บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ตัวกรองคาลมานเป็นเทคนิคการประมาณค่าวิธีหนึ่ง ซึ่งได้รับความสนใจและมีการนำไปประยุกต์ใช้กับสาขาต่าง ๆ มากมาย รวมทั้งด้านวิศวกรรมเคมี ดังนั้นเพื่อให้เห็นถึงบทบาทและความสำคัญของตัวกรองคาลมาน เนื้อหาของบทนี้จึงกล่าวถึงความเป็นมาและวัตถุประสงค์ของตัวกรองคาลมาน รวมไปถึงแนวคิดและทฤษฎีเกี่ยวกับการประมาณค่า สำหรับเนื้อหาเกี่ยวกับทฤษฎีและขั้นตอนวิธีตัวกรองคาลมานจะกล่าวถึงในบทถัดไป

นอกจากนี้ ได้ทำการรวบรวมงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับตัวกรองคาลมาน โดยงานวิจัยที่นำเสนอภายในบทนี้ส่วนใหญ่เป็นงานวิจัยทางด้านวิศวกรรมเคมี ซึ่งถูกนำไปประยุกต์ใช้กับกระบวนการต่าง ๆ

2.1 บทนำและความมุ่งหมายในอดีต

โดยทั่วไป การควบคุมกระบวนการทางเคมีที่เคยปฏิบัติกันเป็นการควบคุมแบบเชิงเส้น ทำให้การควบคุมแบบนี้ได้รับความนิยมอย่างแพร่หลาย ด้วยเหตุนี้ จึงมีการเปรียบเทียบเสถียรภาพและสมรรถนะกันมากขึ้น ดังนั้นความต้องการของการจำลองระบบแบบเชิงเส้นจึงลดน้อยลงไปเมื่อเปรียบเทียบกับการจำลองระบบแบบไม่เป็นเชิงเส้น ทั้งนี้เนื่องจากการใช้เทคนิคนี้มีข้อจำกัดถ้ากระบวนการทางเคมีมีความไม่เป็นเชิงเส้นสูง ในปัจจุบัน ทฤษฎีการควบคุมแบบไม่เชิงเส้นได้เจริญก้าวหน้ามากและนำไปประยุกต์ใช้ในกระบวนการทางเคมีได้เป็นอย่างดี ในช่วงระยะเวลาที่ผ่านมานี้ได้มีเทคนิคระบบควบคุมที่อาศัยแนวความคิดของระบบแบบไม่เป็นเชิงเส้นเกิดขึ้นเป็นจำนวนมาก

ในการออกแบบระบบควบคุม สิ่งสำคัญคือการพัฒนาแบบจำลองทางคณิตศาสตร์ของกระบวนการ แม้แต่เทคนิคการออกแบบระบบควบคุมแบบเชิงเส้นต้องใช้แบบจำลองนี้ ซึ่งได้จากการเลียนแบบของแบบจำลองกระบวนการไม่เชิงเส้น แบบจำลองกระบวนการมีความสำคัญเป็นอย่างยิ่งและความซับซ้อนของแบบจำลองขึ้นอยู่กับแบบจำลองที่สร้างขึ้น

เทคนิคตัวควบคุมแบบไม่เชิงเส้นส่วนใหญ่มักจะสมมติให้มีการวัดหรือประมาณค่าของตัวแปรสเตตทั้งหมด แต่ในทางปฏิบัตินั้น การวัดค่าตัวแปรสเตตทั้งหมดเป็นไปไม่ได้และค่าที่ได้จากการวัดจะมีสัญญาณรบกวนและ/หรืออาจเกิดความคลาดเคลื่อนขึ้นได้ นอกจากนี้แบบจำลองจริง

มักจะไม่ทราบและต้องพบกับปัญหาในการเลือกแบบจำลองกระบวนการที่มีอยู่อย่างมากมาย ซึ่งแบบจำลองต่าง ๆ ทั้งหมดอาจจะไม่ถูกต้องก็ได้ ดังนั้นจึงได้มีการนำเทคนิคการประมาณค่ามาประยุกต์ใช้เพื่อประมาณค่าตัวแปรสเตตที่ไม่ได้วัดหรือวัดไม่ได้และเพื่อลดผลกระทบของสัญญาณรบกวน เทคนิคการประมาณค่าหรือเรียกกันทั่วไปว่า "ตัวกรอง" จะให้ค่าประมาณของค่ากระบวนการจริงจากการวัดกระบวนการที่มีสัญญาณรบกวนและค่าที่คำนวณได้จากแบบจำลองกระบวนการที่เหมาะสม ด้วยเหตุนี้ การประมาณค่าตัวแปรสเตตที่ไม่ได้วัดหรือวัดไม่ได้จึงไม่ใช่ปัญหาเล็ก ๆ แล้ว โดยเฉพาะการนำเทคนิคการประมาณค่าพารามิเตอร์ที่ไม่ทราบค่ามาใช้ ทั้งนี้เนื่องจาก ต้องอาศัยความรู้ความชำนาญทางคณิตศาสตร์เป็นอย่างมากเพื่อแก้ไขแบบจำลองและการเลือกสุ่มค่าเริ่มต้นที่เหมาะสมก่อนทำการประมาณค่าพารามิเตอร์เพื่อให้การประมาณค่าประสบผลสำเร็จ

ขั้นตอนวิธีการประมาณค่าวิธีหนึ่งที่ถูกนำมาใช้กันอย่างกว้างขวาง คือตัวกรองคาลมาน ซึ่งเป็นขั้นตอนวิธีที่ให้ค่าประมาณของตัวแปรระบบที่กำลังถูกควบคุมโดยค่าการวัดของตัวตรวจรับที่เหมาะสม ทฤษฎีตัวกรองคาลมานเป็นเครื่องมือที่สำคัญสำหรับการวิเคราะห์และการแก้ไขปัญหาการประมาณค่าต่าง ๆ สมการตัวกรองคาลมานหรือขั้นตอนวิธีคาลมานช่วยลดขนาดของหน่วยความจำคอมพิวเตอร์โดยให้ค่าประมาณของสัญญาณที่ทันสมัยในช่วงเวลาการวัดโดยปราศจากการเก็บข้อมูลการวัดในอดีต นั่นคือตัวกรองคาลมานมีความยืดหยุ่นเนื่องจากมันสามารถจัดการกับค่าการวัดได้ครั้งละหนึ่งค่าหรือครั้งละหนึ่งชุด ตัวกรองคาลมานเป็นขั้นตอนวิธีการดำเนินการกับข้อมูลแบบเรียกตนเองที่เหมาะสมที่สุด

ทฤษฎีตัวกรองคาลมานเป็นทฤษฎีตัวกรองที่ได้รับการพัฒนาและมีการนำไปประยุกต์ใช้กันอย่างกว้างขวางในระบบเชิงเส้น โดยเฉพาะ การนำเทคนิคการกรองคาลมานไปประยุกต์ใช้กับการกำหนดเส้นทางการบิน, การควบคุมทิศทางและการควบคุม ซึ่งเป็นสาขาแรกที่ได้นำมาใช้ นอกจากนั้น ได้มีการนำไปประยุกต์ใช้กับสาขาต่าง ๆ อาทิเช่น การหาวงโคจร, การติดตามเครื่องบิน, การเคลื่อนที่ของเรือ, อุตสาหกรรมรถยนต์, การควบคุมกระบวนการทางด้านเคมี, การวัดอัตราการไหลในทันทีทันใดและการประมาณค่า และการทำนายตัวแปรที่วัดไม่ได้ในกระบวนการอุตสาหกรรม

ตัวกรองคาลมานมาตรฐานเป็นตัวกรองที่ให้ค่าประมาณเชิงเส้นไม่เอนเอียงตามลำดับที่ดีที่สุด (หรือค่าประมาณที่เหมาะสมที่สุดทั้งหมด) โดยมีสัญญาณรบกวนเป็นแบบเกาส์ ด้วยเหตุนี้กระบวนการสโตแคสติคที่ซับซ้อนจึงมักสมมติให้สัญญาณรบกวนเป็นแบบเกาส์ ซึ่งทำให้การคำนวณทางคณิตศาสตร์ของปัญหาการประมาณค่าง่ายขึ้น วิธีการประมาณค่าที่รู้จักกันดีมี 3 วิธี คือ (1) วิธีกำลังสองน้อยสุด (least-square), (2) วิธีภาวะน่าจะเป็นสูงสุด (maximum-likelihood) และ

(3) วิธีเบส์ (Bayesian method) โดยทั้งสามวิธีให้ค่าประมาณที่ใกล้เคียงกัน ถึงแม้ว่าความไว้วางใจ (reliability) จะแตกต่างกัน

ปัจจัยหนึ่งที่สำคัญและทำให้ทฤษฎีการประมาณค่าและการควบคุมประสบความสำเร็จ คือ เครื่องคอมพิวเตอร์ดิจิตอลที่มีความเร็วสูงและมีขนาดหน่วยความจำมากสำหรับใช้ในการแก้ไขปัญหา

