mathbf{P}_t^{t-1} & \Sigma_t \end{bmatrix}\right). \qquad (5.13)$$

By equations (5.11), (5.12) and the formula (2.39), we can find $\mathbf{X}_t^t$ and $\mathbf{P}_t^t$ as

$$\begin{aligned}
&\mathbf{X}_t^t = E(\mathbf{X}_t|\Upsilon_t), \qquad (5.14)\\
&= E(\mathbf{X}_t|\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_t),\\
&= E(\mathbf{X}_t|\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_{t-1}) + \mathrm{Cov}(\mathbf{X}_t, \epsilon_t|\mathbf{Y}_1, ..., \mathbf{Y}_{t-1})(\mathrm{Var}(\epsilon_t))^{-1}(\epsilon_t - E(\epsilon_t)),\\
&= \mathbf{X}_t^t + \mathbf{P}_t^{t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R})^{-1}\epsilon_t,\\
&= \mathbf{X}_t^{t-1} + \mathbf{K}\epsilon_t,\\
&= \mathbf{X}_t^{t-1} + \mathbf{K}(\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t^{t-1})
\end{aligned}$$

where

$$\mathbf{K} = \mathbf{P}_t^{t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R})^{-1} \tag{5.15}$$

and

$$\begin{aligned}
&\mathbf{P}_t^t = E\{(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t - \mathbf{X}_t^t)^T|\Upsilon_t\}, \qquad (5.16)\\
&= E\{(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t - \mathbf{X}_t^t)^T|\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_{t-1}, \mathbf{Y}_t\},\\
&= E\{(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t - \mathbf{X}_t^t)^T|\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_{t-1}, \epsilon_t\},\\
&= \mathrm{Cov}(\mathbf{X}_t|\Upsilon_{t-1}, \epsilon_t),\\
&= \mathrm{Cov}(\mathbf{X}_t|\Upsilon_{t-1}) - \mathrm{Cov}(\mathbf{X}_t, \epsilon_t)(\mathrm{Cov}(\epsilon_t|\Upsilon_{t-1}))^{-1}(\mathrm{Cov}(\epsilon_t, \mathbf{X}_t)),\\
&= \mathbf{P}_t^{t-1} - \mathbf{P}_t^{t-1}\mathbf{H}^T(\Sigma_t^{-1}\mathbf{H}\mathbf{P}_t^{t-1}),\\
&= \mathbf{P}_t^{t-1} - \mathbf{P}_t^{t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_t^{t-1},\\
&= \mathbf{P}_t^{t-1} - \mathbf{K}\mathbf{H}\mathbf{P}_t^{t-1},\\
&= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^{t-1}.
\end{aligned}$$

For the interval state space model (5.1) and (5.2) with initial conditions $\mathbf{X}_0^0 = \mu_0$ and $\mathbf{P}_0^0 = \Sigma_0$, the equations (5.8),(5.9),(5.14) and (5.16) are called **Interval Kalman Filter** and equation (5.15) called interval Kalman gain.

## 5.5 Derivation of Convexity Interval Kalman Smoother

By Eqn (2.40), we have

$$E[\mathbf{X}_t|\mathbf{X}_{t+1}, \Upsilon_t] = \mathbf{X}_t^t + E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)^T]$$
$$\times E[(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)^T]^{-1} \times (\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t).$$

We can find the first term in the product by substituting for $X_{t+1}$ and using Eqn (5.8):

$$E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)^T] = E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{A}\mathbf{X}_t + w_t - \mathbf{A}\mathbf{X}_t^t)^T]$$
$$= E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t^T\mathbf{A}^T + w_t^T - (\mathbf{X}_t^t)^T\mathbf{A}^T)]$$
$$= E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t^T\mathbf{A}^T - (\mathbf{X}_t^t)^T\mathbf{A}^T) + (\mathbf{X}_t - \mathbf{X}_t^t)w_t^T]$$
$$= E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t - \mathbf{X}_t^t)^T\mathbf{A}^T]$$
$$= \mathbf{P}_t^t\mathbf{A}^T.$$

So, we have

$$E[\mathbf{X}_t|\mathbf{X}_{t+1}, \Upsilon_t] = \mathbf{X}_t^t + \mathbf{P}_t^t\mathbf{A}^T(\mathbf{P}_{t+1}^t)^{-1}((\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t).$$

Let $J_t = \mathbf{P}_t^t\mathbf{A}^T(\mathbf{P}_{t+1}^t)^{-1}$, then we can write

$$q_t = E[\mathbf{X}_t|\mathbf{X}_{t+1}, \Upsilon_t] = \mathbf{X}_t^t + J_t(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t).$$

Since $\Upsilon_t, \mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t$ and $\{v_{t+1}, ..., v_n, w_{t+2}, ..., w_n\}$ generate $\Upsilon_n$, we get

$$\begin{aligned}
\mathbf{X}_t^n &= E[\mathbf{X}_t|\Upsilon_n] = E[q_t|\Upsilon_n] \\
&= E[\mathbf{X}_t^t + J_t(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)|\Upsilon_n] \\
&= \mathbf{X}_t^t + J_t(E[\mathbf{X}_{t+1}|\Upsilon_n] - \mathbf{X}_{t+1}^t \\
&= \mathbf{X}_t^t + J_t(\mathbf{X}_{t+1}^n - \mathbf{X}_{t+1}^t).
\end{aligned}$$

Now, we will derive $\mathbf{P}_t^n$. Using the formula for $\mathbf{X}_t^n$, we have

$$(\mathbf{X}_t - \mathbf{X}_t^n) + J_t\mathbf{X}_{t+1}^n = \mathbf{X}_t - \mathbf{X}_t^t + J_t\mathbf{X}_{t+1}^t.$$

We now multiply both sides by their respective transpose, and take expectations.

$$\begin{aligned}
&E[(\mathbf{X}_t - \mathbf{X}_t^n)(\mathbf{X}_t - \mathbf{X}_t^n)^T] + J_tE[\mathbf{X}_{t+1}^n(\mathbf{X}_t - \mathbf{X}_t^n)^T] \\
&+E[(\mathbf{X}_t - \mathbf{X}_t^n)(\mathbf{X}_{t+1}^n)^T]J_t^T + J_tE[\mathbf{X}_{t+1}^n(\mathbf{X}_{t+1}^n)^T]J_t^T \\
&= E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_t - \mathbf{X}_t^t)^T] + J_tE[\mathbf{X}_{t+1}^t(\mathbf{X}_t - \mathbf{X}_t^t)^T] \\
&+E[(\mathbf{X}_t - \mathbf{X}_t^t)(\mathbf{X}_{t+1}^t)^T]J_t^T + J_tE[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T]J_t^T.
\end{aligned}$$

The second and third term in each side above are clearly equal to 0 since the forecast error at time $t$ is always uncorrelated with the forecast at time $t+1$. Thus

$$\mathbf{P}_t^n = \mathbf{P}_t^t + J_t\{-E[\mathbf{X}_{t+1}^n(\mathbf{X}_{t+1}^n)^T] + E[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T]\}J_t^T \tag{5.17}$$

Now, to simplify the bracketed term,

$$\begin{aligned}
&-E[\mathbf{X}_{t+1}^n(\mathbf{X}_{t+1}^n)^T] + E[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T] \qquad (5.18) \\
&= \{E[\mathbf{X}_{t+1}(\mathbf{X}_{t+1})^T] - E[\mathbf{X}_{t+1}^n\mathbf{X}_{t+1}^n]\} \\
&-\{E[\mathbf{X}_{t+1}(\mathbf{X}_{t+1})^T] - E[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T]\}
\end{aligned}$$

effectively adding and subtracting zero. Focusing on just the left bracket on the RHS of the above equation, we will add and subtract zero again so that

$$E[\mathbf{X}_{t+1}(\mathbf{X}_{t+1})^T] - E[\mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T] \tag{5.19}$$
$$= E[\mathbf{X}_{t+1}(\mathbf{X}_{t+1})^T - \mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T - \mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T + \mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T]$$

