

การหาเอกลักษณ์ของระบบ

System Identification



# การหาเอกลักษณ์ของระบบ System Identification

---

จิตโกมุท ส่งศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย

จิตโกมุท ส่งศิริ

การหาเอกลักษณ์ของระบบ / จิตโกมุท ส่งศิริ

1. เอกลักษณ์ของระบบ.

003.1

ISBN 978-616-590-942-6

สพจ. Q22/09/2565



สรรคุณคำวิชาการ สู้งาม  
Knowledge to All  
www.cupress.chula.ac.th

หนังสือเล่มนี้ผ่านการตรวจคุณภาพวิชาการจากผู้ทรงคุณวุฒิของสำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย

สงวนลิขสิทธิ์เนื้อหาและภาพประกอบของหนังสือเล่มนี้โดย จิตโกมุท ส่งศิริ © พ.ศ. 2565  
ห้ามคัดลอก จัดพิมพ์ หรือทำซ้ำ ก่อนได้รับอนุญาตเป็นลายลักษณ์อักษร

พิมพ์ครั้งที่ 1 จำนวน 200 เล่ม พ.ศ. 2565

จัดทำโดย จิตโกมุท ส่งศิริ  
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

**บรรณาธิการอำนวยการ :** นางอรทัย นันทนาดิษฐ์ และรองศาสตราจารย์ ดร.อรรักษ์ หาญสีบสาย

**กองบรรณาธิการฝ่ายวิชาการ :** ศาสตราจารย์กิตติคุณ ดร.ปิยนาด บุนนาค  
รองศาสตราจารย์ ดร.พิมพ์พันธ์ เดชะคุปต์  
รองศาสตราจารย์ นายแพทย์ชิษณุ พันธุ์เจริญ  
รองศาสตราจารย์ ดร.วิมลวรรณ พิมพ์พันธ์

**ผู้ประสานงาน :** วาสนา ชำเซ็น

**ผู้พิสูจน์อักษร :** บุศรินทร์ สรวัดร

**ออกแบบปก :** กิตติภูมิ ส่งศิริ และอาทิมา อุทัยชลาพันธ์

**ออกแบบรูปเล่ม :** จิตโกมุท ส่งศิริ

**พิมพ์ที่ :** สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย [CUP6506-035]

ถนนพญาไท เขตปทุมวัน กรุงเทพฯ 10330

โทร. 0-2218-3562-3 โทรสาร 0-2218-3547

www.cupress.chula.ac.th

## กิตติกรรมประกาศ

ผู้เขียนขอขอบพระคุณ ผศ. ดร.มานพ วงศ์สายสุวรรณ สำหรับคำแนะนำและข้อมูลที่เป็นประโยชน์ เมื่อผู้เขียนเริ่มสอนวิชาการหาเอกลักษณ์ของระบบในปีพ.ศ.2554 ขอขอบพระคุณ รศ. ดร.วัชรพงษ์ ไชยพิรุณกิจ ผศ. ดร.สุชิน อรุณสวัสดิ์วงศ์ และ ศ. ดร.เดวิด บรรเจิดพงษ์ชัย คณาจารย์ในสาขาระบบควบคุม ภาควิชาวิศวกรรมไฟฟ้า ที่ได้ให้พื้นฐานความรู้ที่เป็นรากฐานอันสำคัญต่อผู้เขียนขณะเป็นนิสิต และให้คำแนะนำอันเป็นประโยชน์ ต่อการสอนและพัฒนาเอกสารทางวิชาการ ในตลอดเวลาที่ผู้เขียนปฏิบัติหน้าที่ในภาควิชาวิศวกรรมไฟฟ้าตั้งแต่ปีพ.ศ. 2554

ขอขอบพระคุณ Professor Lieven Vandenberghе ซึ่งเป็นอาจารย์ที่ปรึกษาของผู้เขียนในการเรียนระดับปริญญาเอก ณ มหาวิทยาลัยแคลิฟอร์เนีย ลอสแอนเจลิส ประเทศสหรัฐอเมริกา อาจารย์ได้จุดประกายให้ผู้เขียนสนใจทำวิจัยด้านเทคนิคการหาค่าเหมาะที่สุดและการเรียนรู้ของเครื่อง ท่านเป็นแบบอย่างที่ดีในด้านการสอนอันมีเนื้อหาที่ทันสมัยจากงานวิจัยเสมอ การมีปรัชญาในการเขียนผลงานในรูปแบบที่ง่ายต่อผู้อ่านเป็นสิ่งสำคัญ สิ่งเหล่านี้ ผู้เขียนได้เรียนรู้และนำมาปรับใช้จนถึงปัจจุบัน

ขอขอบคุณคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้โอกาสผู้เขียนปฏิบัติงานเพิ่มพูนความรู้ทางวิชาการ ตั้งแต่ 1 สิงหาคม พ.ศ. 2563 ถึง 31 กรกฎาคม พ.ศ. 2564 ขอขอบคุณคณาจารย์ในสาขาระบบควบคุมและในภาควิชา ที่ช่วยแบ่งเบาภาระการสอนตลอดระยะเวลาที่ผู้เขียนปฏิบัติงาน เพื่อเพิ่มพูนความรู้ทางวิชาการ ขอขอบคุณ ภาควิชาวิศวกรรมไฟฟ้า ที่สนับสนุนการทำงานบุคลากรอย่างต่อเนื่อง รศ.ดร.แนบบุญ หุ่นเจริญ รศ.ดร.สุพัฒนา เอื้อทวีเกียรติ ผศ.ดร.อภิวัฒน์ เล็กอูทัย และ รศ.ดร.นิศาชล ตั้งเสงี่ยมวิสัย ที่ให้คำแนะนำอันเป็นประโยชน์ต่อการทำวิจัยอันเป็นเนื้อหาที่เกี่ยวข้องในการเขียนหนังสือเล่มนี้ รวมถึงเพื่อนคณาจารย์ในจุฬาลงกรณ์มหาวิทยาลัย ที่แลกเปลี่ยนกำลังใจซึ่งกันและกันในการทำงานตลอดมา ขอขอบคุณนิสิตในที่ปรึกษาระดับป.ตรี ป.โท และป.เอกทุกคน นิสิตในสาขาระบบควบคุมและที่เรียนวิชา 2102531 การหาเอกลักษณ์ของระบบในอดีตจนถึงปัจจุบัน ที่ได้ให้ความเห็นอันเป็นประโยชน์ต่อการสอนและการทำวิจัยร่วมกัน ครูบาอาจารย์ระดับประถมศึกษาจนถึงมหาวิทยาลัยที่ได้อบรมสั่งสอนและเป็นแบบอย่างที่ดีต่อผู้เขียน

ขอขอบคุณกองบรรณาธิการแห่งสำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัยที่ให้คำปรึกษาอย่างดีในการจัดพิมพ์ โดยเฉพาะอย่างยิ่งผู้ทรงคุณวุฒิที่กรุณาอ่านต้นฉบับและให้คำแนะนำที่เป็นประโยชน์ และคุณบุศรินทร์ สรวัดร์ ที่ตรวจพิสูจน์อักษรอย่างถี่ถ้วน

ขอขอบคุณ รศ. ดร.เบญจมาศ พนมรัตน์รักษ์ และ ผศ. ดร.สุดชาย บุญโต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี และ รศ.ดร.พีระยศ แสนโกชณ์ มหาวิทยาลัยเกษตรศาสตร์ ที่ให้คำแนะนำอันเป็นประโยชน์ในการเรียบเรียงเนื้อหา ขอขอบคุณนายณภัทร สามเสาร์ นิสิตที่เรียนวิชาการหาเอกลักษณ์ของ

ระบบ ที่ช่วยตรวจทานแก้ไข สิ่งที่มีผิดพลาดในหนังสือเล่มนี้ และนายปรินทร์ มโนมัยเสาวภาคย์ ที่ได้ช่วยเตรียมรูปผลการทดลองบางส่วนเพื่ออธิบายประกอบเนื้อหา

ผู้เขียนขอขอบคุณโครงการสนับสนุนการเขียนตำรา/หนังสือ/คู่มือ ของคณาจารย์และบุคลากรสายสนับสนุนคณะวิศวกรรมศาสตร์ที่ให้การสนับสนุน และหนังสือ “การหาเอกลักษณ์ของระบบ” เล่มนี้เป็นหนังสือเล่มที่ 32 ในโครงการสนับสนุนการเขียนตำรา/หนังสือ/คู่มือ ของคณาจารย์และบุคลากรสายสนับสนุน คณะวิศวกรรมศาสตร์

ขอขอบคุณแบรนด์ KILTT จากบริษัทคูคิดดีไซน์ #kilttdesign ที่เอื้อเพื่อชุดเฟอร์นิเจอร์ทำงานไม้โอ๊ค ให้ผู้เขียนใช้ทำงานที่บ้าน ตั้งแต่เริ่มมีสถานการณ์โควิด-19 เมื่อเดือนเมษายน พ.ศ. 2562 ทำให้ผู้เขียนมีเครื่องมือที่พร้อมในการเขียนหนังสือและทำงานวิจัยในตลอดเวลาที่ลาปฏิบัติหน้าที่ ขอขอบคุณดีไซน์เนอร์ของ KILTT และคุณอาทิมา อุทัยขลานนท์ ที่ออกแบบกราฟิกของหนังสือเล่มนี้

สุดท้ายนี้ ขอขอบพระคุณ คุณพ่ออภิกิจ สงศิริ คุณแม่ละเมียด สงศิริ คุณพิมพ์ใจ สงศิริ และคุณกิตติภูมิ สงศิริ ที่เป็นกำลังใจซึ่งกันและกันในครอบครัว โดยเฉพาะคุณแม่ที่ช่วยเหลือจัดการธุระในบ้าน น้องสาว น้องชายและภรรยาที่ช่วยเกื้อกูลดูแลความเป็นอยู่และสุขภาพสมาชิกทุกคน ตลอดเวลาสถานการณ์โควิด-19 คุณพ่อที่ได้ช่วยตรวจสอบการเขียนภาษาไทยในหนังสือนี้ ครอบครัวได้เกื้อกูลให้ผู้เขียนได้มีเวลาทุ่มเทกับหนังสือเล่มนี้จนสำเร็จ



จิตโกมุต สงศิริ

ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

พฤษภาคม พ.ศ. 2565

---

# คำนำ

---

**ข้อมูลรอบตัวที่เติบโต.** ในยุคที่ข้อมูลรอบตัวเข้าถึงได้ง่ายและมีปริมาณมากขึ้น การใช้ข้อมูลอันมหาศาลที่สนใจจากงานประยุกต์หนึ่งๆ ด้วยหลักการทางคณิตศาสตร์ จะเป็นเครื่องมือหลักในการตัดสินใจ นโยบายสำคัญของงานประยุกต์นั้น การใช้ข้อมูลสัญญาณเข้า/ออกของระบบพลวัตเพื่อประมาณแบบจำลองระบบ ทำให้ใช้แบบจำลองนั้นในการออกแบบตัวควบคุม การเรียนรู้รูปแบบการใช้พลังงานของยูนิตต่างๆ ร่วมกับการพยากรณ์แหล่งจ่ายพลังงานสะอาด เพื่อจัดสรรกำลังการจ่ายพลังงานจากแหล่งจ่ายต่างๆ ในช่วงเวลาหนึ่งให้มีประสิทธิภาพมากขึ้น การใช้สัญญาณชีพ เช่นสัญญาณสมองของคนไข้ เพื่อวิเคราะห์ภาวะผิดปกติทางสมองเบื้องต้น ก็จะพัฒนาระบบคัดกรองให้มีความเป็นอัตโนมัติมากขึ้น และลดการใช้ทรัพยากรบุคคลไปได้มาก จากตัวอย่างเหล่านี้แสดงให้เห็นว่า ทฤษฎีด้านการประมาณแบบจำลอง การใช้เครื่องมือทางคณิตศาสตร์เพื่อเรียนรู้เชิงสถิติจากข้อมูล กลายเป็นสิ่งจำเป็นในหลายภาคส่วนของงานประยุกต์

**ศาสตร์ที่เกี่ยวข้อง.** การหาเอกลักษณ์ของระบบคือ เทคนิควิธีการประมาณแบบจำลองพลวัตเพื่อนำไปใช้อธิบายระบบกายภาพจริง มีวัตถุประสงค์ร่วมกันกับศาสตร์ด้านการเรียนรู้เชิงสถิติ ในแง่ที่เป็นการศึกษาแบบจำลอง เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตอบสนอง ด้วยการใช้ฟังก์ชันคณิตศาสตร์ ทั้งสองศาสตร์นี้ อาศัยรากฐานเดียวกันคือ ทฤษฎีการประมาณ เพื่อที่จะวิเคราะห์ข้อมูลด้วยหลักการทางสถิติและประมาณพารามิเตอร์ของแบบจำลองนั้น โดยส่วนใหญ่ ปัญหาการประมาณแบบจำลองมักจะอาศัยขั้นตอนเชิงตัวเลขที่ใช้แก้ปัญหาการหาค่าเหมาะที่สุด หนังสือเล่มนี้เน้นที่การใช้แบบจำลองพลวัต เชื่อมโยงกับหลักการเรียนรู้เชิงสถิติในด้านปัญหาถดถอย และให้พื้นฐานที่จำเป็นในด้านทฤษฎีการประมาณ เพื่อให้ผู้อ่านสามารถนำทักษะเหล่านี้ไปสร้างแบบจำลองจากข้อมูลที่ได้รับในงานประยุกต์ที่สนใจ ถึงแม้ว่าเทคนิคการหาค่าเหมาะที่สุดจะเป็นสิ่งที่จำเป็นอย่างยิ่ง แต่หนังสือเล่มนี้จะกล่าวได้พอสังเขปในแง่การใช้เท่านั้น

**กลุ่มผู้อ่านที่คาดหวัง.** หนังสือเล่มนี้สามารถใช้ประกอบการเรียนวิชาการหาเอกลักษณ์ของระบบ (System Identification) อันเป็นวิชาเลือกสำหรับนิสิตระดับปริญญาตรีปี 4 และนิสิตระดับบัณฑิตศึกษา ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เนื้อหาบางส่วนด้านทฤษฎีการประมาณ การถดถอย และการประมาณเชิงสถิติเป็นพื้นฐานสำคัญที่ใช้ในวิชาการอนุมานเชิงสถิติและแบบจำลอง (ภาควิชาวิศวกรรมไฟฟ้า) และวิชาเทคนิคเชิงสถิติสำหรับวิศวกรรมการเงิน (ของหลักสูตรวิศวกรรมการเงิน) เนื้อหาบางประยุกต์ของปัญหาที่ต่อยอดมาจากวิธีกำลังสองน้อยสุดเชิงเส้น ได้มีกล่าวถึงในวิชา การหาค่าเหมาะที่สุดแบบคอนเวกซ์ (ภาควิชาวิศวกรรมไฟฟ้า) หนังสือเล่มนี้ได้เริ่ม

พัฒนาจากเอกสารคำสอนที่ได้ใช้มาตั้งแต่ปีพ.ศ. 2554 ซึ่งอาศัยพื้นฐานด้านพีชคณิตเชิงเส้น เมทริกซ์ ความน่าจะเป็นและสถิติ กระบวนการสุ่ม และระบบเชิงเส้น ผู้อ่านที่มีพื้นฐานเหล่านี้จึงติดตามเนื้อหาได้ไม่ยากนัก ผู้เขียนจัดเนื้อหาหลักการประมาณด้านทฤษฎี บทพิสูจน์ที่สำคัญ เพื่อให้มีความรัดกุมในเชิงคณิตศาสตร์สำหรับวิศวกรรม ส่วนผู้อ่านที่เน้นการใช้งานภาคปฏิบัติ จะมีตัวอย่างเชิงตัวเลขแสดงด้วยโปรแกรม MATLAB เพื่อจำลองข้อมูลด้วยคอมพิวเตอร์ ซึ่งความเชื่อมโยงของหลักการทฤษฎีกับการเรียกใช้งานชุดคำสั่ง และแสดงการแปลผล

**เนื้อหาของหนังสือ** แบ่งเป็น 3 ส่วนหลัก ได้แก่ หนึ่ง การแจกแจงแบบจำลองและการเลือกสัญญาณ กระตุ้นระบบ สอง เทคนิควิธีการประมาณแบบจำลองต่างๆ อันได้แก่ วิธีกำลังสองน้อยสุดเชิงเส้นและปัญหาต่อขยาย วิธีปริภูมิย่อย วิธีความผิดพลาดในการทำนาย การประมาณภาวะน่าจะเป็นสูงสุด และการประมาณภายหลังสูงสุด จากนั้นส่วนสุดท้ายที่สาม จะกล่าวถึง การเลือกและตรวจสอบความสมเหตุสมผลของแบบจำลอง เนื้อหาเหล่านี้ได้นำมาจากหนังสือด้านการหาเอกลักษณ์ของระบบ หรือการเรียนรู้เชิงสถิติหลายเล่ม ตัวอย่างเช่น

- L. Ljung, *System Identification: Theory for the User*, Springer, 1999
- T. Söderström และ P. Stoica, *System Identification*, Prentice Hall, 1989
- T. Hastie and R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2009
- G. James, D. Witten, T. Hastie, และ R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013

เนื้อหาของแบบจำลองในบทที่ 2 จะเสนอกลุ่มแบบจำลองพลวัตเชิงเส้นเวลาวิฤตและประโยชน์เพื่อไปใช้ด้านการทำนายผลตอบ บทที่ 3 จะอธิบายถึงสัญญาณเข้าที่ดีเพื่อวัตถุประสงค์การหาเอกลักษณ์ระบบนั้น ต้องมีสมบัติกระตุ้นระบบที่เพียงพอ ซึ่งการเลือกสัญญาณเข้านี้จะแตกต่างกับวัตถุประสงค์เพื่อวิเคราะห์ผลตอบระบบ หรือเพื่อออกแบบตัวควบคุม ก่อนที่จะอธิบายเรื่องวิธีประมาณแบบจำลองต่างๆ บทที่ 4 จึงปูพื้นฐานด้านทฤษฎีการประมาณให้ทราบสมบัติของตัวประมาณที่ควรรู้ ได้แก่ ความไม่เอนเอียง ความคงเส้นคงวา และประสิทธิภาพ ผลลัพธ์ที่สำคัญที่ใช้กันแพร่หลาย คือสมบัติการแจกแจงเชิงเส้นกำกับแบบปรกติของตัวประมาณเมื่อเรามีข้อมูลมาก ผู้อ่านอาจจะอ่านข้ามบทที่ 4 ก่อน หากต้องการศึกษาเรื่องขั้นตอนการประมาณแบบจำลองโดยตรง แต่การแปลผลเพื่อเปรียบเทียบกับสถิติของแบบจำลอง จำเป็นต้องมีพื้นฐานในบทที่ 4 ที่สามารถศึกษาได้ภายหลัง

บทที่ 5-9 กล่าวถึงรายละเอียดเทคนิคการประมาณต่างๆ เริ่มต้นด้วยวิธีพื้นฐานคือปัญหากำลังสองน้อยสุดเชิงเส้น เพราะในงานวิศวกรรม แบบจำลองเชิงเส้นยังมีการใช้งานอย่างแพร่หลาย บทที่ 5 จึงเริ่มด้วยการจัดรูปแบบปัญหา การหาผลเฉลย สมรรถนะของแบบจำลองถดถอยเชิงเส้น และแง่มุมที่สำคัญในการประยุกต์ใช้งาน การศึกษาในบท 5 นี้อาศัยพีชคณิตเชิงเส้นและการทดสอบสถิติ บทที่ 6 ได้ขยายต่อปัญหากำลังสองน้อยสุดเชิงเส้นที่มีเงื่อนไขบังคับ ฟังก์ชันลงโทษ หรือมีพารามิเตอร์ของรูปแบบปัญหาที่มีความไม่แน่นอน รวมถึงปัญหากำลังสองน้อยสุดไม่เชิงเส้น การอธิบายเนื้อหาในบทนี้ จะเน้นการจัดรูปแบบปัญหาการประมาณ ให้อยู่ในรูปปัญหาการหาค่าเหมาะที่สุดที่ส่วนใหญ่จะเป็นกลุ่มปัญหากำลังสอง (หรือไม่ก็เป็นปัญหาไม่เชิงเส้น) การแก้ปัญหาลักษณะเชิงตัวเลขจึงเป็นการแนะนำชุดคำสั่งใน MATLAB จากนั้น เพื่อความต่อเนื่องของการใช้พื้นฐานของวิธีกำลังสองน้อยสุด จึงเสนอวิธีปริภูมิย่อยในบทที่ 7 ถัดมาอันเป็นวิธีที่ใช้ประมาณแบบจำลองปริภูมิสถานะเชิงเส้นที่ประมาณทั้งตัวแปรสถานะ



และเมทริกซ์ของระบบ จากข้อมูลสัญญาณเข้า/ออกเท่านั้น การหาค่าตอบเชิงตัวเลขใช้วิธีแบบไม่ทำซ้ำ เทคนิคของวิธีนี้อาศัยพื้นฐานด้านพีชคณิตเชิงเส้น การฉายภาพเชิงตั้งฉากและการแยกตัวประกอบของเมทริกซ์ บทที่ 8 เสนอวิธีความผิดพลาดในการทำนาย ที่สามารถประยุกต์ใช้กับแบบจำลองเชิงเส้นในรูปทั่วไป ไม่ว่าจะในรูปฟังก์ชันถ่วงโอน หรือปริภูมิสถานะ หลักการของวิธีนี้คือการใช้ฟังก์ชันสูญเสียที่บ่งชี้ค่าคลาดเคลื่อนของแบบจำลอง พารามิเตอร์ของแบบจำลองจึงถูกเลือกเพื่อให้ฟังก์ชันสูญเสียมีค่าต่ำสุด จัดเป็นปัญหาการหาค่าเหมาะที่สุดและการหาผลเฉลยเชิงตัวเลขจึงใช้ขั้นตอนวิธีทำซ้ำ หลักการทางทฤษฎีของวิธีนี้มีความเป็นทั่วไปสูง โดยเฉพาะในด้านสมบัติเชิงเส้นกำกับของตัวประมาณที่อาศัยพื้นฐานจากบทที่ 4 ผู้เขียนจึงได้ยกตัวอย่างในกรณีเฉพาะเพื่ออธิบายประกอบให้เกิดความเข้าใจได้ง่าย

วิธีประมาณในบทที่ 5-8 อยู่ในกลุ่มที่ไม่ได้อาศัยสมบัติเชิงสถิติของตัวแปรในแบบจำลองในขั้นตอนการประมาณ (แต่เราจำเป็นต้องใช้สมบัติเชิงสถิติในการวิเคราะห์สมรรถนะของแบบจำลอง) บทที่ 9 จะนำเสนอวิธีในกลุ่มที่เรียกว่า การประมาณเชิงสถิติ เพราะว่าแบบจำลองที่สมมติขึ้นจะมีตัวแปร หรือมีพารามิเตอร์ที่สมมติให้มีความไม่แน่นอน การทราบถึงการแจกแจง หรือโมเมนต์บางอันดับของตัวแปรเหล่านั้น จึงนำมาใช้เป็นสมมติฐานส่วนหนึ่งในการจัดรูปแบบปัญหาประมาณ เทคนิควิธีในกลุ่มเหล่านี้คือ การประมาณความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยสุด ภาวะน่าจะเป็นสูงสุด และภายหลังสูงสุด อันเป็นวิธีพื้นฐานในหลายศาสตร์ ทั้งในการเรียนรู้เชิงสถิติ หรือการเรียนรู้ของเครื่อง หลักการของวิธีกลุ่มนี้อาศัยความรู้ทางตัวแปรสุ่มและสถิติ สมบัติเชิงสถิติของค่าประมาณพารามิเตอร์ที่ประมาณได้ขึ้นกับจำนวนตัวอย่างข้อมูลและสมมติฐานที่ใช้ การศึกษาสมบัติเชิงเส้นกำกับของตัวประมาณจะใช้เนื้อหาพื้นฐานทฤษฎีการประมาณจากบทที่ 4

เมื่อผู้อ่านได้เข้าใจหลักการประมาณแบบจำลองหลายวิธีแล้ว บทที่ 10 จึงเสนอประเด็นที่สำคัญเพื่อเลือกแบบจำลองจากหลายแบบในกลุ่มตัวเลือก หลายแบบจำลองนั้นอาจแปรจากกลุ่มโครงสร้างแบบจำลองต่างชนิดกัน หรือแปรจากอันดับของแบบจำลองในโครงสร้างหนึ่งๆ กระบวนการเลือกแบบจำลองอิงกับทฤษฎีสำคัญทางสถิติ คือการค้ำกันระหว่างความเอนเอียงและความแปรปรวนของแบบจำลอง แบบจำลองอย่างง่ายหรือแบบซับซ้อน จะมีทั้งข้อดีและข้อเสียในสองมุมดังกล่าว ผู้เขียนกล่าวถึงกระบวนการเลือก 2 วิธีได้แก่ การใช้เกณฑ์การเลือกแบบจำลองและการตรวจสอบไขว้ อันเป็นวิธีพื้นฐานในศาสตร์ด้านการเรียนรู้เชิงสถิติ เมื่อได้แบบจำลองที่เลือกแล้ว การตรวจสอบแบบจำลองเป็นขั้นตอนสุดท้ายที่จะทดสอบว่าแบบจำลองที่เลือกนั้น นำไปใช้อธิบายข้อมูลชุดใหม่ได้ดีเพียงใด

ผู้เขียนได้คัดเลือกตัวอย่างงานวิจัยของการหาเอกลักษณ์ระบบในด้านการอนุมานพารามิเตอร์ของแบบจำลอง เพื่ออธิบายความสัมพันธ์ของตัวแปรที่มีจำนวนมากในระบบซับซ้อน บทที่ 11 เสนอการใช้แบบจำลองแบบเบาบาง (sparse models) และเทคนิควิธีการประมาณที่ใช้ฟังก์ชันนอร์ม-1 เป็นส่วนหนึ่งของฟังก์ชันสูญเสียของรูปแบบปัญหาประมาณ เนื่องจากการใช้นอร์ม-1 มีสมบัติสนับสนุนให้พารามิเตอร์ของแบบจำลองมีค่าศูนย์เป็นจำนวนมาก งานวิจัยในกลุ่มนี้เป็นที่ใช้แพร่หลายในการเรียนรู้เชิงสถิติและการเรียนรู้ของเครื่องจนถึงปัจจุบัน และมีการพัฒนาอย่างมากในด้านขั้นตอนวิธีเชิงเลขที่ใช้แก้ปัญหาเหมาะที่สุดขนาดใหญ่ ผู้เขียนได้เสนองานประยุกต์ด้านการเรียนรู้ความเชื่อมโยงการทำงาน ของสมองในบริเวณต่างๆ จากหลายรูปแบบปัญหาประมาณด้วยกัน

โครงสร้างของหนังสือจะเริ่มจากการอธิบายหลักการและทฤษฎีบทที่สำคัญ บทพิสูจน์ที่ไม่ยากต่อการทำความเข้าใจจะนำเสนอหลังจากทฤษฎีบท จากนั้นจะแสดงตัวอย่างการประยุกต์เพื่อจัดรูปแบบปัญหา และการคำนวณเชิงตัวเลขที่อิงจากการใช้ชุดคำสั่ง MATLAB ด้วยกล่องเครื่องมือด้านสถิติ และการหาเอกลักษณ์ของระบบ ท้ายบทจะมีแบบฝึกหัดให้ผู้อ่านได้ทบทวนความเข้าใจ บางแบบฝึกหัดจะเป็นการพิสูจน์ผลลัพธ์ของขั้นตอนระหว่างที่อธิบายในหนังสือ บางแบบฝึกหัดเป็นการทดลองของการ

จำลองด้วยคอมพิวเตอร์จากข้อมูลที่ให้ อย่างที่ได้กล่าวไว้ว่าเนื้อหาของหนังสืออาศัยพื้นฐานด้านระบบเชิงเส้น เมทริกซ์ กระบวนการสุ่ม ตัวแปรสุ่ม และสถิติ ผลลัพธ์พื้นฐานเฉพาะที่นำมาใช้ในหนังสือเล่มนี้ จึงได้รวบรวมไว้ในภาคผนวก ผู้อ่านสามารถศึกษาเพิ่มเติมได้จากตำราพื้นฐานทั่วไปที่มีแนะนำไว้ในรายการอ้างอิง

**โปรแกรมและเว็บไซต์.** เนื้อหาที่เกี่ยวข้องกับการคำนวณด้วยคอมพิวเตอร์ จะมีตัวอย่างการใช้ฟังก์ชันใน MATLAB เพื่อให้ผู้อ่านได้ทดลองเขียนโปรแกรมประกอบ โดยการใช้กล่องเครื่องมือ System Identification, Statistics and Machine Learning, และบางส่วนจาก Econometric ผู้อ่านสามารถดาวน์โหลดชุดข้อมูลเพื่อทำการทดลอง สไลด์ประกอบการเรียน และวิดีโอการบรรยายเนื้อหาได้ที่

[http://jitkomut.eng.chula.ac.th/book\\_sysiden.html](http://jitkomut.eng.chula.ac.th/book_sysiden.html)

**สิ่งที่หนังสือนี้ไม่ครอบคลุม.** การหาเอกลักษณ์ของระบบด้วยวิธีไม่อิงแบบจำลอง เช่น การวิเคราะห์สหสัมพันธ์ หรือ ผลตอบสนองเชิงความถี่ การวิเคราะห์ฟูเรียร์หรือสเปกตรัม เทคนิคการเรียนรู้สถิติ ปัญหาด้านการจำแนก หรือการใช้แบบจำลองอื่นในศาสตร์การเรียนรู้ด้วยเครื่อง จะไม่ได้ถูกนำเสนอในตำรานี้ เทคนิคการวิเคราะห์ข้อมูล เช่น การแปลงข้อมูลเป็นโดเมนอื่น การสกัดคุณลักษณะ การลดอันดับของตัวแปร ก่อนนำข้อมูลเข้าไปใช้ในแบบจำลอง ก็เป็นสิ่งสำคัญเช่นเดียวกัน แต่เนื้อหาเหล่านี้สามารถศึกษาได้จากตำราด้านการเรียนรู้เชิงสถิติ หรือการเรียนรู้ของเครื่องเพิ่มเติม ศาสตร์ในด้านการประมาณระบบเกี่ยวข้องอย่างมากกับการใช้วิธีเชิงตัวเลขเพื่อแก้ปัญหาการหาค่าเหมาะที่สุด หนังสือจะกล่าวถึงมุมมองจากขั้นตอนเชิงตัวเลขโดยย่อเท่านั้น นอกจากนี้ เราจะอธิบายวิธีประมาณแบบจำลองพลวัตด้วยชุดข้อมูลโดเมนเวลาเป็นหลัก ถึงแม้ว่าการหาเอกลักษณ์ของระบบจะใช้ข้อมูลโดเมนความถี่ได้เช่นกัน

ผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์ต่อผู้อ่านในด้านพื้นฐานการประมาณแบบจำลอง และการทำวิจัยในงานประยุกต์ต่อไป

จิตโกมุท ส่งศิริ

ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

พฤษภาคม พ.ศ. 2565

---

# สารบัญ

---

กิตติกรรมประกาศ	v
คำนำ	vii
สารบัญ	xi
สารบัญรูป	xiv
สารบัญทฤษฎีบท	xviii
สารบัญรหัสคำสั่ง	xx
สัญลักษณ์และตัวแปร	xxi
อักษรย่อ	xxiv
<b>1 บทนำ</b>	<b>1</b>
1.1 ความสำคัญของการหาเอกลักษณ์ของระบบ . . . . .	2
1.2 องค์ประกอบที่สำคัญ . . . . .	4
1.3 ขั้นตอนการหาเอกลักษณ์ของระบบ . . . . .	7
1.4 ตัวอย่างเบื้องต้น . . . . .	8
1.5 บทสรุป . . . . .	9
<b>2 แบบจำลอง</b>	<b>13</b>
2.1 การจำแนกแบบจำลอง . . . . .	13
2.2 ฟังก์ชันและตัวดำเนินการถ่ายโอน . . . . .	15
2.3 โครงสร้างแบบจำลองทั่วไป . . . . .	16
2.4 การใช้แบบจำลอง . . . . .	17
2.5 ตัวอย่างแบบจำลอง . . . . .	23
2.6 เงื่อนไขของแบบจำลอง . . . . .	32
2.7 รูปแบบของตัวแปรแบบจำลองใน MATLAB . . . . .	32
แบบฝึกหัด . . . . .	36
<b>3 สัญญาณเข้าเพื่อการหาเอกลักษณ์ระบบ</b>	<b>37</b>
3.1 สัญญาณเข้าและคุณสมบัติ . . . . .	39

3.2	ระดับการกระตุ้นคงอยู่ . . . . .	49
3.3	การเลือกสัญญาณเข้าสำหรับการทดลองวงเปิด . . . . .	55
3.4	ตัวอย่างคำสั่ง MATLAB . . . . .	59
	แบบฝึกหัด . . . . .	60
<b>4</b>	<b>พื้นฐานทฤษฎีการประมาณ</b> . . . . .	<b>63</b>
4.1	สมบัติของตัวประมาณ . . . . .	63
4.2	สมบัติของแบบจำลอง . . . . .	66
4.3	ตัวประมาณสุดขีด (Extremum estimator) . . . . .	68
	แบบฝึกหัด . . . . .	72
<b>5</b>	<b>การประมาณกำลังสองน้อยสุดเชิงเส้น</b> . . . . .	<b>75</b>
5.1	รูปแบบปัญหาและตัวอย่างงานประยุกต์ . . . . .	76
5.2	การหาค่าตอบเชิงตัวเลข . . . . .	81
5.3	ปัญหากำลังสองน้อยสุดถ่วงน้ำหนัก . . . . .	83
5.4	สมรรถนะของแบบจำลองถดถอยเชิงเส้น . . . . .	84
5.5	คุณสมบัติเชิงสถิติ . . . . .	85
5.6	การทดสอบนัยสำคัญของตัวประมาณ . . . . .	88
5.7	การเลือกตัวแปรทำนายที่สำคัญต่อตัวแปรตอบสนอง . . . . .	91
5.8	ประเด็นสำคัญเชิงปฏิบัติ . . . . .	94
	แบบฝึกหัด . . . . .	97
<b>6</b>	<b>การขยายปัญหากำลังสองน้อยสุดเชิงเส้น</b> . . . . .	<b>101</b>
6.1	ปัญหากำลังสองน้อยสุดเชิงเส้นที่มีเงื่อนไขบังคับ . . . . .	102
6.2	ปัญหากำลังสองน้อยสุดเชิงเส้นที่มีฟังก์ชันลงโทษ . . . . .	105
6.3	ปัญหากำลังสองน้อยสุดเชิงเส้นเมื่อไม่ทราบค่าพารามิเตอร์ . . . . .	111
6.4	วิธีกำลังสองน้อยสุดไม่เชิงเส้น . . . . .	117
6.5	ตัวอย่างงานประยุกต์ . . . . .	120
6.6	เครื่องมือในการแก้ปัญหาค่าเหมาะที่สุด . . . . .	124
	แบบฝึกหัด . . . . .	129
<b>7</b>	<b>วิธีปริภูมิย่อย</b> . . . . .	<b>135</b>
7.1	ภาพรวมของวิธีปริภูมิย่อย . . . . .	135
7.2	ความสัมพันธ์ของสัญญาณเข้าและสัญญาณออก . . . . .	137
7.3	การกำจัดผลจากสัญญาณเข้า . . . . .	139
7.4	การกำจัดผลจากสัญญาณรบกวน . . . . .	139
7.5	การประมาณ $A$ และ $C$ . . . . .	141
7.6	การประมาณ $B$ และ $D$ . . . . .	144
7.7	วิธีปริภูมิย่อยในมุมมองของปัญหา LS . . . . .	144
7.8	ตัวอย่างคำสั่ง MATLAB . . . . .	149
7.9	ประเด็นต่อยอดและงานประยุกต์ . . . . .	153
	แบบฝึกหัด . . . . .	156
<b>8</b>	<b>วิธีความผิดพลาดในการทำนาย</b> . . . . .	<b>157</b>
8.1	องค์ประกอบของตัวทำนาย . . . . .	157
8.2	การทำนายแบบเหมาะสมที่สุด . . . . .	160
8.3	คุณสมบัติของตัวประมาณ PEM . . . . .	162

8.4	ประเด็นการคำนวณเชิงตัวเลข . . . . .	169
8.5	ตัวอย่างชุดคำสั่งจาก MATLAB . . . . .	171
	แบบฝึกหัด . . . . .	176
<b>9</b>	<b>การประมาณเชิงสถิติ</b>	<b>179</b>
9.1	การประมาณความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด . . . . .	179
9.2	การประมาณภาวะน่าจะเป็นสูงสุด . . . . .	184
9.3	การประมาณภายหลังสูงสุด . . . . .	193
9.4	สมบัติของตัวประมาณ . . . . .	196
	แบบฝึกหัด . . . . .	202
<b>10</b>	<b>การเลือกและตรวจสอบแบบจำลอง</b>	<b>207</b>
10.1	ความเอนเอียงและแปรปรวนของแบบจำลอง . . . . .	209
10.2	การเลือกแบบจำลองเหมาะสมที่สุด . . . . .	214
10.3	การตรวจสอบแบบจำลอง . . . . .	227
10.4	ตัวอย่างคำสั่ง MATLAB . . . . .	232
	แบบฝึกหัด . . . . .	235
<b>11</b>	<b>งานประยุกต์ด้านการเรียนรู้ความสัมพันธ์ของตัวแปรในระบบ</b>	<b>239</b>
11.1	โครงสร้างระบบซับซ้อน . . . . .	239
11.2	ความสัมพันธ์ของตัวแปรนิยามจากอะไร? . . . . .	240
11.3	การประยุกต์ใช้ฟังก์ชันลทอซแบบนอร์ม-1 . . . . .	244
11.4	แบบจำลองเชิงกราฟของตัวแปรเกาส์เซียน . . . . .	244
11.5	การจำลองแบบสมการเชิงโครงสร้าง . . . . .	246
11.6	แบบจำลองเวกเตอร์ถดถอยในตัว . . . . .	248
11.7	กลุ่มแบบจำลองเวกเตอร์ถดถอยในตัว . . . . .	252
11.8	ข้อคิดเห็นและบทสรุป . . . . .	257
	<b>รายการภาคผนวก</b>	<b>261</b>
<b>A</b>	<b>แคลคูลัส</b>	<b>263</b>
A.1	อนุพันธ์ของฟังก์ชันหลายตัวแปร . . . . .	263
A.2	ฟังก์ชันของเมทริกซ์ . . . . .	264
<b>B</b>	<b>เมทริกซ์</b>	<b>267</b>
<b>C</b>	<b>ตัวแปรสุ่มและกระบวนการสุ่ม</b>	<b>273</b>
C.1	การสุ่มของตัวแปรสุ่ม . . . . .	273
C.2	ตัวแปรเกาส์เซียน . . . . .	277
C.3	กระบวนการสุ่ม . . . . .	279
C.4	ความหนาแน่นเชิงสเปกตรัมกำลังของสัญญาณรายคาบ . . . . .	281
	<b>บรรณานุกรม</b>	<b>283</b>
	<b>อภิธานศัพท์</b>	<b>291</b>
	<b>ดัชนี</b>	<b>297</b>
	<b>ประวัติผู้แต่ง</b>	<b>303</b>

---

# สารบัญรูป

---

1.1	ความสัมพันธ์ของสัญญาณเข้ากับสัญญาณออกผ่านระบบที่มีสัญญาณรบกวน . . . . .	1
1.2	ตัวอย่างการใช้แบบจำลองที่ประมาณได้ในการทำนายค่าสัญญาณออก $y$ . . . . .	3
1.3	การหาเอกลักษณ์ของระบบเพื่อนำไปใช้ร่วมกับการออกแบบตัวควบคุม . . . . .	3
1.4	ความรู้เกี่ยวกับสมบัติของสัญญาณรบกวนที่นำไปใช้ร่วมกับการประมาณแบบจำลอง . . . . .	6
1.5	ขั้นตอนการหาเอกลักษณ์ของระบบ . . . . .	8
1.6	ตัวอย่างการประมาณจุดข้อมูลด้วยฟังก์ชันพหุนามอันดับต่างๆ . . . . .	8
1.7	ผลตอบสนอง $y$ ต่อ $u$ ในระบบพลวัต . . . . .	9
1.8	ผลการทดสอบแบบจำลองบนข้อมูลชุดตรวจสอบ สำหรับอันดับแบบจำลองต่างๆ . . . . .	10
1.9	ตำแหน่งของศูนย์ (o) และโพล (x) ในแบบจำลอง ARMAX ที่ประมาณได้ . . . . .	10
2.1	ตัวอย่างผลตอบสนองจากระบบ $y = Gu + He$ โดยกราฟเส้นหนา คือ ผลตอบ $Gu$ และกราฟเส้นประคือผลตอบที่มาจาก $He$ ทั้ง 10 ตัวอย่าง . . . . .	18
2.2	ตัวอย่างการทำนาย 1 ก้าวล่วงหน้า ด้วยการใช้แบบจำลอง ARMAX . . . . .	21
2.3	อนุกรมเวลาที่มีผลของฤดู ค่าเฉลี่ยของกระบวนการนี้เป็นฟังก์ชันเป็นคาบ . . . . .	30
2.4	กระบวนการ AR(1) และกระบวนการผลรวมอันดับต่างๆ . . . . .	30
2.5	อนุกรมเวลาของ log CPI เมื่อผ่านการแปลงหาผลต่าง . . . . .	31
2.6	ระบบผกผัน ระบบต่อเรียง และระบบต่อขนาน . . . . .	35
3.1	ลักษณะของสัญญาณออกของระบบ $G(s)$ ที่มีต่อสัญญาณเข้าชนิดต่างๆ . . . . .	38
3.2	การหาค่าพารามิเตอร์ของแบบจำลองอันดับหนึ่งที่มีค่าประวิงเวลาจากกราฟผลตอบสนองของกระบวนการเมื่อเข้าสู่ภาวะอยู่ตัว . . . . .	40
3.3	สัญญาณไซน์หลายความถี่: $u(t) = \sin(0.4t) + 2\sin(0.8t) + 3\sin(2t)$ สำหรับ $t = 0, 1, 2, \dots, 1000$ . . . . .	43
3.4	สัญญาณรบกวนขาวเกาส์เซียนที่มีความแปรปรวนเป็นหนึ่ง . . . . .	43
3.5	สัญญาณ ARMA ที่สร้างจากกระบวนการ ARMA 1: $G(z) = \frac{1+0.3z^{-1}}{1-0.7z^{-1}-0.2z^{-2}}$ ARMA 2: $G(z) = \frac{1-0.5z^{-1}}{1-1.4z^{-1}+0.8z^{-2}}$ และ ARMA 3: $G(z) = 1 - 1.4z^{-1} + 0.8z^{-2}$ และใช้สัญญาณรบกวนขาวมีความแปรปรวนเท่ากับหนึ่ง . . . . .	44
3.6	แผนผังการสร้างลำดับฐานสองสุ่มเทียม (PRBS) . . . . .	45
3.7	สัญญาณ PRBS ที่ใช้ $n = 5$ มีคาบเท่ากับ $M = 31$ และมีขนาด $\pm 10$ . . . . .	48
3.8	ฟังก์ชันสหสัมพันธ์ของ $y(t)$ เมื่อกระตุ้นด้วยสัญญาณเข้า PRBS และสัญญาณรบกวนขาว ผ่านระบบอันดับหนึ่ง $G(z) = 1/(1 - az^{-1})$ เมื่อ $a = 0.8$ . . . . .	50

3.9	ตัวอย่าง $n = 5$ ที่มี $H_1(z) = 2z^{-1} - 7z^{-2} + 8z^{-3} - 4.5z^{-4} + z^{-5}$ , $H_2(z) = 10z^{-1} - 18.76z^{-2} + 18.97z^{-3} - 7.3z^{-4} + z^{-5}$ , $H_3(z) = z^{-1} - 3.21z^{-2} + 4.53z^{-3} - 3.21z^{-4} + z^{-5}$ ทั้งสามพหุนามนี้มีจำนวนศูนย์ของ $H(z)H(z^{-1})$ ที่อยู่บนวงกลมหนึ่งหน่วยที่ต่างกัน . . . . .	51
3.10	สัญญาณไซน์ความถี่เดียวที่ 11 Hz จำนวน 58 ตัวอย่างที่ให้สมรรถนะประมาณแบบจำลองค่าใกล้เคียงกันที่ 76% แต่ไม่สามารถแยกแยะแบบจำลองแต่ละแบบได้ <i>แฉวบน</i> แผนภาพโบเดของ $\hat{G}$ <i>แฉวกลาง</i> ตำแหน่งศูนย์และสัมประสิทธิ์ของพหุนามเศษของ $\hat{G}$ โดยมีสีที่ต่างกันสำหรับแต่ละค่าศูนย์และสัมประสิทธิ์เศษ <i>แฉวล่าง</i> ตำแหน่งโพลและสัมประสิทธิ์ของพหุนามส่วนของ $\hat{G}$ โดยมีสีที่ต่างกันสำหรับแต่ละค่าโพลและสัมประสิทธิ์ส่วน . . . . .	55
3.11	สัญญาณไซน์หลายความถี่ (multisine) . . . . .	57
3.12	ความไวของแบบจำลองต่อการเปลี่ยนแปลงของพารามิเตอร์และผลการประมาณ . . . . .	58
5.1	ผลเฉลยกำลังสองน้อยสุดทำให้ $X\beta_{ls}$ ใกล้เวกเตอร์ $y$ มากที่สุดในนอร์ม-2 และตั้งฉากกับ $e$ . . . . .	78
5.2	การประมาณฟังก์ชันพหุนามอันดับ 6 เข้ากับชุดข้อมูล . . . . .	79
5.3	การใช้ตัวประมาณกำลังสองน้อยสุดถ่วงน้ำหนักเพื่อหาเส้นตรงที่ประมาณจุดข้อมูล . . . . .	83
5.4	การเลือกเซตย่อยที่ดีที่สุดของตัวแปรทำนายในปัญหาถดถอย . . . . .	93
5.5	การเลือกตัวแปรทำนายแบบเดินหน้าและถอยหลังในปัญหาถดถอย . . . . .	93
5.6	ลักษณะข้อมูลที่ไม่เป็นไปตามสมมติฐานของปัญหาถดถอยเชิงเส้น . . . . .	96
5.7	ค่าความเข้มแสงอาทิตย์และกำลังไฟฟ้าที่ผลิตได้ในวันที่ท้องฟ้าโปร่งใสและมีเมฆมาก . . . . .	98
6.1	ตัวอย่างรูปร่างของเซต $C$ แบบต่างๆ ใน $\mathbf{R}^2$ อันเป็นเซตเงื่อนไขของปัญหา (6.1) และภาพของ $C$ ภายใต้การแปลง $A$ . . . . .	103
6.2	ตัวอย่างรูปแบบผลเฉลยเมื่อมีเทอมลงโทษ $ x $ . . . . .	106
6.3	<i>แฉวบน</i> ฮิสโทแกรมของผลเฉลยจากปัญหา LS ที่ใช้ฟังก์ชันลงโทษนอร์ม-2 และนอร์ม-1 <i>แฉวกลาง</i> ทางเดินของผลเฉลยปัญหา LS ที่ใช้ฟังก์ชันลงโทษนอร์ม-2 และนอร์ม-1 เมื่อแปรค่าพารามิเตอร์ลงโทษ (แต่ละสีคือแต่ละค่า $x_i(\gamma)$ ) <i>แฉวล่าง</i> ปัญหา (6.10) เมื่อมีเงื่อนไขบังคับเป็น $l_2$ -นอร์มบอล และ $l_1$ -นอร์มบอล โดยที่คอนทัวร์ของ $\ Ax - y\ _2^2$ ที่ค่า $(A, y)$ สองชุดแสดงด้วยสีฟ้า และสีม่วง . . . . .	107
6.4	ทางเดินผลเฉลยปัญหาตาข่ายยืดหยุ่นเมื่อใช้ $\alpha = 0.1$ (ซ้าย) และ $\alpha = 1$ คือผลเฉลยแลสโซ (ขวา) . . . . .	109
6.5	(ซ้าย) รูปแบบศูนย์ของผลเฉลยแลสโซกลุ่ม ในแนวตั้งที่ค่า $\gamma$ หนึ่ง จุดสี่เดียวกัน 4 จุดหมายถึง $x_i \in \mathbf{R}^4$ ที่ไม่เท่ากับศูนย์ทั้งบล็อก และมีจำนวน 5 บล็อก เมื่อ $\gamma$ มากขึ้น จำนวนบล็อกใน $x$ ที่เป็นศูนย์มีมากขึ้น (ขวา) ฟังก์ชันลงโทษแบบประกอบนอร์ม-1 กับนอร์ม-2 เมื่อแปรค่า $\gamma$ . . . . .	110
6.6	ทางเดินผลเฉลยปัญหาแลสโซเชื่อม เมื่อ $\gamma_1 = 0$ (ซ้าย) และเมื่อใช้ $\gamma_1 = 2$ (ขวา) . . . . .	111
6.7	สมรรถนะของผลเฉลยต่างๆ ในปัญหา LS เมื่อ $A \in \{A_1, A_2, A_3\}$ . . . . .	112
6.8	นอร์ม-2 ของความคลาดเคลื่อนเลวร้ายที่สุด (6.20) เมื่อค่า $x$ ที่ใช้เป็นค่าต่างๆ . . . . .	115
6.9	ผลเปรียบเทียบนอร์ม-2 ของค่าคลาดเคลื่อนเมื่อใช้ผลเฉลยกำลังสองน้อยสุดทั้ง 3 แบบ . . . . .	117

6.10	ผลการประมาณฟังก์ชันเกาส์เซียนด้วยวิธีกำลังสองน้อยสุดไม่เชิงเส้น โดยเปรียบเทียบข้อมูลวัด (ที่มีสัญญาณรบกวน) และเปรียบเทียบระหว่างฟังก์ชันที่ใช้ค่าพารามิเตอร์จริงกับค่าพารามิเตอร์ที่ประมาณได้ . . . . .	119
6.11	การจำลองแหล่งกำเนิดสัญญาณคลื่นไฟฟ้าสมอง . . . . .	120
7.1	ภาพรวมการหาเอกลักษณ์ของระบบด้วยวิธีปริภูมิย่อย . . . . .	136
7.2	การประมาณแบบจำลองปริภูมิสถานะด้วยวิธี n4sid . . . . .	150
7.3	การประมาณระบบมวล-สปริงด้วยแบบจำลองปริภูมิสถานะที่กำหนดโครงสร้างบางส่วน . . . . .	152
8.1	ผลการประมาณระบบ ARMAX ด้วยการใช้แบบจำลอง ARMA และ ARX . . . . .	172
8.2	ผลการประมาณข้อมูลของมอเตอร์กระแสตรงด้วยแบบจำลองปริภูมิสถานะและ ARX . . . . .	174
8.3	การประมาณแบบจำลอง ARMA ด้วยชุดคำสั่งจากกล่องเครื่องมือ Econometric และ System Identification . . . . .	175
9.1	ตัวอย่างฟังก์ชันแจกแจงความน่าจะเป็นสำหรับพารามิเตอร์ค่าต่างๆ . . . . .	185
9.2	ค่าพารามิเตอร์ที่เลือกในปัญหา การประมาณภาวะน่าจะเป็นสูงสุด Maximum likelihood estimation (MLE) คือค่าที่ทำให้ $\mathcal{L}(\theta)$ มีค่าสูงสุด . . . . .	188
9.3	ฟังก์ชันความหนาแน่นก่อนและภายหลัง . . . . .	194
9.4	ฟังก์ชันความหนาแน่นของตัวแปรเกาส์เซียน (ซ้าย) และลาปลาซ (ขวา) . . . . .	195
10.1	ความยืดหยุ่นและความสามารถในการตีความของแบบจำลอง [HTF09] . . . . .	212
10.2	การประมาณแบบจำลอง AR เมื่อแปรอันดับของแบบจำลอง ค่าฟังก์ชันสูญเสียมีจุดหักที่ประมาณ $p = 5$ . . . . .	212
10.3	ค่าคลาดเคลื่อนของแบบจำลองเมื่อประเมินบนข้อมูลชุดประมาณและชุดทดสอบ . . . . .	213
10.4	ภาวะเข้ากันเกินเมื่อ $f(x)$ มีความไม่เชิงเส้นสูง และ $f(x)$ ใกล้เคียงกับฟังก์ชันเชิงเส้น . . . . .	214
10.5	ผลการเลือกตัวแปรทำนายของแบบจำลองถดถอยเชิงเส้นด้วยเกณฑ์การเลือกแบบจำลองเมื่อใช้ $N = 50$ เกณฑ์ AIC AICc BIC และ $C_p$ เลือกจำนวนตัวแปรทำนายเท่ากับ 15, 4, 4 และ 7 ตามลำดับ Adjusted $R^2$ เลือก 16 ตัวแปร และ $R^2$ เลือกใช้ 20 ตัวแปร . . . . .	220
10.6	ผลการเลือกอันดับของแบบจำลอง AR ด้วยเกณฑ์การเลือกแบบจำลอง . . . . .	221
10.7	การแบ่งข้อมูลสำหรับการตรวจสอบไขว้ . . . . .	225
10.8	การตรวจสอบแบบจำลองพหุนามด้วยวิธี 1) การใช้ข้อมูลตรวจสอบ (ที่มีการทำซ้ำ 7 ครั้ง) 2) การตรวจสอบไขว้แบบ 5 ส่วน (ที่มีการทำซ้ำ 7 ครั้ง) และ 3) การตรวจสอบไขว้แบบ 1 ส่วน . . . . .	226
10.9	การเลือกพารามิเตอร์ลงโทษของปัญหาแลสโซ ในตัวอย่างนี้ที่ใช้จำนวนข้อมูล $N = 200$ วิธี $k$ -fold และ LOOCV เลือกค่า $\gamma$ ที่ใกล้เคียงกัน เกณฑ์ AIC และ AICc เลือกค่า $\gamma$ ที่เท่ากัน ส่วน BIC เลือก $\gamma$ ที่สูงขึ้น ส่งผลให้ผลเฉลยมีศูนย์มากขึ้น . . . . .	227
10.10	ผลการวิเคราะห์ค่าส่วนเหลือด้วยการทดสอบความขาวและฟังก์ชันสหสัมพันธ์ไขว้ . . . . .	231
10.11	ตำแหน่งของโพล (x) และศูนย์ (o) ของแบบจำลองอันดับ 3 และ 6 ด้วยคำสั่ง iopzplot วังวิธีจึงหมายถึงบริเวณความเชื่อมั่นของค่าประมาณโพลและศูนย์ . . . . .	231
11.1	โครงสร้างความสัมพันธ์เชิงกราฟของตัวแปรในระบบขับเคลื่อน . . . . .	240
11.2	ตัวอย่างงานประยุกต์การแสดงความสัมพันธ์ของตัวแปรในระบบที่สนใจ . . . . .	241
11.3	การนำผลประมาณแบบจำลองเพื่ออนุมานโครงสร้างความสัมพันธ์ของตัวแปรในระบบ . . . . .	243



