

วารสารวิศวกรรมศาสตร์

ปัญหาการหาค่าเหมาะสมสุดแบบเบาบางในการหาเอกลักษณ์ของระบบ

จิตโกมุท สงศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพฯ ประเทศไทย 10330

อีเมล: jitkomut.s@chula.ac.th

บทคัดย่อ บทความนี้จะกล่าวถึงปัญหาทางคณิตศาสตร์ที่พบในการหาแบบจำลอง หรือเอกลักษณ์ของระบบ (system identification) ปัญหาดังกล่าว ส่วนใหญ่มักจะเป็นปัญหาการหาค่าเหมาะสมสุด (optimization problems) โดยที่ตัวแปรของปัญหาจะแสดงถึงค่าพารามิเตอร์ของแบบจำลองที่ต้องการประมาณ สำหรับระบบที่ซับซ้อนมากโดยทั่วไปแล้ว แบบจำลองที่จะใช้อธิบายระบบ ก็ย่อมมีจำนวนพารามิเตอร์ เป็นจำนวนมากตามไปด้วย การเลือกใช้แบบจำลองที่มีจำนวนพารามิเตอร์ไม่มาก (parsimonious models) ในการอธิบายระบบที่ซับซ้อนจึงเป็นแนวทางหนึ่งที่จะถูกนำมาพิจารณาดังนั้น ในปัญหาการประมาณแบบจำลอง จึงมุ่งเน้นไปที่รูปแบบปัญหาที่จะทำให้ ตัวแปรของปัญหามีค่าเป็นศูนย์ เป็นจำนวนมาก (sparse solutions) เนื่องจาก ค่าตัวแปรที่เป็นศูนย์จะสอดคล้องกับค่าพารามิเตอร์บางตัวของแบบจำลองเป็นศูนย์นั่นเอง รูปแบบปัญหานี้ เราจึงเรียกว่าการหาเอกลักษณ์ของระบบแบบเบาบาง (sparse system identification) ในบทความนี้จะนำเสนอานิยามของระบบเบาบางในแบบจำลองประเภทต่างๆ รูปแบบปัญหาที่พบบ่อยในกลุ่มปัญหาดังกล่าว และเทคนิคการหาค่าเหมาะสมสุดนี้นำมาพิจารณาในการแก้ปัญหากลุ่มนี้

คำสืบค้น: ระบบเบาบาง, ปัญหาการหาค่าเหมาะสมสุดแบบเบาบาง, การหาเอกลักษณ์ของระบบ

วารสารวิศวกรรมศาสตร์ (ISSN: 1906-3636) ปีที่ 5 ฉบับที่ 1

วันที่ส่ง 7 ธันวาคม 2556

วันที่ตอบรับ 17 มกราคม 2557

วันที่ตีพิมพ์ 29 มกราคม 2557

Online at <http://www.ej.eng.chula.ac.th/>

DOI:10.4186/ejth.2013.5.1.51

Sparse Optimization Problems in System Identification

Jitkomut Songsiri

Department of Electrical Engineering, Chulalongkorn University, Bangkok 10330, Thailand

Email: jitkomut.s@chula.ac.th

Abstract This paper describes various optimization problems considered in system identification. The decision variables of these problems refer to model parameters that need to be estimated. It is known that a complex system should be explained by a high-dimensional model which leads to a large number of parameters to be identified. For this reason, parsimonious models are widely chosen for explaining complex systems. In this paper, we focus on model estimation formulations that promote a sparsity in the parameters and we call this type of problem as sparse system identification. We summarize the definition of sparse systems and associated estimation formulations and describe numerical techniques for solving this type of problems.

Keywords: Sparse systems, sparse optimization, system identification.

Engineering Journal (ISSN: 1906-3636) Volume 5 Issue 1

Received 7 December 2013

Accepted 17 January 2014

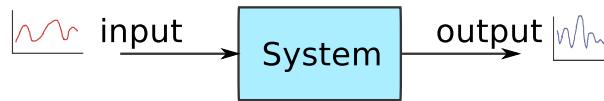
Published 29 January 2014

Online at <http://www.ej.eng.chula.ac.th/>

DOI:10.4186/ejth.2013.5.1.51

1. บทนำ

เมื่อพิจารณาระบบหนึ่งๆ ซึ่งแสดงความสัมพันธ์ระหว่างสัญญาณเข้าและสัญญาณออก ดังแสดงในรูป 1 ปัญหาการหา



รูปที่ 1 ระบบแสดงความสัมพันธ์ระหว่างสัญญาณเข้าและสัญญาณออก

เอกลักษณ์ของระบบ คือปัญหาการเลือกโครงสร้างแบบจำลองทางคณิตศาสตร์ และหาค่าพารามิเตอร์ของแบบจำลอง จากค่าสัญญาณเข้าและสัญญาณออกของระบบ โดยจุดประสงค์ในการหาเอกลักษณ์ของระบบ คือแบบจำลองที่ได้นั้น ควรจะนำมาใช้อธิบายพลวัตหรือลักษณะของระบบที่สนใจได้ดี ในแง่ที่ว่า ผลตอบจากแบบจำลองควรจะใกล้เคียงกับผลตอบจากระบบจริงมากที่สุด ยกตัวอย่างของปัญหาการหาเอกลักษณ์ เช่น การหาฟังก์ชันพหุนามอันดับ n ที่ประมาณจุดข้อมูลที่กำหนดให้ได้ใกล้เคียงที่สุด (polynomial fitting problem) การหาเมทริกซ์ความแปรปรวนร่วม (covariance matrix) ของผลการเลือกตั้งในจังหวัดต่างๆ การหาฟังก์ชันโอนย้าย (transfer function) สำหรับอธิบายพลวัตของระบบมอเตอร์กระแสตรง หรือการหาอนุกรมเวลา (time series) เพื่อใช้อธิบายราคาหุ้นที่เปลี่ยนแปลงตามเวลา เป็นต้น

เป็นที่ทราบกันดีว่า ในขั้นตอนการหาพารามิเตอร์ของแบบจำลองนั้น หากเรามีชุดจำนวนข้อมูลของสัญญาณเข้าและสัญญาณออกเป็นจำนวนมากแล้ว จะทำให้ผลการประมาณค่าพารามิเตอร์นั้น ดีกว่ากรณีที่มีจำนวนชุดข้อมูลที่ใช้ในการประมาณน้อย นอกจากนี้โดยทั่วไปแล้ว ระบบที่มีความซับซ้อนมาก มักจะถูกอธิบายได้ด้วยแบบจำลองทางคณิตศาสตร์ที่มีอันดับสูงๆ ส่งผลให้มีจำนวนตัวแปรหรือจำนวนพารามิเตอร์ที่ต้องหาในแบบจำลองนั้นมีค่ามาก จากเหตุผลดังกล่าว หากเราต้องหาแบบจำลองสำหรับระบบที่ซับซ้อนมาก แต่มีจำนวนชุดข้อมูลที่ใช้ในการประมาณมีจำกัด (ซึ่งข้อจำกัดนี้พบได้ทั่วไปในทางปฏิบัติ) อาจเกิดปัญหาที่เรียกว่า overfitting กล่าวคือ แบบจำลองจะพยายามอธิบายพลวัตส่วนที่มาจากความไม่แน่นอนของสัญญาณรบกวนที่ผ่านรวมมากับสัญญาณออกที่วัดได้ แทนที่จะอธิบายความสัมพันธ์ที่แท้จริงของสัญญาณเข้ากับสัญญาณออกของระบบ ด้วยเหตุนี้ เราจึงเลือกที่จะสนใจแบบจำลองที่มีโครงสร้างที่ไม่ซับซ้อน กล่าวคือ เรามุ่งหาแบบจำลองที่มีจำนวนพารามิเตอร์ไม่มากนัก หรือไม่ซับซ้อน (parsimonious models) แต่ยังสามารถอธิบายพลวัตของระบบได้ดีระดับหนึ่ง แนวทางนี้เป็นที่รู้จักกันดีในหัวข้อว่า การเลือกแบบจำลอง (model selection) ซึ่งเป็นการแลกกันระหว่าง ความสามารถของแบบจำลองในการอธิบายระบบ (goodness of fit) กับความซับซ้อนของแบบจำลอง (model complexity) การใช้แนวทางการเลือกแบบจำลองที่ไม่ซับซ้อนดังกล่าว มีข้อดีคือ 1) ในขั้นตอนการหาพารามิเตอร์ของแบบจำลอง จะสามารถหลีกเลี่ยงปัญหา overfitting ได้ และ 2) โครงสร้างของแบบจำลองที่ไม่ซับซ้อนนั้น จะช่วยแสดงถึง *ความสัมพันธ์ที่เด่นชัดของกลุ่มตัวแปร* ในระบบ ได้ดีกว่าการใช้แบบจำลองที่ซับซ้อน ข้อดีหลังจะเป็นจุดสำคัญที่งานวิจัยในปัจจุบันได้ให้ความสนใจ และเป็นหัวข้อหลักที่จะถูกกล่าวในบทความนี้ ยกตัวอย่างเช่น การหาแบบจำลองอนุกรมเวลา เพื่ออธิบายอัตราการเติบโตของกลุ่มยีนส์ และใช้โครงสร้างเชิงกราฟในการตีความว่า เมื่อยีนส์ตัวหนึ่งๆ มีการเจริญเติบโต จะส่งผลให้ยีนส์ตัวอื่นตัวใดเปลี่ยนแปลงบ้าง (gene expression regulatory networks discovery) [1] หรือ การหาแบบจำลองเพื่ออธิบายระดับกิจกรรมในสมองจากสัญญาณ fMRI แล้วนำแบบจำลองที่ได้มาตีความว่า เมื่อสมองในส่วนหนึ่งๆ เกิดการกระตุ้นจากสิ่งเร้าภายนอก จะส่งผลให้สมองส่วนอื่นใด เกิดการกระตุ้นในการทำงานอีกบ้าง

(functional connectivity in fMRI) [2, 3, 4, 5]