ทฤษฎีสมัยใหม่ของการประมาณค่ามีรากฐานมาจากผลงานชิ้นแรกของ A. N. Kolmogorov และ N. Wiener Kolmogorov, 1941 และ Wiener, 1942 ได้พัฒนาและกำหนดสูตรทฤษฎีการประมาณค่าขั้นพื้นฐานที่สำคัญของการประมาณค่ากำลังสองเฉลี่ยน้อยที่สุดเชิงเส้น (linear minimum mean-square) ซึ่งผลเฉลยของปัญหา Wiener-Kolmogorov เดิมเป็นลักษณะเฉพาะของการกรองสัญญาณรบกวนที่มีเวลาต่อเนื่อง ขณะที่เทคนิคการถดถอยแบบเชิงเส้น (linear regression) ซึ่งอาศัยกำลังสองน้อยสุดแบบถ่วงน้ำหนัก (weighted least-square) หรือหลักการภาวะน่าจะเป็นสูงสุด เป็นลักษณะเฉพาะของการจัดการกับปัญหาการกรองที่มีเวลาไม่ต่อเนื่อง แต่อย่างไรก็ตาม ผลเฉลยของ Weiner-Kolmogorov ที่อยู่ในรูปสมการอินทิกรัลนั้นใช้ได้กับกระบวนการคงที่ (stationary) เท่านั้น จนกระทั่งต้นปี 1960 เมื่อ R. E. Kalman และต่อมา Kalman และ R. Bucy ได้ทำการพัฒนาต่อจนทำให้ตัวกรอง Kalman-Bucy ถูกนำไปประยุกต์ใช้กับระบบเชิงเส้นต่าง ๆ ได้อย่างรวดเร็วและได้พยายามที่จะนำไปประยุกต์ใช้กับระบบไม่เชิงเส้น ซึ่งผลของความพยายามทำให้มีตัวกรองคาลมานแบบยืดขยาย (extended Kalman filters) เกิดขึ้น

ตัวกรองคาลมานเป็นขั้นตอนวิธีการดำเนินการกับข้อมูลแบบเรียกตนเองที่เหมาะสมที่สุด ซึ่งให้ค่าประมาณตัวแปร (หรือสเตต) ของระบบที่กำลังถูกควบคุมโดยการใช้ประโยชน์จากแบบจำลองของระบบ และอาศัยความแตกต่างระหว่างการทำนายจากแบบจำลองกับค่าการวัด ตัวกรองคาลมานจะจัดการกับความคลาดเคลื่อนของระบบและใช้ค่าการวัดที่เหมาะสมทั้งหมดโดยไม่คำนึงถึงความถูกต้องของมันในการประมาณค่าปัจจุบันของตัวแปรที่สนใจ โดยอาศัยข้อเท็จจริงดังต่อไปนี้ คือ (1) ความรู้เกี่ยวกับแบบจำลองระบบและพลศาสตร์ของเครื่องมือการวัด, (2) ค่าสถิติของสัญญาณรบกวนระบบและความไม่แน่นอน และ (3) ข้อมูลเกี่ยวกับเงื่อนไขตอนเริ่มต้นของตัวแปร หน้าที่หลักของตัวกรองคาลมาน คือการประมาณค่าเวกเตอร์สเตตโดยใช้ตัวตรวจรับระบบและข้อมูลการวัดที่มีสัญญาณรบกวน ขั้นตอนวิธีตัวกรองคาลมานถูกนำไปใช้กับปัญหาการประมาณค่า 2 ชนิด คือ (1) การกรอง (หรือการปรับปรุงให้ทันสมัย) และ (2) การทำนาย (หรือการแผ่กระจาย) เมื่อเวลาของค่าประมาณของเวกเตอร์ที่ต้องการตรงกับค่าการวัดในอดีต เราเรียกปัญหาการประมาณค่านี้ว่า *การกรอง (filtering)* หรืออาจกล่าวได้ว่า การกรอง คือการประมาณค่าเวก

เตอร์สเตตที่เวลาปัจจุบัน โดยอาศัยค่าการวัดในอดีต เมื่อเวลาของสิ่งที่สนใจของการประมาณค่าเวกเตอร์สเตตเกิดขึ้นหลังจากค่าการวัดในอดีต ปัญหาการประมาณค่านี้ คือ *การทำนาย (prediction)*

การประยุกต์ใช้ทฤษฎีตัวกรองคาลมานต้องการคำจัดความของแบบจำลองทางคณิตศาสตร์เชิงเส้นของระบบ เมื่อพิจารณาแบบจำลองระบบจะมีความแตกต่างระหว่างแบบจำลองจริงกับแบบจำลองตัวกรอง แบบจำลองจริงเป็นการอธิบายพลศาสตร์ระบบและแบบจำลองทางสถิติของความคลาดเคลื่อนในระบบ ซึ่งสามารถแทนแบบจำลองของระบบจริงได้เหมาะสมและดีที่สุด ส่วนแบบจำลองตัวกรองเป็นแบบจำลองที่หาจากค่าเกนคาลมาน

2.2 แนวคิดและทฤษฎีเกี่ยวกับการประมาณค่า

จุดเริ่มต้นของทฤษฎีการประมาณค่าแบบดั้งเดิมสามารถอ้างอิงได้จาก R. A. Fisher ผู้ซึ่งเสนอวิธีการประมาณค่าด้วยภาวะน่าจะเป็นสูงสุด ซึ่งจากการมองปัญหาเชิงทฤษฎีเป็นวิธีทั่วไปที่สำคัญมากที่สุดของการประมาณค่า หัวข้อนี้จะอธิบายวิธีต่าง ๆ ที่ใช้ในการประมาณค่าพารามิเตอร์แบบสั้น ๆ โดยสนใจเฉพาะในการหาค่าของพารามิเตอร์ระบบ นั่นคือการหาค่าของสัมประสิทธิ์ในแบบจำลองทางคณิตศาสตร์ที่ทราบจากข้อมูลวัดด้านขาเข้าและด้านขาออก อาทิเช่น อาจจะสนใจในการหาค่าอนุพันธ์พลวัตอากาศ (aerodynamic derivative) ของเครื่องบินจากการวัดการบิน อนุพันธ์เหล่านี้ประกอบด้วยสัมประสิทธิ์ของสมการเชิงอนุพันธ์เชิงเส้นสัมพันธ์กับตัวแปรอินพุตและศิลปะของเครื่องบินในการบิน ซึ่งสามารถลดปัญหานี้เป็นการประมาณค่าพารามิเตอร์โดยใช้เครื่องมือของทฤษฎีการประมาณค่า ดังที่อธิบายด้านล่างมีผลเฉลยอยู่หลายวิธี ความแตกต่างของแต่ละวิธีสืบเนื่องมาจากความหลากหลายของสมมติฐานเกี่ยวกับความน่าจะเป็นและหลักการที่เหมาะสมที่สุด ดังนั้นในหัวข้อนี้ เราจะอธิบายการถดถอยเชิงเส้นแบบง่าย (simple linear regression), ตัวประมาณค่า “กำลังสองน้อยสุด”, และตัวประมาณค่าด้วยภาวะน่าจะเป็นสูงสุด ใน 2 วิธีแรก หลักการที่เหมาะสมที่สุดคือปริมาณสเกลาร์อยู่ในพจน์ฟังก์ชันค่า (cost function) การลดฟังก์ชันทำได้โดยการลดผลรวมของกำลังสองของอนุพันธ์ระหว่างจุดข้อมูลและจุดที่ตรงกันซึ่งได้มาจากผลเฉลย วิธีการประมาณค่าด้วยภาวะน่าจะเป็นสูงสุดมีพื้นฐานมาจากการเพิ่มฟังก์ชันความหนาแน่นความน่าจะเป็น (probability density function) ในที่นี้ จะใช้สัญลักษณ์ θ แทนเวกเตอร์พารามิเตอร์ที่ไม่ทราบค่าเพื่อใช้ประมาณค่า และสมมติให้ θ มีค่าคงที่

2.2.1 การถดถอยเชิงเส้นแบบง่าย (simple linear regression)

วัตถุประสงค์คือเพื่อพัฒนาแบบจำลอง ซึ่งใช้ในการประมาณค่าผลตอบสนองของกระบวนการและสามารถจัดช่วงระยะเวลาในระหว่างการทำนายค่าประมาณเหล่านี้ การถดถอยเชิงเส้น

แบบง่ายช่วยให้วัตถุประสงค์นี้สำเร็จได้ พิจารณาปริมาณสุ่ม Y (คือตัวแปรสุ่ม) ซึ่งเป็นฟังก์ชันของตัวแปรอิสระ (เชิงกำหนด) หนึ่งตัวหรือมากกว่านั้น $x_1, x_2, \ldots, x_m$ ในการอธิบายที่นี้ สมมติว่าตัวแปรสุ่ม Y เป็นฟังก์ชันของตัวแปรอิสระ 1 ตัวเท่านั้น และมีความสัมพันธ์แบบเชิงเส้น พจน์เชิงเส้นหมายความว่าค่าเฉลี่ยของ Y (E{Y}) เป็นฟังก์ชันเชิงเส้นกับ x เขียนได้เป็น

$$E\{Y\} = \alpha + \beta x \tag{2.1}$$

เมื่อค่าคงที่ไม่ทราบค่า 2 ค่าคือ α และ β แทนส่วนตัดและความชัน ตามลำดับ ในปัญหาทางกายภาพ ค่าคงที่ α อาจจะแทนทางขึ้น-ลงเวลา (time ramp) ค่าคงที่ไม่ทราบค่าเหล่านี้ (หรือพารามิเตอร์ประชากร) ถูกประมาณค่าจากตัวอย่างของค่า Y ที่มีค่าเกี่ยวพันกันกับ x (ข้อมูลที่ใช้ในการทดสอบและควบคุมกระบวนการเกือบจะเป็นข้อมูลตัวอย่างเสมอหรือสับเซตของข้อมูลประชากร) ด้วยเหตุนี้ เส้นการถดถอยจริงไม่ทราบ นอกจากนั้นเมื่อการสังเกตทั้งหมดไม่ได้วางอยู่บนเส้นอย่างถูกต้อง มีความคลาดเคลื่อนในค่าประมาณเส้นตรงอยู่บ้างอย่างเห็นได้ชัด เพื่อที่จะจัดการความคลาดเคลื่อนนี้ในสมการของการทำนาย Y เมื่อกำหนดค่า x ให้ โดยใช้ความสัมพันธ์

$$Y = \alpha + \beta x + e \tag{2.2}$$

เมื่อ e แทนพจน์ความคลาดเคลื่อน โดยสมมติให้กระจายแบบปกติรอบ ๆ ศูนย์และมีความแปรปรวน σ^2 ค่า e มีการเปลี่ยนแปลงเท่ากันในระดับของ x ทั้งหมดและสมมติให้ค่าความคลาดเคลื่อนเป็นอิสระ สังเกตว่าความแปรปรวนของ e เหมือนกับความแปรปรวนของ Y