11.4	รูปแบบศูนย์ในอินเวอร์สของเมทริกซ์ความแปรปรวนร่วมของตัวแปรเกาส์เซียน: ค่าที่แท้จริง, ค่าที่คำนวณจากตัวอย่าง, ผลเฉลยของปัญหาแลสโซเชิงกราฟ เมื่อใช้ค่าพารามิเตอร์ลงโทษเท่ากับ $\gamma = 0.02$ และ $\gamma = 0.15$ ตามลำดับ . . . . .	245
11.5	โครงสร้างร่วมของความสัมพันธ์ของตัวแปรในระบบมลภาวะจากสถานีต่างๆ เมื่อตีความความคล้ายเปลี่ยน 3 ระดับ (ค่ามากหมายถึงโครงสร้างดังกล่าวพบมากในหลายสถานี) หมายเหตุ: รูปนี้ได้รับอนุญาตให้พิมพ์ซ้ำจาก Elsevier วารสาร Signal Processing ฉบับที่ 166 ปีค.ศ. 2020 ผู้แต่ง A. Pruttiakaravanich และ J. Songsiri บทความเรื่อง Convex Formulation for Regularized Estimation of Structural Equation Models . . . . .	248
11.6	เครือข่ายสมองของกลุ่มผู้ป่วยออทิสติกและกลุ่มปกติ ลิงก์สีคือการเชื่อมโยงที่พบในทั้ง 2 กลุ่ม <b>ลิงก์สีแดง</b> คือการเชื่อมโยงที่พบเฉพาะ <b>กลุ่มออทิสติก</b> และ <b>ลิงก์สีน้ำเงิน</b> คือการเชื่อมโยงที่พบเฉพาะใน <b>กลุ่มปกติ</b> หมายเหตุ: ได้รับอนุญาตให้พิมพ์ซ้ำจาก Elsevier วารสาร Signal Processing ฉบับที่ 166 ปีค.ศ. 2020 ผู้แต่ง A. Pruttiakaravanich และ J. Songsiri บทความเรื่อง Convex Formulation for Regularized Estimation of Structural Equation Models . . . . .	249
11.7	ตัวอย่างรูปแบบศูนย์ในเมทริกซ์สัมประสิทธิ์ของกระบวนการถดถอยในตัวที่ประมาณจากปัญหา sparse VAR (11.12) . . . . .	250
11.8	ข้อมูลแนวโน้มเคสไข้หวัดของกูเกิล (Google Flu Trends) หมายเหตุ: ได้รับอนุญาตให้พิมพ์ซ้ำจาก IEEE บทความที่ประชุมวิชาการ 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) ปีค.ศ. 2013 ผู้แต่ง Jitkomut Songsiri เรื่อง Sparse autoregressive model estimation for learning Granger causality in time series . . . . .	251
11.9	โครงสร้างเชื่อมโยงการทำงานของสมองที่เรียนรู้จากข้อมูลสมอง 2 กลุ่ม . . . . .	253
11.10	การจัดเรียงพารามิเตอร์ของ $K$ แบบจำลองเพื่อออกแบบฟังก์ชันลงโทษ $g$ . . . . .	253
11.11	ตัวอย่างแบบจำลอง VAR 2 แบบที่ให้โครงสร้างเชิงกราฟแบบเกรนเจอร์เหมือนกัน (CGN) หรือต่างกันบางส่วน (DGN, FGN) เมทริกซ์ขนาด $n \times n$ มีกริด $(i, j)$ ที่แสดงความเข้มของ $\ B_{ij}\ _2$ . . . . .	254
11.12	การเชื่อมโยงการทำงานของสมองเมื่อใช้ปัญหาการประมาณกลุ่มแบบจำลองกับข้อมูล ADHD และเลือกแสดงลิงก์สำคัญจากบริเวณ orbitofrontal และระบบลิมบิก [MS22] . . . . .	256
11.13	ขั้นตอนการเลือกพารามิเตอร์ลงโทษในปัญหาการประมาณแบบเบาบาง: การใช้เกณฑ์การเลือกแบบจำลอง และวิธีตรวจสอบไขว้ . . . . .	258
B.1	วงรี $x^T P^{-1} x = \alpha$ . . . . .	269
B.2	การแยกตัวประกอบ QR และ RQ . . . . .	270
C.1	วงรีความเชื่อมั่นของเวกเตอร์เกาส์เซียน . . . . .	278

---

# สารบัญทฤษฎีบท

---

2.1	ประพจน์ (ระบบเชื่อมโยงกัน) . . . . .	35
3.1	ทฤษฎีบท (บทตั้ง C5.3.1 ของ [SS89]) . . . . .	46
3.2	ทฤษฎีบท (Theorem 13.1 ของ [Lju99]) . . . . .	52
4.1	นิยาม (unbiased estimator) . . . . .	64
4.2	นิยาม (consistent estimator) . . . . .	64
4.3	นิยาม (Efficient estimator) . . . . .	65
4.4	นิยาม (Asymptotic distribution of estimators) . . . . .	65
4.5	นิยาม (System identifiable) . . . . .	67
4.6	นิยาม (Parameter identifiable) . . . . .	68
4.1	ทฤษฎีบท (Consistency of global minimum) . . . . .	69
4.2	ทฤษฎีบท (Consistency of local minimum) . . . . .	69
4.3	ทฤษฎีบท (Limit distribution of local minimum) . . . . .	70
5.1	ทฤษฎีบท (ทฤษฎีบทของเกาส์-มาร์คอฟ หรือ Gauss-Markov Theorem) . . . . .	87
7.1	นิยาม (ตัวแปรเชิงเครื่องมือ) . . . . .	139
9.1	ทฤษฎีบท (ตัวประมาณ MMSE) . . . . .	181
9.2	ทฤษฎีบท (สมบัติของตัวประมาณ MLE) . . . . .	196
9.4	ทฤษฎีบท (อสมการของคราเมอร์-ราว Cramér-Rao inequality) . . . . .	200
10.1	ทฤษฎีบท (Bias-Variance decomposition) . . . . .	209
B.1	ประพจน์ (สมบัติของเมทริกซ์สมมาตร) . . . . .	267
B.1	นิยาม (เมทริกซ์บวกกึ่งแน่นอน) . . . . .	267
B.2	นิยาม (รากที่สองของเมทริกซ์) . . . . .	267
B.3	นิยาม (การเปรียบเทียบ 2 เมทริกซ์บวกกึ่งแน่นอน) . . . . .	268
B.4	นิยาม (ส่วนเติมเต็มชัวร์) . . . . .	268
B.1	บทตั้ง (การหาอินเวอร์สของเมทริกซ์ที่แบ่งเป็นบล็อก) . . . . .	268
B.2	บทตั้ง (สูตรอินเวอร์สของ Sherman-Morrison-Woodbury) . . . . .	269
B.5	นิยาม (วงรีใน $\mathbf{R}^n$ ) . . . . .	269
B.2	ประพจน์ (ลักษณะของวงรี) . . . . .	269

C.1	นิยาม (การลู่เข้าเชิงความน่าจะเป็น) . . . . .	273
C.2	นิยาม (การลู่เข้าเชิงการแจกแจง) . . . . .	273
C.1	ทฤษฎีบท (กฎจำนวนมากอย่างอ่อน (weak law of large numbers (WLLN))	274
C.2	ทฤษฎีบท (กฎของลิมิตเชิงความน่าจะเป็น) . . . . .	275
C.3	ทฤษฎีบท (ทฤษฎีบทการส่งต่อเนื่อง—continuous mapping theorem) . .	275
C.4	ทฤษฎีบท (ทฤษฎีบทของ Slutsky—Slutsky’s Theorem) . . . . .	275
C.5	ทฤษฎีบท (กฎลิมิตผลคูณปกติ-product limit normal rule) . . . . .	276
C.6	ทฤษฎีบท (ทฤษฎีบทลิมิตกลางของ Lindeberg-Levy) . . . . .	276
C.7	ทฤษฎีบท (ทฤษฎีบทลิมิตกลางของ Lindeberg-Feller) . . . . .	276
C.1	ประพจน์ (การแปลงสัมพรรคของเกาส์เซียน—affine mapping of Gaussian)	277
C.2	ประพจน์ (การแจกแจงตามขอบของเกาส์เซียน) . . . . .	277
C.3	ประพจน์ (สหสัมพันธ์กับความน่าจะเป็นอิสระต่อกัน) . . . . .	277
C.4	ประพจน์ (ตัวแปรเกาส์เซียนแบบมีเงื่อนไข) . . . . .	277
C.5	ประพจน์ (เกาส์เซียนมาตรฐาน) . . . . .	278
C.6	ประพจน์ (กำลังสองของเกาส์เซียน) . . . . .	278
C.7	ประพจน์ (วงรีเชื่อม) . . . . .	278
C.3	นิยาม (กระบวนการคงที่เชิงกว้าง) . . . . .	279
C.4	นิยาม (สมบัติกระบวนการคงที่เชิงกว้าง) . . . . .	279
C.5	นิยาม (กระบวนการกึ่งคงที่) . . . . .	279
C.8	ทฤษฎีบท (ทฤษฎีบทของวินเนอร์-คินชิน Wiener-Khinchin Theorem) . . .	280
C.9	ทฤษฎีบท (ความหนาแน่นเชิงสเปกตรัมกำลังของสัญญาณออกจากระบบเชิง เส้น) . . . . .	280
C.10	ทฤษฎีบท (ฟังก์ชันแปรปรวนร่วมของสัญญาณออกจากระบบเชิงเส้น) . . . .	280

---

# สารบัญรหัสคำสั่ง

---

2.1	ตัวอย่างผลการทำนายจากระบบที่กำหนดพารามิเตอร์เองกับระบบที่ประมาณ	21
3.1	การสร้างสัญญาณเข้า . . . . .	59
3.2	การเก็บข้อมูลเพื่อหาเอกลักษณ์ระบบแบบวงเปิด . . . . .	60
5.1	การใช้คำสั่ง fitlm . . . . .	82
6.1	ปัญหา LS ที่มีเงื่อนไขบังคับ . . . . .	125
6.2	ปัญหา LS ที่มีเงื่อนไขเป็นนอร์มบอล . . . . .	126
6.3	ปัญหาแลสโซ . . . . .	126
6.4	ปัญหาตาข่ายยืดหยุ่น . . . . .	126
6.5	ปัญหาแลสโซกลุ่ม . . . . .	127
6.6	ปัญหาแลสโซเชื่อม . . . . .	127
6.7	ปัญหาค่ากำลังสองน้อยสุดแบบเกร็ง . . . . .	127
6.8	การประมาณฟังก์ชันด้วยวิธีกำลังสองน้อยสุดไม่เชิงเส้น . . . . .	128
7.1	การประมาณแบบจำลองปริภูมิสถานะ . . . . .	149
7.2	การประมาณแบบจำลองปริภูมิสถานะที่มีโครงสร้าง . . . . .	151
8.1	การประมาณ ARMAX ด้วยวิธี PEM . . . . .	173
8.2	การประมาณแบบจำลองปริภูมิสถานะด้วยวิธี PEM . . . . .	174
8.3	การประมาณแบบจำลอง ARMA . . . . .	175
10.1	การประมาณแบบจำลองด้วยคำสั่ง fit . . . . .	232
10.2	การประมาณแบบจำลองด้วยคำสั่ง fitlm . . . . .	232
10.3	การเลือกพารามิเตอร์ลงโทษในปัญหาแลสโซ . . . . .	233
10.4	การสร้างข้อมูลอนุกรมเวลาและประมาณแบบจำลอง ARMAX . . . . .	233
10.5	การประมาณแบบจำลอง ARMAX และ ARX และตรวจสอบแบบจำลอง . . . . .	234
10.6	การประมาณแบบจำลอง ARMA ด้วยชุดคำสั่ง Econometric . . . . .	234

# สัญลักษณ์และตัวแปร

สัญลักษณ์	คำอธิบาย
$\mathbf{R}$	เซตจำนวนจริง
$\mathbf{R}^n$	ปริภูมิเวกเตอร์ค่าจริงอันดับ $n$
$\mathbf{R}^{m \times n}$	ปริภูมิเมทริกซ์ค่าจริงอันดับ $m \times n$
$\mathbf{S}^n$	ปริภูมิเมทริกซ์สมมาตรค่าจริงขนาด $n \times n$
$\mathbf{C}$	เซตจำนวนเชิงซ้อน
$i$	$\sqrt{-1}$
$\text{Re}(z)$	ค่าจริงของจำนวนเชิงซ้อน $z$
$\text{Im}(z)$	ค่าจินตภาพของจำนวนเชิงซ้อน $z$
$\mathbf{C}^{m \times n}$	ปริภูมิเมทริกซ์ค่าเชิงซ้อนอันดับ $m \times n$
$\mathbf{S}^n$	เซตของเมทริกซ์สมมาตรขนาด $n \times n$
$\mathcal{N}(A)$	ปริภูมิศูนย์ (nullspace) ของเมทริกซ์ $A$
$\mathcal{R}(A)$	ปริภูมิพิสัย (range space) ของเมทริกซ์ $A$ (หรือปริภูมิคอลัมน์)
$\mathcal{M}^\perp$	ส่วนเติมเต็มตั้งฉากของปริภูมิ $\mathcal{M}$
$\text{tr}(A)$	ผลรวมแนวทแยงมุมของ $A$
$\text{diag}(A)$	สมาชิกแนวทแยงมุมของเมทริกซ์ $A$
$\text{rank}(A)$	ลำดับชั้นของเมทริกซ์ $A$
$\sigma(A)$	ค่าเอกฐานของ $A$
$A^\dagger$	ตัวผกผันเทียม (pseudo inverse) ของ $A$
$A^*$	สังยุคเฮอร์มิเทียนของ $A \in \mathbf{C}^{n \times n}$ (Hermitian conjugate) มีค่าเท่ากับ $A^* = \bar{A}^T$
$\mathbf{E}[X]$	ค่าคาดหวังของตัวแปรสุ่ม $X$
$\text{var}[X]$	ความแปรปรวนของตัวแปรสุ่ม $X$ เมื่อ $X$ เป็นสเกลาร์
$\text{cov}(X)$	เมทริกซ์ความแปรปรวนร่วมของ $X$
$\text{dist}(x, S)$	ระยะระหว่างจุด $x$ ใดๆ ไปยังเซต $S$

หนังสือเล่มนี้ใช้ตัวแปรสำหรับเวกเตอร์และเมทริกซ์เป็นหลัก ดังนั้น  $x \in \mathbf{R}^n$  จะหมายถึงเวกเตอร์ที่มี  $n$  สมาชิก (เราจะไม่ใช่  $\vec{x}$ ) ดังนั้นการแยกแยะระหว่างสเกลาร์กับเวกเตอร์จึงพิจารณาอันดับของ  $x$  ได้ตามบริบท เช่น  $xy^T$  เป็นเมทริกซ์  $n \times n$  ส่วน  $x^T y$  เป็นค่าสเกลาร์ หรือ พจน์  $(I + x^T y)^{-1}$  แท้ที่จริงแล้วเป็นสเกลาร์ จึงมีค่าเท่ากับ  $1/(1 + x^T y)$  สัญลักษณ์  $\mathbf{1}$  คือเวกเตอร์ที่ทุกสมาชิกเท่ากับหนึ่ง นอร์มต่างๆ ของเวกเตอร์  $x$  ใช้สัญลักษณ์เป็น  $\|x\|_1, \|x\|_2, \|x\|_\infty$  หมายถึงนอร์ม-1 นอร์ม-2 และนอร์มนอนันต์ ตามลำดับ เราใช้  $\|x\|$  สำหรับนอร์มใดๆ (ขอให้อ้างอิงตามบริบท) การอ้างถึงสมาชิกตัว

ที่  $i$  ของ  $x$  ใช้  $x_i$  ส่วนสัญลักษณ์  $x \succeq y$  หมายถึง  $x_i \geq y_i$  สำหรับทุก  $i$  ฟังก์ชันของเวกเตอร์ที่ดำเนินการทีละสมาชิกของ  $x$  (elementwise operator) จะใช้สัญลักษณ์ที่คล้ายกับคำสั่ง MATLAB เช่น  $\text{sign}(x)$  ให้ค่าเครื่องหมายของแต่ละสมาชิกใน  $x$  เป็น  $\pm 1$  หรือ  $\max(0, x) = x^+$  คือการให้ค่า  $x_i$  หาก  $x_i \geq 0$  และให้ค่าเป็นศูนย์ถ้าหาก  $x_i < 0$

สัญลักษณ์  $X \in \mathbf{R}^{m \times n}$  หมายถึงเมทริกซ์ค่าจริง และ  $x_{ij}$  จะหมายถึงตำแหน่ง  $(i, j)$  ของ  $X$  เมื่อเราใช้  $X^T$  จะหมายถึงเมทริกซ์สลับเปลี่ยนของ  $X$  (transpose) ในพีชคณิตเมทริกซ์ เรามักจะแบ่ง  $X \in \mathbf{R}^{m \times n}$  ออกเป็นแถวหรือคอลัมน์ ดังนี้

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix}, \quad X = [x_1 \quad x_2 \quad \cdots \quad x_n]$$

สัญลักษณ์  $X \succeq 0$  และ  $X \succ 0$  หมายถึง  $X$  เป็นเมทริกซ์บวกกึ่งแน่นอน (positive semidefinite) และเมทริกซ์บวกแน่นอน (positive definite) ตามลำดับ หาก  $X \not\succeq 0$  จะหมายถึง  $X$  เป็นเมทริกซ์ไม่แน่นอน (indefinite matrix) สำหรับ  $X \succeq 0$  เราใช้  $X^{1/2}$  ว่าเป็นรากที่สองของเมทริกซ์  $X$  (ดูนิยาม B.2) ถ้าหาก  $X \succeq 0$  สัญลักษณ์  $X^{-1/2}$  หมายถึงตัวผกผันของ  $X^{1/2}$

ในการอธิบายฟังก์ชัน  $f: X \rightarrow Y$  เป็นฟังก์ชัน  $f$  ที่แปลงค่าจากสมาชิกในเซต  $X$  ไปยังภาพฉายที่อยู่ในเซต  $Y$  สัญลักษณ์  $\text{dom } f$  คือโดเมนของฟังก์ชัน  $f$

ในแต่ละศาสตร์จะมีแนวปฏิบัติการใช้ตัวแปรที่ต่างกันไป เช่น

- ด้านพีชคณิตเชิงเส้นใช้ตัวพิมพ์เล็ก  $x$  เป็นเวกเตอร์ ตัวพิมพ์ใหญ่  $X$  เป็นเมทริกซ์ หรือบางครั้งเป็น เซต/ปริภูมิ
- ด้านระบบควบคุมใช้  $G, H, F$  แทนฟังก์ชันถ่ายโอน และ  $g, h, f$  แทนผลตอบสนองอิมพัลส์ที่สมนัย
- ด้านความน่าจะเป็น ใช้ตัวแปรพิมพ์ใหญ่  $X$  เพื่อแทนตัวแปรสุ่ม ตัวพิมพ์เล็ก  $x$  เป็นค่าที่เป็นได้ค่าหนึ่งของ  $X$  หากเป็นเรื่องกระบวนการสุ่ม เราเลือกใช้สัญลักษณ์  $x(t)$  แทนกระบวนการสุ่ม (ถึงแม้ว่าหากจะกล่าวให้รัดกุมต้องใช้  $\{x(t)\}_{t \in T}$  อันหมายถึงเซตของ  $x(t)$  ทุกเวลา) โดยทั่วไปจะกำหนดให้  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$  เป็นเวกเตอร์กระบวนการสุ่ม

เนื่องจากตำราเล่มนี้ มีเนื้อหาที่ประยุกต์จากหลายศาสตร์ จึงยังไม่ได้ปรับการใช้สัญลักษณ์ให้เป็นไปตามแนวปฏิบัติของแต่ละศาสตร์ อย่างสม่ำเสมอกับทุกบท เพราะอาจทำให้ผู้อ่านสับสน (เช่น การใช้ตัวแปร  $X$  ที่อาจจะหมายถึงเมทริกซ์ หรือตัวแปรสุ่ม) จึงขอให้ผู้อ่านตีความจากบริบทเป็นสำคัญ

---

## อักษรย่อ

---

AIC	เกณฑ์สารสนเทศของอะกะอิเกะ Akaike Information Criterion.
AR	ถดถอยในตัว Autoregressive.
ARIMA	รวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่ Autoregressive Integrated Moving Average.
ARMA	เฉลี่ยเคลื่อนที่ถดถอยในตัว Autoregressive Moving Average.
ARMAX	เฉลี่ยเคลื่อนที่ถดถอยในตัวที่มีสัญญาณนอกระบบ Autoregressive Moving Average with Exogenous Input.
ARX	ถดถอยในตัวที่มีสัญญาณนอกระบบ Autoregressive with Exogenous Input.
BIC	เกณฑ์สารสนเทศของเบส์ Bayesian Information Criterion.
BLUE	ตัวประมาณค่าไม่เอนเอียงเชิงเส้นที่ดีที่สุด Best Linear Unbiased Estimator.
CLT	ทฤษฎีบทลิมิตกลาง Central Limit Theorem.
CR	อสมการคราเมอร์-ราว Cramér-Rao inequality.
CV	การตรวจสอบไขว้ Cross validation.
CVA	Canonical Variate Analysis.
DCM	การจำลองเชิงเหตุผลพลวัต Dynamic causal modeling.
DGP	กระบวนการก่อกำเนิดข้อมูล Data Generating Process.
EEG	อีอีจี Electroencephalogram.
ESS	ผลบวกกำลังสองที่ถูกอธิบาย Explained Sum of Squares.
FIR	ผลตอบสนองอิมพัลส์อันดับจำกัด Finite Impulse Response.
fMRI	เอฟเอ็มอาร์ไอ Functional magnetic resonance imaging.
i.i.d.	ไอไอดี Independently and Identically Distributed.
KL	การลู่ออกของคัลล์แบก-ลีบเลอร์ Kullback–Leibler divergence.
LMSE	การประมาณค่าเชิงเส้นไม่เอนเอียงที่มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด Linear Unbiased MMSE Estimation.
LS	กำลังสองน้อยสุดเชิงเส้น Linear Least-Squares.

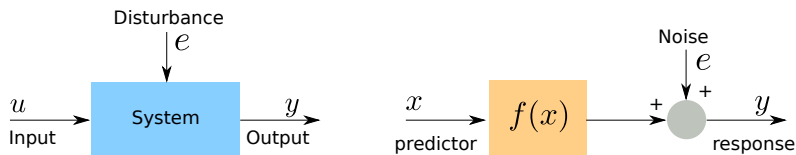
MA	การเฉลี่ยเคลื่อนที่.
MAE	ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย Mean Absolute Error.
MAP	การประมาณภายหลังสูงสุด Maximum a posteriori estimation.
MEG	เอ็มอีจี Magnetoencephalography.
MIMO	หลายสัญญาณเข้าและหลายสัญญาณออก Multiple Input Multiple Output.
MLE	การประมาณภาวะน่าจะเป็นสูงสุด Maximum likelihood estimation.
MMSE	ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด Minimum mean square error.
MSE	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย Mean Square Error.
MVUE	ตัวประมาณไม่เอนเอียงแปรปรวนต่ำสุด Minimum Variance Unbiased Estimator.
N4SID	Numerical Algorithms for Subspace State Space System Identification.
NLS	กำลังสองน้อยสุดไม่เชิงเส้น Nonlinear Least Squares.
OE	ค่าคลาดเคลื่อนที่สัญญาณออก Output Error.
OLS	กำลังสองน้อยสุดสามัญ Ordinary Least-Squares.
PEM	วิธีความผิดพลาดในการทำนาย Prediction Error Method.
PO-MOESP	Past outputs Multivariable Output-Error State-space.
PRBS	ลำดับฐานสองสุ่มเทียม Pseudo Random Binary Sequence.
QP	กำหนดการกำลังสอง Quadratic programming.
RMSE	รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย Root Mean Square Error.
RSE	ความคลาดเคลื่อนส่วนเหลือมาตรฐาน Residual Standard Error.
RSS	ผลบวกกำลังสองของส่วนเหลือ Residual Sum of Squares.
SE	ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ Standard Error of the Coefficient.
SEM	การจำลองแบบสมการเชิงโครงสร้าง Structural equation models.
SISO	สัญญาณเข้าเดี่ยวและสัญญาณออกเดี่ยว Single Input Single Output.
TSS	ผลบวกกำลังสองทั้งหมด Total Sum of Squares.
VAR	เวกเตอร์กระบวนการถดถอยในตัว Vector autoregressive process.
WLLN	กฎจำนวนมากอย่างอ่อน Weak law of large numbers.
WLS	กำลังสองน้อยสุดถ่วงน้ำหนัก Weighted Least-squares.
WSS	คงที่เชิงกว้าง Wide-sense stationary.



# บทที่ 1

## บทนำ

การหาเอกลักษณ์ของระบบ คือการสร้างแบบจำลองเพื่อเป็นตัวแทนของระบบนั้นจากข้อมูลของสัญญาณเข้าและสัญญาณออกที่วัดได้ ด้วยการกระตุ้นสัญญาณเข้าไปยังระบบ จากรูป 1.1 (ซ้าย) แสดงถึงระบบพลวัตทั่วไปที่มีสัญญาณออก (output)  $y$  อันเป็นผลตอบสนองของระบบที่มีต่อสัญญาณเข้า (input)  $u$  และสัญญาณรบกวน (disturbance)  $e$  ผสมกัน เนื่องจากโดยทั่วไป สัญญาณรบกวนเป็นสิ่งที่ไม่ทราบค่า ส่วนสัญญาณเข้า  $u$  นั้นจะเป็นสิ่งที่ผู้ใช้กำหนด หรือ  $u$  สามารถหาค่าได้ รูปแบบปัญหาการหาเอกลักษณ์ของระบบ คือ การมีเซตข้อมูลของ  $(u, y)$  อันจะเรียกว่าข้อมูล I/O ที่พึงตระหนักเสมอว่า  $y$  นั้นจะมีทั้งส่วนที่เป็นผลตอบสนองต่อ  $u$  และผลตอบสนองต่อสัญญาณรบกวน (อันเป็นส่วนไม่พึงประสงค์) ดังนั้น การจะหาแบบจำลอง หรือค่าจำกัดความของระบบ จึงต้องอาศัยทฤษฎีด้านการประมาณแบบจำลอง ตัวอย่างง่าย เช่น ให้ระบบ  $G = 2$  เป็นระบบค่าคงตัว หากเราใส่  $u = 2$  และมีสัญญาณรบกวน  $e = 0.01$  โดยที่สัญญาณรบกวนนั้นเกิดในรูปแบบการบวก (additive noise) กล่าวคือ  $y = Gu + e$  เราจะวัดได้ว่า  $y = 4.01$  ดังนั้นหากเราถือว่า จะหา  $G = y/u$  เราจึงประมาณได้ว่า  $\hat{G} = 2.005$  ซึ่งเกิดความคลาดเคลื่อนของระบบที่ประมาณได้เมื่อเทียบกับระบบจริง อันเป็นผลมาจาก  $e$



รูป 1.1: ความสัมพันธ์ของสัญญาณเข้ากับสัญญาณออกผ่านระบบที่มีสัญญาณรบกวน

รูป 1.1 (ขวา) แสดงถึงการอธิบายระบบในรูปแบบจำลองทางคณิตศาสตร์ในศาสตร์การเรียนรู้เชิงสถิติ สัญญาณออก  $y$  เรียกว่า ตัวแปรตอบสนอง (response) ซึ่งเกิดจากตัวแปรทำนาย (predictor)  $x$  อันส่งผลถึง  $y$  ผ่านทางความสัมพันธ์ที่อธิบายด้วยฟังก์ชัน  $f$  และโดนรบกวนด้วย  $e$  ที่พิจารณาให้เป็นตัวแปรสุ่มอันเป็นลักษณะของสัญญาณรบกวน (noise) การประมาณระบบ จึงหมายถึงการประมาณฟังก์ชัน  $f$  จากชุดข้อมูล  $(x, y)$  หรือ หมายถึงระบบ  $G$  จากชุดข้อมูล  $(u, y)$  นั่นเอง

การหาเอกลักษณ์ของระบบ หรือการประมาณแบบจำลองนั้น มีความเกี่ยวข้องในงานประยุกต์ที่หลากหลายสาขา บทนี้จะกล่าวถึง ความสำคัญของการเข้าใจพื้นฐานหลักการประมาณแบบจำลอง และ

องค์ประกอบ ขั้นตอนที่สำคัญเพื่อบรรลุวัตถุประสงค์ในการหาแบบจำลอง ในตอนท้ายของบทนี้จะเสนอตัวอย่างผลลัพธ์เบื้องต้นเพื่อประกอบการอธิบายขั้นตอนเหล่านั้น

## 1.1 ความสำคัญของการหาเอกลักษณ์ของระบบ

กำหนดให้แบบจำลองทางคณิตศาสตร์ที่แสดงลักษณะของระบบ คือ

$$y = Gu + e \quad (1.1)$$

หรือ

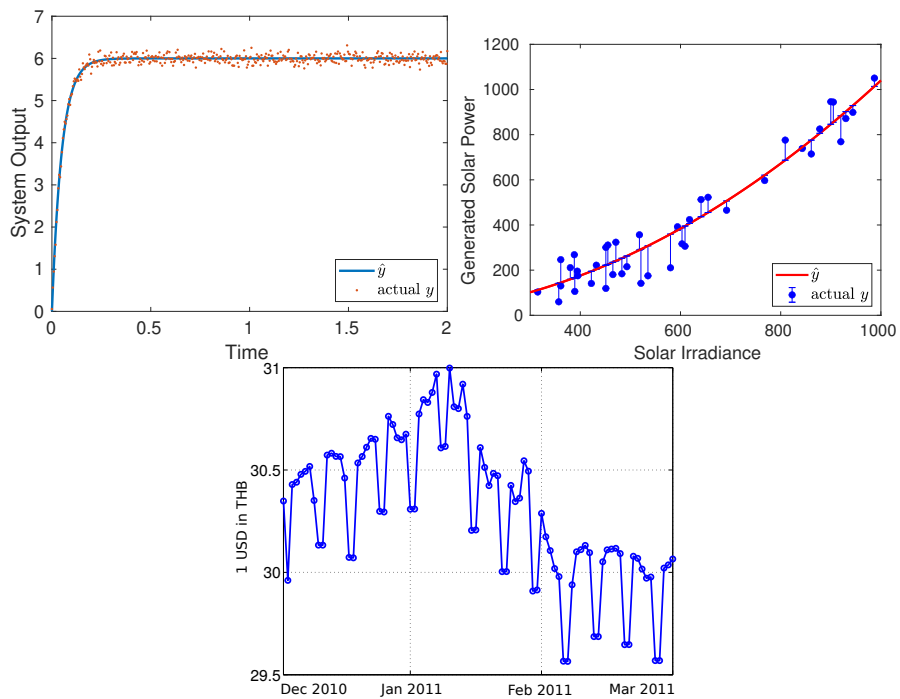
$$y = f(x) + e \quad (1.2)$$

หลังจากการประมาณระบบ นั่นคือ หา  $\hat{G}$  หรือ  $\hat{f}$  ได้ เราสามารถแบ่งผลการนำไปใช้ของการประมาณแบบจำลอง ได้เป็น 3 กลุ่มใหญ่ดังนี้

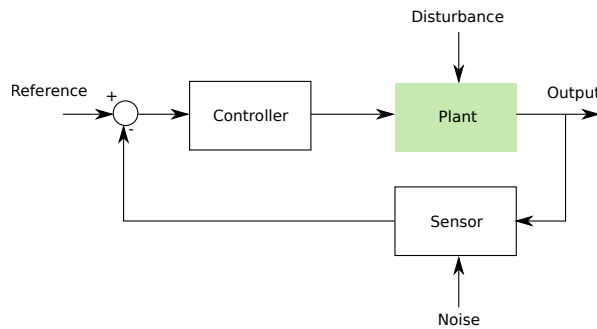
**การทำนาย** ปัญหากลุ่มนี้ คือเราจะมีข้อมูลของ  $u$  หรือ  $x$  แต่เราไม่ทราบค่า  $y$  หรืออาจจะหาได้ยาก ดังนั้น หากเราทราบค่าประมาณของระบบ เราจึงประมาณค่า  $y$  ได้ด้วย  $\hat{y} = \hat{G}u$  หรือ  $\hat{y} = \hat{f}(x)$  ตัวอย่างเช่น หาก  $G$  แสดงถึงระบบมอเตอร์กระแสตรงที่  $y$  คือ ความเร็วเชิงมุมของมอเตอร์ (สมมติให้หน่วยเป็นโวลต์ หรือ V เมื่อแปลงเป็นสัญญาณทางไฟฟ้า) และ  $u$  คือ สัญญาณเข้ามอเตอร์ (V) เมื่อเราประมาณได้ว่า  $\hat{G}(s) = \frac{3}{(0.05s+1)}$  เราก็ประมาณได้ว่า เมื่อใส่  $u(t) = 2$  V สำหรับ  $t \geq 0$  เข้ามอเตอร์ เราจะประมาณความเร็วเชิงมุมของมอเตอร์ในภาวะอยู่ตัวได้เป็น 6 V โดยที่ค่าจริงของ  $y$  ที่วัดได้นั้น อาจจะเบี่ยงเบนจาก 6 จากผลของสัญญาณรบกวนใน (1.1) หรือเราอาจตั้งคำถามว่า เราจะประมาณค่า ความเร็วเชิงมุม ณ วินาทีที่ 0.05 ได้เป็นเท่าใด เราก็สามารถใช้แบบจำลอง  $\hat{G}$  เพื่อประมาณค่า  $\hat{y}(0.05)$  ได้ ดังรูป 1.2 (ซ้าย) ตัวอย่างที่สอง คือความสัมพันธ์ระหว่าง  $y$  ที่เป็นกำลังผลิตไฟฟ้าจากแผงเซลล์สุริยะหนึ่งกับ  $x$  ที่เป็น ความเข้มแสงอาทิตย์ หากเราประมาณความสัมพันธ์นั้นได้เป็น  $y = a_1x + a_2x^2$  โดยที่  $a_1, a_2$  คือค่าคงตัวที่ประมาณได้ เราจึงสามารถประมาณ  $\hat{y}$  ได้ หากเราทราบค่าความเข้มแสงอาทิตย์และอุณหภูมิแผง ณ เวลาหนึ่งของวันที่สนใจ ดังรูป 1.2 (ขวา) ปัญหาการทำนายยังรวมถึงการพยากรณ์ค่า  $y$  ในอนาคต เช่นรูป 1.2 (ล่าง) หากเราสมมติแบบจำลอง เป็น  $\hat{y}_{Apr} = a_1y_{Mar} + a_2y_{Feb}$  และเราประมาณค่า  $a_1, a_2$  ได้ เมื่อทราบค่าอัตราแลกเปลี่ยนเงินบาทในอดีตสองเดือนก่อนหน้า เราจึงสามารถทำนายค่าอัตราแลกเปลี่ยนในเดือนถัดไปได้

**การออกแบบตัวควบคุม** เมื่อพิจารณารูป 1.3 การออกแบบตัวควบคุมโดยทั่วไปนั้น จะอาศัยแบบจำลองคณิตศาสตร์ของระบบ เช่น ฟังก์ชันถ่ายโอน หรือสมการปริภูมิสถานะ เพื่อให้ผู้ออกแบบเข้าใจสมบัติพื้นฐานของระบบ เช่น เสถียรภาพของระบบวงเปิด ตำแหน่งโพล ก่อนที่จะออกแบบตัวควบคุมเพื่อแก้ปัญหาควบคุมใดที่สนใจ ในภาคปฏิบัติ หากเรามีระบบ  $G$  ทางกายภาพ เช่น มอเตอร์กระแสตรง แต่ผู้ใช้ไม่ทราบข้อกำหนดรายละเอียดของฮาร์ดแวร์ที่นำมาประกอบเป็นมอเตอร์นั้น จึงสามารถกำหนดโครงสร้างของแบบจำลอง เช่น การพิจารณาแบบจำลองเชิงเส้นอันดับหนึ่ง และนำข้อมูล  $(u, y)$  มาประมาณ  $G$  เมื่อประมาณได้  $\hat{G}$  แล้ว จึงออกแบบตัวควบคุมที่ใช้ค่าพารามิเตอร์ของ  $\hat{G}$

**การอนุมาน** เมื่อเราใช้แบบจำลองคณิตศาสตร์ในการบรรยายระบบ พารามิเตอร์ของแบบจำลอง หรือฟังก์ชันของพารามิเตอร์นั้น สามารถบ่งชี้บางอย่างทางการเรียนรู้เชิงสถิติ หรือบ่งชี้รูปแบบใดในข้อมูลได้ ตัวอย่างหนึ่งสำหรับ (1.2) เมื่อ  $f$  เป็นฟังก์ชันเชิงเส้น  $y = \beta_1x_1 + \beta_2x_2$  เราสามารถทดสอบ



รูป 1.2: ตัวอย่างการใช้แบบจำลองที่ประมาณได้ในการทำนายค่าสัญญาณออก  $y$



รูป 1.3: การหาเอกลักษณ์ของระบบเพื่อนำไปใช้ร่วมกับการออกแบบตัวควบคุม

สมมติฐานเชิงสถิติว่า  $\beta_1 = 0$  หรือไม่ ซึ่งผลการทดสอบนั้นจะอธิบายว่า ตัวแปรทำนาย  $x_1$  มีความสำคัญต่อการอธิบาย  $y$  หรือไม่ กล่าวคือ รูปแบบของศูนย์ในพารามิเตอร์  $\beta$  ใช้เพื่ออนุมานความสำคัญของตัวแปรทำนาย อันเป็นวิธีพื้นฐานหนึ่งของการคัดเลือกคุณลักษณะ (feature selection) ในปัญหาการเรียนรู้เชิงสถิติ ตัวอย่างที่สอง คือ ระบบพลวัตเชิงเส้นในรูปแบบฟังก์ชันถ่ายโอน (1.1) ที่มีหลายสัญญาณเข้าและหลายสัญญาณออก Multiple Input Multiple Output (MIMO) เช่น

$$G(s) = \begin{bmatrix} G_{11}(s) & G_{12}(s) & \cdots & G_{1m}(s) \\ G_{21}(s) & G_{22}(s) & \cdots & G_{2m}(s) \\ \vdots & \vdots & \ddots & \vdots \\ G_{r1}(s) & G_{r2}(s) & \cdots & G_{rm}(s) \end{bmatrix}$$

หากสัมประสิทธิ์ทั้งหมดของพหุนามเศษในฟังก์ชันถ่ายโอนย่อย  $G_{ij}(s)$  เป็นศูนย์ จึงสรุปได้ว่า ถึงแม้จะใส่สัญญาณเข้า  $u_j \neq 0$  ในระบบนี้ แต่  $y_i$  จะไม่เกิดผลตอบสนองต่อ  $u_j$  แต่อย่างใด กล่าวคือ รูปแบบศูนย์ของฟังก์ชันถ่ายโอนแบบ MIMO จะบอกถึงโครงสร้างความสัมพันธ์ระหว่าง  $u$  กับ  $y$

การนำ  $\hat{G}$  ไปใช้เป็นตัวแทนของ  $G$  จะมีสมรรถนะดีเพียงใด ก็ขึ้นกับความถูกต้องในการประมาณแบบจำลอง ที่จะเกิดจากปัจจัยอันสำคัญ คือ หนึ่ง การใช้ความรู้เกี่ยวกับสัญญาณรบกวน  $e$  ในการประมาณแบบจำลอง สอง การมีสมมติฐานของโครงสร้างแบบจำลองว่าจะใกล้เคียงกับระบบจริงเพียงใด และ สาม เทคนิควิธีในการประมาณแบบจำลองนั้น ในตำรานี้ จะกล่าวถึงรายละเอียดต่อไปสำหรับผลกระทบของปัจจัยทั้งสาม และเทคนิคการประมาณต่างๆ สำหรับระบบเชิงเส้น

## 1.2 องค์ประกอบที่สำคัญ

การประมาณแบบจำลองเกี่ยวข้องกับองค์ประกอบที่สำคัญ 3 ส่วน และมีขั้นตอน ดังนี้

**ส่วนแรก** คือ ข้อมูล  $(u, y)$  สัญญาณเข้า-สัญญาณออก หรือ  $(x, y)$  ตัวแปรทำนาย-ตัวแปรตอบสนอง หากเราประมาณระบบทางกายภาพ เช่น ระบบอุณหภูมิในห้องปิด เราจำเป็นต้องมีตัววัดค่าพลังงานเข้า  $u$  และตัววัดอุณหภูมิ  $y$  ในห้องนั้น การออกแบบสัญญาณเข้าเพื่อไปกระตุ้นระบบให้เกิดสัญญาณออก ก็เป็นหัวข้อสำคัญที่จะกล่าวต่อไป ว่าควรเลือก  $u$  เป็นสัญญาณลักษณะอย่างไร ในภาคปฏิบัติที่ข้อมูลจริงนั้นจะมีความไม่เป็นอุดมคติ เช่น ค่าวัดหายไปในช่วงเวลา ค่า  $y$  คลาดเคลื่อนไปจากความจริง เช่น วัดอุณหภูมิบรรยากาศในกรุงเทพฯ ในหน่วยเซลเซียสออกมาได้เป็นค่าลบ หรือ การวัดสัญญาณไฟฟ้าของคลื่นสมอง (EEG) ก็สามารถคลาดเคลื่อนไปในระดับสูงหากเรากระทบตา หรือขยับกล้ามเนื้อในสัปดาห์นั้น ดังนั้น ขั้นตอนสำคัญก่อนที่จะนำข้อมูลไปใช้ประมาณระบบ มีตัวอย่างเช่น การเติมข้อมูล (missing-value imputation) การจำกัดสิ่งแปลกปน (artifacts) นอกจากนี้ในการพิจารณาแบบจำลองหนึ่งๆ หรือ การใช้เทคนิควิธีเชิงตัวเลขในการประมาณ อาจมีสมมติฐานเกี่ยวกับช่วงของข้อมูล ที่จะส่งผลกับรูปแบบการประมาณแบบจำลอง หรืออาจจะส่งผลกับสมรรถนะของเทคนิควิธีเชิงตัวเลขนั้น ขั้นตอนการจัดการข้อมูลจึงอาจรวมถึง การจัดเรียงข้อมูลให้เข้ากับรูปแบบจำลอง หรือ การแปลงค่าข้อมูลให้อยู่ช่วงมาตรฐาน (normalization) เช่น การกำจัดแนวโน้ม (detrend) และแปลงให้ข้อมูลมีความแปรปรวนเป็นหนึ่ง เป็นต้น ทั้งนี้ ในหลายงานประยุกต์  $x$  อาจจะเป็นปริมาณที่แปลงมาจากตัวแปรต้นตัวอื่น เช่น การจะอธิบายสัญญาณกะพริบของตาจาก EEG เราจะสกัดองค์ประกอบของสัญญาณที่ความถี่ใกล้เคียงกับการกะพริบตา (9-13 ครั้งต่อนาที) ดังนั้น ในที่นี้เราอาจจะพิจารณากำหนดให้  $x$  คือพลังงานของสัญญาณในช่วงความถี่ 0.15-0.22 Hz ซึ่งต้องอาศัยการแปลงเชิงสเปกตรัมจากสัญญาณทางเวลา ให้ไปเป็นสัญญาณเชิงความถี่ก่อน กล่าวโดยสรุปนั้น การจัดการข้อมูลในขั้นต้น (pre-processing) จะเกี่ยวข้องกับ การจัดการความผิดพลาดของข้อมูล การเติมข้อมูล การแปลงข้อมูลให้เป็นเวกเตอร์คุณลักษณะ (feature) ที่สนใจ (หากมี) การจัดเรียงและแปลงข้อมูลในช่วงมาตรฐาน

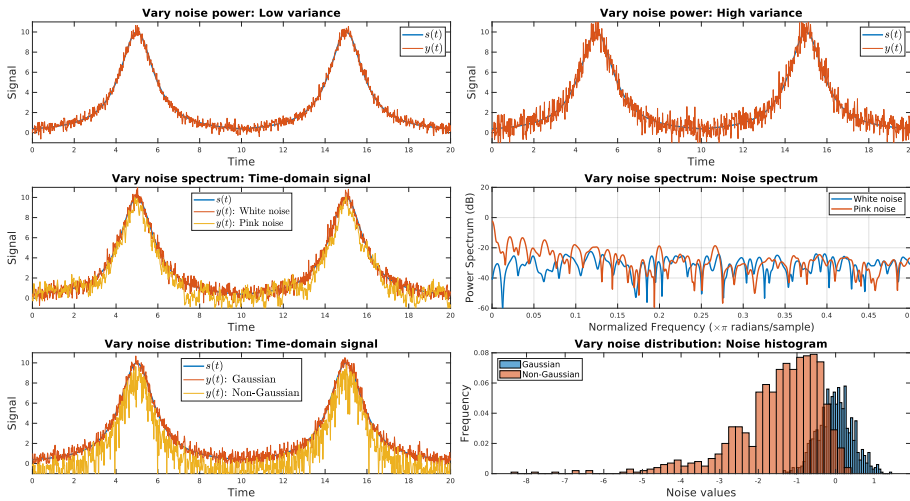
**ส่วนที่สอง** คือ แบบจำลองคณิตศาสตร์ที่เป็นส่วนสำคัญอันส่งผลถึงสมรรถนะของการใช้แบบจำลองอันดับแรก ผู้ออกแบบต้องเลือกโครงสร้างแบบจำลองที่เหมาะสมกับระบบจริงที่ต้องการประมาณ เช่น การอธิบายระบบกายภาพที่ทราบว่ามีคุณสมบัติเชิงเส้นในจุดทำงานหนึ่ง เราเลือกใช้ระบบพลวัตแบบเชิงเส้นที่มีอันดับจำกัดเหมาะสม การเลือกแบบจำลองไม่เชิงเส้นเป็นฟังก์ชันคณิตศาสตร์ที่ได้จากการวิเคราะห์สมการลักษณะเฉพาะของระบบ หรือ การใช้แบบจำลองเชิงความน่าจะเป็น (probabilistic model) สำหรับการอธิบายปรากฏการณ์หนึ่งๆ ของตัวแปรสุ่ม เป็นต้น เราจะเลือกแบบจำลองได้ดี

และเหมาะสม หากเรามีความเข้าใจในตัวแปร  $y$  อย่างถ่องแท้ เพราะการทราบค่า  $y$  มีผลมาจากตัวแปรใด นั่นคือ กำหนดให้ชุดว่าตัวแปรกายภาพใดควรเป็นสัญญาณเข้า (หรือเป็นตัวแปรทำนาย) จะทำให้เราพิจารณาสมการคณิตศาสตร์ที่อธิบายความสัมพันธ์ของ  $(u, y)$  หรือ  $(x, y)$  ได้เหมาะสม ตัวอย่างเช่น เราต้องการหาแบบจำลองเพื่ออธิบายกำลังผลิตไฟฟ้าจากเซลล์สุริยะ เมื่อเราศึกษาหลักการฟิสิกส์ของงานประยุกต์นี้ จะพบว่ากำลังผลิตไฟฟ้าจะขึ้นกับความเข้มแสงอาทิตย์ และอุณหภูมิของแผงเซลล์ แต่เนื่องจากการแผ่รังสีของดวงอาทิตย์เป็นปรากฏการณ์ธรรมชาติ ค่าของความเข้มแสงอาทิตย์จึงขึ้นกับตัวแปรสภาพอากาศหลายตัวแปร เช่น ปริมาณเมฆ (เพราะเมฆจะมาบังแดดได้) ขนาดและทิศทางลม (เพราะลมพัดก้อนเมฆให้ลอยไปจากจุดที่แผงเซลล์ตั้งอยู่) อุณหภูมิ หรือความชื้นสัมพัทธ์ในอากาศ ตัวแปรเหล่านี้จึงกำหนดให้เป็น  $x$  ในการอธิบาย  $y$  ที่เป็นกำลังผลิตไฟฟ้า เป็นต้น นอกจากนี้ การหาแบบจำลองที่เหมาะสมกับชุดข้อมูล ไม่ได้ถูกจำกัดที่ผู้ใช้จะต้องเลือกแค่ประเภทเดียว โดยที่เราสามารถพิจารณาแบบจำลองหลายประเภท เช่น ฟังก์ชันถ่ายโอน ปริภูมิสถานะแบบไม่เชิงเส้น และสามารถทดสอบสมรรถนะของแต่ละแบบจำลอง เพื่อผ่านกระบวนการคัดเลือกแบบจำลอง อันจะได้กล่าวต่อไปในหัวข้อ 10.2

**ส่วนที่สาม** คือ เทคนิคการประมาณแบบจำลอง อันเป็นเนื้อหาหลักของตำรานี้ วิธีการประมาณตั้งอยู่บนหลักการเลือกพารามิเตอร์ ที่ทำให้แบบจำลองมีความเหมาะสมที่สุดด้วยตัวชี้วัดบางอย่าง เช่น วิธีประมาณแบบกำลังสองน้อยสุด จะเลือกค่าพารามิเตอร์ที่ทำให้  $\hat{y} = Gu$  นั้นมีความผิดพลาดจาก  $y$  (ผลตอบระบบจริง) น้อยที่สุด เมื่อวัดความผิดพลาดนั้นด้วยผลรวมกำลังสองของค่าความคลาดเคลื่อนในรูป  $y - \hat{y}$  เทคนิคในการประมาณที่เราสามารถเลือกมาใช้ได้นั้น จะขึ้นกับสมมติฐานที่ใช้ของแบบจำลองหนึ่งๆ ด้วย เช่น หากแบบจำลองที่เราเลือก มีการกำหนดเทอมที่ส่งผลจากสัญญาณรบกวน  $e$  มาয়  $y$  ในลักษณะที่เป็นพลวัต กล่าวคือ  $y$  ขึ้นกับค่า  $e$  ในอดีต ดังนั้น การใช้เทคนิควิธีความผิดพลาดในการทำนาย Prediction Error Method (PEM) หรือ วิธีการหาเอกลักษณ์แบบจำลองปริภูมิสถานะ จึงเป็นตัวเลือกที่ใช้ได้กับแบบจำลองนี้ แต่เราไม่สามารถประยุกต์ใช้วิธีกำลังสองน้อยสุดได้ เป็นต้น ในขั้นตอนการประมาณแบบจำลองนั้น วิธีการประมาณจะแบ่งเป็น 2 กลุ่มใหญ่ คือกลุ่มที่ใช้ข้อมูลของสัญญาณรบกวน  $e$  ในแบบจำลอง กับกลุ่มที่ไม่ต้องใช้สมมติฐานใดเลยจาก  $e$  หากเรามีความรู้เกี่ยวกับ  $e$  มาก เราน่าจะใช้สิ่งนี้ในการช่วยประมาณแบบจำลองได้ดีขึ้น เช่น หาก  $e$  มีความแปรปรวนมาก เราน่าจะเชื่อมั่นได้น้อยลงเกี่ยวกับค่าวัด  $y$  ที่นำมาประมาณแบบจำลอง ความเชื่อมั่นใน  $y$  ที่น้อยลงจะถูกชดเชยอย่างไรในวิธีการประมาณแบบจำลอง ก็จะเป็นรายละเอียดที่จะอภิปรายในหัวข้อ 5.3 และบทที่ 9 สำหรับระดับขั้นความรู้เกี่ยวกับ  $e$  ซึ่งเป็นตัวแปรสุ่ม (หรืออาจจะจำลองเป็นกระบวนการสุ่ม) นั้น เราจะทราบเป็นเพียงแค่บางโมเมนต์ (เช่น ค่าเฉลี่ย ความแปรปรวน) หรือเราจะทราบถึงรูปแบบการแจกแจง (เช่น ฟังก์ชันความหนาแน่นความน่าจะเป็น) ก็จะเป็นปัจจัยที่สอดคล้องต่อหลักการของวิธีประมาณแบบจำลองที่ต่างกันไป