Notice that, since $(\mathbf{X}_{t+1} - (\mathbf{X}^n_{t+1})^T)$ is uncorrelated with $\mathbf{X}^n_{t+1}$,

$$E[(\mathbf{X}_{t+1} - (\mathbf{X}^n_{t+1}))(\mathbf{X}^t_{t+1})^T] = 0$$

Therefore,

$$E[\mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T] = E[(\mathbf{X}_{t+1} - (\mathbf{X}^n_{t+1})^T)(\mathbf{X}^n_{t+1})^T] + E[\mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T] \tag{5.20}$$
$$= E[(\mathbf{X}_{t+1} - \mathbf{X}^n_{t+1} + \mathbf{X}^n_{t+1})\mathbf{X}^n_{t+1}]$$
$$= E[\mathbf{X}_{t+1}\mathbf{X}^n_{t+1}]$$

Similarly

$$E[\mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T] = E[\mathbf{X}^n_{t+1}(\mathbf{X}_{t+1} - (\mathbf{X}^n_{t+1})^T] + E[\mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T] \tag{5.21}$$
$$= E[(\mathbf{X}^n_{t+1}((\mathbf{X}_{t+1})^T - (\mathbf{X}^n_{t+1})^T + (\mathbf{X}^n_{t+1})^T)]$$
$$= E[\mathbf{X}^n_{t+1}\mathbf{X}^T_{t+1}]$$

Now, substituting (5.20) and (5.21) into the third and second terms of the RHS of equation (5.19) respectively shows that

$$E[\mathbf{X}_{t+1}\mathbf{X}^T_{t+1}] - E[\mathbf{X}^n_{t+1}\mathbf{X}^n_{t+1}]$$
$$= E[\mathbf{X}_{t+1}\mathbf{X}^T_{t+1} - \mathbf{X}^n_{t+1}\mathbf{X}^T_{t+1} - \mathbf{X}_{t+1}(\mathbf{X}^n_{t+1})^T + \mathbf{X}^n_{t+1}(\mathbf{X}^n_{t+1})^T]$$
$$= E[(\mathbf{X}_{t+1} - \mathbf{X}^n_{t+1})(\mathbf{X}_{t+1} - \mathbf{X}^n_{t+1})^T].$$

By the same trick, we can simplify the right hand bracket in equation (5.18). Consequently, We find

$$
\begin{aligned}
&-E[\mathbf{X}_{t+1}^n(\mathbf{X}_{t+1}^n)^T] + E[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T] \\
&= \{E[\mathbf{X}_{t+1}\mathbf{X}_{t+1}^T] - E[\mathbf{X}_{t+1}^n(\mathbf{X}_{t+1}^n)^T]\} - \{E[\mathbf{X}_{t+1}\mathbf{X}_{t+1}^T] - E[\mathbf{X}_{t+1}^t(\mathbf{X}_{t+1}^t)^T] = \\
&E[(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^n)(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^n)'] - E[(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)(\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t)^T] \\
&= \mathbf{P}_{t+1}^n - \mathbf{P}_{t+1}^t.
\end{aligned}
$$

Thus,

$$\mathbf{P}_t^n = \mathbf{P}_t^t + J_t(\mathbf{P}_{t+1}^n - \mathbf{P}_{t+1}^t)J_t^T.$$

## 5.6 Summary of the Convexity Interval Kalman Filter and Convexity Interval Kalman Smoother

**Interval Kalman Filter**

$$\mathbf{X}_t^{t-1} = \mathbf{A}\mathbf{X}_{t-1}^{t-1} \tag{5.22}$$

$$\mathbf{P}_t^{t-1} = \mathbf{A}\mathbf{P}_{t-1}^{t-1}\mathbf{A}^T + \mathbf{Q} \tag{5.23}$$

$$\mathbf{X}_t^t = \mathbf{X}_t^{t-1} + \mathbf{K}(\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t^{t-1}) \tag{5.24}$$

$$\mathbf{P}_t^t = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^{t-1}, \tag{5.25}$$

$$\mathbf{K} = \mathbf{P}_t^{t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R})^{-1} \tag{5.26}$$

**Interval Kalman Smoother**

For the ISSM (5.1) and (5.2) with initial conditions $\mathbf{X}_n^n$ and $\mathbf{P}_n^n$ obtained via Eqns (5.24) and (5.25), the convexity interval Kalman smoother is defined by

$$\mathbf{X}_{t-1}^n = \mathbf{X}_{t-1}^{t-1} + J_{t-1}(\mathbf{X}_t^n - \mathbf{A}\mathbf{X}_{t-1}^{t-1}) \tag{5.27}$$

$$\mathbf{P}_{t-1}^n = \mathbf{P}_{t-1}^{t-1} + J_{t-1}(\mathbf{P}_t^n - \mathbf{P}_t^{t-1})J_{t-1}^T \tag{5.28}$$

$$\tag{5.29}$$

where

$$J_{t-1} = \mathbf{P}_{t-1}^{t-1}\mathbf{A}^T(\mathbf{P}_t^{t-1})^{-1}. \tag{5.30}$$

**The Lag-One Covariance Smoother**

For the interval state space model specified in (5.1) and (5.2), with $K_t$, $J_t$, and $\mathbf{P}_n^n$ and with initial condition

$$\mathbf{P}_{n,n-1}^n = (I - K_n\mathbf{H}_n)\mathbf{A}_n\mathbf{P}_{n-1}^{n-1}, \tag{5.31}$$

for $t = n, n-1, ..., 2$,

$$\mathbf{P}_{t-1,t-2}^n = \mathbf{P}_{t-1}^{t-1}J_{t-2}^T + J_{t-1}(\mathbf{P}_{t,t-1}^n - \mathbf{A}_t\mathbf{P}_{t-1}^{t-1})J_{t-2}^T. \tag{5.32}$$

## 5.7 Simulation

Here, we introduce a simulation example of the interval Kalman filter.

We consider the interval state space model

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t \tag{5.33}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t$$

where,

$$\mathbf{A} = \begin{bmatrix} [2,3] & [1,2] & [1,2] \\ [0,1] & [1,2] & [-1,0] \\ [1,2] & [-2,-1] & [1,2] \end{bmatrix}, \mathbf{H} = \begin{bmatrix} [.8,1.2] & 0 & 0 \\ 0 & [.8,1.2] & 0 \\ 0 & 0 & [.8,1.2] \end{bmatrix}$$

The random noise $\mathbf{w}_t$ and $\mathbf{v}_t$ are uncorrelated with $\mathbf{w}_t \sim (0, \mathbf{Q})$ and $\mathbf{v}_t \sim (0, \mathbf{R})$, where $\mathbf{Q} = 0.01 I_{3\times 3}$ and $\mathbf{R} = 0.01 I_{3\times 3}$. We start by the initial state $\mathbf{x}_0 = O_{3\times 1}$ with error covariance $\mathbf{P}_0 = O_{3\times 3}$. We use the convexity definitions of interval operations and interval matrices (see Ch. 4). The true states and measurements for the simulation are generated from these parameters. Our definitions and procedures produce a suboptimal Kalman filter. The numerical approximation is suggested by the convexity definitions that we introduced, namely, to partition the interval $[0, 1]$ into $0 = \alpha_0 < \alpha_1 < ... < \alpha_n = 1$, and choose the point systems determined by these values of $\alpha$. From a numerical point of view, this partitioning procedure produces a

method similar to the well known EKF (see Sec. 3.5). However, our procedures work like an algorithm that is well defined and justified to select the ensembles.

We choose the values of $\alpha = 0, 0.1, 0.2, ..., 1$. When $\alpha = 0$, we get the lower endpoints of all intervals and interval matrices. Hence, we retain the classical state space model and apply KF which gives the values of the true states $\mathbf{x}_t$ at $\alpha = 0, 0.1, ..., 1$, where $t = 1, 2, ..., n$. Also, we get the estimation states $\hat{\mathbf{x}}_t$ at $\alpha = 0, 0.1, 0.2, ..., 1$. We find the true interval states and interval estimation by taking the minimum and the maximum of it. So, we find $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ and $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, ..., \hat{\mathbf{x}}_n$. The error can be computed by the distance between two interval vectors using Definition 28

$$\mathbf{e} = d(\mathbf{x}_t, \hat{\mathbf{x}}_t) = [\max\{|\underline{\mathbf{x}}_t - \underline{\hat{\mathbf{x}}}_t|, |\overline{\mathbf{x}}_t - \overline{\hat{\mathbf{x}}}_t|\}]. \tag{5.34}$$

Figures 5.1 and 5.2 show that the error between the true states and the estimated states is very small.