แบบจำลอง (2.2) เหมาะสำหรับข้อมูลประชากร (คือเซตของค่า x และ Y ที่เป็นไปได้ทั้งหมด) ตัวแปรสุ่ม e ถูกนิยามโดย

$$e = Y - (\alpha + \beta x) \tag{2.3}$$

2.2.2 การประมาณค่ากำลังสองน้อยสุด

พิจารณาตัวอย่างแบบง่าย Y ($Y = Y_1, Y_2, \ldots, Y_m$) เท่านั้น ในการอธิบายต่อไปนี้: เราจะพิจารณาตัวอย่างที่เป็นคู่ (x_j, Y_j) เมื่อ $j = 1, 2, \ldots, n$ วิธีทางคณิตศาสตร์วิธีหนึ่งสำหรับการประมาณค่าพารามิเตอร์การถดถอย α และ β เป็นวิธีดั้งเดิมของกำลังสองน้อยสุด วิธีของกำลังสองน้อยสุดเชิงกำหนดถูกนำมาใช้เป็นครั้งแรกโดยนักคณิตศาสตร์และนักดาราศาสตร์ชาวเยอรมันชื่อ Karl Friedrich Gauss (1777-1855) ในบทความอันเป็นที่รู้จักกันดีของเขาคือ “Theoria Motus” ใน 1809 ซึ่งจัดการกับปัญหาของการหาวงโคจร

การประยุกต์วิธีนี้เพื่อหาเส้นโค้งที่เหมาะสมแบบกำลังสองน้อยสุดโดยต้องการที่จะได้อันดับของพหุนามที่พอดีกับชุดของจุดข้อมูลที่ดีที่สุด ชื่อนี้ได้มาจากคุณสมบัติของการลดผลรวมของอนุพันธ์กำลังสองจากเส้น (คือ $\sum e_i^2$) นั่นคือการเลือกค่าประมาณของ $\hat{\alpha}$ และ $\hat{\beta}$ ดังนั้นผล

รวมของผลต่างกำลังสองระหว่างค่าตัวอย่างสังเกต y_i และค่าคาดหมายโดยประมาณของ Y ($\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$) ลดลง ด้วยเหตุนี้

$$e_i = y_1 - \hat{y}_1 = y_i - (\alpha + \beta x) \tag{2.4}$$

ดังนั้นค่าประมาณกำลังสองน้อยสุด $\hat{\alpha}$ และ $\hat{\beta}$ ของ α และ β ตามลำดับ ได้โดยการลด

$$J = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left[y_i - \left(\hat{\alpha} + \hat{\beta}x\right)\right]^2 \tag{2.5}$$

คู่ตัวอย่าง-ค่า (sample-value pair) คือ $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ ขณะที่ค่า e คือเศษตกค้าง (residuals) ดังนั้นตัวประมาณค่ากำลังสองน้อยสุดของ α และ β สามารถหาได้โดยใช้อนุพันธ์ย่อยของสมการ (2.5) และให้เท่ากับศูนย์ จะได้

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.6}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{2.7}$$

เมื่อ

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{และ} \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

ในบางปัญหาทางกายภาพ เช่น จำนวนข้อมูลลดลง เมื่อต้องจัดการกับข้อมูลทดสอบจำนวนมากจึงต้องประยุกต์ใช้เวกเตอร์-เมทริกซ์เพื่ออธิบายข้างต้น สมมติเหมือนเดิมว่ามีค่าตัวอย่างสังเกต $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ ระบบของสมการการถดถอยสังเกตสามารถเขียนให้อยู่ในรูปของสมการ (2.2) ได้เป็น

$$y_i = \alpha + \beta x_i + e_i \qquad i = 1, 2, \ldots, n \tag{2.8}$$

สมมติแบบจำลองการสังเกตเชิงเส้นมีรูปแบบดังนี้ :

$$z = Hx + \eta$$

เมื่อ η คือเวกเตอร์ของสัญญาณรบกวนการวัดมิติ m และให้

$$C = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \qquad \theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

แล้วสมการ (2.8) สามารถเขียนในรูปเวกเตอร์-เมทริกซ์แบบย่อได้เป็น

$$y = C\theta + e \tag{2.9}$$

นี่คือความสัมพันธ์เชิงเส้นโดยชุดของการวัด y มีความสัมพันธ์เชิงเส้นกับปริมาณไม่ทราบค่า θ (สังเกตว่าเวกเตอร์ y อาจแทนเวกเตอร์ของข้อมูลทดสอบ) จากสมการ (2.5) ผลรวมของเศษตกค้างกำลังสอง (squared residuals) คือ

$$J(\hat{\theta}) = e^T e = (y - C\hat{\theta})^T (y - C\hat{\theta}) \tag{2.10ก}$$

ฟังก์ชันค่าของ $\hat{\theta}$ มักเขียนในรูป

$$j(\hat{\theta}) = \frac{1}{2}(y - C\hat{\theta})^T (y - C\hat{\theta}) \tag{2.10ข}$$

สังเกตว่าฟังก์ชันค่ากำลังสอง (quadratic cost function) สามารถหาค่าได้โดยต้องทราบค่า θ หรือ e ($e = y - C\theta$) มาก่อน

ค่าประมาณกำลังสองน้อยสุดของ θ ($\hat{\theta}$) สามารถหาได้โดยการลด $J(\hat{\theta})$ แสดงว่าเงื่อนไขที่จำเป็นสำหรับตัวประมาณค่ากำลังสองน้อยสุดคือ

$$\left.\frac{\partial J(\hat{\theta})}{\partial \hat{\theta}}\right|_{\theta = \hat{\theta}_{LS}} = 0 \tag{2.11}$$

ก่อนที่จะเริ่มทำการหาค่าประมาณกำลังสองน้อยสุด $\hat{\theta}$ ต้องการทฤษฎีบทประกอบดังนี้ :

ทฤษฎีบทประกอบ : สำหรับเมทริกซ์จตุรัส A มีอนุพันธ์ดังนี้ :

$$\frac{\partial}{\partial x}(x^T A z) = Az$$

$$\frac{\partial}{\partial x}(z^T A x) = A^T z$$

$$\frac{\partial}{\partial x}(x^T A x) = (A + A^T)x$$

การใช้ทฤษฎีบทประกอบข้างต้น และการกระทำกันดังสมการ (2.11) บนสมการ (2.10) ผลเฉลยของ $\hat{\theta}$ คือ

$$\left.\frac{\partial J(\hat{\theta})}{\partial \hat{\theta}}\right|_{\theta = \hat{\theta}} = C^T (y - C\hat{\theta}) = 0$$

หรือ

$$\hat{\theta} = \left[C^T C\right]^{-1} C^T y = C^+ y \tag{2.12}$$

เมื่อ $\left[C^T C\right]^{-1} C^T = C^+$ คือตัวผกผันเทียม (pseudoinverse) ของ C ให้ C มีแถวนอนมากกว่าแถวตั้ง ผลเฉลยมีเพียงหนึ่งเดียวเท่านั้น ถ้า C มีค่าลำดับชั้นเต็ม (full rank) นั่นคือแถวตั้งของ C เป็นอิสระเชิงเส้น (โดยที่ $C^+ C = I$)

สมการ (2.12) แทนค่าประมาณกำลังสองน้อยสุดของ $\hat{\theta}$ ซึ่งหมายถึงการดำเนินการกับการวัด y ที่วัดได้ด้วยกันทั้งหมดที่เวลาหนึ่งที่เรียกว่าวิธี "การดำเนินการแบบเป็นรุ่น (batch processing)" สังเกตว่าตัวผกผันของเมทริกซ์ $C^T C$ มีอยู่ (หมายความว่าตัวกำหนด (determinant) ไม่เป็นศูนย์) ถ้ามี x_i อย่างน้อย 2 ค่าที่แตกต่างกันที่อยู่ในตัวอย่าง สามารถแสดงได้ด้วยสมการ (2.12) เหมือนกับสมการ (2.6) และ (2.7) เพื่อตรวจสอบข้อเท็จจริงนี้ สังเกตการกระทำกันต่อไปนี้:

$$C^TC = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$C^Ty = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

ดังนั้นจะได้ $\hat{\theta}$ เป็น

$$\hat{\theta} = \left[C^TC\right]^{-1} C^Ty = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} - \hat{\beta}\bar{x} \\ \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{bmatrix}$$

กรณีพิเศษของปัญหาที่มีอยู่เดิมเกิดขึ้นเมื่อ n = 2 กรณีนี้เรียกว่าการวิเคราะห์การถดถอยเชิงเส้นและเป็นความพยายามที่จะหาเส้นตรงที่ดีที่สุดเพื่อให้พอดีกับชุดของข้อมูลที่ให้