การหาเอกลักษณ์ของระบบเกี่ยวข้องกับทฤษฎีการประมาณ ตัวแปรสุ่ม และกระบวนการสุ่ม ด้วยสาเหตุที่ว่า ข้อมูล  $(u, y)$  ที่เราเก็บมาถูกสมมติว่ามีสัญญาณรบกวนปนมาด้วย จากองค์ประกอบทั้งสามที่กล่าวมาข้างต้นพบว่า ข้อมูลความรู้เกี่ยวกับสัญญาณรบกวนนั้น สามารถนำไปใช้ร่วมกับขั้นตอนการจัดการข้อมูล การเลือกแบบจำลอง หรือ การเลือกเทคนิคการประมาณแบบจำลองได้ ขึ้นกับว่าเราทราบเกี่ยวกับสัญญาณรบกวนนั้นมากเพียงใด หากสมมติว่า  $y(t) = s(t) + e(t)$  โดยกำหนดให้  $s = Gu$  คือสัญญาณจากระบบที่แท้จริง และ  $e$  คือสัญญาณรบกวน และ  $y$  คือสัญญาณออกจากระบบที่เราวัดได้ ในแต่ละแถวของรูป 1.4 แสดงตัวอย่างของการทราบสมบัติเกี่ยวกับ  $e$  ดังนี้

- การทราบกำลังของ  $e$  (คือทราบความแปรปรวน) เป็นตัวอย่างที่พบได้ทั่วไปในทางปฏิบัติ เพราะ



รูป 1.4: ความรู้เกี่ยวกับสมบัติของสัญญาณรบกวนที่นำไปใช้ร่วมกับการประมาณแบบจำลอง

ความรู้เกี่ยวกับบางโมเมนต์ของตัวแปรสุ่ม (เช่น ค่าเฉลี่ย ความแปรปรวน) เป็นสิ่งที่ประมาณได้จากค่าตัวอย่าง ตัวอย่างเช่น หาก  $e$  แทนสัญญาณรบกวนที่มากับเครื่องมือวัด (sensor noise) การวัดสัญญาณพื้นหลังและดูค่า RMS จึงเสมือนเป็นการประมาณค่าเบี่ยงเบนมาตรฐานของ  $e$  ในแถวที่ 1 ของรูป 1.4 แสดงตัวอย่างการวัด  $y$  จาก  $e$  ที่มีความแปรปรวนมากกับน้อย หากการเก็บข้อมูล  $y$  ในแต่ละค่าตัวอย่างนั้น  $e$  มีความแปรปรวนไม่คงที่ (เช่น อาจเกิดจากการเปลี่ยนเครื่องมือวัด) ในหัวข้อ 5.3 จะอธิบายหลักการว่า เราควรนำความแปรปรวนของ  $e$  ที่ไม่เท่ากันนั้นมาชดเชยได้อย่างไร ในขั้นตอนการประมาณแบบจำลองด้วยวิธีกำลังสองน้อยสุด

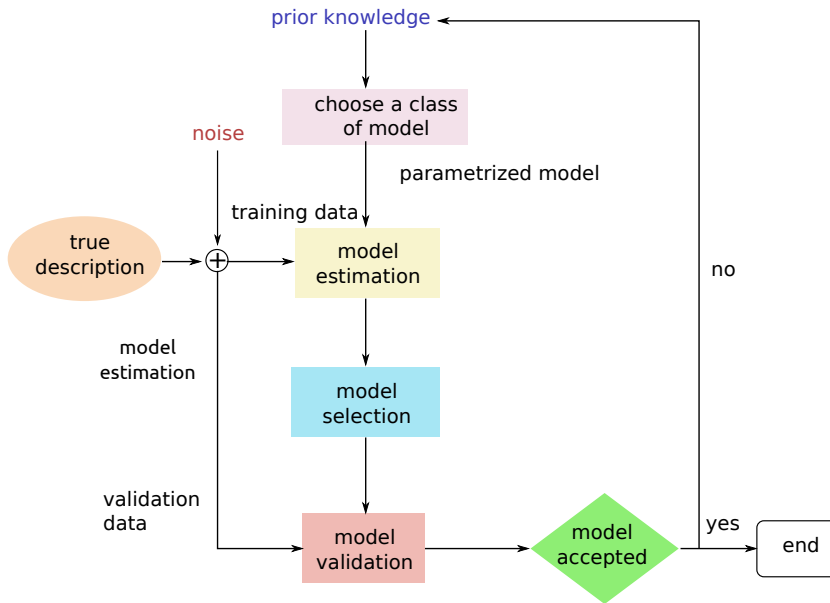
- การทราบช่วงความถี่ของ  $e$  คือคุณสมบัติในโดเมนความถี่ของ  $e$  ว่ามีการแจกแจงพลังงานของ  $e$  อยู่ในช่วงใด ตัวอย่างในแถวสองของรูป 1.4 คือ  $e$  เป็นสัญญาณรบกวนขาวที่มีความหนาแน่นพลังงานในทุกย่านความถี่พอๆ กัน หรือเป็นสัญญาณรบกวนชมพู ที่มีความหนาแน่นสเปกตรัมเป็นส่วนกลับกับความถี่ นั่นคือ มีพลังงานสูงในย่านความถี่ต่ำ คุณสมบัติเชิงความถี่ของ  $e$  ทำให้เราสามารถจัดการข้อมูลได้เหมาะสม เช่น การใช้ตัวกรองเพื่อกำจัด  $e$  ออก ต้องออกแบบย่านความถี่ในตัวกรอง เพื่อไม่ให้บิดเบือน  $s$  ที่ต้องการวัด แต่ต้องกำจัด  $e$  ออกไป
- การทราบการแจกแจงของ  $e$  เช่นแถวสามของรูป 1.4 ว่า  $e$  มีการแจกแจงแบบเกาส์เซียนหรือแบบอื่น เช่น เบต้า แกมมา และเป็นการทราบว่าการแจกแจงนั้นมีพารามิเตอร์อย่างไร เช่น การทราบว่า  $e$  มีการแจกแจงแบบเกาส์เซียนที่มีพารามิเตอร์  $(\mu, \sigma^2)$  เป็น  $(0, 1)$  หรือเป็น  $(2, 3)$  ก็ถือว่าเป็นการแจกแจงที่ต่างกัน สมบัตินี้จะส่งผลโดยตรงต่อ เทคนิคการประมาณแบบจำลองที่เลือกใช้ ดังจะอธิบายในบทที่ 9

สำหรับความรู้เกี่ยวกับ  $e$  นั้น หากทราบการแจกแจงของ  $e$  ก็จะหมายถึง การทราบกำลังของ  $e$  ด้วย (แต่บทกลับไม่จริง) เราจึงคาดเดาได้ว่า การทราบการแจกแจงเป็นความรู้เชิงสถิติที่สมบูรณ์ที่สุดของ  $e$  (เทียบกับการทราบเพียงแค่บางโมเมนต์) ผลการประมาณที่ใช้ความรู้การแจกแจงด้วย น่าจะทำได้ดีกว่า เมื่อเทียบกับการใช้ความรู้บางส่วนของ  $e$  เท่านั้น

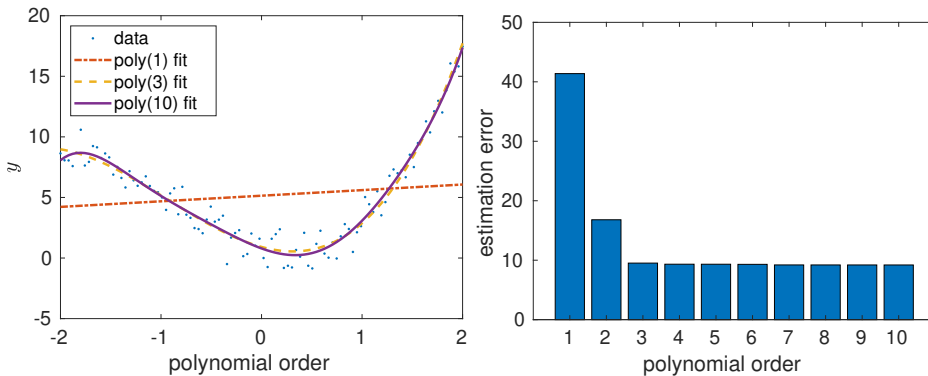
### 1.3 ขั้นตอนการหาเอกลักษณ์ของระบบ

หากกำหนดว่าเรามีข้อมูล  $(u, y)$  หรือ  $(x, y)$  และได้มีการเลือก  $u$  หรือ  $x$  ที่เหมาะสมต่อการทำนาย  $y$  แล้ว ขั้นตอนการหาแบบจำลองที่เหมาะสมกับข้อมูลชุดนี้ สามารถอธิบายได้ด้วยรูป 1.5 ดังนี้

1. การใช้ความรู้เกี่ยวกับงานประยุกต์ หรือระบบกายภาพจริงที่จะประมาณ จะทำให้สามารถเลือกประเภทของแบบจำลองที่เหมาะสม เช่น แบบจำลองเชิงเส้นในรูปแบบฟังก์ชันถ่ายโอน โดยที่อันดับของระบบเป็นสิ่งที่เลือกได้เพื่อป้องกันความยืดหยุ่นของแบบจำลองประเภทนั้น ขั้นตอนนี้มีความสำคัญและส่งผลถึงสมรรถนะของแบบจำลองว่าจะประมาณระบบจริงได้ดีเพียงใด ทั้งนี้โครงสร้างแบบจำลองที่ไม่ตรงกับของระบบจริง ย่อมทำให้เกิดความผิดพลาดในขั้นตอนแรก อันเรียกว่า ความผิดพลาดของแบบจำลอง (model error)
2. ข้อมูล  $y$  ที่วัดจากระบบกายภาพจริงมักถูกรบกวนด้วยสัญญาณรบกวนเสมอ ก่อนจะประมาณแบบจำลอง เราจะแบ่งชุดข้อมูลออกเป็น 2 ชุด อันได้แก่ ข้อมูลชุดฝึกสอน (training) เพื่อใช้ประมาณแบบจำลอง และชุดข้อมูลตรวจสอบ (validation) เพื่อตรวจสอบความสมเหตุสมผลของแบบจำลองต่างๆ ที่ประมาณได้ เราสังเกตว่า ข้อมูล 2 ชุดนี้จะมียิ่งที่เหมือนกันคือ พลวัตหรือคุณลักษณะของระบบจริง แต่สิ่งที่ต่างกันคือ ข้อมูล 2 ชุดถูกรบกวนด้วยสัญญาณรบกวนคนละตัวอย่าง ดังนั้นการใช้ข้อมูลชุดตรวจสอบ เพื่อยืนยันสมรรถนะของแบบจำลองที่ประมาณจากข้อมูลคนละชุด จึงเป็นขั้นตอนหนึ่งที่สำคัญในการออกแบบการทดลอง
3. ในขั้นตอนการประมาณแบบจำลอง เราจะเลือกเทคนิควิธีประมาณพารามิเตอร์ที่เหมาะสม หรือประยุกต์ได้กับโครงสร้างแบบจำลองที่เลือก การแก้ปัญหาค่าประมาณแบบจำลองจะเป็นปัญหาการหาค่าเหมาะที่สุดในรูปแบบหนึ่ง จึงอาศัยวิธีเชิงตัวเลขแบบวนซ้ำ (iterative methods) เพื่อให้ได้มาซึ่งคำตอบของปัญหา เราจะประมาณแบบจำลองที่มีอันดับ หรือมีความซับซ้อนต่างกัน และบันทึกตัวชี้วัดสมรรถนะต่างๆ ไว้
4. ขั้นตอนการเลือกแบบจำลอง คือการพิจารณาตัวเลือกแบบจำลองต่างๆ และเลือกแบบจำลองที่เหมาะสมที่สุด ด้วยเกณฑ์การเลือกแบบจำลอง เกณฑ์นี้ถ่วงน้ำหนักระหว่างความเข้ากันของแบบจำลองและข้อมูล กับระดับความซับซ้อนของแบบจำลอง โดยปกติแล้วแบบจำลองอันดับสูงจะประมาณข้อมูลได้ดี แต่ก็มีพารามิเตอร์ในแบบจำลองมาก ผู้ใช้จึงควรถ่วงดุลให้น้ำหนักสองประเด็นนี้อย่างไร รายละเอียดจะนำเสนอในหัวข้อ 10.2
5. การตรวจสอบความสมเหตุสมผลของแบบจำลอง คือ การนำแบบจำลองที่เลือกนั้นมาทดสอบกับข้อมูลชุดตรวจสอบ ว่ามีคุณลักษณะในแต่ละด้านผ่านเพื่อใช้งานหรือไม่ เช่น แบบจำลองที่ได้มีเสถียรภาพหรือไม่ (หากระบบจริงเป็นระบบเสถียร) แบบจำลองที่ได้อันดับสูงเกินไป จนเกิดการหักล้างระหว่างศูนย์กับโพลหรือไม่ หรือ ความคลาดเคลื่อนส่วนเหลือหลังจากประมาณแบบจำลอง  $\hat{e} = y - \hat{y}$  ควรจะไม่เหลือพลวัตของระบบ  $G$  หากเราใช้  $\hat{G}$  ที่มีโครงสร้างที่ซับซ้อนพอที่จะอธิบาย  $G$  ได้ ดังนั้น  $\hat{e}$  จึงควรมีลักษณะที่ไม่มีสัมพันธ์กับสัญญาณเข้า  $u$  อีก เป็นต้น หากเราตรวจสอบแล้วพบว่า ควรปรับปรุงแบบจำลอง เราจึงควรทบทวนสมมติฐานที่ใช้กำหนดโครงสร้างแบบจำลองเสียใหม่ เพื่อให้เหมาะสมกับระบบที่จะประมาณ



รูป 1.5: ขั้นตอนการหาเอกลักษณ์ของระบบ



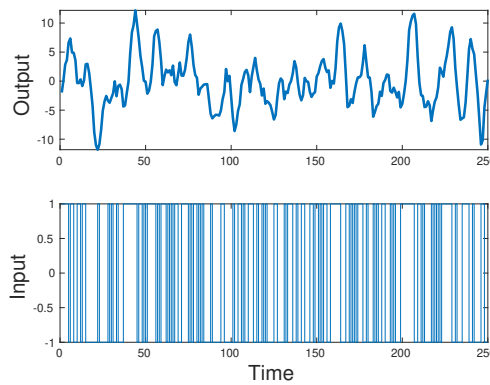
รูป 1.6: ตัวอย่างการประมาณจุดข้อมูลด้วยฟังก์ชันพหุนามอันดับต่างๆ

## 1.4 ตัวอย่างเบื้องต้น

**การปรับเส้นโค้ง (curve fitting).** พิจารณาข้อมูลเป็นคู่ลำดับ  $\{(x_i, y_i)\}_{i=1}^N$  ดังจุดในรูป 1.6 จุดต่างๆ เหล่านั้นเกิดจากฟังก์ชันพหุนามอันดับ 3 แต่ถูกรบกวนจากการวัด เมื่อแก้ปัญหการประมาณจุดข้อมูล (curve fitting) ด้วยการใช้แบบจำลอง  $y = \beta_0 + \beta_1 x + \dots + \beta_n x^n$  สำหรับ  $n = 1, 3, 10$  ได้ผลลัพธ์ คือเมื่อ  $n = 1$  คือการใช้ฟังก์ชันเส้นตรงในการประมาณ จึงพบความเบี่ยงเบนจากกราฟที่ประมาณได้ ออกจากข้อมูลค่อนข้างมาก และเมื่อ  $n = 3, 10$  ผลการประมาณนั้นก็ดีขึ้น แต่เราไม่เห็นความต่างมากของผลการประมาณระหว่างการใช้  $n = 3$  กับ  $n = 10$

หากใช้ نرم-2  $\|y - \hat{y}\|_2$  วัดความคลาดเคลื่อนในการประมาณ เมื่อแปรอันดับของพหุนามที่ใช้ในแบบจำลองดังรูป 1.6 (ขวา) จะเห็นว่าความผิดพลาดนั้นลดลงเมื่อแบบจำลองซับซ้อนขึ้น เพราะความยืดหยุ่นของแบบจำลองที่เพิ่มขึ้นจะเริ่มอธิบายคุณลักษณะของระบบจริงที่อธิบายในข้อมูล เมื่อยิ่งเพิ่ม



รูป 1.7: ผลตอบสนอง  $y$  ต่อ  $u$  ในระบบพลวัต

อันดับของพหุนาม ความซับซ้อนที่มากขึ้นไม่ได้ส่งผลต่อสมรรถนะความเข้ากันของแบบจำลองอย่างมีนัยสำคัญ การเลือกแบบจำลองในที่นี้ คือการเลือกอันดับของพหุนามที่จะเหมาะสม หากค่าผิดพลาดเริ่มไม่ลดลงที่จุดใด ก็จะเลือกอันดับของแบบจำลองที่จุดนั้น (ในที่นี้คือ  $n = 3$ )

**การประมาณแบบจำลองอนุกรมเวลา.** จากชุดข้อมูล  $(u, y)$  ของระบบพลวัตอันดับสองในรูป 1.7 เมื่อ  $u(t)$  เป็นสัญญาณสุ่มที่มีค่า  $\pm 1$  พิจารณาแบบจำลอง ARMAX ที่มีสมการเป็น

$$y(t) + a_1y(t-1) + \dots + a_ny(t-n) = (b_1u(t-1) + \dots + b_nu(t-n) + e(t) + c_1e(t-1) + \dots + c_n e(t-n)) \quad (1.3)$$

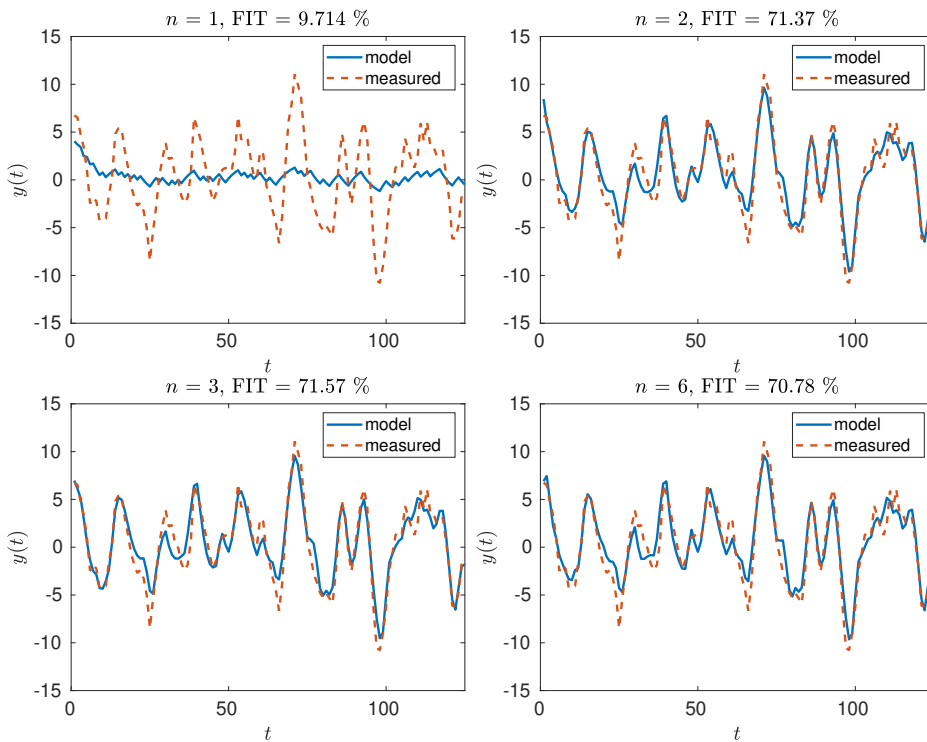
และมีพารามิเตอร์ในแบบจำลองเป็น  $a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n$  ในโครงสร้างนี้  $e$  หมายถึงสัญญาณรบกวนในระบบ และเราสามารถจำลองให้มีฟังก์ชันถ่ายโอนจาก  $e$  ไปยัง  $y$  หากมีสมมติฐานว่าค่า  $e$  ในอดีตจะส่งผลต่อพลวัตของ  $y$  ความซับซ้อนของแบบจำลองจะกำหนดโดย  $n$  และถูกเลือกโดยขั้นตอนการเลือกแบบจำลอง สำหรับรายละเอียดของแบบจำลอง ARMAX จะกล่าวอีกครั้งหนึ่งในหัวข้อ 2.5.2 รวมถึงการประมาณแบบจำลองกลุ่มนี้จะใช้วิธี PEM

แบบจำลองที่ประมาณได้นำมาคำนวณ  $\hat{y}$  และเทียบกับ  $y$  โดยการใช้ข้อมูลชุดตรวจสอบ ดังรูป 1.8 แบบจำลองที่อันดับ  $n = 1$  ยังไม่มีความซับซ้อนมากพอที่จะอธิบายพลวัตทั้งหมดของระบบ แต่เมื่อเริ่มเพิ่มอันดับให้สูงขึ้น ค่าชี้วัดของการเข้ากันได้ระหว่างแบบจำลองและข้อมูล (FIT) มีค่าสูงขึ้น แต่ไม่ได้เพิ่มขึ้นอย่างมีนัยเมื่อ  $n = 3$  และกลับมีค่าลดลงเล็กน้อยเมื่อ  $n = 6$  ในขั้นตอนการเลือกแบบจำลอง จึงควรตัดสินใจจากเกณฑ์การเลือกแบบจำลองว่า อันดับของแบบจำลองที่สูงขึ้นนั้น ได้ค่าชี้วัดคุณภาพแบบจำลองที่สูงขึ้นมาเล็กน้อย แต่จะคุ้มค่ากับการที่มีพารามิเตอร์มากขึ้นหรือไม่

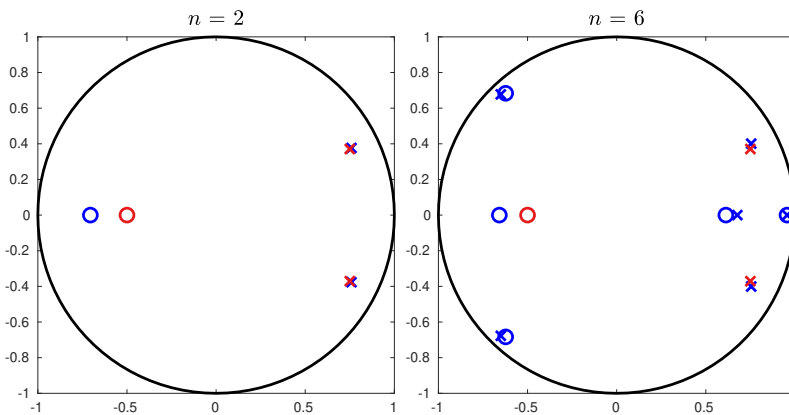
พิจารณาตำแหน่งศูนย์และโพลของแบบจำลองที่ประมาณได้ในรูป 1.9 เมื่อสีแดงคือระบบจริง และสีน้ำเงินคือระบบที่ประมาณได้ เมื่อใช้แบบจำลองอันดับสูง จะมีแนวโน้มของการหักล้างระหว่างศูนย์และโพลในบางคู่ อันบ่งชี้ว่าแบบจำลองอันดับ 6 ไม่น่าจะเหมาะสมในการอธิบายระบบนี้

## 1.5 บทสรุป

บทนี้ได้กล่าวถึงคำจำกัดความของการหาเอกลักษณ์ของระบบ และชี้ให้เห็นว่าการประมาณระบบนั้นไม่ได้จำกัดที่แค่ระบบพลวัต แต่ยังรวมถึงแบบจำลองเชิงสถิติใดๆ องค์ประกอบที่สำคัญของการหา



รูป 1.8: ผลการทดสอบแบบจำลองบนข้อมูลชุดตรวจสอบ สำหรับอันดับแบบจำลองต่างๆ



รูป 1.9: ตำแหน่งของศูนย์ (o) และโพล (x) ในแบบจำลอง ARMAX ที่ประมาณได้

เอกลักษณ์ประกอบไปด้วย ข้อมูลระหว่างสัญญาณเข้า-สัญญาณออก โครงสร้างแบบจำลอง เทคนิควิธีประมาณและวิธีการเลือกแบบจำลอง โดยแสดงตัวอย่างเบื้องต้นของการประมาณชุดข้อมูลด้วยฟังก์ชันพหุนาม และการประมาณแบบจำลองพลวัต ARMAX

ในการทำความเข้าใจเนื้อหาในบทถัดไป จึงเริ่มจากการทำความเข้าใจในโครงสร้างของแบบจำลองประเภทต่างๆ ในบทที่ 2 ที่ผู้อ่านควรมีพื้นฐานของระบบเชิงเส้น จากนั้นในบทที่ 3 จะกล่าวถึงการเลือกสัญญาณที่จะมากระตุ้นระบบกายภาพ เพื่อให้มีชุดสัญญาณเข้าและสัญญาณออกที่เหมาะสมหรือเอื้ออำนวยต่อการประมาณคุณลักษณะที่สำคัญของระบบ บทที่ 4 จะอธิบายพื้นฐานนิยามที่สำคัญของ

---

## บทที่ 2

# แบบจำลอง

---

ในบริบทของการหาเอกลักษณ์ของระบบนั้น แบบจำลองคือตัวแทนของระบบกายภาพจริงที่แสดงความสัมพันธ์ระหว่างสัญญาณเข้าและสัญญาณออกอันบรรยายด้วย  $y = Gu$  ในที่นี้  $G$  จะแสดงถึงระบบ และเราสามารถเลือกใช้แบบจำลองต่างๆ มาอธิบาย  $G$  ได้ อย่างไรก็ตาม ระบบที่สนใจอาจจะเป็นระบบอิสระ (autonomous system) อันหมายถึง ระบบที่ไม่มีสัญญาณเข้า  $u$  มากะดุนก็เป็นได้ หรือในอีกกรณี ที่หมายถึงแบบจำลองแสดงความสัมพันธ์ระหว่างสัญญาณออกและตัวแปรที่มากะดุน จนทำให้เกิดพลวัตใน  $y$

ในบทนี้ จะกล่าวถึงแบบจำลองชนิดต่างๆ ที่แยกประเภทด้วยเกณฑ์ที่ต่างกัน การเข้าใจประเภทและรายละเอียดของแบบจำลอง จะทำให้ทราบโครงสร้างแบบจำลองนั้น และนำไปสู่การเลือกใช้แบบจำลองที่เหมาะสม อันเป็นสมมติฐานที่สำคัญของการเรียนรู้เชิงสถิติต่อไป

### 2.1 การจำแนกแบบจำลอง

ประเภทของแบบจำลองนั้นสามารถจำแนกได้ด้วยสมบัติดังต่อไปนี้

- จำนวนสัญญาณเข้าและสัญญาณออกของระบบ
  - แบบจำลองสัญญาณเข้าเดี่ยวและสัญญาณออกเดี่ยว Single Input Single Output (SISO) มี  $y$  และ  $u$  เป็นสเกลาร์
  - แบบจำลอง MIMO มี  $y$  และ  $u$  เป็นเวกเตอร์
- ความเป็นเชิงเส้นของระบบ: หากพิจารณาความสัมพันธ์ของ  $y$  และ  $u$  ในรูปทั่วไปคือ  $y = f(u)$  เราสามารถแบ่งแบบจำลองได้เป็น
  - แบบจำลองเชิงเส้น คือ แบบจำลองที่ความสัมพันธ์  $f$  นั้น สอดคล้องกับหลักการซ้อนทับ (superposition) หรือ เงื่อนไขการแปลงเชิงเส้น นั่นคือ  $f(u_1 + \alpha u_2) = f(u_1) + \alpha f(u_2)$  สำหรับค่าคงที่  $\alpha$  ใดๆ และ  $u_1, u_2$  ใดๆ ตัวอย่างแบบจำลองเชิงเส้น ได้แก่ แบบจำลองปริภูมิสถานะเชิงเส้น ฟังก์ชันถ่ายโอน
  - แบบจำลองไม่เชิงเส้น คือ แบบจำลองที่  $f$  เป็นการแปลงไม่เชิงเส้น ตัวอย่างเช่น แบบจำลองปริภูมิสถานะทั่วไป บรรยายได้ด้วย  $\dot{x} = f(x, u)$  และ  $y = g(x, u)$  โดยที่  $f, g$  ต่างก็เป็นฟังก์ชันไม่เชิงเส้น

### 3. การที่แบบจำลองมีโครงสร้างที่อิงพารามิเตอร์

- แบบจำลองอิงพารามิเตอร์ (parametric model) มีโครงสร้างและบรรยายได้ด้วยเทอมทางคณิตศาสตร์ และแบบจำลองนั้นสามารถแจกแจงได้ด้วยค่าพารามิเตอร์ต่างๆ ตัวอย่างเช่น  $G = \frac{K(s+a)}{(s+b_1)(s+b_2)}$  ถูกแจกแจง  $G$  ได้หลายแบบจำลอง ด้วยการแปรค่า  $K, a, b_1, b_2$  ที่ต่างกัน ในที่นี้ เราจะเรียก  $K, a, b_1, b_2$  ว่าเป็นพารามิเตอร์ของแบบจำลอง  $G$
- แบบจำลองไม่อิงพารามิเตอร์ (nonparametric model) จะไม่มีโครงสร้างที่เราสามารถแจกแจงพารามิเตอร์ได้ เช่น แผนภาพโบเด (bode plot) เป็นแบบจำลองชนิดหนึ่ง เพราะสามารถอธิบายความสัมพันธ์ระหว่าง  $y$  และ  $u$  ในแง่ของอัตราขยายและเฟสที่แปรต่างกันไปในแต่ละความถี่ แผนภาพโบเดได้มาจากการประมาณผลตอบสนองเชิงความถี่ที่การประมาณนั้นไม่จำเป็นต้องอิงบนโครงสร้างที่แจกแจงพารามิเตอร์แต่อย่างใด

ข้อดีของแบบจำลองอิงพารามิเตอร์ (ที่จะใช้เป็นหลักในหนังสือเล่มนี้) คือความประหยัดในการบรรยายแบบจำลองเพราะสามารถแจกแจงผ่านทางพารามิเตอร์ที่มีจำนวนจำกัดได้ (กรณีระบบมิติจำกัด) เช่น ผลตอบสนองอิมพัลส์ของระบบ  $\{h(k)\}_{k=0}^{\infty}$  ถือเป็นแบบจำลองไม่อิงพารามิเตอร์ และสามารถบรรยายพลวัตของระบบได้ แต่ถ้าทราบพารามิเตอร์ของระบบ เช่น เมทริกซ์ในแบบจำลองปริภูมิสถานะ  $(A, B, C, D)$  ก็สามารถบรรยาย  $h(k)$  ได้เช่นเดียวกันผ่านทางเมทริกซ์ระบบนี้ เช่น สำหรับระบบเวลาวิฤต  $h(k) = D$  เมื่อ  $k = 0$  และ  $h(k) = CA^{k-1}B$  เมื่อ  $k > 0$  อย่างไรก็ตาม การใช้แบบจำลองอิงพารามิเตอร์ควรมีการเลือกโครงสร้างแบบจำลองที่เหมาะสมเสียก่อน และต้องใช้หลักการประมาณแบบจำลอง ที่มักจะซับซ้อนกว่าการประมาณแบบจำลองไม่อิงพารามิเตอร์

### 4. สมบัติแปรเปลี่ยนตามเวลา

- แบบจำลองไม่แปรผันตามเวลา (time-invariant) คือแบบจำลองของระบบที่มีสมบัติไม่แปรผันตามเวลา ตามนิยามหมายถึง เมื่อสัญญาณเข้าที่ถูกประวิงเวลาไป  $\tau$  ผ่านระบบนั้น จะทำให้เกิดสัญญาณออกที่ประวิงเวลาไป  $\tau$  เช่นกัน ตัวอย่างเช่น แบบจำลองปริภูมิสถานะ  $\dot{x}(t) = A(t)x(t) + B(t)u(t)$  และ  $y(t) = C(t)x(t) + D(t)u(t)$  ที่มี  $(A, B, C, D)$  เป็นเมทริกซ์คงที่ และเป็นที่ยอมรับกันว่าเมทริกซ์เปลี่ยนสถานะ (state-transition matrix) ของระบบเชิงเส้นไม่แปรผันตามเวลา จะขึ้นกับระยะเวลาของเวลาเริ่มต้น  $t_0$  กับเวลาที่สนใจ  $t$  เท่านั้น แต่ไม่ใช่ฟังก์ชันโดยตรงของ  $t$
- แบบจำลองแปรผันตามเวลา (time-varying) คือแบบจำลองที่แสดงถึงระบบที่มีสมบัติที่แปรเปลี่ยนตามเวลา เช่น แบบจำลองปริภูมิสถานะ  $\dot{x}(t) = A(t)x(t) + B(t)u(t)$  และ  $y(t) = C(t)x(t) + D(t)u(t)$  เราจะเห็นว่าเมทริกซ์พลวัต  $A(t)$  (และเมทริกซ์อื่นๆ) นั้นมีค่าแปรตามเวลา  $t$  อย่างน้อยเราเห็นได้ชัดว่า เมทริกซ์แปลงสถานะ (state-transition matrix) จะขึ้นกับเวลาที่เริ่มและสิ้นสุดที่สนใจในการพิจารณาผลตอบ  $y(t)$

### 5. โดเมนของระบบ

- แบบจำลองเชิงเวลา (time-domain model) คือแบบจำลองที่อธิบายพลวัตของระบบในโดเมนเวลา เช่น แบบจำลองปริภูมิสถานะ ทั้งเวลาต่อเนื่องและเวลาวิฤต
- แบบจำลองเชิงความถี่ (frequency-domain model) คือแบบจำลองที่อธิบายพลวัตของระบบในโดเมนความถี่ เช่น ฟังก์ชันถ่ายโอน  $G(s) = K(\tau s + 1)$  จะทำให้เราทราบว่าอัตราขยายในย่านความถี่ต่ำจะประมาณได้เท่ากับ  $K$  และมีค่าลดลงเมื่อความถี่สูงเกิน  $1/\tau$  เรเดียน

---

## บทที่ 3

# สัญญาณเข้าเพื่อการศึกษาเอกลักษณ์ระบบ

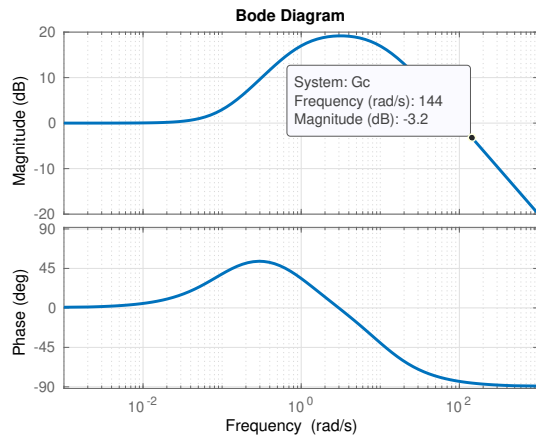
---

ความสามารถในการระบุแบบจำลอง (model identifiability) นั้น เป็นสมบัติสำคัญร่วมระหว่าง การออกแบบการทดลองหาเอกลักษณ์ของระบบ และแบบจำลองที่เลือก สมบัตินี้จะชี้ว่าพารามิเตอร์ ของแบบจำลองจะสามารถประมาณได้อย่างเหมาะสมหรือไม่ ด้วยการเลือกโครงสร้างแบบจำลองและ ด้วยการเลือกชุดข้อมูล  $(y, u)$  บทนี้เสนอปัจจัยที่สำคัญ คือ การเลือกสัญญาณเข้า  $u$  ที่ใช้กระตุ้นระบบ ในการผลิต  $y$  เพื่อที่จะทำให้ชุดข้อมูล  $(y, u)$  เป็นชุดสัญญาณที่เหมาะสมในการประมาณแบบจำลอง

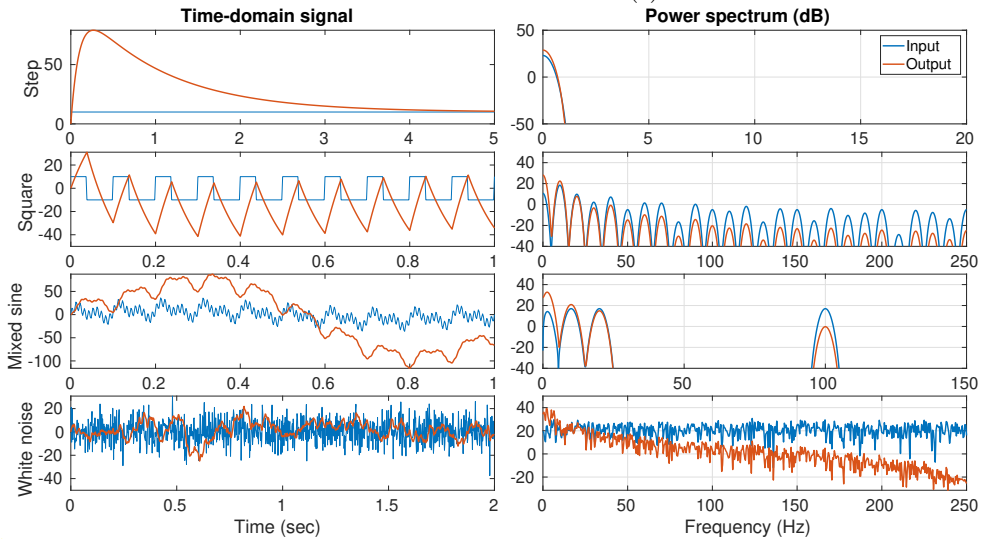
การประมาณฟังก์ชันถ่ายโอนของระบบ  $G(z)$  แบบไม่อิงพารามิเตอร์บนโดเมนความถี่ ทำได้โดย การหาผลการแปลงฟูเรียร์ของ  $(y(t), u(t))$  กำหนดให้เป็น  $(Y(\omega), U(\omega))$  และค่าประมาณของ  $G$  ที่ความถี่หนึ่งๆ นั้นจะเป็น  $\hat{G}(e^{j\omega}) = Y(\omega)/U(\omega)$  (ศึกษาเพิ่มเติมได้ในกรณีวิเคราะห์ฟูเรียร์ การ ประมาณฟังก์ชันถ่ายโอนจากการทดลอง หรือ empirical transfer-function estimation ใน [Lju99, §6]) จากหลักการนี้เราเห็นว่า การประมาณ  $G$  ได้ครบทุกองค์ประกอบความถี่ที่มีในระบบ จึงควรใช้ สัญญาณเข้าที่สามารถไปกระตุ้นระบบได้หลายความถี่เช่นกัน

ตัวอย่างคือ ระบบอันดับสอง  $G(s) = 100(s + 0.1)/(s + 1)(s + 10)$  ที่มีผลตอบสนองเชิง ความถี่เป็นดังรูป 3.1a โดยมีแบนด์วิดท์ที่ 144 rad/s หรือ 22.9 Hz เราทดสอบผลตอบสนองของ ระบบที่มี  $u$ สี่ชนิด อันได้แก่ 1) สัญญาณขั้น 2) สัญญาณคลื่นสี่เหลี่ยมความถี่ 10 Hz 3) ผลรวมของ สัญญาณไซน์ที่มีความถี่ 1, 10, 20 และ 100 Hz และ 4) สัญญาณรบกวนขาวที่มีความแปรปรวน 100 ในรูป 3.1b เราแสดงผลตอบสนองเชิงเวลาและสเปกตรัมของ  $y, u$  ที่เห็นย่านความถี่ของระบบที่โดน กระตุ้นโดย  $u$  ชนิดหนึ่งๆ สำหรับการใช้สัญญาณขั้นที่มีแต่องค์ประกอบไปตรง เป็นที่แน่นอนว่า  $y(t)$  จะมีผลตอบสนองชั่วคราวและเข้าสู่ค่าสถานะอยู่ตัวค่าหนึ่ง สเปกตรัมของ  $y$  จึงมีองค์ประกอบหลักที่ ความถี่ศูนย์ การประมาณ  $G$  (ไม่ว่าจะขนาดหรือเฟส) ที่ความถี่อื่นจึงไม่สามารถทำได้

ตัวอย่างที่สองคือสัญญาณคลื่นสี่เหลี่ยมที่มีองค์ประกอบความถี่ 10 Hz และความถี่ฮาร์มอนิก จะเห็นว่าสัญญาณ  $y$  มีพลังงานไม่ถูกลดทอนลงมากเทียบกับ  $u$  ในช่วงย่านความถี่ต่ำกว่าแบนด์วิดท์ ของระบบ (เนื่องจากอัตราขยายระบบในช่วงนั้นมีค่าสูง) แต่ในย่านความถี่สูงเช่นเกินกว่า 100 Hz จะ เห็นว่าสเปกตรัมของ  $y$  จะมีพลังงานลดลง ตัวอย่างที่สาม คือการใช้สัญญาณไซน์ที่มีองค์ประกอบ 4 ความถี่ ได้แก่ 1, 10, 20, 100 Hz นั้นจะกระตุ้นให้ระบบมีค่า  $y$  ที่มีกำลังสูงในค่าความถี่นั้น และที่ความถี่ 100 Hz ที่สูงกว่าแบนด์วิดท์ของระบบ ทำให้ผลตอบสนอง  $y$  ที่ความถี่นั้นมีการลดลงมา ใน ตัวอย่างนี้จะเห็นว่า ที่ความถี่อื่นๆ นอกเหนือจากสี่ความถี่ของ  $u$  นั้น ระบบจะไม่ถูกกระตุ้นใดๆ เลย



(a) ผลตอบสนองเชิงความถี่ของ  $G(s)$



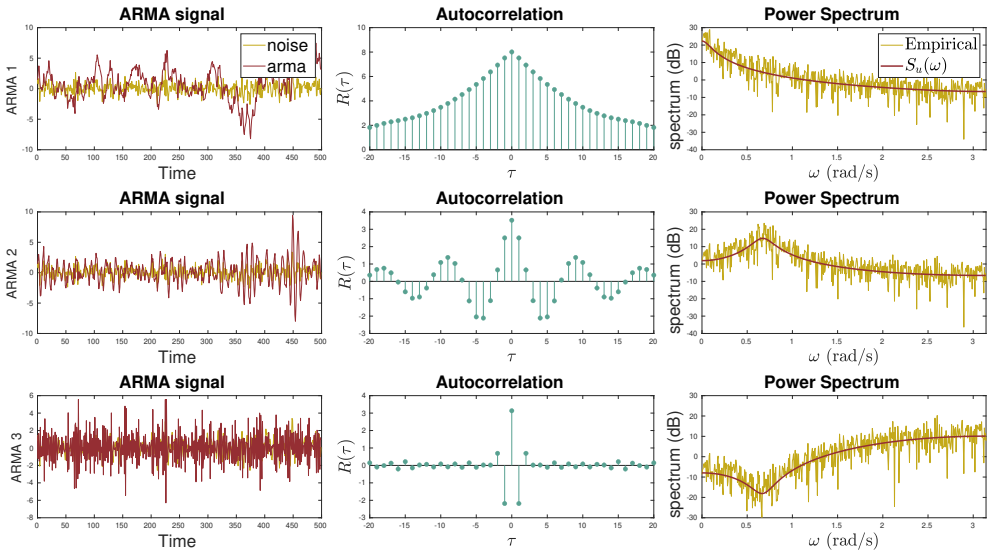
(b) ผลตอบสนองของระบบ  $G(s)$  ที่มีต่อสัญญาณเข้าชนิดต่างๆ

รูป 3.1: ลักษณะของสัญญาณออกของระบบ  $G(s)$  ที่มีต่อสัญญาณเข้าชนิดต่างๆ

ตัวอย่างสุดท้ายคือ การใช้สัญญาณรบกวนขาวที่มีองค์ประกอบความถี่ทุกย่าน (นั่นคือมีสเปกตรัมเป็นค่าคงตัว) สามารถกระตุ้นระบบได้ทุกช่วง และทำให้ผลตอบสนอง  $y$  มีพลังงานสูงในช่วงความถี่ที่อยู่ในแบนด์วิดท์ของระบบ และสเปกตรัมของ  $y$  จะลดลงในย่านความถี่สูงเกินจากแบนด์วิดท์นั้น ทั้งนี้การใช้สัญญาณรบกวนขาวอาจทำได้ในการทดลองจริง เพราะขนาดของ  $u$  อาจเกินช่วงทำงานของตัวขับเร็ว โดยหลักการกว้างจึงควรเลือก  $u$  ที่มีองค์ประกอบถี่ในย่านที่จะกระตุ้นระบบได้ทุกโหมด บทนี้จะศึกษาถึงสมบัติของสัญญาณเข้า  $u$  ชนิดต่างๆ หนึ่งในสมบัตินั้นคือ **ระดับการกระตุ้นของระบบ** อันเป็นดัชนีที่จะบ่งชี้ว่า  $u$  นั้นเหมาะสมในการกระตุ้นระบบที่มีความซับซ้อนระดับหนึ่งหรือไม่

ฟังก์ชันถ่ายโอนจาก  $e$  ไป  $u$  คือ  $G(z) = \frac{U(z)}{E(z)} = \frac{C(z)}{A(z)}$  โดยที่  $A(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p}$  คือพหุนาม AR และ  $C(z) = 1 + c_1z^{-1} + \dots + c_qz^{-q}$  คือพหุนาม MA เมื่อผู้ใช้งานกำหนดสัมประสิทธิ์ของพหุนามทั้งสองนี้ และสร้างสัญญาณ  $e$  ขึ้น เราจะเรียก  $u$  ที่ได้จาก (3.11) ว่าเป็นสัญญาณ ARMA

เงื่อนไขจำเป็นของการเลือกพหุนาม  $A(z), C(z)$  คือ พหุนามทั้งสองนี้ (ในตัวแปร  $z^{-1}$ ) ต้องมีรากอยู่นอกวงกลมหนึ่งหน่วย เพื่อให้กระบวนการ ARMA ที่ได้มีสมบัติผกผันได้และเป็นระบบเชิงเหตุ การเลือกค่าสัมประสิทธิ์ที่ต่างกันจะทำให้ผลตอบสนองเชิงความถี่ของ  $G(z)$  มีรูปร่างและแบนด์วิดท์ที่ต่างกันไป จากความสัมพันธ์ในทฤษฎีบท C.9 หน้า 280:  $S_u(\omega) = G(e^{i\omega})S_e(\omega)G^*(e^{i\omega})$  จะเห็นว่า สเปกตรัมของ  $u$  ก็จะแปรไปตาม  $G(z)$  ที่ผู้ใช้เลือก ดังตัวอย่างในรูป 3.5 กระบวนการ ARMA1 นั้น มีสเปกตรัมกำลังสูงในย่านความถี่ต่ำ เราจึงเห็นสัญญาณ ARMA1 ในโดเมนเวลาที่ได้มีการกระเพื่อมที่ช้าลงเมื่อเทียบกับ  $e$  สำหรับ ARMA3 นั้นสอดคล้องกับสเปกตรัมกำลังที่สูงในย่านความถี่สูง หรืออัตราขยายลดทอนมีในช่วงความถี่แคบเท่านั้น จึงเห็นว่าสัญญาณ ARMA3 มีการเปลี่ยนแปลงที่เร็วพอกับ  $e$  และฟังก์ชันสหสัมพันธ์ก็ยิ่งคล้ายกับของ  $e$  อยู่ (นั่นคือมีค่าที่  $\tau \neq 0$ )



รูป 3.5: สัญญาณ ARMA ที่สร้างจากกระบวนการ ARMA 1:  $G(z) = \frac{1+0.3z^{-1}}{1-0.7z^{-1}-0.2z^{-2}}$  ARMA 2:  $G(z) = \frac{1-0.5z^{-1}}{1-1.4z^{-1}+0.8z^{-2}}$  และ ARMA 3:  $G(z) = 1 - 1.4z^{-1} + 0.8z^{-2}$  และใช้สัญญาณรบกวนขาวมีความแปรปรวนเท่ากับหนึ่ง

### 3.1.5 ลำดับฐานสองคู่เทียบ

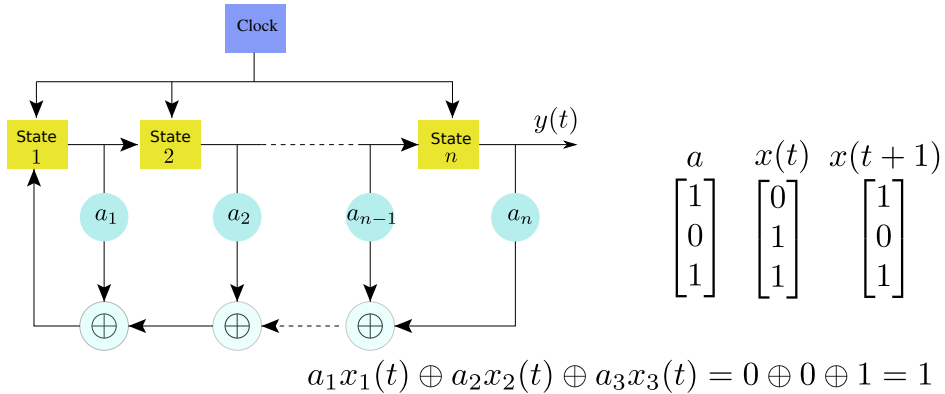
PRBS คือลำดับค่าฐานสองที่สร้างให้เหมือนกับเป็นตัวแปรสุ่ม โดยสังเคราะห์ให้เป็นสัญญาณออกของระบบเชิงเส้นอันดับ  $n$  ที่บรรยายด้วยสมการปริภูมิสถานะดังนี้

$$x(t+1) = \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix} x(t), \quad y(t) = [0 \quad \dots \quad 0 \quad 1] x(t) \quad (3.12)$$

และสมมูลกับสมการ

$$x_1(t + 1) = a_1x_1(t) + a_2x_1(t - 1) + \dots + a_nx_1(t - n + 1)$$

สัญญาณ PRBS นั้นสอดคล้องกับสมการถดถอยในตัวอันดับ  $n$  ที่แสดงเป็นแผนผังดังรูป 3.6



รูป 3.6: แผนผังการสร้างลำดับฐานสองสุ่มเทียม (PRBS)

กล่าวคือ  $x_2(t)$  จะรับค่ามาจาก  $x_1(t - 1)$  ส่วน  $x_n(t)$  จะรับค่ามาจาก  $x_{n-1}(t - 1)$  ตามลำดับ และ  $x_1$  ในเวลาปัจจุบัน ได้จากการหาผลรวมของตัวแปรสถานะทุกตัวในเวลาก่อนหน้าที่ถ่วงน้ำหนักด้วย  $a_k$  จากนั้น สัญญาณ PRBS คือตัวแปร  $y(t)$  จะได้จากการอ่านค่าสถานะ  $x_n(t)$  เรากำหนดให้ระบบ (3.12) มีพารามิเตอร์  $a_1, a_2, \dots, a_n$  นั้นเป็นค่า 0 หรือไม่ก็ 1 ค่าเริ่มต้น  $x(0)$  มีค่าฐานสองที่ไม่เท่ากับศูนย์ และการบวกในสมการ (3.12) จะเป็นการบวกแบบมอดูโลสอง (modulo-two) หรือ XOR ด้วยเงื่อนไขทั้งหมดนี้ เราจะได้ว่าลำดับ  $x(t)$  ที่เป็นผลลัพธ์จะเป็นเวกเตอร์ฐานสองเสมอที่ไม่เท่ากับศูนย์ และมีจำนวนเวกเตอร์ที่เป็นไปได้ทั้งหมดเท่ากับ  $M = 2^n - 1$  ที่ถูกกำหนดให้เรียกว่าเป็น **ความยาวสูงสุดของสัญญาณ PRBS** เมื่อใช้ค่าพารามิเตอร์  $a_k$  ชุดที่ต่างกันจะส่งผลให้คาบของสัญญาณ PRBS ต่างกันได้ เช่น