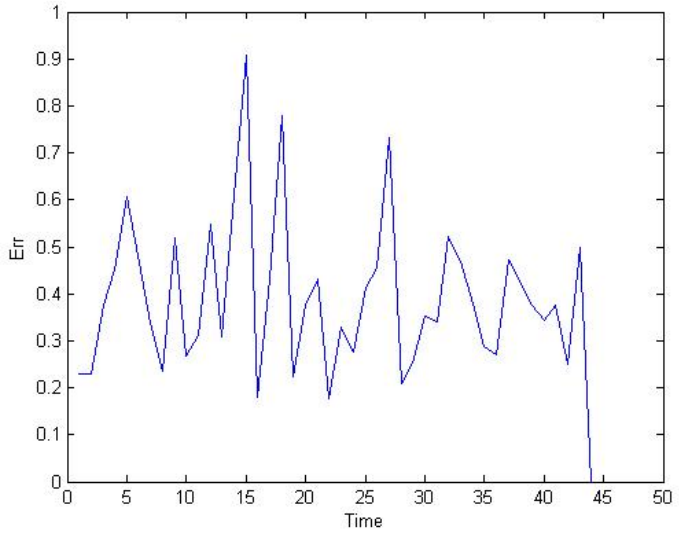


Figure 5.1: Interval Kalman Filter (n=50)

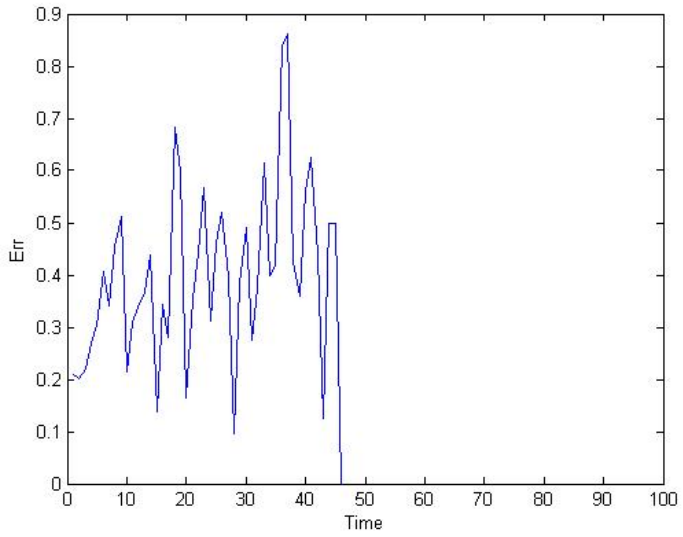


Figure 5.2: Interval Kalman Filter (n=100)

## 5.8 Identification of Interval State Space Models

### 5.8.1 Motivation

We begin by introducing a definition of the differentiation with respect to intervals. The definition is motivated by the results that we obtained on random interval variables and the class of suboptimal Kalman filter which we intend to employ. The notion of the differentiation depends on the definitions and operations on intervals defined in Sec 4.2. We prove two properties of the derivative concept considered here.
Interval differential equations are discussed in several monographs and research papers [2,44,47,48]. Non of these papers use the convexity definitions introduced in this thesis.

**Definition 52** *Let $X \in \mathbb{ID} \subseteq \mathbb{IR}$, $F \in \mathcal{C}(\mathbb{R})$, we define the function defined on an intervals $F : \mathbb{ID} \to \mathbb{IR}$ as*

$$F(X) = \{F(X_\alpha), \alpha \in [0, 1]\}, \tag{5.35}$$

*where $X_\alpha$ defined as in section 4.2.*

It is clear that $F(x)$ is an interval and we denote it by $F(X) = [F^-(X), F^+(X)]$.

**Definition 53** *Let $F : \mathbb{ID} \to \mathbb{IR}$. Then, we define*

$$\lim_{X \to X_0} F(X) = \{\lim_{X_\alpha \to (X_0)_\alpha} F(X_\alpha), \alpha \in [0,1]\}. \tag{5.36}$$

**Lemma 54** *Let $F : \mathbb{ID} \to \mathbb{IR}$. Then,*

$$\lim_{X \to X_0} F(X) = F(X_0) \Leftrightarrow \lim_{X_\alpha \to (X_0)_\alpha} F(X_\alpha) - F((X_0)_\alpha) = 0. \tag{5.37}$$

**Definition 55** *Let $F \in \mathcal{C}^1(\mathbb{R})$. We define $F'(X)$ as*

$$\begin{aligned} F'(X) &= \tfrac{d}{dX} F(X) = \{\tfrac{d}{dX_\alpha} F(X_\alpha), \alpha \in [0,1]\} \\ &= \{\lim_{h \to 0} \tfrac{F(X_\alpha + h) - F(X_\alpha)}{h}, \alpha \in [0,1]\} \end{aligned} \tag{5.38}$$

**Note**: Since $F \in \mathcal{C}^1(\mathbb{R})$, $F'(X)$ is an interval.

**Lemma 56** *Let $F, G : \mathbb{ID} \to \mathbb{IR}$ be differentiable on $\mathbb{R}$ and $I \in \mathbb{ID}$. Then*

*(a) $(F + G)$ is differentiable and*

$$(F + G)'(X) = F'(X) + G'(X) \tag{5.39}$$

*(b) $F'(IX)$ exist and*

$$F'(IX) = IF'(IX). \tag{5.40}$$

**Proof.** To show (a)

$$\frac{d}{dX}(F+G)(X) = \{\frac{d}{dX_\alpha}(F+G)(X_\alpha), \alpha \in [0,1]\}$$

$$= \{\frac{d}{dX_\alpha}(F(X_\alpha) + G(X_\alpha)), \alpha \in [0,1]\}$$

$$= \{\frac{d}{dX_\alpha}F(X_\alpha) + \frac{d}{dX_\alpha}G(X_\alpha), \alpha \in [0,1]\}$$

$$= \{\frac{d}{dX_\alpha}F(X_\alpha), \alpha \in [0,1]\} + \{\frac{d}{dX_\alpha}G(X_\alpha), \alpha \in [0,1]\}$$

$$F'(X) + G'(X).$$

To show (b)

$$F'(IX) = \frac{d}{dX}F(IX) = \{\frac{d}{dX_\alpha}F(I_\alpha X_\alpha), \alpha \in [0,1]\}$$

$$= \{I_\alpha \frac{d}{dX_\alpha}F(I_\alpha X_\alpha), \alpha \in [0,1]\}$$

$$IF'(IX).$$

■

### 5.8.2 Interval Parameters Identification

In this section we present a generalized version of the **EM** algorithm (see Sec. 3.6) for identifying the unknown parameters $\Theta = \{\mathbf{A},\mathbf{H},\mathbf{Q},\mathbf{R}\}$ of the interval state space model (5.1) and (5.2). We use our special definitions of interval operators which we introduced in chapter 4, functions defined on an intervals, and differentiation with respect to intervals which we introduced in previous section. We also recall our definitions of interval random variables and their statistical properties (see section 4.4). To

generalize the EM algorithm in interval setting, we need to show that the basic mathematical operations such as Bayes' formula, trace,...etc that appear in the derivation of the EM algorithm are still valid in interval setting. We give these as lemmas in this section. Our justification of these results will allow us to generalize the EM algorithm to interval parameters identification.

The following lemma is a generalization of Bayes' formula to intervals.

**Lemma 57** ***(Bayes' Rule)*** *Let $X, Y \in \mathbb{IR}$ be interval random variables. Then*

$$P(X,Y) = P(X|Y)P(Y). \tag{5.41}$$

**Proof.**

$$\begin{aligned} P(X,Y) &= \{P(X_\alpha, Y_\alpha), \alpha \in [0,1]\} \\ &= \{P(X_\alpha|Y_\alpha)P(Y_\alpha), \alpha \in [0,1]\} \\ &= \{P(X_\alpha|Y_\alpha), \alpha \in [0,1]\}\{P(Y_\alpha), \alpha \in [0,1]\} \\ &P(X|Y)P(Y). \end{aligned}$$

■

Now, we want to show that the trace is linear for interval matrices.

