การประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM

PARAMETER ESTIMATION USING EM ALGORITHM

นายหิรัญรักษ์ รักษ์ฤทธิกุล รหัส 55364459

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

ปีการศึกษา 2559

ใบรับรองปริญญานิพนธ์

หัวข้อโครงงาน	การประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM
ผู้ดำเนินโครงงาน	นายหิรัญรักษ์ รักษ์ฤทธิกุล รหัส 55364459
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ธนิต มาลากร
สาขาวิชา	วิศวกรรมไฟฟ้า
ภาควิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ปีการศึกษา	2559

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร อนุมัติให้โครงงานฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

....ที่ปรึกษาโครงงาน
(รองศาสตราจารย์ ดร.ธนิต มาลากร)

....กรรมการ
(รองศาสตราจารย์ ดร.ไพศาล มุณีสว่าง)

....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.พรพิศุทธิ์ วรจิรันตน์)

หัวข้อโครงงาน	การประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM
ผู้ดำเนินโครงงาน	นายหิรัญรักษ์ รักษ์ฤทธิกุล รหัส 55364459
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ธนิต มาลากร
สาขาวิชา	วิศวกรรมไฟฟ้า
ภาควิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ปีการศึกษา	2559

....

บทคัดย่อ

วัตถุประสงค์หลักของโครงงานนี้เป็นการใช้ตัวกรองคาลมานร่วมกับขั้นตอนวิธี EM ในการประมาณพารามิเตอร์แบบความควรจะเป็นสูงสุดของแบบจำลองเชิงเฟ้นสุ่มเชิงเส้นโดยข้อมูลที่ใช้เป็นกรณีศึกษาถูกแบ่งเป็น 2 กลุ่มคือข้อมูลที่ได้จากการจำลองและข้อมูลจริง

ข้อมูลที่ได้จากการจำลองถูกใช้ในการทดสอบโปรแกรมที่ได้จัดทำขึ้น โดยผู้จัดทำจะกำหนดพารามิเตอร์ของระบบเพื่อจำลองสัญญาณออกแล้วทำการวนซ้ำเพื่อประมาณพารามิเตอร์ ในโครงงานนี้เลือกใช้แบบจำลองระบบควบคุมความเร็วรถและแบบจำลองตัวกรองดิจิตอลแบบ IIR เป็นกรณีศึกษา จากการทดลองพบว่าพารามิเตอร์ที่ประมาณได้ในทั้งสองตัวอย่างมีค่าใกล้เคียงกับพารามิเตอร์ที่ถูกกำหนดขึ้น

สำหรับข้อมูลจริงที่ถูกนำมาใช้ในแบบจำลองความผันผวนเชิงเฟ้นสุ่มคืออัตราแลกเปลี่ยนดอลลาร์สหรัฐต่อบาทไทย จากการศึกษาพบว่าพารามิเตอร์ที่ได้จากการประมาณลู่เข้าสู่ค่าค่าหนึ่งและฟังก์ชันควรจะเป็นมีค่าเพิ่มสูงขึ้น

Project title	Parameter Estimation using EM Algorithm
Name	Mr. Hirunrak Rakrittikul ID. 55364459
Project advisor	Assc. Prof. Tanit Malakorn, Ph.D.
Major	Electrical Engineering
Department	Electrical and Computer Engineering
Academic year	2016

Abstract

The main purpose of this project is to present the application of the EM algorithm based on the Kalman method to estimate the parameters of stochastic linear models. The EM algorithm has been applied into two types of case studies consisting of simulation data and real data.

The simulation data has been used in order to test the program. The parameters of the system have been assigned to generate output data and estimate the parameters. There are two examples namely a cruise control model and an infinite impulse response filter model. The results of these illustration that parameters computed by the Kalman method in the EM algorithm are slightly different compared to true parameters.

The real data has been applied in a stochastic volatility model using USD/THB exchange rate. The estimated parameters in this case converged and likelihood function has been increased.

กิตติกรรมประกาศ

โครงงานวิศวกรรมไฟฟ้าเรื่องการประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM ฉบับนี้สำเร็จลุล่วงได้ เนื่องมาจากผู้จัดทำได้รับความอนุเคราะห์จากรองศาสตราจารย์ ดร. ธนิต มาลากร ซึ่งเป็นอาจารย์ที่ปรึกษาโครงงานนี้ ที่กรุณาให้คำปรึกษา แนะนำวิธีการทำงาน พร้อมทั้งแนะแนวทางในการแก้ปัญหาต่าง ๆ และคอยกระตุ้นให้ผู้จัดทำโครงงานทำงานอย่างต่อเนื่อง ตลอดจนสละเวลาอันมีค่าเพื่อตรวจสอบและแก้ไขข้อบกพร่องต่างๆ ผู้จัดทำรู้สึกเป็นเกียรติอย่างมากที่ได้รับความอนุเคราะห์จากอาจารย์

ในโอกาสนี้ทางผู้จัดทำโครงงานใคร่ขอขอบคุณรองศาสตราจารย์ ดร.ไพศาล มุณีสว่าง และผู้ช่วยศาสตราจารย์ ดร.พรพิศุทธิ์ วรจิรันตน์ ที่สละเวลามาร่วมเป็นคณะกรรมการในการสอบโครงงานรวมทั้งให้คำแนะนำ ชี้แนะแนวทาง ให้ข้อคิดเห็นต่างๆ ที่เป็นประโยชน์เพื่อให้โครงงานนี้มีความสมบูรณ์ยิ่งขึ้น

ในท้ายที่สุดนี้ผู้จัดทำโครงงานจึงขอขอบพระคุณอาจารย์ทุกท่าน บิดา มารดา ที่คอยสั่งสอน ให้ความรู้จนผู้จัดทำสำเร็จการศึกษา ทั้งนี้ขอขอบคุณเพื่อน ๆ ที่ได้ให้คำแนะนำในการเขียนโปรแกรม และคอยให้กำลังใจ ช่วยให้คำปรึกษาทั้งในเรื่องเรียนและในเรื่องส่วนตัวจนสำเร็จลุล่วงมาด้วยดี

นายหิรัญรักษ์ รักษ์ฤทธิกุล

สารบัญ

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญรูป

รูปที่ หน้า

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของโครงงาน

การประมาณพารามิเตอร์มีบทบาทสำคัญในการวิเคราะห์ระบบเนื่องจากหากพิจารณาในเชิงปฏิบัติแล้วสิ่งที่ผู้ออกแบบทราบคือข้อมูลดิบที่วัดหรือสังเกตออกมาได้จากระบบ การประมาณพารามิเตอร์คือการนำข้อมูลดิบดังกล่าวมาผ่านกระบวนการวนซ้ำเพื่อคำนวณหาพารามิเตอร์ของระบบ โดยกระบวนการวนซ้ำที่นิยมเลือกใช้ในการประมาณพารามิเตอร์ของระบบที่ถูกบรรยายในรูปแบบปริภูมิสถานะ คือ **ขั้นตอนวิธี EM (Expectation-Maximization algorithm)** ซึ่งประกอบไปด้วย 2 ขั้นตอน กล่าวคือ ขั้นตอน E (Expectation step) และขั้นตอน M (Maximization step) ขั้นตอนทั้งสองจะทำงานสลับกันในแต่ละรอบของการวนซ้ำเพื่อทำให้ฟังก์ชันควรจะเป็น (Likelihood function) มีค่าสูงขึ้นในทุกรอบ ด้วยเหตุนี้ขั้นตอนวิธี EM จึงจัดว่าเป็นวิธีการในการหาความควรจะเป็นสูงสุด (Maximum likelihood : ML)

สิ่งแรกที่ขั้นตอนวิธี EM ต้องทำในขั้นตอน E คือการสังเคราะห์ตัวแปรสถานะ x_k จากข้อมูลดิบที่ได้มาซึ่งอยู่ในรูปของอนุกรมเวลา (Time series data) เนื่องจากข้อมูลที่วัดมาได้นั้นมักมีการปนเปื้อนจากสัญญาณรบกวนความถี่สูง ดังนั้นจึงต้องอาศัยทั้งตัวกรอง(Filter) ร่วมกับตัวปรับเรียบ(Smoother) เพื่อทำการสังเคราะห์ตัวแปรสถานะออกมา ภายใต้สมมติฐานที่ว่าระบบภายใต้การพิจารณาเป็นระบบเชิงเส้น ตัวกรองและตัวปรับเรียบที่สามารถลดค่าเฉลี่ยของความผิดพลาดกำลังสอง (Mean square error) ระหว่างตัวแปรสถานะจริงของระบบและตัวแปรสถานะที่ได้จากการประมาณให้ได้น้อยที่สุดคือ **ตัวกรองและตัวปรับเรียบคาลมาน (Kalman filter : KF&Kalman smoother : KS)** อย่างไรก็ตามหากระบบภายใต้การพิจารณาเป็นระบบไม่เชิงเส้น KF & KS ถือว่าเป็นตัวกรองและตัวปรับเรียบเชิงเส้นที่ดีที่สุดในบรรดาตัวกรองและตัวปรับเรียบเชิงเส้นประเภทอื่น ๆ

ในโครงงานนี้จึงศึกษาขั้นตอนการทำงานของ KF & KS ซึ่งเป็นส่วนหนึ่งของขั้นตอนวิธี EM ในการสังเคราะห์พารามิเตอร์ของระบบโดยระบบที่นำมาใช้ศึกษาถูกแบ่งเป็น 2 กลุ่ม นั่นคือ กลุ่มที่ได้จากการจำลอง (Simulation) ซึ่งเป็นกลุ่มที่ใช้ในการทดสอบโปรแกรมที่ได้จัดทำขึ้นและ

กลุ่มที่ 2 คือกลุ่มที่เป็นข้อมูลจริง เช่น ข้อมูลของอัตราแลกเปลี่ยนเงินตราระหว่างประเทศ ข้อมูลของมูลค่าหลักทรัพย์ ข้อมูลของภาพถ่าย หรือข้อมูลทางด้านสถิติชีวภาพ (Biometrics) เป็นต้น

1.2 วัตถุประสงค์ของโครงงาน

1. เพื่อเข้าใจหลักการทำงานของตัวกรองและตัวปรับเรียบคาลมาน(KF & KS) รวมทั้งขั้นตอนวิธี EM
2. เพื่อสามารถประยุกต์ใช้ทฤษฎีด้านความน่าจะเป็นมาใช้กับกระบวนการเชิงเฟ้นสุ่มได้
3. เพื่อฝึกการเขียนโปรแกรม MATLAB ในการประมาณพารามิเตอร์ของระบบ

1.3 ขอบเขตในการดำเนินโครงงาน

1. ระบบที่ศึกษาเป็นระบบเชิงเส้นไม่แปรตามเวลาที่ถูกบรรยายในรูปแบบปริภูมิสถานะในเวลาวิยุต
2. ใช้ขั้นตอนวิธี EM ร่วมกับ KF & KS ในการประมาณพารามิเตอร์ของระบบ
3. ข้อมูลในอนุกรมเวลาเป็นแบบหนึ่งตัวแปร (Univariate) โดยมีสัญญาณรบกวนแบบเกาส์เซียน
4. ทำการจำลองโดยใช้โปรแกรมMATLAB

1.4 ขั้นตอนและแผนการดำเนินงาน

กิจกรรม	ระยะเวลาดำเนินงาน									
	ปี 2558					ปี 2559				
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
1. ศึกษาทฤษฎีความน่าจะเป็นที่เกี่ยวข้องกับกระบวนการเชิงเฟ้นสุ่ม	↔	↔								
2. ศึกษาการทำงานของ KF & KS		↔	↔	↔						
3. ศึกษาการทำงานของขั้นตอนวิธี EM					↔	↔				
4. จำลองผลการทดลอง							↔	↔		
5. สรุปผลการทดลองและเขียนรายงานฉบับสมบูรณ์								↔	↔	↔

1.5 ผลที่คาดว่าจะได้รับจากโครงงาน

1. เข้าใจถึงวิธีการประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM สำหรับระบบเชิงเส้นไม่แปรตามเวลาที่บรรยายในรูปแบบปริภูมิสถานะในเวลาวิยุต
2. สามารถนำความรู้ที่ได้ไปประยุกต์ใช้กับข้อมูลอื่น ๆ เพื่อใช้สร้างแบบจำลองของระบบได้
3. เป็นพื้นฐานในการศึกษาวิธีการประมาณประเภทอื่นที่ใช้กับระบบไม่เชิงเส้น หรือระบบเชิงเส้นที่ปนเปื้อนด้วยสัญญาณรบกวนแบบอื่นทีไม่เป็นเกาส์เซียนได้

1.6 งบประมาณ

1.	ค่าอุปกรณ์ในการทำโครงงาน	500 บาท
2.	จัดทำรูปเล่มปริญญานิพนธ์	500 บาท
	รวมเป็นเงินทั้งสิ้น (หนึ่งพันบาทถ้วน)	1000 บาท

หมายเหตุ : ถัวเฉลี่ยทุกรายการ

บทที่ 2
หลักการและทฤษฎีที่เกี่ยวข้อง

ทฤษฎีการประมาณเป็นสาขาหนึ่งของทฤษฎีการควบคุมและการประมวลผลสัญญาณ ซึ่งศึกษาเกี่ยวกับการสร้างตัวประมาณ (Estimator) เพื่อใช้ในการประมาณพารามิเตอร์ วิธีการสร้างตัวประมาณที่นำเสนอในโครงงานนี้คือขั้นตอนวิธี EM ซึ่งประกอบไปด้วยขั้นตอน E และขั้นตอน M ทั้งนี้การประยุกต์ใช้ขั้นตอนวิธี EM ในการประมาณค่าพารามิเตอร์ของระบบที่บรรยายในรูปแบบปริภูมิสถานะจำเป็นต้องประมาณตัวแปรสถานะ $\hat{x}_k$ ด้วย ในที่นี้ผู้จัดทำได้เลือกใช้การสังเคราะห์ตัวแปรสถานะ $\hat{x}_k$ ด้วยวิธีการของคาลมาน เนื้อหาในบทนี้จึงขอนำเสนอทฤษฎีพื้นฐานที่จำเป็นสำหรับโครงงานนี้ ได้แก่ ทฤษฎีความน่าจะเป็น ทฤษฎีเมทริกซ์ แบบจำลองปริภูมิสถานะ แบบจำลองมาร์คอฟ วิธีการของคาลมาน และขั้นตอนวิธี EM

2.1 ทฤษฎีความน่าจะเป็น

ในหัวข้อนี้จะกล่าวถึงทฤษฎีความน่า (Probability theory) จะเป็นที่มีความสำคัญต่อการศึกษาการทำงานของกระบวนการเชิงเฟ้นสุ่ม (Stochastic process) และขั้นตอนวิธี EM อาทิ เช่น พื้นฐานความน่าจะเป็น ตัวแปรสุ่ม ค่าคาดหมาย ค่าความแปรปรวน ฟังก์ชันการแจกแจงและการแจกแจงความน่าจะเป็นร่วม

2.1.1 ความน่าจะเป็นพื้นฐาน (Basic probability)

เมื่อกล่าวถึงความน่าจะเป็นจำเป็นต้องทราบองค์ประกอบหลักทั้ง 4 กล่าวคือ การทดลองสุ่ม ปริภูมิตัวอย่าง เหตุการณ์และพีชคณิตซิกมา โดยทั้ง 4 องค์ประกอบมีนิยามดังนี้

<u>**นิยาม 2.1**</u> การทดลองสุ่ม (Random experiment) หมายถึงการทดลองซึ่งให้ผลลัพธ์ที่มีความไม่แน่นอน มิอาจคาดเดาผลล่วงหน้าได้ สามารถเปลี่ยนแปลงแตกต่างกันไปในแต่ละครั้งของการทดลอง ถึงแม้จะเป็นการทดลองซ้ำภายใต้สภาพแวดล้อมเดิม ตัวอย่างของการทดลองสุ่ม ได้แก่ การโยนเหรียญ และการโยนลูกเต๋า

<u>**นิยาม 2.2**</u> ปริภูมิตัวอย่าง (Sample space) หมายถึงเซตของผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองสุ่ม ในทางคณิตศาสตร์แทนด้วยสัญลักษณ์ S ยกตัวอย่างเช่น การโยนเหรียญ 1 เหรียญ 1 ครั้งมีปริภูมิตัวอย่างเป็นหัวและก้อยนั่นคือ $S=\{H,T\}$

<u>**นิยาม 2.3**</u> เหตุการณ์ (Event) คือเซตย่อยของปริภูมิตัวอย่าง โดยทั่วไปแทนด้วยสัญลักษณ์ E ยกตัวอย่างเช่น เหตุการณ์ที่ผลบวกตัวเลขของลูกเต๋าสองลูกมีค่าเท่ากับ 4 คือ $E=\{(1,3),(2,2),(3,1)\}$

นิยาม 2.4 พีชคณิตซิกมา (Sigma algebra) หรือสนามซิกมา (Sigma field) หมายถึงเซตของเหตุการณ์ที่เป็นไปได้ทั้งหมดของปริภูมิตัวอย่าง โดยสามารถหาได้จากเซตกำลัง (Power set) ของปริภูมิตัวอย่าง ยกตัวอย่างเช่นในการโยนเหรียญ 1 เหรียญ 1 ครั้ง ซึ่งมีปริภูมิตัวอย่างคือ $S = \{H, T\}$ ดังนั้นพีชคณิตซิกมาในกรณีนี้คือเซต $\{\phi, H, T, \{H, T\}\}$

นิยาม 2.5 กำหนดให้เหตุการณ์ที่ 1 เป็น A และเหตุการณ์ที่ 2 เป็น B ความน่าจะเป็นของการเกิดเหตุการณ์ A โดยเหตุการณ์ B ได้เกิดขึ้นแล้ว สามารถเขียนแทนได้เป็น $P(A|B)$ ซึ่งสามารถหาได้จากสมการ (2.1)

$$P(A|B) = \left.\frac{P(A \cap B)}{P(B)}\right\}; P(B) > 0 \tag{2.1}$$

นิยาม 2.6 เหตุการณ์ A และเหตุการณ์ B จัดว่ามีความเป็นอิสระต่อกันก็ต่อเมื่อเหตุการณ์ทั้งสองสอดคล้องกับสมการ (2.2)

$$P(A \cap B) = P(A)P(B) \tag{2.2}$$

2.1.2 ตัวแปรสุ่ม (Random variable)

ในการทดลองสุ่มนั้นสามารถให้ผลลัพธ์ที่เป็นตัวเลขหรือไม่ใช่ตัวเลขก็ได้ เช่น การทอดลูกเต๋าให้ผลลัพธ์เป็นแต้มตั้งแต่ 1 ถึง 6 แต่ในขณะที่การโยนเหรียญให้ผลลัพธ์เป็นหัวหรือก้อย อย่างไรก็ตามสิ่งที่สนใจมักไม่ใช่ผลลัพธ์ที่มาจากการทดลองโดยตรง หากแต่เป็นตัวเลขที่มีการกำหนดให้กับผลลัพธ์ที่เกิดขึ้นในแต่ละรูปแบบ นี่จึงเป็นที่มาของนิยามต่อไปนี้