- ถ้ากำหนดให้  $n = 3$  และเริ่มต้นด้วย  $x(0) = (1, 0, 0)$  เมื่อ  $a = (1, 1, 0)$  จะได้ว่า  $x(t)$  เมื่อ  $t = 1, 2, 3, \dots$ , มีค่าเป็น

$$(1, 0, 0), (1, 1, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)$$

นั่นคือลำดับ  $x(t)$  จะมีคาบเท่ากับ 3

- แต่ถ้าหากกำหนดให้  $a = (1, 0, 1)$  จะได้ว่า  $x(t)$  เมื่อ  $t = 1, 2, 3, \dots$ , จะมีค่าเป็น

$$(1, 0, 0), (1, 1, 0), (1, 1, 1), (0, 1, 1), (1, 0, 1), (0, 1, 0), (0, 0, 1), (1, 0, 0)$$

นั่นคือลำดับ  $x(t)$  จะมีคาบเท่ากับ 7 และเท่ากับความยาวสูงสุดคือ  $2^3 - 1$

จากตัวอย่างข้างต้น ลำดับ  $x(t)$  ที่มีคาบเท่ากับ  $M$  จึงเป็นกรณีที่น่าสนใจ เพราะนั่นหมายความว่า  $y(t)$  มีค่าที่เป็นไปได้หลากหลายมากที่สุด ในทฤษฎีบทดังต่อไปนี้ จึงกล่าวถึงเงื่อนไขของพารามิเตอร์  $a_1, \dots, a_n$  ที่จะนำไปสู่สัญญาณ PRBS ที่มีความยาวสูงสุด



### สเปกตรัมของ $u$

สัญญาณ  $u$  ควรมีสเปกตรัมที่มีองค์ประกอบหลายความถี่ให้เป็นจำนวนเหมาะสมกับความซับซ้อนของแบบจำลองที่ประมาณ ดังที่อธิบายไว้ในหัวข้อ 3.2 หากใช้สัญญาณไซน์มีจำนวนความถี่เท่ากับ  $m$  จะมีระดับการกระตุ้นคงอยู่ประมาณ  $2m$  (จึงควรเลือกให้  $m$  มากพอ) หากใช้สัญญาณ PRBS ระดับการกระตุ้นคงอยู่เท่ากับความยาวสูงสุด  $M = 2^n - 1$  (จึงควรเลือก  $M$  ให้สูงด้วยการเลือก  $n$ ) ตัวอย่างสเปกตรัมของสัญญาณไซน์หลายความถี่ แสดงดังรูป 3.11a เราควรใช้ความรู้เบื้องต้นของระบบ เช่น แบนด์วิดท์ ช่วงความถี่หักมุม เพื่อเลือกความถี่ของสัญญาณไซน์

ตัวอย่างระบบ  $G(z) = \frac{0.5z^{-1} + 0.3z^{-2} - 0.1z^{-3}}{1 - 1.5z^{-1} + 0.5z^{-2} + 0.1z^{-3}}$  ที่มีแบนด์วิดท์ประมาณ 42 rad/s หรือ 6.6 Hz ความถี่หักมุมแรกประมาณ 4 Hz (สามารถอ่านค่าเหล่านี้ได้จากแผนภาพโพล) เมื่อสร้างสัญญาณไซน์สี่แบบ แต่ละแบบคือสัญญาณไซน์ความถี่เดียวที่ความถี่ 1, 11, 23, 41 Hz ตามลำดับ และนำไปกระตุ้นระบบ  $G$  เพื่อเก็บข้อมูล  $\{y(t), u(t)\}_{t=1}^{300}$  แล้วประมาณแบบจำลอง  $y(t) = \frac{N(L)}{D(L)}u(t) + e(t)$  ที่มีจำนวนศูนย์และโพลเท่ากับ 3 (ตามโครงสร้างระบบจริง) จะได้ผลลัพธ์การประมาณดังรูป 3.11b ที่แสดงให้เห็นว่า สำหรับระบบนี้ สัญญาณไซน์ความถี่ต่ำค่า 1 Hz ให้ผลการประมาณที่ดีกว่าการใช้สัญญาณความถี่สูง (เมื่อเทียบโดยประมาณกับแบนด์วิดท์) ดังนั้น เราเลือกใช้สัญญาณไซน์หลายความถี่ที่แปรค่าในย่านความถี่ 1-10 Hz น่าจะเป็นตัวเลือกในทางปฏิบัติสำหรับระบบนี้

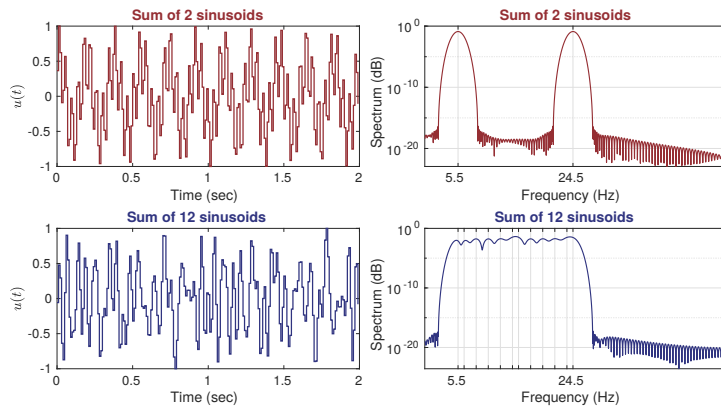
นอกจากนี้ เราควรมีหลักการเกี่ยพลังงานของ  $u$  ในแต่ละย่านความถี่ที่เลือกไว้  $u$  ควรมีพลังงานสูงในย่านความถี่ที่ผลตอบสนองเชิงความถี่ของระบบมีเกรเดียนต์สูงเมื่อเทียบกับพารามิเตอร์ ตัวอย่างง่าย เช่น แผนภาพโพลเดของระบบอันดับหนึ่งจะมีการหักมุมเปลี่ยนแปลงที่ความถี่หักมุมและที่ความถี่นั้นเองจะเป็นสิ่งที่บอกตำแหน่งโพลของระบบ ความถี่สัญญาณไซน์ที่เลือกควรครอบคลุมช่วงความถี่หักมุมนั้น เป็นต้น การจัดสรรกำลังของ  $u$  ในย่านความถี่ที่สำคัญมีเหตุผลรองรับทางคณิตศาสตร์พิสูจน์ไว้ใน [Lju99, §13.3]

พิจารณาการประมาณระบบ  $G(z) = \frac{0.8z^{-1} + 0.1z^{-2} - 0.3z^{-3}}{1 - 1.2z^{-1} + 0.1z^{-2} + 0.4z^{-3} - 0.2z^{-4}}$  การทดลองนี้ใช้  $u$  สี่แบบได้แก่ 1) สัญญาณไซน์ 10 ความถี่ในช่วง 1-19 Hz 2) สัญญาณไซน์ 10 ความถี่ในช่วง 21-49 Hz 3) PRBS ที่มีความยาวสูงสุด 15 และ 4) สัญญาณขั้น ผลการทดลองรูป 3.12a แสดงว่า  $G(e^{i\omega})$  มีเกรเดียนต์สูงเทียบกับสัมประสิทธิ์พหุนามเศษและส่วน (คู่กับพจน์  $z^{-1}, z^{-2}$  ตามลำดับ) ในช่วงย่านความถี่ไม่เกิน 50 rad/s หรือประมาณ 8 Hz (กราฟอัตราขยายของ  $G + \Delta G$  มีค่าแปรเปลี่ยนไปสูง) จึงควรจัดสรรกำลังของ  $u$  ให้มีค่าสูงไปในช่วงความถี่ต่ำ ผลการประมาณดังรูป 3.12b จึงสะท้อนว่า สัญญาณไซน์ 10 ความถี่ในช่วงความถี่ต่ำ ให้ผลการประมาณออกมาที่ดีกว่าการเลือกใช้ความถี่ย่านสูง

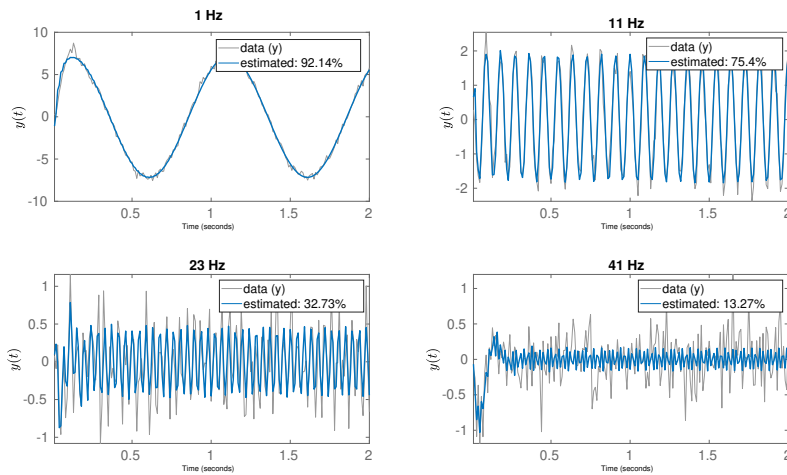
การเลือก  $u$  ที่มีสเปกตรัมกว้างและครอบคลุมย่านความถี่ของระบบที่เราสนใจ จึงเป็นตัวเลือกที่เหมาะสม เช่น การใช้สัญญาณรบกวนขาว หรือในรูป 3.12b ที่ใช้ PRBS คาบเท่ากับ 15 ซึ่งผลเทียบกับการใช้สัญญาณขั้นที่กระตุ้นระบบได้ไม่เต็มที่ จึงเห็นว่าผลการประมาณมีความคลาดเคลื่อนสูงกว่ากรณีใช้ PRBS ในทางทฤษฎีนั้น สัญญาณรบกวนขาวเกาส์เซียนมีค่าไม่จำกัด ในทางปฏิบัติจึงต้องมีการจำกัดค่าขอบเขต ซึ่งส่งผลทำให้สเปกตรัมเปลี่ยนไป วิธีหนึ่งที่ย่าง คือใช้สัญญาณรบกวนขาวที่ผ่านตัวกรอง เช่นเลือกพารามิเตอร์ในตัวกรอง ARMA ดัง (3.11) ให้สัญญาณมีรูปร่างสเปกตรัมที่เหมาะสม รวมไปถึงการเลือกใช้สัญญาณ PRBS ที่มีคาบ  $M$  สูงมากพอ

### ลักษณะของสัญญาณ

การเก็บข้อมูล  $\{y(t), u(t)\}$  จากการทดลองการหาเอกลักษณ์ของระบบ ควรเป็นไปให้ชุดข้อมูลมี SNR (signal-to-noise-ratio) ที่สูง เราจึงควรเลือก  $u$  ให้มีกำลังของสัญญาณค่าสูง แต่เนื่องจาก



(a) สัญญาณไซน์และสเปกตรัมจากคำสั่ง idinput



(b) ผลของความถี่ไซน์ที่เลือกต่อสมรรถนะการประมาณ

รูป 3.11: สัญญาณไซน์หลายความถี่ (multisine)

ข้อจำกัดทางปฏิบัติที่เราไม่สามารถสังเคราะห์ขนาดของ  $u$  เกินขีดจำกัดของตัวขับเราได้ การเลือกรูปแบบของสัญญาณ  $u$  ที่เหมาะสม จึงมีการนิยามในรูปของ **ตัวประกอบเครสต์ (Crest factor)** [Lju99, S13.3] สำหรับสัญญาณ  $u$  ที่มีค่าเฉลี่ยเป็นศูนย์

$$C_r = \sqrt{\frac{\max_t u^2(t)}{\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u^2(t)}} \quad (3.23)$$

ตัวประกอบ  $C_r$  แสดงอัตราส่วนระหว่างค่ายอดสูงสุดกับค่า RMS หากค่า  $C_r$  ต่ำสุดเท่ากับ 1 จึงเปรียบเสมือนสัญญาณไม่มีค่ายอด เช่น สัญญาณฐานสองสมมาตร  $u(t) = \pm 1$  (คลื่นสี่เหลี่ยม) สำหรับสัญญาณไซน์ขนาด  $A$  จะมีค่า  $C_r = \sqrt{2} \approx 1.414$  เราจะเห็นว่าตัวส่วนของ  $C_r$  มีความหมายเป็นกำลังของสัญญาณเข้า  $u$  (ที่ต้องการให้มีค่ามากเท่าที่เป็นไปได้) ค่า  $C_r$  ที่ต่ำจึงเป็นสิ่งที่พึงประสงค์

สำหรับการหาเอกลักษณ์ของระบบเชิงเส้น สรุปลงได้ว่าการเลือก  $u$  ควรเลือกให้องค์ประกอบความถี่ที่ครอบคลุม มีรูปร่างสเปกตรัมเพื่อจัดสรรกำลังของ  $u$  ในย่านความถี่ที่เหมาะสม และมีค่า  $C_r$  ที่น้อย

---

## บทที่ 4

# พื้นฐานทฤษฎีการประมาณ

---

บทนี้จะกล่าวถึงพื้นฐานที่จำเป็นเพื่อทำความเข้าใจสมบัติต่างๆ ของตัวประมาณ สัญลักษณ์  $\theta$  คือพารามิเตอร์ที่ต้องการประมาณ และสัญลักษณ์  $\hat{\theta}$  คือค่าประมาณของ  $\theta$  โดยทั่วไป การจะประมาณ  $\theta$  ได้นั้นต้องใช้ข้อมูลวัดที่มีความเกี่ยวข้องกับ  $\theta$  ข้อมูลวัดมักจะมีสัญญาณรบกวน จึงมีลักษณะเป็นตัวแปรสุ่ม หรือในบางงานประยุกต์ พารามิเตอร์ที่ต้องการประมาณเป็นตัวแปรสุ่ม จึงสามารถกล่าวโดยทั่วไปว่าข้อมูลตัวอย่างเป็นตัวแปรสุ่ม จึงทำให้  $\hat{\theta}$  เป็นตัวแปรสุ่มเช่นกัน สารสำคัญของบทนี้มีไว้เพื่อให้ผู้อ่านคุ้นเคยกับนิยามที่จะใช้วัดคุณภาพหรืออภิปรายสมบัติของ  $\hat{\theta}$  บทนี้ใช้หลักการทางสถิติ และนิยามการลู่เข้าแบบต่างๆ ของตัวแปรสุ่ม ซึ่งบรรยายไว้ในภาคผนวก C.1 ในบริบทของการประมาณแบบจำลองนั้น  $\hat{\theta}$  จะหมายถึงพารามิเตอร์ของแบบจำลองที่เราประมาณได้ หรือพิจารณาในรูปคณิตศาสตร์ว่า ค่าของแบบจำลองที่แทนด้วยพารามิเตอร์ที่ประมาณได้ก็เสมือนเป็นฟังก์ชันของ  $\hat{\theta}$  ดังนั้น การประมาณระบบ  $G(\theta)$  ด้วย  $\hat{G}(\hat{\theta})$  เราต้องการให้  $G(\theta) = \hat{G}(\hat{\theta})$  หรือเราสนใจพารามิเตอร์ที่เท่ากัน  $\theta = \hat{\theta}$  สิ่งเหล่านี้หมายความว่า เรายังต้องการนิยามที่จะกล่าวถึงสมบัติของแบบจำลองด้วย อันเป็นเนื้อหาในหัวข้อ 4.2

วิธีการประมาณแบบจำลองในหนังสือเล่มนี้เป็นแนวทางอิงพารามิเตอร์ (parametric approach) ที่จะประมาณพารามิเตอร์  $\theta$  ด้วยการกำหนดฟังก์ชันสูญเสียอันบ่งชี้ความต่างระหว่างแบบจำลองกับข้อมูลจากระบบ เราจะเลือก  $\theta$  ค่าที่ *เหมาะสมที่สุด* ในความหมายว่าจะทำให้ฟังก์ชันสูญเสียมีค่าต่ำสุด ตัวประมาณที่ได้จากวิธีกลุ่มนี้จะเรียกว่าตัวประมาณสุดขีด (extremum estimator) สมบัติเชิงสถิติในหัวข้อ 4.3 เป็นผลลัพธ์พื้นฐานที่ใช้ทำความเข้าใจในเนื้อหาขั้นสูงต่อไป แต่ผู้อ่านเริ่มต้นสามารถข้ามหัวข้อ 4.3 ก่อน เพื่อทำความเข้าใจวิธีประมาณแบบต่างๆ แล้วจึงค่อยยกกลับมาอ่านหัวข้อนี้อีกครั้ง ก่อนศึกษาเนื้อหาในบทนี้ ผู้อ่านควรมีความคุ้นเคยเกี่ยวกับนิยามการลู่เข้าต่างๆ ของตัวแปรสุ่มเสียก่อน ซึ่งอธิบายไว้ในภาคผนวก C.1 หน้า 273

### 4.1 สมบัติของตัวประมาณ

ค่า  $\hat{\theta}$  ที่ประมาณได้จากข้อมูลตัวอย่างชุดหนึ่งจะเรียกว่า **ค่าประมาณแบบจุด** (point estimate) เมื่อเปลี่ยนข้อมูลชุดตัวอย่างไป เราย่อมได้ค่า  $\hat{\theta}$  ที่เปลี่ยนไป ตัวอย่างเช่น ค่าเฉลี่ยตัวอย่าง (sample mean) เป็นค่าประมาณแบบจุด เมื่อใช้ข้อมูลตัวอย่าง 2 ชุด ย่อมให้ค่าเฉลี่ยตัวอย่างที่ไม่เท่ากัน เมื่อ

เข้าใจว่า  $\hat{\theta}$  เป็นตัวแปรสุ่มแล้วนั้น โดยที่ยังไม่ทราบการแจกแจงที่แน่ชัด เราอาจนึกถึง 2 โมเมนต์แรก อันได้แก่ ค่าคาดหวัง และความแปรปรวนของ  $\hat{\theta}$  เราคิดว่า ค่าคาดหวังของ  $\hat{\theta}$  ควรจะใกล้เคียงกับ พารามิเตอร์จริง ส่วนความแปรปรวนควรจะมีค่าน้อย เพราะส่งผลถึงความน่าเชื่อถือของค่าประมาณ แบบจุดจากการใช้ข้อมูลชุดหนึ่งๆ

หัวข้อนี้จะเสนอนิยามของ 3 สมบัติที่สำคัญของตัวประมาณ  $\hat{\theta}$  ได้แก่

- ความไม่เอนเอียง (unbiased) จะกล่าวถึงค่าคาดหวังของ  $\hat{\theta}$  (เมื่อเทียบกับการแจกแจงของ  $\hat{\theta}$  เอง) ว่าควรมีค่าเท่ากับพารามิเตอร์ที่แท้จริง
- ความคงเส้นคงวา (consistency) เป็นสมบัติที่จะกล่าวว่าหากมีข้อมูลตัวอย่างจำนวนมากขนาด  $N \rightarrow \infty$  แล้ว ความน่าจะเป็นที่  $\hat{\theta}$  จะเข้าใกล้ค่าพารามิเตอร์จริงจะมีค่าเข้าใกล้หนึ่งหรือไม่
- ประสิทธิภาพ (efficiency) เป็นสมบัติที่พิจารณาเมทริกซ์ความแปรปรวนร่วมของ  $\hat{\theta}$  ในกรณีที่ มีข้อมูลจำนวนมาก ว่าความแปรปรวนจะมีค่าต่ำที่สุดเท่าที่เป็นไปได้หรือไม่

**นิยาม 4.1** (unbiased estimator).  $\hat{\theta}$  เป็น *ตัวประมาณไม่เอนเอียง* ของ  $\theta$  ถ้าหาก  $\mathbf{E}[\hat{\theta}] = \mathbf{E}[\theta]$



**ตัวอย่าง 4.1** กำหนดให้  $\{X_i\}_{i=1}^N$  เป็นข้อมูล i.i.d. ที่มีค่าเฉลี่ยเป็น  $\mu$  และความแปรปรวนเท่ากับ  $\sigma^2$  (ซึ่งพารามิเตอร์  $(\mu, \sigma^2)$  ที่แท้จริงนั้น เราไม่ทราบค่า) ตัวประมาณหนึ่งของ  $\mu$  คือ ค่าเฉลี่ยตัวอย่าง  $\bar{X} = (1/N) \sum_{i=1}^N X_i$  เพื่อจะตรวจสอบสมบัติความไม่เอนเอียงของ  $\bar{X}$  เราจึงหาค่าคาดหวังของ  $\bar{X}$  และใช้สมบัติว่า  $X_i$  ต่างก็มีค่าคาดหวังเท่ากับ  $\mu$

$$\mathbf{E}[\bar{X}] = \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N X_i\right] = \frac{1}{N} \sum_{i=1}^N \mathbf{E}[X_i] = \frac{1}{N} \sum_{i=1}^N \mu = \mu$$

จากผลลัพธ์ข้างต้น ค่าเฉลี่ยตัวอย่างจึงเป็นตัวประมาณไม่เอนเอียงของ  $\mu$

**นิยาม 4.2** (consistent estimator). ตัวประมาณ  $\hat{\theta}_N$  ที่คำนวณจากข้อมูล  $N$  ตัวอย่างเป็น *ตัวประมาณคงเส้นคงวา* กับ  $\theta$  ถ้าหาก  $\hat{\theta}_N$  เข้าใจความน่าจะเป็นเข้าหาค่าพารามิเตอร์ที่แท้จริง  $\theta$  นั่นคือ  $\hat{\theta}_N \xrightarrow{P} \theta$

$$\forall \epsilon > 0, \quad \lim_{N \rightarrow \infty} P(|\hat{\theta}_N - \theta| \leq \epsilon) = 1$$



**ตัวอย่าง 4.2** กำหนดให้  $\{X_i\}_{i=1}^N$  คือตัวอย่าง i.i.d. ที่สุ่มมาจากการแจกแจง  $\mathcal{N}(0, \sigma^2)$  และ  $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$  คือค่าเฉลี่ยจากข้อมูล  $N$  ตัวอย่าง เนื่องจาก  $\bar{X}_N$  เป็นตัวแปรเกาส์เซียนเช่นกัน (เพราะเป็นผลรวมเชิงเส้นของ  $X_i$ ) และมีความแปรปรวนเป็น  $\sigma^2/N$  เราคำนวณได้ว่า

$$P(|\bar{X}_N - \mu| \geq \epsilon) = P\left[\frac{\sqrt{N}|\bar{X}_N - \mu|}{\sigma} \geq \sqrt{N}\epsilon/\sigma\right] = \left(1 - \Phi\left(\frac{\sqrt{N}\epsilon}{\sigma}\right)\right) \rightarrow 0$$

เมื่อ  $N \rightarrow \infty$  โดยที่  $\Phi(\cdot)$  คือฟังก์ชันการแจกแจงสะสม (cdf) ของตัวแปรเกาส์เซียนมาตรฐาน ค่าเฉลี่ยตัวอย่างของตัวแปรเกาส์เซียนมีสมบัติคงเส้นคงวากับค่าเฉลี่ยที่แท้จริง

## บทที่ 5

# การประมาณกำลังสองน้อยสุดเชิงเส้น

ปัญหาการประมาณกำลังสองน้อยสุดเชิงเส้น Linear Least-Squares (LS) นั้นเป็นปัญหาพื้นฐานแบบหนึ่งของกลุ่มวิธีการประมาณแบบจำลอง เราเริ่มจากสมมติฐานที่ว่าตัวแปรตอบสนอง<sup>1</sup>  $y \in \mathbf{R}$  นั้น ถูกอธิบายจากตัวแปรทำนาย<sup>2</sup>  $n$  ตัวแปร ผ่านทางความสัมพันธ์เชิงเส้น ในรูป

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (5.1)$$

เราจะเรียก  $\beta_k \in \mathbf{R}$  สำหรับ  $k = 1, 2, \dots, n$  ว่าเป็น สัมประสิทธิ์การถดถอย (regression coefficients) และเป็นพารามิเตอร์ที่ต้องการประมาณ ตัวอย่างการใช้แบบจำลองถดถอยใน (5.1) เช่น  $y$  เป็นยอดขายผลิตภัณฑ์หนึ่งของบริษัท (หน่วยเป็นบาท) ผู้ใช้ต้องการทราบว่า ยอดขายที่ได้แตกต่างกันในแต่ละเดือนมาจากการลงทุนโฆษณาในช่องทางใดบ้างหากเลือก  $x_1, x_2, x_3$  คือ งบการโฆษณา (หน่วยเป็นบาท) บน Facebook, Google, และทางทีวี ตามลำดับ ดังนั้น ค่า  $\beta_1$  บ่งชี้ว่า หากลงทุนงบโฆษณาบน Facebook ไปหนึ่งหน่วย (ซึ่งในที่นี้คือ 1 บาท) จะทำให้ยอดขายผลิตภัณฑ์เพิ่มขึ้น  $\beta_1$  บาท เมื่องบโฆษณาช่องทางอื่นมีค่าคงที่

แบบจำลองดังสมการ (5.1) มีตัวแปรทำนายหลายตัวแปร จึงเรียกว่า แบบจำลองถดถอยเชิงเส้นหลายตัวแปร (multiple linear regression) หากกำหนดให้  $\{y_i, x_{i1}, x_{i2}, \dots, x_{in}\}_{i=1}^m$  คือเซตของข้อมูลที่วัดได้ จึงสามารถเขียน (5.1) ในรูปเวกเตอร์ดังนี้

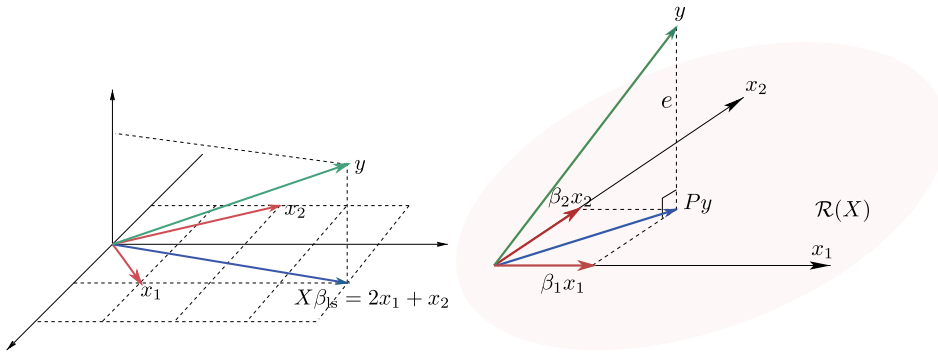
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \dots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \triangleq y = X\beta \quad (5.2)$$

เราเรียก  $X$  ว่าเป็นเมทริกซ์ถดถอย (regressor matrix) และ  $\beta$  ว่าเป็นเวกเตอร์สัมประสิทธิ์ (coefficient vector)

จุดหมายของการใช้แบบจำลองถดถอยเชิงเส้นคือ เมื่อทราบ  $y$  และ  $X$  จากข้อมูล จึงต้องการหาค่า  $\beta$  ที่ทำให้ระบบสมการ (5.2) นั้นมีคำตอบ หรือถ้าหากไม่มีคำตอบ ก็หมายถึงว่า ต้องการเลือก  $\beta$  ให้  $y \approx X\beta$  ด้วยความหมายการ “ใกล้เคียง” บางอย่าง การหาคำตอบของระบบสมการเชิงเส้น (5.2)

<sup>1</sup>  $y$  ยังมีชื่ออื่น เช่น ตัวแปรตาม (dependent variable) หรือตัวแปรถูกอธิบาย (explained variable)

<sup>2</sup>  $x$  ยังมีชื่ออื่น เช่น ตัวแปรอิสระ (independent variable) ตัวแปรอธิบาย (explanatory variable)



รูป 5.1: ผลเฉลยกำลังสองน้อยสุดทำให้  $X\beta_{ls}$  ใกล้เวกเตอร์  $y$  มากที่สุดในนอร์ม-2 และตั้งฉากกับ  $e$

แต่เนื่องจาก  $Py \in \mathcal{R}(X)$  เราจึงกล่าวได้ว่า  $Py$  ก็ยังสามารถเขียนได้ในรูปของผลรวมเชิงเส้นของคอลัมน์ใน  $X$  เป็น  $Py = \sum_{j=1}^n \beta_j x_j$  เมื่อนำพจน์นี้ไปแทนในเงื่อนไขตั้งฉาก จะได้ว่า

$$x_i^T y = \sum_{j=1}^n x_i^T x_j \beta_j, \quad i = 1, 2, \dots, n$$

อันเป็นระบบสมการเชิงเส้นที่มี  $\beta_j$  เป็นพารามิเตอร์ หากเราเขียนในรูปเวกเตอร์ จะได้ว่า

$$\begin{bmatrix} x_1^T y \\ x_2^T y \\ \vdots \\ x_n^T y \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & x_1^T x_n \\ x_2^T x_1 & x_2^T x_2 & \cdots & x_2^T x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n^T x_1 & x_n^T x_2 & \cdots & x_n^T x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \triangleq X^T y = X^T X \beta$$

อันเป็นสมการเดียวกันกับสมการปกติ (normal equation) ใน (5.4) จึงสรุปได้ว่า ผลเฉลยกำลังสองน้อยสุด  $\beta_{ls}$  ใน (5.5) นั้น ทำให้ค่าประมาณของ  $y$  ที่จะเรียกว่า

$$\hat{y} = X\beta_{ls} = X(X^T X)^{-1} X^T y := Py$$

เป็นเวกเตอร์ภาพฉายตั้งฉากจาก  $y$  ไปยัง  $\mathcal{R}(X)$  ดังแสดงในรูป 5.1 นั้นเอง และพบว่า

$$P = X(X^T X)^{-1} X^T \quad (5.6)$$

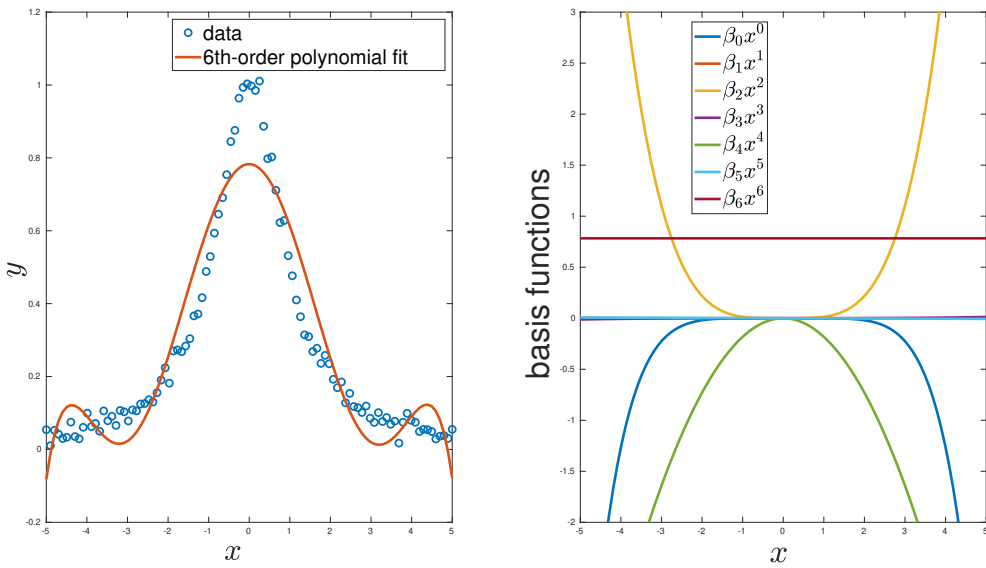
เป็นเมทริกซ์ภาพฉายตั้งฉาก (orthogonal projection) เนื่องจาก  $P$  มีคุณสมบัติ  $P = P^T$  และ  $P^2 = P$  ตัวอย่างปัญหากำลังสองเชิงเส้นน้อยสุดมีดังนี้

**ตัวอย่าง 5.1** พิจารณาการประมาณจุดข้อมูล  $\{(t_i, y_i)\}_{i=1}^m$  ด้วยฟังก์ชัน (data fitting) และจะประมาณความสัมพันธ์  $y$  กับ  $t$  ด้วยการแตกองค์ประกอบของ  $y$  เป็นผลรวมของฟังก์ชัน

$$y = \hat{g}(t) := \beta_1 g_1(t) + \beta_2 g_2(t) + \cdots + \beta_n g_n(t) \quad (5.7)$$

เราเรียก  $g_k(t) : \mathbf{R} \rightarrow \mathbf{R}$  ว่าเป็นฟังก์ชันมูลฐาน (basis function) เช่น ฟังก์ชันพหุนาม  $g_1(t) = 1, g_2(t) = t, g_3(t) = t^2, \dots, g_n(t) = t^{n-1}$  หรือ ฟังก์ชันไซน์และโคไซน์ที่มีความถี่  $\omega_1, \dots, \omega_n$

$$g_{11}(t) = \cos(\omega_1 t), g_{12}(t) = \sin(\omega_1 t), \dots, g_{n1}(t) = \cos(\omega_n t), g_{n2}(t) = \sin(\omega_n t)$$



รูป 5.2: การประมาณฟังก์ชันพหุนามอันดับ 6 เข้ากับชุดข้อมูล

เราสามารถจัดรูปแบบของสมการเชิงเส้นถดถอยได้เป็น

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} g_1(t_1) & g_2(t_1) & \cdots & g_n(t_1) \\ g_1(t_2) & g_2(t_2) & \cdots & g_n(t_2) \\ \vdots & \vdots & \cdots & \vdots \\ g_1(t_m) & g_2(t_m) & \cdots & g_n(t_m) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \triangleq y = X\beta$$

ในงานประยุกต์จะมีข้อมูล  $\{(t_i, y_i)\}_{i=1}^m$  และผู้ใช้กำหนดตัวเลือกของฟังก์ชันมูลฐานที่จะใช้ ดังนั้นเราจะทราบค่าเมทริกซ์  $X$  ด้วยการแทนค่า  $t_i$  ลงไปใน  $g_k(t_i)$  ในกรณีที่  $m \gg n$  หมายความว่าเราต้องการแบบจำลองที่ไม่ซับซ้อนมากนัก เพราะมีจำนวนพารามิเตอร์น้อยเมื่อเทียบกับจำนวนข้อมูล ในการอธิบายข้อมูล  $y$  ที่วัดได้ จึงแก้ปัญหา LS ด้วยการใส่  $y$  และ  $X$  ดังข้างต้น

ผลของการประมาณความสัมพันธ์  $y = \hat{g}(t)$  ที่แตกเป็นองค์ประกอบของฟังก์ชันพหุนาม แสดงในรูป 5.2 ชุดข้อมูลนั้นสร้างจากฟังก์ชันไม่เชิงเส้น  $g(t) = 1/(1 + t^2)$  (ซึ่ง ณ ขั้นตอนการประมาณเราไม่ทราบฟังก์ชันจริงนี้) เมื่อประมาณด้วยวิธี LS  $\hat{g}(t)$  สามารถแตกเป็นฟังก์ชันมูลฐาน  $x^k$  สำหรับ  $k = 0, 1, \dots, 6$  ดังรูป 5.2

**ตัวอย่าง 5.2** แบบจำลอง FIR ดังตาราง 2.1 มีความสัมพันธ์ระหว่างสัญญาณเข้า/ออกเป็น

$$y(t) = \sum_{k=0}^{n-1} h(k)u(t - k) \tag{5.8}$$

ในปัญหานี้ สัญญาณเข้า  $u$  เป็นสัญญาณที่ทราบค่าที่ผู้ใช้ตั้งใจออกแบบให้ไปกระตุ้นระบบ  $H$  และ วัดสัญญาณออก  $y$  ออกมา ค่าของผลตอบสนองอิมพัลส์  $h(k)$  สำหรับทุก  $k = 0, \dots, n-1$  เป็นพารามิเตอร์ที่ต้องการประมาณ กำหนดให้มีข้อมูล  $\{y(t), u(t)\}_{t=0}^m$  เมื่อเห็นว่า ความสัมพันธ์

จึงเห็นผลดังนี้ หาก  $m$  สูงขึ้น  $X^T X$  จะเป็นเมทริกซ์ที่ใหญ่ขึ้น ด้วยการเทียบแบบเมทริกซ์บวกแน่นอน ( $A \succeq B$  ก็ต่อเมื่อ  $A - B \succeq 0$ ) เนื่องจากเมทริกซ์ถดถอยที่ใช้  $m$  ข้อมูล สามารถแยกเป็น 2 ส่วนคือ

$$X^T X = \sum_{i=1}^{m-1} x_i x_i^T + x_m x_m^T$$

ดังนั้น  $X^T X \succeq \sum_{i=1}^{m-1} x_i x_i^T$  เพราะว่า  $x_m x_m^T \succeq 0$  นอกจากนี้ เราทราบว่าสำหรับ  $A > 0, B > 0$  ใดๆ หาก  $A \succeq B$  แล้ว  $B^{-1} \succeq A^{-1}$  ดังนั้น หาก  $m$  มีค่าสูงขึ้น  $X^T X$  จะใหญ่ขึ้น และทำให้  $(X^T X)^{-1}$  เล็กลง การใช้ข้อมูลจำนวนมากขึ้น ในปัญหา LS จึงช่วยลดความแปรปรวนของ  $\beta_{ls}$  นั่นเอง  
ในหัวข้อต่อไป จะนำเสนอทฤษฎีบทที่สำคัญของตัวประมาณ LS

**ทฤษฎีบท 5.1** (ทฤษฎีบทของเกาส์-มาร์คอฟ หรือ Gauss-Markov Theorem). สำหรับ  $y$  ที่มาจาก DGP (5.20) นั้น ถ้าหาก  $e$  เป็นสัญญาณรบกวนขาว ที่มีค่าเฉลี่ยเป็นศูนย์ และมีเมทริกซ์ความแปรปรวนร่วมเป็น  $\sigma^2 I$  แล้ว ตัวประมาณ  $\beta_{ls} = (X^T X)^{-1} X^T y$  จะเป็นตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด Best Linear Unbiased Estimator (BLUE)

**พิสูจน์.** ในที่นี้ตัวประมาณที่ดีที่สุด (best) คือ ตัวประมาณที่มีเมทริกซ์ความแปรปรวนร่วมต่ำสุด (แบบเมทริกซ์บวกแน่นอน) ในการพิสูจน์จะสมมติให้มีตัวประมาณเชิงเส้นอื่นใดๆ กำหนดให้เป็น  $\hat{\beta}$  ดังนั้น  $\hat{\beta}$  จะอยู่ในรูป  $\hat{\beta} = B y$  หาก  $\hat{\beta}$  ก็เป็นตัวประมาณค่าไม่เอนเอียง จะได้ว่า  $\mathbf{E}[\hat{\beta}] = B \mathbf{E}[y] = B \mathbf{E}[X\beta + e] = B X \beta + 0 = \beta$  ดังนั้น จะได้เงื่อนไขว่า  $B X = I$  เมื่อคำนวณเมทริกซ์ความแปรปรวนร่วมของแต่ละตัวประมาณ จะได้ว่า

$$\begin{aligned} \mathbf{cov}(\hat{\beta}) &= \mathbf{E}[(B y - \beta)(B y - \beta)^T] \\ &= \mathbf{E}[(B X \beta + B e - \beta)(B X \beta + B e - \beta)^T] \\ &= \mathbf{E}[B e e^T B^T] = \sigma^2 B B^T \\ \mathbf{cov}(\beta_{ls}) &= \sigma^2 B X (X^T X)^{-1} X^T B^T \quad (\text{ใช้ว่า } B X = I) \end{aligned}$$

ดังนั้น เมื่อเทียบเมทริกซ์ความแปรปรวนร่วม

$$\mathbf{cov}(\hat{\beta}) - \mathbf{cov}(\beta_{ls}) = \sigma^2 B (I - X (X^T X)^{-1} X^T) B^T$$

และใช้คุณสมบัติของเมทริกซ์ภาพฉายเชิงตั้งฉาก  $P := X (X^T X)^{-1} X^T$  ว่า  $\|P x\| \leq \|x\|$  หรือ  $x^T x - x^T P x \geq 0$  ซึ่งหมายความว่า  $I - P \succeq 0$  ทำให้สรุปได้ว่า

$$\mathbf{cov}(\hat{\beta}) - \mathbf{cov}(\beta_{ls}) = \sigma^2 B (I - P) B^T \succeq 0 \text{ เสมอ สำหรับ } B \text{ ใดๆ}$$

เมื่อเทียบกับตัวประมาณเชิงเส้นไม่เอนเอียงตัวอื่นทั้งหมดแล้ว  $\beta_{ls}$  จะมีค่า  $\mathbf{cov}(\beta_{ls})$  ที่น้อยที่สุด ■

หากสมมติฐานภาวะแปรปรวนเท่ากันในทฤษฎีบท 5.1 ไม่เป็นจริง จะแสดงให้เห็นว่า ตัวประมาณ LS จะยังมีสมบัติ BLUE หรือไม่ หากไม่ใช่ ตัวประมาณ BLUE ควรีรูปแบบอย่างไร



ซึ่งหากความน่าจะเป็นของตัวแปร  $F_{1,N-1}$  เกินกว่าค่าสถิติที่คำนวณจากข้อมูลตัวอย่างนั้นน้อยกว่า  $\alpha$  ที่กำหนด เราจะปฏิเสธ  $H_0$  ในชุดคำสั่งแก้ปัญหาคาดก้อยทั่วไปนั้น จะรายงานค่า  $F$  ของแบบจำลองเต็ม กับแบบจำลองลดรูปที่เป็นแบบจำลองค่าคงที่

**ตัวอย่าง 5.5** เราสร้าง DGP ที่เป็น  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + e$  นั่นคือมีตัวแปรต้นคือ ค่าคงที่  $x_1$  และ  $x_2$  ที่ส่งผลในการอธิบาย  $y$  และมีค่าพารามิเตอร์จริงเป็น  $\beta = (1, 1, 1)$  ในขณะที่ผู้ใช้สร้างแบบจำลองถดถอย คิดว่าตัวแปร  $y$  จะมี 4 ตัวแปรทำนาย จึงได้แก้ปัญหาคาดก้อยแบบจำลอง  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$  ซึ่งทั้งนี้  $x_3$  อาจจะเป็นตัวแปรทำนายที่ผู้ใช้เข้าใจผิด ว่าส่งผลต่อ  $y$  จึงเก็บข้อมูลมาใช้ประมาณแบบจำลอง สมมติว่า ข้อมูลมี 20 ตัวอย่าง เมื่อเราใช้ชุดคำสั่ง fitlm ใน MATLAB ได้ผลลัพธ์ดังแสดงด้านล่าง คอลัมน์ Estimate คือค่า  $\beta_{ls}$  ที่ประมาณได้ SE คือความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ (5.25) ส่วน tStat คือค่าสถิติ  $t$  ที่ทดสอบความสำคัญของ  $\beta_k$  และแสดงค่า  $p$  ที่สอดคล้องกัน ในตัวอย่างนี้ หากกำหนดให้ระดับนัยสำคัญเท่ากับ  $\alpha = 0.05$  จะเห็นว่า ค่า  $p$  ที่น้อยกว่า  $\alpha$  จะสอดคล้องกับตัวแปรต้นที่เป็นค่าคงที่  $x_1, x_2$  เราจึงปฏิเสธ  $H_0$  (ปฏิเสธว่าตัวแปรเหล่านั้นไม่มีนัยสำคัญ) และจะเห็นว่าค่า  $p$  ของ  $\beta_3$  ซึ่งอ่านได้ 0.23487 ซึ่งมีค่ามากกว่า  $\alpha$  เราจึงไม่มีหลักฐานที่จะปฏิเสธ  $H_0$  (นั่นคือยอมรับว่า  $\beta_3 = 0$ ) ทั้งนี้ ในตอนท้ายของผลลัพธ์ค่าสถิติ  $F$  เป็นการบอกเทียบแบบจำลองเต็ม กับแบบจำลองค่าคงที่ ที่มีค่าเท่ากับ 16.1 และค่า  $p$  เท่ากับ  $4.35 \times 10^{-5}$  จึงบ่งชี้ว่าเราปฏิเสธ  $H_0$  นั่นคือ ค่าสถิติ  $F$  ที่คำนวณจากข้อมูลตัวอย่างนั้นมีค่าสูงมากพออย่างมีนัย อันแปลว่าการใช้แบบจำลองเต็มมีความผิดพลาดที่ต่างกับแบบจำลองค่าคงที่อย่างมีนัยสำคัญ เราจึงเลือกใช้แบบจำลองเต็มในตัวอย่างนี้

Linear regression model:				
$y \sim 1 + x_1 + x_2 + x_3$				
Estimated Coefficients:				
	Estimate	SE	tStat	pValue
(Intercept)	0.99678	0.22554	4.4196	0.0004294
x1	0.98566	0.25086	3.9292	0.0011979
x2	0.64226	0.27465	2.3384	0.032676
x3	-0.30756	0.24915	-1.2344	0.23487

Number of observations: 20, Error degrees of freedom: 16  
 Root Mean Squared Error: 0.988  
 R-squared: 0.751, Adjusted R-Squared: 0.704  
 F-statistic vs. constant model: 16.1, p-value = 4.35e-05

## 5.7 การเลือกตัวแปรทำนายที่สำคัญต่อตัวแปรตอบสนอง

การใช้แบบจำลองเชิงเส้นถดถอยหลายตัวแปรนั้น อันดับแรก ต้องเลือกตัวแปรทำนาย  $x_1, \dots, x_n$  ที่จะมาใช้อธิบาย  $y$  ซึ่งอาศัยความรู้เบื้องต้นของความสัมพันธ์ของตัวแปรในงานประยุกต์นั้นๆ ตัวอย่างเช่น หากปัญหาคือการทำนายกำลังผลิตไฟฟ้าจากเซลล์สุริยะ ( $y$ ) เราพบว่าแบบจำลองคณิตศาสตร์ของเซลล์สุริยะที่ใช้หลักการทางฟิสิกส์และวงจรไฟฟ้ามีตัวแปรต้นหลัก คือ ความเข้มแสงอาทิตย์ ( $x_1$ ) และ อุณหภูมิของแผงเซลล์ ( $x_2$ ) เมื่อใช้ความรู้ด้านอุณหุนิยมวิทยาเพิ่มเติมว่า ปริมาณความเข้มแสงอาทิตย์ยังขึ้นกับตัวแปรสภาพอากาศอื่นๆ เช่น ความหนาแน่นของเมฆ อุณหภูมิบรรยากาศ ความชื้นสัมพัทธ์ และ

ความเข้มแสงอาทิตย์ของเวลาในอดีต เป็นต้น ดังนั้น ขั้นตอนแรกคือการเลือกตัวแปรทำนายที่คาดว่าจะเกี่ยวข้อง ในการอธิบายตัวแปรตอบสนอง คำถามสำคัญที่ถัดมาคือ ตัวแปรทำนายต่างๆ เหล่านี้จำเป็นทั้งหมดหรือไม่ หรือการใช้แค่บางตัวแปรก็เพียงพอในการอธิบาย  $y$  แล้ว กระบวนการนี้เทียบได้กับ การคัดเลือกเวกเตอร์คุณลักษณะ (feature selection) ในปัญหาการเรียนรู้ของเครื่อง การเลือกตัวแปรทำนายในปัญหาถดถอย มีขั้นตอนพื้นฐาน 2 แบบ [HTF09] อันได้แก่ การเลือกเซตย่อยของตัวแปรที่ดีที่สุด (best subset selection) และการถดถอยทีละขั้น (stepwise regression) ในวิธีเหล่านี้เราอ้างอิงถึงแบบจำลองสองแบบ นั่นคือ แบบจำลองว่าง (null model) ซึ่งหมายถึงแบบจำลองที่ไม่มีตัวแปรทำนายอยู่เลย (มีเพียงแต่เทอมค่าคงที่) และแบบจำลองเต็ม (full model) อันหมายถึงแบบจำลองที่มีทุกตัวแปรทำนาย สมมติให้เป็น  $n$  ตัวแปร

การเลือกเซตย่อยที่ดีที่สุดมุ่งหา  $k$  ตัวแปรทำนายที่ดีที่สุด (โดยที่  $k < n$ ) ดังนั้น สำหรับ  $k$  ค่าหนึ่ง จะมีรูปแบบของแบบจำลองทั้งหมดเป็น  $n$  เลือก  $k$  แบบ ดังนั้นจำนวนแบบจำลองที่เป็นไปได้ทั้งหมดคือ  $\sum_{k=1}^n \binom{n}{k} = 2^n$  แบบ ตัวอย่างเช่น  $n = 3$  นั่นคือมี  $x_1, x_2, x_3$  แบบจำลองย่อยทั้งหมดจึงมีตัวแปรทำนายเป็นไปได้ในแต่ละกรณี ดังนี้

$$\{\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}$$

ในแต่ละแบบจำลองย่อย เราจะประมาณแบบจำลองถดถอย และคำนวณคะแนนเกณฑ์การเลือกแบบจำลอง (model selection criterion) เช่น เกณฑ์สารสนเทศของอะกะอิเกะ Akaike Information Criterion (AIC) หรือ เกณฑ์สารสนเทศของเบส์ Bayesian Information Criterion (BIC) และเลือกแบบจำลองย่อยที่ทำให้เกณฑ์การเลือกแบบจำลองนั้นดีที่สุด

---

**Algorithm 5.1** การเลือกเซตย่อยที่ดีที่สุด: เลือกเซตย่อยของ  $\{x_1, x_2, \dots, x_n\}$

---

```

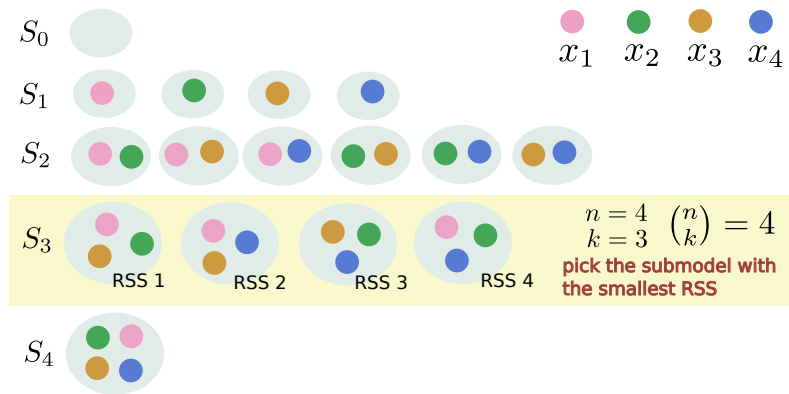
for  $k = 1 : n$  do
   $m = \binom{n}{k}$ 
  for  $j = 1 : m$  do
    เลือกเซตย่อยของตัวแปรทำนายจาก  $\{x_1, x_2, \dots, x_n\}$  ที่มีจำนวน  $k$  ตัวแปร
    ประมาณแบบจำลองถดถอย  $M_j$  ที่มีตัวแปรทำนายตามทีเลือกไว้
    คำนวณตัวชี้วัดคุณภาพแบบจำลอง  $M_j$  เช่น  $R^2$  หรือ RSS
  end for
  เลือกแบบจำลองที่มีตัวชี้วัดคุณภาพดีที่สุดจากแบบจำลองย่อย  $m$  แบบจำลอง และเรียกว่า  $S_k$ 
end for
คำนวณเกณฑ์ AIC หรือ BIC ของ  $S_1, S_2, \dots, S_n$ 
เลือกแบบจำลองที่มีเกณฑ์การเลือกแบบจำลองต่ำที่สุด

```

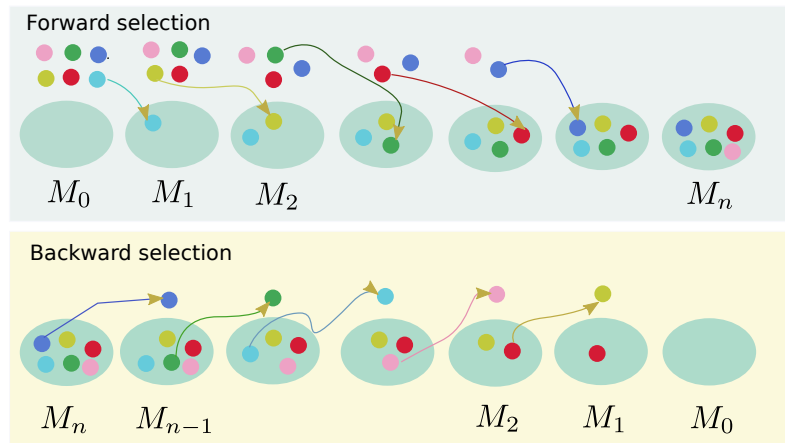
---

ตัวอย่างรูป 5.4 คือการแจกแจงแบบจำลองย่อยกรณี  $n = 4$  เพื่อคำนวณตามขั้นตอนดังกล่าวข้างต้น จะเห็นว่า  $m$  คือจำนวนแบบจำลองที่เป็นไปได้อันมีจำนวนตัวแปรทำนายเท่ากับ  $k$  ในบรรดา  $m$  แบบจำลองนี้ ทุกตัวล้วนมีความซับซ้อนเท่ากัน (เพราะมี  $k$  ตัวแปร) ลูบย่อยจึงคำนวณตัวชี้วัดคุณภาพของ  $M_j$  เพื่อจะเปรียบเทียบว่า เซตของตัวแปรทำนายที่เลือกไว้แบบใดจะอธิบายข้อมูลได้ดีกว่ากัน จากนั้น ลูบออกจึงนำตัวเลือก  $S_1, S_2, \dots, S_n$  ที่มีความต่างกันทั้งความซับซ้อนและคุณภาพของแบบจำลองด้วยเกณฑ์ AIC หรือ BIC (รายละเอียดทางทฤษฎีจะกล่าวอีกครั้งในหัวข้อ 10.2.1)