**Lemma 58** *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{IR}^{n\times n}$. Then*

$$\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}). \tag{5.42}$$

**Proof.**

$$\begin{aligned}
\mathrm{tr}(\mathbf{A}+\mathbf{B}) &= \{\mathrm{tr}(A_\alpha + B_\alpha), \alpha \in [0,1]\} \\
&= \{\mathrm{tr}A_\alpha + \mathrm{tr}B_\alpha, \alpha \in [0,1]\} \\
&= \{\mathrm{tr}A_\alpha, \alpha \in [0,1]\} + \{\mathrm{tr}B_\alpha, \alpha \in [0,1]\} \\
&= \mathrm{tr}\mathbf{A} + \mathrm{tr}\mathbf{B}
\end{aligned}$$

■

**Lemma 59** *Let* $X = [a_1, b_1], Y = [a_2, b_2] \in \mathbb{IR}$ *and* $a_1, a_2 > 0$. *Then*

$$\mathrm{Log}(XY) = \mathrm{Log}X + \mathrm{Log}Y. \tag{5.43}$$

**Proof.**

$$\begin{aligned}
\mathrm{Log}(XY) &= \{\mathrm{Log}(X_\alpha Y_\alpha), \alpha \in [0,1]\} \\
&= \{\mathrm{Log}X_\alpha + \mathrm{Log}Y_\alpha, \alpha \in [0,1]\} \\
&= \{\mathrm{Log}X_\alpha, \alpha \in [0,1]\} + \{\mathrm{Log}Y_\alpha, \alpha \in [0,1]\} \\
&= \mathrm{Log}X + \mathrm{Log}Y
\end{aligned}$$

■

Now, let $\{\Upsilon_n, \mathbf{X}_n\}$ be the complete data where the states $\mathbf{X}_n = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ and the observations $\Upsilon_n = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$ are interval data.

Since all operations and properties that used to maximize Eqn (3.48) are well defined in interval setting by the convexity definition of intervals, we can use the same steps

that introduced in linear identification in Sec 3.6 for interval parameters identification of interval state space model (5.1) and (5.2). Under the Gaussian distribution assumption, the probability density $P(\mathbf{x}_0)$ is given by

$$P(\mathbf{x}_0) = (2\pi)^{-k/2}|\Sigma_0|^{-1/2}\exp(-1/2(\mathbf{x}_0-\mu_0)^T\Sigma_0^{-1}(\mathbf{x}_0-\mu_0)) \tag{5.44}$$

Based on (5.1) and (5.2), we can write the conditional densities for the interval state and output

$$\begin{aligned} P(\mathbf{y}_t|\mathbf{x}_t) &= (2\pi)^{-k/2}|\mathbf{R}|^{-1/2}\exp(-1/2(\mathbf{y}_t-\mathbf{H}\mathbf{x}_t)^T\mathbf{R}^{-1}(\mathbf{y}_t-\mathbf{H}\mathbf{x}_t)) \\ P(\mathbf{x}_{t+1}|\mathbf{x}_t) &= (2\pi)^{-k/2}|\mathbf{Q}|^{-1/2}\exp(-1/2(\mathbf{x}_t-\mathbf{A}\mathbf{x}_{t-1})^T\mathbf{Q}^{-1}(\mathbf{x}_t-\mathbf{A}\mathbf{x}_{t-1})). \end{aligned} \tag{5.45}$$

By the Bayes's formula and the same steps in (3.40), the joint likelihood for the complete data is given by

$$P(\Upsilon_n, \mathbf{X}_n) = \prod_{t=1}^n P(\mathbf{y}_t|\mathbf{x}_t)\prod_{t=1}^n P(\mathbf{x}_{t+1}|\mathbf{x}_t)P(\mathbf{x}_0). \tag{5.46}$$

and the joint log likelihood can be written as

$$\begin{aligned} &\mathrm{Log}P(\Upsilon_n, \mathbf{X}_n;\Theta) = -n/2\mathrm{Log}|\mathbf{R}| - \textstyle\sum_{t=1}^n(\mathbf{y}_t-\mathbf{H}\mathbf{x}_t)^T\mathbf{R}^{-1}(\mathbf{y}_t-\mathbf{H}\mathbf{x}_t) - n/2\mathrm{Log}|\mathbf{Q}| \\ &- \textstyle\sum_{t=1}^n(\mathbf{x}_t-\mathbf{A}\mathbf{x}_{t-1})^T\mathbf{Q}^{-1}(\mathbf{x}_t-\mathbf{A}\mathbf{x}_{t-1}) - 1/2\mathrm{Log}|\Sigma_0| - 1/2(\mathbf{x}_0-\mu_0)^T\Sigma_0^{-1}(\mathbf{x}_0-\mu_0) + \mathrm{Const.} \end{aligned} \tag{5.47}$$

Similar in section 3.6, we can calculate the conditional expectation of the joint likelihood of the complete data at iteration $j, j = 1, 2, ...$

$$
\begin{aligned}
&G(\Theta|\Theta^{(j-1)}) = E[\mathrm{Log}P(Y_n, X_n; \Theta)|Y_n, \Theta^{(j-1)}] \\
&= -n/2\mathrm{Log}|\mathbf{R}| - \mathrm{tr}\{\mathbf{R}^{-1}\sum_{t=1}^{n}[(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^n)(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^n)^T \\
&+\mathbf{H}\mathbf{P}_t^n\mathbf{H}^T]\} - n/2\mathrm{Log}|\mathbf{Q}| - 1/2\mathrm{tr}\{\mathbf{Q}^{-1}[\mathbf{D} - \mathbf{C}\mathbf{A}^T - \mathbf{A}\mathbf{C}^T + \mathbf{A}\mathbf{B}\mathbf{A}^T]\} - 1/2\mathrm{Log}|\Sigma_0| \\
&-1/2\mathrm{tr}|\{\Sigma_0^{-1}[\mathbf{x}_0^n - \mu_0)(\mathbf{x}_0^n - \mu_0)^T + \mathbf{P}_0^n]\} + \mathrm{Const},
\end{aligned}
\tag{5.48}
$$

where,

$$\mathbf{B} = \sum_{t=1}^{n}[\mathbf{x}_{t-1}^n(\mathbf{x}_{t-1}^n)^T + \mathbf{P}_{t-1}^n], \tag{5.49}$$

$$\mathbf{C} = \sum_{t=1}^{n}[\mathbf{x}_t^n(\mathbf{x}_{t-1}^n)^T + \mathbf{P}_{t,t-1}^n], \tag{5.50}$$

$$\mathbf{D} = \sum_{t=1}^{n}[\mathbf{x}_t^n(\mathbf{x}_t^n)^T + \mathbf{P}_t^n]. \tag{5.51}$$

In the above equations, the components $\mathbf{x}_{t-1}^n, \mathbf{x}_t^n, \mathbf{P}_{t-1}^n, \mathbf{P}_t^n$ and $\mathbf{P}_{t,t-1}^n$ can be calculated by using the CIKF equations and the CIKS equations (5.22)-(5.32).

Therefore, we find at iteration $j$ the update estimations of the unknown interval

parameters as

$$\mathbf{A} = \mathbf{C}\mathbf{B}^{-1} \tag{5.52}$$

$$\mathbf{Q} = 1/n(\mathbf{D} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T) \tag{5.53}$$

$$\mathbf{H} = (\sum_{t=1}^{n}(\mathbf{y}_t(\mathbf{x}_t^n)^T))(\sum_{t=1}^{n}(\mathbf{x}_t^n(\mathbf{x}_t^n)^T + \mathbf{P}_t^n))^{-1} \tag{5.54}$$

$$\mathbf{R} = 1/n\{\sum_{t=1}^{n}\mathbf{y}_t\mathbf{y}_t^T - (\sum_{t=1}^{n}\mathbf{y}_t(\mathbf{x}_t^n)^T)(\sum_{t=1}^{n}\mathbf{x}_t^n(x_t^n)^T + \mathbf{P}_t^n)^{-1}(\sum_{t=1}^{n}\mathbf{x}_t^n\mathbf{y}_t^T)\} \tag{5.55}$$

$$\mu_0 = \mathbf{x}_0^n, \tag{5.56}$$

The iterative procedure to obtain the maximum likelihood of the interval parameters of interval state space model is summarized as follows:

1. Initialize the procedure by selecting starting values for the parameters $\Theta^{(0)} = \{\mathbf{A}^{(0)}, \mathbf{H}^{(0)}, \mathbf{R}^{(0)}, \mathbf{Q}^{(0)}, \mu_0\}$ and use Eqns (5.22)-(5.32) to estimate the smoothed values $\mathbf{x}_t^n, \mathbf{P}_t^n$ and $\mathbf{P}_{t,t-1}^n$ with initial parameters.