นิยาม 2.7 ให้ S เป็นปริภูมิตัวอย่างแล้วตัวแปรสุ่มคือฟังก์ชันที่กำหนดค่าตัวเลขจำนวนจริง $X(\varsigma)$ ให้กับผลลัพธ์ ς แต่ละแบบในปริภูมิตัวอย่างของการทดลองสุ่ม ยกตัวอย่างเช่น การโยนเหรียญมีปริภูมิตัวอย่าง $S = \{H, T\}$ หากกำหนดให้ผลลัพธ์ของการออก H เป็น 0 และผลลัพธ์ของการออก T เป็น 1 กล่าวคือ $X(H) = 0$ และ $X(T) = 1$ แล้ว $X(\varsigma)$ คือตัวแปรสุ่มนั่นเอง

1. ตัวแปรสุ่มวิยุต (Discrete random variable)

ตัวแปรสุ่ม X ถูกกล่าวว่าเป็นตัวแปรสุ่มวิยุตก็ต่อเมื่อเรนจ์ของ X สามารถนับได้ (Countable set) สำหรับฟังก์ชันมวลความน่าจะเป็น (Probability mass function : PMF) ของตัวแปรสุ่มวิยุต X คือฟังก์ชันที่มีคุณสมบัติดังนี้

1) $P(X = x_i) = f(x_i) > 0$
2) $\sum_{k=i}^{n} f(x_k) = 1$
3) $P(X \in B) = \sum_{x \in B} f(x)$ เมื่อ B คือเหตุการณ์ที่สนใจ

2. ตัวแปรสุ่มต่อเนื่อง (Continuous random variable)

ตัวแปรสุ่ม X ถูกกล่าวว่าเป็นตัวแปรสุ่มต่อเนื่องก็ต่อเมื่อเรนจ์ของ X มีค่าได้ไม่จำกัดและไม่สามารถนับได้ (Uncountable set) สำหรับฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function : PDF) ของตัวแปรสุ่มต่อเนื่อง X คือฟังก์ชันที่มีคุณสมบัติดังนี้

1) $f(x) > 0$ ทุกค่า $x \in S$

2) $\int_{-\infty}^{\infty} f(x) = 1$

3) $P(X \in B) = \int_B f(x)dx$ เมื่อ B เป็นช่วงใดๆบนจำนวนจริง

2.1.3 ค่าคาดหมายและความแปรปรวน **(Expectation value and variance)**

พารามิเตอร์ที่ใช้ในการอธิบายพฤติกรรมของตัวแปรสุ่มที่สำคัญและใช้กันอย่างแพร่หลายคือ ค่าคาดหมาย (Expectation value) หรือ ค่าเฉลี่ยทางสถิติ (Statistical mean) และค่าความแปรปรวน (Variance) โดยมีรายละเอียดดังนี้

1. ค่าคาดหมายของตัวแปรสุ่ม (Expectation value of random variable)

กำหนดให้ X เป็นตัวแปรสุ่ม ค่าคาดหมายหรือค่าเฉลี่ยทางสถิติใช้สัญลักษณ์แทนด้วย $E[X]$ สามารถหาได้ดังนี้

1) เมื่อ X เป็นตัวแปรสุ่มวิยุตแล้ว

$$E[X] = \sum_k x_k f(x_k) \tag{2.3a}$$

2) เมื่อ X เป็นตัวแปรสุ่มต่อเนื่องแล้ว

$$E[X] = \int_{-\infty}^{\infty} x f(x)dx \tag{2.3b}$$

2. ความแปรปรวนของตัวแปรสุ่ม (Variance of random variable)

เนื่องจากการทราบเพียงค่าคาดหมายของข้อมูลเพียงอย่างเดียวยังไม่สามารถอธิบายคุณลักษณะของของตัวแปรสุ่มได้ พารามิเตอร์ที่สำคัญอีกตัวที่จะช่วยในการบ่งบอกถึงการแกว่งตัวหรือการกระจายตัวของข้อมูลคือ ความแปรปรวน

กำหนดให้ X เป็นตัวแปรสุ่ม ความแปรปรวนของตัวแปรสุ่มใช้สัญลักษณ์แทนด้วย $V[X]$ สามารถหาได้ดังนี้

$$V[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \tag{2.4}$$

2.1.4 ฟังก์ชันการแจกแจง (Distribution function)

ในหัวข้อนี้นำเสนอตัวอย่างของการแจกแจงที่สำคัญของตัวแปรสุ่มทั้งแบบวิยุตและแบบต่อเนื่องดังนี้

1. ตัวอย่างฟังก์ชันการแจกแจงของตัวแปรสุ่มวิยุต

การแจกแจงแบบเบอร์นูลี (Bernoulli distribution)

การแจกแจงแบบเบอร์นูลลีเป็นการแจกแจงที่อธิบายถึงสถานการณ์ของการทดลองที่มีผลการทดลองเป็น 2 ประเภท คือ เกิดเหตุการณ์ที่สนใจหรือเกิดผลสำเร็จ (Success) และเกิดเหตุการณ์ที่ไม่สนใจหรือไม่เกิดผลสำเร็จ (Failure) ถ้าให้ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจแล้วได้รับความสำเร็จเท่ากับ p แล้วความน่าจะเป็นที่เกิดเหตุการณ์ที่สนใจแล้วไม่ได้รับความสำเร็จเท่ากับ $1-p$ ดังนั้นการแจกแจงแบบเบอร์นูลีของตัวแปรสุ่มวิยุต X คือ

$$P(X=x)=f(x;p)=p^x(1-p)^{1-x} \quad \text{เมื่อ} \quad x\in\{0,1\} \tag{2.5}$$

โดยที่ $E[X]=p$ และ $V[X]=p(1-p)$

การแจกแจงแบบเรขาคณิต (Geometric distribution)

กำหนดให้ X แทนจำนวนครั้งที่ต้องทำการทดลองแบบเบอร์นูลีซ้ำๆกัน โดยแต่ละครั้งเป็นอิสระต่อกัน จนกว่าจะได้ความสำเร็จเป็นครั้งแรก ให้ $p\in[0,1]$ เป็นความน่าจะเป็นที่ได้รับความสำเร็จแล้วการแจกแจงแบบเรขาคณิตของ X คือ

$$P(X=x)=f(x;p)=p(1-p)^{x-1} \quad \text{เมื่อ} \quad x\in\{1,2,\ldots\} \tag{2.6}$$

โดยที่ $E[X]=\dfrac{1}{p}$ และ $V[X]=\dfrac{1-p}{p^2}$

การแจกแจงแบบทวินาม (Binomial distribution)

ให้ $p\in[0,1]$ และ $n\in Z^+$ เป็นจำนวนครั้งของการทดลองแล้วการแจกแจงทวินามของตัวแปรสุ่มวิยุต X คือ

$$P(X=x)=f(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x} \text{ เมื่อ } x\in\{1,2,\ldots,n\} \tag{2.7}$$

โดยที่ $E[X]=np$ และ $V[X]=np(1-p)$

2. ตัวอย่างฟังก์ชันการแจกแจงของตัวแปรสุ่มต่อเนื่อง

การแจกแจงเอกรูป (Uniform distribution)

ให้ X เป็นตัวแปรสุ่มต่อเนื่องแล้วการแจกแจงเอกรูปของ X ในช่วง (a,b) คือ

$$f(x)=\frac{1}{b-a} \qquad ; \; a<x<b \tag{2.8}$$

โดยที่ $E[X]=\frac{b+a}{2}$ และ $\sigma_x^2=\frac{(b-a)^2}{12}$

การแจกแจงแบบแกมมา (Gamma distribution)

ในการหาการแจกแจงแบบแกมมาต้องอาศัยฟังก์ชันแกมมา (Gamma function) ซึ่งมีนิยามดังนี้

$$\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy \qquad ; \; \alpha>0 \tag{2.9}$$

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องแล้วการแจกแจงแบบแกมมาของ X ที่มีพารามิเตอร์ α และ β คือ

$$f(x;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}} \quad ; \; x>0;\alpha>0;\beta>0 \tag{2.10}$$

โดยที่ $E[X]=\alpha\beta$ และ $V[X]=\alpha\beta^2$

กราฟของฟังก์ชันจะขึ้นอยู่กับพารามิเตอร์คือ α และ B เมื่อ α คือพารามิเตอร์บ่งรูปร่าง (Shape parameter) และ β คือพารามิเตอร์บ่งขนาด (Scale parameter)

การแจกแจงแบบปกติ (Normal distribution)

การแจกแจงแบบปกติหรือการแจกแจงแบบเกาส์ (Gaussian distribution) เป็นการแจกแจงที่มีความสำคัญในทฤษฎีทางสถิติและวิธีทางสถิติในการวิเคราะห์ข้อมูล การแจกแจงแบบปกติมีลักษณะเป็นรูประฆังคว่ำ โดยมีพารามิเตอร์ที่ใช้วัดการกระจายตัวของข้อมูลอยู่ 2 ตัวคือ ค่าเฉลี่ยและความแปรปรวนโดยทั่วไปฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 นิยมแทนด้วย $\mathcal{N}(\mu,\sigma^2)$ โดยที่

$$\mathcal{N}(\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad ; \; \sigma^2>0 \tag{2.11}$$

โดยมี $E[X]=\mu$ และ $V[X]=\sigma^2$

2.1.5 ฟังก์ชันการแจกแจงร่วม (Joint distribution function)

ในหัวข้อที่ผ่านมาเป็นการศึกษาการแจกแจงของตัวแปรสุ่มเพียงตัวเดียว ในกรณีที่มีตัวแปรสุ่มมากกว่า 1 ตัวแปร นิยามต่างๆที่ได้ศึกษามาแล้ว จำเป็นต้องมีการปรับเปลี่ยนเพื่อให้สามารถใช้ได้กับกรณีของตัวแปรสุ่มหลายตัวแปรได้โดยมีรายละเอียดดังนี้

1. เวกเตอร์สุ่มแบบวิยุต (Discrete random vector)

กำหนดให้ $X=\{x_1,x_2,...\}$ และ $Y=\{y_1,y_2,...\}$ เป็นแบบเวกเตอร์สุ่ม $(X,Y)=\{(x_j,y_k),j,k=1,2,...\}$ ฟังก์ชันมวลของความน่าจะเป็นร่วม (Joint probability mass function : Joint PMF) ของ (X,Y) สามารถหาได้ดังสมการ (2.12)

$$P(X=x_j,Y=y_k)=f(x_j,y_k) \tag{2.12}$$

โดยที่ฟังก์ชันมวลความน่าจะเป็นตามขอบ (Marginal PMF) ของ X และ Y หาได้จาก

$$g(x_j)=\sum_{k=1}^{\infty} f(x_j,y_k), j=1,2,... \tag{2.13a}$$

$$h(y_k)=\sum_{j=1}^{\infty} f(x_j,y_k), k=1,2,... \tag{2.13b}$$

2. เวกเตอร์สุ่มแบบต่อเนื่อง (Continuous random vector)

กำหนดให้ X และ Y เป็นฟังก์ชันที่มีค่าอยู่ใน R โดยในการพิจารณา X ในช่วง (a,b) และ Y เป็นฟังก์ชันที่มีค่าอยู่ในช่วง (c,d) ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วม (Joint probability density function : Joint PDF) ของ (X,Y) สามารถหาได้ดังสมการ (2.14)

$$P(X,Y)=\int_c^d\int_a^b f(x,y)dxdy \tag{2.14}$$

โดยที่ฟังก์ชันความหนาแน่นความน่าจะเป็นตามขอบ (Marginal PDF) ของ X และ Y หาได้จาก

$$g(x)=\int_c^d f(x,y)dy \qquad , \ a<x<b \tag{2.15a}$$

$$h(y)=\int_a^b f(x,y)dx \qquad , \ c<y<d \tag{2.15b}$$

3. การกระจายตัวแบบมีเงื่อนไข (Conditional distributions)

กำหนดให้ X และ Y เป็นแบบเวกเตอร์สุ่ม $(X,Y)=\{(x_j,y_k),j,k=1,2,...\}$ โดยมีฟังก์ชันมวลความน่าจะเป็นร่วม $P(x_j,y_k)$ ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไขของ Y โดยทราบเงื่อนไข $X=x_j$ สามารถหาได้ดังสมการ (2.16)

$$P(y_k \mid x_j) = \frac{P(x_j, y_k)}{P(x_j)} \tag{2.16}$$

กำหนดให้ X และ Y เป็นฟังก์ชันที่มีค่าอยู่ใน R โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม $P(x,y)$ ฟังก์ชันความหนาแน่นความน่าจะเป็นแบบมีเงื่อนไขของ Y โดยทราบเงื่อนไข $X = x$ สามารถหาได้ดังสมการ (2.17)

$$P(y \mid x) = \frac{P(x,y)}{P(x)} \tag{2.17}$$

4. ค่าคาดหมายแบบมีเงื่อนไข (Conditional expectation)

กำหนดให้ X และ Y เป็นตัวแปรสุ่ม ค่าคาดหมายแบบมีเงื่อนไขของ X โดยทราบเงื่อนไข Y สามารถหาได้ดังนี้

1) ในกรณีที่ $X = x_j,\ \ j = 1,2,\ldots$ และ $Y = y_k,\ \ k = 1,2,\ldots$

$$E[X \mid Y = y_k] = \sum_{j=1}^{\infty} x_j P(x_j \mid Y = y_k) \tag{2.18}$$

2) ในกรณีที่ X และ Y เป็นฟังก์ชันต่อเนื่องในช่วง $(-\infty,\infty)$

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} xP(x \mid y)dx \tag{2.19}$$

5. ความแปรปรวนแบบมีเงื่อนไข (Conditional variance)

กำหนดให้ X และ Y เป็นตัวแปรสุ่ม ความแปรปรวนแบบมีเงื่อนไขของ X โดยทราบเงื่อนไข Y สามารถหาได้ดังนี้

$$V[X \mid Y] = E[(X - E[X \mid Y])^2 \mid Y] = E[X^2 \mid Y] - \left(E[X \mid Y]\right) \tag{2.20}$$

6. ความแปรปรวนร่วมเกี่ยว (Covariance)

กำหนดให้ X และ Y เป็นตัวแปรสุ่ม ความแปรปรวนร่วมเกี่ยวระหว่าง X และ Y สามารถหาได้ดังนี้

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] \tag{2.21}$$

7. สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

กำหนดให้ X และ Y เป็นตัวแปรสุ่ม สัมประสิทธิ์สหสัมพันธ์ ระหว่าง X และ Y สามารถหาได้ดังนี้

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \tag{2.22}$$

2.2 ทฤษฎีเมทริกซ์

ในหัวข้อนี้ขอนำเสนอคุณสมบัติและการดำเนินการทางเมทริกซ์ที่เกี่ยวข้องกับโครงงาน อาทิเช่น รอยเมทริกซ์ (Trace matrix) รูปแบบกำลังสองของเมทริกซ์ (Quadratic form of matrix) และอนุพันธ์ของเมทริกซ์ (Derivative of matrix)

โดยในที่นี้ ให้ $A=\left[a_{ij}\right]_{\substack{1\le i\le m\\1\le j\le n}}\in \mathrm{R}^{m\times n}$ เป็นเมทริกซ์ที่มี m แถวและ n หลัก ถ้า $\boldsymbol{m=n}$ จะเรียก A ว่าเป็นเมทริกซ์จัตุรัส

2.2.1 รอยเมทริกซ์ (Trace matrix)

<u>**นิยาม 2.8**</u> รอยเมทริกซ์หรือผลบวกแนวเฉียงของเมทริกซ์จัตุรัส $A\in \mathrm{R}^{n\times n}$ คือผลรวมของสมาชิกที่อยู่บนเส้นทแยงมุมหลักของเมทริกซ์ A ซึ่งสามารถหาได้ดังนี้

$$tr(A)=a_{11}+a_{22}+\ldots+a_{nn}=\sum_{i=1}^{n}a_{ii} \tag{2.23}$$

คุณสมบัติของรอยเมทริกซ์ (Properties of trace matrix)

กำหนดให้ $A,B\in \mathrm{R}^{n\times n}$ และ c เป็นปริมาณสเกลาร์แล้วต่อไปนี้เป็นคุณสมบัติของรอยเมทริกซ์

1. $tr(A+B)=tr(A)+tr(B)$
2. $tr(cA)=c\cdot tr(A)$
3. $tr(A)=tr(A^T)$
4. $tr(AB)=tr(BA)$

2.2.2 รูปกำลังสองของเมทริกซ์ (Quadratic form of matrix)

<u>**นิยาม 2.9**</u> ฟังก์ชันที่มีการส่งจาก $\mathrm{R}^n \to \mathrm{R}$ ที่อยู่ในรูปของ $f(x)=x^TAx=\sum_{i,j=1}^{n}a_{ij}x_ix_j$ จะเรียกว่ารูปกำลังสอง

โดยทั่วไป หากมีการพิจารณารูปกำลังสองของเมทริกซ์ A นิยมนิยามตั้งสมมติฐานว่า A เป็นเมทริกซ์สมมาตร เนื่องจาก

$$\begin{aligned} x^T Ax &= \frac{1}{2}x^T Ax + \frac{1}{2}x^T Ax \\ &= \frac{1}{2}x^T Ax + \frac{1}{2}\left[x^T A^T x\right]^T \quad \text{(คุณสมบัติเมทริกซ์สลับเปลี่ยน)} \\ &= \frac{1}{2}x^T Ax + \frac{1}{2}x^T A^T x \quad \text{(คุณสมบัติสเกลาร์)} \\ &= x^T\left(\frac{A+A^T}{2}\right)x \\ &= x^T Px \end{aligned} \tag{2.24}$$

โดยที่ $P = \dfrac{A+A^T}{2}$ เรียกว่า ส่วนสมมาตรของ A (Symmetric part of A)

<u>ตัวอย่าง 2.1</u> รูปกำลังสองกับส่วนสมมาตรของเมทริกซ์

กำหนดให้ $A = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$ รูปกำลังสองของ A คือ $x^T Ax = 2x_1^2 + 3x_1x_2 + 3x_2^2$ หากจัดรูปให้อยู่ในรูปของเมทริกซ์สมมาตร จะได้ว่า $P = \dfrac{A+A^T}{2} = \begin{bmatrix} 2 & 3/2 \\ 3/2 & 3 \end{bmatrix}$ ซึ่งมีรูปกำลังสองเท่ากับรูปกำลังสองของ A นั่นคือ $x^T Px = 2x_1^2 + 3x_1x_2 + 3x_2^2$

2.2.3 อนุพันธ์ของเมทริกซ์ (Derivative of matrix)

ในหัวข้อนี้ขอนำเสนอการหาอนุพันธ์ของเมทริกซ์และตัวดำเนินการทางเมทริกซ์ในแบบต่างๆ

1. อนุพันธ์ของฟังก์ชันเวกเตอร์ (Derivative of vector function)

<u>นิยาม 2.10</u> กำหนดให้ $A \in \mathrm{R}^m$ และ $B \in \mathrm{R}^n$ โดยที่ $a = f(b)$ อนุพันธ์ของเวกเตอร์ A เทียบกับเวกเตอร์ B คือ

$$\frac{\partial A}{\partial B} = \begin{bmatrix} \frac{\partial a_1}{\partial b_1} & \frac{\partial a_2}{\partial b_1} & \cdots & \frac{\partial a_m}{\partial b_1} \\ \frac{\partial a_1}{\partial b_2} & \frac{\partial a_2}{\partial b_2} & \cdots & \frac{\partial a_m}{\partial b_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_1}{\partial b_n} & \frac{\partial a_2}{\partial b_n} & \cdots & \frac{\partial a_m}{\partial b_n} \end{bmatrix} \tag{2.25}$$