เมื่อพิจารณาการหาเอกลักษณ์ของระบบ ด้วยแนวทางเชิงพารามิเตอร์ (parametric approach) จะพบว่า ปัญหาทางคณิตศาสตร์ที่เกี่ยวข้องในการหาพารามิเตอร์ของแบบจำลองนั้น ส่วนใหญ่มักจะเป็นปัญหาการหาค่าเหมาะสมสุด (optimization problems):

$$\underset{x}{\text{minimize}} \ f(x) \quad (1)$$

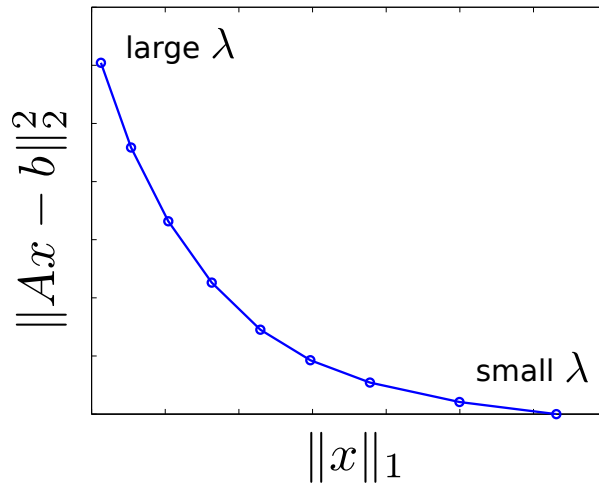
โดยที่ตัวแปรของปัญหา คือ $x \in \mathbb{R}^n$ จะแสดงถึงพารามิเตอร์ของแบบจำลองที่ต้องการประมาณ และ $f(x)$ จะเป็นฟังก์ชันที่อธิบายว่าแบบจำลองที่ใช้ค่าพารามิเตอร์เท่ากับ x นั้น จะอธิบายพลวัตของระบบได้ดีเพียงใด (goodness of fit) ยกตัวอย่างเช่น f อาจจะเป็นฟังก์ชันสูญเสียกำลังสอง (quadratic loss) หรือค่าลบของฟังก์ชันเป็นไปได้อันเป็นลบ (negative loglikelihood) สำหรับระบบที่ซับซ้อนมากโดยทั่วไปแล้ว แบบจำลองที่จะใช้อธิบายระบบ ก็ย่อมมีพารามิเตอร์เป็นจำนวนมากตามไปด้วย (x จะเป็นเวกเตอร์อันดับสูง หรือ n มีค่ามาก) ดังนั้นในปัญหาการประมาณแบบจำลองที่มุ่งเน้นในการหาแบบจำลองที่ไม่ซับซ้อน (parsimonious models) จึงสอดคล้องกับการหาแก้ปัญหา (1) ที่มีผลตอบของปัญหามีค่าเป็นศูนย์เป็นจำนวนมาก (sparse solutions) เนื่องด้วย ค่าตัวแปรที่เป็นศูนย์จะสอดคล้องกับค่าพารามิเตอร์บางตัวของแบบจำลองเป็นศูนย์นั่นเอง เราจึงเรียกปัญหาเหล่านี้ว่าเป็นการหาเอกลักษณ์ของระบบแบบเบาบาง (sparse system identification) และรูปแบบปัญหาการหาค่าเหมาะสมสุดที่จะส่งผลให้ผลตอบมีค่าเป็นศูนย์เป็นจำนวนมาก จะถูกเรียกว่า การหาค่าเหมาะสมสุดแบบเบาบาง (sparse optimization) หนึ่งในรูปแบบปัญหาในแนวทางดังกล่าว ที่นิยมใช้กันมากมีรูปแบบดังนี้

$$\underset{x}{\text{minimize}} \ f(x) + \lambda \|x\|_1 \quad (2)$$

กล่าวคือ เราเพิ่มเทอม 1-นอร์มในฟังก์ชันจุดประสงค์ และกำหนด λ ซึ่งเป็นค่าสเกลาร์บวก ใช้ในการถ่วงน้ำหนักที่จะให้ความสำคัญกับเทอม $f(x)$ หรือ $\|x\|_1$ มากกว่ากัน การใช้รูปแบบปัญหาดังข้างต้น หากใช้ค่า λ ที่มากพอ จะส่งผลให้สมาชิกหลายตัวของ x เป็นศูนย์ [6, 7] ดังนั้นรูปแบบข้างต้น จึงเป็นการแลกกันระหว่าง การเลือก x ที่มีศูนย์มากๆ (แบบจำลองที่ได้จะเป็นแบบจำลองที่ง่าย) หรือการเลือก x ที่ทำให้ค่า f ต่ำๆ (แบบจำลองอธิบายพลวัตระบบได้ดี) โดยผ่านการเลือกค่า λ ดังแสดงในรูป (2) ซึ่งเป็นตัวอย่างกรณีที่ $f(x) = \|Ax - b\|_2^2$ กล่าวคือ f เป็นฟังก์ชันกำลังสองของค่าคลาดเคลื่อนระหว่างผลตอบของแบบจำลอง กับผลตอบที่วัดได้จริง

ในกรณีที่ $f(x) = \|Ax - b\|_2^2$ ปัญหา (2) จะเป็นที่รู้จักกันดีในนามว่าปัญหา LASSO [7, 8, 9] หรือ ℓ_1 -regularized least-squares ที่ถูกนำมาพิจารณากันมากในปัญหา compressed sensing [10, 6, 11], ปัญหาการกู้สัญญาณ (signal recovery) หรือปัญหาผกผัน (inverse problem) [12, 13, 14, 15, 16] และปัญหาการหาเอกลักษณ์ของระบบ (ซึ่งจะกล่าวถึงงานวิจัยที่ผ่านมาโดยละเอียดต่อไป) นอกจากนั้นงานวิจัยที่เกี่ยวข้องกับปัญหา (2) ในอีกแง่มุมหนึ่ง คือการหาเทคนิคการหาค่าเหมาะสมสุดสำหรับ (2) เมื่อขนาดของปัญหามีขนาดใหญ่ (มากกว่าหลักหมื่น) โดยมีขั้นตอนวิธีที่นำเสนอก่อนหน้านี้ใช้แค่เกรเดียนต์ของฟังก์ชันจุดประสงค์ (gradient-based algorithms) หรือขั้นตอนวิธีที่เหมาะสมกับการคำนวณแบบกระจาย (distributed programming) เป็นจำนวนมาก [17, 18, 19, 14, 20, 21]

เมื่อกล่าวถึงรูปแบบปัญหา (2) ที่พบในการหาเอกลักษณ์ของระบบนั้น สามารถแบ่งออกได้เป็นสามหัวข้อใหญ่ๆ จากงานวิจัยที่ผ่านมา คือ 1) การหาเอกลักษณ์ของระบบเบาบาง (sparse systems) และ 2) การใช้รูปแบบการหาค่าเหมาะสมสุดแบบเบาบาง (sparse optimization formulation) และ 3) เทคนิคการหาค่าเหมาะสมสุดแบบเบาบาง ในการประมาณ



รูปที่ 2 ตัวอย่างแสดงผลการแก้ปัญหา (2) เมื่อ $f(x) = \|Ax - b\|_2^2$ สำหรับค่า λ ที่หลายๆ ค่า จะพบว่าเมื่อเลือก λ ให้มีค่ามาก จะได้ว่า $\|x\|_1$ มีค่าน้อยเพราะ x จะกลายเป็นเวกเตอร์เบาบางมาก และทำให้ $\|Ax - b\|_2^2$ มีค่าสูง เนื่องจากค่า x ดังกล่าวทำให้เกิดความคลาดเคลื่อนระหว่างแบบจำลองและข้อมูลจริงนั้นมีค่าสูงขึ้น

(sparse optimization techniques in estimation) ดังนั้น จุดประสงค์หลักของบทความนี้คือการเลือกและสรุปงานวิจัยปัจจุบันที่น่าสนใจ ในภายใต้หัวข้อทั้งสาม แต่เนื่องจากหัวข้อเรื่องเทคนิคการหาค่าเหมาะสมที่สุดนั้น เป็นหัวข้อที่มีรายละเอียดในทางคณิตศาสตร์มาก จึงขอกล่าวถึงหัวข้อนี้แบบจำกัดเท่านั้น

สัญลักษณ์ทางคณิตศาสตร์. \mathbb{R}^n คือปริภูมิของเวกเตอร์ค่าจริงขนาด n , $\mathbb{R}^{m \times n}$ คือปริภูมิของเมทริกซ์ค่าจริงขนาด $m \times n$, $\|\cdot\|$ หมายถึงนอร์มใดๆ ของเวกเตอร์หรือเมทริกซ์ สำหรับนอร์ม-1 ของ $x \in \mathbb{R}^n$ นั้น $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$ และ $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$, สำหรับ $X \in \mathbb{R}^{m \times n}$ นั้น x_{ij} จะหมายถึงสมาชิกตัวที่ (i, j) ของ X และ $\text{rank}(X)$ หมายถึงค่าลำดับชั้น (rank) ของเมทริกซ์ X , $\|X\|_*$ คือนอร์มนิวเคลียร์ของ X ซึ่งเป็นผลบวกของค่าเอกฐาน (singular value) ของ X ส่วน $\mathcal{O}(g(k))$ หมายถึง บิ๊กโอของฟังก์ชัน g (Big O of g)

2. ระบบแบบเบาบาง

ระบบแบบเบาบาง (sparse systems) นั้น ในที่นี้เราจะขออนุญาตว่า หมายถึงระบบที่มีพารามิเตอร์หลายตัวมีค่าเป็นศูนย์ ในที่นี้เราจะสนใจระบบเชิงเส้นเป็นส่วนใหญ่ และจะยกตัวอย่างระบบเบาบาง ด้วยการอธิบายจากแบบจำลองหลายๆ แบบ สำหรับแบบจำลองพลวัต (dynamical models) นั้น เราจะใช้สัญลักษณ์ $y \in \mathbb{R}^m$ ว่าเป็นสัญญาณออกของระบบ $z \in \mathbb{R}^n$ ว่าเป็นตัวแปรสถานะ และ $u \in \mathbb{R}^p$ เป็นสัญญาณเข้าของระบบ ส่วนระบบพลวัต (static models) นั้น เราจะใช้สัญลักษณ์ $y \in \mathbb{R}^m$ และ $u \in \mathbb{R}^n$ ว่าเป็นสัญญาณออกและสัญญาณเข้าของระบบ ตามลำดับ จุดมุ่งหมายของส่วนนี้คือ การรวบรวมคำจำกัดความของระบบแบบเบาบางที่น่าสนใจ รวมถึงการตีความของความเบาบางดังกล่าว ในงานวิจัยที่ผ่านมา