2. Calculate the conditional expectation of the log-likelihood with Equation (5.48)

3. (**E**-step) Use Eqns (5.22)-(5.32) to estimate the smoothed values $\mathbf{x}_t^n, \mathbf{P}_t^n$ and $\mathbf{P}_{t,t-1}^n$, for $t = 1, 2, ..., n$, with parameters$\Theta^{(j-1)}, (j = 1, 2, ...)$.Use the smoothed value to calculate **B**, **C**, **D** in (5.49)-(5.51).

4. (**M**-step) Update the estimates,$\Theta$ using the eqns.(5.52)-(5.56), to obtain $\Theta^{(j)}$.

5. Repeat 2 to 4 above until the estimates and the log likelihood function are stable.

### 5.8.3 Simulation

In this subsection, we give a software simulation using MATLAB for interval parameters identification and the convexity interval Kalman filter.

In order to start the simulation, we consider the interval state space model (5.1) and (5.2) with assumptions (5.3)-(5.7), the parameters $\Theta = \{\mathbf{A}, \mathbf{H}, \mathbf{Q}, \mathbf{R}\}$ are unknown. Our goal is to identify these interval parameters using the **EM** algorithm discussed in subsection 5.8.2. As well, we will estimate the interval states $\hat{\hat{\mathbf{x}}}_t$ using interval Kalman filter simulation and compare the new state estimation with the estimation interval states $\hat{\mathbf{x}}_t$ introduced in section 5.7.

This comparison is defined by the distance formula given in Definition 25:

$$d[\hat{\hat{\mathbf{x}}}_t, \hat{\mathbf{x}}_t] = \max\{|\underline{\hat{\hat{\mathbf{x}}}_t} - \underline{\hat{\mathbf{x}}}_t|, |\overline{\hat{\hat{\mathbf{x}}}}_t - \overline{\hat{\mathbf{x}}}_t|\}.$$

Now, we start our simulation by the following initial values of $\Theta^{(0)}$:

$$\mathbf{A}^{(0)} = \begin{bmatrix} [2,3] & [.9,2] & [.5,1.5] \\ [0,1] & [.9,2] & [-.9,0] \\ [1,2] & [-2,-1] & [.9,1.5] \end{bmatrix}, \mathbf{H}^{(0)} = \begin{bmatrix} [0,1] & [.9,1.5] & [1,2] \end{bmatrix}$$

$$\mathbf{Q}^{(0)} = \begin{bmatrix} [.9,1] & [0] & [0] \\ [0] & [1,1.1] & [0] \\ [0] & [0] & [0] \end{bmatrix}, \mathbf{R}^{(0)} = \begin{bmatrix} [.9,1.1] \end{bmatrix}, \mu^{(0)} = 0_{3\times 1}.$$

We use the convexity definitions of interval operations and interval matrices (see Ch. 4).

As in section 5.7, we use the convexity definitions to partition the interval $[0,1]$ into $0 = \alpha_0 < \alpha_1 < ... < \alpha_n = 1$ and then choose the values of $\alpha = 0, 0.1, 0.2, ..., 1$. When $\alpha = 0$, we get the lower endpoints of all intervals and interval matrices. Hence, we retain the classical state space model and apply KF and the **EM** algorithm which gives the values of

$$A_{\alpha=0}, H_{\alpha=0}, Q_{\alpha=0}, R_{\alpha=0}.$$

Again, we take $\alpha = 0.1$ and we can get

$$A_{\alpha=0.1}, H_{\alpha=0.1}, Q_{\alpha=0.1}, R_{\alpha=0.1},$$

and so on until we get

$$A_{\alpha=1}, H_{\alpha=1}, Q_{\alpha=1}, R_{\alpha=1}.$$

Now, we find the intervals value by taking the minimum and the maximum for the above parameters. Hence, we have

$$A^I, H^I, Q^I, R^I.$$

Using these parameters, we can find out the estimation states $\hat{\hat{\mathbf{x}}}_t$.

Figures 5.3 and 5.4 show the error between the estimation states $\hat{\mathbf{x}}_t$ and the estimation sates in $\hat{\hat{\mathbf{x}}}_t$. The X-axis is the error and Y-axis is the time $t = 1, 2, ..., n$.

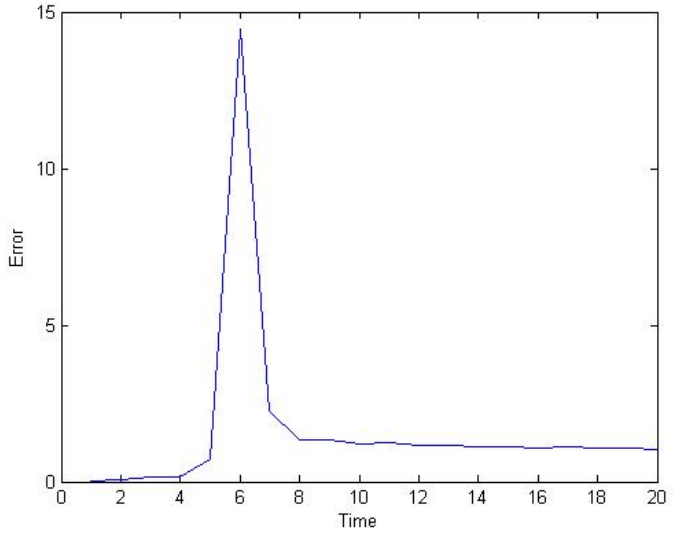


Figure 5.3: Identification of Interval Parameters (n=20)

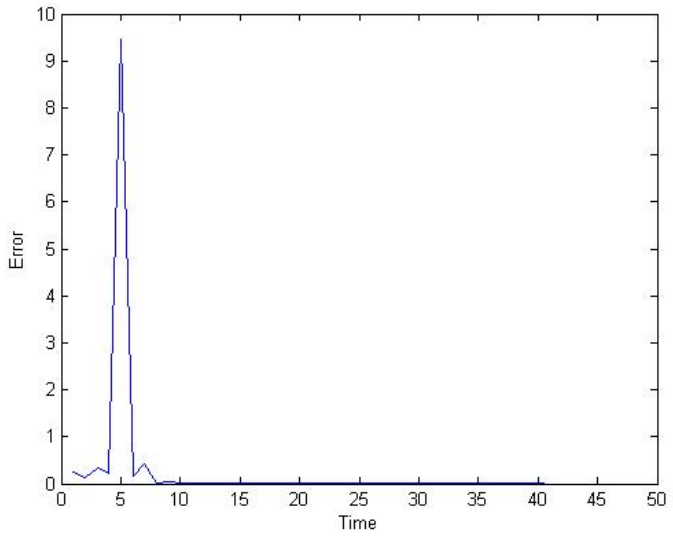


Figure 5.4: Identification of Interval Parameters (n=50)

## 5.9 Application to Weather Prediction

In this section we present an application of the convexity interval Kalman filter to the prediction of the temperature in Turaif which is located in the north of Saudi Arabia. Temperature records are obtained from the Presidency of Metrology and Environment (PME) in KSA. We use the IKF to predict a temperature interval rather than a single temperature. The IKF approach was applied to the prediction of daily temperature based on 10 years of observation data (1996-2005). The measured data was converted into intervals by adding and subtracting one degree to each average daily temperature. Let $y_t^I = \{y_1^I, y_2^I, ..., y_n^I\}$ be the observation data for $n$ days. Our goal is to find the prediction for the next day (for $t = n + 1$). This prediction will be an interval.