<u>นิยาม 2.11</u> กำหนดให้ $B \in \mathrm{R}^{m \times n}$ และ a เป็นปริมาณสเกลาร์ อนุพันธ์ของ a เทียบกับเมทริกซ์ B คือ

$$\frac{\partial a}{\partial B} = \begin{bmatrix} \frac{\partial a}{\partial b_{11}} & \frac{\partial a}{\partial b_{12}} & \cdots & \frac{\partial a}{\partial b_{1n}} \\ \frac{\partial a}{\partial b_{21}} & \frac{\partial a}{\partial b_{22}} & \cdots & \frac{\partial a}{\partial b_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a}{\partial b_{m1}} & \frac{\partial a}{\partial b_{m2}} & \cdots & \frac{\partial a}{\partial b_{mn}} \end{bmatrix} \tag{2.26}$$

<u>นิยาม 2.12</u> กำหนดให้ $A \in \mathrm{R}^{m \times n}$ และ b เป็นปริมาณสเกลาร์ อนุพันธ์ของเมทริกซ์ A เทียบกับ b คือ

$$\frac{\partial A}{\partial b} = \begin{bmatrix} \frac{\partial a_{11}}{\partial b} & \frac{\partial a_{12}}{\partial b} & \cdots & \frac{\partial a_{1n}}{\partial b} \\ \frac{\partial a_{21}}{\partial b} & \frac{\partial a_{22}}{\partial b} & \cdots & \frac{\partial a_{2n}}{\partial b} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{m1}}{\partial b} & \frac{\partial a_{m2}}{\partial b} & \cdots & \frac{\partial a_{mn}}{\partial b} \end{bmatrix} \tag{2.27}$$

2. อนุพันธ์ของดีเทอร์มิแนนต์ (Derivative of determinant)

<u>นิยาม 2.13</u> ให้ $A \in \mathrm{R}^{n \times n}$ และ x เป็นปริมาณสเกลาร์ อนุพันธ์ของดีเทอร์มิแนนต์ A เทียบกับ x คือ

$$\frac{\partial \det(A)}{\partial x} = \det(A) tr\left(A^{-1} \frac{\partial A}{\partial x} \right) \tag{2.28}$$

ให้ $A, B, X \in \mathrm{R}^{n \times n}$ แล้วอนุพันธ์ดีเทอร์มิแนนต์ของเมทริกซ์เทียบกับ X คือ

$$\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^T \tag{2.29}$$

$$\frac{\partial \det(AXB)}{\partial X} = \det(AXB)(X^{-1})^T = \det(AXB)(X^T)^{-1} \tag{2.30}$$

3. อนุพันธ์ของเมทริกซ์ผกผัน (Derivative of inverse matrix)

<u>นิยาม 2.14</u> ให้ $A \in \mathrm{R}^{n \times n}$ และ x เป็นปริมาณสเกลาร์ อนุพันธ์ของเมทริกซ์ผกผัน A เทียบกับ x คือ

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1} \tag{2.31}$$

ให้ $A, B, X \in \mathrm{R}^{n\times n}$ แล้วอนุพันธ์เมทริกซ์ผกผันของเมทริกซ์เทียบกับ X คือ

$$\frac{\partial tr(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^T \tag{2.32}$$

$$\frac{\partial tr((X+A)^{-1})}{\partial X} = -((X+A)^{-1}(X+A)^{-1})^T \tag{2.33}$$

4. อนุพันธ์ของรอยเมทริกซ์ (Derivative of trace matrix)

<u>นิยาม 2.15</u> กำหนดให้ $A, B, X \in \mathrm{R}^{n\times n}$ อนุพันธ์ของรอยเมทริกซ์ A เทียบกับเมทริกซ์ X มีดังนี้

$$\frac{\partial tr(A)}{\partial X} = tr\left(\frac{\partial A}{\partial X}\right) \tag{2.34}$$

$$\frac{\partial tr(XA)}{\partial X} = A^T \tag{2.35}$$

$$\frac{\partial tr(AXB)}{\partial X} = A^T B^T \tag{2.36}$$

$$\frac{\partial tr(A^T XB^T)}{\partial X} = \frac{\partial tr(BX^T A)}{\partial X} = AB \tag{2.37}$$

$$\frac{\partial tr(XAX^T)}{\partial X} = \frac{\partial tr(AX^T X)}{\partial X} = \frac{\partial tr(X^T XA)}{\partial X} = X(A+A^T) \tag{2.38}$$

2.3 แบบจำลองปริภูมิสถานะ

ระบบพลวัตส่วนใหญ่มักถูกอธิบายด้วยสมการเชิงอนุพันธ์อันดับใดๆ ต่อมาได้มีแนวคิดในการแปลงสมการเชิงอนุพันธ์อันดับใดๆ ให้เป็นสมการเชิงอนุพันธ์อันดับหนึ่ง โดยการกำหนดตัวแปรขึ้นมาเรียกว่าตัวแปรสถานะ (State variables) เขียนแทนด้วย $x \in \mathrm{R}^n$ ซึ่งตัวแปรสถานะแต่ละตัวทำหน้าที่เป็นฐานหลักในปริภูมิสถานะ (State space) ทั้งนี้นิยมแทนสัญญาณควบคุมหรือสัญญาณขาเข้าด้วย $u \in \mathrm{R}^m$ และแทนผลตอบสนองของระบบหรือสัญญาณขาออกด้วย $y \in \mathrm{R}^p$

2.3.1 แบบจำลองปริภูมิสถานะเวลาต่อเนื่อง (Continuous time state space model)

แบบจำลองปริภูมิสถานะเวลาต่อเนื่องสามารถเขียนในรูปทั่วไปได้ดังนี้

$$\Sigma_c := \begin{cases} \dot{x}(t) = f(x,u;t) \\ y(t) = g(x,u;t) \end{cases} \quad ; \ t \in \mathrm{R}^+ \tag{2.39}$$

ในกรณีของระบบเชิงเส้นที่พารามิเตอร์ของระบบไม่แปรเปลี่ยนตามเวลา ระบบใน (2.39) สามารถเขียนใหม่ได้ดังนี้

$$\Sigma_c := \begin{cases} \dot{x}(t) = A_c x(t) + B_c u(t) \\ y(t) = C_c x(t) + D_c u(t) \end{cases} \tag{2.40}$$

โดยที่ $A_c \in \mathrm{R}^{n\times n}$ เป็นเมทริกซ์ระบบ (System matrix)

$B_c \in \mathrm{R}^{n\times m}$ เป็นเมทริกซ์สัญญาณขาเข้า (Input matrix)

$C_c \in \mathrm{R}^{p\times n}$ เป็นเมทริกซ์สัญญาณขาออก (Output matrix)

$D_c \in \mathrm{R}^{p\times m}$ เป็นเมทริกซ์ป้อนไปข้างหน้า (Feedforward matrix)

2.3.2 แบบจำลองปริภูมิสถานะเวลาวิยุต (Discrete time state space model)

แบบจำลองปริภูมิสถานะเวลาวิยุตสามารถเขียนในรูปทั่วไปดังนี้

$$\Sigma_d := \begin{cases} x(k+1) = f(x,u;k) \\ y(k) = g(x,u;k) \end{cases} \quad ; \; k \in Z^{+} \tag{2.41}$$

ในกรณีของระบบเชิงเส้นที่พารามิเตอร์ของระบบไม่แปรเปลี่ยนตามเวลา ระบบใน (2.41) สามารถเขียนใหม่ได้ดังนี้

$$\Sigma_d := \begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k) \end{cases} \tag{2.42}$$

2.3.3 ความสัมพันธ์ระหว่างแบบจำลองปริภูมิสถานะเวลาต่อเนื่องและเวลาวิยุต

โดยทั่วไประบบที่อธิบายด้วยแบบจำลองปริภูมิสถานะมักอยู่ในรูปแบบเวลาต่อเนื่อง แต่เนื่องจากการหาตัวประมาณจำเป็นต้องดำเนินการในรูปแบบเวลาวิยุต ดังนั้นในหัวข้อนี้ขอนำเสนอความสัมพันธ์ระหว่างแบบจำลองปริภูมิสถานะเวลาต่อเนื่องและเวลาวิยุต

พิจารณาแบบจำลองในรูปแบบเวลาต่อเนื่องทางเวลาตามสมการที่ (2.40) ซึ่งมีผลเฉลยที่เวลา $t = kT$ และ $t = (k+1)T$ ดังนี้

$$x((k+1)T) = e^{A_c(k+1)T}x(0) + e^{A_c(k+1)T}\int_0^{(k+1)T} e^{-A_c\tau}B_c u(\tau)d\tau \tag{2.43}$$

$$x(kT) = e^{A_c kT}x(0) + e^{A_c kT}\int_0^{kT} e^{-A_c\tau}B_c u(\tau)d\tau \tag{2.44}$$

คูณสมการ (2.44) ด้วย $e^{A_c T}$ จะได้

$$e^{A_c T}x(kT) = e^{A_c(k+1)T}x(0) + e^{A_c(k+1)T}\int_0^{kT} e^{-A_c\tau}B_c u(\tau)d\tau \tag{2.45}$$

จากสมการ (2.43) และ (2.45) จะได้ว่า

$$x((k+1)T) = e^{A_c T}x(kT) + e^{A_c(k+1)T}\int_{kT}^{(k+1)T} e^{-A_c\tau}B_c u(\tau)d\tau \tag{2.46}$$

เนื่องจากสัญญาณเข้าต้องถูกคงค่าไว้ให้คงที่ตลอดคาบของการชักตัวอย่างนั่นคือ $u(t) = u(kT)$ โดยที่ $\tau \in [kT,(k+1)T)$ สมการ (2.46) สามารถเขียนใหม่ได้ดังนี้

$$x((k+1)T) = e^{A_c T}x(kT) + e^{A_c(k+1)T}\left[\int_{kT}^{(k+1)T} e^{-A_c\tau}d\tau\right]B_c u(kT) \tag{2.47}$$

กำหนดให้ $\lambda = (k+1)T - \tau$ ดังนั้น $d\lambda = -d\tau$ โดยที่ λ มีค่าตั้งแต่ T ถึง 0 จะได้ดังนี้

$$x((k+1)T) = e^{A_c T}x(kT) + \left[\int_0^T e^{A_c\lambda}d\lambda\right]B_c u(kT) \qquad \text{เมื่อ } \lambda \in [0,T) \tag{2.48}$$

ให้ $A = e^{A_c T}$ และ $B = \left[\int_0^T e^{A_c\lambda}d\lambda\right]B_c = A_c^{-1}\left(e^{A_c T} - I\right)B_c$ ดังนั้นสมการ (2.48) จึงเขียนใหม่ได้เป็น

$$x((k+1)T) = Ax(kT) + Bu(kT) \tag{2.49a}$$

สำหรับสัญญาณออก $y(t)$ ซึ่งวัด ณ เวลาเดียวกับการชักตัวอย่างจะได้

$$y(kT) = Cx(kT) + Du(kT) \tag{2.49b}$$

โดยทั่วไปไม่นิยมเขียนคาบของการชักตัวอย่าง T เพื่อความกระชับของเนื้อหา ดังนั้นสมการ (2.49) จึงเป็นสมการที่ใช้บรรยายระบบเชิงเส้นที่พารามิเตอร์ของระบบไม่แปรเปลี่ยนตามเวลาดังสมการ (2.42)

2.4 วิธีการของคาลมาน

ในการประมาณพารามิเตอร์ของกระบวนการเชิงเฟ้นสุ่มที่บรรยายในรูปแบบของปริภูมิสถานะ ซึ่งมีตัวแปรสถานะ x_k ที่ไม่สามารถวัดหรือสังเกตออกมาจากระบบได้โดยตรง จึงต้องมีการสังเคราะห์ตัวแปรสถานะ $\hat{x}_k$ จากสัญญาณ y_k ที่วัดหรือสังเกตได้จากระบบผ่านทางวิธีการกรองและการปรับเรียบ

ในกรณีของกระบวนการเชิงเฟ้นสุ่มภายใต้การพิจารณาซึ่งมีรูปแบบเชิงเส้นและสัญญาณรบกวนมีฟังก์ชันการแจกแจงเป็นแบบปกติ วิธีการของคาลมานได้รับการยอมรับว่าเป็นวิธีที่สามารถลดค่าเฉลี่ยความผิดพลาดกำลังสองระหว่างตัวแปรสถานะจากการสังเคราะห์และตัวแปรสถานะจริงได้น้อยที่สุด สำหรับกระบวนการเชิงเฟ้นสุ่มไม่เชิงเส้นหรือมีฟังก์ชันการแจกแจงของสัญญาณรบกวนที่ไม่อยู่ในรูปแบบปกติ วิธีการของคาลมานถือว่าเป็นวิธีที่ดีที่สุดในบรรดาวิธีการเชิงเส้นประเภทอื่นๆ ในหัวข้อนี้จะกล่าวถึงหลักการทำงานของตัวกรองและตัวปรับเรียบคาลมานพอสังเขปดังนี้

2.4.1 ตัวกรองคาลมาน (Kalman filter)

ขั้นตอนการทำงานของตัวกรองคาลมานเป็นขั้นตอนแบบวนซ้ำ(Recursive) ซึ่งประกอบไปด้วย 2 ขั้นตอน คือ ขั้นตอนการทำนาย (Prediction) และขั้นตอนการกรอง (Filtering) โดยทั้ง 2 ขั้นตอนจะทำงานสลับกันจนครบตามจำนวนข้อมูลที่มี โดยกำหนดให้ $y_{1:k} = \{y_1,\ldots,y_k\}$ และ $x_{1:k} = \{x_1,\ldots,x_k\}$

พิจารณากระบวนการเฟ้นสุ่มเชิงเส้นดังนี้

$$x_{k+1} = Ax_k + Bu_k + w_k \qquad ; \; w_k \sim \mathcal{N}(0,Q) \tag{2.50}$$

$$y_k = Cx_k + Du_k + v_k \qquad ; \; v_k \sim \mathcal{N}(0,R) \tag{2.51}$$

1. ขั้นตอนการทำนาย (Prediction)

ในขั้นตอนนี้จะใช้ข้อมูลทั้งหมดตั้งแต่เริ่มต้นจนถึง ณ เวลา k มาใช้ทำนายค่าคาดหมายและเมทริกซ์ความแปรปรวนร่วมเกี่ยว ณ เวลา $k+1$ ดังนี้

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k \tag{2.52}$$

$$P_{k+1|k} = AP_{k|k}A^T + Q \tag{2.53}$$

โดยที่ $\hat{x}_{k+1|k} = E[x_{k+1} \mid y_{1:k}]$ คือค่าคาดหมายของตัวแปรสถานะ ณ เวลา $k+1$ เมื่อมีข้อมูลถึง ณ เวลา k

$P_{k+1|k} = E[(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T]$ คือเมทริกซ์ความแปรปรวนร่วมเกี่ยวของตัวแปรสถานะที่เกิดขึ้น ณ เวลา $k+1$ เมื่อมีข้อมูลถึง ณ เวลา k

2. ขั้นตอนการกรอง (Filtering)

ขั้นตอนนี้จะทำการปรับปรุงตัวแปรสถานะจากขั้นตอนการทำนาย โดยใช้สัญญาณ y_{k+1} จากการวัดหรือสังเกตจากระบบ ณ เวลา $k+1$ ซึ่งสามารถคำนวณได้ดังสมการ

$$K_{k+1} = P_{k+1|k} C^T (C P_{k+1|k} C^T + R)^{-1} \tag{2.54}$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k} - Du_k) \tag{2.55}$$

$$P_{k+1|k+1} = (I - CK_{k+1}) P_{k+1|k} \tag{2.56}$$

โดยที่ K_{k+1} คืออัตราขยายของคาลมาน ณ เวลา $k+1$

$\hat{x}_{k+1|k+1} = E[x_{k+1} \mid y_{1:k+1}]$ คือค่าคาดหมายของตัวแปรสถานะ ณ เวลา $k+1$ เมื่อมีข้อมูลถึง ณ เวลา $k+1$

$P_{k+1|k+1} = E[(x_{k+1} - \hat{x}_{k+1|k+1})(x_{k+1} - \hat{x}_{k+1|k+1})^T]$ คือเมทริกซ์ความแปรปรวนร่วมเกี่ยวของตัวแปรสถานะที่เกิดขึ้น ณ เวลา $k+1$ เมื่อมีข้อมูลถึง ณ เวลา $k+1$

2.4.2 ตัวปรับเรียบคาลมาน (Kalman smoother)

หากมีการกำหนดข้อมูลทั้งหมดมาให้ นั่นคือทราบ $y_1, y_2, \ldots, y_N$ การประมาณหาตัวแปรสถานะนิยมเลือกใช้ตัวปรับเรียบ ในกรณีของวิธีการของคาลมาน ขั้นตอนของตัวปรับเรียบต้องอาศัยตัวแปรสถานะที่สังเคราะห์มาจากตัวกรองคาลมาน นั่นคือ $\hat{x}_{k|k}, \hat{x}_{k+1|k}, P_{k|k}$ และ $P_{k+1|k}$ โดยตัวปรับเรียบคาลมานอาศัยกระบวนการทำซ้ำแบบย้อนกลับ โดยเริ่มจากที่ $k = N-1$ ลดลงมาจนถึง $k = 1$ โดยอาศัย 3 สมการต่อไปนี้

$$J_k = P_{k|k} A^T P_{k+1|k}^{-1} \tag{2.57}$$

$$\hat{x}_{k|N} = \hat{x}_{k|k} + J_k(\hat{x}_{k+1|N} - \hat{x}_{k+1|k}) \tag{2.58}$$

$$P_{k|N} = P_{k|k} + J_k(P_{k+1|N} - P_{k+1|k}) J_k^T \tag{2.59}$$

โดยที่ J_k คืออัตราขยายของตัวปรับเรียบคาลมาน ณ เวลา k

$\hat{x}_{k|N} = E[x_k \mid y_{1:N}]$ คือค่าคาดหมายของตัวแปรสถานะ ณ เวลา k เมื่อมีข้อมูลทั้งหมด N ตัว

$P_{k|N} = E[(x_k - \hat{x}_{k|N})(x_k - \hat{x}_{k|N})^T]$ คือเมทริกซ์ความแปรปรวนร่วมเกี่ยวของตัวแปรสถานะที่เกิดขึ้น ณ เวลา k เมื่อมีข้อมูลทั้งหมด N ตัว

เนื่องจากการประยุกต์ใช้วิธีการของคาลมานร่วมกับขั้นตอนวิธี EM จำเป็นต้องคำนวณหาเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่เกิดขึ้น ณ เวลา k และ $k-1$ เมื่อมีข้อมูลทั้งหมด N ตัว โดยเป็นกระบวนการทำซ้ำแบบย้อนกลับเช่นเดียวกับตัวปรับเรียบคาลมานโดยเริ่มจากที่ $k = N-1$ ลดลงมาจนถึง $k = 2$ โดยอาศัยสมการต่อไปนี้

$$P_{N,N-1|N} = (I - K_N C) A P_{N-1|N-1} \tag{2.60}$$

$$P_{k,k-1|N} = P_{k|k} J_{k-1}^T + J_k (P_{k+1,k|N} - A P_{k|k}) J_{k-1}^T \tag{2.61}$$