วิธีการหาแบบจำลองย่อยที่ดีที่สุด มีจำนวนครั้งของการคำนวณที่สูงเป็นฟังก์ชันเลขชี้กำลังของ  $n$  หาก  $n$  มีค่ามาก เช่น  $n = 13, 2^n = 8192$  จำนวนรอบในการคำนวณก็สูงขึ้นตาม อีกทางเลือกหนึ่ง



รูป 5.4: การเลือกเซตย่อยที่ดีที่สุดของตัวแปรทำนายในปัญหาถดถอย



รูป 5.5: การเลือกตัวแปรทำนายแบบเดินหน้าและถอยหลังในปัญหาถดถอย

คือการใช้วิธีการถดถอยทีละขั้น (stepwise regression) ที่เริ่มต้นจากแบบจำลองที่ไม่มีตัวแปรทำนายชุดหนึ่ง แล้วค่อยๆ เพิ่มหรือลดตัวแปรทำนายตัวใหม่ที่ละขั้น วิธีนี้มีสมบัติแบบเกือบเหมาะที่สุด (sub-optimal) เพราะไม่ได้หาแบบจำลองย่อยที่ดีที่สุด การถดถอยทีละขั้นแบ่งออกเป็นสองวิธีตามรูป 5.5 ดังนี้ โดยมีตัวอย่างในรูปคือ  $n = 6$

การเลือกเดินหน้า (forward selection) คือ การเริ่มต้นด้วยแบบจำลองว่าง ที่ไม่มีตัวแปรทำนายโดยอยู่เลย แต่ละขั้นตอนจะเพิ่มตัวแปรทำนายใหม่จากจำนวนตัวแปรที่เหลือที่ทำให้แบบจำลองมี RSS น้อยที่สุด เมื่อให้  $M_k$  คือแบบจำลองของขั้นตอน  $k$  ที่มีจำนวนตัวแปรทำนาย  $k$  ตัวแปร ณ ขั้นตอนนี้ จะมีตัวแปรให้เลือกได้เพิ่มจำนวน  $n - k$  ตัวเลือก เมื่อเลือกตัวแปรใหม่ที่ให้ RSS น้อยที่สุดแล้วนำมาเพิ่มจะเรียกว่าแบบจำลอง  $M_{k+1}$  จากนั้นจะเลือกแบบจำลองในบรรดา  $M_0, M_1, \dots, M_n$  ที่ให้ค่าเกณฑ์การเลือกแบบจำลองเช่น AIC/BIC ที่มีค่าต่ำสุด แต่ละขั้นตอนจะให้แบบจำลองซ้อนใน (nested model) เพราะในขั้นถัดไป แบบจำลองจะมีตัวแปรที่มาจากขั้นตอนก่อนหน้าเสมอ

การเลือกถดถอยหลัง (backward selection) คือการเริ่มต้นด้วยแบบจำลองเต็ม ที่มีตัวแปรทำนายทุกตัวอยู่ แต่ละขั้นจะลดตัวแปรทำนายทีละตัวที่มีผลน้อยที่สุดต่อคะแนนความเข้ากันของแบบจำลอง ณ ขั้นตอน  $k$  เรามีแบบจำลอง  $M_k$  ที่มีตัวแปรทำนาย  $k$  ตัว และมีตัวเลือก  $k$  ตัวแปรที่จะละทิ้ง

## บทที่ 6

# การขยายปัญหาค่าลิ่งสองน้อยสุดเชิงเส้น

งานประยุกต์ของปัญหาค่าลิ่งสองน้อยสุดเชิงเส้น (LS) จากบทที่ 5 มีสมมติฐานเกี่ยวกับพารามิเตอร์ของปัญหาและพารามิเตอร์ของแบบจำลองที่ต่างกันไป สมมติฐานเหล่านั้นสามารถจัดให้อยู่ในรูปของเงื่อนไขบังคับในปัญหาการหาค่าเหมาะที่สุดได้ งานประยุกต์อาจมีสมมติฐานเกี่ยวกับตัวประมาณ LS เช่น เวกเตอร์ผลเฉลยมีศูนย์เป็นจำนวนมากเพื่อบังคับถึงแบบจำลองที่ไม่ซับซ้อน หรือ การที่เวกเตอร์ผลเฉลยมีค่าเท่ากันเป็นกลุ่ม เราสามารถจัดให้สมมติฐานเหล่านี้ ถูกสะท้อนในรูปแบบปัญหาประมาณได้ โดยการเพิ่มฟังก์ชันลงโทษในรูปแบบปัญหา LS อันเป็นหัวข้อเรื่องการทำให้ตามเกณฑ์ (regularization) และเป็นปัญหาการประมาณเชิงสถิติแบบหนึ่ง ในปัญหา LS ที่มีพารามิเตอร์ที่ไม่ทราบค่า เราก็สามารถจัดรูปแบบปัญหาให้ครอบคลุมความไม่แน่นอนของพารามิเตอร์เหล่านั้น ด้วยการใช่วิธีเชิงสถิติหรือพิจารณาสมรรถนะผลการประมาณกรณีเลวร้ายที่สุดได้ กล่าวโดยรวม เนื้อหาบทนี้พิจารณาการต่อยอดปัญหา LS ในแง่มุมต่างๆ ซึ่งผลลัพธ์สุดท้ายจะอยู่ในรูปแบบปัญหาการหาค่าเหมาะที่สุดแบบคอนเวกซ์ ที่สามารถหาผลตอบเชิงตัวเลขได้จากหลายขั้นตอนวิธีเชิงตัวเลขที่มีอยู่ และเรายังกล่าวถึงวิธีกำลังสองน้อยสุดไม่เชิงเส้น ซึ่งเป็นปัญหาการหาค่าเหมาะที่สุดไม่เชิงเส้น (nonlinear optimization)

ในแบบจำลองเชิงเส้นที่แสดงในบทที่ 5 เราใช้การประมาณ  $y = X\beta$  โดยที่  $\beta$  เป็นพารามิเตอร์ที่ต้องการประมาณ และเรียกผลเฉลยของปัญหาแบบไม่มีเงื่อนไขใน  $\beta$  นี้ว่าได้จากวิธีกำลังสองน้อยสุดสามัญ Ordinary Least-Squares (OLS) เพื่อแยกให้เห็นความต่างของผลเฉลยปัญหาต่อยอดของ LS สำหรับบทนี้ เราจะเปลี่ยนสัญลักษณ์เป็นการใช้แบบจำลอง  $y = Ax$  โดยที่  $A$  เป็นเมทริกซ์ถดถอย และ  $x$  จะเป็นพารามิเตอร์ของแบบจำลอง เพื่อให้  $x$  เป็นตัวแปรของปัญหาประมาณ และสัญลักษณ์คณิตศาสตร์ของรูปแบบปัญหาการหาค่าเหมาะที่สุดที่ได้นั้นสอดคล้องกับแหล่งศึกษาอื่นทั่วไป ปัญหาที่นำมาพิจารณาในบทนี้เป็นกลุ่มปัญหาคอนเวกซ์

**นิยาม 6.1.** เราจะกล่าวว่า  $C$  เป็นเซตคอนเวกซ์ หากสำหรับ  $x, y \in C$  ใดๆ และ  $0 \leq \theta \leq 1$  เราจะได้ว่า  $\theta x + (1 - \theta)y \in C$  กล่าวคือ เซตที่ผลรวมคอนเวกซ์ของสองจุดใดๆ ในเซตยังคงอยู่ในเซตนั้น

**นิยาม 6.2.** กำหนดให้  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  เราจะกล่าวว่า  $f$  เป็นฟังก์ชันคอนเวกซ์ ถ้าหากโดเมนของ  $f$  (ใช้สัญลักษณ์เป็น  $\text{dom } f$ ) เป็นเซตคอนเวกซ์ และ

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y), \quad 0 \leq \theta \leq 1, \quad \forall x, y \in \text{dom } f$$

ตัวอย่างที่เกี่ยวข้องคือ  $\|x\|$ ,  $\|y - Ax\|_2$  หรือ  $\|y - Ax\|_2^2$  ล้วนแต่เป็นฟังก์ชันคอนเวกซ์ ปัญหา minimize  $f(x)$  โดยที่  $x$  ต้องอยู่ในเซตคอนเวกซ์ เรียกว่าปัญหาคอนเวกซ์ ซึ่งมีสมบัติของการมีค่าต่ำสุดวงกว้าง (global minimum) ผู้อ่านสามารถทบทวนรายละเอียดเกี่ยวกับเซตคอนเวกซ์ ฟังก์ชันคอนเวกซ์ ปัญหาคอนเวกซ์ และตัวอย่างได้จาก [BV04]

## 6.1 ปัญหากำลังสองน้อยสุดเชิงเส้นที่มีเงื่อนไขบังคับ

กำหนดให้  $C$  เป็นเซตใดๆ การมีเงื่อนไขบังคับในผลเฉลย LS สามารถจัดในรูปแบบปัญหา

$$\begin{aligned} & \text{minimize} && \|Ax - y\|_2 \\ & \text{subject to} && x \in C \end{aligned} \quad (6.1)$$

ที่มีตัวแปรคือ  $x \in \mathbf{R}^n$  ความหมายของการบังคับให้  $x \in C$  มีได้หลากหลายเช่น การมีสมมติฐานเฉพาะในเวกเตอร์  $x$  หรือการเจาะจงให้  $x \in C$  เพื่อให้การประมาณ  $y$  ด้วยฟังก์ชันเชิงเส้น  $y \approx Ax$  นั้นสมเหตุสมผล เป็นต้น หัวข้อนี้จะยกตัวอย่างเซตคอนเวกซ์  $C$  ที่มีความหมายในงานประยุกต์ต่างๆ

**ความไม่เป็นลบ** หรือ nonnegative constraints คือการที่  $x_i \geq 0$  ทุก  $i$  หรือ บรรยายด้วย

$$C = \{ x \in \mathbf{R}^n \mid x \geq 0 \}$$

(เราใช้สัญลักษณ์  $\geq$  เพื่อหมายถึงเครื่องหมายมากกว่าสำหรับทุกสมาชิกของเวกเตอร์) หากเวกเตอร์  $x$  ในแบบจำลองมีความหมายถึงปริมาณที่เป็นบวกทางกายภาพ เช่น กำลังไฟฟ้า ปริมาณน้ำ หรือ ราคา เราอาจพิจารณาวิธีอย่างง่ายเพื่อบังคับให้  $x \geq 0$  ว่าเป็นการแก้ปัญหาค่า OLS มาก่อน แล้วปรับค่า  $x_i$  ที่มีค่าลบให้กลายเป็นศูนย์ แต่คำตอบที่ได้ย่อมสูญเสียความเป็นคำตอบที่เหมาะสมที่สุดไป การจัดรูปแบบปัญหาตาม (6.1) จึงเป็นการบังคับให้  $x$  มีความหมายตามต้องการ และเลือกผลเฉลยที่ดีที่สุดตามฟังก์ชันจุดประสงค์แบบกำลังสองน้อยสุด จากที่  $Ax = a_1x_1 + \dots + a_nx_n$  โดยที่  $a_k$  คือคอลัมน์ที่  $k$  ของ  $A$  และการที่  $x_i \geq 0$  จึงเห็นว่า  $Ax$  คือผลรวมถ่วงน้ำหนักบวก (หรือ conic combination) ของคอลัมน์ใน  $A$  ดังนั้นปัญหา OLS เดิมคือการหาภาพฉายของ  $y$  ลงบน  $\mathcal{R}(A)$  แต่ความหมายของปัญหา (6.1) คือ การหาภาพฉายของ  $y$  ลงบนกรวย (cone) ที่สร้างจากคอลัมน์ของ  $A$

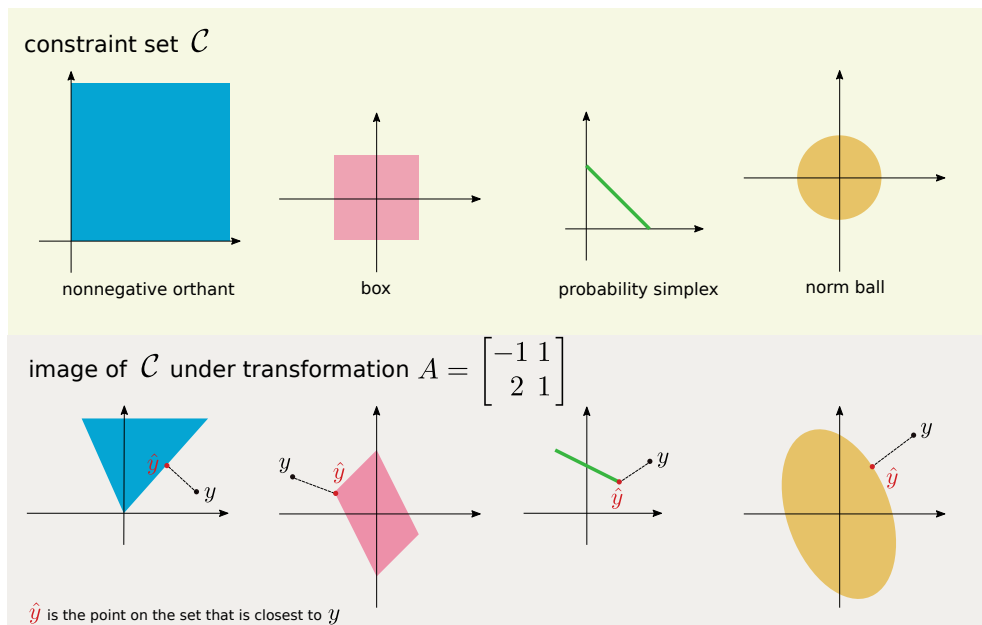
**การอยู่ในช่วงมีขอบเขต** คือการที่  $x_i$  ทุก  $i$  มีค่าขอบบนและขอบล่าง

$$C = \{ x \in \mathbf{R}^n \mid l \leq x \leq u \} \quad (6.2)$$

โดยที่  $l, u \in \mathbf{R}^n$  เป็นเวกเตอร์ขอบเขตที่กำหนดโดยผู้ใช้ ตัวอย่างของเงื่อนไข เช่น  $x_i$  เป็นระยะความยาวของชิ้นงานตอนที่  $i$  ซึ่งต้องอยู่ในช่วงระยะไม่เกินตามข้อกำหนด เซต  $C$  มีรูปร่างเป็นกล่อง ความหมายของ  $Ax$  คือภาพของกล่องที่ผ่านการแปลงเชิงเส้น  $A$  ดังนั้น ความหมายของปัญหา (6.1) คือ การหาภาพฉายของ  $y$  บนภาพของกล่องที่ผ่านการแปลง  $A$  ซึ่งจะเป็รูปทรงหลายหน้า (polyhedron)

**การแจกแจงความน่าจะเป็น** คือ การที่เวกเตอร์  $x$  มีสมบัติเหมือนฟังก์ชันการแจกแจงความน่าจะเป็น (probability mass function) นั่นคือ  $x_i \geq 0$ ,  $\sum_i x_i = 1$  ที่บรรยายได้ด้วย

$$C = \{ x \in \mathbf{R}^n \mid x \geq 0, \quad \mathbf{1}^T x = 1 \} \quad (6.3)$$



รูป 6.1: ตัวอย่างรูปร่างของเซต  $C$  แบบต่างๆ ใน  $\mathbf{R}^2$  อันเป็นเซตเงื่อนไขของปัญหา (6.1) และภาพของ  $C$  ภายใต้การแปลง  $A$

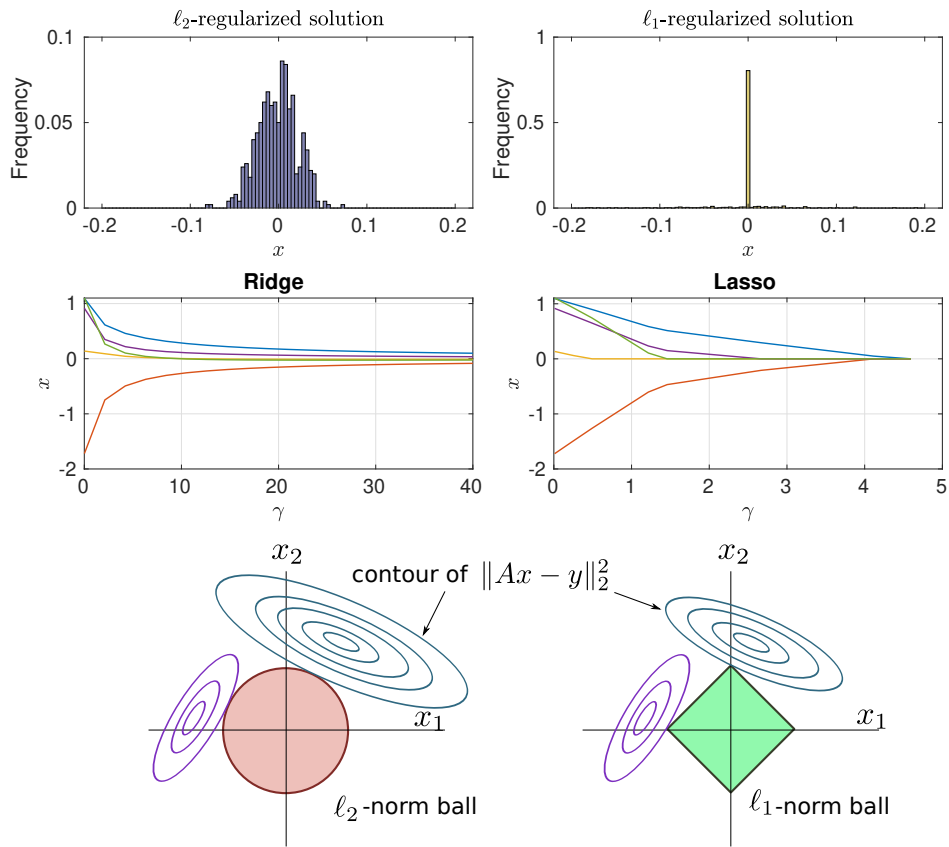
เซตนี้มีชื่อเรียกว่าซิมเพลกซ์ความน่าจะเป็น (probability simplex) ในบางงานประยุกต์  $x_i$  มีความหมายเป็นสัดส่วนเทียบกับหนึ่ง เช่น  $x_i$  คือสัดส่วนการลงทุนในหุ้น  $i$  หรือ สัดส่วนอาหารที่บริโภคในหมวดหมู่  $i$  หรือ สัดส่วนการผลิตกำลังไฟฟ้าจากโรงผลิตกำลังชนิด  $i$  ในปัญหาประมาณ เราต้องตัดสินใจว่า จะลงทุนหรือเกลี้ยสัดส่วนนี้อย่างไร สำหรับเซต  $C$  นี้ จะได้ว่า  $Ax$  มีความหมายเป็นผลรวมคอนเวกซ์ของคอลัมน์ใน  $A$  ดังนั้น ความหมายของปัญหา (6.1) คือการประมาณ  $y$  ที่ดีที่สุด (วัดด้วยนอร์ม-2) ที่ประมาณด้วยการใช้ผลรวมคอนเวกซ์ของคอลัมน์ใน  $A$

**การอยู่ในนอร์มบอล** คือการที่  $x$  อยู่ในย่านใกล้กับเวกเตอร์  $x_0$  ในระยะทางไม่เกิน  $d$

$$C = \{ x \in \mathbf{R}^n \mid \|x - x_0\| \leq d \} \quad (6.4)$$

พารามิเตอร์  $x_0 \in \mathbf{R}^n$  และระยะ  $d > 0$  เป็นสิ่งที่ผู้ใช้กำหนด ในที่นี้  $x_0$  มีความหมายเป็นค่าคาดเดา เริ่มต้นว่า  $x$  ควรีค่าไม่ห่างจาก  $x_0$  มากนัก ตัวอย่างเช่น  $Ax$  มีความหมายว่าเป็นการใช้แบบจำลองเชิงเส้น การประมาณดังกล่าว มักจะตามมาด้วยสมมติฐานบางอย่างเสมอ เช่น การประมาณระบบไม่เชิงเส้นให้เป็นเชิงเส้น (linearization) ที่ผู้ใช้ต้องกำหนดจุดทำงาน (หรือจุดสมดุล) เพื่อที่แบบจำลองเชิงเส้นที่ประมาณได้มีความสมเหตุสมผล หรือใกล้เคียงระบบจริงรอบจุดทำงานนั้น ด้วยเหตุนี้ จึงได้เงื่อนไขบังคับว่า  $x$  ต้องอยู่ในย่านใกล้เคียงกับ  $x_0$  เพื่อความเหมาะสมในการเลือกแบบจำลองดังกล่าว ในเงื่อนไขนอร์มบอลนั้น เราอาจใช้นอร์ม-1 นอร์ม-2 หรือนอร์มใดๆ ตามแต่ผู้ใช้จะเลือกตามรูปร่างเชิงเรขาคณิตของย่านใกล้เคียงนั้น

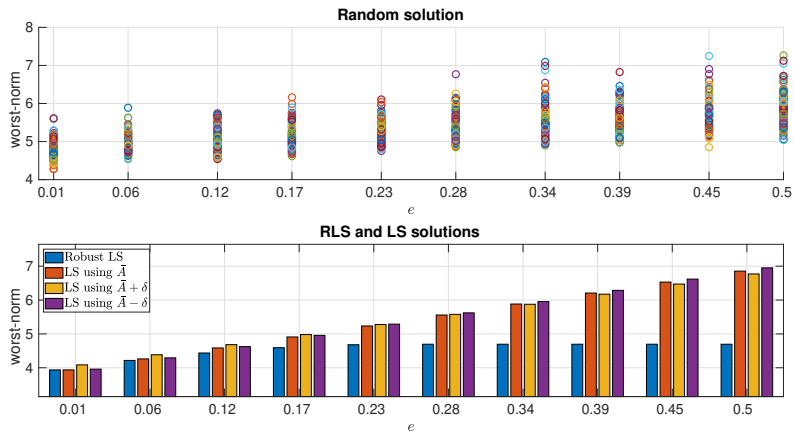
รูป 6.1 แสดงตัวอย่างของเซต  $C$  และภาพของ  $C$  ภายใต้การแปลง  $A$  การหาคำตอบของปัญหา (6.1) คือการหาจุดในภาพของ  $C$  ที่ใกล้กับเวกเตอร์  $y$  มากที่สุด นั่นคือ เราหา  $\hat{x}$  ที่เป็นผลเฉลยทำให้  $\hat{y} = A\hat{x}$  มีค่า  $\|y - \hat{y}\|_2$  น้อยที่สุด



รูป 6.3: แฉวบน ฮิสโทแกรมของผลเฉลยจากปัญหา LS ที่ใช้ฟังก์ชันลงโทษนอร์ม-2 และนอร์ม-1 แฉวกลาง ทางเดินของผลเฉลยปัญหา LS ที่ใช้ฟังก์ชันลงโทษนอร์ม-2 และนอร์ม-1 เมื่อแปรค่าพารามิเตอร์ลงโทษ (แต่ละสีคือแต่ละค่า  $x_i(\gamma)$ ) แฉวล่าง ปัญหา (6.10) เมื่อมีเงื่อนไขบังคับเป็น  $l_2$ -นอร์มบอล และ  $l_1$ -นอร์มบอล โดยที่คอนทัวร์ของ  $\|Ax - y\|_2^2$  ที่ค่า  $(A, y)$  สองชุดแสดงด้วยสีฟ้า และสีม่วง

และเรียกว่า**ปัญหากำลังสองน้อยสุดที่ทำให้ตามเกณฑ์ด้วยนอร์ม-1** หรือ **แลสโซ (lasso)** [HTF09, §3.4] [HTW15] ผลเฉลยของ (6.9) คือการหาเวกเตอร์เบาบาง  $x$  โดยที่ดีกรีความเบาบางถูกกำหนดโดยการเลือกพารามิเตอร์ลงโทษ  $\gamma$  และในขณะเดียวกันผลเฉลย  $x$  นี้ให้การประมาณ  $Ax$  ที่ใกล้เคียงกับ  $y$  ตัวอย่างการจำลองแก้ปัญหาลาสโซกับปัญหาคถอยแบบบริจด์ ด้วยการใส่  $A \in \mathbf{R}^{m \times n}, y \in \mathbf{R}^m$  เมื่อ  $m = 100, n = 500, \gamma = 0.2$  เพื่อดูการกระจายตัวของสมาชิกในเวกเตอร์ผลเฉลยแสดงในรูป 6.3 (แฉวบน) จะเห็นว่าผลเฉลยบริจด์นั้น ขนาดของ  $x_i$  จะมีการกระจายตัวในช่วง  $(-0.1, 0.1)$  คือมีทั้งสมาชิกที่มีค่าน้อยเกือบศูนย์และมีค่าไม่เป็นศูนย์ แต่ผลเฉลยแลสโซมีสมาชิกจำนวนมากกระจุกตัวรอบศูนย์ สำหรับอีกตัวอย่างจำลองที่ใช้  $n = 5$  และเราแปรค่า  $\gamma$  ตั้งแต่ศูนย์ไปจนถึงค่าที่ทำให้  $x$  เป็นศูนย์ทั้งหมด รูป 6.3 (แฉวกลาง) แสดงทางเดินผลเฉลยที่มี 5 สมาชิกของปัญหาแลสโซและบริจด์สำหรับผลเฉลยแลสโซนั้น เมื่อค่า  $x_i$  ไต่เท่ากับศูนย์ที่  $\gamma$  หนึ่งๆ นั้น หลังจาก  $\gamma$  เพิ่มมากขึ้นเรื่อยๆ  $x_i$  ตำแหน่งนั้นก็ยังคงศูนย์ไปเรื่อยๆ และพบว่าสมาชิกของ  $x$  นั้นเป็นศูนย์แน่ชัด สำหรับทางเดินผลเฉลยบริจด์นั้น  $x_i$  จะไม่ได้เป็นศูนย์แน่ชัด แต่จะค่อยๆ มีค่าลดลงจนเข้าสู่หาศูนย์

เราจะเห็นว่าฟังก์ชันจุดประสงค์ของ (6.9) เป็นฟังก์ชันคอนเวกซ์ แต่การมีเทอม  $|\cdot|$  ทำให้ฟังก์ชัน



รูป 6.8: นอร์ม-2 ของความคลาดเคลื่อนเลวร้ายที่สุด (6.20) เมื่อค่า  $x$  ที่ใช้เป็นค่าต่างๆ

ใช้ค่า  $x$  สุ่มใดๆ อันหมายถึงการไม่ได้ใช้รูปแบบปัญหาประมาณ (ที่เหมาะสมที่สุดในหลักการหนึ่ง) เพื่อให้ได้ตัวเลือก  $x$  หนึ่ง ย่อมทำให้ค่านอร์มเลวร้ายที่สุดที่เป็นไปได้อยู่ในช่วงกว้างมาก

**รูปแบบปัญหาสโตแคสติก** หากเรามีสมมติฐานว่า  $A$  เป็นเมทริกซ์สุ่ม และพิจารณารูปแบบที่ง่ายที่สุดคือ  $A$  เป็นตัวแปรสุ่มอันเกิดจากการบวก (หรือมองว่าเป็นการแปลงเชิงเส้นจากตัวแปรสุ่มอื่น) ดังนี้

$$A = \bar{A} + U \tag{6.22}$$

โดยที่  $\bar{A}$  เป็นค่าเชิงกำหนด (deterministic) และ  $U$  เป็นเมทริกซ์สุ่มที่มีค่าคาดหวังเป็นศูนย์ ดังนั้นความหมายของ  $\bar{A}$  จึงเป็นค่าเฉลี่ยของ  $A$  ในเบื้องต้นนั้น หาก  $A$  เป็นเมทริกซ์สุ่ม ปริมาณใดๆ ที่เป็นฟังก์ชันของ  $A$  ย่อมเป็นตัวแปรสุ่ม ดังนั้นเราจึงควรพิจารณา  $\|Ax - y\|_2$  ที่เป็นตัวแปรสุ่มว่าจะแปรปรวนน้อยอย่างไรเพื่อให้เราเลือก  $x$  ที่เหมาะสมได้ แนวทางที่มักนิยมใช้ คือเราเปลี่ยนจุดประสงค์ของปัญหาประมาณมาเป็นการหา  $x$  ที่ทำให้ ค่าคาดหวัง ของ  $\|Ax - y\|_2$  น้อยที่สุด

$$\text{minimize}_x \mathbf{E}\|Ax - y\|_2^2 \tag{6.23}$$

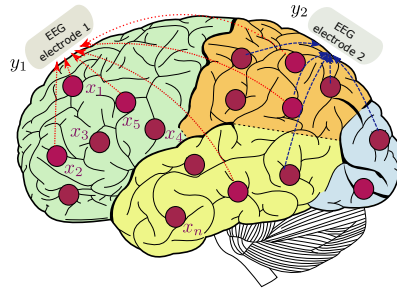
การจะหา  $x$  ที่เป็นผลเฉลยของปัญหากำลังสองน้อยสุดแบบสโตแคสติกนี้ จึงขึ้นกับว่าค่าคาดหวังของ  $\|Ax - y\|_2$  นั้นเป็นฟังก์ชันอย่างไรของ  $x$  ดังนี้

$$\begin{aligned} \mathbf{E}\|Ax - y\|_2^2 &= \mathbf{E}\|\bar{A}x - y + Ux\|_2^2 = (\bar{A}x - y)^T(\bar{A}x - y) + \mathbf{E}[x^T U^T U x] \\ &= \|\bar{A}x - y\|_2^2 + x^T P x \end{aligned} \tag{6.24}$$

(โดยเรากำหนดให้  $P = \mathbf{E}[U^T U]$ ) นั่นคือ จะเห็นว่า  $\|Ax - y\|_2^2$  เป็นฟังก์ชันกำลังสองของตัวแปรสุ่ม  $U$  ดังนั้นการหาค่าคาดหวังจึงอาศัยเพียงแค่การทราบโมเมนต์ที่สองของ  $U$  โดยที่ไม่จำเป็นต้องทราบการแจกแจงของ  $U$  พิจารณาเทอม  $x^T P x$  เนื่องจาก  $\mathbf{E}[U^T U]$  เป็นเมทริกซ์บวกเสมอ ดังนั้น  $x^T P x > 0$  จึงทำให้ด้านขวามือของ (6.24) นั้นมีค่าเป็นบวก สอดคล้องกับความหมายของค่าเฉลี่ย  $\|Ax - y\|_2^2$  นอกจากนี้ การที่  $P \succ 0$  จึงทำให้  $P$  จึงมีรากที่สองและกำหนดให้เป็น  $P^{1/2}$  (ดูนิยามในนิยาม B.2 หน้า 267) ดังนั้น ปัญหา (6.23) จึงจัดรูปเป็น

$$\text{minimize}_x \|\bar{A}x - y\|_2^2 + \|P^{1/2}x\|_2^2 \tag{6.25}$$





รูป 6.11: การจำลองแหล่งกำเนิดสัญญาณคลื่นไฟฟ้าสมอง

## 6.5 ตัวอย่างงานประยุกต์

การใช้รูปแบบปัญหากำลังสองน้อยสุดร่วมกับฟังก์ชันลงโทษหรือเงื่อนไขบังคับ มีงานประยุกต์แพร่หลาย หัวข้อนี้จะยกตัวอย่างพอสังเขป ส่วนงานประยุกต์กลุ่มที่นำพารามิเตอร์ที่ประมาณได้ไปอนุมาณโครงสร้างความสัมพันธ์ของตัวแปรในระบบ จะถูกกล่าวถึงอีกครั้งในบทที่ 11

### 6.5.1 การบอกตำแหน่งแหล่งกำเนิดของสัญญาณคลื่นไฟฟ้าสมอง

สัญญาณคลื่นไฟฟ้าสมองหรือ อีอีจี Electroencephalogram (EEG) คือความต่างศักย์ไฟฟ้าที่วัดบนผิวหนังศีรษะตำแหน่งต่างๆ มีการจำลองว่าเป็นผลรวมของสัญญาณมาจากแหล่งกำเนิด (source) ด้วยการแปลงเชิงเส้น ที่เรียกว่าสมการสนามนำ (lead-field equation):  $y(t) = Lx(t) + v(t)$  โดยที่  $y \in \mathbf{R}^m$  คือ EEG และ  $x \in \mathbf{R}^n$  คือแหล่งกำเนิดกระแสโดยที่  $x_i$  คือแหล่งกำเนิดที่ตำแหน่ง  $i$  ต่างๆ ภายในสมอง การแปลง  $L$  ทำหน้าที่รวมผลจากแหล่งกำเนิดต่างๆ เรียกว่าเมทริกซ์สนามนำ (lead-field matrix) ที่จะมีค่าแตกต่างกันไปตามแบบจำลองรูปทรงศีรษะ (head model) อันส่งผลถึงสมบัติการนำทางไฟฟ้าในสมอง และ  $v$  ถือเป็นสัญญาณรบกวนที่จุดวัดของอิเล็กโทรด รูป 6.11 อธิบายการวัด EEG อันเป็นความต่างศักย์ที่วัดได้บนผิวหนังศีรษะ และมีสมมติฐานว่าเป็นผลรวมมาจากแหล่งกำเนิดหลายแหล่ง จำนวนช่องสัญญาณ EEG หรือ  $m$  (ซึ่งก็คือจำนวนอิเล็กโทรดวัด) ที่ใช้ในงานประยุกต์มักจะถูกอยู่ในหลักสิบ เช่น 32 หรือ 64 ช่อง (แต่มีเครื่องวัดที่มีจำนวนช่องได้ละเอียดเป็นหลักร้อย) ส่วนจำนวนแหล่งกำเนิด หรือ  $n$  อยู่ในช่วงหลักสิบถึงหลักร้อย ขึ้นกับจุดประสงค์การวิเคราะห์ข้อมูล เช่น ความสัมพันธ์ของการทำงานบริเวณต่างๆ ของสมอง จะพิจารณาจาก  $x$  [dSFK<sup>+</sup>16] (ไม่ใช่ที่ระดับผิวหนังศีรษะจากตัวแปร  $y$ ) ดังนั้น การสมมติให้มี  $x_i$  กระจายไปทั่วหลายบริเวณสมองที่สนใจ (regions of interest) เช่น แบ่งตามมาตรฐานแบบ AAL [TMLP<sup>+</sup>02] จะมี  $n = 116$  เป็นหลักร้อย

เมื่อเราวัด  $y$  ได้ และทราบค่า  $L$  จากแบบจำลองศีรษะ การประมาณ  $x$  จากสมการสนามนำจะเรียกว่าปัญหาผกผัน (inverse problem) หรือการประกอบให้คืนสภาพของแหล่งกำเนิด (source reconstruction) จะเห็นว่าเมื่อ  $n > m$  สมการสนามนำเป็นสมการกำหนดขาด (under-determined equation) ไม่สามารถประมาณ  $x$  ได้อย่างเป็นหนึ่งเดียว การประมาณ  $x$  ต้องมีสมมติฐานเพิ่มเติมเกี่ยวกับลักษณะของ  $x$  ที่ต้องการ นำไปสู่ปัญหาการประมาณที่ทำให้ตามเกณฑ์ ดังตัวอย่างที่สรุปไว้ใน [Hau12, §2] และ [MB19] เช่น วิธี WMNE (weighted minimum-norm estimate) ใช้ฟังก์ชันลงโทษนอร์ม-2 ที่มีเมทริกซ์ถ่วงน้ำหนัก  $W$  อันต่อยอดจากปัญหา (6.7) ดังนี้

$$\underset{x}{\text{minimize}} \quad \|y - Lx\|_2^2 + \gamma \|Wx\|_2^2 \quad (6.34)$$

## บทที่ 7

# วิธีปริภูมิย่อย

การศึกษาพลวัตและสมบัติ เช่น สภาพสังเกตได้ สภาพเข้าถึงได้ และการออกแบบตัวควบคุมขั้นสูง สำหรับระบบ MIMO ที่ซับซ้อนมีอันดับสูง มักจะใช้การวิเคราะห์ปริภูมิสถานะ วิธีการประมาณแบบจำลองในกลุ่มนี้จึงมีความจำเป็น วิธีหาเอกลักษณ์แบบปริภูมิย่อย (subspace identification) นำมาประยุกต์ใช้กับการประมาณแบบจำลองปริภูมิสถานะเชิงเส้นเวลาวิฤต (2.21) ดังนี้

$$x(t+1) = Ax(t) + Bu(t) + w(t), \quad y(t) = Cx(t) + Du(t) + v(t) \quad (7.1)$$

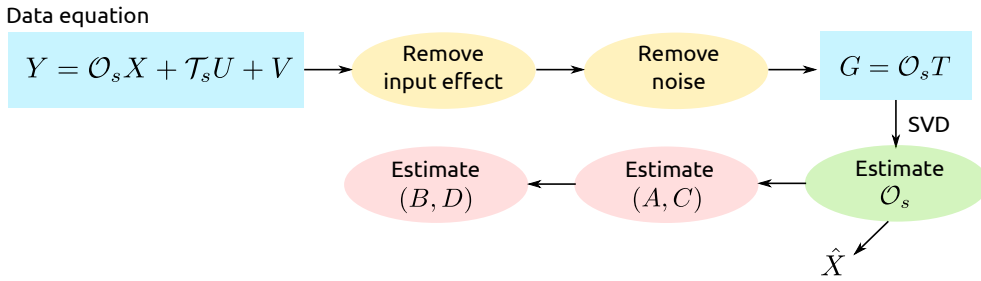
เมื่อกำหนดให้อันดับของตัวแปรเป็น  $x \in \mathbf{R}^n, y \in \mathbf{R}^p, u \in \mathbf{R}^m$  ข้อกำหนดของปัญหานี้คือ การทราบข้อมูล  $\{u(t), y(t)\}_{t=1}^N$  ผู้ใช้กำหนดอันดับของระบบ  $n$  และผลลัพธ์คือเมทริกซ์  $(A, B, C, D)$  ที่ประมาณได้เทคนิคของวิธีนี้คือการเขียนสมการของ  $y$  และ  $u$  ที่เกี่ยวข้องกับเมทริกซ์ฮังคัลแบบบล็อก (block Hankel matrix) ทำให้เราหาปริภูมิย่อยที่เชื่อมโยงไปเมทริกซ์ระบบ  $(A, B, C, D)$  ได้ผ่านทางเมทริกซ์สภาพสังเกตได้ (observability matrix) ผู้อ่านจึงควรมีพื้นฐานด้านระบบปริภูมิสถานะเชิงเส้น ในรูประบบเชิงกำหนด (deterministic system) แล้วต่อยอดไปยังระบบสโตแคสติก หรือจะเริ่มจากระบบอิสระ (autonomous system) แล้วต่อยอดไปยังระบบที่มีสัญญาณเข้า

วิธีนี้มีการประมาณตัวแปรสถานะที่ใช้หลักการของตัวกรองคาลมาน (Kalman filter) เครื่องมือทางคณิตศาสตร์อันสำคัญจึงเป็นด้านพีชคณิตเชิงเส้น ได้แก่ การหาค่าเอกฐาน (SVD) ผลเฉลยกำลังสองน้อยสุด และการแยกตัวประกอบ RQ (RQ factorization) ที่ผู้อ่านควรทบทวนจาก (B.9) ในบทนี้จึงขอเสนอหลักการโดยย่อของวิธีปริภูมิย่อยจาก [Lju99, §7.3, §10.6] และ [VW07, §9] ส่วนรายละเอียดทางเทคนิคนั้นศึกษาเพิ่มเติมได้จาก [Kat06, OM12]

### 7.1 ภาพรวมของวิธีปริภูมิย่อย

**พื้นฐานระบบเชิงเส้น.** เรานิยามเมทริกซ์สภาพสังเกตได้แบบขยาย ( $\mathcal{O}_s$ ) หรือ extended observability matrix สำหรับระบบ (7.1) ดังนี้

$$\mathcal{O}_s = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-1} \end{bmatrix} \quad (7.2)$$



รูป 7.1: ภาพรวมการหาเอกลักษณ์ของระบบด้วยวิธีปริภูมิย่อย

สำหรับกรณีที่  $s = n$  นิยาม  $\mathcal{O}_s$  จึงเป็นเมทริกซ์สภาพสังเกตได้ (observability matrix) แต่เมื่อ  $s > n$  เมทริกซ์  $\mathcal{O}_s$  จึงหมายถึงการเมทริกซ์สภาพสังเกตได้ด้วยการเพิ่มแถวจนถึง  $s$  บล็อก หากระบบ (7.1) มีผลสัมฤทธิ์เฉพาะกลุ่มต่ำสุด (minimal realization) ซึ่งสมมูลกับสมบัติระบบควบคุมได้และสังเกตได้ เราจึงสรุปได้ว่า  $\text{rank}(\mathcal{O}_s) = n$  เมื่อระบบ (7.1) ถูกแปลงไปยังระบบที่สมมูลกัน ณ พิกัดอื่น ด้วยการแปลงภาวะคล้าย (similarity transform) โดยที่ใช้  $z = T^{-1}x$  จะทำให้สมการพลวัตระบบเป็น

$$z(t+1) = T^{-1}ATz(t) + T^{-1}Bu(t) + \tilde{w}(t), \quad y(t) = CTz(t) + Du(t) + v(t) \quad (7.3)$$

โดยที่  $T \in \mathbf{R}^{n \times n}$  เป็นเมทริกซ์ไม่เอกฐานใดๆ เมื่อแปลงพิกัดแล้ว เมทริกซ์สภาพสังเกตได้แบบขยายของ (7.3) จึงสัมพันธ์กับของระบบ (7.1) ดังนี้

$$\tilde{\mathcal{O}}_s = \mathcal{O}_s T \quad (7.4)$$

หัวข้อ 7.2 จะแสดงว่าเราสามารถเขียนระบบ (7.1) ให้อยู่ในรูปความสัมพันธ์ของข้อมูลสัญญาณกับเมทริกซ์ระบบดังสมการ  $Y = \mathcal{O}_s X + \mathcal{T}_s U + V$  ที่จะเรียกว่า **สมการข้อมูล** โดย  $Y$  คือสัญญาณออก  $X$  คือตัวแปรสถานะ  $U$  คือสัญญาณเข้า และ  $V$  คือสัญญาณรบกวน เทคนิคของวิธีปริภูมิย่อยจะประกอบไปด้วย การกำจัดผลของ  $U$  ด้วยการใช้เมทริกซ์สภาพฉายตั้งฉาก การกำจัดผลของ  $V$  ด้วยการหาเมทริกซ์ที่ไม่มีสหสัมพันธ์กับ  $V$  มาคูณกับสมการข้อมูล (เราจะเรียกเทอมดังกล่าวว่า ตัวแปรเชิงเครื่องมือ) ผลลัพธ์ที่ได้คือ ความสัมพันธ์ของ  $\mathcal{O}_s$  กับเมทริกซ์  $G$  และ  $T$  หนึ่งๆ ดังสมการ

$$G = \mathcal{O}_s T \quad (7.5)$$

โดยที่เมทริกซ์ผืนผ้า  $G$  ขึ้นอยู่กับข้อมูลที่วัดได้เพียงอย่างเดียว ส่วน  $T$  ที่เป็นเมทริกซ์ผืนผ้า นั้นจะขึ้นกับตัวแปรสถานะที่ไม่ทราบค่า โดยทั่วไปเราทราบจาก (7.5) ว่า  $\mathcal{R}(G) \subseteq \mathcal{R}(\mathcal{O}_s)$  ประเด็นที่สำคัญของวิธีปริภูมิย่อยคือ ถ้าหาก  $\text{rank}(G) = \text{rank}(\mathcal{O}_s) = n$  จะทำให้เราสามารถสรุปได้ว่า

$$\mathcal{R}(G) = \mathcal{R}(\mathcal{O}_s) \quad (7.6)$$

การแยก SVD แบบลดรูปของ  $G$  ที่มี  $U_n \Sigma_n V_n^T = G$  จึงบอกได้ว่าแต่ละคอลัมน์ใน  $U_n$  เป็นเวกเตอร์มูลฐานของ  $\mathcal{R}(G)$  ดังนั้น จากการใช้ (7.6) เราจึงเขียนได้ว่า

$$U_n \Sigma_n V_n^T = \mathcal{O}_s T \quad \Rightarrow \quad U_n = \mathcal{O}_s T V_n \Sigma_n^{-1} = \mathcal{O}_s \tilde{T} = \tilde{\mathcal{O}}_s \quad (7.7)$$

กล่าวคือ ถึงแม้ว่าเราจะไม่ทราบ  $X$  (ทำให้เราไม่ทราบ  $T$ ) แต่เราสามารถพิจารณา  $\tilde{\mathcal{O}}_s$  ว่าเป็นเมทริกซ์สภาพสังเกตได้แบบขยายที่นิยามบนระบบพิกัดอื่นๆ เราจะใช้นิยามของ  $\mathcal{O}_s$  ที่สัมพันธ์กับ  $(A, C)$  ในการประมาณ  $(A, C)$  จาก  $U_n$  ที่คำนวณได้ รูปแบบของ  $G$  ที่เป็นไปตาม (7.5) นั้น จะขึ้นกับข้อกำหนดปัญหาประมาณ อันแบ่งได้เป็น 3 กรณี ดังนี้

1. หากระบบไม่มีสัญญาณเข้า (autonomous system) และไม่มีสัญญาณรบกวนแล้ว  $G$  จะขึ้นกับ  $Y$  และ  $T$  จะขึ้นกับ  $X$  การประมาณ  $(A, C)$  จะทำได้อย่างแม่นยำตรง
2. หากระบบมีสัญญาณเข้าแต่ไม่มีสัญญาณรบกวน เมทริกซ์  $G$  และ  $T$  จะได้จากการกำจัดผลของ  $U$  ออกจากสมการข้อมูล และใช้การแยกตัวประกอบ RQ เพื่อปรับปรุงให้การคำนวณเชิงตัวเลขมีประสิทธิภาพ
3. หากระบบมีสัญญาณเข้าและมีสัญญาณรบกวน เมทริกซ์  $G$  และ  $T$  จะได้จากการกำจัดผลของ  $U$  และสัญญาณรบกวน  $V$  ออกจากสมการข้อมูล ด้วยการใช้เทคนิคตัวแปรเชิงเครื่องมือ (instrumental variable) ตัวแปรนี้ยังคงเป็นฟังก์ชันของข้อมูลที่วัดได้อยู่ จึงสามารถสังเคราะห์ได้ และจากนั้นใช้การแยกตัวประกอบ RQ เช่นเดียวกัน เพื่อคำนวณ  $G$  อย่างมีประสิทธิภาพ ในกรณีที่มีสัญญาณรบกวนนี้ ข้อสรุปทางทฤษฎีที่ทำให้ (7.5) เป็นจริงนั้น จะสรุปได้ในความหมายเชิงเส้นกำกับ นั่นคือ  $G$  จะอยู่ในรูปลิมิต เมื่อ  $N \rightarrow \infty$

เมื่อเราประมาณ  $(A, C)$  ได้แล้ว เราสามารถจัดรูปสมการค่าทำนายของ  $y$  จาก (7.1) ให้อยู่ในรูปของแบบจำลองเชิงเส้นในพารามิเตอร์  $(B, D)$  ได้ และใช้วิธี LS ในการประมาณ ส่วนการประมาณตัวแปรสถานะนั้น ทำได้โดยแก้ปัญหา LS เพื่อประมาณ  $Y$  ในอนาคตด้วยข้อมูล  $Y$  ในอดีต และ  $U$  ในอนาคต และได้ความสัมพันธ์เชิงเส้นที่ประมาณเป็น  $\hat{Y} = \mathcal{O}_s \hat{X}$  จึงใช้การแยกตัวประกอบ SVD ของ  $\hat{Y}$  เพื่อหา  $\hat{X}$  จากนั้นเมื่อเราทราบค่าประมาณตัวแปรสถานะและเมทริกซ์ของระบบ จากสมการ (7.1) เราจึงประมาณ  $w(t)$  และ  $v(t)$  ได้จาก

$$\hat{w}(t) = \hat{x}(t+1) - \hat{A}\hat{x}(t) - \hat{B}u(t), \quad \hat{v}(t) = y(t) - \hat{C}\hat{x}(t) - \hat{D}u(t)$$

และคำนวณเมทริกซ์แปรปรวนร่วมตัวอย่างของ  $\hat{w}$  และ  $\hat{v}$

หัวข้อดังต่อไปนี้จะอธิบายถึงรายละเอียดทางคณิตศาสตร์ของขั้นตอนต่างๆ ตามรูป 7.1

## 7.2 ความสัมพันธ์ของสัญญาณเข้าและสัญญาณออก

จากสมการ (7.1) ผลตอบของตัวแปรสถานะจะอยู่ในรูป

$$x(t) = A^t x(0) + \sum_{\tau=0}^{t-1} A^{t-1-\tau} [Bu(\tau) + w(\tau)]$$

เราสามารถเขียนความสัมพันธ์ระหว่างชุดข้อมูลสัญญาณออก  $\{y(k)\}_{k=0}^{s-1}$  กับสัญญาณเข้า  $\{u(k)\}_{k=0}^{s-1}$  และสัญญาณรบกวนได้ดังนี้

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(s-1) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{s-1} \end{bmatrix}}_{\mathcal{O}_s} x(0) + \underbrace{\begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & & 0 \\ \vdots & & & \ddots & \ddots \\ CA^{s-2}B & CA^{s-3}B & \cdots & CB & D \end{bmatrix}}_{\mathcal{T}_s} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \\ u(s-1) \end{bmatrix} + \begin{bmatrix} v(0) \\ Cw(0) + v(1) \\ CAw(0) + Cw(1) + v(2) \\ \vdots \\ CA^{s-2}w(0) + CA^{s-3}w(1) + \cdots + Cw(s-2) + v(s-1) \end{bmatrix} \quad (7.8)$$

## บทที่ 8

# วิธีความผิดพลาดในการทำนาย

วิธีความผิดพลาดในการทำนาย (prediction error method หรือ PEM) คือการใช้แบบจำลองที่ให้ค่าทำนายของผลตอบ และเลือกพารามิเตอร์ของแบบจำลองที่ทำให้ค่าคลาดเคลื่อนระหว่างผลตอบจริงกับค่าทำนายมีค่าต่ำที่สุด บทนี้จะอธิบายเกี่ยวกับองค์ประกอบที่สำคัญของตัวทำนาย วิธีการเลือกพารามิเตอร์ของแบบจำลอง และคุณสมบัติของแบบจำลองที่ประมาณได้

### 8.1 องค์ประกอบของตัวทำนาย

กำหนดให้แบบจำลองมีพารามิเตอร์เป็น  $\theta$  และให้สัญลักษณ์

$$\hat{y}(t + p|t; \theta) \quad (8.1)$$

คือค่าทำนายของ  $y$  ณ เวลา  $t + p$  (โดยที่  $p > 0$ ) ที่เป็นสัญญาณออกของแบบจำลองที่มีพารามิเตอร์  $\theta$  แบบจำลองทำนายใช้ข้อมูลของ  $y$  ในอดีตจนถึงเวลา  $t$  เราจะเรียก (8.1) ว่าเป็น *ค่าทำนาย (หรือค่าพยากรณ์)  $p$  ช่วงเวลาล่วงหน้าของ  $y$*  พารามิเตอร์  $p$  นี้เป็นสิ่งที่ทราบจากข้อกำหนดของปัญหา เช่น ปัญหาการพยากรณ์กำลังไฟฟ้าแสงอาทิตย์ในช่วงวันเดียวกัน (intra-day forecasting) ข้อกำหนดปัญหาหมักกำหนดให้  $t$  เป็นตัวชี้เวลารายชั่วโมง และ  $p$  มีค่า 1 ชั่วโมง นั่นคือ เมื่อขณะนี้ข้อมูลถึงเวลา 10:00 น. ผู้ใช้ต้องการทราบค่าพยากรณ์ ณ เวลา 11:00 น. เป็นต้น หากข้อกำหนดปัญหาเปลี่ยนเป็นการพยากรณ์กำลังไฟฟ้าหลายวันล่วงหน้า (day-ahead forecasting) ข้อกำหนดปัญหาอาจจะกำหนดให้  $t$  เป็นตัวชี้เวลารายวัน และ  $p$  มีค่า 3 วัน นั่นคือ หาก ณ ขณะนี้คือวันที่ 1 มกราคม การจะทราบค่าพยากรณ์ของวันที่ 4 มกราคม จึงใช้ข้อมูลค่าวัดในอดีตจนถึงวันที่ 1 มกราคมเท่านั้นในการคำนวณค่าพยากรณ์ จากตัวอย่างดังกล่าว เห็นได้ชัดว่า การพยากรณ์ที่จะทำนายค่าในอนาคตอันไกลขึ้นย่อมทำได้ยากกว่าเพราะมีความไม่แน่นอนเริ่มมากขึ้น ปัญหาการทำนายจึงยากขึ้นหาก  $p$  มีค่าสูงขึ้น รวมถึงเป็นที่คาดหมายได้ว่า การทำนายในช่วงเวลา  $p$  ล่วงหน้าที่สูงขึ้น จะมีความผิดพลาดในการทำนายที่สูงขึ้น