The interval parameters can be identified by the EM algorithm which was introduced in subsection 5.8.2. Applying the convexity interval Kalman filter to the interval state space model with Known parameters as introduced in Sec. 5.7, we can predict $(\hat{x}^I)_{n+1}^n$. This prediction $(\hat{x}^I)_{n+1}^n$ is used to compute the predicted value $\hat{y}_{n+1}^I$ which is compared with the measurement $y_{n+1}^I$ by computing the distance between the two intervals. We use various dimensions for the state vector ($k = 1, 2, 3, 4, 5$) with appropriate dimensions for the system parameters and show their results. Comparison between the results is based on computing the standard deviation of the errors. So, we compute

the standard deviation for

$$e = \min(|y - \hat{y}_L|, |y - \hat{y}_U|), \tag{5.57}$$

where $\hat{y}_L$ and $\hat{y}_U$ are the lower end point and the upper end point for $\hat{y}^I$, respectively. We summarize the procedure of this application as follows:

1. Convert the measured data into intervals.

2. Identify the parameters for interval state space model using interval parameters estimation introduced in subsection 5.8.2.

3. Apply IKF equations introduced in section 5.7 to find $(\hat{x}^I)_{n+1}^n$.

4. Substitute $(\hat{x}^I)_{n+1}^n$ in the measurement equation to compute $\hat{y}_{n+1}^I$.

5. Compare between $y_{n+1}^I$ and $\hat{y}_{n+1}^I$ by computing the distance between the two intervals.

6. Change the dimension of states to determine the best prediction model using the standard deviation of the errors.

The comparison results indicate the prediction intervals are in good agreement with the measurements intervals. The best model being obtained for system dimension of 1 to 4.

**Numerical Results:**

**Case 1 ($k = 1$):** Using the interval parameter identification introduced in subsection 5.8.2, we take the initial guess of the parameters of interval state space model (5.1) and

(5.2) as follows: $A^I = [0.8, 1.1]$, $H^I = [0.8, 1]$, $Q^I = [0.02]$ and $R^I = [0.01]$ with initial $x_0^I = [0]$, $\mu_0^I = [0]$ and $P_0^I = [0]$. The random noise $\mathbf{w}_t$ and $\mathbf{v}_t$ are uncorrelated with $\mathbf{w}_t \sim (0, Q^I)$ and $\mathbf{v}_t \sim (0, R^I)$. Let $y_t^I = \{y_1^I, y_2^I, ..., y_{2000}^I\}$ be observation temperatures which written in Excel file. Our goal is to find the prediction for the next day $(\hat{y}^I)_{2001}$. We obtained the numerical results $(\hat{y}^I)_{2001} = [12.1, 16.6]$. The measurement value of the the day 2001 is $y_{2001} = 13.6$, which clearly is included in the predicted interval. The prediction intervals are in a good agreement with the measurement (see Figure 5.5). Using Eqn (5.57), we have

$$\begin{aligned} \text{std}(e_t) &= \text{srd}(\min(|y_t - \hat{y}_t L|, |y_t - \hat{y}_t U|)), t = 1, 2, ..., 2000. \\ &= .78 \end{aligned} \tag{5.58}$$

**Case 2** ($k = 2$)**:** With same assumptions in case 1, we start with initial guess of the parameters of the interval state space model as follows:

$$A^I = \begin{bmatrix} [.7, 1] & [0] \\ [0] & [.7, 1] \end{bmatrix}, H^I = \begin{bmatrix} [0.8, 1] & [0.8, 1.1] \end{bmatrix}$$
$$Q^I = \begin{bmatrix} [.2] & [0] \\ [0] & [.2] \end{bmatrix}, R^I = \begin{bmatrix} 0.1, .3 \end{bmatrix}.$$

We obtained the numerical results $(\hat{y}^I)_{2001} = [11.2, 15.9]$. The measurement value of the the day 2001 is $y_{2001} = 13.6$, which clearly is included in the predicted interval

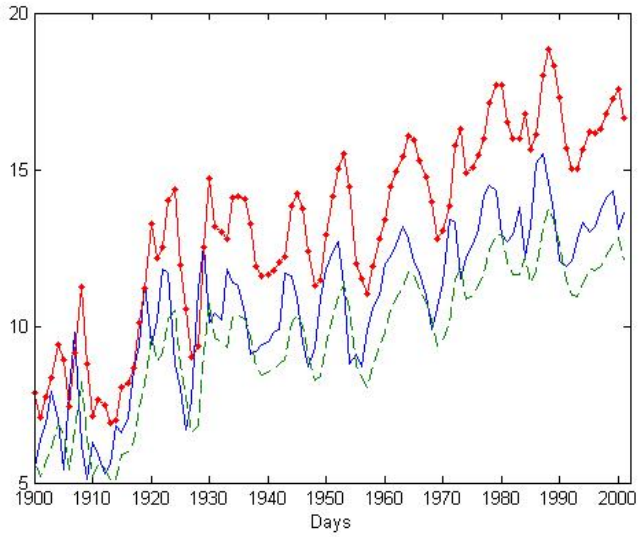


Figure 5.5: Case 1 of Interval Prediction(– observed value, - - the lower of interval prediction,-. the upper of interval prediction)

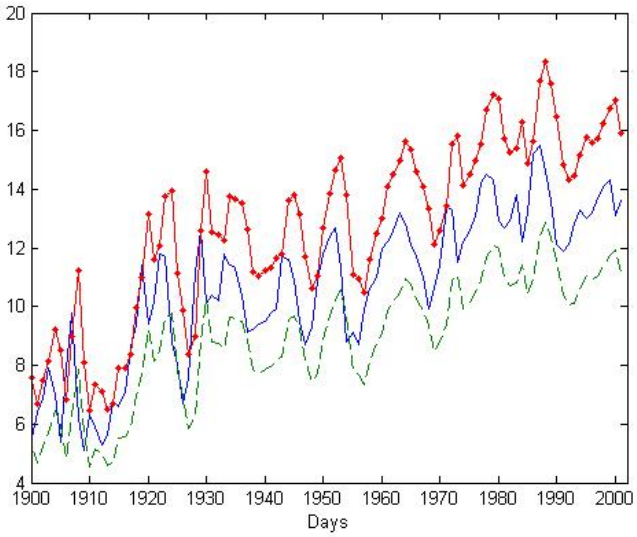


Figure 5.6: Case 2 of Interval Prediction(– observed value, - - the lower of interval prediction,-. the upper of interval prediction)

(see Figure 5.6). Using Eqn (5.57), we have

$$\text{std}(e_t) = \text{std}(\min(|y_t - \hat{y}_t L|, |y_t - \hat{y}_t U|)), t = 1, 2, ..., 2000. \tag{5.59}$$

$$= .84$$

**Case 3** ($k = 3$)**:** As in Case 1, we take the initial guess of the parameters of interval

state space model as follows:

$$A^I = \begin{bmatrix} [.7, 1.1] & [0] & [0] \\ [0] & [.8, 1] & [0] \\ [0] & [0] & [.7, 1.1] \end{bmatrix}, H^I = \begin{bmatrix} [0.8, 1] & [0.8, 0.9] & [.8, 9] \end{bmatrix}$$
$$Q^I = .03 I_{3\times 3}, R^I = \begin{bmatrix} 0.2, .03 \end{bmatrix}.$$

We obtained the numerical results $(\hat{y}^I)_{2001} = [11.1, 15.5]$. The measurement value of the the day 2001 is $y_{2001} = 13.6$, which clearly is included in the predicted interval. The prediction intervals are in a good agreement with the measurement (see Figure 5.7). Using Eqn (5.57), we have

$$\begin{aligned} &\text{std}(e_t) = \text{std}(\min(|y_t - \hat{y}_t L|, |y_t - \hat{y}_t U|)), t = 1, 2, ..., 2000. \\ &= .81 \end{aligned} \tag{5.60}$$

**Case 4 ($k = 4$):** Also, we take the initial guess of the parameters of interval state space model as follows:

$$A^I = \begin{bmatrix} [.7, 1.1] & [0] & [0] & [0] \\ [0] & [.8, 1] & [0] & [0] \\ [0] & [0] & [.8, 1.1] & [0] \\ [0] & [0] & [0] & [.8, 1] \end{bmatrix}, H^I = \begin{bmatrix} [0.8, 1] & [0.8, 1.1] & [.8, .9] & [.8, .9] \end{bmatrix}$$
$$Q^I = .03 I_{4\times 4}, R^I = \begin{bmatrix} 0.2 \end{bmatrix}.$$