State-space models. หากระบบพลวัตหนึ่งๆ อธิบายได้ด้วยสมการปริภูมิสถานะ

$$\dot{z} = Az + Bu, \quad y = Cz + Du \quad (3)$$

เราจะกล่าวว่า ระบบ (3) นั้นเป็นระบบเบาบาง หาก A หรือเมทริกซ์พลวัตนั้น เป็นเมทริกซ์เบาบาง (มีศูนย์เป็นจำนวนมาก) การใช้นิยามระบบเบาบางในลักษณะนี้พบได้ใน [22] ซึ่งประมาณระบบไม่เชิงเส้นด้วยระบบเชิงเส้น และใช้เทคนิคกำลังสองเชิงเส้นต่ำสุด (linear least-squares) เพื่อประมาณหา A โดยมีเงื่อนไขบังคับคือจำนวนสมาชิกที่ไม่เป็นศูนย์ใน A ต้องไม่เกินค่าๆ หนึ่ง และนำเทคนิคนี้ไปประยุกต์ในการหาแบบจำลองของระบบเฮลิคอปเตอร์ (quad rotor helicopter) อีกกลุ่มงานวิจัยที่น่าสนใจคือ การใช้สมการปริภูมิสถานะเพื่ออธิบายพลวัตของข้อมูลพันธุกรรม (genetic data) โดยที่ z_i บอกถึงปริมาณกิจกรรมการถอดรหัส (transcription activity) ของยีนส์ i ใน [23] ได้อธิบายไว้ว่า a_{ij} คือค่าอัตราขยายของยีนส์ j ที่จะส่งผลให้เกิดการเปลี่ยนแปลงในยีนส์ i ดังนั้น ถ้าหาก $a_{ij} = 0$ จะกล่าวว่ายีนส์ i กับยีนส์ j นั้นไม่มีปฏิสัมพันธ์กัน และจะเรียก A ว่าเป็นเมทริกซ์แสดงโครงข่ายการกำกับพันธุกรรม (genetic regulation network) ใน [24, 25] ได้กล่าวว่า รูปแบบของศูนย์ของ A จึงสามารถนำไปใช้อธิบายโครงข่ายความสัมพันธ์ของตัวแปรเชิงชีวภาพ (biological network) ได้ ซึ่งมักจะมีลักษณะเป็นโครงข่ายแบบเบาบางอยู่แล้วตามธรรมชาติ โดยปัญหาที่งานวิจัยนี้สนใจคือ การหา A ที่เบาบางสำหรับข้อมูลยีนส์ และในขณะเดียวกันได้พิจารณาเงื่อนไขเสถียรภาพของ A ไว้ในปัญหาการเอกลักษณ์อีกด้วย โดยผ่านทางการใช้สมการเลี่ยนฟูเฟเป็นหนึ่งในเงื่อนไขบังคับของปัญหาการประมาณ

Finite impulse response (FIR) model. หากสัญญาณออกของระบบเชิงเส้นหนึ่งๆ อธิบายได้ด้วยสมการสังวัตนาการ (convolution equation)

$$y(t) = \sum_{k=0}^M H(t-k)u(k) + e(t) \quad (4)$$

โดยที่ $H(\cdot) \in \mathbb{R}^{m \times p}$ เป็นผลตอบอิมพัลส์ปลายตัด (truncated impulse response) อันดับ M นั่นคือ เราสมมติว่า $H(k) = 0$ สำหรับ $k > M$. เราจะกล่าวว่าระบบ (4) เป็นระบบเบาบาง ถ้าหาก H เป็นเมทริกซ์เบาบาง นั่นคือหมายความว่า เราสนใจหาระบบที่ สัญญาณเข้า u ที่ตำแหน่ง j ใดๆ จะมีผลกับสัญญาณออก y_i ที่บางตำแหน่ง i เท่านั้น ในขั้นตอนการประมาณ เราสามารถแบ่งปัญหาออกเป็นกรณีการประมาณแต่ละแถวของ H ได้ โดยที่แต่ละแถวนั้นจะเป็นเวกเตอร์แบบเบาบาง [26, 27]

สำหรับกรณีที่ (4) เป็นระบบสัญญาณเข้าและสัญญาณออกเดี่ยว (SISO) จะได้ว่าระบบเบาบางหมายถึง ค่าของ $H(k)$ (ซึ่งเป็นสเกลาร์และมักจะใช้ตัวแปร $h(k)$ หรือ h_k แทน) เป็นศูนย์อยู่จำนวนมาก เราจะพบว่า นิยามความเบาบางของแบบจำลอง FIR เช่นนี้มีประโยชน์ เพราะว่าในทางปฏิบัติเราจะไม่ทราบค่า M ล่วงหน้าว่าควรใช้ค่าเท่าใด จึงจะมีประสิทธิภาพในการประมาณมากพอ ดังนั้นเราจึงเลือก M ให้มีค่าใหญ่ แต่มีสมมติฐานให้ $h(k)$ นั้นเป็นศูนย์เยอะแทน [28] กรณีนี้จะนำไปใช้อย่างมากในงานประยุกต์ด้านตัวกรอง ที่ต้องการหาตัวกรอง FIR ที่มีค่า $h(k) \neq 0$ สำหรับบาง k อยู่เป็นจำนวนน้อย เพื่อลดความซับซ้อนของตัวกรอง แนวทางเช่นนี้ มีประโยชน์ในงานด้านแถวลำดับตัวรับรู้ (sensor array) [29] เพราะจะลดต้นทุนจำนวนแถวลำดับที่ต้องติดตั้งในทางปฏิบัติ สำหรับแนวทางนี้เป็นที่รู้จักกันในนามว่า การออกแบบตัวกรองแบบเบาบาง (sparse filter design) นอกจากนี้ แบบจำลอง FIR เบาบาง ยังถูกนำไปประยุกต์กับการประมาณช่องสัญญาณ (channel) อีกด้วย [30]

นอกจากความเบาบางของสัญญาณอิมพัลส์จะเป็นตัวบอกความซับซ้อนของแบบจำลองนี้แล้ว [31] ยังได้อธิบายถึงความสัมพันธ์ของคุณสมบัติดังกล่าว กับอัตราขยาย l_0 ของระบบเชิงเส้นเวลาวิยุตว่า แท้ที่จริงแล้ว อัตราขยาย l_0 ของระบบคือจำนวนสัมประสิทธิ์ในผลตอบอิมพัลส์ที่ไม่เป็นศูนย์ นั่นเอง และได้อธิบายถึงความเกี่ยวข้องกันกับปัญหาการหาตัว

ควบคุมที่ทำให้ระบบวงปิดมีผลตอบกลับพัลส์แบบเบาบาง

Autoregressive exogenous models. รูปแบบของแบบจำลองเป็นดังนี้

$$y(t) + a_1y(t-1) + \cdots + a_ny(t-n) = b_1u(t-1) + \cdots + b_mu(t-m) + e(t) \quad (5)$$

โดยที่ $x = (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m) \in \mathbb{R}^{n+m}$ คือพารามิเตอร์ที่ต้องการหา และ e คือสัญญาณรบกวน (noise) ใน [32, 33] ได้พิจารณาปัญหาการประมาณที่ทำให้ x เป็นเวกเตอร์เบาบาง นั่นคือ สัมประสิทธิ์ a_k หรือ b_k บางตัวเป็นศูนย์ การกระทำดังกล่าว คือการพิจารณาแบบจำลองที่ใช้จำนวนพารามิเตอร์น้อยเท่าที่เป็นไปได้ เพื่อที่จะลดผลของความแปรปรวนของค่าประมาณพารามิเตอร์นั้น แต่ในขณะเดียวกัน แบบจำลองก็ต้องมีโครงสร้างที่ซับซ้อนพอ (ซึ่งก็คือ มีจำนวนพารามิเตอร์มากพอ) ที่จะทำให้อธิบายพลวัตของระบบได้ระดับหนึ่ง

Multivariate autoregressive moving average model. สมการของแบบจำลองเป็นดังนี้

$$y(t) + A_1y(t-1) + \cdots + A_py(t-p) = B_0e(t) + B_1e(t-1) + \cdots + B_qe(t-q) \quad (6)$$

โดยที่ $A_1, \dots, A_p, B_0, B_1, \dots, B_q \in \mathbb{R}^{m \times m}$ เป็นเมทริกซ์สัมประสิทธิ์ที่ต้องการหา และ e คือสัญญาณรบกวน เราจะกล่าววาระบบ (6) เป็นระบบเบาบางเมื่อ เมทริกซ์ผกผันของความหนาแน่นเชิงสเปกตรัม (*the inverse of spectral density matrix*) เป็นเมทริกซ์แบบเบาบาง [34, 35] การนิยามเช่นนี้ มีความหมายในเชิงสถิติคือ สมมติให้ $S^{-1}(\omega)_{ij}$ เป็นสมาชิกตัวที่ (i, j) ของเมทริกซ์ผกผันของความหนาแน่นเชิงสเปกตรัม เราสามารถแสดงให้เห็นได้ว่า ถ้าหากสมมติให้ e เป็นสัญญาณรบกวนแบบเกาส์เซียน การที่ $S^{-1}(\omega)_{ij} = 0$ นั้นสมมูลกับเงื่อนไขที่ว่าสมาชิกตัวที่ i กับตัวที่ j ของ $y(t)$ นั้นเป็นอิสระต่อกัน เมื่อทราบสมาชิกตัวที่เหลือทั้งหมดของ y (y_i and y_j are conditionally independent) [36] นอกจากนี้ สำหรับกรณีเฉพาะของแบบจำลอง (6) เมื่อให้ B_k ทุกตัวเป็นศูนย์ยกเว้น $B_0 = I$ ซึ่งแสดงได้ดังนี้

$$y(t) + A_1y(t-1) + \cdots + A_py(t-p) = e(t) \quad (7)$$

เราจะเรียก (7) ว่าเป็นแบบจำลองถดถอยอัตโนมัติหลายตัวแปร (multivariate autoregressive model) สำหรับระบบนี้ ยังมีนิยามความเป็นระบบเบาบางอีกนิยามหนึ่ง กล่าวคือ เราเรียกระบบ (7) ว่าเป็นระบบเบาบาง เมื่อ A_k เป็นเมทริกซ์เบาบาง ที่มีรูปแบบศูนย์ร่วมกัน สำหรับ $k = 1, 2, \dots, p$ นิยามนี้ถูกเรียกว่าเป็นเงื่อนไขของเกรนเจอร์ (Granger constraint) [37] ซึ่งเป็นการอธิบายความสัมพันธ์ในเชิงเหตุและผลของสมาชิกสองตัวใดๆ ใน y กล่าวคือ เหตุภาพของเกรนเจอร์ (Granger causality) ได้แสดงให้เห็นว่าถ้าหาก $(A_k)_{ij} = 0$ ทุกๆ k แล้วนั้น จะได้ว่ากรูค่า y_j จะไม่มีผลต่อการทำนาย y_i สำหรับการใช้นิยามเหตุภาพของเกรนเจอร์กับแบบจำลองถดถอยอัตโนมัตินั้น ได้ถูกนำมาใช้อย่างแพร่หลาย ในการหาแบบจำลองของระบบทางชีวภาพ เพื่อหาโครงสร้างความสัมพันธ์ของตัวแปรในระบบ ซึ่งพบในงานวิจัยเช่น [1, 38, 39, 40, 41, 42, 5] เงื่อนไขเกรนเจอร์บนแบบจำลอง (7) ดังกล่าว ยังถูกพิจารณาในโดเมนเชิงความถี่อีกด้วย นั่นคือ [43] พิจารณาการหาตัวกรองวินเนอร์แบบหลายช่อง (multi-channel Wiener filter) ที่มีสัมประสิทธิ์หลายตัวเป็นศูนย์ และนำรูปแบบของศูนย์ที่ได้มาอธิบายเป็นแบบจำลองเชิงกราฟ (graphical models)