โดยที่ $P_{k,k-1|N} = E[(x_k - \hat{x}_{k|N})(x_{k-1} - \hat{x}_{k-1|N})^T$ คือเมทริกซ์ความแปรปรวนร่วมเกี่ยวของตัวแปรสถานะที่เกิดขึ้น ณ เวลา k และ $k+1$ เมื่อมีข้อมูลทั้งหมด N ตัว

2.5 แบบจำลองมาร์คอฟ

ในการศึกษากระบวนการเฟ้นสุ่มแบบจำลองมาร์คอฟเป็นพื้นฐานสำคัญสำหรับการประยุกต์ใช้ในการประมาณพารามิเตอร์ในรูปแบบต่างๆ ในหัวข้อนี้ขอนำเสนอองค์ความรู้ที่เกี่ยวข้องกับขั้นตอนวิธี EM ของแบบจำลองมาร์คอฟ

2.5.1 แบบจำลองมาร์คอฟ (Markov model)

พิจารณากระบวนการเชิงเฟ้นสุ่มตามสมการ (2.50) และ (2.51) ซึ่งสามารถบรรยายพฤติกรรมด้วยฟังก์ชันการแจกแจงได้ดังนี้

$$x_k \sim P(x_k \mid x_{k-1}) \tag{2.62}$$

$$y_k \sim P(y_k \mid x_k) \tag{2.63}$$

โดยที่ $P(x_k \mid x_{k-1})$ คือแบบจำลองความน่าจะเป็นของตัวแปรสถานะซึ่งเป็นฟังก์ชันการแจกแจง แบบมีเงื่อนไขของตัวแปรสถานะ ณ เวลา k เมื่อทราบตัวแปรสถานะ ณ เวลา $k-1$

$P(y_k \mid x_k)$ คือแบบจำลองความน่าจะเป็นของสัญญาณออกซึ่งเป็นฟังก์ชันการแจกแจงแบบมีเงื่อนไขของสัญญาณออก ณ เวลา k เมื่อทราบตัวแปรสถานะ ณ เวลา k

คุณสมบัติของแบบจำลองมาร์คอฟซ่อน

1. คุณสมบัติมาร์คอฟสำหรับตัวแปรสถานะ (Markov property of state)

เซตของตัวแปรสถานะ $\{x_k\}$ เมื่อ $k = 1,2,\ldots$ เป็นห่วงโซ่มาร์คอฟ (Markov chain) ก็ต่อเมื่อข้อมูลที่จำเป็นต่อการคำนวณตัวแปรสถานะ ณ เวลา k ถูกเก็บไว้ในตัวแปรสถานะ ณ เวลา $k-1$ ด้วยเหตุนี้จึงไม่มีความจำเป็นที่ต้องเก็บข้อมูลของกระบวนการในอดีต ซึ่งสามารถเขียนในทางคณิตศาสตร์ได้ดังนี้

$$P(x_k \mid x_{1:k-1}, y_{1:k-1}) = P(x_k \mid x_{k-1}) \quad \text{เมื่อ } k = 2,3,\ldots,N \quad (2.64)$$

เช่นเดียวกันกับกระบวนการคำนวณแบบย้อนกลับ ข้อมูลที่จำเป็นต่อการคำนวณตัวแปรสถานะ ณ เวลา $k-1$ ถูกเก็บไว้ในตัวแปรสถานะ ณ เวลา k ด้วยเหตุนี้จึงไม่มีความจำเป็นที่ต้องเก็บข้อมูลของกระบวนการในอดีต ซึ่งสามารถเขียนในทางคณิตศาสตร์ได้ดังนี้

$$P(x_{k-1} \mid x_{k:N}, y_{k:N}) = P(x_{k-1} \mid x_k) \quad \text{เมื่อ } k = N, N-1,\ldots,2 \quad (2.65)$$

2. ความเป็นอิสระแบบมีเงื่อนไขของสัญญาณออก (Conditional independence of output)

ตัวแปรสถานะและสัญญาณออกในอดีตไม่ส่งผลต่อสัญญาณออก ณ เวลาปัจจุบันมีแต่เฉพาะตัวแปรสถานะ x_k เท่านั้นที่ส่งผลต่อ y_k ซึ่งสามารถเขียนในทางคณิตศาสตร์ได้ดังนี้

$$P(y_k \mid x_{1:k-1}, y_{1:k-1}) = P(y_k \mid x_k) \quad \text{เมื่อ } k = 2,3,\ldots,N \quad (2.66)$$

2.6 ขั้นตอนวิธี EM

ขั้นตอนวิธี EM (Expectation-Maximization algorithm) เป็นทางเลือกหนึ่งในการหาความควรจะเป็นสูงสุด (Maximum likelihood : ML) เพื่อใช้ในการประมาณพารามิเตอร์ หลักการทำงานของขั้นตอนวิธี EM ประกอบไปด้วย 2 ขั้นตอน กล่าวคือขั้นตอน E (Expectation step) และ ขั้นตอน M (Maximization step) ทั้งสองขั้นตอนจะทำงานสลับกันเพื่อทำให้ความควรจะเป็นมีค่าสูงสุด

หลักการพื้นฐานของขั้นตอนวิธี EM คือการคำนวณหาขอบเขตล่าง (Lower bound) ของฟังก์ชันควรจะเป็น $L(\theta)$ แทนการคำนวณ $L(\theta)$ โดยตรง จากนั้นพยายามทำให้ขอบเขตล่างของ $L(\theta)$ มีค่าสูงขึ้นในแต่ละรอบของการวนซ้ำจึงส่งผลให้ $L(\theta)$ มีค่ามากขึ้นด้วยเช่นกัน โดยทั่วไปค่าสูงสุดที่ได้จากขั้นตอนวิธี EM จะเป็นค่าสูงสุดเฉพาะที่ (Local maxima) สำหรับวิธีการคำนวณหาขอบเขตล่าง $L(\theta)$ มีดังนี้

กำหนดให้ θ แทนพารามิเตอร์ใดๆ และให้ θ_n แทนพารามิเตอร์ที่ประมาณได้ในแต่ละรอบของการทำซ้ำและพิจารณาผลต่างของฟังก์ชันควรจะเป็นดังนี้

$$
\begin{aligned}
L(\theta)-L(\theta_n) &= \ln P(y_{1:N}\mid\theta)-\ln P(y_{1:N}\mid\theta_n)\\
&= \ln\left[\int P(y_{1:N}\mid x_{1:N},\theta)P(x_{1:N}\mid\theta)dx_{1:N}\right]-\ln P(y_{1:N}\mid\theta_n)\\
&= \ln\left[\int \frac{P(y_{1:N}\mid x_{1:N},\theta)P(x_{1:N}\mid\theta)}{P(y_{1:N}\mid\theta_n)}dx_{1:N}\right]\cdot\left[\frac{P(x_{1:N}\mid y_{1:N},\theta_n)}{P(x_{1:N}\mid y_{1:N},\theta_n)}\right]\\
&= \ln\left[\int P(x_{1:N}\mid y_{1:N},\theta_n)\cdot\frac{P(y_{1:N}\mid x_{1:N},\theta)P(x_{1:N}\mid\theta)}{P(y_{1:N}\mid\theta_n)P(x_{1:N}\mid y_{1:N},\theta_n)}dx_{1:N}\right] \qquad (2.67)
\end{aligned}
$$

เนื่องจากฟังก์ชันลอการิทึมเป็นฟังก์ชันเว้าจึงสามารถประยุกต์ใช้ทฤษฎีบทต่อไปนี้ได้

<u>ทฤษฎีบทที่ 2.1</u> (อสมการของเจนเสน (Jensen's inequality)) ให้ x เป็นตัวแปรสุ่มใดๆ และ $f(x)$ เป็นฟังก์ชันเว้าแล้ว

$$f[E(x)]\geq E[f(x)] \qquad (2.68)$$

เมื่อประยุกต์สมการของเจนเสนกับสมการ (2.67) จะได้ว่า

$$
\begin{aligned}
L(\theta)-L(\theta_n) &\geq \int P(x_{1:N}\mid y_{1:N},\theta_n)\ln\left(\frac{P(y_{1:N}\mid x_{1:N},\theta)P(x_{1:N}\mid\theta)}{P(y_{1:N}\mid\theta_n)P(x_{1:N}\mid y_{1:N},\theta_n)}\right)dx_{1:N}\\
&= \int P(x_{1:N}\mid y_{1:N},\theta_n)\ln\left(\frac{P(x_{1:N}\mid y_{1:N},\theta)}{P(x_{1:N}\mid y_{1:N},\theta_n)}\right)dx_{1:N}\\
&:= \Delta(\theta\mid\theta_n) \qquad (2.69)
\end{aligned}
$$

นั่นคือ

$$L(\theta)\geq L(\theta_n)+\Delta(\theta\mid\theta_n) \qquad (2.70)$$

เมื่อแทน θ ด้วย θ_n ลงใน $\Delta(\theta\mid\theta_n)$ จะพบว่า

$$\Delta(\theta_n\mid\theta_n)=\int P(x_{1:N}\mid y_{1:N},\theta_n)\ln\left(\frac{P(x_{1:N}\mid y_{1:N},\theta_n)}{P(x_{1:N}\mid y_{1:N},\theta_n)}\right)dx_{1:N}=0$$

ดังนั้น $L(\theta_n)+\Delta(\theta\mid\theta_n)$ จึงเป็นขอบเขตล่างของ $L(\theta)$ และจะมีค่าเท่ากับ $L(\theta)$ เมื่อพารามิเตอร์ θ มีค่าเท่ากับ θ_n ด้วยเหตุนี้หากสามารถหา θ_n ทำให้ $\Delta(\theta\mid\theta_n)$ มีค่ามากขึ้นในแต่ละรอบของการวนซ้ำ ซึ่งค่า $L(\theta)$ ก็จะมีค่ามากขึ้นเช่นกัน โดย θ_{n+1} สามารถหาได้ดังนี้

$$
\begin{aligned}
\theta_{n+1} &= \arg\max_{\theta}\{\Delta(\theta\mid\theta_n)\}\\
&= \arg\max_{\theta}\left\{\int P(x_{1:N}\mid y_{1:N},\theta_n)\ln\left(\frac{P(x_{1:N}\mid y_{1:N},\theta)}{P(x_{1:N}\mid y_{1:N},\theta_n)}\right)dx_{1:N}\right\} \qquad (2.71)
\end{aligned}
$$

เนื่องจาก $P(x_{1:N}, y_{1:N} \mid \theta_n)$ ไม่เป็นฟังก์ชันของ θ จึงไม่มีผลต่อการคำนวณหา θ ดังนั้นสมการ (2.71) สามารถลดรูปได้ดังนี้

$$\begin{aligned}\theta_{n+1} &= \arg\max_{\theta}\left\{\int P(x_{1:N} \mid y_{1:N}, \theta_n)\ln\left(P(x_{1:N} \mid y_{1:N}, \theta)\right)dx_{1:N}\right\} \\ &= \arg\max_{\theta}\left\{E_{x_{1:N}\mid y_{1:N},\theta_n}\left[\ln P(x_{1:N}, y_{1:N} \mid \theta)\right]\right\} \\ &= \arg\max_{\theta}\left\{Q(\theta)\right\} \qquad (2.72)\end{aligned}$$

โดยกำหนดให้ $Q(\theta)$ มีค่าเท่ากับ $E_{x_{1:N}\mid y_{1:N},\theta_n}\left[\ln P(x_{1:N}, y_{1:N} \mid \theta)\right]$ และสำหรับเนื้อหาที่เหลือจะใช้สัญลักษณ์ $E[\cdot]$ แทน $E_{x_{1:N}\mid y_{1:N},\theta_n}[\cdot]$ เพื่อความกระชับของเนื้อหา

พิจารณา $Q(\theta)$ จากสมการ (2.72) ดังนี้

$$Q(\theta) = E\left[\ln\left(P(x_1 \mid \theta)\prod_{k=2}^{N}P(x_k \mid x_{1:k-1}, \theta)\prod_{k=1}^{N}P(y_k \mid x_{1:k}, \theta)\right)\right] \qquad (2.73)$$

ทั้งนี้จากคุณสมบัติของมาร์คอฟ ดังสมการ (2.64) - (2.66) สามารถลดรูปสมการ (2.73) ได้ดังนี้

$$\begin{aligned}Q(\theta) &= E\left[\ln\left(P(x_1 \mid \theta)\prod_{k=2}^{N}P(x_k \mid x_{k-1}, \theta)\prod_{k=1}^{N}P(y_k \mid x_k, \theta)\right)\right] \\ &= E\left[\ln P(x_1 \mid \theta) + \sum_{k=2}^{N}\left[\ln P(x_k \mid x_{k-1}, \theta)\right] + \sum_{k=1}^{N}\left[\ln P(y_k \mid x_k, \theta)\right]\right] \qquad (2.74)\end{aligned}$$

2.6.1 ขั้นตอน E (Expectation step)

สำหรับการประมาณพารามิเตอร์ของกระบวนการเชิงเฟ้นสุ่มที่ถูกบรรยายในรูปแบบปริภูมิสถานะ ในขั้นตอน E เป็นกระบวนการหาค่าของฟังก์ชันควรจะเป็น โดยใช้สัญญาณ y_k จากการวัดหรือสังเกตและตัวแปรสถานะ $\hat{x}_{k|N}$ จากตัวปรับเรียบคาลมาน ในหัวข้อนี้ขอนำเสนอการหาค่าของฟังก์ชันควรจะเป็นร่วมกับวิธีการของคาลมาน

เนื่องจากในโครงงานนี้พิจารณารูปแบบของปริภูมิสถานะแบบเชิงเส้นและมีฟังก์ชันการแจกแจงรูปแบบปกติ โดยพิจารณาในรูปแบบของเมทริกซ์ ฟังก์ชันการแจกแจกสามารถหาได้ดังนี้

$$f(x) = \frac{1}{(2\pi)^{1/2}\left|\sigma^2\right|^{1/2}}e^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)^T} \qquad (2.75)$$

จากสมมติฐานที่ว่า $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ ดังนั้น

$$\ln P(x_1 \mid \theta) = \ln\left(\frac{1}{(2\pi)^{1/2}\left|\sigma_1^2\right|^{1/2}} e^{-\frac{1}{2}(x_1-\mu_1)(\sigma_1^2)^{-1}(x_1-\mu_1)^T}\right)$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma_1^2\right| - \frac{1}{2}(x_1-\mu_1)(\sigma_1^2)^{-1}(x_1-\mu_1)^T \quad (2.76)$$

จากสมการ (2.50) พบว่า

$$E[x_k \mid x_{k-1}] = Ax_{k-1} + Bu_{k-1} \text{ และ } E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = Q$$

ดังนั้น

$$\ln P(x_k \mid x_{k-1}, \theta) = \ln\left(\frac{1}{(2\pi)^{1/2}|Q|^{1/2}} e^{-\frac{1}{2}(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T}\right)$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|Q| - \frac{1}{2}(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T \quad (2.77)$$

จากสมการ (2.51) พบว่า

$$E[y_k \mid x_k] = Cx_k + Du_k \text{ และ } E[(y_k - \hat{y}_k)(y_k - \hat{y}_k)^T] = R$$

ดังนั้น

$$\ln P(y_k \mid x_k, \theta) = \ln\left(\frac{1}{(2\pi)^{1/2}|R|^{1/2}} e^{-\frac{1}{2}(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T}\right)$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|R| - \frac{1}{2}(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T \quad (2.78)$$

แทนค่าสมการ (2.76) - (2.78) ลงในสมการ (2.74) จะได้ดังนี้

$$Q(\theta) = E\Bigg[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma_1^2\right| - \frac{1}{2}(x_1-\mu_1)(\sigma_1^2)^{-1}(x_1-\mu_1)^T$$
$$+ \sum_{k=2}^{N}\left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|Q| - \frac{1}{2}(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T\right)\Bigg]$$
$$+ \sum_{k=1}^{N}\left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln|R| - \frac{1}{2}(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T\right)$$

$$\left.\begin{aligned}&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma_1^2\right| - \frac{1}{2}E\left[(x_1-\mu_1)(\sigma_1^2)^{-1}(x_1-\mu_1)^T\right] - \frac{N}{2}\ln(2\pi)\\ &\quad -\frac{N-1}{2}\ln|Q| - \frac{1}{2}\sum_{k=2}^{N}E\left[(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T\right]\\ &\quad -\frac{N}{2}\ln|R| - \frac{1}{2}\sum_{k=1}^{N}E\left[(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T\right] - \frac{N-1}{2}\ln(2\pi)\end{aligned}\right\} \quad (2.79)$$

กำหนดให้ $\widetilde{x}_k = x_k - \hat{x}_k$ จะได้

$$x_k = \hat{x}_k + \widetilde{x}_k \quad (2.80)$$

$$E\left[\widetilde{x}_k \middle| y_k, \theta\right] = 0 \quad (2.81)$$

จากสมการ (2.79) พิจารณาที่ $E\left[(x_1-\mu_1)(\sigma_1^2)^{-1}((x_1-\mu_1)^T)\right]$ เมื่อแทนสมการ (2.80) และ (2.81) ลงไปจะได้ค่าคาดหมายดังนี้

$$\begin{aligned}E\left[(x_1-\mu_1)(\sigma_1^2)^{-1}((x_1-\mu_1)^T)\right] &= E\left[tr\left((\sigma_1^2)^{-1}(\hat{x}_1 + \widetilde{x}_k - \mu_1)((\hat{x}_1 + \widetilde{x}_k - \mu_1)^T\right)\right]\\ &= tr\left[(\sigma_1^2)^{-1}\left((\hat{x}_1 - \mu_1)((\hat{x}_1 - \mu_1)^T + \sigma_1^2\right)\right]\end{aligned} \quad (2.82)$$

จากสมการ (2.79) พิจารณาที่ $E\left[(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T\right]$ เมื่อแทนสมการ (2.80) และ (2.81) ลงไปจะได้ค่าคาดหมายดังนี้

$$\begin{aligned}&E\left[(x_k - Ax_{k-1} - Bu_{k-1})Q^{-1}(x_k - Ax_{k-1} - Bu_{k-1})^T\right]\\ &\quad = E\left[tr\left(Q^{-1}(\hat{x}_k + \widetilde{x}_k - A\hat{x}_{k-1} - A\widetilde{x}_{k-1} - Bu_{k-1})(\hat{x}_k + \widetilde{x}_k - A\hat{x}_{k-1} - A\widetilde{x}_{k-1} - Bu_{k-1})^T\right)\right]\\ &\quad = tr\left[Q^{-1}\left((\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})(\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})^T + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T\right)\right]\end{aligned} \quad (2.83)$$

จากสมการ (2.79) พิจารณาที่ $E\left[(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T\right]$ เมื่อแทนสมการ (2.80) และ (2.81) ลงไปจะได้ค่าคาดหมายดังนี้

$$\begin{aligned}&E\left[(y_k - Cx_k - Du_k)R^{-1}(y_k - Cx_k - Du_k)^T\right]\\ &\quad = E\left[tr\left(R^{-1}(y_k - C\hat{x}_k - C\widetilde{x}_k - Du_k)(y_k - C\hat{x}_k - C\widetilde{x}_k - Du_k)^T\right)\right]\\ &\quad = tr\left[R^{-1}\left((y_k - C\hat{x}_k - Du_k)(y_k - C\hat{x}_k - Du_k)^T + CP_kC^T\right)\right]\end{aligned} \quad (2.84)$$