สมการ (2.12) สามารถถูกปรับปรุงได้โดยการรวมเมทริกซ์ถ่วงน้ำหนัก W ผลที่ได้คือสมการ (2.12) ให้รูปแบบปรับปรุงเป็น

$$\hat{\theta} = \left[C^TWC\right]^{-1} C^TWy \tag{2.13}$$

เมื่อสมมติให้ W เป็นบวกแน่นอน (positive definite) และสมมาตร (mxm) เมทริกซ์ถ่วงน้ำหนักนี้อาจนำมาใช้เพื่อกำหนดค่า (cost) ของความคลาดเคลื่อน $(y - C\hat{\theta})_i$ แต่ละค่าแตกต่างกันไป สำหรับการถ่วงน้ำหนัก y แต่ละค่าเท่า ๆ กัน เมทริกซ์ถ่วงน้ำหนัก W กลายเป็นเมทริกซ์เอกลักษณ์ (นั่นคือ $W = I$) วิธีนี้ถูกอ้างถึงหลายครั้งว่าเป็นวิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนัก แต่ถ้าต้องการจับคู่จุดข้อมูลบางจุดให้ใกล้กันมากกว่าจุดอื่น ๆ สามารถกำหนดสัมประสิทธิ์ถ่วงน้ำหนักให้กับแต่ละพจน์ในผลบวกได้เพื่อให้มีค่าลดลง ค่าปริมาณกำลังสองน้อยสุดแบบถ่วงน้ำหนักอาจหาได้โดยการลดฟังก์ชันค่ากำลังสอง

$$J(\hat{\theta}) = \frac{1}{2}(y - C\hat{\theta})^T W (y - C\hat{\theta}) \tag{2.14}$$

ก่อนจบหัวข้อนี้ การหาความแปรปรวนของความคลาดเคลื่อนในการประมาณค่ากำลังสองน้อยสุดเป็นเรื่องที่น่าสนใจ พิจารณาแบบจำลองการวัด (คือสมการ (2.9)) ความคลาดเคลื่อนในการประมาณค่าคือ

$$\tilde{\theta}_{LS} \triangleq \theta - \hat{\theta}_{LS} = \theta - \left[C^TWC\right]^{-1} C^TWy = -\left[C^TWC\right]^{-1} C^TWe \tag{2.15}$$

และความแปรปรวนของความคลาดเคลื่อนในการประมาณค่ากำลังสองน้อยสุดเท่ากับ

$$\mathrm{var}\{\hat{\theta}_{LS}\} = E\{\tilde{\theta}_{LS}\} = -[C^T WC]^{-1} C^T WE\{e\} \tag{2.16}$$

ที่นี้เราต้องการให้ค่าคาดหมายของตัวประมาณค่า $\tilde{\theta}_{LS}$ เท่ากับค่าคาดหมายของพารามิเตอร์ θ เมื่อตัวประมาณค่าเป็นไปตามคุณสมบัตินี้ เรียกตัวประมาณค่านี้ว่าไม่เอนเอียง โดยเฉพาะตัวประมาณค่าไม่เอนเอียง ค่าคาดหมายมีค่าเท่ากับปริมาณที่กำลังถูกประมาณค่า (คือ $E\{\tilde{\theta}\} = E\{\theta\}$ สำหรับค่า θ ทั้งหมด)

จากการอธิบายข้างต้น สังเกตว่าถ้าค่าคาดหมายของ $\tilde{\theta}_{LS}$ คือ

$$E\{\tilde{\theta}_{LS}\} = E\{e\} = 0 \tag{2.17}$$

แล้ว

$$\begin{aligned}\mathrm{var}\{\theta_{LS}\} &= \mathrm{var}\left\{-\left(C^T WC\right)^{-1} C^T We\right\} \\ &= \left(C^T WC\right)^{-1} C^T WV_e WC\left(C^T WC\right)^{-1}\end{aligned} \tag{2.18}$$

เมื่อ V_e คือความแปรปรวนของ e (คือ $E\{ee^T\}$) ถ้าให้แบบจำลองการวัดคือ $z = Hx+\eta$ แล้วสมการ (2.18) อาจจะเขียนให้อยู่ในรูป

$$\mathrm{var}\{\tilde{\theta}_{LS}\} = E\{\tilde{\theta}_{LS}\tilde{\theta}_{LS}^T\} = \left(H^T WH\right)^{-1} H^T WV_\eta WH\left(H^T WH\right)^{-1} \tag{2.19}$$

อีกครั้งเมื่อ $E\{\eta\eta^T\} = V_\eta$ ส่วนขยายของการอธิบายข้างต้นนำไปสู่แนวความคิดของการประมาณค่าความแปรปรวนน้อยสุดเชิงเส้น (linear minimum-variance, LMV) สำหรับตัวประมาณค่าความแปรปรวนน้อยสุดเชิงเส้นที่ไม่มีข้อมูลเก่าเกี่ยวกับ θ (ซึ่ง $V^{-1} = 0$) ตัวประมาณค่าถูกให้โดย

$$\tilde{\theta}_{LMV} = \left(H^T V^{-1} H\right)^{-1} H^T V^{-1} z \tag{2.20}$$

และความแปรปรวนของความคลาดเคลื่อนในการประมาณค่าคือ

$$\mathrm{var}\{\tilde{\theta}_{LMV}\} = E\{\tilde{\theta}_{LMV}\tilde{\theta}_{LMV}^T\} = \left(H^T V^{-1} H\right)^{-1} \tag{2.21}$$

สังเกตว่าผลเฉลยของสมการ (2.20) ต้องการการผกผันของเมทริกซ์ $H^T V^{-1} H$ มีมิติ NxN นอกจากนี้ สังเกตว่าถ้า H = I แล้วความแปรปรวนความคลาดเคลื่อนมีค่าเท่ากับ V (ความแปรปรวนสัญญาณรบกวน) ความแปรปรวนความคลาดเคลื่อนไม่ได้ขึ้นอยู่กับการวัด z ดังนั้นสามารถคำนวณหาความแปรปรวนนอกเส้นได้ ถ้าความถูกต้องยอมรับได้ก่อนใช้การวัด การเปรียบเทียบสมการ (2.12) และ (2.20) สังเกตว่าถ้า $W = V^{-1}$ แล้วตัวประมาณค่ากำลังสองน้อยสุดจะเป็นตัวประมาณค่าความแปรปรวนน้อยสุดเชิงเส้น ผลที่ตามมาคือเมื่อค่าประมาณกำลังสองน้อยสุดเป็นเชิงเส้นและตัวประมาณค่าความแปรปรวนน้อยสุดเชิงเส้นมีความแปรปรวนน้อยที่สุด สรุปได้ว่า

$$\mathrm{var}\{\tilde{\theta}_{LMV}\} \le \mathrm{var}\{\tilde{\theta}\} \tag{2.22}$$

สำหรับ W ทั้งหมด

พิจารณาการประมาณค่าคงที่สเกลาร์ โดยทั่วไปการประมาณค่าคงที่เวกเตอร์ x ที่มีมิติ N เราทำการวัดเชิงเส้น ซึ่งถูกรบกวนโดยสัญญาณรบกวนแบบเกาส์ที่มีค่าเฉลี่ยศูนย์มีรูปแบบเป็น

$$z_k = H_k x + \eta_k \tag{2.23}$$

เมื่อ k คือเวลา, z_k เป็นเวกเตอร์การสังเกตมีมิติ M และ H_k และ u_k คือเมทริกซ์มิติ MxN และ Mx1 ตามลำดับ

ดังนั้น

$$z_k^T = [z_1, \ z_2, \ \ldots, \ z_M]$$
$$H_k^T = [H_1, \ H_2, \ \ldots, \ H_M]$$
$$\eta_k^T = [\eta_1, \ \eta_2, \ \ldots, \ \eta_M]$$

เมื่อกำหนด $\hat{x}_k$ ให้ ค่าประมาณของ x สามารถหาได้โดยลดฟังก์ชันค่า $J(\hat{x}_k)$ ได้เป็น

$$J(\hat{x}_k) = \frac{1}{2}(z_k - H_k\hat{x}_k)^T(z_k - H_k\hat{x}_k)$$

เมื่อ R_k^{-1} คือเมทริกซ์ถ่วงน้ำหนัก MxM การใช้เกรเดียนต์ (gradient) ของ $J(\hat{x}_k)$

$$\left.\frac{\partial J(\hat{x}_k)}{\partial \hat{x}_k}\right|_{\hat{x}_k = \hat{x}_{LS}} = H_k^T R_k^{-1}\left(z_k - H_k\hat{x}_{k_{LS}}\right) = 0$$

หรือ

$$\hat{x}_{k_{LS}} = \left[H_k^T R_k^{-1} H_k\right]^{-1} H_k^T R_k^{-1} z_k$$

ตัวประกอบในวงเล็บเหลี่ยม คือเมทริกซ์ความแปรปรวน NxN นั่นคือ

$$P_k \triangleq \left[H_k^T R_k^{-1} H_k\right]^{-1}$$

หรือ

$$P_k^{-1} \triangleq H_k^T R_k^{-1} H_k$$

ถ้าใช้การวัด x เพิ่มเติมแล้ว

$$z_{k+1} = H_{k+1} x + \eta_{k+1}$$

ดังนั้น

$$\hat{x}_{k+1_{LS}} = \left[H_{k+1}^T R_{k+1}^{-1} H_{k+1}\right]^{-1} H_{k+1}^T R_{k+1}^{-1} z_{k+1}$$

หรือ

$$P_{k+1}^{-1} = P_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}$$

ดังนั้น

$$P_{k+1} = \left[P_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}\right]^{-1}$$

จากการผกผันทฤษฎีประกอบเมทริกซ์ (matrix lemma inversion) ที่รู้จักดี เรามี

$$P_{k+1} = P_k + P_k H_{k+1}^T \left(R_{k+1} + H_{k+1} P_k H_{k+1}^T\right)^{-1} H_{k+1} P_k$$

เวลานี้สมมติให้ x ที่ไม่ทราบเป็นสเกลาร์และ R_k^{-1} เป็นเมทริกซ์เอกลักษณ์ แล้วสำหรับอนุกรมของการสังเกตสเกลาร์ k คือ

$$z_i = x + \eta_i$$

เมื่อ $H = [1, 1, \ldots, 1]^T$ ดังนั้นจาก

$$\hat{x}_{k_{LS}} = \left[H^T H\right]^{-1} H^T z_k$$
$$= \frac{1}{k} \sum_{i=1}^{k} z_i$$