ดังนั้นปัญหาการประมาณที่รวมเงื่อนไขความเบาบางของ $S^{-1}(\omega)$ (ซึ่งจะปรากฏเป็นเงื่อนไขไม่เชิงเส้นในเมทริกซ์สัมประสิทธิ์ A_k และ B_k) หรือ ปัญหาที่รวมเงื่อนไขเหตุภาพของเกรนเจอร์ (ซึ่งจะปรากฏเป็นเงื่อนไขเชิงเส้นในเมทริกซ์

A_k) จะทำให้เราสามารถนำผลของรูปแบบศูนย์ของเมทริกซ์ที่เกี่ยวข้อง ไปตีความเป็นรูปแบบโครงสร้างความสัมพันธ์ในตัวแปร y ได้

Linear static models. พิจารณาความสัมพันธ์

$$y = Au + e \quad (8)$$

โดยที่ y คือสัญญาณออก และ u คือสัญญาณเข้า และเป็นสองสัญญาณที่ทราบค่า และ e เป็นสัญญาณรบกวนที่ไม่ทราบค่า ปัญหาการหาเอกลักษณ์ระบบในที่นี้คือการหาเมทริกซ์ A (ค่าคงที่) ที่เบาบาง ซึ่งหมายถึงตำแหน่งหนึ่งๆ ของ u จะมีผลต่อบางตำแหน่งใน y เท่านั้น ถ้าหากจะพิจารณาในกรณีทั่วๆ ไปของ (8) คือความสัมพันธ์เชิงเส้นทั่วไปแบบผสมของตัวแปรต่อเนื่องและตัวแปรวิฤต (hybrid generalized linear model หรือ GLM) [9] ที่มีสมการดังนี้

$$y = A^{cc}x + A^{dc}d + b^c + e_c, \quad z = \text{pos}(A^{cd}x + A^{dd}d + b^d + e_d) \quad (9)$$

โดยที่ x เป็นสัญญาณเข้าต่อเนื่อง d เป็นสัญญาณเข้าวิฤต y เป็นสัญญาณออกต่อเนื่อง z เป็นสัญญาณออกวิฤต e_c และ e_d เป็นสัญญาณรบกวนต่อเนื่องและวิฤตตามลำดับ และฟังก์ชัน pos นิยามคือ $\text{pos}(x)_i = 1$ ถ้า $x_i \geq 0$ และ $\text{pos}(x)_i = 0$ ถ้า $x_i < 0$

ในที่นี้ [44] กล่าวว่าระบบ (9) เป็นระบบเบาบางเมื่อเมทริกซ์ $A^{cc}, A^{dc}, A^{cd}, A^{dd}, b^c, b^d$ มีศูนย์อยู่เป็นจำนวนมาก หรือเป็นเมทริกซ์แบบเบาบางนั่นเอง

Nonlinear Volterra model. สำหรับระบบไม่เชิงเส้นที่อยู่ในรูป

$$y(t) = f(u(t), u(t-1), \dots, u(1))$$

โดยที่ u และ y เป็นสัญญาณเข้าและสัญญาณออกตามลำดับ และ f เป็นฟังก์ชันไม่เชิงเส้น จะเห็นว่าสัญญาณออกที่เวลา t จะเป็นฟังก์ชันของสัญญาณเข้าในเวลาที่อดีตจนถึงปัจจุบัน (infinite memory) แต่ถ้าหากเรามีการสมมติฐานเกี่ยวกับความจำที่จำกัด กล่าวคือ $y(t)$ นั้นเป็นฟังก์ชันของ $f(x(t), x(t-1), \dots, x(t-L+1))$ โดยที่ L เป็นค่าจำกัด สำหรับแนวทางนี้ แบบจำลองโวลเทอรา (Volterra model) เป็นอีกแบบจำลองหนึ่งซึ่งมักนำมาใช้อธิบายระบบไม่เชิงเส้น ที่มีความจำจำกัด (finite memory) ซึ่งบรรยายได้ด้วยสมการดังนี้

$$y(t) = h_0 + \sum_{p=1}^P \sum_{k_1=0}^{L-1} \cdots \sum_{k_p=0}^{L-1} h_p(k_1, k_2, \dots, k_p) \times \prod_{i=1}^p u(t-k_i) + e(t) \quad (10)$$

เราเรียก h_p ว่าเป็นเคอร์เนลโวลเทอราอันดับ p (Volterra kernel) เรียก P ว่าเป็นอันดับในการกระจายแบบโวลเทอรา (order of Volterra expansion) และ e จะเป็นสัญญาณรบกวนจุดประสงค์ในการหาเอกลักษณ์ของระบบนี้คือการประมาณ $h_p(k_1, k_2, \dots, k_p)$ สำหรับ $p = 0, 1, \dots, P$ และ k_i สำหรับ $i = 0, 1, 2, \dots, L-1$ จะสังเกตเห็นว่า ถึงแม้แบบจำลองโวลเทอราเป็นแบบจำลองไม่เชิงเส้น แต่สัญญาณออกของระบบเป็นฟังก์ชันเชิงเส้นในสัมประสิทธิ์เคอร์เนล h_p ซึ่งสามารถใช้เทคนิคกำลังสองเชิงเส้นต่ำสุด (linear least-squares) ในการหาพารามิเตอร์ของ h_p ที่มีจำนวนพารามิเตอร์เท่ากับ L^p ได้

สำหรับแบบจำลองโวลเทอร่า (10) ได้ถูกพิจารณาใน [45, 46, 47] ซึ่งได้กล่าวไว้ว่า ระบบดังกล่าวเป็นระบบแบบเบาบาง ถ้าหากหลายสัมประสิทธิ์เคอร์เนลใน $h_p(k_1, k_2, \dots, k_p)$ เป็นศูนย์ ซึ่งเป็นสมมติฐานที่พบได้สำหรับตัวอย่างประยุกต์เช่น แบบจำลองของช่องสื่อสารของแซทเทลไลท์ (satellite channel) ตัวขยายกำลังสูง (high-power amplifier) ลำโพง หรือในแบบจำลองวีเนอร์ (Wiener model)

กล่าวโดยสรุป ในส่วนนี้เราได้รวบรวมนิยามของแบบจำลองเบาบาง ตามลักษณะของแบบจำลองต่างๆ กัน โดยที่ จะพบว่าความเป็นแบบจำลองแบบเบาบางนั้นจะเกี่ยวข้องกับการที่พารามิเตอร์ของแบบจำลองมีค่าเป็นศูนย์จำนวนมาก การเลือกใช้แบบจำลองแบบเบาบาง จะส่งผลให้ผลการประมาณของระบบจริงอาจจะแย่งบ้าง แต่สิ่งที่ได้ที่ จะเป็นผลดี คือ ความแปรปรวนของพารามิเตอร์ที่ประมาณได้นั้นก็จะลดลง การแลกเปลี่ยนระหว่างข้อดีกับข้อเสียเช่นนี้ เป็นที่ทราบ กันดีว่าเกี่ยวข้องกับ หลักการแลกกันระหว่างค่าเอนเอียงกับค่าความแปรปรวน (trade off between bias and variance) นั้นเอง

3. รูปแบบการหาค่าเหมาะสมสุดแบบเบาบาง

ในหัวข้อนี้ เราจะบรรยายถึงรูปแบบปัญหาการหาค่าเหมาะสมสุดแบบเบาบาง ที่พบในปัญหาการหาเอกลักษณ์ของระบบ ตัวอย่างปัญหาเหล่านี้ได้แก่ การประมาณแบบจำลองของระบบเบาบาง การเลือกแบบจำลองอันดับต่ำ และการประมาณ แบบจำลองเชิงเส้นแปรผันตามเวลา

การประมาณแบบจำลองของระบบเบาบาง. ตัวอย่างปัญหาในกลุ่มนี้จะเป็นการประยุกต์การใช้ 1-นอร์มแบบตรง ไปตรงมา ในปัญหาการประมาณ เพื่อให้พารามิเตอร์ของแบบจำลองมีศูนย์เป็นจำนวนมาก อันดับแรก เราจะเห็นว่า สำหรับแบบจำลองใดๆ ที่สัญญาณออกของระบบเป็นฟังก์ชันเชิงเส้นในพารามิเตอร์ ที่ต้องการประมาณ ในกรณีนี้ จะได้ ว่า เราสามารถใช้วิธีกำลังสองต่ำสุดในปัญหาการประมาณได้ นั่นคือการหาค่าต่ำสุดของ $\|Ax - y\|_2^2$ โดยที่ y จะเป็น เวกเตอร์ของสัญญาณออกที่เวลาต่างๆ A จะเป็นเมทริกซ์ตัวถดถอย (regressor matrix) ที่จะประกอบไปด้วยสัญญาณ ออก และสัญญาณเข้าที่เวลาต่างๆ ในอดีต และ x จะเป็นเวกเตอร์ของพารามิเตอร์ที่ต้องการประมาณ ดังนั้น สำหรับจุด ประสงค์ที่ต้องการระบบแบบเบาบาง หรือ คือการทำให้พารามิเตอร์เป็นศูนย์จำนวนมาก รูปแบบปัญหาการประมาณจึง พิจารณาในรูป

$$\underset{x}{\text{minimize}} \quad \|Ax - y\|_2^2 + \lambda \|x\|_1 \quad (11)$$

หรือที่เรียกว่า ปัญหากำลังสองต่ำสุดแบบทำให้ปกติด้วย 1-นอร์ม (ℓ_1 -regularized least squares) ตัวอย่างแบบจำลอง และงานวิจัยที่พิจารณาปัญหานี้คือ แบบจำลองผลตอบหิมพัลส์ (4) ใน [26, 27, 28] แบบจำลอง ARX (5) ใน [32, 33] แบบจำลอง AR หลายตัวแปร (7) ที่พิจารณาเงื่อนไขเกรนเจอร์ [48, 1, 38, 39, 40, 41, 49] แบบจำลองโวลเทอร่า (10) ใน [45, 47] ทั้งนี้ ปัญหา (11) อาจจะถูกใช้ในรูป

$$\underset{x}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad \|Ax - y\| \leq \epsilon \quad (12)$$