วิธี PEM ที่จะกล่าวในบทนี้ ใช้ค่าพยากรณ์ 1 ช่วงเวลาล่วงหน้า หรือ  $\hat{y}(t|t-1; \theta)$  และกำหนดให้

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|t-1; \theta) \quad (8.2)$$

คือค่าคลาดเคลื่อนการทำนาย  $y$  (หรือความผิดพลาดในการทำนาย prediction error) ณ เวลา  $t$  โดยที่ตัวทำนายใช้ข้อมูลได้จนถึงเวลา  $t-1$  หลักการใหญ่ของ PEM คือการหา  $\theta$  ที่ทำให้  $\varepsilon(t, \theta)$  รวม

เฉลี่ยจากหลายเวลา  $t$  นั้นมีค่าต่ำสุด ที่วัดได้จากฟังก์ชันสูญเสีย (loss function) องค์ประกอบแรกที่ สำคัญของตัวทำนาย คือ **โครงสร้างของตัวทำนายในรูปทั่วไป**

$$\hat{y}(t|t-1; \theta) = N(L; \theta)y(t) + M(L; \theta)u(t) \quad (8.3)$$

โดยที่  $N$  และ  $M$  นั้นเป็นตัวกรองหรือฟังก์ชันถ่ายโอนที่ต้องมีตัวดำเนินการหน่วง (delay operator) อย่างน้อย 1 อันดับ อันกล่าวได้เป็นเงื่อนไขคือ  $N(0; \theta) = M(0; \theta) = 0$  เงื่อนไขนี้มาจาก ข้อกำหนดที่ว่า  $\hat{y}(t|t-1; \theta)$  ใน (8.3) นั้นต้องใช้ข้อมูลจากในอดีตจนถึงเวลา  $t-1$  เท่านั้น การ กำหนดโครงสร้าง (8.3) คือการเลือก  $N, M$  ว่าเป็นฟังก์ชันถ่ายโอนอันดับใด องค์ประกอบที่สอง คือ การเลือกโครงสร้างของแบบจำลองรูปทั่วไปดัง (2.27) เราต้องกำหนด  $G, H, \Lambda$  ดังรายละเอียดในบท ที่ 2 เมื่อกำหนดโครงสร้างทั้ง 2 องค์ประกอบแล้วนั้น จะทำให้ทราบว่า  $\varepsilon(t, \theta)$  นั้นเป็นฟังก์ชันอย่างไร ในพารามิเตอร์  $\theta$  องค์ประกอบที่สามเป็นการกำหนดฟังก์ชันสูญเสียที่นำมาใช้ประเมินค่าคลาดเคลื่อน ว่ามีค่าน้อยเพียงใด เราจะนิยาม **เมทริกซ์ความแปรปรวนร่วมของค่าคลาดเคลื่อน** เป็น

$$R(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t, \theta) \varepsilon^T(t, \theta) \quad (8.4)$$

และจะเห็นว่า  $R(\theta) \succeq 0$  เสมอ หาก  $N$  มีค่ามากเพียงพอ  $R(\theta)$  มักจะกลายเป็นเมทริกซ์บวกแน่นอน

**ฟังก์ชันสูญเสีย** (loss function)  $f: \mathbf{S}^n \rightarrow \mathbf{R}$  คือฟังก์ชันบ่งชี้ความผิดพลาดในการประมาณ ฟังก์ชันนี้ส่งจากเมทริกซ์ความแปรปรวนร่วมที่มีสมบัติสมมาตรไปยังค่าสเกลาร์ โดยมีสมบัติคือ

1.  $f(X + \Delta X) = f(X) \iff \Delta X = 0$  และ
2.  $f$  เป็นฟังก์ชันเพิ่ม (increasing function) กล่าวคือ สำหรับ  $X \succ 0, \Delta X \succeq 0$  ใดๆ

$$f(X + \Delta X) \geq f(X)$$

สมบัติดังกล่าวสะท้อนความเป็นฟังก์ชันสูญเสีย หากเลือกค่า  $\theta$  ไม่ดี การทำนายจะแยกลง นั่นคือ  $\varepsilon(t, \theta)$  มีค่าสูงขึ้น ทำให้  $R(\theta)$  เป็นเมทริกซ์ที่ใหญ่ขึ้น (แบบเมทริกซ์บวกกึ่งแน่นอน) ค่าของฟังก์ชันสูญเสีย  $f$  จึงควรมีค่าเพิ่มขึ้น

**ตัวอย่าง 8.1** ฟังก์ชันสูญเสียเชิงเส้น  $f(X) = \text{tr}(WX)$  สำหรับ  $W \succ 0$  ใดๆ ที่กำหนดให้

$$f(X + \Delta X) = \text{tr}(WX) + \text{tr}(W\Delta X) \geq f(X)$$

จากสมบัติ  $A \succeq 0, B \succeq 0$  แล้วจะได้ว่า  $\text{tr}(AB) \geq 0$  เพราะว่า  $B$  สามารถแยกตัวประกอบ ได้เป็น  $B = LL^T$  ดังนั้น  $\text{tr}(AB) = \text{tr}(ALL^T) = \text{tr}(L^TAL) \geq 0$  เมื่อนำมาใช้กับเทอม  $\text{tr}(W\Delta X)$  เนื่องจากทั้ง  $W \succ 0$  และ  $\Delta X \succeq 0$  จึงสรุปได้ว่า  $f(X)$  เป็นฟังก์ชันเพิ่ม และสามารถเลือกให้เป็นฟังก์ชันสูญเสียได้ สำหรับ  $R(\theta) \succ 0$  ใดๆ จะเห็นว่า  $\text{tr}(R(\theta))$  คือผลรวม แนวทางแยกมุมของ  $R(\theta)$  อันหมายถึงผลรวมค่าคลาดเคลื่อนกำลังสองเฉลี่ย Mean Square Error (MSE) ของ  $\varepsilon_i(t, \theta)$  ในทุก  $i$  (เนื่องจาก  $\varepsilon$  เป็นเวกเตอร์)

ตัวอย่างข้างต้นเป็นฟังก์ชันสูญเสียที่เป็นเชิงเส้นในตัวแปร  $X$  ในตัวอย่างถัดไปจะพิจารณาฟังก์ชันไม่เชิงเส้น

เทียบกับ  $\theta$  จึงมองว่า  $\nabla_{\theta}^2 \varepsilon(t, \hat{\theta})$  ได้จากการใช้ตัวกรองตัวอื่นบน  $y$  และ  $u$  เมื่อ  $N$  มีค่ามาก เทอมแรกใน (8.28) จึงใช้กฎจำนวนมาก และการที่  $\nabla_{\theta}^2 \varepsilon(t, \hat{\theta})$  และ  $\varepsilon(t, \hat{\theta})$  เป็นอิสระต่อกัน จึงได้ว่า

$$\frac{2}{N} \sum_{t=1}^N \varepsilon(t, \hat{\theta}) \nabla_{\theta}^2 \varepsilon(t, \hat{\theta}) \xrightarrow{P} \frac{2}{N} \sum_{t=1}^N \mathbf{E}[\varepsilon(t, \hat{\theta}) \nabla_{\theta}^2 \varepsilon(t, \hat{\theta})] = \frac{2}{N} \sum_{t=1}^N \mathbf{E}[\varepsilon(t, \hat{\theta})] \mathbf{E}[\nabla_{\theta}^2 \varepsilon(t, \hat{\theta})] = 0$$

ตัวอย่างการคำนวณเกรเดียนต์ของ  $V_N$  สำหรับแบบจำลอง ARMAX สามารถดูได้ใน [SS89, §7.6]

การใช้วิธีเชิงตัวเลขเพื่อหาค่าที่เหมาะสมที่สุด (ปัญหาไม่เชิงเส้น) มักจะการันตีได้เพียงการพบค่าเหมาะสมที่สุดเฉพาะที่ (local optimum) ซึ่งอาจพบผลเฉลยต่างกันจากการใช้ค่าเริ่มต้นของตัวแปรที่ต่างกัน ความซับซ้อนเชิงคำนวณและอัตราการลู่เข้าจะขึ้นกับวิธีเชิงตัวเลข การลู่เข้าของผลเฉลยขึ้นกับตัวเลือกของค่าเริ่มต้นด้วย วิธี PEM จึงมักใช้ผลการประมาณจากวิธีอื่นมาเป็นค่าเริ่มต้น เช่น วิธีปริภูมิย่อยที่เป็นวิธีไม่วนซ้ำ (non-iterative) แล้วใช้วิธี PEM ปรับพารามิเตอร์ให้เหมาะสมอีกรอบหนึ่ง

## 8.5 ตัวอย่างชุดคำสั่งจาก MATLAB

หัวข้อนี้แสดงการเรียกใช้คำสั่ง pem ด้วยสามตัวอย่าง อันได้แก่ การประมาณแบบจำลอง ARMAX การประมาณแบบจำลองปริภูมิสถานะ ที่จะใช้ PEM ประมาณพารามิเตอร์เพื่อปรับจากค่าประมาณเริ่มต้นจากวิธีปริภูมิย่อย และตัวอย่างเปรียบเทียบการประมาณแบบจำลอง ARMA ด้วยกล่องเครื่องมือกลุ่มอื่น

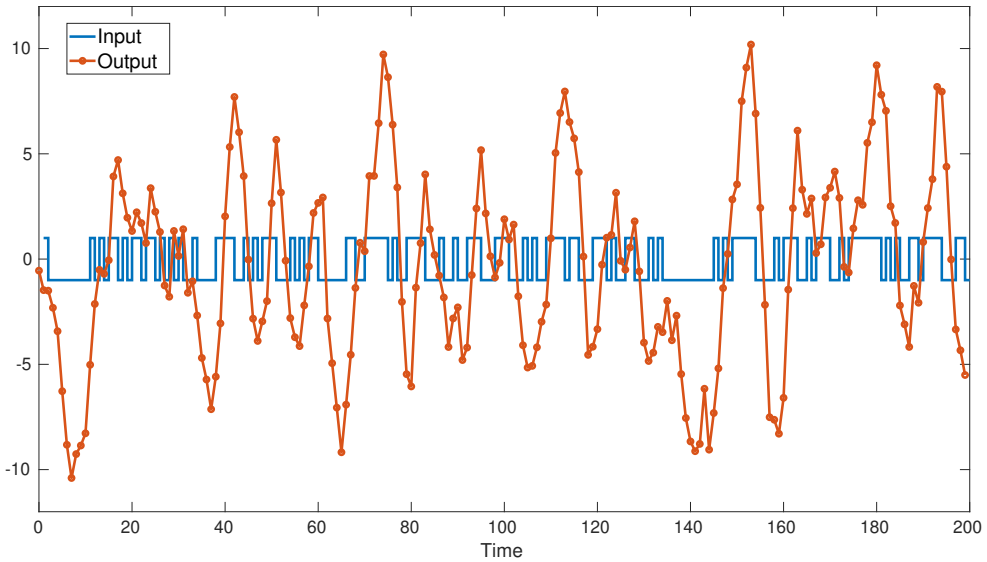
**ตัวอย่าง 8.7** ตัวอย่างนี้แสดงการประมาณ PEM ด้วยคำสั่ง pem จากกล่องเครื่องมือ System Identification Toolbox ใน MATLAB เราเริ่มจากการสมมติให้ระบบจริงคือ กระบวนการ ARMAX

$$(1 - 1.5L + 0.7L^2)y(t) = (1.0L + 0.5L^2)u(t) + (1 - 1.0L + 0.2L^2)e(t)$$

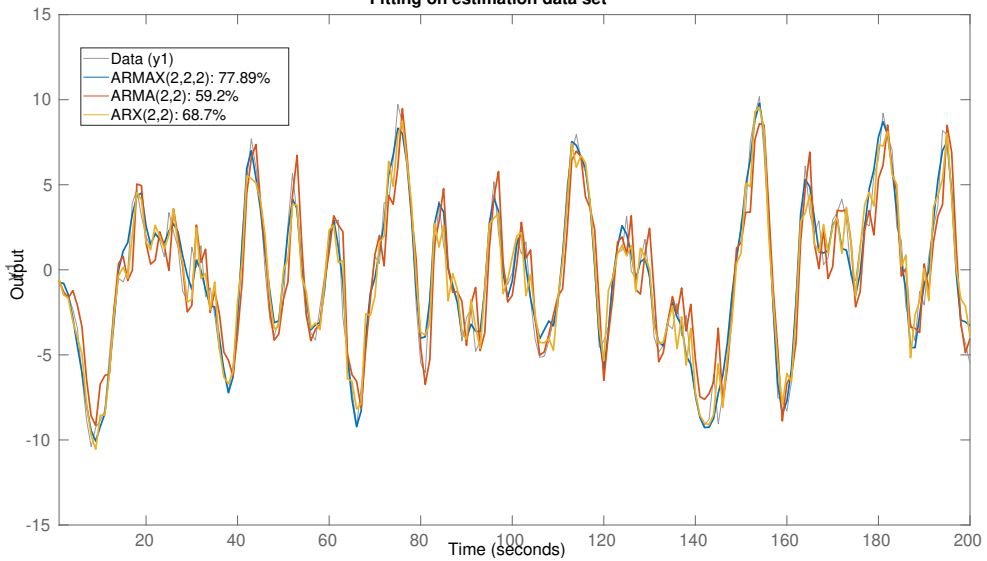
สัญญาณเข้า  $u$  เป็นสัญญาณ 2 ระดับ และเป็นอิสระกับ  $e$  ที่เป็นสัญญาณรบกวนขาว ทั้ง 2 สัญญาณมีความแปรปรวนเป็นหนึ่ง ในขั้นตอนการประมาณแบบจำลอง เราสมมติโครงสร้างแบบจำลอง 3 ชนิด ได้แก่

- แบบจำลอง ARMAX(2,2,2) ตรงตามระบบจริง
- แบบจำลอง ARMA(2,2) นั่นคือ เราไม่นำข้อมูล  $u$  มาช่วยในการประมาณระบบ
- แบบจำลอง ARX(2,2) นั่นคือเราไม่ได้คำนึงถึงพลวัตของ  $e$  ในอดีตว่าจะมีผลต่อ  $y$

เมื่อสังเคราะห์ข้อมูลเป็นจำนวน 200 จุดเวลา จึงประมาณแบบจำลองด้วยคำสั่ง armax (อันต้องใช้วิธี PEM) และ arx (ซึ่งใช้วิธี LS ได้) และได้ผลลัพธ์ของการใช้ข้อมูลชุดประมาณในรูป 8.1 การสมมติใช้โครงสร้างระบบที่ซับซ้อนต่ำกว่าระบบจริง (ARX) หรือการละเลยผลของ  $u$  ต่อการประมาณ (ARMA) จะทำให้ผลการประมาณแย่กว่าแบบจำลอง ARMAX ซึ่งมีโครงสร้างและอันดับที่ตรงกับระบบจริง การประมาณด้วยวิธี PEM ใช้คำสั่ง pem สำหรับแบบจำลองโครงสร้างทั่วไป หรือใช้คำสั่ง polyst สำหรับแบบจำลองพหุนาม แต่ในตัวอย่างนี้ เราสามารถใช้คำสั่ง armax ที่ออกแบบมาเพื่อโครงสร้าง ARMAX โดยเฉพาะ ตัวอย่างโปรแกรมดูได้จากรหัสคำสั่ง 8.1



(a) สัญญาณเข้าที่ใช้กระตุ้นระบบและสัญญาณออกจากระบบจริง  
Fitting on estimation data set



(b) ผลการประมาณ

รูป 8.1: ผลการประมาณระบบ ARMAX ด้วยการใช้แบบจำลอง ARMA และ ARX



## รหัสคำสั่งที่ 8.1: การประมาณ ARMAX ด้วยวิธี PEM

```

% Generate the data
N = 200; Ts = 1; t = (0:Ts:Ts*(N-1))'; noise_var = 1;
a = [1 -1.5 0.7]; b = [0 1 .5]; c = [1 -1 0.2];
u = sign(randn(N,1));
e = sqrt(noise_var)*randn(N,1);
dgp = idpoly(a,b,c,1,1,noise_var,Ts); % data generating process
y = sim(dgp,[u e]);

% Another way to generate the data
% u = idinput(N,'PRBS'); u = idinput(N,'RBS');
% opt = simOptions('AddNoise',true,'NoiseData',e); y1 = sim(dgp,u,opt);

% Identification
m = armax(DAT,[2 2 2 1]); % [na nb nc nk] ARMAX(2,2,2)
m1 = armax(DAT,[2 0 2 1]); % ARMA(2,2)
m2 = arx(DAT,[2 2 1]); % ARX(2,2) uses the LS method

% Another way to estimate a polynomial model
% p = polyest(DAT,[2 2 2 0 0 1]); % [na nb nc nd nf nk]

% Compare the measured output and the model output
compare(DAT,m,m1,m2,1); % Use '1' to compare the 1-step prediction

```

ตัวอย่างต่อไป คือการประมาณแบบจำลองปริภูมิสถานะกับข้อมูลมอดเตอร์กระแสตรง ดังแสดงในรหัสคำสั่ง 8.2 ข้อมูลของสัญญาณเข้าและสัญญาณออกได้แบ่งออกเป็นชุดประมาณ และชุดทดสอบแบบจำลองปริภูมิสถานะสามารถประมาณโดยใช้วิธีปริภูมิย่อย เช่น N4SID และนำผลลัพธ์มาเป็นแบบจำลองเริ่มต้นเพื่อใช้วิธี PEM ในการปรับค่าพารามิเตอร์เพิ่มเติม เราเปรียบเทียบสมรรถนะแบบจำลอง `init_sys` (วิธีปริภูมิย่อยที่ใช้คำสั่ง `n4sid`) และ `estss` (วิธี PEM ที่ใช้คำสั่ง `pem`) ด้วยค่า `FitPercent` เท่ากับ 93.42% และ 93.86 % ตามลำดับ (ค่าเปอร์เซ็นต์ความเข้ากัน ยิ่งสูงยิ่งดี บ่งชี้ถึงแบบจำลองอธิบายข้อมูลได้ดีขึ้น) นอกจากนี้ พิจารณาการใช้แบบจำลอง ARX อันดับต่างๆ เพื่อเปรียบเทียบผลการประมาณ ผลลัพธ์ในรูป 8.2 พบว่า

- แบบจำลองปริภูมิสถานะอันดับ 2 และ 3 แสดงโดย  $SS(2)$  และ  $SS(3)$  ในกราฟ ให้ผลการประมาณที่มีสมรรถนะใกล้เคียงกัน แต่แบบจำลอง  $SS(1)$  ยังไม่ซับซ้อนพอที่จะอธิบายข้อมูลได้ดี ทั้งนี้ มีข้อสังเกตว่า แบบจำลองปริภูมิสถานะที่ประมาณได้อาจจะมีตัวแปรสถานะอยู่ในระบบพิกัดอื่น (ที่ไม่ใช่ระบบพิกัดเดิมตามลักษณะทางกายภาพของระบบจริง)
- แบบจำลอง  $ARX(2,2)$  ใช้เทอมอันเป็นผลจากช่วงเวลาที่ผ่านมา (lag order) ของ  $u$  อันดับ 2 จะให้ผลประมาณที่ดีขึ้นกว่าการเลือกใช้อันดับ 1 ในแบบจำลอง  $ARX(2,1)$  หรือ  $ARX(1,1)$  บ่งชี้ว่าสัญญาณเข้าส่งผลมากในการเรียนรู้พลวัตของ  $y$  (มากกว่าการใช้ข้อมูลในอดีตของ  $y$  มาประมาณค่า  $y$ )
- สุดท้ายเราเปรียบเทียบ 2 แบบจำลองในข้อมูลตรวจสอบ อันเป็นข้อมูลชุดใหม่ที่ไม่ได้ใช้ในขั้นตอนการประมาณ พบว่า แบบจำลอง  $ARX(2,2)$  และ  $SS(2)$  ให้สมรรถนะที่ใกล้เคียงกัน โพลของทั้ง 2 แบบจำลองก็อยู่บนตำแหน่งที่ใกล้เคียงกัน ได้แก่ ตำแหน่ง 0.2298, 0.8506 สำหรับแบบจำลองปริภูมิสถานะ และตำแหน่ง 0.2498, 0.8462 สำหรับแบบจำลอง ARX ทั้งนี้ โพลที่ประมาณได้ทั้งหมดอยู่ในวงกลมหนึ่งหน่วย บ่งชี้ว่าแบบจำลองที่ประมาณได้มีเสถียรภาพ การหาโพลของแบบจำลอง ARX ทำได้จากการหารากของพหุนามส่วนของฟังก์ชันถ่ายโอน ส่วนการหาโพลของแบบจำลองปริภูมิสถานะ หาได้จากค่าเฉพาะของเมทริกซ์พลวัต  $A$  ที่ประมาณได้

---

## บทที่ 9

# การประมาณเชิงสถิติ

---

วิธีการประมาณแบบจำลองที่กล่าวมาดังปัญหา LS หรือวิธีปริภูมิย่อย ไม่ได้ใช้ความรู้เชิงสถิติของตัวแปรไม่แน่นอนในระบบ ในขั้นตอนการประมาณ ถึงแม้ว่าวิธีเหล่านี้มีสมมติฐานของการมีสัญญาณรบกวน  $e$  (อันเป็นตัวแปรสุ่ม) ในระบบจริง แต่สมบัติเชิงสถิติของ  $e$  ไม่ได้นำมาใช้ในกรรมวิธีของการหาค่า  $\hat{\theta}$  เช่น วิธี LS หาค่า  $\hat{\theta}$  ที่ทำให้ความคลาดเคลื่อนกำลังสองน้อยที่สุด หรือวิธีปริภูมิย่อยที่ใช้หลักการหาภาพฉายตั้งฉากเพื่อกำจัดเทอมสัญญาณรบกวนจากสมการข้อมูล ไม่ว่า  $e$  ที่รบกวนในระบบจริงจะมีการแจกแจงอย่างไร หรือมีความแปรปรวนสูงเพียงใด<sup>1</sup> บทนี้จึงนำเสนอแนวทางวิธีการประมาณที่อาศัยความรู้เชิงสถิติ เช่น สมบัติการแจกแจง ปริมาณโมเมนต์ต่างๆ ของตัวแปรสุ่มที่เกี่ยวข้องทั้งหมดในแบบจำลอง

กำหนดให้ตัวแปรสุ่ม  $x$  และ  $y$  มีฟังก์ชันแจกแจงความน่าจะเป็นร่วมคือ  $f(x, y)$  โดยที่  $y$  เป็นตัวแปรที่ทราบค่าวัด การกำหนดเช่นนี้ หมายถึงว่า ค่าของ  $y$  ที่เกิดขึ้นย่อมมีผลมาจาก  $x$  ด้วยว่าเป็นค่าใด บทนี้จะอธิบายรูปแบบปัญหาการประมาณต่างๆ ที่ใช้สมบัติเชิงสถิติจากตัวแปรสุ่ม 2 ตัวนี้<sup>2</sup>

### 9.1 การประมาณความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด

หาก  $x$  เป็นตัวแปรสุ่มที่มีฟังก์ชันแจกแจงความน่าจะเป็นคือ  $f(x)$  เมื่อมีคำถามว่า เราจะคาดเดาหรือประมาณค่า  $x$  เป็นเท่าใดจึงจะเหมาะสม เราอาจเดาว่า ค่าคาดหมายของ  $x$  น่าจะเป็นตัวประมาณที่ดีในนัยหนึ่ง เมื่อขยายสถานการณ์เป็นการกำหนดให้  $x$  และ  $y$  มีฟังก์ชันแจกแจงความน่าจะเป็นร่วมคือ  $f(x, y)$  กำหนดให้เราไม่ทราบค่า  $x$  แต่ทราบค่าวัด  $y$  และในทำนองคล้ายกัน หากเรามีคำถามว่า เมื่อทราบค่าวัดของ  $y$  ค่าหนึ่ง จะประมาณค่า  $x$  เป็นเท่าใดจึงจะเหมาะสม เราอาจเดาเหมือนกับตัวอย่างแรกว่า ควรใช้ค่าคาดหมายของ  $x$  ที่มีเงื่อนไขว่าทราบ  $y$  (conditional expectation of  $x$  given  $y$ ) น่าจะเป็นตัวประมาณที่ดีในสถานการณ์นี้

ในเบื้องต้น ก่อนที่จะพิสูจน์ค่าประมาณของ  $x$  อย่างที่กล่าว ว่าจะเป็นค่าที่เหมาะสมหรือไม่ ขอให้

---

<sup>1</sup>ขอให้พึงสังเกตว่าการใช้สมบัติเชิงสถิติของ  $e$  ในขั้นตอนการประมาณเพื่อหาค่า  $\hat{\theta}$  กับในขั้นตอนวิเคราะห์สมบัติเชิงสถิติของ  $\hat{\theta}$  ที่ต้องอาศัยสมมติฐานของ DGP นั้นเป็นคนละประเด็นกัน

<sup>2</sup>ในบทนี้ เราเลือกใช้สัญลักษณ์ตัวพิมพ์เล็ก  $x, y$  ให้แทนตัวแปรสุ่ม (แทนการใช้ตัวพิมพ์ใหญ่) เพราะไม่มีประเด็นรายละเอียดในการแยกความแตกต่างระหว่างตัวแปรสุ่มและค่าที่เป็นไปได้ จึงขอให้อ่านตามบริบท

พิจารณาสมบัติของค่าคาดหวังมีเงื่อนไข จากนิยาม

$$\mathbf{E}[x|y] = \int x f(x|y) dx$$

โดยที่  $f(x|y)$  คือฟังก์ชันแจกแจงความน่าจะเป็นมีเงื่อนไขของ  $x$  ที่ทราบ  $y$  และมีความสัมพันธ์กับฟังก์ชันการแจกแจงร่วมและฟังก์ชันแจกแจงตามขอบเป็น  $f(x|y) = f(x, y)/f(y)$  สมบัติที่สำคัญของ  $\mathbf{E}[x|y]$  มีดังนี้

1.  $\mathbf{E}[x|y]$  เป็นฟังก์ชันของ  $y$  สมมติให้  $y = 1$  เราย่อมได้  $\mathbf{E}[x|y = 1]$  เป็นค่าหนึ่ง และเมื่อให้  $y = 2$  เราย่อมได้  $\mathbf{E}[x|y = 2]$  เป็นอีกค่าหนึ่งที่แตกต่างกัน เราจึงเห็นได้ว่า ค่าคาดหวังมีเงื่อนไขของ  $x$  เมื่อทราบค่า  $y$  เป็นฟังก์ชันของ  $y$  เพราะค่าคาดหวังนั้นเป็นการหาค่าเฉลี่ยบนตัวแปรสุ่ม  $x$  ไปแล้ว ณ สำหรับค่า  $y$  ค่าหนึ่ง
2. ค่าคาดหวังรวม (Total expectation) หาได้จากกฎค่าคาดหวังทำซ้ำ (law of iterated expectation)

$$\mathbf{E}[\mathbf{E}[x|y]] = \mathbf{E}[x] \quad (9.1)$$

กล่าวคือ จากสมบัติแรก  $\mathbf{E}[x|y]$  เป็นฟังก์ชันของ  $y$  เมื่อพิจารณาเทอมทางซ้ายมือ การหาค่าคาดหวังเทียบตัวแปร  $y$  ของ  $\mathbf{E}[x|y]$  จะได้เท่ากับค่าคาดหวังของ  $x$  พอดี

3. สำหรับฟังก์ชันสเกลาร์  $g$  ใดๆ ที่ทำให้  $\mathbf{E}[|g(y)|] < \infty$  เราขยายผลของค่าคาดหวังรวมได้ว่า

$$\mathbf{E}[(x - \mathbf{E}[x|y])g(y)] = 0 \quad (9.2)$$

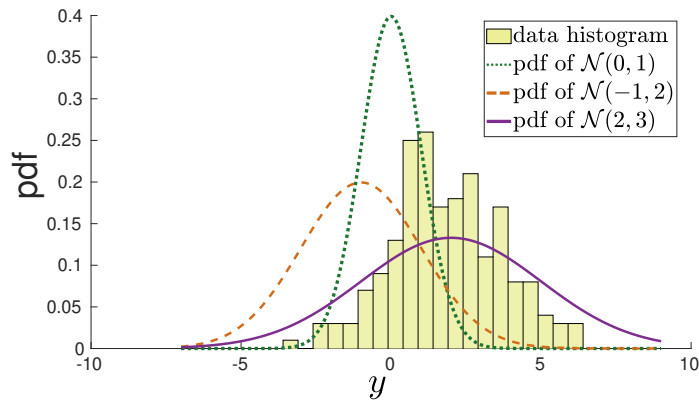
**พิสูจน์.** เมื่อใช้นิยามของ  $f_{x|y}(x|y) = f_{xy}(x, y)/f_y(y)$  จะได้ว่า

$$\begin{aligned} \mathbf{E}[x|y]g(y) &= \int_x x f_{x|y}(x|y) dx \cdot g(y) = \int_x x \cdot \frac{f_{xy}(x, y)}{f_y(y)} dx \cdot g(y) \\ \mathbf{E}[\mathbf{E}[x|y]g(y)] &= \int_y \int_x x g(y) \frac{f_{xy}(x, y)}{f_y(y)} f_y(y) dx dy \\ &= \int_y \int_x x f_{xy}(x, y) g(y) dx dy = \mathbf{E}[xg(y)] \end{aligned}$$

เราอธิบายความหมายของ (9.2) ได้จากการใช้ค่าสหสัมพันธ์เพื่อบอกความตั้งฉากของตัวแปรสุ่ม 2 ตัวใดๆ (ที่มีค่าเฉลี่ยเป็นศูนย์) กล่าวคือ  $\mathbf{E}[uv] = 0$  บ่งชี้ว่า  $u$  และ  $v$  ไม่มีสหสัมพันธ์ต่อกัน และเรียกว่า  $u$  ตั้งฉาก<sup>3</sup>กับ  $v$  เทอม  $(x - \mathbf{E}[x|y])$  คือความคลาดเคลื่อนส่วนเหลือ (residual error) ของการประมาณ  $x$  ด้วยค่าคาดหวังมีเงื่อนไข ความคลาดเคลื่อนนี้จะตั้งฉากหรือไม่มีสหสัมพันธ์กับฟังก์ชันใดๆ ของ  $y$  สมบัตินี้บ่งชี้ว่า  $\mathbf{E}[x|y]$  เป็นตัวประมาณที่ดีที่สุดของ  $x$  เพราะว่าการคลาดเคลื่อนไม่เหลือข้อมูลที่สามารถอธิบายได้อีกจาก  $y$

สมมติให้  $y$  เป็นข้อมูลที่วัดได้ การประมาณความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด หรือ MMSE นั้น มีข้อกำหนดปัญหาคือ การหาฟังก์ชันของ  $y$  สมมติเป็น  $h(y)$  ที่ค่าของฟังก์ชันนั้นประมาณ  $x$  ได้ดีที่สุด ด้วยการทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ย  $\mathbf{E}|x - h(y)|^2$  มีค่าต่ำสุด เมื่อ  $x$  และ  $y$  เป็นเวกเตอร์สุ่ม เราสามารถขยายข้อกำหนดปัญหา และแสดงผลลัพธ์ที่ได้ดังทฤษฎีบท 9.1

<sup>3</sup>หาก  $u$  และ  $v$  เป็นเวกเตอร์สุ่ม นิยามการตั้งฉากจะใช้  $\mathbf{E}[u^T v] = 0$



รูป 9.1: ตัวอย่างฟังก์ชันแจกแจงความน่าจะเป็นสำหรับพารามิเตอร์ค่าต่างๆ

ว่าเป็น ฟังก์ชันภาวะน่าจะเป็น (likelihood function) อันบ่งชี้ถึงความน่าจะเป็นของข้อมูลตัวอย่าง  $y$  เมื่อกำหนดโครงสร้างฟังก์ชันแจกแจงสำหรับพารามิเตอร์  $\theta$  หนึ่งๆ แต่ในมุมมองปัญหาการประมาณ เราพิจารณาว่า  $\ell(\theta; y)$  คือฟังก์ชันของพารามิเตอร์ว่ามีความเป็นไปได้มากน้อยเพียงใดสำหรับค่า  $\theta$  นั้น เมื่อกำหนด  $y$  มาให้ พิจารณา  $\mathcal{L}(\theta; y)$  ให้เป็นการแปลงลอการิทึมของ  $\ell(\theta; y)$

$$\mathcal{L}(\theta; y) = \sum_{i=1}^N \log f(y_i; \theta) \quad (9.14)$$

เราจะเรียก  $\mathcal{L}$  ว่าเป็น ฟังก์ชันลอการิทึมภาวะน่าจะเป็น (log-likelihood function) ของพารามิเตอร์  $\theta$  ซึ่งให้ความหมายเช่นเดียวกับฟังก์ชันภาวะน่าจะเป็น เนื่องจากฟังก์ชันลอการิทึมเป็นฟังก์ชันเพิ่ม (increasing function) ดังนั้น สำหรับ  $\theta$  หนึ่งๆ หาก  $\ell$  มีค่าสูง ฟังก์ชัน  $\mathcal{L}$  ย่อมมีค่าสูงตามไปด้วย ทั้งนี้ จะเห็นจากตัวอย่างต่อไปว่า การแปลงลอการิทึมให้ความสะดวกในการวิเคราะห์ เพราะมีการแจกแจงหลายชนิดอยู่ในกลุ่มของฟังก์ชันเลขชี้กำลัง (exponential family) สุดท้าย สังเกตว่า  $\mathcal{L}$  นั้นขึ้นกับจำนวนข้อมูลตัวอย่าง  $N$  ที่นำมาใช้คำนวณด้วย

การประมาณภาวะน่าจะเป็นสูงสุด หรือ Maximum likelihood estimation (MLE) คือการเลือก  $\theta$  ที่ทำให้ฟังก์ชันลอการิทึมภาวะน่าจะเป็นมีค่าสูงสุด

$$\hat{\theta}_{\text{mle}} = \operatorname{argmax}_{\theta} \mathcal{L}(\theta; y) \quad (9.15)$$

และใช้สัญลักษณ์  $\hat{\theta}_{\text{mle}}$  ว่าเป็นตัวประมาณของพารามิเตอร์ค่าจริง  $\theta^*$  ด้วยวิธี MLE ปัญหา (9.15) เป็นการหาค่าเหมาะที่สุดแบบไม่มีเงื่อนไข ผลเฉลยจึงต้องสอดคล้องกับเงื่อนไขเกรเดียนต์เท่ากับศูนย์

$$\nabla \mathcal{L}(\theta; y) = \sum_{i=1}^N \nabla_{\theta} \log f(y_i; \theta) = 0 \quad (9.16)$$

เงื่อนไขเหมาะที่สุดของปัญหา MLE ดัง (9.16) นั้น โดยทั่วไปจะอยู่ในรูปสมการไม่เชิงเส้นใน  $\theta$  การหาผลเฉลยจึงมักต้องใช้วิธีเชิงตัวเลข หรือบางปัญหาอาจมีผลเฉลยในรูปแบบปิดก็เป็นได้

(exact likelihood function) ซึ่งจะมีรูปแบบเป็น

$$f(y_{p+1}, y_{p+2}, \dots, y_N | Z^p) = \left( \prod_{t=p+1}^N f_e(y_t - g(t, Z^{t-1})) \right) \cdot f(y_1, y_2, \dots, y_p) \quad (9.27)$$

โดยที่  $Z^p = (y_1, y_2, \dots, y_p)$  ถือเป็นตัวแปรสุ่มเช่นกันที่ต้องหาพารามิเตอร์ของฟังก์ชันการแจกแจง หากสมมติให้เป็นเกาส์เซียน เราจำเป็นต้องหาความเมทริกซ์ความแปรปรวนร่วมของ  $Z^p$  ซึ่งเป็นฟังก์ชันของพารามิเตอร์ในแบบจำลอง ดูรายละเอียดของตัวอย่างแบบจำลอง AR ได้ใน [BJ16, Appendix A7.4] (กรณีเกาส์เซียน) ที่พบว่าฟังก์ชันภาวะน่าจะเป็นแม่นยำนั้นเป็นฟังก์ชันไม่เชิงเส้นใน  $(A_1, \dots, A_p)$  การใช้ฟังก์ชันภาวะน่าจะเป็นมีเงื่อนไขในตัวอย่าง 9.4 จึงเป็นทางเลือกที่ง่ายลงทางหนึ่งของการประมาณฟังก์ชันภาวะน่าจะเป็นแม่นยำ

### 9.3 การประมาณภายหลังสูงสุด

หัวข้อการประมาณ MLE กำหนดให้  $\theta$  เป็นพารามิเตอร์ของ  $f(y; \theta)$  ที่ต้องประมาณ ในหัวข้อนี้เรามีสสมมติฐานว่า

1.  $\theta$  เป็นตัวแปรสุ่มที่มีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ  $f_\theta(\theta)$
2.  $\theta$  และ  $y$  มีฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม คือ  $f(y, \theta)$

เราเรียก  $f_{\theta|y}(\theta|y)$  ว่าเป็นฟังก์ชันความหนาแน่นภายหลัง (posterior density function) เพราะตัวอย่างค่า  $\theta$  หนึ่งที่เราเห็นย่อมขึ้นกับ  $y$  ( $\theta$  และ  $y$  มีการแจกแจงร่วมกัน) ดังนั้น  $f_{\theta|y}(\theta|y)$  จึงมีความหมายว่า หลังจากที่เราสังเกตค่า  $y$  แล้ว จะช่วยให้เราตัดสินใจความน่าจะเป็นของ  $\theta$  ค่าหนึ่งได้อย่างไร

การประมาณภายหลังสูงสุด หรือ Maximum a Posteriori Estimation (MAP) คือการทำ  $\theta$  ที่ทำให้ฟังก์ชันความหนาแน่นความน่าจะเป็นภายหลังมีค่าสูงสุด

$$\hat{\theta}_{\text{map}} = \operatorname{argmax}_\theta f_{\theta|y}(\theta|y) \quad (9.28)$$

จากกฎของเบย์ (Bayes' rule)

$$f_{\theta|y}(\theta|y) = \frac{f_{y,\theta}(y, \theta)}{f_y(y)} = \frac{f_{y|\theta}(y|\theta)f_\theta(\theta)}{f_y(y)}$$

และพบว่า  $f_{\theta|y}(\theta|y) \propto f_{y|\theta}(y|\theta) \cdot f_\theta(\theta)$  ปัญหา MAP จึงสมมูลกับ

$$\hat{\theta}_{\text{map}} = \operatorname{argmax}_\theta f_{y|\theta}(y|\theta) \cdot f_\theta(\theta) \quad (9.29)$$

อันหมายความว่า สำหรับ  $\theta$  ที่กำหนดค่าหนึ่ง หากทำให้  $f_{y|\theta}(y|\theta)$  มีค่าสูงก็แปลว่า  $\theta$  ค่านั้นเป็นตัวเลือกที่ดี (ตามนัยของตัวประมาณ MLE) แต่เสริมด้วยการถ่วงน้ำหนักด้วยฟังก์ชันความหนาแน่นของ  $\theta$  ค่านั้นไว้ด้วย เทอม  $f_\theta(\theta)$  จึงเรียกว่า ฟังก์ชันความหนาแน่นก่อน (prior density function) เพราะหมายถึงความรู้ที่เรามีเกี่ยวกับพารามิเตอร์  $\theta$  ซึ่งในงานประยุกต์จริง ต้องอาศัยประสบการณ์ของผู้ใช้ในการสมมติ หรือหากไม่มีเหตุผลสนับสนุนเพียงพอว่าจะเลือก  $f_\theta(\theta)$  เป็นการแจกแจงหนึ่งที่คิดว่ารูปแบบอื่น อาจจะเลือก  $f_\theta(\theta)$  ที่มีสมบัติสังยุคก่อน (conjugate prior)<sup>6</sup> ได้

<sup>6</sup>  $f_\theta(\theta)$  ที่เป็นสมาชิกของคลาสการแจกแจงหนึ่งๆ มีสมบัติสังยุคก่อน ถ้าหากฟังก์ชันความหนาแน่นภายหลัง  $f_{\theta|y}(\theta|y)$  ยังคงเป็นสมาชิกของคลาสการแจกแจงนั้น ดูเพิ่มเติมใน [Bar12, §8.5]

---

## บทที่ 10

# การเลือกและตรวจสอบแบบจำลอง

---

บทนี้บรรยาย 2 กระบวนการที่สำคัญเกี่ยวกับแบบจำลอง อันได้แก่

1. การเลือกแบบจำลอง (model selection) คือการเลือกแบบจำลองจากกลุ่มตัวเลือกหนึ่งโดยใช้เกณฑ์การเลือกที่เหมาะสม กลุ่มแบบจำลองนั้นมาจากโครงสร้างแบบจำลองหนึ่งที่ใช้เลือก แต่มีความซับซ้อนของแบบจำลองด้วยดีกรีต่างกัน ตัวอย่างเช่น แบบจำลองฟังก์ชันพหุนามดีกรี  $n$  แบบจำลองเชิงเส้นปริภูมิสถานะอันดับ  $n$  ฟังก์ชันถ่ายโอนที่มีอันดับ  $n$  (อันดับของพหุนามส่วน) หรือแบบจำลอง ARMAX  $(p, q, r)$  กลุ่มแบบจำลองยังรวมถึงแบบจำลองชนิดที่ต่างกัน เช่น จะเลือกใช้แบบจำลองเชิงเส้น หรือแบบจำลองต้นไม้ หรือโครงข่ายประสาทเทียม เป็นต้น
2. การประเมินแบบจำลอง (model assessment) หรือการตรวจสอบแบบจำลอง (model validation) คือ การประเมินสมรรถนะของแบบจำลองที่เลือกไว้ว่ามีค่าคลาดเคลื่อนน้อยทั่วไป (generalization error) มากน้อยเพียงใด สิ่งสำคัญคือ การประเมินนี้จะทำบนข้อมูลชุดใหม่ ที่ไม่ได้ใช้ในการประเมินแบบจำลอง

สำหรับงานประยุกต์ที่มีข้อมูลล้นเหลือ แนวทางสำหรับ 2 กระบวนการข้างต้นคือ การแบ่งข้อมูลเป็น 3 ชุด ได้แก่

1. ข้อมูลชุดประมาณ หรือชุดฝึกสอน (estimation or train data) เพื่อใช้ในการประมาณแบบจำลอง จากเนื้อหาด้านทฤษฎีการประมาณในบทที่ 4 และบทที่ 9 จะพบว่า ความแม่นยำของแบบจำลอง (ที่วัดได้จากค่าเอนเอียงหรือ bias) ส่วนหนึ่งจะขึ้นกับจำนวนข้อมูลที่ใช้ประมาณ
2. ข้อมูลชุดตรวจสอบ (validation data) คือชุดที่ใช้ประมาณความผิดพลาดในการทำนายค่า  $y$  ของแบบจำลอง เพื่อวัตถุประสงค์การเลือกแบบจำลอง
3. ข้อมูลชุดทดสอบ (test data) คือข้อมูลที่เก็บแยกไว้เพื่อการประเมินสมรรถนะของแบบจำลองที่ได้เลือกไว้ ข้อมูลชุดนี้จะถูกใช้เป็นขั้นตอนสุดท้ายของการวิเคราะห์ข้อมูล และไม่ยุ่งเกี่ยวใดในขั้นตอนการประมาณหรือเลือกแบบจำลอง

การเลือกอัตราส่วนของข้อมูล 3 ชุดควรเลือกให้ชุดประมาณมีสัดส่วนที่มากกว่าส่วนอื่น เพราะส่งผลต่อความเอนเอียงของแบบจำลอง แต่จะเลือกเป็นจำนวนเท่าใดที่เหมาะสมนั้น ขึ้นกับปัจจัยด้านอื่นด้วย เช่น อัตราสัญญาณรบกวนต่อสัญญาณ (SNR) หรือเส้นโค้งการเรียนรู้ของวิธีประมาณเชิงสถิติที่ใช้ด้วย

(แต่ละวิธีมีสมรรถนะการประมาณขึ้นกับจำนวนข้อมูลด้วยอัตราที่ต่างกัน) อัตราส่วนของข้อมูล 3 ส่วนที่ลองใช้ได้คือ 50:25:25

ในทางกลับกัน หากงานประยุกต์ที่เราสนใจมีข้อมูลจำกัด และไม่สามารถแบ่งเป็น 3 ส่วนข้างต้นด้วยเหตุผลเช่น จำนวนข้อมูลที่ใช้ประมาณจะมีน้อยเกินไป อีกทางเลือกคือ เราสามารถแบ่งข้อมูลเป็น 2 ชุด (เรียกชุดประมาณและชุดทดสอบ) และใช้ชุดทดสอบเพื่อวัตถุประสงค์การเลือกแบบจำลอง เมื่อได้แบบจำลองที่เลือกแล้ว ใช้ค่าคลาดเคลื่อนทดสอบ เพื่อประมาณค่าคลาดเคลื่อนที่แท้จริง เนื้อหาบทนี้จึงเป็นเสนอแนวทางในสถานการณ์ข้อมูลจำกัดนี้ การเลือกแบบจำลองในหัวข้อ 10.2 มี 2 วิธี คือ การใช้เกณฑ์ในรูปแบบฟังก์ชันคณิตศาสตร์ที่เรียกว่า เกณฑ์การเลือกแบบจำลองและวิธีการตรวจสอบไขว้ จากนั้นจะอธิบายกระบวนการประเมิน (หรือตรวจสอบ) แบบจำลองพลวัต ว่าควรผ่านเกณฑ์ใดบ้าง ก่อนที่จะนำแบบจำลองนั้นไปใช้

หัวใจสำคัญของกระบวนการเลือกแบบจำลองและการประเมินแบบจำลอง คือการถ่วงดุลความเอนเอียงและความแปรปรวนของแบบจำลอง อันเป็น 2 สิ่งที่คานกันเสมอ ดังที่จะอธิบายในหัวข้อ 10.1

**ปัจจัยในการเลือกแบบจำลอง** วัตถุประสงค์หลักของการเลือกแบบจำลอง คือ การได้แบบจำลองที่นำไปอธิบายระบบได้ดี และมีต้นทุนในการคำนวณที่เหมาะสม ปัจจัยที่ส่งผลต่อวัตถุประสงค์ 2 ข้อนี้ได้แก่

1. **คุณภาพของแบบจำลอง:** คือปริมาณที่บ่งชี้ว่าแบบจำลองหนึ่งๆ อธิบายระบบที่สนใจได้ดีเพียงใด สามารถวัดได้ด้วยตัวบ่งชี้เช่น MSE ค่าฟังก์ชันลึกลับภาวะน่าจะเป็น เป็นต้น หากแบบจำลองมีความซับซ้อน ทำให้โครงสร้างของแบบจำลองมีความยืดหยุ่น (flexible) ในการอธิบายระบบได้ดีกว่าระบบที่ง่าย เช่น แบบจำลองไม่เชิงเส้นย่อมอธิบายระบบจริงที่มีความไม่เชิงเส้น ได้ดีกว่าการใช้แบบจำลองเชิงเส้น และสะท้อนได้จากค่า MSE ของแบบจำลองไม่เชิงเส้นที่ควรจะต่ำกว่า
2. **ต้นทุนของแบบจำลอง:** ขั้นตอนการประมาณแบบจำลองมักเกี่ยวข้องกับการแก้ปัญหาการหาค่าเหมาะที่สุด ความซับซ้อนเชิงคำนวณจึงขึ้นกับสมบัติของปัญหาคณิตศาสตร์นั้นๆ โดยทั่วไป โครงสร้างแบบจำลอง ความไม่เชิงเส้น และรูปแบบปัญหาประมาณ (เช่นฟังก์ชันสูญเสีย) ส่งผลกับรูปแบบปัญหาการหาค่าเหมาะที่สุด และส่งผลต่อวิธีเชิงตัวเลขที่นำมาประยุกต์แก้ปัญหาได้ เช่น แบบจำลองที่มีความเป็นเชิงเส้นในพารามิเตอร์สามารถใช้รูปแบบปัญหา LS ได้ จึงมีต้นทุนในการคำนวณที่ไม่ซับซ้อน หรือตัวอย่างแบบจำลอง ARMAX ที่มีความซับซ้อนกว่าแบบจำลอง ARX (มีคุณภาพที่ดีกว่า) แต่รูปแบบปัญหาที่ประยุกต์ใช้ได้กับการประมาณ ARMAX คือวิธี PEM จึงทำให้มีความซับซ้อนเชิงคำนวณที่สูงกว่า
3. **การนำแบบจำลองไปใช้:** วัตถุประสงค์ที่ต่างกันของการใช้แบบจำลอง มีความเกี่ยวข้องกับการคำนวณที่ต่างกัน กำหนดให้ตัวแปรตอบสนองหรือสัญญาณออกของแบบจำลองที่มีพารามิเตอร์  $\theta$  คือ  $\hat{y}$  และตัวแปรทำนายคือ  $x$ 
  - การคำนวณค่าพยากรณ์  $\hat{y}$  มีความซับซ้อนในการคำนวณตามฟังก์ชันคณิตศาสตร์ของแบบจำลอง หากใช้แบบจำลองพลวัต การคำนวณ  $\hat{y}$  นี้จะเกี่ยวข้องกับการบันทึก (cache) ค่าของตัวแปรที่เกี่ยวข้องในอดีตด้วย
  - การนำแบบจำลองไปใช้แทนระบบจริงในการออกแบบตัวควบคุม ต้องมีปัจจัยที่คำนึงถึงในขั้นตอนการประมาณเพิ่มเติม เพื่อให้แบบจำลองเหมาะสม เช่น อาจต้องมีการเพิ่มเงื่อนไขของ  $\theta$  ที่ทำให้แบบจำลองที่ได้มีเสถียรภาพ (หากระบบจริงมีเสถียรภาพ)
  - การทดสอบสมมติฐานที่สนใจผ่าน  $\theta$  ผู้ใช้แบบจำลองอาจจะไม่ได้ต้องการนำ  $\theta$  หรือแบบจำลองไปใช้โดยตรง แต่มีค่าทดสอบสถิติ (test statistic) ที่สนใจในปัญหาประยุกต์หนึ่งๆ

และค่าทดสอบนั้นเป็นฟังก์ชันของ  $\theta$  ตัวอย่างคือ การทดสอบนัยสำคัญของตัวประมาณในปัญหา LS ดังหัวข้อ 5.6 ที่บ่งชี้ว่าตัวแปรทำนายใดมีความสำคัญต่อการอธิบาย  $y$

- การอนุมานสมบัติของระบบ ผู้ใช้อาจพบว่าปริมาณทางสถิติที่สนใจ เป็นฟังก์ชันบางอย่างของ  $\theta$  ตัวอย่างใน บทที่ 11 จะแสดงงานประยุกต์ที่นำค่าประมาณพารามิเตอร์ ไปใช้เพื่ออนุมานโครงสร้างความสัมพันธ์ของตัวแปรในระบบได้

## 10.1 ความเอนเอียงและแปรปรวนของแบบจำลอง

การแปรระดับความซับซ้อนของแบบจำลองจะเป็นส่วนหนึ่งของการเลือกแบบจำลองเสมอ หัวข้อนี้แสดงทฤษฎีบทที่สำคัญเกี่ยวกับค่าเอนเอียง (bias) และความแปรปรวนของแบบจำลอง ผลลัพธ์นี้จะอธิบายว่า MSE ของการประมาณตัวแปรตอบสนองนั้น ประกอบไปด้วย 2 ค่าดังกล่าวที่คานกันเสมอ แบบจำลองที่ซับซ้อนจะมีค่าเอนเอียงต่ำ แต่จะมีความแปรปรวนสูง

**ทฤษฎีบท 10.1** (Bias-Variance decomposition). กำหนดให้ตัวแปรตอบสนอง  $Y$  ที่อธิบายด้วยตัวแปรทำนาย  $X$  นั้นเป็นไปตาม DGP บรรยายด้วย  $Y = f(X) + e$  โดยที่  $f$  แสดงถึงความสัมพันธ์ของ  $Y$  และ  $X$  ที่แท้จริง มี  $e$  เป็นสัญญาณรบกวนที่มี  $\mathbf{E}[e] = 0$  และ  $\mathbf{cov}(e) = \sigma^2$  ในขั้นตอนการจำลองระบบนั้น ผู้ใช้เลือก  $\hat{f}(X)$  ในการอธิบายระบบ เราจะพบว่า MSE ของการประมาณถดถอย  $\hat{f}(X)$  ที่ค่า  $X = x$  นั้นประกอบไปด้วย 3 เทอมดังนี้

$$\text{MSE} = \mathbf{E}[(Y - \hat{f}(X))^2 | X = x] = \sigma^2 + \text{Bias}^2(\hat{f}(x)) + \text{Var}(\hat{f}(x)) \quad (10.1)$$