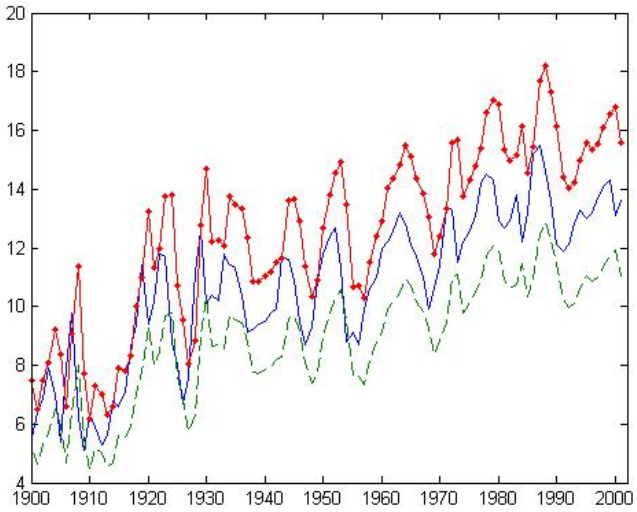


Figure 5.7: Case 3 of Interval Prediction(– observed value, - - the lower of interval prediction,-. the upper of interval prediction)

We obtained the numerical results $(\hat{y}^I)_{2001} = [12.5, 15.6]$. The measurement value of the the day 2001 is $y_{2001} = 13.6$, which clearly is included in the predicted interval. Using Eqn (5.57), we have

$$\text{std}(e_t) = .73 \tag{5.61}$$

**Case 5 ($k = 5$):** As in previous cases, we take the initial guess of the parameters of interval state space model as follows:

$$A^I = \begin{bmatrix} [.9, 1.1] & [0] & [0] & [0] & [0] \\ [0] & [.9, 1.1] & [0] & [0] & [0] \\ [0] & [0] & [.9, 1.1] & [0] & [0] \\ [0] & [0] & [0] & [.9, 1.1] & [0] \\ [0] & [0] & [0] & [0] & [.9, 1.1] \end{bmatrix},$$

$$H^I = \begin{bmatrix} [0.8, 1] & [0.8, 1.1] & [.9, 1.1] & [.8, 1] & [.8, 1] \end{bmatrix}, Q^I = .02 I_{5\times 5}, R^I = \begin{bmatrix} 0.2 \end{bmatrix}.$$

We obtained the numerical results $(\hat{y}^I)_{2001} = [14.5, 18]$. The measurement value of the the day 2001 is $y_{2001} = 13.6$,

Using Eqn (5.57), we have

$$\text{std}(e_t) = 1.3 \tag{5.62}$$

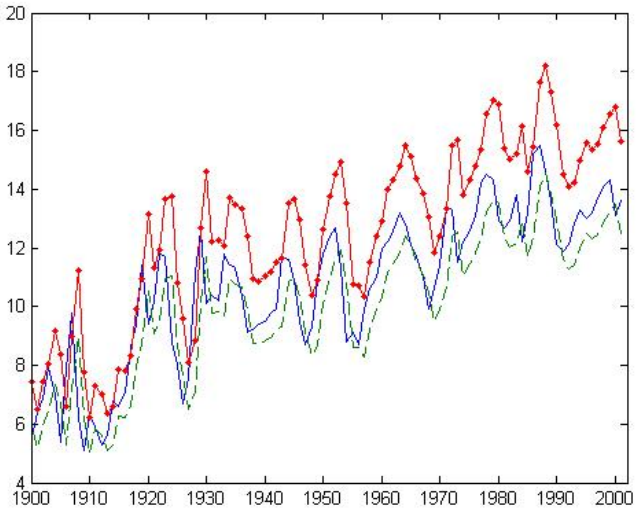


Figure 5.8: Case 4 of Interval Prediction(– observed value, - - the lower of interval prediction,-. the upper of interval prediction)

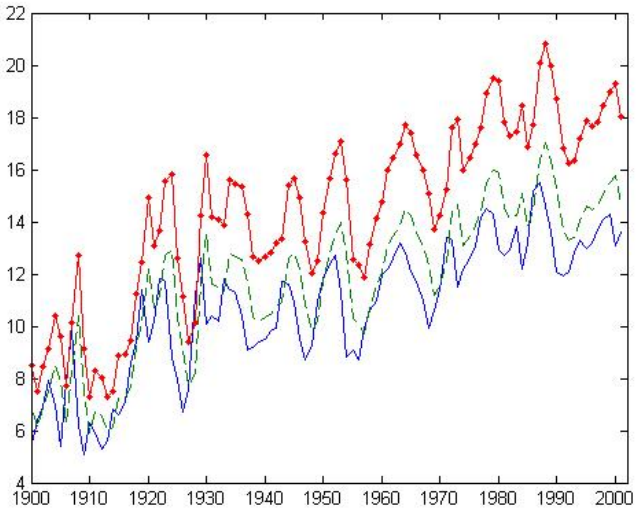


Figure 5.9: Case 5 of Interval Prediction(– observed value, - - the lower of interval prediction,-. the upper of interval prediction)

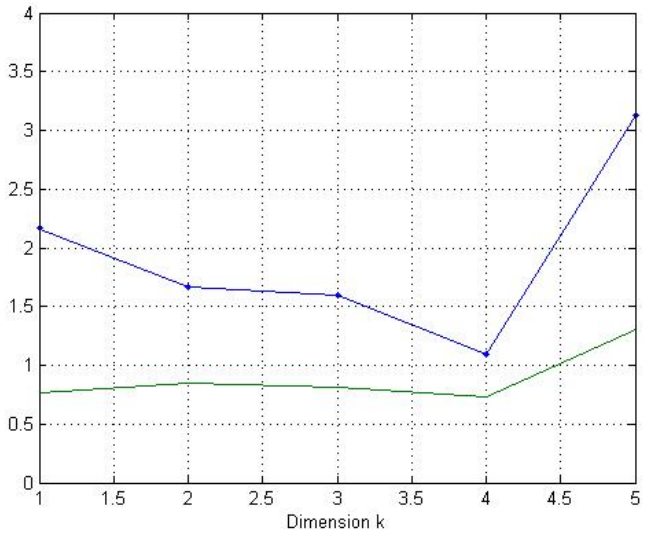


Figure 5.10: Comparison between dimensions (– the standard deviation, -. the error between $y_{2001}$ and $\hat{y}_{2001}$ )

**Conclusion :** .Comparison between the results for the previous cases by computing the standard deviation for the errors by Eqn (58)-(62). The standard deviation for their errors indicate the dimensions from 1 to 4 gives the best agreement between the interval prediction and measurements (see Figure 5.10).

# Bibliography

[1] G. Alefeld and J. Herzberger, Introduction to Interval Computations, Academic Press, New York 1983.

[2] G. Alferd, G. Mayer, Interval Analysis: Theory and Applications, Tournal of Computitional and Applied Mathimatics, 121(2000) 421-464.

[3] Andrew C. Harvey, Forecasting, structural time series models and the Kalman Filter, Cambrigde University Press, 1989.

[4] A. Andrews, M. Grewal, Kalman Filtering, Theory and Practice using Matlab. Wiley, 2001.

[5] Bartlomiej Jacek Kubica, Krzysztof Malinowski, "Interval Random Variables and Their Application in Queueing Systems with Long Tailed Service Times", Advances in Soft Computing, 2006, Volume 37/2006, 393-403, Innovations in hybrid intelligent systems.

[6] A.H. Bentbib, "Conjugate directions method for solving interval linear systems", Numerical Algorithms 21 (1999) 79-86.

[7] Brockwell P, Davis R. Time series: Theory and methods. 2nd ed. New York: Springer Verlag; 1991.

[8] Charles K., G. Chen, Kalman filtering with real time applications, Springer, 2009.

[9] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck. Analytic Moment-based Gaussian Process Filtering.

[10] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum lik elihood from uncomplete data via the EM algorithm, J. Royal ststist. Socser.B, 39, 1977.

[11] A. Doucet and V. B. Tadic. Parameter estimation in general state-space models using particle methods. Annals of the Institute of Statistical Mathematics, 55:409-422, 2003.

[12] Durbin J. and Koopman S.J. (2001). Time Series Analysis by State Space Methods, Oxford University Press.