นั่นคือ เปลี่ยนมาหาค่าต่ำสุดของ 1-นอร์มของ x โดยตรง เพื่อให้ได้พารามิเตอร์แบบจำลองเป็นศูนย์จำนวนมาก แล้ว ควบคุมความคลาดเคลื่อนของแบบจำลองกับสัญญาณออกจริง ผ่านทางเงื่อนไขบังคับแบบ ตัวอย่างเช่นปัญหาใน [33,

50) ถ้าหาก $\epsilon = 0$ รูปแบบปัญหานี้เป็นที่รู้จักกันดีในนามว่า basis pursuit [51, §6.3.2] หรือเราอาจจะพบรูปแบบปัญหา

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2 \quad \text{subject to} \quad \|x\|_1 \leq \delta \quad (13)$$

ซึ่งหมายความว่า เราต้องการหาพารามิเตอร์ที่ทำให้ความคลาดเคลื่อนของแบบจำลองต่ำสุด และเราสามารถควบคุมความเบาบางของ x ผ่านทางการปรับค่าสเกลาร์ δ ทั้งนี้ เราอาจกล่าวได้ว่า การเลือกใช้รูปแบบปัญหา (11) หรือ (12) หรือ (13) นั้น ไม่ได้ส่งผลต่อการประยุกต์เทคนิคเชิงเลขมาเพื่อหาคำตอบมากนัก (รูปแบบปัญหาทั้งสามมีความยากง่ายต่อการหาคำตอบเชิงเลขพอๆ กัน)

สำหรับวิธีการประมาณ (estimation methods) ยังมีวิธีอื่นนอกจากวิธีกำลังสองต่ำสุด เช่น การใช้ฟังก์ชันล็อกของความเป็นไปได้ (loglikelihood function) ในการบอกว่าแบบจำลองที่ใช้ค่าพารามิเตอร์ที่เลือกขึ้น จะอธิบายระบบจริงได้ดีเพียงใด (ถ้าค่าฟังก์ชันมีค่ามาก หมายความว่า พารามิเตอร์ชุดนั้นทำให้แบบจำลองใกล้เคียงกับระบบจริงมาก) สำหรับวิธีนี้ รูปแบบปัญหาการประมาณจะอยู่ในรูป

$$\underset{x}{\text{minimize}} \quad -\mathcal{L}(x) + \lambda\|x\|_1 \quad (14)$$

โดยที่ $\mathcal{L}(x)$ คือฟังก์ชันล็อกความเป็นไปได้ และเทอม 1-นอร์มที่เพิ่มเข้ามา จึงเป็นการลงโทษเพื่อให้ได้พารามิเตอร์บางตัวเป็นศูนย์ ตัวอย่างปัญหารูปแบบนี้พบได้ใน [44] รูปแบบ (14) นั้นมีความหมายเชิงสถิติคือ เป็นปัญหาการประมาณหลังประสบการณ์สูงสุด (maximum a posteriori estimation หรือ MAP) โดยที่มีฟังก์ชันการแจกแจงก่อนประสบการณ์ (prior distribution function) ของพารามิเตอร์ x เป็นแบบลาปลาเซียน (Laplacian distribution)

นอกจากนี้ ยังมีปัญหาการประมาณที่ไม่ได้พิจารณาความเบาบางของพารามิเตอร์โดยตรง แต่เป็นความเบาบางของฟังก์ชันของพารามิเตอร์ เช่น ต้องการแบบจำลอง (7) ที่เมทริกซ์ผกผันของสเปกตรัมมีศูนย์อยู่จำนวนมาก [34, 52] หรือแบบจำลองของตัวแปรสุ่ม โดยที่เราต้องการประมาณเมทริกซ์ความแปรปรวนร่วมที่มีศูนย์อยู่ในเมทริกซ์ผกผันเป็นจำนวนมาก [53] รวมถึงปัญหาที่มีเงื่อนไขบังคับเพิ่มเติมนอกจากเงื่อนไขความเบาบางของพารามิเตอร์ ซึ่งขึ้นกับงานประยุกต์ที่พิจารณา เช่น แบบจำลองที่ต้องมีเสถียรภาพ [25], จำนวนลิงก์ในแบบจำลองเชิงกราฟของยีนส์ถูกจำกัดภายใต้ค่าหนึ่งๆ [23], แบบจำลองที่มีค่าคงที่อยู่ตัว (steady-state value) เป็นค่าที่กำหนดให้ [25], แบบจำลองที่ได้มีความคลาดเคลื่อนน้อยกว่าแบบจำลองที่ได้จากวิธีประมาณพื้นฐานอื่น [33], แบบจำลองที่มีพารามิเตอร์ที่ประมาณได้มีค่าใกล้เคียงกับค่าประมาณมีประสิทธิภาพเชิงเส้นกำกับ (asymptotically efficient estimate) ในระดับหนึ่ง [50] เป็นต้น ดังนั้นเราจึงอาจกล่าวโดยสรุปว่า ปัญหาการประมาณแบบจำลองสำหรับระบบเบาบาง มีรูปแบบปัญหาทั่วไปคือ

$$\underset{x}{\text{minimize}} \quad f(x) + \lambda\|g(x)\|_1 \quad \text{subject to} \quad x \in \mathcal{C} \quad (15)$$

โดยที่ฟังก์ชัน f จะบอกถึงความคลาดเคลื่อนของแบบจำลองกับระบบจริง g เป็นฟังก์ชันที่นิยามถึงความเบาบางของแบบจำลอง และ \mathcal{C} จะบ่งบอกถึงเซตของพารามิเตอร์ที่เป็นไปได้ โดยขึ้นกับเงื่อนไขบังคับที่พิจารณาในแต่ละปัญหา

การเลือกแบบจำลองอันดับต่ำ. สำหรับหัวข้อนี้ จะใช้หลักการพื้นฐานเกี่ยวกับเมทริกซ์ผกผันใดๆ $X \in \mathbb{R}^{m \times n}$ ที่สำคัญ คือ 1) เวนิยาม $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ หรือผลบวกของค่าเอกฐาน (singular value) ของ X ว่าเป็น

นอร์มนิวเคลียร์ (nuclear norm) ของ X 2) ค่าลำดับชั้นของ X หรือ $\mathbf{rank}(X)$ คือจำนวนของค่าเอกฐานของ X ที่ไม่เป็นศูนย์ จากนิยามและข้อเท็จจริงดังกล่าว [54] จึงได้นำเสนอว่าปัญหาที่ต้องการให้ผลเฉลยมีค่าลำดับชั้นต่ำ:

$$\text{minimize } f(X) + \lambda \mathbf{rank}(X)$$

ซึ่งเป็นปัญหาไม่คอนเวกซ์นั้น สามารถแก้ได้เชิงฮิวริสติก (heuristic) ผ่านทางการประยุกต์ใช้รูปแบบปัญหาการค่าต่ำสุดของนอร์มนิวเคลียร์

$$\text{minimize } f(X) + \lambda \|X\|_* \quad (16)$$

กล่าวคือปัญหา (16) จะให้ผลตอบ X มีค่าเอกฐานหลายตัวที่เป็นศูนย์ซึ่งหมายความว่า X จะเป็นเมทริกซ์ที่มีค่าลำดับชั้นต่ำ เราจะเห็นว่า (16) เป็นเสมือนปัญหาต่อขยายของ (2) จากปัญหาของเวคเตอร์มาเป็นปัญหาของเมทริกซ์นั่นเอง

สำหรับงานประยุกต์ของผลลัพธ์ดังกล่าวข้างต้น ในการหาเอกลักษณ์ของระบบอันดับต่ำนั้น ได้ถูกพิจารณาใน [54] คือปัญหาการประมาณหาเมทริกซ์เรขาคณิต R_i ใน ฟังก์ชันโอนถ่าย $H(s) = R_0 + \sum_{i=1}^N \frac{R_i}{s-p_i}$ ซึ่งมีดีกรีแมคมิลแลน (McMillan degree) นิยามว่าเป็น

$$\deg(H) = \sum_{i=1}^N \mathbf{rank}(R_i)$$

ดังนั้น จุดประสงค์ของงาน [54] คือการหา R_i สำหรับ $i = 1, 2, \dots, N$ ที่ทำให้ $\deg(H)$ มีค่าต่ำสุด โดยพิจารณาเทอมที่เป็นนอร์มนิวเคลียร์ของ R_i ไว้ในฟังก์ชันจุดประสงค์ของปัญหาการประมาณนั่นเอง

อีกปัญหาที่เกี่ยวข้อง คือปัญหาการประมาณแบบจำลองอันดับต่ำของระบบเชิงเส้นเวลาวิฤต

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) \quad (17)$$

ในการนี้ [55] ได้เสนอผลลัพธ์ที่สำคัญว่า ถ้าให้ผลตอบฮิมพัลส์ $h_1, \dots, h_n \in \mathbb{R}$ แล้ว จะมีแบบจำลอง (A, B, C, D) อันดับต่ำสุดเท่ากับ r ที่ $h_k = CA^{k-1}B$ สำหรับ $k = 1, 2, \dots$ ก็ต่อเมื่อ

$$r = \min_{h_{n+1}, \dots, h_{2n-1} \in \mathbb{R}} \mathbf{rank}(H_n) \quad \text{โดยที่} \quad H_n = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_2 & h_3 & h_4 & \cdots & h_{n+1} \\ h_3 & h_4 & h_5 & \cdots & h_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n+1} & h_{n+2} & \cdots & h_{2n-1} \end{bmatrix},$$

เป็นเมทริกซ์แฮนเคลที่มี n พารามิเตอร์แรกเป็น h_1, h_2, \dots, h_n ที่กำหนดให้ และมีพารามิเตอร์จำนวน $n-1$ ท้าย (h_{n+1}, \dots, h_{2n-1}) เป็นตัวแปรอิสระ

ผลลัพธ์ดังกล่าวข้างต้นจึงนำมาประยุกต์ในการประมาณแบบจำลองอันดับต่ำได้ โดยตัวแปรที่ต้องการหา คือ h_k สำหรับ $k = 1, 2, \dots, 2n-1$ จากนั้นพิจารณาการหาค่าต่ำสุดของค่าลำดับชั้นของ H_n โดยมีเงื่อนไขบังคับคือ h_1, \dots, h_n จะทำให้ $\sum_{i=1}^k h_i$ (ซึ่งคือผลตอบต่อสัญญาณขั้น) มีค่าใกล้เคียงกับค่าที่เก็บจากข้อมูลจริง และให้ h_{n+1}, \dots, h_{2n-1} เป็นตัวแปรอิสระ