แทนสมการ (2.82) - (2.84) ลงในสมการ (2.79) จะได้

$$
\left.
\begin{aligned}
Q(\theta) = & -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma_1^2\right| - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln|R| - \frac{N-1}{2}\ln(2\pi) \\
& -\frac{N-1}{2}\ln|Q| - \frac{1}{2}tr\left[(\sigma_1^2)^{-1}\left((\hat{x}_1-\mu_1)((\hat{x}_1-\mu_1)^T+\sigma_1^2\right)\right] \\
& -\frac{1}{2}\sum_{k=2}^{N} tr\left[Q^{-1}\left((\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})(\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})^T\right.\right. \\
& \left.\left. + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T\right)\right] \\
& -\frac{1}{2}\sum_{k=1}^{N} tr\left[R^{-1}\left((y_k - C\hat{x}_k - Du_k)(y_k - C\hat{x}_k - Du_k)^T + CP_kC^T\right)\right]
\end{aligned}
\right\} \tag{2.85}
$$

2.6.2 ขั้นตอน M (Maximization Step)

หลังจากคำนวณ $Q(\theta)$ ในขั้นตอน E เป็นการคำนวณพารามิเตอร์ตัวใหม่โดยอาศัยขั้นตอน M เพื่อคำนวณพารามิเตอร์ตัวใหม่ที่ทำให้ขอบเขตล่างของ $L(\theta)$ ค่าเพิ่มสูงขึ้น โดยการหาอนุพันธ์ย่อยเทียบแต่ละพารามิเตอร์และกำหนดให้มีค่าเป็นศูนย์ ดังนั้นหัวข้อนี้ขอนำเสนอการหาอนุพันธ์ย่อยของ $Q(\theta)$ เทียบกับพารามิเตอร์ โดยเซตของพารามิเตอร์คือ $\theta = \{A,B,C,D,Q,R,\mu_1,\sigma_1^2\}$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ A

$$
\begin{aligned}
0 &= \frac{\partial}{\partial A}\left\{tr\left[Q^{-1}\left((\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})(\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})^T + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T\right)\right]\right\} \\
&= -\frac{1}{2}\sum_{k=2}^{N}\frac{\partial}{\partial A}\left\{tr\left[-Q^{-1}A(P_{k-1,k}+\hat{x}_{k-1}\hat{x}_k^T) - Q^{-1}(P_{k,k-1}+\hat{x}_k\hat{x}_{k-1}^T)A^T + Q^{-1}A(P_{k-1}+\hat{x}_{k-1}\hat{x}_{k-1}^T)A^T\right.\right. \\
&\quad \left.\left. + Q^{-1}Bu_{k-1}\hat{x}_{k-1}^TA^T + Q^{-1}A\hat{x}_{k-1}u_{k-1}^TB^T\right]\right\} \\
&= -\frac{1}{2}\sum_{k=2}^{N} tr\left[-Q^{-1}(P_{k-1,k}+\hat{x}_{k-1}\hat{x}_k^T) - Q^{-1}(P_{k,k-1}+\hat{x}_k\hat{x}_{k-1}^T) + 2Q^{-1}A(P_{k-1}+\hat{x}_{k-1}\hat{x}_{k-1}^T) + Q^{-1}Bu_{k-1}x_{k-1}^T\right. \\
&\quad \left. + Q^{-1}\hat{x}_{k-1}u_{k-1}^TB^T\right] \\
&= -\frac{1}{2}\sum_{k=2}^{N}\left(-2Q^{-1}(P_{k,k-1}+\hat{x}_k\hat{x}_{k-1}^T) + 2Q^{-1}A(P_{k-1}+\hat{x}_{k-1}\hat{x}_{k-1}^T) + 2Q^{-1}Bu_{k-1}\hat{x}_{k-1}^T\right) \\
&= Q^{-1}\sum_{k=2}^{N}\left(P_{k,k-1}+\hat{x}_k\hat{x}_{k-1}^T - A(P_{k-1}+\hat{x}_{k-1}\hat{x}_{k-1}) - Bu_{k-1}\hat{x}_{k-1}^T\right)
\end{aligned} \tag{2.86}
$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ B

$$0=\frac{\partial}{\partial B}\left\{-\frac{1}{2}\sum_{k=2}^{N}tr\left[Q^{-1}\left((\hat{x}_k-A\hat{x}_{k-1}-Bu_{k-1})(\hat{x}_k-A\hat{x}_{k-1}-Bu_{k-1})^T+P_k-AP_{k-1,k}-P_{k,k-1}A^T+AP_{k-1}A^T\right)\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=2}^{N}\frac{\partial}{\partial B}\left\{tr\left[-Q^{-1}Bu_{k-1}\hat{x}_k^T+Q^{-1}Bu_{k-1}\hat{x}_{k-1}^TA^T-Q^{-1}\hat{x}_ku_{k-1}^TB^T+Q^{-1}A\hat{x}_{k-1}u_{k-1}^TB^T\right.\right.$$
$$\left.\left.+Q^{-1}Bu_{k-1}u_{k-1}^TB^T\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=2}^{N}tr\left[-Q^{-1}u_{k-1}\hat{x}_k^T+Q^{-1}u_{k-1}\hat{x}_{k-1}^TA^T-Q^{-1}\hat{x}_ku_{k-1}^T+Q^{-1}A\hat{x}_{k-1}u_{k-1}^T+2Q^{-1}Bu_{k-1}u_{k-1}^T\right]$$

$$=-\frac{1}{2}\sum_{k=2}^{N}\left(-2Q^{-1}\hat{x}_ku_{k-1}^T+2Q^{-1}A\hat{x}_{k-1}u_{k-1}^T+2Q^{-1}Bu_{k-1}u_{k-1}^T\right)$$

$$=Q^{-1}\sum_{k=2}^{N}\left(\hat{x}_ku_{k-1}^T-A\hat{x}_{k-1}u_{k-1}^T-Bu_{k-1}u_{k-1}^T\right) \tag{2.87}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ C

$$0=\frac{\partial}{\partial C}\left\{-\frac{1}{2}\sum_{k=1}^{N}tr\left[R^{-1}\left((y_k-C\hat{x}_k-Du_k)(y_k-C\hat{x}_k-Du_k)^T+CP_kC^T\right)\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=1}^{N}\frac{\partial}{\partial C}\left\{tr\left[-R^{-1}C\hat{x}_ky_k^T-R^{-1}y_k\hat{x}_k^TC^T+R^{-1}C(P_k+\hat{x}_k\hat{x}_k^T)C^T+R^{-1}Du_k\hat{x}_k^TC^T+R^{-1}C\hat{x}_ku_k^TD^T\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=1}^{N}tr\left[-R^{-1}\hat{x}_ky_k^T-R^{-1}y_k\hat{x}_k^T+2R^{-1}C(P_k+\hat{x}_k\hat{x}_k^T)+R^{-1}Du_k\hat{x}_k^T+R^{-1}\hat{x}_ku_k^TD^T\right]$$

$$=-\frac{1}{2}\sum_{k=1}^{N}\left(-2R^{-1}y_k\hat{x}_k^T+2R^{-1}C(P_k+\hat{x}_k\hat{x}_k^T)+2R^{-1}Du_k\hat{x}_k^T\right)$$

$$=R^{-1}\sum_{k=1}^{N}\left(y_k\hat{x}_k^T-C(P_k+\hat{x}_k\hat{x}_k^T)-Du_k\hat{x}_k^T\right) \tag{2.88}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ D

$$0=\frac{\partial}{\partial D}\left\{-\frac{1}{2}\sum_{k=1}^{N}tr\left[R^{-1}\left((y_k-C\hat{x}_k-Du_k)(y_k-C\hat{x}_k-Du_k)^T+CP_kC^T\right)\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=1}^{N}\frac{\partial}{\partial D}\left\{tr\left[-R^{-1}Du_ky_k^T+R^{-1}Du_k\hat{x}_k^TC^T-R^{-1}y_ku_k^TD^T+R^{-1}C\hat{x}_ku_k^TD^T+R^{-1}Du_ku_k^TD^T\right]\right\}$$

$$=-\frac{1}{2}\sum_{k=1}^{N}tr\left[-R^{-1}u_ky_k^T+R^{-1}u_k\hat{x}_k^TC^T-R^{-1}y_ku_k^T+R^{-1}C\hat{x}_ku_k^T+2R^{-1}Du_ku_k^T\right]$$

$$=-\frac{1}{2}\sum_{k=1}^{N}\left(-2R^{-1}y_ku_k^T+2R^{-1}C\hat{x}_ku_k^T+2R^{-1}Du_ku_k^T\right)$$

$$= R^{-1}\sum_{k=1}^{N}\left(y_k u_k^T - C\hat{x}_k u_k^T - Du_k u_k^T\right) \tag{2.89}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ Q

$$0 = \frac{\partial}{\partial Q^{-1}}\left\{-\frac{1}{2}\sum_{k=2}^{N} tr\left[Q^{-1}\left((\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})(\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})^T + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T\right)\right]\right.$$
$$\left. - \frac{(N-1)}{2}\ln|Q|\right\}$$
$$= -\frac{1}{2}\sum_{k=2}^{N}\left((x_k - A\hat{x}_{k-1} - Bu_{k-1})(x_k - A\hat{x}_{k-1} - Bu_{k-1})^T + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T\right) + \frac{(N-1)}{2}Q \tag{2.90}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ R

$$0 = \frac{\partial}{\partial R^{-1}}\left\{-\frac{1}{2}\sum_{k=1}^{N} tr\left[R^{-1}\left((y_k - C\hat{x}_k - Du_k)(y_k - C\hat{x}_k - Du_k)^T + CP_kC^T\right)\right] - \frac{N}{2}\ln|R|\right\}$$
$$= -\frac{1}{2}\sum_{k=1}^{N}\left((y_k - C\hat{x}_k - Du_k)(y_k - C\hat{x}_k - Du_k)^T + CP_kC^T\right) + \frac{N}{2}R \tag{2.91}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ μ_1

$$0 = \frac{\partial}{\partial \mu_1}\left\{-\frac{1}{2}tr\left[(\sigma_1^2)^{-1}\left((\hat{x}_1 - \mu_1)((\hat{x}_1 - \mu_1)^T + \sigma_1^2\right)\right]\right\}$$
$$= -\frac{1}{2}\frac{\partial}{\partial \mu_1} tr\left[-(\sigma_1^2)^{-1}\hat{x}_1\mu_1^T - (\sigma_1^2)^{-1}\mu_1\hat{x}_1^T + (\sigma_1^2)^{-1}\mu_1\mu_1^T\right]$$
$$= -\frac{1}{2}tr\left[-(\sigma_1^2)^{-1}\hat{x}_1 - (\sigma_1^2)^{-1}\hat{x}_1^T + 2(\sigma_1^2)^{-1}\mu_1\right]$$
$$= -\frac{1}{2}\left(-2(\sigma_1^2)^{-1}\hat{x}_1 + 2(\sigma_1^2)^{-1}\mu_1\right)$$
$$= (\sigma_1^2)^{-1}\hat{x}_1 - (\sigma_1^2)^{-1}\mu_1 \tag{2.92}$$

พิจารณาอนุพันธ์ย่อยของ $Q(\theta)$ ในสมการ (2.85) เทียบกับพารามิเตอร์ σ_1^2

$$0 = \frac{\partial}{\partial(\sigma_1^2)^{-1}}\left\{-\frac{N}{2}\ln\left|\sigma_1^2\right| - \frac{1}{2}tr\left[(\sigma_1^2)^{-1}\left((\hat{x}_1 - \mu_1)((\hat{x}_1 - \mu_1)^T + \sigma_1^2\right)\right]\right\}$$
$$= \frac{N}{2}\sigma_1^2 - \frac{1}{2}\left((\hat{x}_1 - \mu_1)(\hat{x}_1 - \mu_1)^T + \sigma_1^2\right) \tag{2.93}$$

พิจารณาพารามิเตอร์ A และ B ในสมการ (2.86) และ (2.87) โดยสามารถนำมาเขียนในรูปแบบเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} \sum_{k=2}^{N} P_{k,k-1} + \hat{x}_k \hat{x}_{k-1}^T & \sum_{k=2}^{N} \hat{x}_k u_{k-1}^T \end{bmatrix} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} \sum_{k=2}^{N} P_{k-1} + \hat{x}_{k-1}\hat{x}_{k-1}^T & \sum_{k=2}^{N} \hat{x}_{k-1} u_{k-1}^T \\ \sum_{k=2}^{N} u_{k-1}\hat{x}_{k-1}^T & \sum_{k=2}^{N} u_{k-1} u_{k-1}^T \end{bmatrix}$$

$$\begin{bmatrix} A_{new} & B_{new} \end{bmatrix} = \begin{bmatrix} \sum_{k=2}^{N} P_{k,k-1} + \hat{x}_k \hat{x}_{k-1}^T & \sum_{k=2}^{N} \hat{x}_k u_{k-1}^T \end{bmatrix} \begin{bmatrix} \sum_{k=2}^{N} P_{k-1} + \hat{x}_{k-1}\hat{x}_{k-1}^T & \sum_{k=2}^{N} \hat{x}_{k-1} u_{k-1}^T \\ \sum_{k=2}^{N} u_{k-1}\hat{x}_{k-1}^T & \sum_{k=2}^{N} u_{k-1} u_{k-1}^T \end{bmatrix}^{-1} \quad (2.94)$$

พิจารณาพารามิเตอร์ C และ D ในสมการ (2.88) และ (2.89) โดยสามารถนำมาเขียนในรูปแบบเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} \sum_{k=1}^{N} y_k \hat{x}_k^T & \sum_{k=1}^{N} y_k u_k^T \end{bmatrix} = \begin{bmatrix} C & D \end{bmatrix} \begin{bmatrix} \sum_{k=1}^{N} P_k + \hat{x}_k \hat{x}_k^T & \sum_{k=1}^{N} \hat{x}_k u_k^T \\ \sum_{k=1}^{N} u_k \hat{x}_k^T & \sum_{k=1}^{N} u_k u_k^T \end{bmatrix}$$

$$\begin{bmatrix} C_{new} & D_{new} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{N} y_k \hat{x}_k^T & \sum_{k=1}^{N} y_k u_k^T \end{bmatrix} \begin{bmatrix} \sum_{k=1}^{N} P_k + \hat{x}_k \hat{x}_k^T & \sum_{k=1}^{N} \hat{x}_k u_k^T \\ \sum_{k=1}^{N} u_k \hat{x}_k^T & \sum_{k=1}^{N} u_k u_k^T \end{bmatrix}^{-1} \quad (2.95)$$

พิจารณาพารามิเตอร์ Q ในสมการ (2.90) ดังนี้

$$\frac{(N-1)}{2} Q = \frac{1}{2} \sum_{k=2}^{N} tr\left[Q^{-1} \left((\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})(\hat{x}_k - A\hat{x}_{k-1} - Bu_{k-1})^T + P_k - AP_{k-1,k} - P_{k,k-1}A^T + AP_{k-1}A^T \right) \right]$$

$$\begin{aligned} Q_{new} = \frac{1}{(N-1)} \sum_{k=2}^{N} \Big(& \hat{x}_k \hat{x}_k^T + P_k - A_{new}(\hat{x}_{k-1}\hat{x}_k^T + P_{k-1,k}) - B_{new} u_{k-1} \hat{x}_k^T - (\hat{x}_k \hat{x}_{k-1}^T + P_{k,k-1}) A_{new}^T - \hat{x}_k u_{k-1}^T B_{new}^T \\ & + A_{new}(\hat{x}_{k-1}\hat{x}_{k-1}^T + P_{k-1}) A_{new}^T + B_{new} u_{k-1} \hat{x}_{k-1}^T A_{new}^T + A_{new} \hat{x}_{k-1} u_{k-1}^T B_{new}^T + B_{new} u_{k-1} u_{k-1}^T B_{new}^T \Big) \end{aligned}$$

จากสมการ (2.86) และ (2.87) พารามิเตอร์ Q_{new} สามารถลดรูปได้ดังนี้

$$Q_{new} = \frac{1}{(N-1)} \sum_{k=2}^{N} \left(\hat{x}_k \hat{x}_k^T + P_k - A_{new}(\hat{x}_{k-1}\hat{x}_k^T + P_{k-1,k}) - B_{new} u_{k-1} \hat{x}_k^T \right) \quad (2.96)$$

พิจารณาพารามิเตอร์ R ในสมการ (2.91) ดังนี้

$$\frac{N}{2}R = \frac{1}{2}\sum_{k=1}^{N} tr\left[R^{-1}\left((y_k - C\hat{x}_k - Du_k)(y_k - C\hat{x}_k - Du_k)^T + CP_kC^T\right)\right]$$

$$R_{new} = \frac{1}{N}\sum_{k=1}^{N}\Big(y_k y_k^T - y_k\hat{x}_k^T C_{new}^T - y_k u_k^T D_{new}^T - C_{new}\hat{x}_k y_k^T + C_{new}\hat{x}_k\hat{x}_k^T C_{new}^T + C_{new}\hat{x}_k u_k^T D_{new}^T - D_{new}u_k y_k^T$$
$$+ D_{new}u_k\hat{x}_k^T C_{new}^T + D_{new}u_k u_k^T D_{new}^T + C_{new}P_k C_{new}^T\Big)$$

$$R_{new} = \frac{1}{N}\sum_{k=1}^{N}\Big(y_k y_k^T - C_{new}\hat{x}_k y_k^T - D_{new}u_k y_k^T - (y_k\hat{x}_k^T - C_{new}(P_k + \hat{x}_k\hat{x}_k^T C_{new}^T) - D_{new}u_k\hat{x}_k^T)C_{new}^T$$
$$- (y_k u_k^T - C_{new}\hat{x}_k u_k^T - D_{new}u_k u_k^T)D_{new}^T\Big)$$

พารามิเตอร์ R_{new} สามารถลดรูปได้โดยใช้สมการ (2.88) และ (2.89) ดังนี้

$$R_{new} = \frac{1}{N}\sum_{k=1}^{N}\left(y_k y_k^T - C_{new}\hat{x}_k y_k^T - D_{new}u_k y_k^T\right) \tag{2.97}$$

พิจารณาพารามิเตอร์ μ_1 ในสมการ (2.92) ดังนี้

$$\mu_{1,new} = \hat{x}_1 \tag{2.98}$$

พิจารณาพารามิเตอร์ σ_1^2 ในสมการ (2.93) ดังนี้

$$\sigma_{1,new}^2 = (\hat{x}_1 - \mu_1)(\hat{x}_1 - \mu_1)^T + \sigma_1^2 \tag{2.99}$$

จากการหาอนุพันธ์ย่อยของฟังก์ชันควรจะเป็นเทียบกับพารามิเตอร์ A, B, C, D, Q, R, μ_1 และ σ_1^2 โดยสามารถหาพารามิเตอร์ตัวใหม่ได้ตามสมการ (2.94) - (2.99)