ผลคูณ $H^T z$ รวมการวัดหารด้วย k เท่ากับค่าประมาณ ในลักษณะแบบเดียวกัน ถ้าทำการวัดค่าอื่น ๆ แล้ว

$$\hat{x}_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} z_i$$

2.2.3 ตัวประมาณค่าด้วยภาวะน่าจะเป็นสูงสุด

ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดเป็นค่าคาดหมายของปริมาณที่กำลังถูกประมาณ เมื่อกำหนดข้อมูลการวัดมาให้ วิธีการประมาณค่าด้วยภาวะน่าจะเป็นสูงสุดเริ่มมาจาก K. F. Gauss และนำมาใช้อย่างกว้างขวางโดย R. A. Risher, 1922 ตัวประมาณค่าด้วยภาวะน่าจะเป็นสูงสุดเป็นวิธีหนึ่ง ซึ่งเป็นประโยชน์มากที่สุดที่จะอธิบายในหัวข้อนี้เนื่องจากความง่ายของวิธีนี้, ความต้องการข่าวสาร (information) ทางสถิติน้อยที่สุด, และการจัดหาค่าประมาณไม่เอนเอียง โดยเฉพาะคุณสมบัติที่สำคัญข้อหนึ่งของตัวประมาณค่าด้วยภาวะน่าจะเป็นสูงสุดคือ ความยืนยง (invariance) ของมันภายใต้การแปลงแบบหาตัวผกผันได้ (invertible transformation)

สมมติอีกครั้งว่าลำดับการสังเกต z เป็นฟังก์ชันเชิงเส้นกับ θ

$$z = H\theta + \eta \tag{2.26}$$

เมื่อ $z = (z_1, z_2, \ldots, z_l)$ เป็นลำดับการสังเกต lxl, H คือเมทริกซ์ lxn, θ คือเวกเตอร์พารามิเตอร์ไม่ทราบค่า nx1 และ η คือเวกเตอร์สัญญาณรบกวน lx1 การแก้ปัญหาด้วยภาวะน่าจะเป็นสูงสุดวิธีหนึ่งคือพิจารณาการประมาณค่าเบส์ เมื่อกำหนดพารามิเตอร์ไม่ทราบค่า θ ให้ ฟังก์ชันความหนาแน่นความน่าจะเป็นแบบมีเงื่อนไข $p(z|\theta)$ ของการสังเกต z ถูกสมมติว่าทราบค่า นอกจากนั้นยังสนใจในการหาค่าฟังก์ชันความหนาแน่นแบบมีเงื่อนไขค่าใหม่ที่มีรูปแบบ pθ|z) โดยใช้กฎเบส์ เราสามารถเขียนความหนาแน่นแบบมีเงื่อนไข $p(\theta|z)$ ได้ในรูปแบบ

$$p(\theta|z) = \frac{p(z|\theta)p(\theta)}{p(z)} \tag{2.27}$$

เมื่อ $p(\theta)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็นค่าเก่าของ θ และ $p(z)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็นของการสังเกต

ก่อนที่จะทำการพัฒนาตัวประมาณค่าด้วยภาวะน่าจะเป็นสูงสุด จำเป็นต้องนิยามตัวประมาณค่า maximum a posterior (MAP) $\hat{\theta}_{MAP}$ ซึ่งแสดงให้เห็นในที่นี้ว่าแนวทางกำลังสองน้อยสุดเชิงกำหนดมีค่าเท่ากับ (หรือเทียบเท่า) แนวทาง MAP เมื่อตั้งสมมติฐานเหมาะสม โดย $p(z)$ ไม่ได้ขึ้นอยู่กับ θ คำจำกัดความเทียบเท่าของการประมาณค่า MAP คือ

$$p(\hat{\theta}_{MAP})p(z|\hat{\theta}_{MAP}) \geq p(\hat{\theta})p(z|\hat{\theta}) \quad \text{สำหรับ } \hat{\theta} \neq \hat{\theta}_{ML} \text{ ทั้งหมด} \qquad (2.28)$$

ในหลาย ๆ กรณีที่น่าสนใจ $\hat{\theta}_{MAP}$ เกิดขึ้นที่จุดคงที่ของ $p(\theta|z)$ แสดงว่า

$$\left.\frac{\partial p(\theta|z)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{MAP}} = 0 \qquad (2.29\text{ก})$$

หรือเทียบเท่า

$$\left.\frac{\partial p(\theta|z)p(\theta)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{MAP}} = 0 \qquad (2.29\text{ข})$$

ในปัญหาแบบเกาส์ โดยปกติจะหาจุดคงที่ของ $\ln p(\theta|z)$ มากกว่า $p(\theta|z)$ เมื่อ $\ln(\cdot)$ คือฟังก์ชันเพิ่มขึ้นทางเดียว ดังนั้น $\hat{\theta}_{MAP}$ ถูกให้โดย

$$\left.\frac{\partial \ln p(\theta|z)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{MAP}} = \left.\frac{\partial \ln p(z|\theta)}{\partial\theta} + \frac{\partial \ln p(\theta)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{MAP}} = 0 \qquad (2.30)$$

ด้วยเหตุนี้ จึงมักใช้ลอการิทึม ($\ln p(\theta|z)$) ซึ่งเรียกว่าฟังก์ชันภาวะน่าจะเป็นแบบล็อก (log likelihood function)

นิยามของค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดแบบปกติเป็นดังนี้ : สำหรับการสังเกต z $\hat{\theta}_{ML}(z)$ คือค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ θ ถ้า

$$p(z|\hat{\theta}_{ML}) \geq p(z|\hat{\theta}) \qquad (2.31)$$

สำหรับค่าประมาณ $\hat{\theta} \neq \hat{\theta}_{ML}$ นั่นคือ $\hat{\theta}_{ML}(z)$ เพิ่มฟังก์ชันภาวะน่าจะเป็น $p(\theta|z)$ เมื่อกำหนด z ให้ จากการอธิบายข้างต้น ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดหาได้โดยการเพิ่มฟังก์ชันภาวะน่าจะเป็นแบบล็อกบนพารามิเตอร์ไม่ทราบค่าตามด้วยการแทนที่ตัวแปร x ด้วยการสังเกต ดังนั้นจากสมการ (2.27) เราสามารถเขียนได้ว่า

$$\left.\frac{\partial \ln p(z|\theta)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{MAP}} = 0 \qquad (2.32)$$

ในทำนองเดียวกัน เรามี

$$\left.\frac{\partial \ln p(z|\theta)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{ML}} = 0 \qquad (2.33)$$

แสดงว่า $\hat{\theta}_{MAP} = \hat{\theta}_{ML}$ ดังนั้นถ้าเราไม่ทราบค่า θ มาก่อน แล้วค่าประมาณด้วยภาวะน่าจะเป็นสูงสุด และ MAP เท่ากัน ในอีกด้านหนึ่งคือค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดสมนัยกับกรณีจำกัดของค่าประมาณ MAP โดยค่าที่ทราบมาก่อนเข้าใกล้ศูนย์

การแก้ไขปัญหาการประมาณค่าด้วยภาวะน่าจะเป็นสูงสุดด้วยวิธีอื่น ๆ คือพิจารณาฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขสำหรับ z ภายใต้ค่า θ ที่กำหนดให้ ในการอธิบายที่นี้ η คือการสังเกตถูกรบกวนแบบเกาส์มีค่าเฉลี่ยเป็นศูนย์มีเมทริกซ์ความแปรปรวนร่วม R ดังนั้นความหนาแน่นความน่าจะเป็นแบบมีเงื่อนไขสามารถถูกเขียนเป็น

$$p(z|\theta) = \frac{1}{(2\pi)^{n/2}|R|^{1/2}}\exp\left[-\frac{1}{2}(z-H\theta)^T R^{-1}(z-H\theta)\right] \qquad (2.34)$$

เมื่อเมทริกซ์ความแปรปรวนร่วม R^{-1} มีมิติเป็น nxn และสอดคล้องกับเมทริกซ์ถ่วงน้ำหนัก W ในสมการ (2.13) ในการเพิ่ม p(z|θ) ต้องลดเลขชี้กำลังในวงเล็บ ฟังก์ชันภาวะน่าจะเป็นแบบล็อกของสมการ (2.34) คือ

$$\log p(z|\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det R - \frac{1}{2}(z-H\theta)^T R^{-1}(z-H\theta) \qquad (2.35)$$

ถ้าทราบ R ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุด $\hat{\theta}_{ML}$ ตรงกับค่าประมาณกำลังสองน้อยสุด ซึ่งสมนัยกับ R^{-1} ถ่วงน้ำหนัก ด้วยเหตุนี้

$$\hat{\theta}_{ML} = \left(H^T R^{-1} H\right)^{-1} H^T R^{-1} z \qquad (2.36)$$

(เปรียบเทียบกับสมการ (2.30)) โน้ตว่าเพื่อให้ได้ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดในสมการ (2.36) เมทริกซ์การสังเกต (หรือการวัด) H ต้องมีค่าลำดับชั้นเต็ม (full rank) โดยที่ $\hat{\theta}$ ลดเกณฑ์กำลังสองน้อยสุด ซึ่งลดสมการ (2.35) เมื่อกำหนด R และการสังเกต z ให้

$$\hat{\theta} \rightarrow \frac{1}{2}(z-H\theta)^T R^{-1}(z-H\theta)$$

โดยทั่วไปคุณสมบัติของค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดที่ซึ่งยังใช้ได้อยู่แม้ว่าคลาดเคลื่อนไปบ้างถูกอ้างอิงเป็นเชิงเส้นกำกับ ภายใต้เงื่อนไขทั่วไปที่เหมาะสมสามารถพิสูจน์สิ่งต่าง ๆ ได้ดังต่อไปนี้ :