**พิสูจน์.** จากข้อกำหนด  $f$  เป็นฟังก์ชันเชิงกำหนด (deterministic) และเมื่อทราบค่า  $X = x$  (ตัวแปรทำนาย ซึ่งเป็นตัวแปรสุ่ม มีค่าเป็น  $x$ ) จะได้ว่า  $\mathbf{var}(Y|X = x) = \sigma^2$  และ  $\mathbf{E}[Y|X = x] = f(x)$  ส่วนค่าฟังก์ชันของแบบจำลองนั้นซึ่งก็คือ  $\hat{f}(x)$  จะเป็นตัวแปรสุ่ม (เพราะโดยทั่วไป  $\hat{f}$  จะขึ้นกับ  $Y$  และพารามิเตอร์ในแบบจำลองจะขึ้นกับ  $Y$  ซึ่งเป็นตัวแปรสุ่มด้วย) เมื่อเราเริ่มจากนิยามของ MSE

$$\mathbf{E}[(Y - \hat{f}(X))^2 | X = x] = \mathbf{E}[Y^2 | X = x] + \mathbf{E}[\hat{f}(X)^2 | X = x] - \mathbf{E}[2Y\hat{f}(X) | X = x]$$

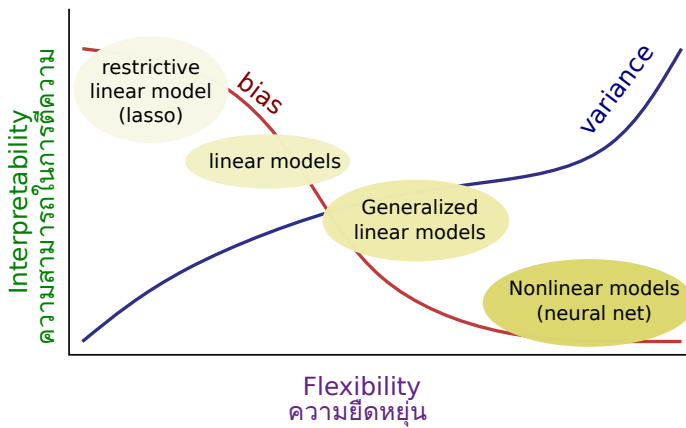
และใช้สัญลักษณ์ที่เป็นเงื่อนไขว่า  $X = x$  ด้วยการใช้เพียงสัญลักษณ์  $|X$  เราจะได้ว่า ค่า MSE สามารถเขียนได้เป็น

$$\begin{aligned} \text{MSE} &= \mathbf{var}(Y|X) + \mathbf{E}[Y|X]^2 + \mathbf{var} \hat{f}(x) + \mathbf{E}[\hat{f}(x)]^2 - 2f(x)\mathbf{E}[\hat{f}(x)] \\ &= \mathbf{var}(Y|X) + f(x)^2 + \mathbf{var} \hat{f}(x) + \mathbf{E}[\hat{f}(x)]^2 - 2f(x)\mathbf{E}[\hat{f}(x)] \\ &= \sigma^2 + \mathbf{var} \hat{f}(x) + (f(x) - \mathbf{E}[\hat{f}(x)])^2 \\ &= \sigma^2 + \mathbf{var} \hat{f}(x) + (\mathbf{E}[f(x) - \hat{f}(x)])^2 \\ &= \sigma^2 + \mathbf{var} \hat{f}(x) + [\text{Bias}(\hat{f}(x))]^2 \end{aligned}$$

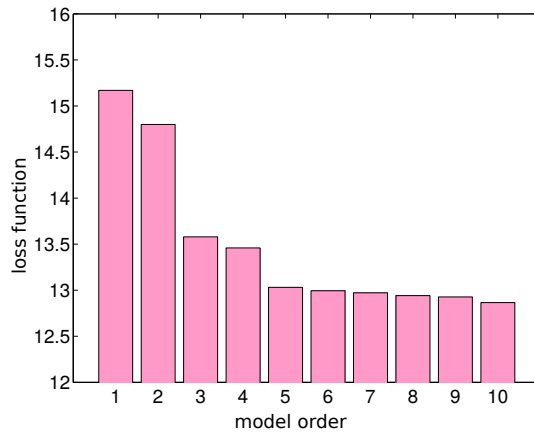


ทฤษฎีบท 10.1 อธิบายว่า หากข้อมูล  $Y$  นั้นมาจากกระบวนการ DGP ที่มีสัญญาณรบกวน ไม่ว่าเราจะใช้แบบจำลอง  $\hat{f}$  ใดเพื่อประมาณ  $Y$  ค่า MSE ที่ได้เมื่อใช้ค่า  $X = x$  หนึ่งๆ จะไม่สามารถลดลงได้จนเป็นศูนย์ ค่าคลาดเคลื่อนที่ลดไม่ได้ (irreducible error) มีค่าเท่ากับความแปรปรวนของสัญญาณรบกวนเท่ากับ  $\sigma^2$  สำหรับแบบจำลองที่มีซับซ้อน หรือยืดหยุ่นสูงนั้น จะมีค่าเอนเอียงที่ต่ำ แต่ผลลัพธ์





รูป 10.1: ความยืดหยุ่นและความสามารถในการตีความของแบบจำลอง [HTF09]



รูป 10.2: การประมาณแบบจำลอง AR เมื่อแปรอันดับของแบบจำลอง ค่าฟังก์ชันสูญเสียมีจุดหักที่ประมาณ  $p = 5$

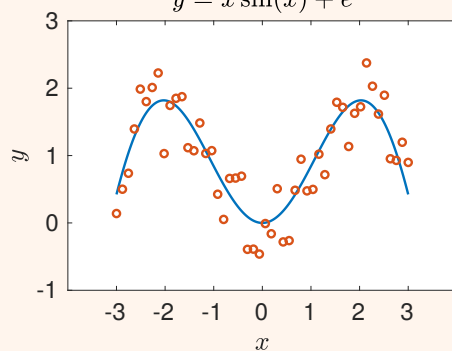
**ความเข้ากันเกิน.** *ความเข้ากัน* คือระดับความสอดคล้องกันระหว่างข้อมูลกับแบบจำลอง (fitting) ที่สามารถวัดได้จากตัวชี้วัดเช่น หากแบบจำลองเข้ากันกับข้อมูลดี จะมี MSE ต่ำ และฟังก์ชันล๊อคภาวน่าจะเป็นมีค่าสูง การใช้แบบจำลองอันดับสูงเกินไป อาจไม่ได้ให้ผลการประมาณดีขึ้นอย่างมีนัยสำคัญเสมอไป เนื่องจากว่า หากแบบจำลองมีความซับซ้อนเกินจำเป็น แบบจำลองนั้นจะพยายามเรียนรู้รายละเอียดการเปลี่ยนแปลงพลวัตของสัญญาณออก ที่จะเริ่มมาจากความไม่แน่นอนของสัญญาณรบกวนในภาวะเช่นนั้น ความเข้ากันอาจจะไม่ได้ดีขึ้นอย่างมีนัยสำคัญ รูป 10.2 แสดงตัวอย่างที่ข้อมูลนั้นเกิดจากระบบ AR อันดับ 5 ในขั้นตอนการประมาณนั้น สมมติว่าใช้แบบจำลอง AR อันดับ  $p$  และใช้วิธี LS ในการประมาณพารามิเตอร์ เมื่อแปรค่า  $p = 1, 2, \dots, 10$  พบว่าฟังก์ชันสูญเสียกำลังสองมีค่าลดลงอย่างมีนัยสำคัญเมื่อ  $1 \leq p \leq 5$  นั่นคือในช่วงนี้แบบจำลองมีโครงสร้างที่ซับซ้อนขึ้น และมีความสามารถในการอธิบายพลวัตหรือคุณลักษณะที่สำคัญของระบบได้ดีขึ้น เมื่อ  $p \geq 6$  ฟังก์ชันสูญเสียลดลงเพียงเล็กน้อย การใช้จุดหักของกราฟของฟังก์ชันสูญเสียเทียบกับความซับซ้อนของแบบจำลองที่ลดลงอย่างมีนัยสำคัญ จึงแนะนำของอันดับของแบบจำลองที่เหมาะสมได้

ว่าเป็น จำนวนยังผลของพารามิเตอร์ [HTF09, §7.6] หรืออาจใช้คำว่า องศาเสรี (degree of freedom) [HTW15, §2.5] จะเห็นว่า เมื่อแบบจำลองซับซ้อนมากขึ้นย่อมทำให้  $\hat{y}$  อธิบายแนวโน้มของ  $y$  ได้ดีขึ้น ส่งผลให้ค่าความแปรปรวนร่วมระหว่าง  $\hat{y}$  กับ  $y$  สูงขึ้น จึงทำให้ (10.14) เป็นตัวบ่งชี้ระดับความซับซ้อนของแบบจำลอง

- สำหรับกรณีเฉพาะที่การประมาณเป็นเชิงเส้นในตัวแปรข้อมูล นั่นคือ  $\hat{y} = Py$  สำหรับ  $P$  ที่ขึ้นกับตัวแปรทำนายอย่างเดียว (ไม่ขึ้นกับ  $y$ ) เราแสดงให้เห็นได้ว่า  $df(\hat{y}) = \text{tr}(P)$  (ดูแบบฝึกหัด 10.6) จากผลลัพธ์นี้ เราจึงเห็นว่าสำหรับแบบจำลองถดถอยเชิงเส้น  $y = X\beta$  ที่มี  $n$  ตัวแปรทำนาย และ  $\hat{y} = X(X^T X)^{-1} X^T y$  ดังนั้น  $df(\hat{y}) = \text{tr}(P) = \text{tr}(X(X^T X)^{-1} X^T) = \text{tr}(I_n) = n$  จำนวนยังผลของพารามิเตอร์จึงเท่ากับจำนวนตัวแปรทำนายที่ใช้ในแบบจำลองเชิงเส้น นั่นเอง สำหรับปัญหาถดถอยแบบบริดจ์เราพบว่า  $df = \sum_i \frac{\lambda_i}{\lambda_i + \gamma}$  เมื่อ  $\lambda_i$  คือค่าเฉพาะ (eigenvalue) ของ  $X^T X$  (ดูแบบฝึกหัด 10.7)
- สำหรับปัญหาแลสโซที่ใช้ค่าพารามิเตอร์ลงโทษ  $\lambda$  ค่าหนึ่ง จำนวนสัมประสิทธิ์ผลเฉลยที่ไม่เป็นศูนย์ ที่สมมติให้เท่ากับ  $k_\lambda$  นั้น เป็นค่าประมาณไม่เอนเอียงของ  $df$  [HTW15, §2.5] และสำหรับกรณีเฉพาะของแลสโซ เมื่อเมทริกซ์ถดถอย (regressor matrix) เป็นเมทริกซ์ตั้งฉาก จะได้ว่า  $df = k_\lambda$  (แบบฝึกหัด 2.8 ของ [HTW15]) อันเป็นผลลัพธ์ที่สอดคล้องกับความเข้าใจว่า หากผลเฉลยแลสโซเป็นเวกเตอร์เบาบาง เราจึงควรนับเฉพาะสมาชิกที่ไม่เป็นศูนย์ว่าเป็นพารามิเตอร์ยังผลของแบบจำลอง

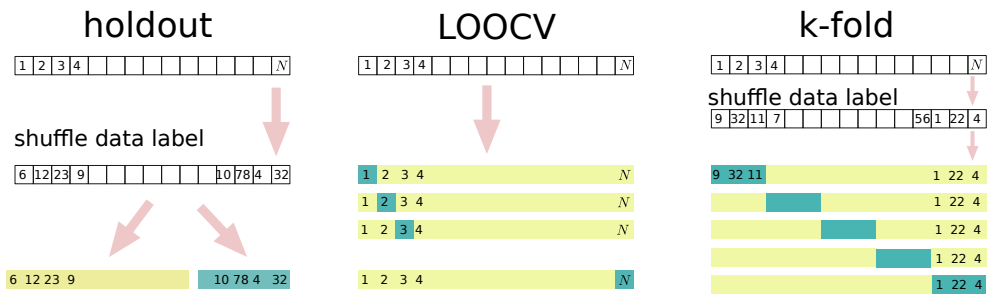
ตัวอย่าง 10.4 (การเลือกจำนวนตัวแปรในแบบจำลองถดถอยเชิงเส้น)

ข้อมูล  $y$  สร้างจากฟังก์ชัน  $y = x \sin(x) + e$  เมื่อ  $\text{var}[e] = 0.1$  เราประมาณจุดข้อมูลด้วย  $\hat{y} = \sum_{k=1}^n a_k \cos(kx) + b_k \sin(kx)$  เมื่อ  $n = 1, 2, \dots, 20$  และจำนวนข้อมูลเท่ากับ  $N = 50$  การเลือกแบบจำลองจึงหมายถึงการเลือกจำนวนของฟังก์ชันมูลฐานว่าจะเป็นฟังก์ชัน  $\cos(\cdot)$  และ  $\sin(\cdot)$  ที่ความถี่ใดบ้าง เราใช้  $d = 2n$  ซึ่งคือจำนวนเทอมของ  $\cos(kx), \sin(kx)$  ที่ใช้สำหรับ  $\hat{y}$



รูป 10.5 แสดงการจำลองเมื่อ  $N$  มีค่าไม่สูงมาก จะเห็นว่า AIC เลือกใช้จำนวนตัวแปรทำนายที่มาก (แบบจำลองซับซ้อน) และ AICc กับ BIC เลือกแบบจำลองที่ง่ายลง โดยที่ตัวแปรทำนาย 4 ตัวที่เลือกสอดคล้องกับฟังก์ชันมูลฐาน  $\sin(x), \cos(x), \sin(2x), \cos(2x)$  สำหรับค่า  $C_p$  ที่เลือก 7 ตัวแปรกับ 4 ตัวแปรนั้นมีค่าค่อนข้างใกล้เคียงกัน การใช้ adjusted  $R^2$  ยังคงเลือกแบบจำลองที่ซับซ้อนพอกับการใช้ AIC อยู่ และจะเห็นว่าการใช้  $R^2$  ไม่ให้ข้อมูลที่เป็นประโยชน์ต่อการเลือกตัวแปรทำนายเลย เพราะค่า  $R^2$  จะสูงขึ้นเสมอ เช่นเดียวกับค่า MSE ฝึกสอน (train MSE) ที่จะลดลงเสมอ แต่จะสังเกตว่าเมื่อเลือกตัวแปรทำนาย 4 ตัวแรกแล้ว ค่า MSE ทดสอบ (test MSE) ไม่ได้ลดลงอย่างมีนัยสำคัญ ณ จุดที่  $n = 4$  เราก็สังเกตเช่นเดียวกันว่าค่า MSE ทดสอบ ก็มีค่าสูงขึ้นเล็กน้อย บ่งชี้ว่าหากใช้  $n \geq 4$  แบบจำลองที่ได้จะมีภาวะเข้ากันเกิน

ในตัวอย่างต่อไป เราจะแสดงผลการเลือกแบบจำลองพลวัตที่ประมาณกับข้อมูลอนุกรมเวลา



รูป 10.7: การแบ่งข้อมูลสำหรับการตรวจสอบไขว้

$i = 1, 2, \dots, N$  เราจะใช้ค่า

$$CV_{\text{loocv}} = \frac{1}{N} \sum_{i=1}^N \text{MSE}_i = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (10.21)$$

อันเป็นค่าเฉลี่ยของ MSE บน  $N$  ชุดตรวจสอบ เพื่อเป็นค่าประมาณของ MSE ทดสอบ การตรวจสอบแบบนี้ เป็นกรณีเฉพาะของการใช้ข้อมูลตรวจสอบ เพราะเป็นการแบ่งจำนวนข้อมูลเพียง 1 ตัวอย่างบนชุดตรวจสอบ จึงมีชื่อเรียกว่า leave-one-out cross validation หรือ LOOCV ดังรูป 10.7

ข้อดีของวิธีนี้เมื่อเทียบกับวิธีการใช้ข้อมูลตรวจสอบ คือ ชุดข้อมูลประมาณมี  $N - 1$  ตัวอย่าง ซึ่งเกือบจะเป็นข้อมูลทั้งหมด ดังนั้น LOOCV จึงไม่ค่อยจะประเมินค่าคลาดเคลื่อนทดสอบสูงเกินไป ข้อดีที่สองคือ แบบจำลองที่ประมาณทั้ง  $N$  ครั้งนั้นมีความสัมพันธ์กันสูง เพราะข้อมูลมีความเหลื่อมกันสูง ค่า  $\text{MSE}_i$  ที่คำนวณในแต่ละชุดตรวจสอบจึงมีสหสัมพันธ์ร่วมกันสูง ข้อเสียของวิธี LOOCV คือมีการวนซ้ำการประมาณ  $N$  ครั้ง ดังนั้น หาก  $N$  มีค่ามาก และแบบจำลองที่ใช้มีความซับซ้อนในการคำนวณที่สูง จะใช้เวลาในการคำนวณ  $\hat{y}_i$  ตาม (10.21) ที่นาน

### การตรวจสอบไขว้แบบ $k$ ส่วน

วิธีตรวจสอบไขว้แบบ  $k$  ส่วน ( $k$ -fold cross validation) แบ่งข้อมูลอย่างสุ่มออกเป็น  $k$  ส่วนๆ ละจำนวนตัวอย่างที่พอเท่ากัน ข้อมูลส่วนแรกจะเก็บไว้เป็นชุดตรวจสอบ ส่วน  $k - 1$  ที่เหลือเป็นชุดประมาณ ค่า MSE ประเมินบนชุดตรวจสอบถูกเก็บไว้เป็น  $\text{MSE}_1$  จากนั้น กระบวนการจะวนซ้ำ  $k$  ครั้ง โดยที่จะใช้ข้อมูลส่วนที่สอง สาม และ  $k$  เป็นชุดตรวจสอบ ตามลำดับดังรูป 10.7 เราจึงได้ค่า  $\text{MSE}_1$   $\text{MSE}_2$  จนถึง  $\text{MSE}_k$  จากนั้นเราใช้ค่าเฉลี่ยดังสมการ

$$CV_{k\text{-fold}} = \frac{1}{k} \sum_{i=1}^k \text{MSE}_i \quad (10.22)$$

เพื่อเป็นค่าประมาณของค่าคลาดเคลื่อนทดสอบ เราพบว่า เมื่อกำหนดให้  $k = N$  วิธีนี้จึงเป็นวิธีเดียวกับ LOOCV (LOOCV เป็นกรณีเฉพาะของวิธีตรวจสอบไขว้แบบ  $k$  ส่วน) เมื่อใช้วิธีนี้ อัตราส่วนข้อมูลประมาณจะเป็น  $(k - 1)/k\%$  ของข้อมูลทั้งหมด ในทางปฏิบัติหากเราเลือกให้  $k$  อยู่ในช่วง 5–10 ก็จะเป็นการใช้ข้อมูลประมาณ 80–90 % ของข้อมูลทั้งหมด แบบจำลองที่ประมาณจากการวน  $\text{MSE}$  ส่วน  $i$  หนึ่งๆ มีความเอนเอียงมากกว่าวิธี LOOCV (เพราะใช้ข้อมูล  $\frac{k-1}{k}\%$  ซึ่งน้อยกว่า  $\frac{N-1}{N}\%$ )

รูป 10.8 แสดงตัวอย่างการตรวจสอบแบบจำลองพหุนามทั้ง 3 วิธี เมื่อมีข้อมูล  $N = 500$  โดยที่วิธี  $k$ -fold และวิธี holdout นั้น จะทำซ้ำ 7 ครั้งเพื่อสังเกตความแปรปรวนของ MSE ที่ได้จาก 3

## 10.4 ตัวอย่างคำสั่ง MATLAB

จากเนื้อหาที่ได้อธิบายมานั้น ปริมาณที่ส่งผลต่อการเลือกแบบจำลองได้แก่  $\hat{\theta}$ ,  $\text{cov}(\hat{\theta})$  ฟังก์ชันล๊อกรภาวะน่าจะเป็นของ  $\hat{\theta}$  เกณฑ์ AIC, AICc หรือ BIC ค่า MSE การคำนวณ  $\hat{y}$  ที่เป็นค่าประมาณของ  $y$  จากแบบจำลองที่ประมาณได้ การหาค่าคลาดเคลื่อนส่วนเหลือ  $\varepsilon = y - \hat{y}$  ปริมาณเหล่านี้ ส่วนใหญ่จะส่งกลับมาจากชุดคำสั่งที่ใช้ประมาณแบบจำลองของ MATLAB

**กล่องเครื่องมือ** curvefit มีตัวอย่างง่ายคือ แบบจำลองที่ประมาณจากคำสั่ง fit

รหัสคำสั่งที่ 10.1: การประมาณแบบจำลองด้วยคำสั่ง fit

```
load census; x = cdate ; y = pop ; % use MATLAB example data
[f,gof,output] = fit(x,y,'poly2'); % fit with a quadratic polynomial
yhat = feval(f,x);
plot(f);
```

ตัวแปร f จะรายงาน  $\hat{\theta}$  และช่วงความเชื่อมั่น การคำนวณ  $\hat{y}$  สามารถใช้คำสั่ง feval กับตัวแปร f โดยตรง หากจะคำนวณ  $\hat{y}$  ที่ใช้ข้อมูลใหม่ก็เพียงเปลี่ยนตัวแปร x การจะดูกราฟผลการประมาณใช้คำสั่ง plot และการรายงานค่าผลการประมาณเช่น RSS  $R^2$  หรือ ค่า  $R^2$  ที่ปรับแก้ สามารถดูได้จาก gof

**กล่องเครื่องมือ** Statistics and Machine Learning มีตัวอย่างคือ แบบจำลองเชิงเส้นที่ประมาณจากคำสั่ง fitlm ที่รายงานทุกปริมาณไว้ในตัวแปรขาออก (m) เพียงตัวเดียว

รหัสคำสั่งที่ 10.2: การประมาณแบบจำลองด้วยคำสั่ง fitlm

```
N=50; n = 5; X = randn(N,n) ; y = randn(N,1); % Generate data
m = fitlm(X,y) ; % fit the linear model including the intercept term
m.Coefficients ; % the estimated parameter
m.CoefficientCovariance % the covariance matrix of estimated parameter
m.Fitted % yhat estimated from the training data
m.LogLikelihood % loglikelihood value (Gaussian assumption)
m.ModelCriterion % AIC, AICc, BIC
m.Residuals % Residuals
m.Rsquared % Both ordinary R^2 and adjusted R^2
m.SSE % Residual sum squares or RSS
m.MSE % Mean squared error computed on training data
```

เกณฑ์การเลือกแบบจำลอง และค่าฟังก์ชันล๊อกรภาวะน่าจะเป็นที่รายงานใน MATLAB เป็นสูตรเดียวกับ (10.16) การคำนวณ  $\hat{y}$  จากข้อมูลชุดใหม่จากแบบจำลองนี้ สามารถใช้คำสั่ง feval(m,new\_x) ส่วนค่า MSE ที่รายงานนั้น คือปริมาณ  $\text{RSS}/(N - n - 1)$

สำหรับการตรวจสอบแบบจำลองด้วยวิธีตรวจสอบไขว้ หรือการใช้เกณฑ์การเลือกแบบจำลอง จะยกตัวอย่างเป็นการเลือกพารามิเตอร์ลงโทษ  $\gamma$  สำหรับปัญหาแลสโซ ตามที่อธิบายในตัวอย่าง 10.6 ชุดคำสั่งเริ่มด้วยการสร้างสัมประสิทธิ์ของแบบจำลองในตัวแปร r ที่มีสมาชิกไม่เป็นศูนย์เพียง 3 ตัว (จาก 50 ตัว) และเก็บรูปแบบศูนย์ที่แท้จริงไว้ในตัวแปร truesupport การประมาณแลสโซใช้คำสั่ง lasso ที่มีตัวเลือกการใช้ตรวจสอบไขว้เพื่อเลือก  $\gamma$  (ในคำสั่งใช้ชื่อพารามิเตอร์ว่า  $\lambda$  แทน  $\gamma$ ) ออกมาในช่วงหนึ่งๆ จากนั้นการประมาณ LS อีกครั้งด้วยเงื่อนไขบังคับของรูปแบบศูนย์ของผลเฉลยสำหรับ  $\lambda$  ค่าหนึ่งๆ จะใช้วนลูป ซึ่งควรใช้ parfor ร่วมด้วย หากจำนวน  $\lambda$  ที่พิจารณามีมาก จากนั้นนำค่าเกณฑ์การเลือกแบบจำลอง และการคำนวณมาพิจารณาว่าจะมีค่าต่ำสุดเมื่อ  $\lambda$  เป็นค่าใด และเปรียบเทียบ

## บทที่ 11

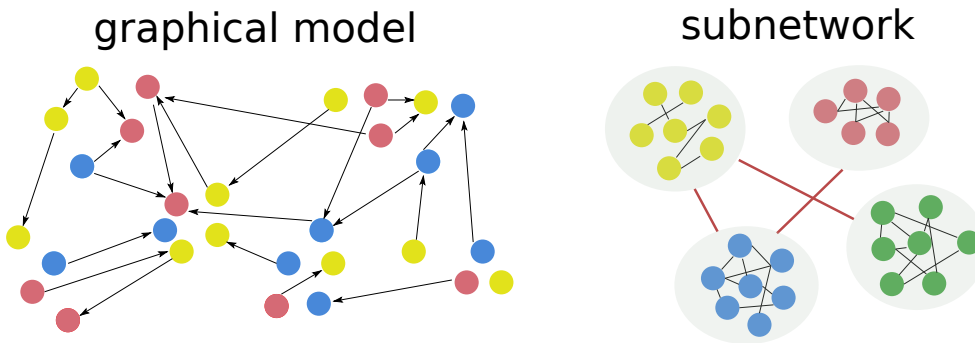
# งานประยุกต์ด้านการเรียนรู้ความสัมพันธ์ของตัวแปรในระบบ

จากหัวข้อ 1.1 นอกจากการประมาณพารามิเตอร์ของแบบจำลองจะเป็นผลลัพธ์โดยตรงที่ต้องการการอนุมานรูปแบบบางอย่างจากพารามิเตอร์ของแบบจำลอง ก็เป็นงานประยุกต์หนึ่งที่สำคัญ บทนี้จึงเสนองานวิจัยด้านการเรียนรู้ความสัมพันธ์เชิงโครงสร้างของตัวแปรในระบบอันซับซ้อน ที่สรุป หรือสกัดได้จากฟังก์ชันทางคณิตศาสตร์ที่ต่างกันไปของพารามิเตอร์แบบจำลอง โครงสร้างเหล่านั้นได้แก่ ความสัมพันธ์เชิงกราฟของเวกเตอร์ส่มหรือเวกเตอร์อนุกรมเวลา ที่นิยามด้วยความเป็นอิสระต่อกันแบบไม่มีเงื่อนไข และความสัมพันธ์เชิงกราฟที่อธิบายด้วยเหตุกาณ์ของเกรนเจอร์ (Granger causality)

### 11.1 โครงสร้างระบบซับซ้อน

งานประยุกต์ที่ศึกษาความสัมพันธ์รูปแบบง่ายของหลายตัวแปรในระบบซับซ้อน ดังรูป 11.1 แสดงโครงสร้างความสัมพันธ์นั้นในรูปแบบของแบบจำลองเชิงกราฟ (graphical model) อันมีวงกลมแทนโหนดหรือตัวแปรของระบบ หากตัวแปรเหล่านั้นไม่มีสมบัติเชิงสถิติที่เปลี่ยนไปตามเวลา เราอาจจำลองเป็นเวกเตอร์ส่มที่ใช้สัญลักษณ์เป็น  $X = (X_1, X_2, \dots, X_n)$  หรือหากตัวแปรเหล่านั้นมีพลวัตที่เปลี่ยนไปตามเวลา เราอาจจำลองเป็นเวกเตอร์อนุกรมเวลา (หรือเวกเตอร์กระบวนการส่ม) ที่ใช้สัญลักษณ์เป็น  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$  ลูกศรจากโหนด  $i$  ไปยังโหนด  $j$  จะหมายถึงความสัมพันธ์จากตัวแปร  $X_i$  ไปยังตัวแปร  $X_j$  (หรือหมายถึงกระบวนการ  $x_i(t)$  ไปยังกระบวนการ  $x_j(t)$ ) คำถามที่สำคัญคือ การมีความเชื่อมโยงระหว่าง 2 ตัวแปรนั้น จะอธิบายได้ด้วยคณิตศาสตร์เชิงสถิติอย่างไร จากหัวข้อ 1.1 เรามีตัวอย่างในเชิงระบบพลวัตว่า เราสามารถอนุมานรูปแบบของศูนย์จากฟังก์ชันถ่ายโอนของระบบตามความสัมพันธ์  $y = Gu$  เพราะหาก  $G_{ij} = 0$  ย่อมหมายถึงว่า  $u_j$  ไม่ส่งผลโดยตรงสัญญาณออก  $y_i$  เป็นต้น โครงสร้างเครือข่าย (network) ในแบบจำลองเชิงกราฟอาจจะมีการจัดเรียงตัวแปรเป็นกลุ่ม (cluster) อันบังชี้ถึงระบบย่อยภายใต้ระบบรวม การมีสมมติฐานโครงสร้างเครือข่ายย่อยอันแสดงในรูป 11.1 ก็นำมาเป็นสมมติฐานที่ใช้ในการประมาณแบบจำลองได้เช่นกัน

รูป 11.2a แสดงการประยุกต์ใช้แบบจำลองเชิงกราฟเพื่ออธิบายการเชื่อมโยงการทำงานของสมอง โหนดจะหมายถึงบริเวณต่างๆ ของสมอง และเส้นเชื่อมจากโหนด A ไปยังโหนด B จะแสดงว่า บริเวณ A ของสมองส่งผลบางอย่างกับกิจกรรมของสมองในบริเวณ B ด้วย การทราบโครงสร้างดังกล่าวย่อมทำให้ผู้เชี่ยวชาญในงานด้านสมอง เข้าใจฟังก์ชันการทำงานของสมองได้ดีขึ้น เช่น หากมีสมองของคน



รูป 11.1: โครงสร้างความสัมพันธ์เชิงกราฟของตัวแปรในระบบซับซ้อน

ปกติ และคนไข้ทางสมองบางอย่าง (เช่น ผู้ป่วยโรคอัลไซเมอร์ หรือโรคจิตเภท) และเราสกัดโครงสร้างการทำงานของสมองจากผู้ป่วย 2 กลุ่มได้ต่างกันในช่วงบริเวณ ทำให้ตีความได้ว่าโรคดังกล่าว มีความเกี่ยวข้องกับบริเวณสมองนั้นๆ อย่างไรก็ดี ระดับการเชื่อมโยงที่ซับซ้อนมากจนกราฟมีลิงก์หนาแน่นสูง หากเทียบกับระดับการเชื่อมโยงที่เบาบางลงจนเห็นกราฟเฉพาะลิงก์เชื่อมที่สำคัญ ผู้ใช้งานย่อมแปลผลของการเชื่อมโยงอันเบาบางได้ง่ายกว่า ข้อสังเกตนี้จึงเป็นเหตุสำคัญที่เราควรเลือกระหว่าง แบบจำลองที่ซับซ้อนซึ่งอธิบายข้อมูลได้ดีแต่แปลผลยาก กับแบบจำลองง่ายซึ่งอธิบายข้อมูลได้แยกลง แต่แปลผลได้ง่ายกว่า ดังที่อธิบายในหัวข้อ 10.1 นั่นเอง

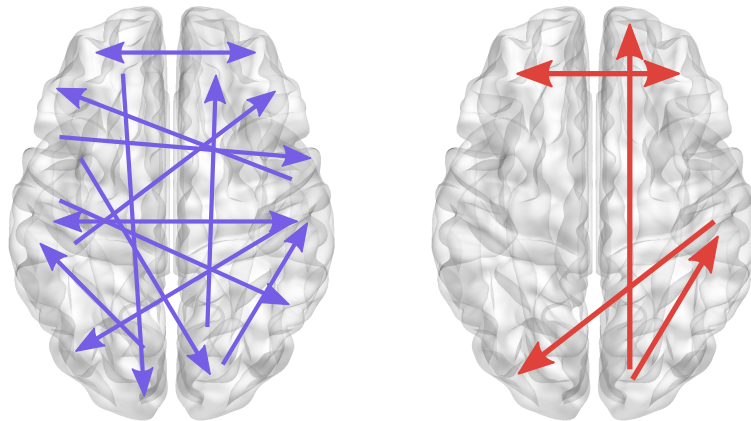
งานประยุกต์ด้านเศรษฐศาสตร์ มีการศึกษาโครงสร้างความสัมพันธ์ของกลุ่มตัวแปรทางการเงินของประเทศต่างๆ ดังรูป 11.2c เช่น โหนดหมายถึงราคาหุ้นสำคัญของประเทศที่สนใจ [SZLW14] เป็นต้น ตัวอย่างในรูป 11.2b คือ การศึกษาตัวแปรมลภาวะและตัวแปรทางอากาศ เช่น คาร์บอนมอนนอกไซด์ ซัลเฟอร์ไดออกไซด์ ฝุ่น PM2.5 ความเข้มข้นแสงอาทิตย์ อุณหภูมิ ความชื้น ว่ามีความสัมพันธ์กันอย่างไร โครงสร้างความสัมพันธ์นี้อาจจะแปรเปลี่ยนไปตามพื้นที่หรือไม่ เช่น ฝุ่นในกรุงเทพมหานครกับฝุ่นทางภาคเหนืออาจเกิดมาจากคนละปัจจัย การเรียนรู้โครงสร้างนั้นจากข้อมูลจึงช่วยให้ผู้ใช้งานเข้าใจพลวัตของระบบสิ่งแวดล้อมได้ดีขึ้น

## 11.2 ความสัมพันธ์ของตัวแปรนิยามจากอะไร?

ในแบบจำลองเชิงกราฟนั้น การที่ลิงก์ความสัมพันธ์ของตัวแปรคู่หนึ่งขาดหายไป จะแสดงให้เห็นได้ว่าสมมูลกับเงื่อนไขคณิตศาสตร์อย่างใดอย่างหนึ่งของพารามิเตอร์แบบจำลอง เช่น การใช้แบบจำลอง  $y = Gu$  ถ้าหาก  $G_{ij} = 0$  แล้วจะไม่มีลิงก์เชื่อมระหว่าง  $u_j$  กับ  $y_i$  เราจะกล่าวว่าเงื่อนไขนี้เป็นสมมติฐานก่อน (prior assumption) ที่เรามีต่อแบบจำลอง เงื่อนไขดังกล่าว สามารถนิยามได้หลากหลายในงานวิจัย ขึ้นอยู่กับสมมติฐานเกี่ยวกับตัวแปรในระบบและหลักการทางสถิติว่าผู้ใช้ตีความลิงก์ดังกล่าวจากมุมใด [PS16] มุมมองของเงื่อนไขความสัมพันธ์สามารถแบ่งได้หลายแบบ ดังนี้

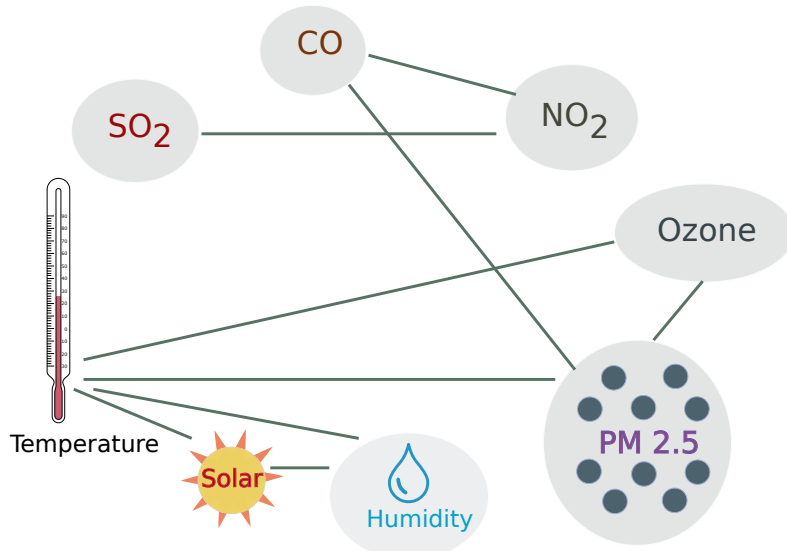
- ประเภทของตัวแปรในระบบ เช่น ตัวแปรสุ่ม หรืออนุกรมเวลา
- ทิศทางของความสัมพันธ์ เช่น มีทิศทางแบบทางเดียว (directional) หรือแบบสมมาตร
- การที่เงื่อนไขความสัมพันธ์นั้นอิงกับแบบจำลองที่ใช้หรือไม่

ตัวอย่างแรก คือการใช้เมทริกซ์ความแปรปรวนร่วม  $\Sigma$  เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรสุ่ม ที่กล่าวว่าตัวแปร  $X_i$  และ  $X_j$  ไม่มีสหสัมพันธ์ต่อกันหาก  $\Sigma_{ij} = 0$  ตัวอย่างนี้ เป็นความสัมพันธ์แบบ

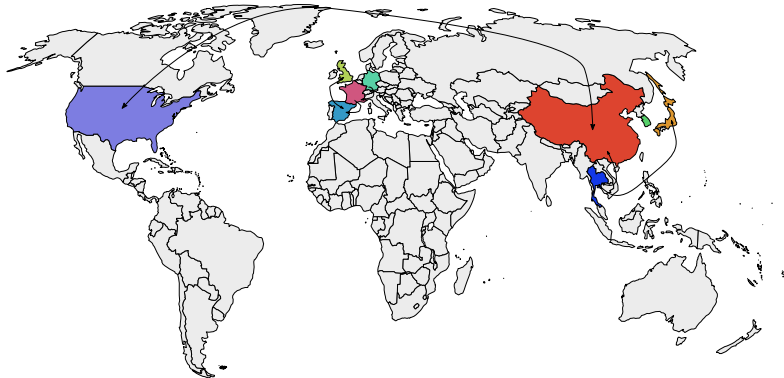


dense and sparse brain networks

(a) การเชื่อมโยงของบริเวณต่างๆ ในสมอง



(b) ตัวแปรในระบบมลภาวะทางอากาศ



(c) ดัชนีทางเศรษฐกิจระหว่างประเทศต่างๆ

รูป 11.2: ตัวอย่างงานประยุกต์การแสดงความสัมพันธ์ของตัวแปรในระบบที่สนใจ

ภาคผนวก





---

## ภาคผนวก A

### แคลคูลัส

---

#### A.1 อนุพันธ์ของฟังก์ชันหลายตัวแปร

กำหนดให้  $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$  และ  $x \in \text{int dom } f$  ( $x$  เป็นจุดภายในของโดเมนของ  $f$ ) เราใช้สัญลักษณ์  $Df(x) \in \mathbf{R}^{m \times n}$  ว่าเป็น **อนุพันธ์** หรือ **จาโคเบียน (Jacobian)** ของ  $f$  ณ ค่า  $x$

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

ตัวอย่างเช่น  $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ ,  $f(x) = Ax + b$  เราจะได้ว่า  $Df(x) = A$

หาก  $f$  เป็นฟังก์ชันสเกลาร์ (หรือ  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ ) จะได้ว่า  $Df(x)$  กลายเป็นเวกเตอร์แถวขนาด  $1 \times n$  ในกรณีนี้ เราจะเรียกเวกเตอร์สลับเปลี่ยน (transpose) ของจาโคเบียนว่าเป็น **เกรเดียนต์** ของฟังก์ชัน  $f$

$$\nabla f(x) = Df(x)^T, \quad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \dots, n$$

ซึ่งจะเป็นคอลัมน์เวกเตอร์ใน  $\mathbf{R}^n$  สำหรับ  $f$  ที่เป็นสเกลาร์ เราเรียกอนุพันธ์อันดับสองว่าเป็น **เมทริกซ์เฮสเซียน (Hessian matrix)** ของ  $f$  ณ ค่า  $x$  และนิยามโดย

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, n$$

**ฟังก์ชันกำลังสอง.** กำหนดให้  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  โดยที่  $f(x) = (1/2)x^T P x + q^T x + r$  และมี  $P \in \mathbf{S}^n$ ,  $q \in \mathbf{R}^n$  และ  $r \in \mathbf{R}$  เราพบว่า  $\nabla f(x) = P x + q$  และ  $\nabla^2 f(x) = P$

**กฎลูกโซ่ (Chain rule).** สำหรับฟังก์ชัน  $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$  และ  $g : \mathbf{R}^m \rightarrow \mathbf{R}^p$  โดยที่  $f$  สามารถหาอนุพันธ์ได้ที่  $x \in \text{int dom } f$  และ  $g$  สามารถหาอนุพันธ์ได้ที่  $f(x) \in \text{int dom } g$  หากเรากำหนดให้  $h : \mathbf{R}^n \rightarrow \mathbf{R}^p$  เป็นฟังก์ชันประกอบของ  $g$  และ  $f$  นั่นคือ  $h(x) = g(f(x))$  จะได้ว่า  $h$  สามารถหาอนุพันธ์ได้ที่  $x$  และมีอนุพันธ์เป็น

$$Dh(x) = Dg(f(x))Df(x)$$

## ภาคผนวก B

### เมทริกซ์

ในภาคผนวกนี้จะสรุปสมบัติที่สำคัญเกี่ยวกับเมทริกซ์สมมาตร หรือเมทริกซ์บวกกึ่งแน่นอน ผลลัพธ์เหล่านี้ศึกษาเพิ่มเติม หรือดูบทพิสูจน์ได้จากตำราทางพีชคณิตเชิงเส้น หรือเมทริกซ์วิเคราะห์ทั่วไป เช่น [HJ13]

**ประพจน์ B.1** (สมบัติของเมทริกซ์สมมาตร). กำหนดให้  $A \in \mathbf{S}^n$  อันเป็นเมทริกซ์สมมาตรค่าจริง ขนาด  $n \times n$  กำหนดให้  $\lambda_i$  และ  $u_i$  คือค่าเฉพาะ (eigenvalue) และเวกเตอร์เฉพาะ (eigenvector) ที่คู่กันของ  $A$  ตัวที่  $i$  ใดๆ เราพบว่า  $A$  มีสมบัติดังนี้

- สำหรับ  $x \in \mathbf{R}^n$  ใดๆ จะพบว่า  $x^T Ax = x^T \left( \frac{A+A^T}{2} \right) x$  เสมอ นั่นคือฟังก์ชันกำลังสองของ  $x$  ที่กำหนดด้วย  $A$  จะคำนวณได้จากการใช้ส่วนสมมาตร  $(A + A^T)/2$  เท่านั้น
- ค่าเฉพาะของ  $A$  เป็นค่าจริงเสมอ
- $u_i$  และ  $u_j$  จะตั้งฉากกันเสมอ เมื่อ  $i \neq j$  ถ้าหาก  $A$  มีค่าเฉพาะที่ซ้ำกัน เรายังสามารถหาเวกเตอร์เฉพาะ 2 ตัวที่เป็นอิสระเชิงเส้นต่อกันและทำให้ตั้งฉากกันได้เสมอ
- $A$  สามารถแยกตัวประกอบค่าเฉพาะได้เป็น  $A = UDU^T$  โดยที่  $U^T U = U U^T = I$  และ  $D$  เป็นเมทริกซ์ทแยงมุมประกอบไปด้วยค่าเฉพาะของ  $A$
- สำหรับ  $x \in \mathbf{R}^n$  ใดๆ  $\lambda_{\min}(A) \|x\|_2^2 \leq x^T Ax \leq \lambda_{\max}(A) \|x\|_2^2$

**นิยาม B.1** (เมทริกซ์บวกกึ่งแน่นอน). สำหรับ  $A \in \mathbf{S}^n$  เราจะกล่าวว่า  $A \succeq 0$  ถ้าหาก

$$x^T Ax \geq 0 \quad \text{สำหรับทุกๆ } x \in \mathbf{R}^n$$

และใช้สัญลักษณ์ว่า  $A \succ 0$  (เมทริกซ์บวกแน่นอน) ถ้าหาก  $x^T Ax > 0$  สำหรับทุกๆ  $x \neq 0$  ■

เราพิสูจน์ให้เห็นได้ว่า  $A \succeq 0$  ก็ต่อ  $\lambda_i(A) \geq 0$  สำหรับทุก  $i$  (ค่าเฉพาะของ  $A$  ไม่เป็นลบทุกตัว)

**นิยาม B.2** (รากที่สองของเมทริกซ์). สำหรับ  $A \succeq 0$  เราเรียก  $A^{1/2}$  ว่าเป็นเมทริกซ์รากที่สองของ  $A$  หาก  $A^{1/2} A^{1/2} = A$  ■

เราสามารถหา  $A^{1/2}$  ได้หลายวิธี หนึ่งในนั้นคือการแยกตัวประกอบค่าเฉพาะของ  $A$  เป็น  $A = UDU^T$  (eigenvalue decomposition) และกำหนดให้

$$A^{1/2} = U D^{1/2} U^T, \quad D^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}) \quad (\text{B.1})$$

## ภาคผนวก C

# ตัวแปรสุ่มและกระบวนการสุ่ม

### C.1 การลู่เข้าของตัวแปรสุ่ม

เนื้อหาส่วนนี้จะกล่าวถึงนิยามการลู่เข้าของลำดับตัวแปรสุ่ม  $\{X_n : n = 1, 2, \dots\}$  โดยที่  $n$  เป็นตัวชี้ของลำดับ ว่ามีการลู่เข้าในความหมายต่างๆ รวมถึงจะอธิบายสมบัติของลิมิตของการลู่เข้านั้น ในทางสถิติ หัวข้อนี้เป็นเรื่องทฤษฎีเชิงเส้นกำกับ (asymptotic theory) และเมื่อประยุกต์ใช้กับทฤษฎีการประมาณ ลำดับตัวแปรสุ่มในที่นี้คือตัวประมาณ  $\hat{\theta}_N$  โดยที่  $N$  เป็นจำนวนข้อมูลที่ใช้และเป็นตัวชี้ลำดับในบริบทการลู่เข้านี้

**นิยาม C.1** (การลู่เข้าเชิงความน่าจะเป็น). ลำดับตัวแปรสุ่ม  $X_n$  **ลู่เข้าเชิงความน่าจะเป็น** (converge in probability) ไปหาตัวแปรสุ่ม  $X$  ถ้าสำหรับ  $\epsilon > 0$  แล้ว

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

เราใช้สัญลักษณ์  $X_n \xrightarrow{p} X$  แทนการลู่เข้าเชิงความน่าจะเป็นนี้ และเรียก  $X$  ว่าเป็น **ลิมิตความน่าจะเป็น** (probability limit หรือ plim) ของ  $X_n$  :  $\text{plim } X_n = X$  ■

การลู่เข้าเชิงความน่าจะเป็นไปยังค่าคงที่  $c$  ก็สามารถนิยามได้เช่นเดียวกันกับนิยาม C.1

**นิยาม C.2** (การลู่เข้าเชิงการแจกแจง). ลำดับตัวแปรสุ่ม  $X_n$  **ลู่เข้าเชิงการแจกแจง** (converge in distribution) ไปยังตัวแปรสุ่ม  $X$  โดยที่ใช้สัญลักษณ์ว่า  $X_n \xrightarrow{d} X$  ถ้าหาก

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad \forall x \in \mathbf{R}$$

เมื่อ  $F_n$  และ  $F$  คือฟังก์ชันการแจกแจงสะสมของ  $X_n$  และของ  $X$  ตามลำดับ นอกจากนี้ เราจะเรียก  $F(x)$  ว่าเป็น **การแจกแจงขีดจำกัด** (limiting distribution) ■

ทฤษฎีบทลิมิตกลาง เป็นตัวอย่างสำคัญของผลลัพธ์การลู่เข้าของตัวแปรสุ่มเชิงการแจกแจง อันมีใจความว่า หาก  $X_n$  เป็นตัวอย่างสุ่ม i.i.d. มาจากการแจกแจงหนึ่งๆ ที่มีค่าเฉลี่ยเป็น  $\mu$  ผลลัพธ์คือ  $\sqrt{N}(\bar{X}_N - \mu)$  จะลู่เข้าเชิงการแจกแจงสู่ตัวแปรเกาส์เซียน  $\mathcal{N}(0, \sigma^2)$  ตัวอย่างอื่นของการลู่เข้าเชิงการแจกแจง เช่น 1) ตัวแปร  $t_{n-1} \xrightarrow{d} \mathcal{N}(0, 1)$  ตัวแปร  $t$  ลู่เข้าหาการแจกแจงเกาส์เซียน หรือ 2)

โดยที่เรากำหนดให้  $\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$  จากนั้น เมื่อนำเทอมกำลังสองข้างต้นไปแทนใน  $f_{x|y}(x|y)$  และการใช้  $\det \Sigma = \det S \cdot \det \Sigma_{yy}$  เราจึงได้ว่า

$$f_{x|y}(x|y) = \frac{1}{(2\pi)^n(\det S)^{n/2}} \exp(x - \mu_{x|y})^T S^{-1}(x - \mu_{x|y})$$

$f_{x|y}$  เป็นฟังก์ชันแจกแจงแบบเกาส์เซียน  $\mathcal{N}(\mu_{x|y}, S)$  ■

**ประพจน์ C.5** (เกาส์เซียนมาตรฐาน). กำหนดให้  $X \sim \mathcal{N}(\mu, \Sigma)$  ที่มี  $\Sigma \succ 0$  ถ้า  $A$  เป็นตัวประกอบกำลังสองของ  $\Sigma$  กล่าวคือ  $AA^T = \Sigma$  เวกเตอร์  $X$  แปลงให้เป็นเกาส์เซียนมาตรฐานได้ภายใต้การแปลง  $Z = A^{-1}(X - \mu) \sim \mathcal{N}(0, I)$

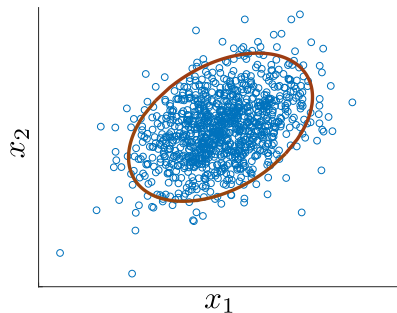
**ประพจน์ C.6** (กำลังสองของเกาส์เซียน). ถ้าหาก  $X \sim \mathcal{N}(0, I_n)$  เราจะได้ว่า  $X^T X \sim \chi^2(n)$  นอกจากนี้ เมื่อกำหนดให้  $A$  เป็นเมทริกซ์สมมาตรและนิพจน์ (idempotent) ( $A^2 = A$ ) ที่มี  $m = \text{tr}(A)$  แล้ว จะได้ว่า  $X^T A X \sim \chi^2(m)$

**พิสูจน์.** เนื่องจาก  $A$  มีความสมมาตร จึงแยกตัวประกอบได้เป็น  $A = UDU^T$  อันมี  $U$  เป็นเมทริกซ์ตั้งฉาก และ  $D$  เป็นเมทริกซ์ทแยงมุมที่มีค่าเป็น 1 และ 0 เท่านั้น (จากสมบัตินิพจน์) หากกำหนดให้  $Y = U^T X$  จึงสรุปจากประพจน์ C.1 ได้ว่า  $Y \sim \mathcal{N}(0, I)$  เราจึงได้ผลลัพธ์ว่า  $X^T A X = X^T UDU^T X = Y^T D Y = \sum_{i=1}^n d_{ii} Y_i^2$  เนื่องจากค่าของ  $d_{ii}$  ที่เท่ากับ 1 มีจำนวนเป็น  $m$  เทอม เราจึงมอง  $X^T A X$  ว่าเป็นผลรวม  $m$  เทอมของกำลังสองของเกาส์เซียน อันมีการแจกแจงเป็น  $\chi^2(m)$  ■

**ประพจน์ C.7** (วงรีเชื่อมั่น). กำหนดให้  $X = (X_1, \dots, X_n) \sim \mathcal{N}(0, P)$  และให้  $x \in \mathbf{R}^n$  คือค่าของเวกเตอร์สุ่ม  $X$  ความน่าจะเป็นที่ตัวอย่างของ  $X$  จะอยู่ในวงรี  $\mathcal{E}_{P,\alpha}$  (ดูนิยาม B.5) นั้นหาได้จาก

$$P(X \in \mathcal{E}_{P,\alpha}) = P(x^T P^{-1} x \leq \alpha) = P(\chi^2(n) \leq \alpha)$$

หากจะการันตีได้ว่าตัวอย่างของ  $X$  จะอยู่ในวงรีดังกล่าวด้วยความน่าจะเป็นมากกว่า  $c$  เราต้องเลือก  $\alpha \geq F^{-1}(c)$  โดยที่  $F$  คือฟังก์ชันแจกแจงสะสมของ  $\chi^2(n)$



รูป C.1: วงรีเชื่อมั่นของเวกเตอร์เกาส์เซียน

**พิสูจน์.** ถ้า  $P$  แยกตัวประกอบค่าเฉพาะเป็น  $P = UDU^T$  และกำหนดให้  $Z = D^{-1/2}U^T X$  เมื่อใช้ประพจน์ C.1 จะพบว่า  $Z \sim \mathcal{N}(0, I)$  ดังนั้น  $\sum_{k=1}^n z_k^2$  จึงเป็นตัวแปรไคกำลังสอง (chi-squared)

$$P(x^T P^{-1} x \leq \alpha) = P(z^T z \leq \alpha) = P(\chi^2(n) \leq \alpha) \geq c \iff \alpha \geq F^{-1}(c)$$

■

---

## บรรณานุกรม

---

- [ABVN11] A. A. Bolstad, B. Van Veen, and R. Nowak. Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing*, 59(6):2628–2641, 2011.
- [AM79] B.D.O. Anderson and J.B. Moore. *Optimal filtering*. Dover, 1979.
- [BA02] K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, second edition, 2002.
- [Bar12] D. Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- [BBP18] C.O. Becker, D.S. Bassett, and V.M. Preciado. Large-scale dynamic modeling of task-fMRI signals via subspace system identification. *Journal of neural engineering*, 15(6):066016, 2018.
- [BJ16] G.E.P. Box and G. Jenkins. *Time Series Analysis, Forecasting and Control*. Wiley, 5 edition, 2016.
- [Bre17] M. Breakspear. Dynamic models of large-scale brain activity. *Nature neuroscience*, 20(3):340–352, 2017.
- [BS14a] L. Barnett and A. K. Seth. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223:50 – 68, 2014.
- [BS14b] L. Barnett and A.K. Seth. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of neuroscience methods*, 223:50–68, 2014.
- [BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. Available: [www.stanford.edu/~boyd/cvxbook](http://www.stanford.edu/~boyd/cvxbook).
- [CGR18] S. Cekić, D. Grandjean, and O. Renaud. Time, frequency, and time-varying Granger-causality measures in neuroscience. *Statistics in medicine*, 37(11):1910–1931, 2018.
- [CM09] P.S.P. Cowpertwait and A.V. Metcalfe. *Introductory time series with R*. Springer Science & Business Media, 2009.