อีกงานวิจัยใน [56, 57] ได้พิจารณาแบบเชิงเส้นที่อธิบายด้วยสมการปริภูมิสถานะวิฤต (17) และพิจารณาสมการสัญญาณเข้าสัญญาณออก

$$Y = OX + HU \quad (18)$$

ที่ U และ Y เป็นเมทริกซ์แฮนเคิลบล็อก (block hankel matrix) ที่สร้างมาจากสัญญาณเข้าและสัญญาณออกจากสมการปริภูมิสถานะ X สร้างมาจากสัญญาณตัวแปรสถานะ H และ O จะประกอบไปด้วยเมทริกซ์ของระบบ (A, B, C, D) สมการ (18) นั้น ได้ถูกใช้ในวิธีการหาเอกลักษณ์ปริภูมิย่อย (subspace identification) ซึ่งเป็นวิธีที่ต้องการหาอันดับและเมทริกซ์ในสมการปริภูมิสถานะ (17) ในวิธีนี้ จะหาภาพฉาย (projection) ของแถวของ Y ลงบนเคอร์เนลของ U และอันดับของแบบจำลองจะบอกได้จากค่าลำดับชั้น (rank) ของภาพฉายของ Y ดังนั้น งานวิจัย [56, 57, 58] จึงได้นำเสนอรูปแบบปัญหาคอนเวกซ์ที่เพิ่มเติมที่เป็นนอร์มนิวเคลียร์ ไว้ในฟังก์ชันจุดประสงค์ ดังนี้

$$\text{minimize } \|YQ\|_* + \lambda \sum_t \|y(t) - y_m(t)\|^2 \quad (19)$$

ที่มีตัวแปรคือ $y(t)$ และ Y (ซึ่งเป็นฟังก์ชันของ A, B, C, D) โดย คอลัมน์ของ Q สร้างมาจากเซตมูลฐาน (basis) ของเคอร์เนลของ U , y คือสัญญาณออกจริง และ y_m คือสัญญาณออกที่วัดได้จากข้อมูลจริง เราจะเห็นว่าการใช้รูปแบบปัญหา (19) จึงเป็นการหาเมทริกซ์ของระบบ (A, B, C, D) ที่ทำให้ค่าลำดับชั้นของภาพฉายของ Y (หรืออันดับของแบบจำลอง) มีค่าน้อย โดยเลือกค่า λ ให้มีค่ามากพอนั่นเอง อนึ่ง [59] ได้เสนอรูปแบบทั่วไปของปัญหา (19) กล่าวคือ ปัญหาการหาค่าต่ำสุดของค่าลำดับชั้นของเมทริกซ์แฮนเคิล และนำเสนอวิธีแก้ปัญหาคอนเวกซ์ในการหาคำตอบเชิงเลข นอกจากนี้ [60] ได้นำเสนองานที่ใกล้เคียง คือการพิจารณาปัญหาการประมาณผลตอบฮิมพัลส์ h_1, h_2, \dots ที่ทำให้ $\sum_{t=1}^N (y(t) - \sum_{k=1}^{\infty} b_k u(t-k))^2$ มีค่าต่ำสุด และมีเงื่อนไขคือ ค่าลำดับชั้นของเมทริกซ์แฮนเคิล ซึ่งสร้างมาจากผลตอบฮิมพัลส์ h_1, h_2, \dots ต้องมีค่าเท่ากับ n (เลือกผลตอบฮิมพัลส์ที่อันดับจำกัด) และพิจารณาอีกหนึ่งปัญหา คือการหาแบบจำลอง ARX ที่อันดับต่ำ โดยใช้วิธีกำลังสองต่ำสุด ที่มีเงื่อนไขบังคับคือค่าลำดับชั้นของเมทริกซ์แฮนเคิลที่เกี่ยวข้องมีค่าตามที่กำหนดให้ หรืออีกงานใน [28] ที่พิจารณาปัญหาประมาณแบบจำลอง FIR แบบเบาบาง (4) และทำให้ระบบมีอันดับต่ำด้วยไปพร้อมๆ กัน โดยการพิจารณาปัญหากำลังสองต่ำสุดที่รวม 1-นอร์มของเวกเตอร์ผลตอบฮิมพัลส์ และเทอมนอร์มนิวเคลียร์ของเมทริกซ์แฮนเคิลที่เกี่ยวข้อง ดังนั้น สามปัญหานี้จึงสามารถนำมาประยุกต์ใช้นอร์มนิวเคลียร์กับเมทริกซ์แฮนเคิลได้

การประมาณแบบจำลองเชิงเส้นแปรผันตามเวลา สำหรับตัวอย่างแรก คือแบบจำลองถดถอยอัตตะภายนอก (autoregressive exogenous models) ที่มีรูปแบบการดังนี้

$$y(t) + a_1(t)y(t-1) + \dots + a_n(t)y(t-n) = b_1(t)u(t-1) + \dots + b_m(t)u(t-m) + e(t)$$

โดย e เป็นสัญญาณรบกวน ในงานของ [61] ได้พิจารณาว่า ถ้าหาก $x = (a_1, \dots, a_n, b_1, \dots, b_m)$ เป็นพารามิเตอร์ที่เปลี่ยนแปลงตามเวลาได้ ในลักษณะที่ว่า $x(t) = x_k$ สำหรับ $t_k \leq t \leq t_{k+1}$ กล่าวคือ $x(t)$ เป็นค่าคงที่เป็นช่วง (piecewise constant) ดังนั้นถ้าพิจารณาเวกเตอร์ผลต่างของพารามิเตอร์ $(x(N) - x(N-1), x(N-1) - x(N-2), \dots, x(2) - x(1))$ แล้ว ควรจะมีลักษณะเป็นเวกเตอร์แบบเบาบาง ด้วยข้อสังเกตนี้ [61] จึงได้นำเสนอรูปแบบปัญหาประมาณ $x(t)$ สำหรับ $t = 1, 2, \dots, N$ ที่เพิ่มฟังก์ชันลงโทษ (penalty function) ดังนี้

$$\text{minimize}_{x(t), t=1,2,\dots,N} \sum_{t=1}^N \|y(t) - A(t)x(t)\|_2^2 + \lambda \sum_{t=2}^N \|x(t) - x(t-1)\|_{\text{reg}} \quad (20)$$

โดย $A(t)$ เป็นเมทริกซ์ตัวถดถอย (regressor matrix) ที่ถูกสร้างจากสัญญาณ y และ u ที่เวลาในอดีต และ $\|\cdot\|_{\text{reg}}$ เป็นนอร์มใด ๆ ที่จะใช้ในฟังก์ชันลงโทษ ข้อสังเกตสำหรับปัญหา (20) มีดังนี้ 1) ในฟังก์ชันลงโทษนั้นจะใช้ผลรวมของนอร์ม (แต่ไม่ใช่ผลรวมของนอร์มกำลังสอง) การทำเช่นนี้เสมือนเป็นการหา 1-นอร์มของเวกเตอร์

$$(\|x(N) - x(N-1)\|_{\text{reg}}, x(N-1) - x(N-2), \dots, \|x(2) - x(1)\|_{\text{reg}})$$

หรือรู้จักกันว่าเป็น การทำให้เป็นปกติแบบผลรวมของนอร์ม หรือ *sum-of-norms regularization* หรือเรียกว่า Group Lasso [62] ซึ่งถือว่าเป็นปัญหาในรูปทั่วไปของ (2) ดังนั้นหากเทอมฟังก์ชันลงโทษใน (20) มีค่าน้อย (โดยการเลือก λ ให้มากพอ) จะทำให้ $\|x(t) - x(t-1)\|_{\text{reg}}$ มีค่าเป็นศูนย์ที่หลายๆ t หรือกล่าวอีกนัยหนึ่ง $x(t)$ มีค่าคงที่เป็นช่วงๆ นั่นเอง 2) ถ้าหากเลือก $\|\cdot\|_{\text{reg}}$ ให้เป็น 1-นอร์ม (กรณีนี้ได้ถูกพิจารณาใน [63]) จะได้ว่า ถึงแม้ ณ บางเวลา t เทอม $\|x(t) - x(t-1)\|_1$ มีค่าน้อย (แต่ไม่เท่ากับศูนย์) จะทำให้ พารามิเตอร์บางตัวในเวกเตอร์ $x(t)$ มีค่าคงที่ แต่ถ้าหากเลือก $\|\cdot\|_{\text{reg}}$ ให้เป็น 2-นอร์ม หรือนอร์มอื่น จะได้ว่า พารามิเตอร์ทุกตัวในเวกเตอร์ $x(t)$ ยังไม่เปลี่ยนแปลง

ตัวอย่างที่สองจาก [64] เป็นปัญหาการหาพารามิเตอร์ในแบบจำลอง ARX เหมือนกับตัวอย่างแรก แต่มีสมมติฐานที่ว่า เวกเตอร์พารามิเตอร์ x มีค่าที่เป็นไปได้ s ค่า กล่าวคือ [64] ได้เขียนปัญหาการประมาณแบบจำลองในรูป ดังนี้

$$y(t) = a^T(t)x_i + e(t)$$

โดยที่ $y(t) \in \mathbb{R}$ เป็นค่าสัญญาณออกจากระบบที่วัดได้ และ $a(t) \in \mathbb{R}^n$ เป็นเมทริกซ์ตัวถดถอย (regressor matrix) ที่ทราบค่าและมักจะประกอบไปด้วยค่าของสัญญาณออกและสัญญาณเข้าของระบบในอดีต ส่วน $x_i \in \mathbb{R}^n$ นั้นเป็นพารามิเตอร์ของแบบจำลอง โดยที่ ตัวชี้ $i \in \{1, 2, \dots, s\}$ เป็นค่าที่บ่งบอกว่า ในขณะที่เวลา t นั้น สมการของระบบจะถูกอธิบายด้วยค่าพารามิเตอร์ชุดใด และ e เป็นสัญญาณรบกวน ปัญหาการประมาณแบบจำลอง สามารถใช้วิธีกำลังสองต่ำสุดได้ ดังนี้

$$\min_{x_1, \dots, x_s} \sum_{t=1}^N \min_{i=1, 2, \dots, s} (y(t) - a^T(t)x_i)^2$$

กล่าวคือ ณ เวลา t หนึ่งๆ เราต้องการหาว่า ควรจะใช้ค่าพารามิเตอร์ของชุดใด (เลือก i จาก $\{1, 2, \dots, s\}$) และหลังจากนั้น จึงจะหาค่าของพารามิเตอร์ควรจะเป็นค่าใด (การหาค่าของ x_i) ที่จะทำให้แบบจำลองอธิบายระบบได้ดีที่สุด เราเห็นได้ว่าการแก้ปัญหานี้เกี่ยวข้องกับการค้นหาเชิงการจัด (combinatorial search) บนเซต $\{1, 2, \dots, s\}$ ดังนั้นใน [64] จึงได้เสนอข้อสังเกตว่า ถ้าเรานิยาม

$$\phi(x) = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} - \begin{bmatrix} a^T(1) \\ \vdots \\ a^T(N) \end{bmatrix} x - \begin{bmatrix} e(1) \\ \vdots \\ e(N) \end{bmatrix}$$