1. ผลเฉลยของสมการภาวะน่าจะเป็น

$$\left.\frac{\partial \ln p(z|\theta)}{\partial\theta}\right|_{\hat{\theta}=\hat{\theta}_{ML}} = 0$$

ในความน่าจะเป็นจะลู่เข้าสู่ค่าที่ถูกต้องของ θ เมื่อ $N \to \infty$ ค่าประมาณที่มีคุณสมบัตินี้เรียกว่า *คงเส้นคงวา* (consistent)

2. ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดเป็นแบบเกาส์อย่างเชิงเส้นกำกับ

3. ค่าประมาณด้วยภาวะน่าจะเป็นสูงสุดมีประสิทธิภาพอย่างเชิงเส้นกำกับ นั่นคือ

$$\lim_{N\to\infty} \frac{\operatorname{var}\left\{\hat{\theta}_{ML} - \theta\right\}}{\left(-E\left\{\frac{\partial^2 \ln p(z|\theta)}{\partial\theta^2}\right\}\right)^{-1}} = 1 \tag{2.37}$$

สุดท้ายจาก*อสมการ Cramer-Rao* สำหรับตัวประมาณค่า*ไม่เอนเอียง*ที่ซึ่งความแปรปรวนความคลาดเคลื่อนมีเงื่อนไขเป็น

$$\operatorname{var}\left\{\hat{\theta}_{ML} - \theta\right\} \geq \frac{1}{E\left\{\left[\frac{\partial \ln p(z|\theta)}{\partial\theta}\right]^2\right\}} \tag{2.38ก}$$

หรือ

$$\operatorname{var}\left\{\hat{\theta}_{ML} - \theta\right\} \geq \left[-E\left\{\frac{\partial^2}{\partial\theta^2} \ln p(z|\theta)\right\}\right]^{-1} \tag{2.38ข}$$

ขอบเขต Cramer-Rao ให้นิยามและหาค่าประมาณที่มีประสิทธิภาพ สังเกตว่าขอบเขต Cramer-Rao ประยุกต์ใช้กับค่าประมาณของสเกลาร์เท่านั้น มีรูปแบบของขอบเขตสำหรับเวกเตอร์การประมาณค่า แต่ซับซ้อนมากกว่า ดังนั้นสำหรับค่าประมาณไม่เอนเอียง $\hat{\theta}$ ของสเกลาร์ ความแปรปรวนแบบมีเงื่อนไขถูกจำกัดเขตโดย

$$\operatorname{var}\left\{\hat{\theta}|\theta\right\} \geq \left[-E\left\{\frac{\partial^2}{\partial\theta^2} \ln p(z|\theta)\right\}\right]^{-1} \tag{2.38ค}$$

จากส่วนขยายของขอบเขต Cramer-Rao และการสมมติฟังก์ชันเวกเตอร์ สังเกตว่าสมาชิกแนวเส้นทแยงมุมของเมทริกซ์ความแปรปรวนแบบมีเงื่อนไข $\operatorname{var}\{\hat{\theta}|\theta\}$ แต่ละค่ามีค่ามากกว่าหรือเท่ากับสมาชิกแนวทแยงมุมที่ตรงกันของ J^{-1} เมื่อ J คือเมทริกซ์จตุรัสให้โดย

$$J = E\left[\frac{\partial \ln p(z|\theta)}{\partial\theta}\left(\frac{\partial \ln p(z|\theta)}{\partial\theta}\right)^T\right]$$

$$J = -E\left[\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta} \ln p(z|\theta)\right)^T\right] \tag{2.38ง}$$

โดยปกติเมทริกซ์ J เรียกกันว่า *เมทริกซ์ Fisher information* และสามารถเขียนสมการ (2.38ข) ได้เป็น

$$E\left\{\left(\hat{\theta}-\theta\right)\left(\hat{\theta}-\theta\right)^{T}\right\} \geq J^{-1}$$

ถ้า J^{-1} คงอยู่

2.2.4 การประมาณค่าของเบส์

ในการประมาณค่าพารามิเตอร์จะเกี่ยวข้องกับปัญหาของการหาค่าหรือการประมาณค่าสำหรับพารามิเตอร์ระบบ โดยเฉพาะถ้าเราแทนพารามิเตอร์ด้วย θ และค่าประมาณพารามิเตอร์ด้วย $\hat{\theta}$ เมื่อ θ คือตัวแปรสุ่มต่อเนื่อง เราสามารถกำหนดให้ค่าเป็นคู่ทั้งหมด $C[\theta,\hat{\theta}]$ บนช่วงที่สนใจ ในหลาย ๆ กรณีที่น่าสนใจทางวิศวกรรม มันเป็นไปได้ที่จะสมมติว่าค่าขึ้นอยู่กับความคลาดเคลื่อนของค่าประมาณเท่านั้น ซึ่งนิยามความคลาดเคลื่อนการประมาณค่านี้เป็น

$$\tilde{\theta} = \theta - \hat{\theta} \qquad (2.47)$$

สมการนี้บอกว่าความคลาดเคลื่อนเท่ากับค่าจริงลบด้วยค่าประมาณ ค่าคาดหมายของค่าให้โดยการเสี่ยงแบบเบส์ (Bayes risk) ในรูปแบบ

$$B(\hat{\theta}) = E\left\{C\left[\theta,\hat{\theta}(z)\right]\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C\left[\theta,\hat{\theta}(z)\right]p(\theta,z)\,d\theta\,dz \qquad (2.48)$$

สมการ (2.48) บอกว่าเราต้องกำหนดฟังก์ชันค่าและความน่าจะเป็นโดยหลักเกณฑ์หนึ่งครั้ง จึงสามารถลดการเสี่ยง ถ้าเขียน $p(\theta,z)$ ด้วย $p(z|\theta)p(\theta)$ แล้วสมการ(2.48) สามารถเขียนได้เป็น

$$B(\hat{\theta}) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} C\left[\theta,\hat{\theta}(z)\right]p(z|\theta)\,dz\right]p(\theta)\,d\theta \qquad (2.49)$$

สำหรับ $\hat{\theta}$ ที่ระบุไว้

$$B \triangleq E\left\{C(\hat{\theta})\right\} = E\left\{C(\theta-\hat{\theta})\right\} \qquad (2.50)$$

แสดงว่าการเสี่ยงแบบเบส์ คือค่าคาดหมายของความคลาดเคลื่อนในการประมาณค่า ซึ่งสามารถลดการเสี่ยงได้โดยการเลือก $\tilde{\theta}$ ที่เหมาะสม ดังนั้นค่าประมาณของเบส์ คือค่าประมาณซึ่งลดการเสี่ยงจากสมการ(2.49) สังเกตว่าอินทิกรัลภายใน (inner integral) คือค่าแบบมีเงื่อนไขเมื่อกำหนด θ ให้ และสามารถเขียนใหม่ให้อยู่ในรูปที่แทนค่าคาดหมายของค่าด้วยค่า θ ที่ให้ได้เป็น

$$B(\theta|\hat{\theta}) = E\left\{C(\theta,\hat{\theta})|\theta\right\} = \int_{-\infty}^{\infty} C\left[\theta,\hat{\theta}(z)\right]p(z|\theta)\,dz \qquad (2.51)$$

สังเกตว่าในสมการ(2.51) ค่าคาดหวัง (expectation) อยู่บนตัวแปรสุ่ม θ และตัวแปรสังเกต z

เราต้องการหาค่าการเสี่ยงที่สมนัยกับค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error, MSE) ก่อนที่จะพัฒนาเกณฑ์ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เราเขียนสมการ (2.47) ใหม่ให้อยู่ในรูป

$$B(\hat{\theta}) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} C\left[\theta,\hat{\theta}(z)\right]p(\theta|z)d\theta\right]p(z)dz \tag{2.52}$$

เมื่อแทนความหนาแน่น $p(\theta,|z)$ ด้วย $p(\theta|z)$ สำหรับการสังเกต z ถ้าเลือกที่จะลด $\hat{\theta}$ แล้ว

$$B(\hat{\theta}|z) = \int_{-\infty}^{\infty} C\left[\theta,\hat{\theta}(z)\right]p(\theta|z)d\theta \tag{2.53}$$

การแทนสมการ (2.47) ลงในสมการ (2.52) โดยมีค่ากำลังสองเฉลี่ย

$$C_{MS}(\theta,\hat{\theta}) = \left\|\theta-\hat{\theta}\right\|^2 = \sum_{i=1}^{l}\left(\theta_i-\hat{\theta}_i\right)^2 \tag{2.54}$$

จะได้นิพจน์

$$B_{MS}(\hat{\theta}|z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\|\theta-\hat{\theta}\right\|^2 p(\theta|z)d\theta \tag{2.55}$$

จากสมการ (2.53) จะได้

$$B_{MS}(\hat{\theta}|z) = \int_{-\infty}^{\infty}\left\|\theta-\hat{\theta}\right\|^2 p(\theta|z)d\theta \tag{2.56}$$

ดังนั้นตัวประมาณค่าที่เหมาะสมที่สุดสำหรับเกณฑ์ค่ากำลังสองเฉลี่ยให้โดย

$$\hat{\theta}_{MSE} = \int_{-\infty}^{\infty}\theta p(\theta|z)d\theta \tag{2.57}$$

ซึ่งเป็นค่าเฉลี่ยแบบมีเงื่อนไขของ θ ที่ให้การสังเกต z และเป็นค่าต่ำสุดเพียงตัวเดียวเท่านั้นเพราะว่าในที่สุดอนุพันธ์ที่สองของอินทิกรัลภายของสมการ (2.55) กลับเป็นเท่ากับสอง