- [CNLK08] N. Chuersuwat, S. Nimrat, S. Lekphet, and T. Kerdkumrai. Levels and major sources of PM<sub>2.5</sub> and PM<sub>10</sub> in Bangkok metropolitan region. *Environment international*, 34(5):671–677, 2008.
- [CT05] A.C. Cameron and P.V. Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.
- [CVX12] CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0. <http://cvxr.com/cvx>, August 2012.
- [Dah00] R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51(2):157–172, 2000.
- [Das08] A. DasGupta. *Asymptotic theory of statistics and probability*. Springer Science & Business Media, 2008.
- [dBG08] A. d’Aspremont, O. Banerjee, and L.E. Ghaoui. First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications*, 30(1):56–66, 2008.
- [DH12] G. Deshpande and X. Hu. Investigating effective brain connectivity from fMRI data: past findings and current issues with reference to Granger causality analysis. *Brain connectivity*, 2(5):235–245, 2012.
- [DS12] M.H. Degroot and M. J. Schervish. *Probability and statistics*. Pearson, 4 edition, 2012.
- [dSFk+16] F. Van de Steen, L. Faes, E. Karahan, J. Songsiri, P.A. Valdes-Sosa, and D. Marinazzo. Critical comments on EEG sensor space dynamical connectivity analysis. *Brain Topography*, pages 1–12, 2016.
- [dV00] A.W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- [Faz02] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.
- [FHP03] K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.
- [FHT08] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- [FKBR14] K.J. Friston, J. Kahan, B. Biswal, and A. Razi. A DCM for resting state fMRI. *Neuroimage*, 94:396–407, 2014.
- [GB08] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. [http://stanford.edu/~boyd/graph\\_dcp.html](http://stanford.edu/~boyd/graph_dcp.html).
- [GL97] L.E. Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on matrix analysis and applications*, 18(4):1035–1064, 1997.

- [GMP<sup>+</sup>09] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [Gra69] C.W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- [Gre00] W.H. Greene. *Econometric Analysis*. Pearson, 2000.
- [Hau12] S. Haufe. *Towards EEG Source Connectivity Analysis*. PhD thesis, Technische Universität Berlin, Germany, 2012.
- [HD12] E.J. Hannan and M. Deistler. *The statistical theory of linear systems*. SIAM, 2012.
- [He13] B. He. *Neural engineering*. Springer, 2013.
- [HJ13] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.
- [HLC<sup>+</sup>15] A. Hartmann, J. Lemos, R.S. Costa, J. Xavier, and S. Vinga. Identification of switched ARX models via convex optimization and expectation maximization. *Journal of Process Control*, 28:9–16, 2015.
- [HLS<sup>+</sup>10] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, E. Reiman, and Alzheimer’s Disease Neuroimaging Initiative. Learning brain connectivity of alzheimer’s disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.
- [HNK08] S. Haufe, G. Nolte, and N. Kräemer. Sparse causal discovery in multivariate time series. *Proceedings of JMLR Workshop and Conference*, 6:97–106, 2008.
- [HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd edition, 2009.
- [HTW15] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC, 2015.
- [IM10] R. Isermann and M. Münchhof. *Identification of dynamic systems: an introduction with applications*. Springer, 2010.
- [JS06] K.G. Jöreskog and D. Sörbom. LISREL 8.80. *Lincolnwood, IL: Scientific Software International Inc*, 2006.
- [JWHT13] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning: with Applications in R*. Springer, 2013.
- [Kai80] T. Kailath. *Linear systems*, volume 156. Prentice-Hall Englewood Cliffs, NJ, 1980.
- [Kat06] T. Katayama. *Subspace methods for system identification*. Springer Science & Business Media, 2006.
- [Kay93] Steven M. Kay. *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.



- [Kee11] K.J. Keesman. *System identification: an introduction*. Springer, 2011.
- [KK08] S. Konishi and G. Kitagawa. *Information criteria and statistical modeling*. Springer Science and Business Media, 2008.
- [KMT11] H. Kobayashi, B.L. Mark, and W. Turin. *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press, 2011.
- [LALR09] A.C. Lozano, N. Abe, Y. Liu, and S. Rosset. Grouped graphical Granger modeling for gene expression regulatory networks discovery. *Bioinformatics*, 25:110–118, 2009.
- [LG09] A. Leon-Garcia. *Probability, statistics, and random processes for electrical engineering*. Pearson Education, 3 edition, 2009.
- [LHV13] Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62(8):605–612, 2013.
- [Lju99] L. Ljung. *System identification: Theory for the User*. Springer, 1999.
- [LSDA15] R.B. Leandro, A.S.e. Silva, I.C. Decker, and M.N. Agostini. Identification of the oscillation modes of a large power system using ambient data. *Journal of Control, Automation and Electrical Systems*, 26(4):441–453, 2015.
- [Lüt05] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, 2005.
- [LZP+15] H. Liu, L. Zhu, Z. Pan, F. Bai, Y. Liu, Y. Liu, M. Patel, E. Farantatos, and N. Bhatt. ARMAX-based transfer function model identification using wide-area measurement for adaptive and coordinated damping control. *IEEE Transactions on Smart Grid*, 8(3):1105–1115, 2015.
- [MB06] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34(3):1436–1462, 2006.
- [MB19] C.M. Michel and D. Brunet. EEG source imaging: a practical review of the analysis steps. *Frontiers in neurology*, 10:325, 2019.
- [MNS21] P. Manomaisaowapak, A. Nartkulpat, and J. Songsiri. Granger causality inference in EEG source connectivity analysis: A state-space approach. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [MPS08] D. Marinazzo, M. Pellicoro, and S. Stramaglia. Kernel method for nonlinear Granger causality. *Physical review letters*, 100(14):144103, 2008.
- [MS22] P. Manomaisaowapak and J. Songsiri. Joint learning of multiple Granger causal networks via non-convex regularizations: Inference of group-level brain connectivity. *Neural Networks*, 2022.
- [Mul09] S.A. Mulaik. *Linear causal modeling with structural equations*. CRC Press, 2009.

- [MZS16] P. Mattsson, D. Zachariah, and P. Stoica. Recursive identification method for piecewise ARX models: A sparse estimation approach. *IEEE Transactions on Signal Processing*, 64(19):5082–5093, 2016.
- [Nor09] J.P. Norton. *An Introduction to Identification*. Dover, 2009.
- [OL13] H. Ohlsson and L. Ljung. Identification of switched linear regression models using sum-of-norms regularization. *Automatica*, 49(4):1045–1050, 2013.
- [OLB10] H. Ohlsson, L. Ljung, and S. Boyd. Segmentation of ARX-models using sum-of-norms regularization. *Automatica*, 46(6):1107–1111, 2010.
- [OM12] P. Van Overschee and B. De Moor. *Subspace identification for linear systems: Theory–Implementation–Applications*. Springer Science & Business Media, 2012.
- [PB14] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.
- [PCC+22] G. Pillonetter, T. Chen, A. Chiuso, G.D. Nicolao, and L. Ljung. *Regularized System Identification: Learning Dynamic Models from Data*. Springer, 2022.
- [PLS13] A. Pongrattarakul, P. Lerdkultanon, and J. Songsiri. Sparse system identification for discovering brain connectivity from fMRI time series. In *Proceedings of SICE Annual Conference*, pages 949–954, 2013.
- [PMN11] R.A. Poldrack, J.A. Mumford, and T.E. Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.
- [PS16] A. Pruttiakaravanich and J. Songsiri. A review on exploring brain networks from fMRI data. *Engineering Journal*, 20(3):1–28, 2016.
- [PS20] A. Pruttiakaravanich and J. Songsiri. Convex formulation for regularized estimation of structural equation models. *Signal Processing*, 166:107237, 2020.
- [QME21] M. Qandil, O. Mohamed, and W.A. Elhajja. A comparative study of wind turbine-generator modeling techniques: Physical modeling, subspace identification, and dynamic neural networks. *Wind Engineering*, 2021.
- [RFG11] A. Roebroeck, E. Formisano, and R. Goebel. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage*, 58(2):296–302, 2011.
- [RSZ+17] A. Razi, M.L. Seghier, Y. Zhou, P. McColgan, P. Zeidman, H. Park, O. Sporns, G. Rees, and K.J. Friston. Large-scale DCMs for resting-state fMRI. *Network Neuroscience*, 1(3):222–241, 2017.
- [SBB15] A.K. Seth, A.B. Barrett, and L. Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.
- [SDV09] J. Songsiri, J. Dahl, and L. Vandenberghe. Graphical models of autoregressive processes. In Y. Eldar and D. Palomar, editors, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2009.

- [SF21] A. Shojaie and E.B. Fox. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9, 2021.
- [SL04] R.E. Schumacker and R.G. Lomax. *A Beginner's Guide to Structural Equation Modeling*. Psychology Press, 2004.
- [SM12] K. Scheinberg and S. Ma. Optimization methods for sparse inverse covariance selection. In *Optimization for machine learning*, chapter 17, pages 455–477. MIT Press, 2012.
- [SMED11] D.E. Seborg, D.A. Mellichamp, T.F. Edgar, and F.J. Doyle. *Process dynamics and control*. John Wiley and Sons, 2011.
- [Son13a] J. Songsiri. Sparse autoregressive model estimation for learning Granger causality in time series. In *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3198–3202, 2013.
- [Son13b] J. Songsiri. Sparse optimization problems in system identification. *Engineering Journal (Thai)*, 5(1):51–75, 2013.
- [SS89] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, London, 1989.
- [SS10] R.H. Shumway and D.S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.
- [SV19] B. Siquin and M. Verhaegen. K4SID: Large-scale subspace identification with kronecker modeling. *IEEE Transactions on Automatic Control*, 64(3):960–975, 2019.
- [SZLW14] K. Sukcharoen, T. Zohrabyan, D. Leatham, and W. Wu. Interdependence of oil prices and stock market indices: A copula approach. *Energy Economics*, 44:331–339, 2014.
- [Tak85] A. Takeshi. *Advanced econometrics*. Harvard university press, 1985.
- [TMLP<sup>+</sup>02] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.
- [Ver19] M. Verhaegen. The identification of network connected systems. In J. Bailieul and T. Samad, editors, *Encyclopedia of systems and control*, volume 320. Springer London, UK, 2019.
- [VH16] M. Verhaegen and A. Hansson. N2SID: nuclear norm subspace identification of innovation models. *Automatica*, 72:57–63, 2016.
- [VV07] M. Verhaegen and V. Verdult. *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.
- [WBK20] J. Watt, R. Borhani, and A.K. Katsaggelos. *Machine learning refined: foundations, algorithms, and applications*. Cambridge University Press, 2020.
- [YCV19] C. Yu, J. Chen, and M. Verhaegen. Subspace identification of individual systems in a large-scale heterogeneous network. *Automatica*, 109, 2019.

- [YL07] M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19, 2007.
- [You12] P. C. Young. *Recursive estimation and time-series analysis: an introduction*. Springer Science & Business Media, 2012.
- [YV18] C. Yu and M. Verhaegen. Subspace identification of individual systems operating in a network (SI2 ON). *IEEE Transactions on Automatic Control*, 63(4):1120–1125, 2018.
- [YZT+15] Y. Yang, F. Zhang, K. Tao, B. Sanchez, H. Wen, and Z. Teng. An improved crest factor minimization algorithm to synthesize multisines with arbitrary spectrum. *Physiological measurement*, 36(5):895, 2015.
- [ZD98] K. Zhou and J.C. Doyle. *Essentials of robust control*, volume 104. Prentice hall Upper Saddle River, NJ, 1998.



# อภิธานศัพท์

ตาราง 1: ศัพท์บัญญัติภาษาไทยที่ใช้ในหนังสือ

ภาษาอังกฤษ	ภาษาไทย	ที่มา
a posteriori	หลังประสบการณ์	วิทยาศาสตร์ (พ.ศ. 2547)
additive noise	สัญญาณรบกวนเชิงบวก	
affine	สัมพรรค	คณิตศาสตร์ (พ.ศ. 2559)
almost everywhere	เกือบทุกแห่ง	คณิตศาสตร์ (พ.ศ. 2559)
artifact	สิ่งแปลกปน	วิทยาศาสตร์ (พ.ศ. 2546)
asymptotic	เชิงเส้นกำกับ	สถิติศาสตร์ (พ.ศ.2561)
asymptotic distribution	การแจกแจงเชิงเส้นกำกับ	สถิติศาสตร์ (พ.ศ.2561)
autocorrelation	สหสัมพันธ์ในตัว	สถิติศาสตร์ (พ.ศ.2561)
autocorrelation function	ฟังก์ชันสหสัมพันธ์ในตัว	สถิติศาสตร์ (พ.ศ.2561)
autocovariance	ความแปรปรวนร่วมในตัว	สถิติศาสตร์ (พ.ศ.2561)
autoregressive	ถดถอยในตัว	สถิติศาสตร์ (พ.ศ.2561)
autoregressive integrated moving average model	แบบจำลองรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่	สถิติศาสตร์ (พ.ศ.2561)
autoregressive moving average process	กระบวนการเฉลี่ยเคลื่อนที่ถดถอยในตัว	สถิติศาสตร์ (พ.ศ.2561)
autoregressive process	กระบวนการถดถอยในตัว	สถิติศาสตร์ (พ.ศ.2561)
autoregressive with exogeneous input process	กระบวนการถดถอยในตัวที่มีสัญญาณนอกระบบ	
bandwidth	แบนด์วิธ, ความกว้างแถบความถี่	คอมพิวเตอร์และเทคโนโลยีสารสนเทศ (พ.ศ.2549)
Bayes estimator	ตัวประมาณของเบส์	สถิติศาสตร์ (พ.ศ.2561)
Bernoulli	แบร์นูลลี	สถิติศาสตร์ (พ.ศ.2561)
biased estimator	ตัวประมาณเอนเอียง	สถิติศาสตร์ (พ.ศ.2561)
categorical variable	ตัวแปรจำแนกประเภท	สถิติศาสตร์ (พ.ศ.2561)
Cauchy-Schwarz inequality	อสมการโคชี-ชวาร์ซ	คณิตศาสตร์ (พ.ศ. 2547)
causal system	ระบบเชิงเหตุ	
companion matrix	เมทริกซ์คู่เคียง	คณิตศาสตร์ (พ.ศ. 2559)
conditional expectation	ค่าคาดหวังมีเงื่อนไข	สถิติศาสตร์ (พ.ศ.2561)

ตาราง 1 – ศัพท์บัญญัติภาษาไทยที่ใช้ในหนังสือ

ภาษาอังกฤษ	ภาษาไทย	ที่มา
conditional probability	ความน่าจะเป็นมีเงื่อนไข	สถิติศาสตร์ (พ.ศ.2561)
consistent equations	สมการต้องกัน	คณิตศาสตร์ (พ.ศ. 2559)
consistent estimator	ตัวประมาณคงเส้นคงวา	สถิติศาสตร์ (พ.ศ.2561)
continuous mapping theorem	ทฤษฎีบทการส่งต่อเนื่อง	
continuous-time system	ระบบเวลาต่อเนื่อง	
convergence in distribution	การลู่เข้าเชิงการแจกแจง	สถิติศาสตร์ (พ.ศ.2561)
convergence in probability	การลู่เข้าเชิงความน่าจะเป็น	สถิติศาสตร์ (พ.ศ.2561)
convex relaxation	การผ่อนคลายแบบคอนเวกซ์	คณิตศาสตร์ (พ.ศ. 2559)
covariance	ความแปรปรวนร่วม	คณิตศาสตร์ (พ.ศ. 2559)
covariate	ตัวแปรร่วม	สถิติศาสตร์ (พ.ศ.2561)
cross correlation	สหสัมพันธ์ไขว้	สถิติศาสตร์ (พ.ศ.2561)
cross covariance	ความแปรปรวนร่วมไขว้	สถิติศาสตร์ (พ.ศ.2561)
cumulative density function	ฟังก์ชันความหนาแน่นสะสม	สถิติศาสตร์ (พ.ศ.2561)
cumulative distribution function	ฟังก์ชันการแจกแจงสะสม	สถิติศาสตร์ (พ.ศ.2561)
cumulative probability function	ฟังก์ชันความน่าจะเป็นสะสม	สถิติศาสตร์ (พ.ศ.2561)
curve fitting	การปรับเส้นโค้ง	คณิตศาสตร์ (พ.ศ. 2559)
delay	ประวิง	คอมพิวเตอร์และ เทคโนโลยีสารสนเทศ (พ.ศ.2549)
delay operator	ตัวดำเนินการหน่วง	คอมพิวเตอร์และ เทคโนโลยีสารสนเทศ (พ.ศ.2549)
density function	ฟังก์ชันความหนาแน่น	คณิตศาสตร์ (พ.ศ. 2559)
deterministic	เชิงกำหนด	คณิตศาสตร์ (พ.ศ. 2559)
discrete-time system	ระบบเวลายืด	วิทยาศาสตร์ (พ.ศ. 2546)
effective number of parameters	จำนวนยังผลของพารามิเตอร์	
efficient estimator	ตัวประมาณประสิทธิภาพ	สถิติศาสตร์ (พ.ศ.2561)
error	ค่าคลาดเคลื่อน, ความคลาด เคลื่อน, ค่าผิดพลาด	คณิตศาสตร์ (พ.ศ. 2559), สถิติศาสตร์ (พ.ศ.2561), เทคโนโลยีสารสนเทศ (พ.ศ. 2545)
estimator	ตัวประมาณ	สถิติศาสตร์ (พ.ศ.2561)
exogeneous variable	ตัวแปรนอกระบบ	สถิติศาสตร์ (พ.ศ.2561)
expectation	ค่าคาดหวัง	คณิตศาสตร์ (พ.ศ. 2559)
Fisher information matrix	เมทริกซ์สารสนเทศของฟิชเชอร์	สถิติศาสตร์ (พ.ศ.2561)
fitting	ความเข้ากัน	
Frobenius	โฟรเบนียุส	
global minimum	ค่าต่ำสุดวงกว้าง	คณิตศาสตร์ (พ.ศ. 2559)
Hankel matrix	เมทริกซ์ฮังเคิล	คณิตศาสตร์ (พ.ศ. 2559)
Hermitian matrix	เมทริกซ์เฮอร์มิเทียน	คณิตศาสตร์ (พ.ศ. 2559)

มีต่อหน้าถัดไป

ตาราง 1 – ศัพท์บัญญัติภาษาไทยที่ใช้ในหนังสือ

ภาษาอังกฤษ	ภาษาไทย	ที่มา
heterogeneous	วิวิธพันธุ์	วิทยาศาสตร์ (พ.ศ. 2546)
heteroscedastic	แปรปรวนต่าง	สถิติศาสตร์ (พ.ศ.2561)
heteroscedasticity	ภาวะแปรปรวนต่าง	สถิติศาสตร์ (พ.ศ.2561)
homogeneous	เอกพันธุ์	วิทยาศาสตร์ (พ.ศ. 2546)
homoscedastic	แปรปรวนเท่ากัน	สถิติศาสตร์ (พ.ศ.2561)
homoscedasticity	ภาวะแปรปรวนเท่ากัน	สถิติศาสตร์ (พ.ศ.2561)
integrated process	กระบวนการรวม	สถิติศาสตร์ (พ.ศ.2561)
Kronecker	โครเนกเกอร์	
Kullback-Leibler divergence	การลู่ออกของคัลล์แบก-ลีบเลอร์	
k-step predictor	ตัวทำนาย k ก้าวล่วงหน้า	
lag	ช่วงเวลาที่ยากกว่ากัน	สถิติศาสตร์ (พ.ศ.2561)
lag operator	ตัวดำเนินการล่าหลัง	
lasso	แลสโซ	
least-squares	กำลังสองน้อยสุด	คณิตศาสตร์ (พ.ศ. 2559)
left inverse	เมทริกซ์ผกผันซ้าย	
Lyapunov	เลียปูนอฟ	
Leibnitz	ไลบ์นิทซ์	คณิตศาสตร์ (พ.ศ. 2559)
likelihood function	ฟังก์ชันภาวะน่าจะเป็น	สถิติศาสตร์ (พ.ศ.2561)
limiting distribution	การแจกแจงขีดจำกัด	สถิติศาสตร์ (พ.ศ.2561)
loglikelihood function	ฟังก์ชันล็อกภาวะน่าจะเป็น	สถิติศาสตร์ (พ.ศ.2561)
machine learning	การเรียนรู้ของเครื่อง	
marginal distribution	การแจกแจงตามขอบ	สถิติศาสตร์ (พ.ศ.2561)
Markov	มาร์คอฟ	คณิตศาสตร์ (พ.ศ. 2559)
matrix	เมทริกซ์	คณิตศาสตร์ (พ.ศ. 2559)
maximum a posteriori estimation	การประมาณภายหลังสูงสุด	
maximum likelihood estimation	การประมาณภาวะน่าจะเป็นสูงสุด	สถิติศาสตร์ (พ.ศ.2561)
maximum likelihood estimator	ตัวประมาณภาวะน่าจะเป็นสูงสุด	สถิติศาสตร์ (พ.ศ.2561)
moving average	ค่าเฉลี่ยเคลื่อนที่	สถิติศาสตร์ (พ.ศ.2561)
moving average process	กระบวนการเฉลี่ยเคลื่อนที่	สถิติศาสตร์ (พ.ศ.2561)
multicollinearity	ความสัมพันธ์เชิงเส้นพหุ	สถิติศาสตร์ (พ.ศ.2561)
objective function	ฟังก์ชันจุดประสงค์	คณิตศาสตร์ (พ.ศ. 2559)
odds	ออดด์	สถิติศาสตร์ (พ.ศ.2561)
optimality condition	เงื่อนไขเหมาะสมที่สุด	คณิตศาสตร์ (พ.ศ. 2559)
optimization	การหาค่าเหมาะสมที่สุด	คณิตศาสตร์ (พ.ศ. 2559)
orthogonal complement	ส่วนเติมเต็มตั้งฉาก	คณิตศาสตร์ (พ.ศ. 2559)
orthogonal matrix	เมทริกซ์เชิงตั้งฉาก	คณิตศาสตร์ (พ.ศ. 2559)
orthonormal matrix	เมทริกซ์เชิงตั้งฉากปรกติ	คณิตศาสตร์ (พ.ศ. 2559)
orthogonal projection	ภาพฉายเชิงตั้งฉาก	คณิตศาสตร์ (พ.ศ. 2559)
outlier	ค่านอกเกณฑ์	สถิติศาสตร์ (พ.ศ.2561)

มีต่อหน้าถัดไป



ตาราง 1 – ศัพท์บัญญัติภาษาไทยที่ใช้ในหนังสือ

ภาษาอังกฤษ	ภาษาไทย	ที่มา
output error model	แบบจำลองค่าคลาดเคลื่อนที่ สัญญาณออก	
over-determined system	ระบบกำหนดเกิน	
path matrix	เมทริกซ์วิถี	
periodic	เป็นคาบ	วิทยาศาสตร์ (พ.ศ. 2546)
persistently exciting of order	ระดับกระตุ้นคงอยู่	
perturbation	เพอร์เทอร์เบชัน	คณิตศาสตร์ (พ.ศ. 2559)
point estimator	ตัวประมาณแบบจุด	สถิติศาสตร์ (พ.ศ.2561)
positive definite matrix	เมทริกซ์บวกแน่นอน	คณิตศาสตร์ (พ.ศ. 2559)
posterior distribution	การแจกแจงภายหลัง	สถิติศาสตร์ (พ.ศ.2561)
posterior odds	อัตราส่วนภายหลัง	สถิติศาสตร์ (พ.ศ.2561)
power spectral density	ความหนาแน่นเชิงสเปกตรัม กำลัง	
predictor variable	ตัวแปรทำนาย	สถิติศาสตร์ (พ.ศ.2561)
prior density function	ฟังก์ชันความหนาแน่นก่อน	
prior probability	ความน่าจะเป็นก่อน	สถิติศาสตร์ (พ.ศ.2561)
probability density function	ฟังก์ชันความหนาแน่นความน่า จะเป็น	สถิติศาสตร์ (พ.ศ.2561)
probability distribution	การแจกแจงความน่าจะเป็น	สถิติศาสตร์ (พ.ศ.2561)
probability distribution func- tion	ฟังก์ชันการแจกแจงความน่า จะเป็น	สถิติศาสตร์ (พ.ศ.2561)
pseudo random binary se- quence	ลำดับฐานสองสุ่มเทียม	
quadratic programming	กำหนดการกำลังสอง	คณิตศาสตร์ (พ.ศ. 2559)
quasi-stationary process	กระบวนการกึ่งคงที่	
rank	ค่าลำดับชั้น	คณิตศาสตร์ (พ.ศ. 2559)
rank-nullity theorem	ทฤษฎีบทค่าลำดับชั้น-ศูนย์ภาพ	คณิตศาสตร์ (พ.ศ. 2559)
recursive equation	สมการเวียนเกิด	คณิตศาสตร์ (พ.ศ. 2559)
regression	การถดถอย	สถิติศาสตร์ (พ.ศ.2561)
regressor	ตัวแปรถดถอย	สถิติศาสตร์ (พ.ศ.2561)
regularity conditions	เงื่อนไขตามเกณฑ์	สถิติศาสตร์ (พ.ศ.2561)
regularization	การทำให้ตามเกณฑ์	
regularized solution	ผลเฉลยตามเกณฑ์	
relaxation	ผ่อนคลาย, ผ่อนปรน	คณิตศาสตร์ (พ.ศ. 2559)
resampling	การเลือกตัวอย่างซ้ำ	สถิติศาสตร์ (พ.ศ.2561)
residual	ส่วนเหลือ, ตกค้าง	วิทยาศาสตร์ (พ.ศ. 2546), สถิติศาสตร์ (พ.ศ.2561)
residual error	ความคลาดเคลื่อนส่วนเหลือ	สถิติศาสตร์ (พ.ศ.2561)
response variable	ตัวแปรตอบสนอง	สถิติศาสตร์ (พ.ศ.2561)
ridge estimator	ตัวประมาณแบบริดจ์	สถิติศาสตร์ (พ.ศ.2561)
ridge regression	การถดถอยแบบริดจ์	สถิติศาสตร์ (พ.ศ.2561)
robust estimator	ตัวประมาณแกร่ง	สถิติศาสตร์ (พ.ศ.2561)

มีต่อหน้าถัดไป

ตาราง 1 – ศัพท์บัญญัติภาษาไทยที่ใช้ในหนังสือ

ภาษาอังกฤษ	ภาษาไทย	ที่มา
robust least-squares	กำลังสองน้อยสุดแบบแกร่ง	
robust regression	การถดถอยแกร่ง	
root mean square	รากกำลังสองเฉลี่ย	คณิตศาสตร์ (พ.ศ. 2559)
root mean square error	รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย	สถิติศาสตร์ (พ.ศ.2561)
sampling	การชักตัวอย่าง	สถิติศาสตร์ (พ.ศ.2561)
sparse	เบาบาง	
standard error	ความคลาดเคลื่อนมาตรฐาน	สถิติศาสตร์ (พ.ศ.2561)
standard error of the regression	ความคลาดเคลื่อนมาตรฐานของการถดถอย	สถิติศาสตร์ (พ.ศ.2561)
standardized coefficient	สัมประสิทธิ์ปรับเข้ามาตรฐาน	
state-space model	แบบจำลองปริภูมิสถานะ	
stationary process	กระบวนการคงที่	สถิติศาสตร์ (พ.ศ.2561)
stationary time series	อนุกรมเวลาแบบคงที่	สถิติศาสตร์ (พ.ศ.2561)
stepwise regression	การถดถอยทีละขั้น	สถิติศาสตร์ (พ.ศ.2561)
stochastic	สโตแคสติก	คณิตศาสตร์ (พ.ศ. 2559)
stratified model	แบบจำลองแบ่งชั้นภูมิ	สถิติศาสตร์ (พ.ศ.2561)
strict-sense stationary	กระบวนการคงที่โดยแท้	
strictly	โดยแท้	คณิตศาสตร์ (พ.ศ. 2559)
thresholded solution	ผลเฉลยโดนขีดแบ่ง	
Toeplitz matrix	เมทริกซ์เทอพลิตซ์	คณิตศาสตร์ (พ.ศ. 2559)
unbiased estimator	ตัวประมาณไม่เอนเอียง	สถิติศาสตร์ (พ.ศ.2561)
under-determined system	ระบบกำหนดขาด	
underparametrization	การแจกแจงพารามิเตอร์ขาด	
uniform stability	เสถียรภาพแบบเอกรูป	
uniformly stable	เสถียรอย่างเอกรูป	
variance	ความแปรปรวน	สถิติศาสตร์ (พ.ศ.2561)
vector	เวกเตอร์	คณิตศาสตร์ (พ.ศ. 2559)
weak law of large numbers	กฎจำนวนมากอย่างอ่อน	สถิติศาสตร์ (พ.ศ.2561)
wide-sense stationary	กระบวนการคงที่เชิงกว้าง	

สิ้นสุด



---

# ดัชนี

---

- $k$ -fold cross validation, 225
- $q$ -norm, 154
- Adjusted  $R^2$ , 218
- AIC, 215
- AICc, 216
- Akaike information criterion, 215
- Akaike's final prediction error, 217
- alternative hypothesis, 88
- asymptotic covariance, 66
- asymptotic distribution, 66, 199
- autoregressive, 25
  - integrated moving average, 31
  - moving average, 25
  - moving average with exogenous input, 25
  - with exogenous input, 25
- backward selection, 93
- basis function, 78
- Bayes estimator, 194
- Bayes factor, 217
- Bayes' rule, 190, 193
- Bayesian information criterion, 216
- best linear unbiased estimator, 184
- best subset selection, 92
- BIC, 216
- categorical, 94, 252
- Cauchy–Schwarz inequality, 114
- causality, 239
- composite of norms, 108, 121, 154
- confidence interval, 89
- consistent estimator, 89
- converge
  - almost sure, 274
  - in  $r$ -th mean, 274
  - in distribution, 273
  - in probability, 273
  - with probability 1, 274
- convex, 69, 102–104
- convex relaxation, 155, 247
- correlation function, 16
- Crest factor, 57
- cross validation, 224
- cross-validated error, 259
- curve fitting, 8, 119
- CVA, 147
- cyclic coordinate descent, 108
- data equation, 138
- data fitting, 78
- data generating process, 85
- delay operator, 15
- deterministic, 112
- deterministic model, 15
- difference operator, 111
- dummy variable, 94
- effective number of parameters, 215, 218
- elastic net, 109
- estimated asymptotic covariance, 89
- estimation
  - maximum a posteriori, 105
  - maximum likelihood, 185
- estimator
  - best linear unbiased, 184
  - consistent, 64, 89
  - efficient, 65
  - extremum, 68
  - minimum variance unbiased, 184, 200
  - unbiased, 64, 86

- exact likelihood function, 193
- explained sum of squares, 84
- feature selection, 92
- finite impulse response, 25
- flat spectrum, 42
- forward selection, 93
- FPE, 217
- Frobenius norm, 97, 139
- full model, 92
- Gauss-Markov Theorem, 87
- Granger
  - causality, 239, 242
  - graphical model, 250
- hold out, 224
- homoscedasticity, 86
- homoskedastic, 96
- hypothesis test, 88
- identified model, 33
- impulse response, 15
- inequality
  - Cauchy–Schwarz, 114
  - Chebyshev, 274
  - Cramér–Rao, 199
  - Sylvester rank, 142
- innovation, 20
- instrumental variable, 137, 140
- integrated model, 30
- integrated moving average, 31
- interior-point, 108
- invertible, 26
- Kronecker, 155
- Kullback–Leibler divergence, 215
- lasso
  - fused, 111
  - graphical, 245
  - group, 110
  - overlapped, 110
  - sparse group, 110
- law of iterated expectation, 180
- least mean square error, 184
- least-squares
  - $\ell_1$ -regularized, 107
  - $\ell_2$ -regularized, 105
  - feasible generalized, 88
  - generalized, 88
  - linear, 75
  - robust, 84
  - Tikhonov regularized, 105
  - weighted, 83
- leave-one-out cross validation, 224
- left inverse, 82
- likelihood function, 185
- limiting distribution, 273
- log-likelihood
  - conditional, 191
  - function, 185
  - ratio test, 217
  - score, 197
- Lyapunov equation, 281
- Mallow’s  $C_p$ , 217
- matrix
  - band, 155
  - companion, 28
  - difference, 124
  - extended observability, 136
  - Fisher information, 198
  - orthogonal, 81
  - path, 243, 246
  - Toeplitz, 123
  - idempotent, 278
  - state-transition, 14
- maximum a posteriori estimation, 105
- maximum likelihood estimation, 185
- minimum mean square error, 184
- minimum variance unbiased
  - estimator, 184
- model
  - assessment, 207
  - full, 92
  - generalized linear, 188
  - graphical, 246
  - identifiability, 37
  - integrated, 30
  - nonparametric, 14
  - null, 92
  - output error, 35
  - parametric, 14
  - probabilistic, 15
  - selection, 207
  - state-space, 23
  - state-space in innovation form, 24
  - stratified, 122
- moving average, 25
- multicollinearity, 100
- multiple linear regression, 75

- N4SID, 147
- network
  - heterogeneous, 155
  - homogenous, 155
- non-stationary, 29
- nonnegative constraints, 102
- normal equation, 78
- nuclear norm, 155
- null hypothesis, 88
- null model, 92
- oblique projection, 147
- one-step prediction, 18
- operator
  - delay, 15
  - difference, 111
- optimal predictor, 159, 161
- orthogonal complement, 139
- orthogonal projection, 78, 139
- orthogonality condition, 77
- orthonormal, 142, 269
- outlier, 83, 84
- overfitting, 213
- overparametrized, 168
- parameter identifiable, 68
- partial correlation function, 242
- penalty function, 105
- persistently exciting, 50
- PO-MOEPS, 147
- point estimate, 86
- pointwise supremum, 112
- Poisson regression, 188
- polyhedron, 102, 124
- posterior density function, 193
- power of a test, 88
- power spectral density, 22, 242, 279
- prediction error, 158
- prior density function, 193
- probability limit, 70, 273
- probability simplex, 103
- process reaction curve, 39
- product limit normal rule, 276
- proximal algorithm, 108
- pseudo random binary sequence, 44
- QP, 104, 124, 125, 127
- QR factorization, 269
- quadratic programming, 104
- qualitative variable, 94
- quantitative variable, 94
- recursive equation, 24
- regression coefficient, 75
- regression fit, 90
- regularity conditions, 197
- resampling, 214
- residual error, 76
- residual sum of squares, 84
- Riccati equation, 148
- ridge regression, 105
- robust least-squares, 84
- row space, 139
- RQ factorization, 135, 269
- Sample-size corrected Akaike
  - information, 216
- Schur complement, 268
- second-order cone programming, 113
- sparse estimation, 257
- standard error of the coefficient, 90
- standard error of the regression, 89
- stationary
  - quasi, 279
  - second-order, 279
  - strict-sense, 279
  - wide-sense, 28, 279
- stepwise regression, 92, 93
- strictly convex, 109
- subset selection, 108
- subspace identification, 135
- sum of squares
  - explained, 84
  - residual, 84
  - total, 84
- Sylvester rank inequality, 142
- system identifiable, 67
- test statistic, 88
- Theorem
  - central limit, 276
  - continuous mapping, 275
  - Gauss-Markov, 87
  - Rank-nullity, 84
  - Slutsky, 275
  - Wiener-Khinchin, 280
- time series, 25
- total expectation, 180
- total sum of squares, 84
- total variation reconstruction, 111
- type I error, 88
- type II error, 88
- variable

- dependent, 75
- dummy, 94
- explained, 75
- explanatory, 75
- independent, 75
- instrumental, 137, 140
- qualitative, 94
- quantitative, 94
  
- weak law of large numbers, 274
- weighted least squares, 83
- weighted norm, 116
- white noise, 42
- Wiener-Hopf equation, 182
  
- กฎของลิมิตเชิงความน่าจะเป็น, 275
- กฎของเบส์, 190, 193
- กฎค่าคาดหวังทำซ้ำ, 180
- กฎจำนวนมาก, 166, 171
- กฎจำนวนมากอย่างอ่อน, 71, 140, 274
- กฎลิมิตผลคูณปกติ, 276
- กระบวนการก่อกำเนิดข้อมูล, 85
- การแจกแจง
  - ซีตจำกั๊ด, 273
- กำลังสองน้อยสุด, 87, 101, 195, 248, 253, 257
  - ที่ทำให้ตามเกณฑ์ด้วยนอร์ม-1, 107
  - ที่ทำให้ตามเกณฑ์ด้วยนอร์ม-2, 105
  - ที่ทำให้ตามเกณฑ์แบบทีโฮนอนอฟ, 105
  - น้อยทั่วไป, 88
  - น้อยทั่วไปที่เป็นไปได้, 88
  - เชิงเส้น, 75
  - แบบแกร่ง, 84, 114
  - ไม่เชิงเส้น, 117
- กำหนดการกำลังสอง, 104, 124, 125
- ขั้นตอนวิธีจุดภายใน, 108
- ขั้นตอนวิธีใกล้เคียง, 108
- ข้อมูล
  - ชุดตรวจสอบ, 207
  - ชุดทดสอบ, 207
  - ชุดประมาณ, 207
  - ชุดฝึกสอน, 207
- คงที่
  - กึ่ง, 279
  - อันดับสอง, 279
  - โดยแท้, 279
  - เชิงกว้าง, 279
- คงเส้นคงวา, 163, 196
- ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์, 90
- ความคลาดเคลื่อนส่วนเหลือ, 83
- ความคลาดเคลื่อนส่วนเหลือมาตรฐาน, 84
- ความคลาดเคลื่อนเลวร้ายที่สุด, 112, 114, 116
- ความผิดพลาดมาตรฐานของการถดถอย, 89
- ความผิดพลาดในการทำนาย, 158
- ความสัมพันธ์เชิงเส้นพหุ, 100
- ความหนาแน่นเชิงสเปกตรัมกำลัง, 22, 42, 242
- ความเข้ากันเกิน, 211, 213
- ความเป็นเชิงเส้นร่วม, 109
- คอนเวกซ์, 69, 102–104, 111, 112, 116, 124, 132
- ค่าคลาดเคลื่อนส่วนเหลือ, 83
- ค่าคาดหวังรวม, 180
- ค่านอกเกณฑ์, 83
- ค่าประมาณแบบจุด, 86
- ค่าผิดพลาดไขว้, 259
- ค่าเอกฐาน, 143
- จำนวนยังผลของพารามิเตอร์, 215, 218–220
- จำแนกประเภท, 94, 252
- ฉายภาพเอียง, 147
- ช่วงความเชื่อมั่น, 89
- ซิมเพลกซ์ความน่าจะเป็น, 103
- ฐานสองสุมเทียม, 39
- ตรวจสอบ
  - แบบจำลอง, 207
  - ไขว้, 224, 257
  - ไขว้บน 1 ส่วน, 224
  - ไขว้แบบ  $k$  ส่วน, 225
- ตัวดำเนินการ
  - ผลต่าง, 111
  - หนึ่ง, 15
- ตัวทำนายเหมาะที่สุด, 161
- ตัวประกอบ
  - QR, 269
  - RQ, 269
  - เครสต์, 57
  - เบส์, 217
- ตัวประมาณ
  - คงเส้นคงวา, 64, 89
  - ประสิทธิภาพ, 65, 167, 200
  - สุดขีด, 68, 166, 201
  - เชิงเส้นไม่เอนเอียง, 182, 183
  - ไม่เอนเอียง, 64, 86, 199

- ไม่เอนเอียงเชิงเส้นที่สุด, 87  
 ไม่เอนเอียงแปรปรวนต่ำสุด, 184, 200
- ตัวแปร
- เชิงคุณภาพ, 94
  - เชิงปริมาณ, 94
  - เชิงเครื่องมือ, 137, 140
- ตาข่ายยัดหยุ่น, 109
- ถดถอย
- ทีละขั้น, 92, 93
  - ปัวส์ซอง, 188
  - ลอจิสติก, 201
  - เชิงเส้น, 75, 76, 121
  - แบบบริดจ์, 105, 107, 109, 114, 116, 117, 123, 195, 205, 219, 236
  - แบบแกร่ง, 117, 131
- ทดสอบสมมติฐาน, 88
- ทรงหลายหน้า, 102, 124
- ทฤษฎีบท
- การส่งต่อเนื่อง, 275
  - ค่าลำดับขั้น-ศูนย์กลาง, 76, 84
  - ลิมิตกลาง, 71, 276
  - วินเนอร์-คินซิน, 42, 280
  - สลุตสกี, 275
  - เกาส์-มาร์คอฟ, 87
- ทำนาย 1 ก้าวล่วงหน้า, 18
- นอร์ม- $q$ , 154, 254
- นอร์มถ่วงน้ำหนัก, 116
- นอร์มนิวเคลียร์, 155
- นอร์มโพรเบนิอุส, 97, 113, 132, 139, 253
- ประมาณ
- ของเบส์, 194
  - ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด, 180
  - ภายหลังสูงสุด, 105
  - ภาชนะน่าจะเป็นสูงสุด, 185
  - แบบเบาบาง, 257
- ประเมินแบบจำลอง, 207
- ปรับเส้นโค้ง, 8, 119, 128
- ปริภูมิย่อย, 135, 173
- ปริภูมิแถว, 139
- ปัญหาการเรียงอันดับสอง, 113, 114, 116
- ปัญหาสโตแคสติก, 115
- ปัญหาแบบแกร่ง, 112
- ผลตอบสนองอิมพัลส์, 14, 15, 19
- ผลตอบสนองอิมพัลส์อันดับจำกัด, 25
- ผลบวกกำลังสอง
- ของส่วนเหลือ, 84
  - ทั้งหมด, 84
- ที่ถูกอธิบาย, 84
- ผอนคล้ายแบบคอนเวกซ์, 155, 247
- พารามิเตอร์ลงโทษ, 107, 117
- ฟังก์ชัน
- มูลฐาน, 78
  - สหสัมพันธ์บางส่วน, 242
  - สหสัมพันธ์ในตัว, 54
  - สหสัมพันธ์ไขว้, 54
  - ความหนาแน่นก่อน, 193
  - ความหนาแน่นภายหลัง, 193
  - คอนเวกซ์โดยแท้, 109
  - ประกอบ, 250
  - ประกอบของนอร์ม, 108, 121, 154
  - ภาชนะน่าจะเป็น, 185
  - ภาชนะน่าจะเป็นแม่นยำ, 193
  - ลงโทษ, 105
  - ลือกภาชนะน่าจะเป็น, 185, 245, 246
  - ลือกภาชนะน่าจะเป็นมีเงื่อนไข, 191
  - สหสัมพันธ์ในตัว, 16
- ภาวะความต่าง, 96
- ภาวะแปรปรวนเท่ากัน, 86, 89, 96
- ระดับการกระตุ้นคงอยู่, 50, 51, 53, 55
- ระบบกำหนดเก็น, 76
- ระบุหาพารามิเตอร์ได้, 68
- ลำดับขั้นเต็ม, 76
- ลำดับฐานสองสุ่มเทียม, 44, 45
- ขีดความน่าจะเป็น, 70, 273
- คู่เขา
- ในค่าเฉลี่ยที่  $r$ , 274
  - ด้วยความน่าจะเป็น, 274
  - เกือบแน่นอน, 274
  - เชิงการแจกแจง, 273
  - เชิงความน่าจะเป็น, 273
- ลือกภาชนะน่าจะเป็น
- คะแนน, 197
  - ฟังก์ชัน, 185
  - อัตรา, 217
- วนพิศเคลื่อนลง, 108
- สมการ
- เลียปูนอฟ, 154, 281
  - ข้อมูล, 136, 138
  - ปกติ, 76, 78
  - รีคคาติ, 148
  - วินเนอร์-ฮอฟ, 182
  - เวียนเกิด, 24, 25, 161
- สร้างใหม่แบบแปรผันทั้งหมด, 111
- สัมประสิทธิ์การถดถอย, 75
- ส่วนเติมเต็มซูร์, 133, 184, 210, 268



- ส่วนเติมเต็มตั้งฉาก, 139  
 หาเอกลักษณ์ได้, 67
- อสมการ  
   คราเมอร์-ราว, 196, 199  
   เซฟปีเซฟ, 274  
   ค่าลำดับชั้นของซิลเวสเตอร์, 142  
   โคซี-ชวาร์ซ, 53, 114
- อัตราขยายกาลมาน, 145, 148
- เกรนเจอร์  
   เหตุภาพ, 242, 252  
   แบบจำลองเชิงกราฟ, 250
- เครือข่าย  
   วิวิธพันธุ์, 155  
   เอกพันธุ์, 155
- เงื่อนไข  
   การตั้งฉาก, 77  
   ปกติ, 197, 198  
   เหมาะสมที่สุด, 76
- เชิงกำหนด, 85, 112, 115
- เชิงตั้งฉากปกติ, 142
- เบาบาง, 106–108, 117, 219
- เมทริกซ์  
   คู่เคียง, 28  
   นิจพล, 278  
   ผลต่าง, 124  
   แกรมเมียน, 154  
   แถบ, 155  
   ความหนาแน่นเชิงสเปกตรัมกำลัง, 242  
   ความแปรปรวนร่วมเชิงเส้นกำกับ, 66, 89  
   ถดถอย, 75  
   ถ่วงน้ำหนัก, 83, 143, 148  
   ภาพถ่าย, 77  
   ภาพถ่ายเชิงตั้งฉาก, 78, 87, 139  
   วิถี, 243, 246  
   สภาพสังเกตได้แบบขยาย, 136  
   สารสนเทศของพีชเซอร์, 196, 198  
   เชิงตั้งฉาก, 81  
   เทอพลิทซ์, 123  
   แปลงสถานะ, 14
- เลือกตัวอย่างซ้ำ, 214
- เลือกถอยหลัง, 93
- เลือกเซตย่อย, 108
- เลือกเดินทาง, 93
- เลือกแบบจำลอง, 207, 257
- เหตุภาพของเกรนเจอร์, 239
- แจกแจง  
   เชิงเส้นกำกับ, 66, 163, 196, 199
- แจกแจงเกิน, 168
- แบบจำลอง  
   รวมการถดถอยในตัวกับการเคลื่อนที่  
     เคลื่อนที่, 31  
   เชิงกราฟ, 245, 246  
   เวกเตอร์ถดถอยในตัว, 248  
   ข้อมูลใหม่, 145  
   ความน่าจะเป็น, 15  
   ซ่อนใน, 93, 217  
   ถดถอยปัวส์ซอง, 188  
   ถดถอยเชิงเส้นหลายตัวแปร, 75  
   ถดถอยในตัว, 25  
   ถดถอยในตัวกับการเคลื่อนที่เคลื่อนที่ที่มี  
     สัญญาณเข้านอกระบบ, 25  
   ถดถอยในตัวที่มีสัญญาณนอกระบบ, 25  
   ปริภูมิสถานะในรูปข้อมูลใหม่, 24  
   อิงพารามิเตอร์, 14  
   เคลื่อนที่เคลื่อนที่, 25  
   เคลื่อนที่เคลื่อนที่ถดถอยในตัว, 25  
   เชิงกราฟของเกรนเจอร์, 250  
   เชิงกำหนด, 15  
   เชิงเส้นนัยทั่วไป, 188  
   แน่น, 216  
   แบ่งชั้นภูมิ, 122  
   ไม่อิงพารามิเตอร์, 14
- แลสโซ, 107, 108, 117, 205, 219, 226, 236  
   เชิงกราฟ, 245  
   กลุ่ม, 110, 248  
   กลุ่มเบาบาง, 110  
   คาบเกี่ยว, 110  
   เชื่อม, 111, 255
- โครเนกเกอร์, 155

## ประวัติผู้แต่ง



จิตโกมุท ส่งศิริ



เกิดที่จังหวัดนครสวรรค์ สำเร็จการศึกษาระดับมัธยมศึกษาต้นจากโรงเรียนตาคลีประชาสรรค์ และโรงเรียนสตรีนครสวรรค์ ระดับมัธยมปลายจากโรงเรียนเตรียมอุดมศึกษา ระดับปริญญาตรีและโท จากสาขาระบบควบคุม ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย หลังจากทำงานเป็นผู้ช่วยวิจัย ที่ภาควิชาวิศวกรรมไฟฟ้า ได้รับทุนกระทรวงวิทยาศาสตร์ ไปศึกษาต่อระดับดุษฎีบัณฑิตจาก University of California, Los Angeles (UCLA) ในสาขาวิศวกรรมไฟฟ้า เริ่มปฏิบัติหน้าที่เป็นอาจารย์ที่ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เดือนพฤษภาคม พ.ศ.2554 จนถึงปัจจุบัน รับผิดชอบงานสอนวิชาในกลุ่มพีชคณิตเชิงเส้น การวิเคราะห์เลขเชิงซ้อน กระบวนการสุ่ม ระบบควบคุมเชิงเส้น การหาเอกลักษณ์ระบบ การเรียนรู้เชิงสถิติ และเทคนิคการแก้ปัญหาค่าเหมาะที่สุด มีความสนใจงานวิจัยด้านการเรียนรู้เชิงสถิติ การพัฒนาแบบจำลองของระบบพลวัต วิธีเชิงตัวเลขของปัญหาค่าเหมาะที่สุด ที่นำมาประยุกต์ในด้านการเรียนรู้ความเชื่อมโยงของเครือข่ายสมองจากสัญญาณคลื่นไฟฟ้าสมอง หรือ fMRI การพยากรณ์กำลังไฟฟ้าของเซลล์สุริยะ หรือปัญหาประมาณแบบจำลองของข้อมูลอนุกรมเวลาในงานประยุกต์



[jitkomut.s@chula.ac.th](mailto:jitkomut.s@chula.ac.th)



<http://jitkomut.eng.chula.ac.th/>



[https://scholar.google.com/citations?hl=en&user=I3\\_Mc9MAAAAJ](https://scholar.google.com/citations?hl=en&user=I3_Mc9MAAAAJ)



[https://www.researchgate.net/profile/Jitkomut\\_Songsiri](https://www.researchgate.net/profile/Jitkomut_Songsiri)



[https://www.youtube.com/channel/UC\\_IYuZ-vlJ\\_jAbcQdjs3Zuw](https://www.youtube.com/channel/UC_IYuZ-vlJ_jAbcQdjs3Zuw)

งานด้านบริการวิชาการ เป็นผู้อ่านบทความวิจัยให้กับวารสารระดับนานาชาติ ได้แก่ IEEE Transactions on Signal Processing, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Automatic Control, Neural Networks, และ Biocybernetics and Biomedical Engineering