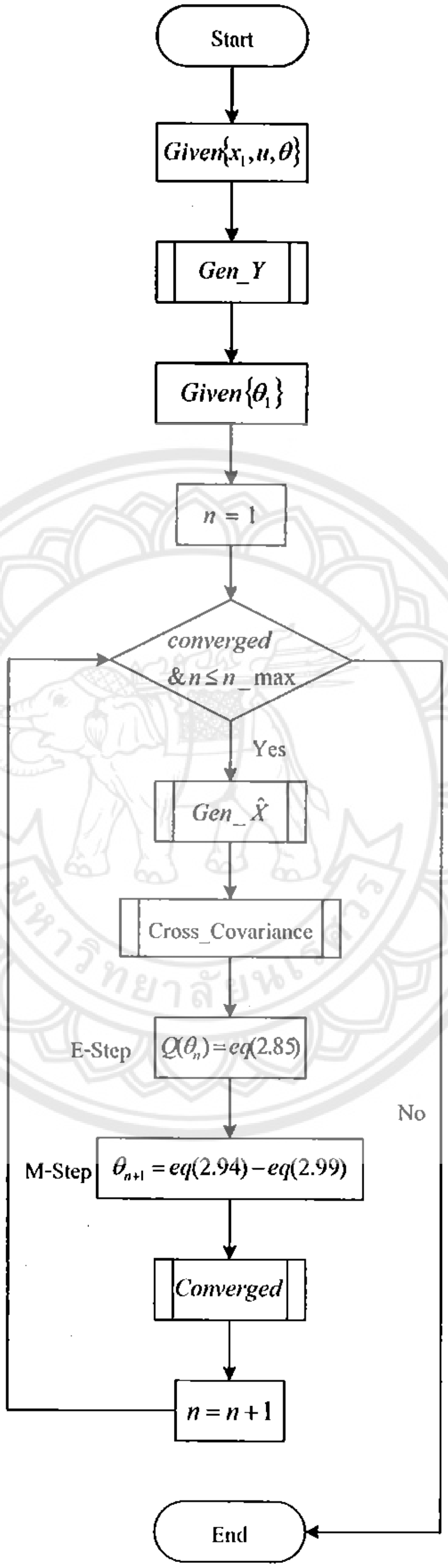
บทที่ 3
ขั้นตอนการดำเนินงาน

ในการทดลองโปรแกรม ให้สมมติว่าผู้ออกแบบทราบพารามิเตอร์ของระบบ θ จากนั้นจึงป้อนสัญญาณเข้า u_k และวัดสัญญาณออก y_k นำ $\{u_k, y_k\}$ ป้อนเข้าสู่ขั้นตอนวิธี EM เพื่อทำการสังเคราะห์พารามิเตอร์ $\hat{\theta}$

การทำงานของโปรแกรมจะวนซ้ำจนขอบเขตล่างของฟังก์ชันควรจะเป็นเพิ่มขึ้นได้เพียงเล็กน้อยหรือครบจำนวนรอบที่ต้องใช้จึงจบโปรแกรม รหัสเทียมและแผนภาพการทำงานของการประมาณพารามิเตอร์ด้วยขั้นตอนวิธี EM แสดงไว้ในรูปที่ 3.1 และ 3.2

```
Main()
    input ( x_1, u, θ )
    Gen_Y()
    input ( θ_1 )
    while not converged & n ≤ n_max
        Gen_X̂()
        Cross_Covariance()
        Q(θ_n) = calculate according to eq (2.85)
        θ_{n+1} = update parameter according to eq (2.94) – (2.99)
        Converged()
        n = n + 1
    end
end
```

รูปที่ 3.1 รหัสเทียมของฟังก์ชันหลัก



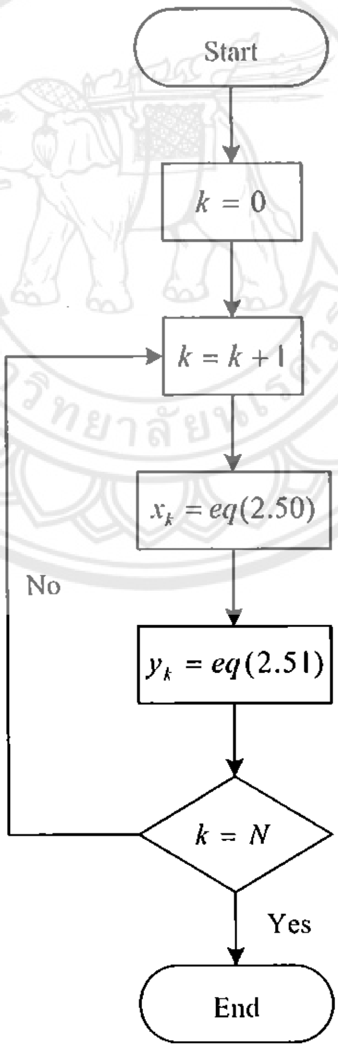
รูปที่ 3.2 แผนภาพการทำงานของฟังก์ชันหลัก

3.1 การสร้างสัญญาณออก

สังเคราะห์สัญญาณออก y_k จากระบบด้วยสมการ (2.50) และ (2.51) โดยมีรหัสเทียมและแผนภาพการทำงานดังรูปที่ 3.3 และ 3.4

```
Gen_Y()
        for k = 1 : N
                x_k = calculate state variable according to eq (2.50)
                y_k = calculate output according to eq (2.51)
        end
end
```

รูปที่ 3.3 รหัสเทียมของการจำลองสัญญาณออก



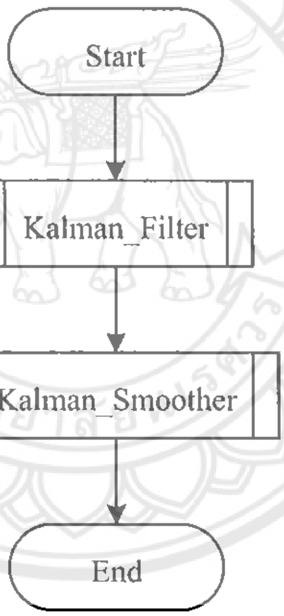
รูปที่ 3.4 แผนภาพการทำงานของการจำลองสัญญาณออก

3.2 การสังเคราะห์ตัวแปรสถานะด้วยวิธีการของคาลมาน

กำหนดค่าพารามิเตอร์เริ่มต้นแล้วนำ $\{u_k, y_k\}$ มาสังเคราะห์ตัวแปรสถานะด้วยวิธีการของคาลมาน โดยสามารถแบ่งออกเป็น 2 ขั้นตอนหลัก กล่าวคือ ตัวกรองคาลมาน (Kalman_Filter) และตัวปรับเรียบคาลมาน (Kalman_Smoother) ซึ่งสามารถอธิบายในรูปแบบของรหัสเทียมและแผนภาพดังรูปที่ 3.5 และ 3.6

```
Gen_X̂()
        Kalman_Filter()
        Kalman_Smoother()
end
```

รูปที่ 3.5 รหัสเทียมของการสังเคราะห์ตัวแปรสถานะ



รูปที่ 3.6 แผนภาพการทำงานของการสังเคราะห์ตัวแปรสถานะ

3.2.1 สังเคราะห์ตัวแปรสถานะผ่านตัวกรองคาลมาน

กำหนดพารามิเตอร์เริ่มต้น $\theta_1 = \{A, B, C, D, Q, R, \mu_1, \sigma_1^2\}$ และเงื่อนไขเริ่มต้น $\hat{x}_{1|0} = \mu_1, P_{1|0} = \sigma_1^2$ เพื่อเข้าสู่กระบวนการสังเคราะห์ตัวแปรสถานะ $\hat{x}_{k+1|k}, \hat{x}_{k+1|k+1}$ และเมทริกซ์ความแปรปรวนร่วมเกี่ยว $P_{k+1|k}, P_{k+1|k+1}$ ผ่านตัวกรองคาลมาน โดยสามารถเขียนเป็นรหัสเทียมและแผนภาพดังรูปที่ 3.7 และ 3.8

Kalman_Filter()

for k = 1 : N

$\hat{x}_{k+1|k}$ = calculate a priori state estimate according to eq (2.52)

$P_{k+1|k}$ = calculate a priori estimate covariance according to eq (2.53)

K_{k+1} = calculate Kalman gain according to eq (2.54)

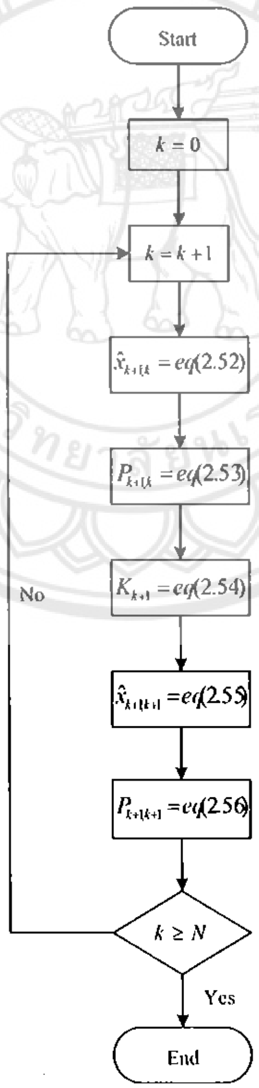
$\hat{x}_{k+1|k+1}$ = calculate a posteriori state estimate according to eq (2.55)

$P_{k+1|k+1}$ = calculate a posteriori estimate covariance according to eq (2.56)

end

end

รูปที่ 3.7 รหัสเทียมของตัวกรองคาลมาน



รูปที่ 3.8 แผนภาพการทำงานของตัวกรองคาลมาน

3.2.2 สังเคราะห์ตัวแปรสถานะผ่านตัวปรับเรียบคาลมาน

กำหนดให้เงื่อนไขเริ่มต้นของตัวปรับเรียบคาลมานเป็นดังนี้ $\hat{x}_{N|N}$ และ $P_{N|N}$ เพื่อเข้าสู่กระบวนการสังเคราะห์ตัวแปรสถานะ $\hat{x}_{k|N}$ และหาเมทริกซ์ความแปรปรวนร่วมเกี่ยว $P_{k|N}$ ผ่านตัวปรับเรียบคาลมาน โดยสามารถอธิบายในรูปแบบรหัสเทียมและแผนภาพได้ดังรูปที่ 3.9 และ 3.10

Kalman_Smoother()

    for k = N : -1 : 1

        J_k = calculate smoother gain according to eq (2.57)

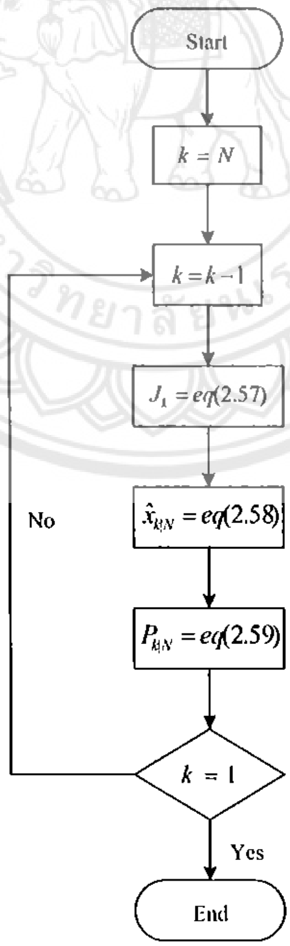
        $\hat{x}_{k|N}$ = calculate smoothed state estimate according to eq (2.58)

        $P_{k|N}$ = calculate smoothed estimate covariance according to eq (2.59)

    end

end

รูปที่ 3.9 รหัสเทียมของตัวปรับเรียบคาลมาน



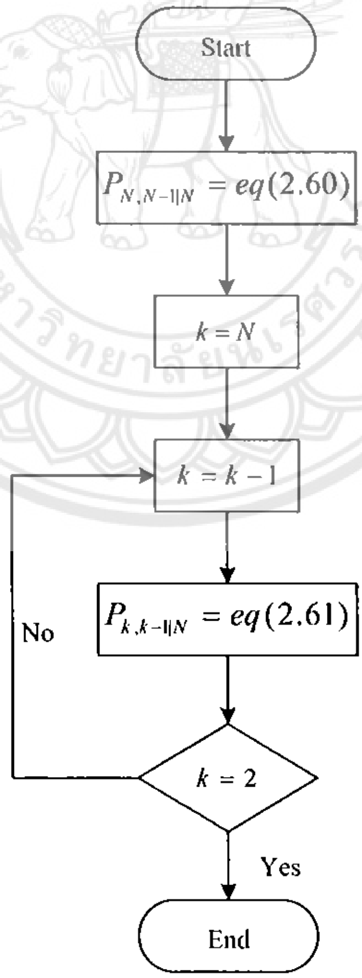
รูปที่ 3.10 แผนภาพการทำงานของตัวปรับเรียบคาลมาน

3.3 คำนวณเมทริกซ์ความแปรปรวนร่วมเกี่ยว

เมทริกซ์ความแปรปรวนร่วมเกี่ยวระหว่างตัวแปรสถานะ ณ เวลา k และตัวแปรสถานะ ณ เวลา $k-1$ $P_{k,k-1|N}$ สามารถคำนวณได้จากสมการ (2.60) และ (2.61) โดยสามารถเขียนรหัสเทียมและแผนภาพได้ดังรูปที่ 3.11 และ 3.12

```
Cross_Covariance()
    P_{N,N-1|N} = calculate first cross covariance according to eq (2.60)
    for k = N-1 : -1 : 2
        P_{k,k-1|N} = calculate cross covariance according to eq (2.61)
    end
end
```

รูปที่ 3.11 รหัสเทียมของการคำนวณความแปรปรวนร่วมเกี่ยว



รูปที่ 3.12 แผนภาพการทำงานของการคำนวณความแปรปรวนร่วมเกี่ยว

3.4 ขั้นตอน E

ขั้นตอน E เป็นการหาขอบเขตล่างของฟังก์ชันควรจะเป็น $Q(\theta_n)$ เพื่อใช้ตรวจสอบเงื่อนไขการจบการทำงานของโปรแกรม โดยอาศัยข้อมูลที่ได้จากวิธีการของคาลมาน $\{y_1, y_2, \ldots, y_N\}$, $\{\hat{x}_{1|N}, \hat{x}_{2|N}, \ldots, \hat{x}_{N|N}\}$, $\{P_{1|N}, P_{2|N}, \ldots, P_{M|N}\}$ และ $\{P_{2,1|N}, P_{3,2|N}, \ldots, P_{N,N-1|N}\}$ซึ่งสามารถหาได้จากสมการ (2.85)

3.5 ขั้นตอน M

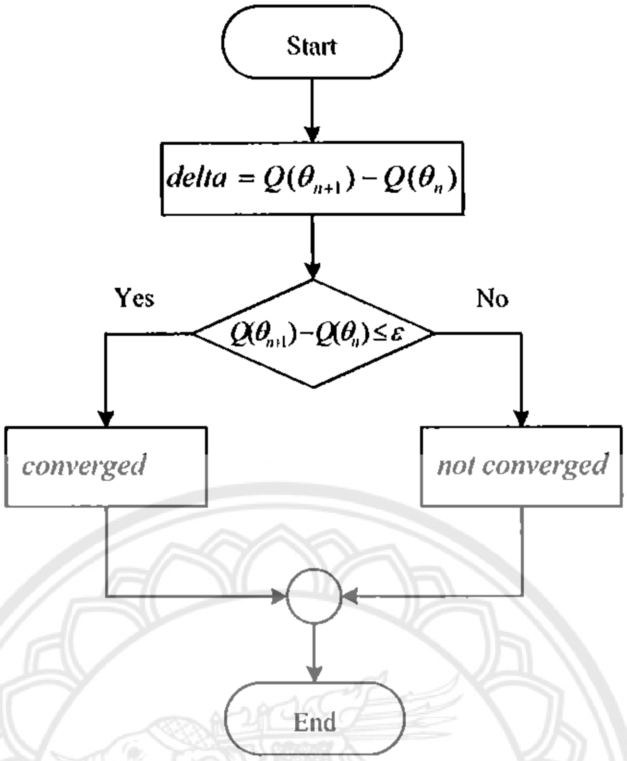
ขั้นตอน M เป็นการคำนวณหาพารามิเตอร์ตัวใหม่ที่ทำให้ขอบเขตล่างของฟังก์ชันควรจะเป็นมีค่าเพิ่มสูงขึ้น โดยอาศัยข้อมูลจากการคำนวณในวิธีการของคาลมาน โดยแทนพารามิเตอร์ตัวใหม่ด้วย $\theta_{n+1} = \{A_{new}, B_{new}, C_{new}, D_{new}, Q_{new}, R_{new}, \mu_{1,new}, \sigma^2_{1,new}\}$ ซึ่งสามารถหาได้จากสมการ (2.94) – (2.99)

3.6 ตรวจสอบเงื่อนไข

นำ θ_{n+1} ที่ได้จากขั้นตอน M คำนวณหาขอบเขตล่างของฟังก์ชันควรจะเป็นตัวใหม่ $Q(\theta_{n+1})$ เพื่อตรวจสอบผลต่างระหว่าง $Q(\theta_{n+1})$ และ $Q(\theta_n)$ เมื่อผลต่างที่ได้มีค่าค่าเป็นศูนย์หรือมีค่าน้อยมากๆ โปรแกรมจะหยุดทำงาน โดยสามารถเขียนเป็นเงื่อนไขดังนี้ $Q(\theta_{n+1}) - Q(\theta_n) \le \varepsilon$ แต่ถ้าผลต่างที่ได้ไม่สอดคล้องกับเงื่อนไข โปรแกรมทำงานซ้ำในขั้นตอน 3.2 ถึงขั้นตอน 3.5 จนกว่าผลต่างดังกล่าวจะเป็นไปตามเงื่อนไขโปรแกรมจึงจะหยุดทำงาน โดยสามารถอธิบายในรูปแบบรหัสเทียมและแผนภาพได้ดังรูปที่ 3.13 – 3.14

```
Converged()
        delta = Q(θ_{n+1}) − Q(θ_n)
        if delta ≤ ε
                converged
        else
                not converged
        end
end
```

รูปที่ 3.13 รหัสเทียมของโปรแกรมตรวจสอบเงื่อนไข



รูปที่ 3.14 แผนภาพการทำงานของโปรแกรมตรวจสอบเงื่อนไข

บทที่ 4

ผลการทดลอง

ในบทนี้นำหลักการประมาณพารามิเตอร์ที่ศึกษาในบทที่ 2 มาเขียนชุดคำสั่งด้วยโปรแกรม MATLAB® ตามแผนภาพการทำงานในบทที่ 3 โดยในบทนี้จะนำเสนอวิธีการประมาณพารามิเตอร์ โดยใช้ข้อมูล 2 ประเภทกล่าวคือ (1) ข้อมูลที่ได้จากการจำลอง ซึ่งได้จากแบบจำลองระบบควบคุมความเร็วรถและแบบจำลองตัวกรองดิจิตอลแบบ IIR และ (2) ข้อมูลจริงที่ได้จากตลาดอัตราแลกเปลี่ยนสกุลเงินดอลลาร์สหรัฐต่อสกุลเงินบาทไทย ซึ่งนำมาใช้ประมาณพารามิเตอร์ของแบบจำลองความผันผวนเชิงเฟ้นสุ่ม โดยมีรายละเอียดดังนี้