2.2 เอกสารและผลงานที่เกี่ยวข้องกับตัวกรองคาลมาน

Kolmogorov, 1930 ได้กำหนดสูตรทฤษฎีของสัจพจน์ความน่าจะเป็น (axiomatic theory of probability) ซึ่งเป็นการค้นพบทฤษฎีของกระบวนการสโตแคสติค และในปี 1941 เขาและ Wiener ได้สร้างทฤษฎีการประมาณค่า (estimation theory) สำหรับระบบพลวัต ในต้นศตวรรษที่ 18 ตระกูลของ Bernoulli, Bayes, Gauss และคนอื่น ๆ ได้ตีพิมพ์บทความเรื่องทฤษฎีความน่าจะเป็นและวิธีกำลังสองน้อยสุดของการประมาณค่าพารามิเตอร์

ต่อมาในปี 1950 และ 1960 Kalman และ Bucy (1961) ได้เสนอเทคนิคการประมาณค่าสเตต โดยตัวกรองคาลมานได้รับการศึกษากันอย่างกว้างขวางนับตั้งแต่มีการเผยแพร่เอกสารและ

ผลงานของ Kalman (1960) และKalman และ Bucy (1961) ตัวกรองคาลมานถูกนำไปประยุกต์ใช้ในสาขาต่าง ๆ มากมาย อาทิเช่น การบิน, การเดินเรือทางทะเล, การสร้างแบบจำลองเพื่อศึกษาเรื่องประชากร, การผลิต เป็นต้น ถึงแม้ว่าจะมีเอกสารและผลงานที่เกี่ยวข้องกับตัวกรองคาลมานอยู่มากมาย แต่การนำตัวกรองคาลมานไปประยุกต์ใช้กับปัญหาทางด้านวิศวกรรมเคมีค่อนข้างจะเป็นเรื่องใหม่ การประยุกต์ใช้ตัวกรองคาลมานในอดีตที่ผ่านมาเป็นการศึกษาการจำลอง รวมไปถึงแบบจำลองกระบวนการอย่างง่าย ๆ แต่ในปัจจุบันตัวกรองคาลมานถูกนำไปประยุกต์ใช้กับปัญหาทางด้านวิศวกรรมเคมีมากขึ้น

ในเอกสารที่เกี่ยวข้องจะกล่าวถึงการหาค่าของตัวกรองไม่เชิงเส้นชนิดต่าง ๆ เช่น ตัวกรองคาลมานแบบยืดขยายสำหรับการประมาณค่าสเตตและพารามิเตอร์

การศึกษาการจำลองประกอบด้วยการนำไปประยุกต์ใช้กับกระบวนการต่าง ๆ อย่างเช่น

- ประยุกต์ใช้กับเครื่องปฏิกรณ์ถังกวนแบบต่อเนื่องโดย Seinfeld และคณะ (1969), Seinfeld (1970) และ Wells (1971)

- ประยุกต์ใช้กับเครื่องปฏิกรณ์แบบท่อและเบดหยุดนิ่งโดย Gavalas และ Seinfeld (1969), Joffe และ Sargent (1972), McGreavy และ Vago (1972) และ Vakel และคณะ (1972)

- ประยุกต์ใช้กับเครื่องแลกเปลี่ยนความร้อนโดย Coggan และ Noton (1970) และ Coggan และ Wilson (1971)

- ประยุกต์ใช้กับเตาเผาแบบใช้ออกซิเจนขั้นพื้นฐาน โดย Wells (1970)

จากการศึกษาหลาย ๆ ครั้งที่ผ่านมา อาทิเช่น การศึกษาของ Seinfeld (1970) และ Wells และ Larson (1970) ตัวกรองถูกปรับปรุงให้กลายไปเป็นส่วนหนึ่งของแผนการควบคุมแบบป้อนกลับและทำให้การควบคุมมีประสิทธิภาพดีขึ้นเป็นอย่างมาก ความเป็นไปได้ของการนำตัวกรองคาลมานไปประยุกต์ใช้กับเครื่องคอมพิวเตอร์ควบคุมกระบวนการถูกรายงานโดย Coggan และ Wilson (1971)

Goldmann และ Sargent (1971) ได้รายงานผลของการศึกษาเกี่ยวกับปัจจัยที่มีผลกระทบต่อสมรรถนะของตัวกรองคาลมานสำหรับการจำลองกระบวนการทางเคมี 2 กระบวนการ โดยพิจารณาสัญญาณรบกวนการวัดในการจำลองเท่านั้น และหาความไวของเทคนิคนี้ต่อความคลาดเคลื่อนของเมทริกซ์ในการออกแบบ, ความคลาดเคลื่อนในการสร้างแบบจำลองแพลนท์และสัญญาณรบกวนการวัด

การประยุกต์ใช้การกรองคาลมานในระบบอุตสาหกรรมมีเพียงเล็กน้อยเท่านั้น ซึ่งมีรายงานในสาขาของการควบคุมกระบวนการ Åström (1970), Sastry และ Vetter (1969) และ Sastry และคณะ (1969) เกี่ยวข้องกับการนำตัวกรองคาลมานไปประยุกต์ใช้ในอุตสาหกรรมการผลิตกระดาษ ขณะที่ Noton และคนอื่น ๆ ในปี 1968 และ 1970 ได้รายงานการใช้ตัวกรองคาลมานแบบยืดขยายสำหรับประมาณค่าสเตตและพารามิเตอร์ในระบบอุตสาหกรรมที่มีเครื่องปฏิกรณ์หลายเครื่อง Wells และ Wismer (1971) ได้นำตัวกรองคาลมานไปใช้ในการประมาณค่าของปริมาณคาร์บอนภายในเตาเผาแบบใช้ออกซิเจนขั้นพื้นฐาน

การออกแบบตัวควบคุมแบบป้อนกลับไม่ต่อเนื่องที่ลดฟังก์ชันเชิงเส้นของความแปรปรวนของค่าเบี่ยงเบนศิลปะจากเป้าหมายขึ้นอยู่กับเงื่อนไขบังคับที่เป็นไปได้บนความแปรปรวนของอินพุต สำหรับระบบเชิงเส้นขึ้นอยู่กับการรบกวนสโตแคสติค ซึ่งปฏิบัติได้จากข้อคิดเห็น 2 ข้อคือ (1) การใช้แบบจำลองฟังก์ชันถ่ายโอน (transfer function model) เพื่อแสดงคุณลักษณะพลศาสตร์กระบวนการและแบบจำลองค่าเฉลี่ยเคลื่อนที่แบบถดถอยเอง (autoregressive-moving-average model) เพื่อแสดงคุณลักษณะการรบกวนแบบสโตแคสติค และแล้วการแก้ไขปัญหาการควบคุมที่เหมาะสมที่สุดโดยใช้แนวทางเนื่องจาก box และ jenkins และทฤษฎี Wiener-Newton ที่มีชุดไม่ต่อเนื่อง และ (2) การใช้แบบจำลองตัวแปรสเตตเพื่อแสดงคุณลักษณะทั้งพลศาสตร์และส่วนสโตแคสติคของระบบ และแล้วการแก้ไขปัญหาการควบคุมที่เหมาะสมที่สุดโดยใช้ผลของการทำโปรแกรมคอมพิวเตอร์พลวัตและการกรองคาลมานรายงานโดย Macgregor, J. F. (1973)

ปี 1973 James C. Hamilton, Dale E. Seborg และ D. Grant Fisher ได้รายงานสมรรถนะของตัวกรองคาลมานในแผนการควบคุมคอมพิวเตอร์แบบหลายตัวแปรสำหรับ pilot scale evaporator โดยให้ความสนใจต่อปัจจัยทั่วไปที่มีผลกระทบต่อสมรรถนะของตัวกรองคาลมานรวมไปถึงค่าประมาณที่ไม่ถูกต้องของเมทริกซ์ความแปรปรวนร่วมสัญญาณรบกวน, การรบกวนกระบวนการที่วัดไม่ได้, ค่าประมาณสเตตเริ่มต้นที่ไม่ถูกต้อง และความคลาดเคลื่อนของพารามิเตอร์ในแบบจำลอง

ต่อมาในปี 1983 Leigh, J. R. ได้แสดงเทคนิคการคำนวณซึ่งสามารถประมาณค่าตัวแปรกระบวนการที่วัดไม่ได้จากข้อมูลแพลนท์ที่ไม่ถูกต้อง การประยุกต์ใช้เทคนิคนี้ซึ่งรวมไปถึงการใช้ขั้นตอนวิธีตัวกรองคาลมานถูกแสดงสำหรับการผลิตเหล็กกล้าและการหมัก

Chen, Chun-Yu และ Sun, Chang-Chuen (1988) ได้ศึกษาและเปรียบเทียบตัวประมาณค่าที่แตกต่างกัน 3 วิธี คือตัวกรองคาลมาน, ตัวกรองเชิงเส้นและไม่เชิงเส้น โดยนำมาใช้กับเครื่องปฏิกรณ์แบบเบดหยุดนิ่งเพื่อประมาณค่าศิลปะที่วัดไม่ได้ในเบื้องต้น โดยตำแหน่งของจุดร้อน (hot spot) เคลื่อนเมื่อเงื่อนไขการป้อนเปลี่ยนไป ซึ่งมีผลต่อความถูกต้องของผลที่ประมาณได้จากตัวกรองคาลมานและตัวประมาณค่าเชิงเส้น

ในปี 1986 Shi-Shang Jang, Babu Joseph และ Hiro Mukai ได้ทำการเปรียบเทียบวิธีการประมาณค่าสเตตและพารามิเตอร์ของระบบไม่เชิงเส้นแบบออนไลน์ 2 วิธี วิธีแรก คือ วิธีของตัวกรองคาลมานแบบยืดขยาย ส่วนวิธีที่สองคือการนำวิธีการหาค่าเหมาะที่สุด (optimization) แบบไม่เชิงเส้นมาประยุกต์ใช้เพื่อให้ฟังก์ชันของความคลาดเคลื่อนในการประมาณค่ามีค่าน้อยที่สุด โดยกระบวนการเคมีตัวอย่างที่ถูกนำมาใช้เปรียบเทียบทั้ง 2 วิธี คือระบบของเครื่องปฏิกรณ์ถังกวนแบบต่อเนื่องที่มีปฏิกิริยาคายความร้อนอันดับหนึ่ง