แล้ว จะพบว่า เมื่อใดก็ตามที่ $x = x_i$ สำหรับค่า i หนึ่งๆ ใน $\{1, 2, \dots, s\}$ แล้วนั้น เวกเตอร์ $\phi(x)$ จะกลายเป็นเวกเตอร์แบบบาง นอกจากนี้การที่ $y(t) - a^T(t)x_i = 0$ นั้น สมมูลกับ $b^T(t)z_i = 0$ โดยที่ $b(t) = [y(t) \ -a^T(t)]^T$ และ $z_i = [1 \ x_i^T]^T$ เราสังเกตว่า $\phi(x) = [b^T(1)z \ b^T(2)z \ \dots \ b^T(N)z]$ คือการทำฉายภาพเวกเตอร์ข้อมูล $b(t)$ ลงบน z ดังนั้น อาจกล่าวได้ว่าการหาว่า z_i ชุดใดจะทำให้ $\phi(z_i)$ เป็นเวกเตอร์แบบบางมากที่สุดนั้น ทำได้

โดยการแก้ปัญหาค่าต่ำสุดแบบ 1-นอร์ม

$$\underset{x}{\text{minimize}} \quad \|\phi(x)\|_1$$

ซึ่งเป็นปัญหาประเภทเดียวกับ (2) นั่นเอง

กล่าวโดยสรุป รูปแบบปัญหาการหาค่าเหมาะสมสุดแบบเบาบางที่ใช้ในการหาเอกลักษณ์ของระบบนั้น มีหลากหลาย ทั้งการประมาณหาแบบจำลองแบบเบาบางโดยตรง หรือโดยอ้อม ซึ่งหมายถึงว่า เป็นการนำเทคนิคที่ต้องการให้ผลเฉลย ปัญหาเป็นคำตอบแบบเบาบาง มาประยุกต์ใช้กับการหาเอกลักษณ์ของระบบแปรผันตามเวลา หรือ ระบบที่มีอันดับต่ำ เราอาจจะเห็นว่า เทคนิคนี้ อาจจะนำไปใช้ประยุกต์กับปัญหาทางการออกแบบระบบควบคุมก็ได้เช่นกัน ตัวอย่างเช่น หากเราต้องการออกแบบสัญญาณควบคุม $u(t)$ ให้มีค่าไม่เป็นศูนย์เป็นช่วงๆ สำหรับบาง t อาจจะได้ด้วยเหตุผลเพื่อประหยัด พลังงาน หรือลดต้นทุนในการติดต่อกับตัวขับเคลื่อน (actuator) การออกแบบเช่นนี้ก็สามารถทำได้โดยจัดรูปแบบปัญหาให้ เวกเตอร์ u ที่เวลาต่างๆ ออกมาเป็นเวกเตอร์แบบเบาบาง เป็นต้น ทั้งนี้ สิ่งที่สำคัญที่สุดในการนิยามรูปแบบปัญหา คือ ต้องกำหนดว่าต้องการปริมาณหรือลักษณะใดของระบบให้มีความเบาบาง

4. เทคนิคการหาค่าเหมาะสมสุดแบบเบาบาง

เมื่อพิจารณารูปแบบปัญหาการหาค่าเหมาะสมสุดแบบเบาบาง อาจกล่าวได้ว่ามีรูปทั่วไปคือ (15) ดังนี้

$$\underset{x}{\text{minimize}} \quad f(x) + h(x) \quad \text{subject to } x \in \mathcal{C} \quad (21)$$

โดย $h(x) = \lambda \|g(x)\|_1$ เราจะเห็นว่าลักษณะของปัญหาในกลุ่มนี้ คือ **ข้อแรก** ฟังก์ชันจุดประสงค์มีเทอม h ซึ่ง เกี่ยวข้องกับ 1-นอร์ม (หรืออาจจะอยู่ในเงื่อนไขบังคับ) และเป็นที่ทราบกันดีว่าฟังก์ชันนอร์มนั้น เป็นฟังก์ชันที่ไม่สามารถ หาอนุพันธ์ได้ที่ $x = 0$ ปัญหาในกลุ่มนี้จึงเป็นปัญหาไม่เรียบ (nonsmooth optimization problem) จึงจำเป็นต้องหาขั้นตอนวิธีเชิงเลขสำหรับกลุ่มปัญหานี้ [65, §3] **ข้อสอง** เมื่อฟังก์ชัน f เป็นฟังก์ชันคอนเวกซ์ และเซต \mathcal{C} เป็นเซตคอนเวกซ์ แล้ว ปัญหา (15) จะเป็นปัญหาคอนเวกซ์ ที่สามารถนำขั้นตอนวิธีที่พัฒนาไว้ [51] แล้วมาประยุกต์ได้ทันที เราอาจกล่าวได้ว่าความเป็นคอนเวกซ์ของปัญหามักจะพบได้ในงานประยุกต์ของปัญหาการประมาณโดยทั่วไป เพราะฟังก์ชัน f มักจะเป็นฟังก์ชันกำลังสอง หรือค่าลบของฟังก์ชันความเป็นไปได้ ซึ่งเหล่านี้ ล้วนแต่เป็นฟังก์ชันคอนเวกซ์ **ข้อสาม** เนื่องจากตัวแปร x แสดงถึงค่าพารามิเตอร์ในแบบจำลอง ดังนั้น สำหรับการหาเอกลักษณ์ของระบบซับซ้อน จึงได้ว่า x มักจะเป็นตัวแปรที่มีอันดับสูง (high-dimensional variable) ขั้นตอนวิธีที่นำมาพิจารณาจึงควรจะต้องเหมาะสมกับปัญหาที่มีขนาดใหญ่ (large-scale optimization) [66] หรือมีลักษณะการคำนวณแบบกระจาย (distributed computing) [20]

ดังนั้น ในส่วนนี้จะขอบรรยายถึง ตัวอย่างของขั้นตอนวิธีที่ใช้แก้ปัญหา (15) ที่เป็นปัญหาคอนเวกซ์และเหมาะสมกับการแก้ปัญหามาตรฐานใหญ่ ที่ได้รับความสนใจในช่วงสิบปีที่ผ่านมาจนถึงปัจจุบัน

- [67] นำเสนอวิธีจุดภายใน (interior-point method) ที่คำนึงถึงโครงสร้างของเมทริกซ์ฮessian อันดับสอง (Hessian matrix) ในการก้าวนิวตัน (Newton step) เพื่อช่วยเพิ่มประสิทธิภาพในการคำนวณ หรือใน [68] ได้พิจารณาวิธีจุดภายในเช่นกัน สำหรับปัญหาการหาเอกลักษณ์ของระบบที่เกี่ยวข้องกับเทอมที่มีนอร์มนิวเคลียร์

- วิธีซิดแบ่งและหดตัวแบบวนซ้ำ (iterative shrinking-thresholding method) ที่นำเสนอใน [69, 70] ซึ่งมีหลักการคือ การพิจารณาการเกรเดียนท์ (gradient step) ที่เกรเดียนท์ถูกคำนวณมาจากเทอม f (เทอมที่หาอนุพันธ์ได้เท่านั้น) และนำการเกรเดียนท์นั้นมาใส่ในฟังก์ชันซิดแบ่ง (thresholding) วิธีนี้มีข้อดีคือ การคำนวณอาศัยเพียงเกรเดียนท์เท่านั้น (ไม่ใช่อนุพันธ์อันดับสอง) และการคำนวณฟังก์ชันซิดแบ่ง ก็มักจะเป็นการคำนวณที่ไม่ซับซ้อน จึงเหมาะสมกับปัญหาที่มีขนาดใหญ่ โดยมีอัตราการลู่เข้าเป็น $\mathcal{O}(1/k)$ ซึ่ง k คือตัวชี้การวนซ้ำ (iteration index) วิธีนี้อาจใช้ร่วมกับการเลือกใช้ระยะก้าว (step size) แบบค้นหาเส้นตรง (line search) หรือวิธี Barzilai-Borwein [71] ที่ให้ผลการลู่เข้าที่เร็วในทางปฏิบัติ
- วิธีอันดับหนึ่งแบบเหมาะสมที่สุด (optimal first-order methods) หรือบางครั้งเรียกว่า วิธีเกรเดียนท์แบบเร่ง (accelerated gradient methods) วิธีนี้จะคล้ายกับวิธี Iterative shrinking-thresholding ในแง่ที่ว่า อาศัยเพียงแค่เกรเดียนท์ของ f แต่มีการนิยามตัวแปรเสริมในการวนซ้ำเพิ่มขึ้นมา โดยที่การคำนวณตัวแปรเสริมนั้น เป็นเพียงแค่ฟังก์ชันเชิงเส้นของตัวแปรหลัก จึงไม่ทำให้มีการคำนวณเพิ่มอย่างมีนัยยะแต่อย่างใด อย่างไรก็ตาม วิธีนี้เป็นวิธีที่เหมาะสมที่สุด เนื่องจากมีอัตราการลู่เข้าเป็น $\mathcal{O}(1/k^2)$ ซึ่ง [17, 65] ได้พิสูจน์ไว้ว่า เป็นอัตราการลู่เข้าที่ดีที่สุดที่ทำได้ ในกลุ่มปัญหาไม่เรียบที่ f เป็นฟังก์ชันต่อเนื่องแบบลิปชิตซ์ (Lipchitz continuous) สำหรับวิธีในกลุ่มนี้ได้มี [14, 17] นำเสนอขึ้น และมีงานหลายงานนำไปประยุกต์และพัฒนาต่อ ตัวอย่างเช่น [18, 19, 52, 72]
- วิธีสลับทิศของตัวคูณ (alternating direction method of multipliers) หรือ ADMM [20] เป็นวิธีที่ใช้หลักการจากวิธีการลากรางเจียนแต่งเติม (augmented Lagrangian) ที่มาประยุกต์กับรูปแบบปัญหาที่สมมูลกับ (21) ที่มีการเพิ่มตัวแปรเสริมและเพิ่มสมการเชิงเส้นในเงื่อนไขบังคับ สำหรับปัญหา 1-นอร์มสามารถประยุกต์วิธีนี้ได้โดยตรง และมีข้อดีคือ ขั้นตอนวิธีที่ได้ จะสามารถคำนวณแบบกระจายได้ (distributed programming) อัตราการลู่เข้าของวิธีนี้ในทางปฏิบัตินั้นเป็นที่น่าพอใจ กล่าวคือ สำหรับหลายๆ ปัญหา มีการลู่เข้าที่เร็วเทียบเท่า หรือเร็วกว่าวิธีอันดับหนึ่งแบบเหมาะสมที่สุด