4.1 แบบจำลองระบบควบคุมความเร็วรถ

แบบจำลองของระบบควบคุมความเร็วรถ (Cruise Control) เขียนบรรยายได้ดังนี้

$$\Sigma := \begin{cases} \dot{x}(t) = -\dfrac{b}{m}x(t) + \dfrac{500}{m}u(t) \\ z(t) = x(t) \end{cases} \tag{4.1}$$

เมื่อ m คือมวลของรถมีหน่วยเป็นกิโลกรัม (kg) b คือสัมประสิทธิ์การหน่วงในตัวหน่วงกันสั่นสะเทือน (Shock absorber) มีหน่วยเป็นนิวตัน-วินาทีต่อเมตร $(N \cdot m/s)$ และ $x(t)$ คือความเร็วรถในที่นี้เลือกเป็นตัวแปรสถานะ

โดยทั่วไป อุปกรณ์ในการวัดความเร็วมักมีการปนเปื้อนจากสัญญาณรบกวนความถี่สูง ดังนั้นสิ่งที่ผู้ออกแบบวัดได้คือ $y(t) = z(t) + v(t)$ โดยในที่นี้สมมติให้ $v(t)$ มีการแจกแจงแบบปกติ นั่นคือ $v(t) \sim \mathcal{N}(0,R)$ นอกจากนี้ กระบวนการภายในระบบมักมีการรบกวนด้วยสัญญาณรบกวนความถี่สูงซึ่งสมมติให้มีการแจกแจงปกติด้วยเช่นกัน นั่นคือ $w(t) \sim \mathcal{N}(0,Q)$ ดังนั้น ระบบ Σ ใน (4.1) จึงเขียนใหม่ให้อยู่ในรูปแบบของกระบวนการเชิงเฟ้นสุ่มดังนี้

$$\Sigma_1 := \begin{cases} \dot{x}(t) = -\dfrac{b}{m}x(t) + \dfrac{500}{m}u(t) + w(t) \\ y(t) = x(t) + v(t) \end{cases} \tag{4.2}$$

สมมติว่าผู้ออกแบบสามารถวัดความเร็วของรถได้และต้องการทราบพารามิเตอร์ m และ b จึงเลือกใช้ขั้นตอนวิธี EM มาประมาณพารามิเตอร์ทั้งสองโดยต้องทำการแปลงระบบ Σ_1 ใน (4.2)

ให้เป็นระบบที่บรรยายในเวลาวิยุตโดยอาศัยหลักการแปลงดังแสดงไว้ในบทที่ 2 ดังนั้นระบบที่ถูกนำมาใช้ในการทดลองจึงเขียนได้ดังนี้

$$\Sigma_1^d := \begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = x_k + v_k \end{cases} \tag{4.3}$$

โดยที่

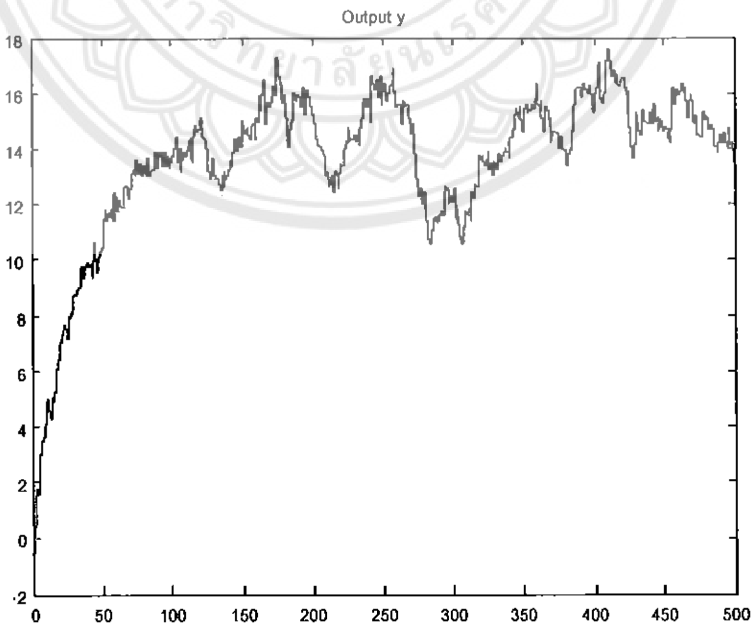
$$A = e^{-\frac{b}{m}T} \text{ และ } B = \frac{500}{b}\left(1 - e^{-\frac{b}{m}T}\right) \tag{4.4}$$

เมื่อ T คือคาบเวลาของการชักตัวอย่างซึ่งในที่นี้เลือกให้ $T = 1$ วินาที

ในการทดลอง สมมติให้ $m = 1{,}075\ kg$, $b = 35\ N \cdot m/s$, $R = 0.05$ และ $Q = 0.1$ เมื่อแทนลงในระบบ Σ_1^d ใน (4.3) จึงได้ว่า

$$\Sigma_1^d := \begin{cases} x_{k+1} = 0.9680x_k + 0.4576u_k + w_k \\ y_k = x_k + v_k \end{cases} \tag{4.5}$$

เมื่อนำระบบ Σ_1^d ใน (4.5) มาจำลองด้วยโปรแกรม โดยป้อนสัญญาณเข้า u_k อยู่ในรูปแบบของฟังก์ชันขั้นบันไดหนึ่งหน่วยจะได้ผลตอบสนอง y_k ดังแสดงในรูปที่ 4.1

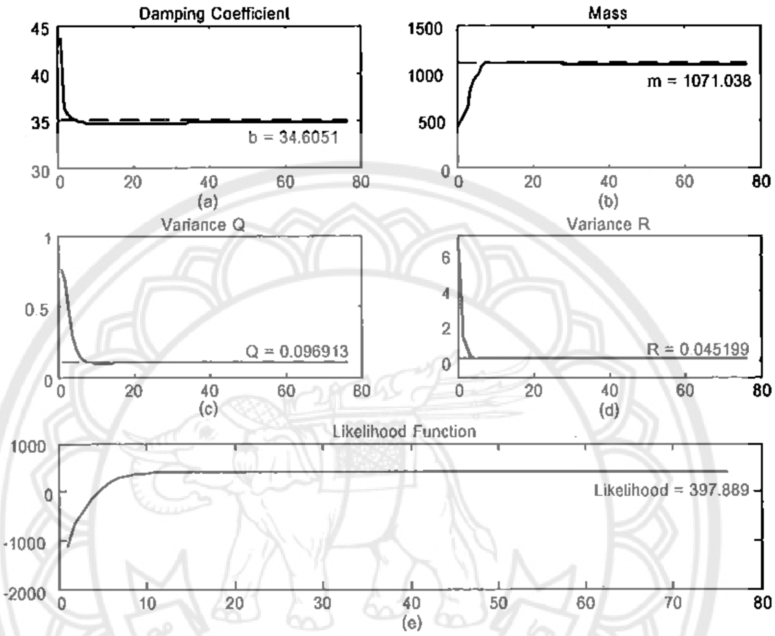


รูปที่ 4.1 ผลตอบสนองของระบบควบคุมความเร็วรถ

เมื่อนำผลตอบสนอง y_k มาเข้าขั้นตอนวิธี EM เพื่อประมาณพารามิเตอร์ โดยกำหนดให้เงื่อนไขเริ่มต้นดังนี้

$$A_1 = 0.5,\ B_1 = 1,\ Q_1 = 1, R_1 = 1 \text{ และ } x_1 \sim \mathcal{N}(0,0.1)$$

จะได้พารามิเตอร์แสดงไว้ในรูปที่ 4.2(a) – 4.2(d) โดยมีฟังก์ชันควรจะเป็นแสดงไว้ในรูป 4.2(e)



รูปที่ 4.2 พารามิเตอร์และฟังก์ชันควรจะเป็นของระบบควบคุมความเร็วรถ

เมื่อนำพารามิเตอร์ที่ได้จากขั้นตอนวิธี EM มาเปรียบเทียบกับค่าจริงของระบบจะได้ดังตารางที่ 4.1

ตารางที่ 4.1 การเปรียบเทียบพารามิเตอร์จากการประมาณและค่าจริงของระบบควบคุมความเร็วรถ

พารามิเตอร์	พารามิเตอร์จริง	พารามิเตอร์จากการประมาณ	ร้อยละความคลาดเคลื่อน
สัมประสิทธิ์การหน่วง	$b = 35$	$\hat{b} = 34.6051$	$\left\lvert\frac{\hat{b}-b}{b}\right\rvert \times 100 = 1.13\%$
มวล	$m = 1075$	$\hat{m} = 1071.04$	$\left\lvert\frac{\hat{m}-m}{m}\right\rvert \times 100 = 0.37\%$
ความแปรปรวน Q	$Q = 0.1$	$\hat{Q} = 0.0969$	$\left\lvert\frac{\hat{Q}-Q}{Q}\right\rvert \times 100 = 3.10\%$
ความแปรปรวน R	$R = 0.05$	$\hat{R} = 0.0452$	$\left\lvert\frac{\hat{R}-R}{R}\right\rvert \times 100 = 9.60\%$

จากผลการทดลองในตารางที่ 4.1 พบว่าพารามิเตอร์จากการประมาณด้วยขั้นตอนวิธี EM มีค่าใกล้เคียงกับค่าจริงโดยมีร้อยละความคลาดเคลื่อนค่อนข้างต่ำเมื่อเทียบกับค่าจริง ในขณะเดียวกันความแปรปรวน R มีร้อยละความคลาดเคลื่อนค่อนข้างสูง

4.2 แบบจำลองตัวกรองแบบ IIR

แบบจำลองตัวกรองแบบ IIR (Infinite Impulse Response Filter : IIR Filter) สามารถอธิบายในรูปแบบสมการเชิงผลต่าง ได้ดังนี้

$$z_k = K_1 z_{k-2} + K_2 z_{k-1} + K_3 u_{k-2} + K_4 u_{k-1} \tag{4.6}$$

กำหนดให้

$$x_k^1 = K_2 x_{k-1}^2 + K_4 u_{k-1}$$
$$x_k^2 = K_1 x_{k-1}^2 + x_{k-1}^1 + K_3 u_{k-1}$$
$$z_k = x_k^2$$

จะได้ว่า

$$\begin{aligned} z_k &= x_k^2 \\ &= K_1 x_{k-1}^2 + x_{k-1}^1 + K_3 u_{k-1} \\ &= K_1 x_{k-1}^2 + \left[K_2 x_{k-2}^2 + K_4 u_{k-2}\right] + K_3 u_{k-1} \\ &= K_1 z_{k-1} + K_2 z_{k-2} + K_3 u_{k-1} + K_4 u_{k-2} \end{aligned}$$

ดังนั้นสมการเชิงผลต่างใน (4.6) จึงเขียนในรูปแบบปริภูมิสถานะ ได้เป็นดังสมการ (4.7) โดยกำหนดให้ $\mathrm{x}_k = \begin{bmatrix} x_k^1 & x_k^2 \end{bmatrix}^T$

$$\Sigma_2 := \begin{cases} \mathrm{x}_k = \begin{bmatrix} 0 & K_2 \\ 1 & K_1 \end{bmatrix} \mathrm{x}_{k-1} + \begin{bmatrix} K_4 \\ K_3 \end{bmatrix} u_{k-1} \\ z_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \mathrm{x}_k \end{cases} \tag{4.7}$$

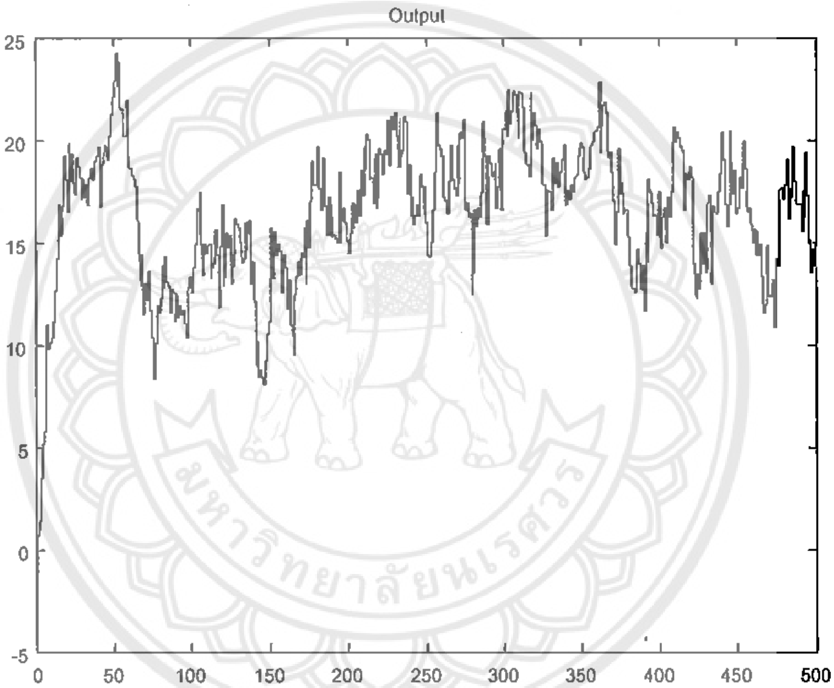
โดยทั่วไป อุปกรณ์ในการวัดมักมีการปนเปื้อนจากสัญญาณรบกวนความถี่สูง ดังนั้นสิ่งที่วัดได้คือ $y_k = z_k + v_k$ โดยในที่นี้สมมติให้ v_k มีการแจกแจงแบบปกติ นั่นคือ $v_k \sim \mathcal{N}(0, R)$ ระบบในสมการ (4.7) จึงเขียนใหม่ให้อยู่ในรูปแบบของกระบวนการเชิงเฟ้นสุ่มดังนี้

$$\Sigma_2 := \begin{cases} \mathrm{x}_k = \begin{bmatrix} 0 & K_2 \\ 1 & K_1 \end{bmatrix} \mathrm{x}_{k-1} + \begin{bmatrix} K_4 \\ K_3 \end{bmatrix} u_{k-1} \\ y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \mathrm{x}_k + v_k \end{cases} \tag{4.8}$$

ในการทดลองสมมติให้ $K_1 = 0.2142$, $K_2 = 0.6735$, $K_3 = 1.2873$, $K_4 = 0.5967$ และ $R = 0.2$ แทนค่าระบบในสมการ (4.8) จะได้ดังนี้

$$\Sigma_2 := \begin{cases} \mathrm{x}_{k+1} = \begin{bmatrix} 0 & 0.6735 \\ 1 & 0.2142 \end{bmatrix} \mathrm{x}_k + \begin{bmatrix} 0.5967 \\ 1.2873 \end{bmatrix} u_k \\ y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \mathrm{x}_k + v_k \end{cases} \tag{4.9}$$

นำระบบ Σ_2 ในสมการ (4.9) มาจำลองหาผลตอบสนอง y_k โดยป้อนสัญญาณเข้า u_k เป็นแบบขบวนอิมพัลส์ จะได้ y_k ดังแสดงในรูปที่ 4.3

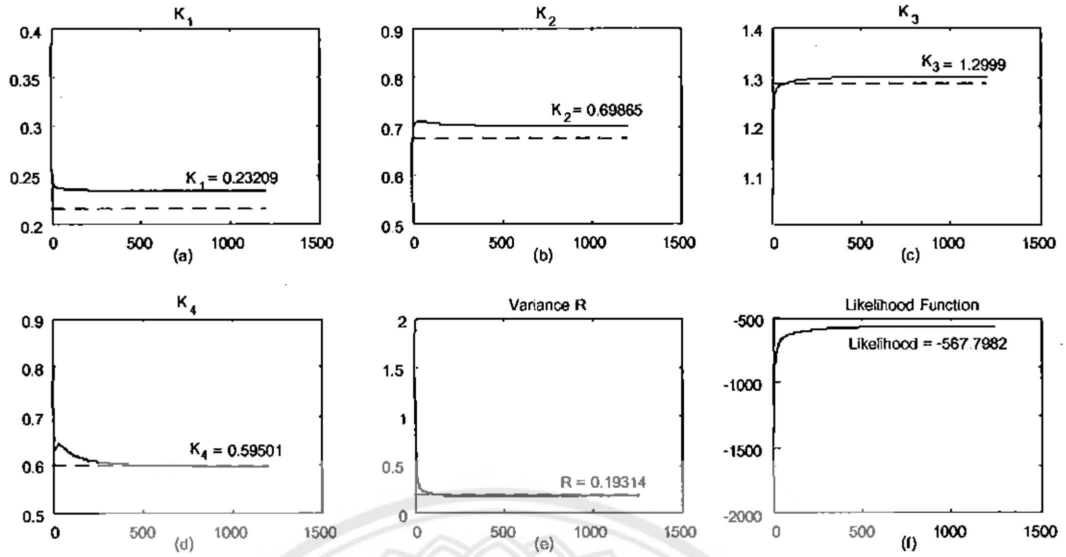


รูปที่ 4.3 ผลตอบสนองของตัวกรองแบบ IIR

นำ y_k ในรูปที่ 4.3 มาเข้าขั้นตอนวิธี EM เพื่อประมาณพารามิเตอร์โดยกำหนดเงื่อนไขเริ่มต้นดังนี้

$$K_{1,1} = 0.5\,, K_{2,1} = 0.5\,, K_{3,1} = 1\,, K_{4,1} = 1\,, R_1 = 1 \text{ และ } x_1 \sim \mathcal{N}(0,1)$$

จะได้พารามิเตอร์ดังรูปที่ 4.4(a) – 4.4(e) และฟังก์ชันควรจะเป็นดังรูปที่ 4.4(f)



รูปที่ 4.4 พารามิเตอร์และฟังก์ชันควรจะเป็นของตัวกรองดิจิตอลแบบ IIR

เมื่อนำพารามิเตอร์ที่ได้จากขั้นตอนวิธี EM มาเปรียบเทียบกับค่าจริง จะได้ดังตารางที่ 4.2

ตารางที่ 4.2 การเปรียบเทียบพารามิเตอร์จากการประมาณและค่าจริงของตัวกรองดิจิตอลแบบ IIR

พารามิเตอร์	พารามิเตอร์จริง	พารามิเตอร์จากการประมาณ	ร้อยละความคลาดเคลื่อน
พารามิเตอร์ K_1	$K_1 = 0.2142$	$\hat{K}_1 = 0.2321$	$\left\|\dfrac{\hat{K}_1 - K_1}{K_1}\right\| \times 100 = 8.36\%$
พารามิเตอร์ K_2	$K_2 = 0.6735$	$\hat{K}_2 = 0.6987$	$\left\|\dfrac{\hat{K}_2 - K_2}{K_2}\right\| \times 100 = 3.74\%$
พารามิเตอร์ K_3	$K_3 = 1.2873$	$\hat{K}_3 = 1.2999$	$\left\|\dfrac{\hat{K}_3 - K_3}{K_3}\right\| \times 100 = 0.98\%$
พารามิเตอร์ K_4	$K_4 = 0.5967$	$\hat{K}_4 = 0.5950$	$\left\|\dfrac{\hat{K}_4 - K_4}{K_4}\right\| \times 100 = 0.28\%$
ความแปรปรวน R	$R = 0.2$	$\hat{R} = 0.1931$	$\left\|\dfrac{\hat{R} - R}{R}\right\| \times 100 = 3.45\%$