จากนั้น Agarwal, M. และ Bonvin, D. (1989) ได้ทำการศึกษาและรายงานถึงข้อจำกัดของตัวกรองคาลมานแบบยืดขยายสำหรับเครื่องปฏิกรณ์แบบกะ โดยการประมาณค่าสเตตและพารามิเตอร์สำหรับเครื่องปฏิกรณ์ที่ประสบผลสำเร็จนั้นได้ซ่อนข้อจำกัดไว้ รวมไปถึงวิธีการทำให้เป็นเชิงเส้น ซึ่งอาจเป็นเพราะว่าพารามิเตอร์ที่ทราบค่าหรือการเปลี่ยนแปลงจุดดำเนินการของกระบวนการเล็กน้อย ซึ่งไม่จริงสำหรับเครื่องปฏิกรณ์แบบกะ เช่น ในระบบการเกิดพอลิเมอร์ ซึ่งแสดงการประมาณค่าความเข้มข้นในระบบแบบออนไลน์จากสมดุลความร้อนที่มีค่าคงที่ปฏิกิริยาไม่ทราบค่า

ปี 1989 De Valliere, P. และ Bonvin, D. ได้ศึกษาการประมาณค่าสเตตและพารามิเตอร์ในเครื่องปฏิกรณ์แบบกะโดยใช้ตัวกรองคาลมานไม่เชิงเส้น มีจุดประสงค์เพื่อที่จะประมาณค่าสเตตของระบบปฏิกิริยาจากการวัดอุณหภูมิเท่านั้น

ต่อจากนั้นได้มีการนำตัวกรองคาลมานมาประยุกต์ใช้เพื่อลดความคลาดเคลื่อนในการสังเกตที่เกิดจากสัญญาณรบกวนในการวัด โดย Carroll, T. A. and Ramirez, W. F. (1990)

M. A. Myers and R. H. Luecke (1991) ได้ศึกษาและอธิบายถึงประสิทธิภาพของตัวกรองคาลมานแบบยืดขยายสำหรับการควบคุมระบบพลวัตต่อเนื่องที่มีการวัดไม่ต่อเนื่อง โดยระบบที่ถูกนำมาทดสอบเป็นระบบถังกวนต่อเนื่องแบบไอโซเทอร์มัลที่มีปฏิกิริยาอันดับหนึ่งผันกลับไม่ได้

Dhingra, J. S., R. L. Moose, H. Vanlandingham และ T. A. Lauzon (1992) ได้แสดงเทคนิคการประมาณค่าสำหรับระบบไม่เชิงเส้น โดยอ้างถึงวิธีของตัวอย่างเชิงตัวเลขว่าขั้นตอนวิธีที่นำเสนอมีชื่อว่าเทคนิคเมทริกซ์ค่ากระโดด (Jump Matrix Technique, JMT) ให้ผลการประมาณค่าดีกว่าตัวกรองคาลมานแบบยืดขยาย ขณะที่ M. Boutayeb และ M. Darouach, 1994 ได้แสดงให้เห็นว่าการปรับปรุงตัวกรองคาลมานแบบยืดขยายให้ดีขึ้นจะทำให้การประมาณค่ามีประสิทธิภาพดีกว่า

เทคนิคเมทริกซ์ค่ากระโดด โดยได้นำตัวอย่างเชิงตัวเลขเหมือนกับในบทความของ Dhingra แต่ใช้การประมาณแบบรัง-กัตตา (Runge-Kutta approximation) แทนการทำให้เป็นเชิงเส้นแบบอนุกรมเทย์เลอร์ ดังนั้นประสิทธิภาพของตัวกรองคาลมานแบบยืดขยายขึ้นอยู่กับวิธีการในการประมาณค่าที่นำมาใช้

ต่อมาในปี 1994 Kershenbaum, L. S. และ Kittisupakorn, P. ได้ทำการศึกษาการควบคุมอุณหภูมิของเครื่องปฏิกรณ์ถังกวนแบบกะ โดยใช้ตัวควบคุมเจเนริกโมเดล (generic model controller) ในการศึกษาครั้งนี้ได้นำตัวกรองคาลมานแบบยืดขยายมาใช้ประมาณค่าปริมาณความร้อนที่คายออกมาจากปฏิกิริยา โดยที่ตัวกรองคาลมานแบบยืดขยายไวต่อแพลนท์และ/หรือแบบจำลองผิดพลาดมากกว่าที่จะนำไปใช้ในการทำนายจากการจำลองเพียงอย่างเดียว ซึ่งมีผลโดยตรงต่อสมรรถนะของตัวควบคุมเจเนริกโมเดล และในปีถัดมาพวกเขาทั้งสองได้ทำการทดสอบการควบคุมอุณหภูมิของเครื่องปฏิกรณ์แบบกะคายความร้อนโดยใช้ตัวควบคุมเจเนริกโมเดล ในการทดสอบนี้ได้นำตัวกรองคาลมานแบบยืดขยายไปประยุกต์ใช้สำหรับประมาณค่าปริมาณความร้อนที่คายออกมาจากปฏิกิริยา และนำค่าประมาณนี้ไปใช้ในตัวควบคุมเจเนริกโมเดล ผลการจำลองแสดงว่าตัวกรองคาลมานแบบยืดขยายให้ค่าประมาณของปริมาณความร้อนที่คายออกมาถูกต้องและให้ผลการควบคุมที่ทนทานและไว้วางใจได้

ปี 1995 Hong Zhao และ Mogens Kummel ได้ศึกษาปัญหาของการประมาณค่าสเตตและพารามิเตอร์ในกระบวนการตะกอนกัมมันต์แบบสลับ (alternating activated sludge process) สำหรับการกำจัดฟอสฟอรัสทางชีวภาพ โดยใช้ตัวประมาณค่าสเตตของขั้นตอนวิธีตัวกรองคาลมานแบบยืดขยาย

Kiparissides, Verros และ Pertsinedis (1996) ได้นำการประมาณค่าพารามิเตอร์มาใช้ในเครื่องปฏิกรณ์แบบท่อสำหรับการผลิตโพลิเอทิลีนชนิดความหนาแน่นต่ำโดยใช้กระบวนการแบบความดันสูง (high-pressure low-density polyethylene tubular reactor)

นอกจากนั้นตัวกรองคาลมานได้ถูกนำไปทำงานร่วมกับตัวควบคุมแบบต่าง ๆ ด้วย อาทิเช่น ตัวควบคุมแบบป้อนกลับสเตต (state-feedback controller), ตัวควบคุมแบบโมเดลพรีดิกทีฟ (model predictive controller) และตัวควบคุมแบบข่ายงานนิวรัล (neural network controller) เป็นต้น เพื่อให้การควบคุมมีประสิทธิภาพมากขึ้น

จากงานวิจัยที่ผ่านมา จะเห็นว่าตัวกรองคาลมานได้รับความสนใจเป็นอย่างมากและมีการนำไปประยุกต์ใช้กันอย่างกว้างขวางนับตั้งแต่เริ่มมีการเผยแพร่ผลงานของ Kalman จนกระทั่งใน

ปัจจุบันนี้ ตัวกรองคาลมานได้ถูกพัฒนามากขึ้นสามารถนำไปประยุกต์ใช้กับระบบไม่เชิงเส้นได้ แต่เนื่องจากตัวกรองคาลมานเป็นเทคนิคการประมาณค่าสเตตและพารามิเตอร์แบบเรียกตัวเอง ซึ่งเป็นการทำงานแบบวนซ้ำและมีการคำนวณทางคณิตศาสตร์ที่ค่อนข้างซับซ้อน และจำเป็นต้องอาศัยเครื่องคอมพิวเตอร์เข้ามาช่วยในการคำนวณ แต่เนื่องจากการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการประมาณค่าด้วยขั้นตอนวิธีตัวกรองคาลมานได้รับความสนใจไม่มากนัก เป็นผลทำให้การประยุกต์ใช้ในอุตสาสหกรรมจริงมีน้อย แต่หากได้รับการพัฒนามากขึ้น จะทำให้มีการนำตัวกรองคาลมานไปประยุกต์ใช้ในกระบวนการต่าง ๆ ทางอุตสาหกรรมจริงมากขึ้น

ในงานวิจัยนี้ จึงได้ทำการออกแบบและเขียนโปรแกรมคอมพิวเตอร์สำหรับการประมาณค่าสเตตและพารามิเตอร์ โดยใช้ขั้นตอนวิธีตัวกรองคาลมาน โปรแกรมคอมพิวเตอร์นี้มีชื่อเรียกว่าโปรแกรม “kSTAPEN” โดยภาษาคอมพิวเตอร์ที่ใช้สำหรับการออกแบบและเขียนโปรแกรม คือภาษา Borland C++ ซึ่งเป็นภาษาที่รวมเอาความสามารถของภาษา C ที่มีความคล่องตัว, ความเร็วสูง และความสามารถใหม่เพิ่มขึ้นมา คือความเป็นโปรแกรมเชิงวัตถุ (object-oriented program) ซึ่งเป็นการออกแบบโครงสร้างการเขียนโปรแกรมที่มีประสิทธิภาพสูงสามารถแก้ไขหรือนำมาพัฒนาต่อในภายหลังได้ง่าย และมีเครื่องมือต่าง ๆ ที่ช่วยในการสร้างหน้าต่าง (window)