[13] Evensen, G., 2007: Data assimilation: the ensemble Kalman filter. Springer.

[14] K. Ganesan and P. Veeramani, On Arithmetic Operations of Interval Numbers, International Journal of Uncertainty, Fuzziness and Knowledge - Based Systems, 13 (6) (2005), 619 - 631.

[15] K. Ganesan, On Some Properties of Interval Matrices, International Journal of Computational and Mathematical Sciences, 1 (2) (2007), 92-99.

[16] Gene H. Golub and Charles F. van Loan. Matrix Computations. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[17] Z. Ghahramani and G.E. Hinton. Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, University of Toronto, 1996.

[18] G.Ghen, J.Wang,L.Shieh. Interval Kalman Filter. IEEE VOL 33,N0 1,January 1997.

[19] E. R. Hansen. Global Optimization Using Interval Analysis. Marcel Dekker, New York, NY, 1992.

[20] E. R. Hansen. On solving systems of equations using interval arithmetic. Mathematics of Computation, 22(102):374(384), April ,1968.

[21] E. R. Hansen and R.R. Smith, Interval Matrix Computations, Part 2, SIAM. Journal of Numerical Analysis, Vol 4, p 1-9, 1967.

[22] G. I. Hargreaves, Interval Analysis in MATLAB, Manchester centre for computational mathmatics, Numerical Analysis Report, 2002.

[23] Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge, Cambridge University Press.

[24] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied Interval Analysis. Springer Verlag, London, 2001.

[25] Jean. Pierre Aubin, H. Frankowska, Set Valued Analysis, Birkhanser, 1990.

[26] Jeff A. Bilmes, A Gentle Tutorial of the EM algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical report TR-97-021, ICSI, 1997.

[27] Jensen J.L., Petersen, N.V., (1999). Asymptotic normality of the maximum likelihood estimator in state space models, Annals of Statistics, 27, 514-535.

[28] De Jong, P. (1988). The likelihood for a state space model, Biometrika, 75, 165-169.

[29] Juang, J.-N., Phan, M.Q., Horta, L.G., and Longman, R.W., "Identification of Observer/Kalman Filter Markov Parameters: Theory and Experiments," Journal of Guidance, Control, and Dynamics, Vol. 16, No. 2, 1993, pp. 320-329.

[30] Juang, J.-N., Applied System Identification, Prentice-Hall, Englewood Cliffs, NJ, 1994, pp. 175-252.

[31] T. Kailath, A. Sayed, B. Hassibi, "Linear Estimation", Prentice Hall, 2000.

[32] R. E. Kalman, "A new approach to linear filter and prediction problems,"J. Basic Eng., vol. 82, pp. 35-45, 1960.

[33] R. E. Kalman and R. S. Bucy, "New results of linear filtering and prediction theory," J. Basic Eng., vol. 83, pp. 95-108, 1961.

[34] E. Kalnay, Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 2003.

[35] M. Kieffer, L. Jaulin, and E. Walter. Guaranteed recursive nonlinear state bounding using interval analysis. International Journal of Adaptative Control and Signal Processing, 6(3):193-218, 2002.

[36] M. Kieffer and E. Walter. Guaranteed nonlinear state estimator for cooperative systems. Numerical Algorithms, 2003.

[37] L. Kolev. A method for outer interval solution of linear parametric systems. Reliable Computing, 10, pp. 227-239, 2004.

[38] L.V. Kolev. Solving Linear Systems whose elements are nonlinear functions of interval parameters. Numerical Algorithms, 37: 199-212, 2004.

[39] L.V. Kolev. An Improved interval linearization for solving non-linear problems. Numerical Algorithms, 37: pp. 213-224, 2004.

[40] N. R. Kristensen, H. Madsen, and S. B. Jorgensen. Parameter estimation in stochastic grey-box models. Automatica, 40(2):225-237, February 2004.

[41] L. Ljung. System identification, Theory for the user. System sciences series. Prentice Hall, Upper Saddle River, NJ, USA, second edition, 1999

[42] L. Ljung and A. Vicino, editors. Special Issue 'System Identification: Linear vs Nonlinear'. IEEE Transactions on Automatic Control, 2005.

[43] M. Lyashko, The optimal soution of an interval system of linear algebraic equations, Reliable computing, 11: 105127, 2005.

[44] D. Moens and D. Vandepitte. A survey of non-probabilistic uncertainty treatment in .nite element analysis. Computer Methods in Applied Mechanics and Engineering, 194(14-16):1527-1555, 2005.

[45] R. Moore. Interval Arithmetic. Prentice-Hall, Englewood Clis, NJ, USA, 1966.

[46] R. E. Moore. Methods and Applications of Interval Analysis. SIAM, Philadelphia, PA, 1979.

[47] N. S. Nedialkov, K. R. Jackson, J. D. Pryce, An eective high order interval method for validating existence and uniqueness of the solution of an IVP for an ODE, Reliable Computing, 7, No. 6 (2001), 449-465.

[48] Neumaier, A. (1990). Interval Methods for Systems of Equations. Cambridge University Press, Cambridge, England.

[49] S. Ning and R. B. Kearfott. A comparison of some methods for solving linear interval equations. SIAM Journal on Numerical Analysis, 34(4):1289(1305), August 1997

[50] T. Nirmala, D. Datta, H.S. Kushwaha, K. Ganesan. Inverse Interval Matrix: A New Approach. Applied Mathematical Sciences, Vol. 5, 2011, no. 13, 607- 624.

[51] Phan, M.Q., Horta, L.G., Juang, J.-N., and Longman, R.W., "Improvement of Observer/Kalman Filter Identification (OKID) by Residual Whitening, " Journal of Vibrations and Acoustics, Vol. 117, April 1995, pp. 232-239.

[52] R. Pintelon and J. Schoukens. System Identification . A Frequency Domain Approach. IEEE Press, New York, 2001.

[53] R. Pintelon, J Schoukens, and G Vandersteen. Frequency domain system identification using arbitrary signals. IEEE Trans. Automatic Control, 42:17171.1720, 1997.

[54] J. Rohn, Inverse interval matrix, SIAM Journal of Numerical Analysis, 3 (1993), 864 - 870.

[55] S. Ross, Introduction to Probability Models. 9e. Academic Press, 2007.

[56] J. Schoukens and R. Pintelon. identi cation of Linear Systems: A Practical Guideline to Accurate Modeling. Pergamon Press, London (U.K.), 1991.

[57] G. Seber and A. Lee. Linear Regression Analysis. John Wiley and Sons, 2002.

[58] R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications With R Examples, Springer, 2006.

[59] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis, 3(4):253-264, 1982.

[60] I. Skalna, M.V. Rama Rao, A. Pownuk, Systems of fuzzy equations in structural mechanics, Journal of Computational and Applied Mathematics, 218(2008), 149-156.

[61] T. Soderstrom and P. Stoica. System identification. Systems and Control Engineering. Prentice Hall, 1989.

[62] Stoffer, D.S. and K.D.Wall (1991). Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. J. Am. Stat. Assoc., 86, 1024-1033.

[63] E. Walter and M.Kieffer,Interval Analysis for Guaranteed Nonlinear Parameter and State Etimation,Draft Article ,France,2004.

[64] Wei Chen and Shaohua Tan, "Robust Portfolio Selection using Interval Random Programming", FUZZ-IEEE, Korea 2009, August 20-24, 2009.

[65] Xiaodong Luo, Recursive Bayesian Filters for Data Assimilation, Ph.D thesis, Oxford University, 2009.

# VITAE

- OBAID JEFAIN ALGAHTANI
- Born in Bisha, Saudi Arabia on July 1, 1978.
- Received B.S. degree in Mathematics from King Khaled University, in 2001.
- Received M.S. degree in Mathematics from King Abdul Aziz University, Jeddah, in 2007.
- Faculty member, Mathematics Department, Teacher College, King Saud University, Riyadh, since 2001.
- Joined, in 2007, Ph.D. program in the Department of Mathematics and statistics, King Fahd University of Petroleum and Minerals.
- **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:

  **1** obaid8888@hotmail.com

  **2** g200704370@kfupm.edu.sa

- **Permanent Address:** Department of Mathematics, King Saud University, Riyadh, Saudi Arabia.