จากผลการทดลองในตารางที่ 4.2 พบว่าร้อยละความคลาดเคลื่อนมีค่าสูงกว่าผลการทดลองในตารางที่ 4.1 เล็กน้อย ซึ่งเป็นผลมาจากการดำเนินการในรูปแบบเมทริกซ์และจำนวนพารามิเตอร์ในการประมาณมีมากกว่าการทดลองก่อนหน้า แต่อย่างไรก็ตามร้อยละความคลาดเคลื่อนยังมีค่าอยู่ในเกณฑ์ที่ผู้จัดทำโครงงานยอมรับได้

4.3 แบบจำลองความผันผวนเชิงเฟ้นสุ่ม

แบบจำลองความผันผวนสามารถจำแนกได้เป็น 2 กลุ่ม กล่าวคือ กลุ่มที่ 1 แบบจำลองความผันผวนเชิงกำหนด (Deterministic volatility model : DV) ตัวอย่างของแบบจำลองกลุ่มนี้ได้แก่ แบบจำลอง GARCH ต่อมาได้มีการพัฒนาเพื่อให้ได้แบบจำลองที่เหมาะสมกับการประยุกต์ใช้งาน เช่น NGARCH, IGARCH และ TGARCH และกลุ่มที่ 2 แบบจำลองความผันผวนเชิงเฟ้นสุ่ม (Stochastic volatility model : SV) ตัวอย่างของแบบจำลองกลุ่มนี้ได้แก่ แบบจำลองเทย์เลอร์ แบบจำลองเฮสตัน แบบจำลองอัลฟา-เบตา-โร เชิงเฟ้นสุ่มและแบบจำลองเชน ในการนำแบบจำลองความผันผวนมาประยุกต์ใช้กับดัชนีราคาหุ้น [12], [26] และอัตราแลกเปลี่ยนเงินตราต่างประเทศ [11], [24] นิยมใช้แบบจำลอง GARCH ในการวิเคราะห์ เท่าที่ผู้จัดทำทราบงานวิจัย [4] ถือได้ว่าเป็นงานวิจัยชิ้นแรกที่นำแบบจำลองความผันผวนเชิงเฟ้นสุ่มมาใช้บรรยายพฤติกรรมความผันผวนของอัตราแลกเปลี่ยนสกุลเงินดอลลาร์สหรัฐต่อสกุลเงินบาทไทย อย่างไรก็ตามในงานวิจัยดังกล่าวเลือกใช้แบบจำลองของ Taylor [25] ในรูปแบบไม่เชิงเส้น ซึ่งต่างจากในโครงงานนี้ซึ่งเลือกใช้วิธีการประมาณแบบจำลองไม่เชิงเส้นให้อยู่ในรูปแบบของแบบจำลองเชิงเส้นเพื่อสามารถประยุกต์ใช้วิธีการของคาลมานได้ โดยใช้ข้อมูลของอัตราแลกเปลี่ยนสกุลเงินดอลลาร์สหรัฐต่อสกุลเงินบาทไทย (USD/THB) ตั้งแต่วันที่ 4 ม.ค. 2554 จนถึง 29 เม.ย. 2559

กำหนดให้ R_k คืออัตราแลกเปลี่ยน และ $r_k = \log\left(\frac{R_k}{R_{k-1}}\right)$ คือผลตอบแทนแบบลอการิทึมของ R_k แบบจำลองความผันผวนสามารถเขียนบรรยายได้ดังนี้

$$\Sigma_{NL} := \begin{cases} x_k = \phi x_{k-1} + w_{k-1} \\ r_k = \beta \exp(0.5 x_k)\varepsilon_k \end{cases} \tag{4.10}$$

เนื่องจากการพิจารณาแบบจำลองตามวิธีการของคาลมานจำเป็นต้องพิจารณาในรูปแบบเชิงเส้นเท่านั้น ดังนั้นจึงต้องทำการแปลงแบบจำลองข้างต้นให้อยู่ในรูปแบบเชิงเส้นโดยอาศัยฟังก์ชันลอการิทึม จึงจะได้ดังนี้

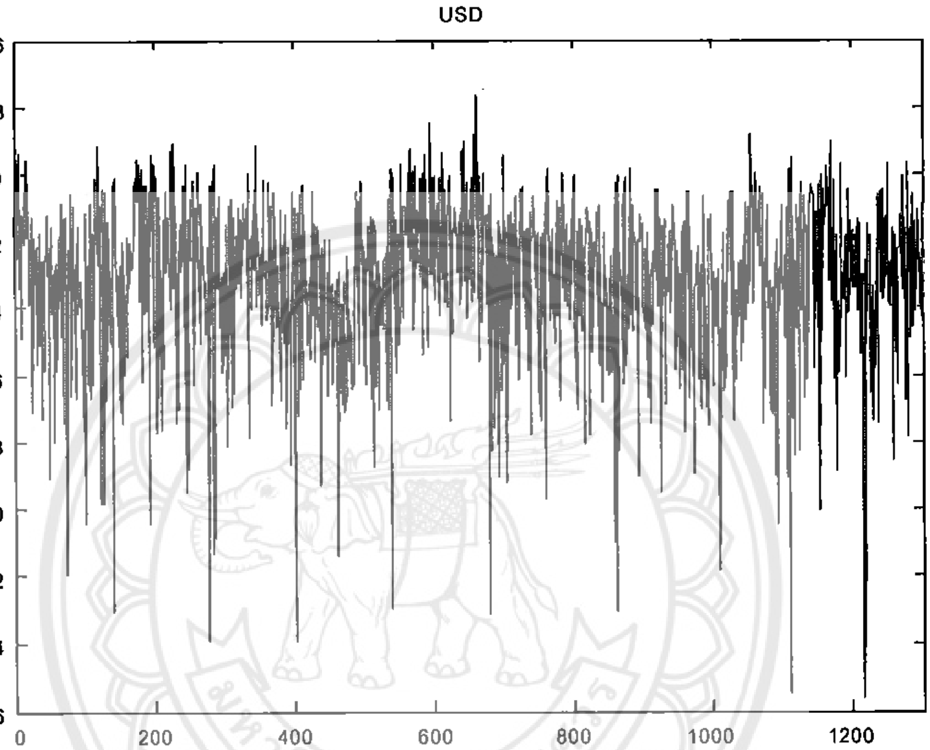
$$\Sigma_3 := \begin{cases} x_k = \phi x_{k-1} + w_{k-1} \\ y_k = \alpha + x_k + v_k \end{cases} \tag{4.11}$$

โดยที่ $y_k = \log r_k^2, \alpha = \log\beta^2 + E[\log\varepsilon_k^2]$ และ $v_k = \log\varepsilon_k^2 - E[\log\varepsilon_k^2]$

พิจารณาสมการ (4.11) พบว่าผลการแปลงทำให้ v_k มีการแจกแจงเป็นลอการิทึมไคกำลังสอง (Log-Chi Squared Distribution) ที่มีค่าเฉลี่ยเป็น 0 และมีความแปรปรวนเป็น $0.5\pi^2$ มีงานวิจัย

หลายชิ้นที่นำวิธีการของคาลมานมาประยุกต์ใช้กับแบบจำลองในสมการ (4.11) โดยได้ประมาณให้ $v_k \sim \mathcal{N}(0,0.5\pi^2)$

เมื่อนำข้อมูลอัตราแลกเปลี่ยน USD/THB ตั้งแต่วันที่ 4 ม.ค. 2554 จนถึง 29 เม.ย. 2559 มาหา y_k ซึ่งจะเป็นค่าลอการิทึมของผลตอบแทนยกกำลังสอง จะได้กราฟดังรูป 4.5

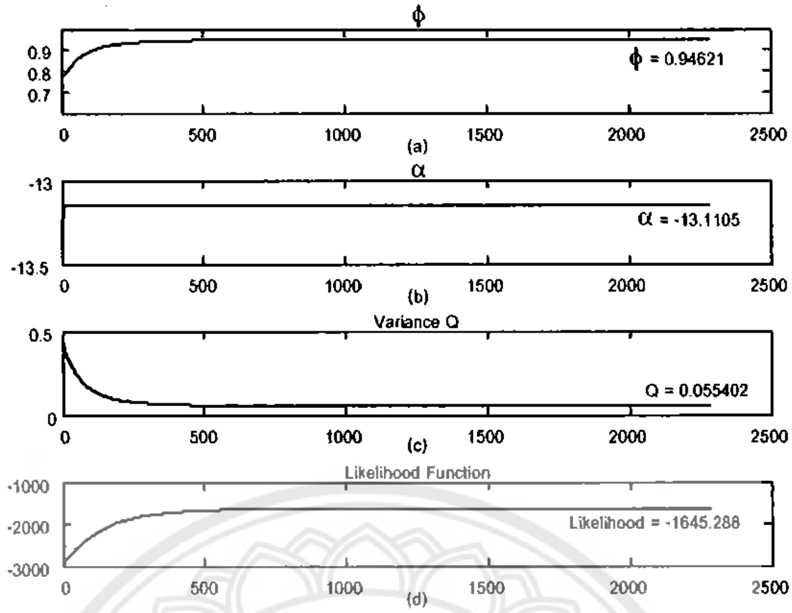


รูปที่ 4.5 ค่าลอกาลิทึมของผลตอบแทนยกกำลังสอง

นำสัญญาณ y_k ในรูป 4.5 มาเข้าขั้นตอนวิธี EM เพื่อประมาณพารามิเตอร์ ϕ , α และ Q โดยกำหนดให้เงื่อนไขเริ่มต้นดังนี้

$$\phi_1 = 0.9 \ , \ \alpha_1 = -13.5 \ , \ Q_1 = 0.5 \text{ และ } x_1 \sim \mathcal{N}(0,1)$$

จะได้พารามิเตอร์ดังรูปที่ 4.6(a) – 4.6(c) และฟังก์ชันควรจะเป็นดังรูปที่ 4.6(d)



รูปที่ 4.6 พารามิเตอร์และฟังก์ชันควรจะเป็นของแบบจำลองความผันผวนเชิง**เฟ้นสุ่ม**

ตารางที่ 4.3 พารามิเตอร์จากการประมาณของแบบจำลองความผันผวนเชิงเ**ฟ้นสุ่ม**

พารามิเตอร์	พารามิเตอร์จากการประมาณ
พารามิเตอร์ ϕ	$\hat{\phi} = 0.94621$
พารามิเตอร์ α	$\hat{\alpha} = -13.1105$
ความแปรปรวน Q	$\hat{Q} = 0.0554$

บทที่ 5
สรุปและข้อเสนอแนะ

5.1 สรุปผลการดำเนินงาน

โครงงานนี้ศึกษาวิธีการประมาณค่าพารามิเตอร์ด้วยขั้นตอนวิธี EM สำหรับระบบที่เขียนบรรยายในรูปแบบปริภูมิสถานะและทำการจำลองผลการศึกษาโดยใช้โปรแกรม MATLAB จากกรณีศึกษาทั้ง 2 กลุ่มข้อมูลพบว่าพารามิเตอร์จาการประมาณเมื่อใช้ข้อมูลจากการจำลองมีค่าใกล้เคียงกับพารามิเตอร์จริง เมื่อนำข้อมูลจริงมาใช้ประมาณพารามิเตอร์ดังแสดงในหัวข้อ 4.3 พิจารณาฟังก์ชันควรจะเป็นพบว่ามีค่าเพิ่มสูงขึ้น ในทางทฤษฎีเมื่อฟังก์ชันควรจะเป็นมีค่าเพิ่มสูงขึ้นจะทำให้พารามิเตอร์จากการประมาณลู่เข้าสู่ค่าจริงและแบบจำลองที่ได้จะมีความใกล้เคียงกับแบบจำลองจริง เมื่อพิจารณาพารามิเตอร์จากการประมาณพบว่ามีการลู่เข้าสู่ค่าค่าหนึ่งซึ่งเป็นไปตามทฤษฎี

5.2 ข้อเสนอแนะและแนวทางการพัฒนาโครงงาน

1. ทดลองใช้ตัวกรองและตัวปรับเรียบในรูปแบบอื่นๆ มาสังเคราะห์ตัวแปรสถานะแล้วนำผลที่ได้มาเปรียบเทียบกับวิธีการของคาลมาน

2. การสังเคราะห์ตัวแปรสถานะด้วยวิธีการของคาลมานสามารถทำงานได้อย่างมีประสิทธิภาพภายใต้สมมติฐานความเป็นเชิงเส้นของระบบและสัญญาณรบกวนอยู่ในรูปแบบการแจกแจงปกติเท่านั้น สำหรับระบบไม่เชิงเส้นที่มีสัญญาณรบกวนที่ไม่ใช่การแจกแจงปกติจำเป็นต้องหาวิธีการกรองและการปรับเรียบวิธีอื่นมาใช้แทน เช่นในกลุ่มของ Sequential Monte Carlo

3. ควรทดลองใช้สัญญาณเข้าในรูปแบบอื่น เช่น สัญญาณรูปไซน์ (Sinusoidal signal) สัญญาณฟันเลื่อย (Sawtooth signal) และสัญญาณพัลส์ (Pulse signal) มาป้อนให้กับระบบแล้วใช้ขั้นตอนวิธี EM มาประมาณพารามิเตอร์ของระบบ

เอกสารอ้างอิง

[1] กฤษติพงษ์ เพ็ญภาคและขวัญใจ สมรัก. (2556). *ตัวกรองคาลมานและการประยุกต์ใช้งานในด้านวิศวกรรมอุตสาหการกรณีศึกษาการพยากรณ์ความต้องการของลูกค้า*.ปริญญานิพนธ์ วศ.บ. (วิศวกรรมอุตสาหการ),มหาวิทยาลัยนเรศวร.พิษณุโลก.

[2] คณดิถ เจษฎ์พัฒนานนท์และพรชัย พฤกษ์ภัทรานนต์. (2553). *สัญญาณและระบบ*.สงขลา : ภาควิชาวิศวกรรมไฟฟ้า มหาวิทยาลัยสงขลานครินทร์.

[3] ณัฐพงศ์ พงศ์ไพจิตรวงศ์และวีระพล บัวแก้ว. (2553). *การปรับค่าตัวควบคุม PID ด้วยขั้นตอนวิธีอาณานิคมมด*. ปริญญานิพนธ์ วศ.บ. (วิศวกรรมคอมพิวเตอร์), มหาวิทยาลัยนเรศวร. พิษณุโลก.

[4] ธนภัทร เอี่ยมตาล. (2559). *ขั้นตอนวิธีอีเอ็มสำหรับการประมาณพารามิเตอร์ของแบบจำลองความผันผวนเชิงเฟ้นสุ่ม*. วิทยานิพนธ์ ปร.ด. (วิศวกรรมไฟฟ้า), มหาวิทยาลัยนเรศวร. พิษณุโลก.

[5] ปิยะ โควินท์ทวีวัฒน์. (2552). *สัญญาณและระบบกับการประยุกต์ใช้โปรแกรม SCILAB*. นครปฐม : โปรแกรมวิศวกรรมโทรคมนาคม คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม.

[6] มงคล สีห์โสภณ. (2534). *พีชคณิตเชิงเส้น*. กรุงเทพฯ : ประกายพรึก.

[7] สิทธิโชค ใจยอด. (2550). *การออกแบบระบบเชิงเส้นโดยใช้วิธีอสมการเมทริกซ์เชิงเส้น*. ปริญญานิพนธ์ วศ.บ. (วิศวกรรมไฟฟ้า), มหาวิทยาลัยนเรศวร, พิษณุโลก.

[8] อำพน ศรัณยชัย. (2519). *ทฤษฎีเมทริกซ์และการแปรรูปเชิงเส้น*.กรุงเทพฯ : มหาวิทยาลัยศิลปกร.

[9] Alkin and Oktay. (2014). *Signals and systems : a MATLAB integrated approach*. Boca Raton: CRC Press.

[10] Andrew C. Harvey. (1990) . *Forecasting structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

[11] Dahiru A. Bala and Joseph O. Asemota. (2013). Exchange-Rate Volatility in Nigeria: Application of GRACH Models with Exogenous Break. *CBN Journal of Applied Statistics: Vol. 4* (pp. 89–116).

[12] Dana AL-Najjar. (2016). Modelling and Estimation of Volatility Using ARCH/GARCH Models in Jordan's Stock Market. *Asian Journal of Finance & Accounting: Vol. 8* (pp. 152-167)

[13] Fridman M. and Harris L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economics Statistics*, 1(3), 284-291.

[14] Geoffrey J. McLachlan and Thriyambakam Krishnan. (2008). *The EM algorithm and extensions* (2nd ed). Hoboken, N.J.: Wiley-Interscience.

[15] Gopal M. (2008). *Control systems: principles and design*. Boston: McGraw-Hill.

[16] Gopal M. (2009). *Digital control and state variable methods: conventional and intelligent control systems* (3rd ed). New Delhi: McGraw- Hill.

[17] Jeongeun Kim. (2005). *Parameter Estimation In Stochastic Volatility Models With Missing Data Using Particle Methods And The EM Algorithm.* Thesis Ph. D. (Statistics), University of Pittsburgh, Pittsburgh.

[18] Katsuhiko Ogata. (1995). *Discrete-time control systems (2nd ed).* Englewood Cliffs, N.J.: Prentice-Hall Interantional.

[19] Mohinder S. Grewal and Angus P. Andrews. (2008).*Kalman filtering:theory and practice using MATLAB* (3rd ed). Hoboken, N.J.: Wiley.

[20] Nahvi and Mahmood. (2014). *Signals and systems.* New York, N.Y.: McGraw Hill.

[21] Norman S. Nise. (2011). *Control systems engineering* (6th ed). Hoboken, N.J.: Wiley.

[22] PeterOlofsson andMikaelAndersson. (2012). *Probability, statistics, and stochastic processes*(2nd ed). Hoboken, N.J.: Wiley-Interscience.

[23] Shuo Chen. (2006). *The Application of the Expectation-Maximization Algorithm to the Identification of Biological Models.* Thesis M. Eng. (Electrical Engineering), Virginia polytechnic institute and state university, Blacksburg.

[24] Suliman Zakaria Suliman Abdalla. (2012). Modelling Exchange Rate Volatility using GARCH Model: Empirical Evidence from Arab Countries. *International Journal of Economics and Finance: Vol. 4* (pp. 216–229).

[25] Taylor S. J. (1982). Financial returns modeled by the product of two stochastic processes–a study of daily sugar prices 1961–79. *In Time Series Analysis: Theory and Practice: Vol. 1*(pp. 203–226). New York: Elsevier Science Publishing.

[26] Wei Jiang. (2012).*Using the GARCH model to analyze and predict the different stock markets.* Thesis M.S. (Statistics), Uppsala University, Uppsala

ประวัติผู้ดำเนินโครงงาน

ชื่อ นายหิรัญรักษ์ รักษ์ฤทธิกุล

ภูมิลำเนา 360 หมู่ 3 ต.อรัญญิก อ.เมือง จ.พิษณุโลก 65000

ประวัติการศึกษา

- จบระดับมัธยมศึกษาจากโรงเรียนพิษณุโลก-พิทยาคม
- ปัจจุบันกำลังศึกษาในระดับปริญญาตรีชั้นปีที่ 4 สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

E-mail : **hirunrakr55@gmail.com**