สำหรับขั้นตอนวิธีสามวิธีหลังข้างต้น สามารถเขียนอธิบายได้ในเชิงรวม ด้วยขั้นตอนวิธีแบบพรอกซิมอล (proximal algorithms) กล่าวคือ ขั้นตอนปรับตัวแปรจะเกี่ยวข้องกับตัวดำเนินการพรอกซิมอล (proximal operator) โดยที่ตัวดำเนินการนี้ สำหรับปัญหา 1-นอร์ม จะเกี่ยวข้องกับตัวดำเนินการหดตัว (shrinkage operator) ซึ่งจะมีรูปแบบปิด (closed-form) ที่คำนวณได้อย่างมีประสิทธิภาพ สำหรับรายละเอียดวิธีนี้สามารถค้นได้ใน [21]

อนึ่งเราขอกล่าวว่า ยังมีขั้นตอนวิธีอีกมากมายสำหรับปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบาง ที่ไม่ได้นำมาบรรยายในบทความนี้ ตัวอย่างเช่น วิธี Orthogonal matching pursuit (OMP) [73], วิธี Least angle regression (LAR) [74], วิธีฮอโมโทปี และสามารถดูการเปรียบเทียบของหลายๆ วิธีได้ใน [75] หรือสามารถใช้คำสั่งสำเร็จรูปได้จาก [19]

5. ตัวอย่างการประมาณแบบจำลองเบาบาง

ในหัวข้อนี้จะยกตัวอย่างการประมาณแบบจำลอง FIR ใน (4) ดังนี้ สมมติให้ระบบจริงที่ต้องการประมาณเป็นระบบเชิงเส้นที่อธิบายด้วยปริภูมิสถานะ

$$\begin{aligned}x(t+1) &= \begin{bmatrix} 0.5 & -1 \\ 0 & 0.9 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t) \\ y(t) &= [1 \quad 1] x(t)\end{aligned}$$

จะได้ว่าผลตอบกลับพัลส์ของระบบจริงคือ

$$h(0) = 0, \quad \text{เมื่อ } k = 0, \quad h(k) = 3.5 \cdot 0.5^{k-1} - 1.5 \cdot 0.9^{k-1}, \quad \text{สำหรับ } k \geq 1 \quad (22)$$

แบบจำลอง FIR ใน (4) นั้น สามารถเขียนได้ใหม่ ดังนี้

$$y(t) = \sum_{k=0}^M u(t-k)h(k) + e(t) \quad (23)$$

เพื่อให้่ายต่อการทำให้ $h(k)$ เป็นตัวแปรที่ต้องการประมาณ ในที่นี้ e เป็นสัญญาณรบกวนเข้ามาที่สัญญาณออก เมื่อกำหนดให้สัญญาณออกที่วัดได้คือ $y(0), y(1), \dots, y(T)$ และให้ตัวแปรที่ต้องการประมาณคือ

$$h(0), h(1), \dots, h(M)$$

ดังนั้น เมื่อเขียน (23) ในรูปสมการเชิงเส้นจะได้ว่า

$$b = Ax + e \quad (24)$$

โดยที่

$$b = \begin{bmatrix} y(M) \\ y(M+1) \\ \vdots \\ y(T) \end{bmatrix}, \quad A = \begin{bmatrix} u(M) & u(M-1) & \cdots & u(0) \\ u(M+1) & u(M) & \cdots & u(1) \\ \vdots & \vdots & \ddots & \vdots \\ u(T) & u(T-1) & \cdots & u(T-M) \end{bmatrix}, \quad x = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(M) \end{bmatrix}$$

หากเรามีจำนวนสัญญาณออกที่วัดได้มากพอ จะได้ว่าระบบสมการ (24) เป็นระบบสมการแบบจำนวนสมการมากกว่าจำนวนตัวแปร (overdetermined equations) ดังนั้น เมื่อใช้วิธีกำลังสองต่ำสุดในการประมาณจะได้ว่า ปัญหาที่ได้คือ

$$\underset{x}{\text{minimize}} \quad \|b - Ax\|_2$$

เทคนิคกำลังสองต่ำสุดเป็นวิธีพื้นฐานที่ง่าย อย่างไรก็ตาม การใช้วิธีนี้ เราจำเป็นต้องรู้ค่า M (effective length of the FIR model) เสียก่อน กล่าวโดยง่ายคือ M เป็นค่าจำนวนเต็มต่ำที่สุดที่ $h(M)$ มีค่าใกล้ศูนย์ การเลือก M ที่น้อยเกินไปก็อาจจะทำให้ผลการประมาณไม่ดี (เพราะแบบจำลองจะง่ายเกินไป) แต่ถ้าหากเลือก M มากเกินไปก็อาจจะทำให้เกิดผลเสียในการประมาณ เนื่องจากเมทริกซ์ A ซึ่งมีขนาดเป็น $(T - M + 1) \times (M + 1)$ กลายเป็นเมทริกซ์

อ้วน (fat matrix) ดังนั้นระบบ (24) จะกลายเป็นระบบแบบจำนวนสมการน้อยกว่าจำนวนตัวแปร (underdetermined system) ซึ่งหมายความว่าถ้าจะมีผลเฉลย x หลายค่าที่ทำให้ (24) เป็นจริง

การใช้แบบจำลองแบบบางสำหรับ (23) คือ การมีสมมติฐานว่า $h(k)$ มีศูนย์เป็นจำนวนมาก สำหรับบาง k ซึ่งก็สอดคล้องกับลักษณะสมบัติของผลตอบอิมพัลส์ของระบบจริง เนื่องจากสำหรับระบบเสถียรแล้วนั้น $h(k) \rightarrow 0$ เมื่อ k มีค่ามากพอ (ดังแสดงให้เห็นใน (22) เช่นกัน) ดังนั้น เราจึงเพิ่มฟังก์ชันลงโทษในปัญหาการประมาณ ด้วยการใช้ 1-นอร์มของเวกเตอร์ผลตอบอิมพัลส์ ปัญหาการประมาณแบบจำลอง FIR จึงจัดรูปแบบได้เป็น

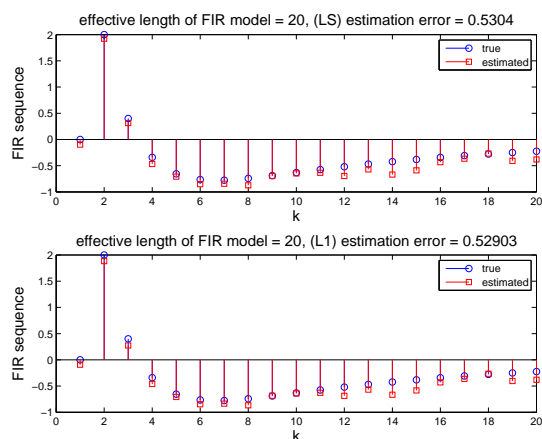
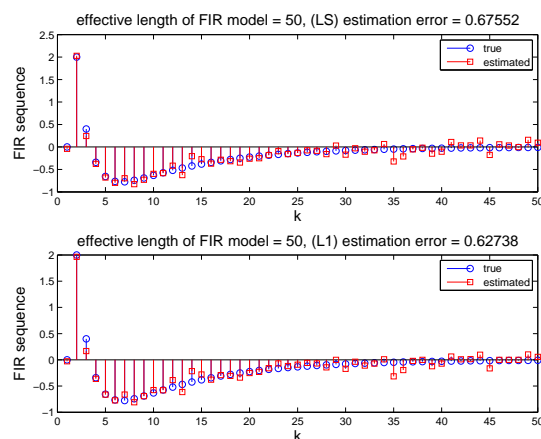
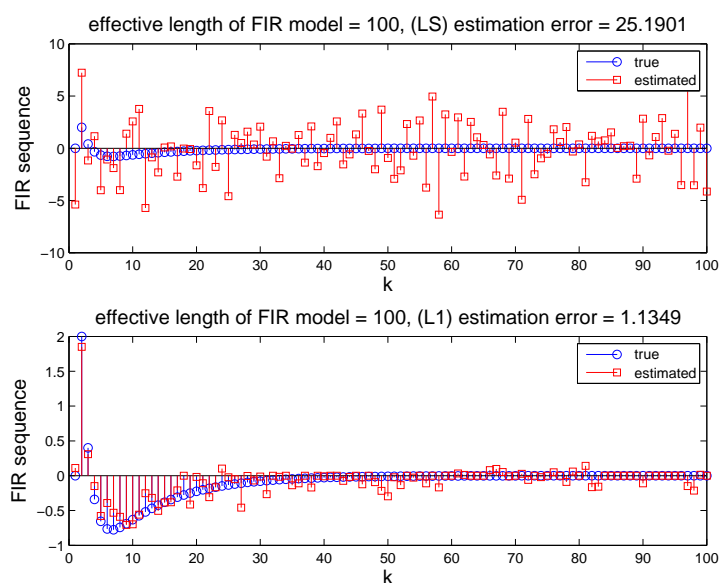
$$\underset{x}{\text{minimize}} (1/2)\|b - Ax\|_2^2 + \lambda\|x\|_1 \quad (25)$$

เราแก้หาผลตอบเชิงเลขของ (25) ด้วยการใช้ตัวแก้ (solver) สำเร็จรูป และผ่านทางโปรแกรม CVX ซึ่งเป็นตัวจัดรูปแบบปัญหาคอนเวกซ์บนภาษา MATLAB [76, 77]

ผลการทดลอง. ในการทดลองเราได้สร้างสัญญาณออกจากระบบจริงจากการกระตุ้นด้วยสัญญาณเข้าที่เป็น PRBS (Pseudo random binary sequence) ตั้งแต่ $t = 1$ ถึง $t = 200$ และสัญญาณออกที่ได้นั้นจะถูกกรองด้วยสัญญาณรบกวนแบบเกาส์เซียนที่มีค่าความแปรปรวนเท่ากับ 0.5 เราได้ปรับอันดับของแบบจำลอง FIR เป็น 3 ค่าด้วยกันคือ $M = 20, 50$ และ 100 การแก้ปัญหา (25) เราใช้ค่า $\lambda = 1$ (สำหรับการเลือกค่าพารามิเตอร์ลงโทษนี้ เป็นอีกหัวข้อหนึ่งที่สำคัญ สามารถศึกษาได้จาก [9]) จากผลการทดลองในรูปที่ 3 จะเห็นว่าเมื่อ M มีค่าน้อยเกินไป ผลตอบอิมพัลส์ที่ประมาณได้ยังไม่เข้าสู่ศูนย์ แต่ค่าความคลาดเคลื่อนก็อยู่ในเกณฑ์ที่พอยอมรับได้ เมื่อ $M = 50$ เราเริ่มสังเกตเห็นว่าผลตอบอิมพัลส์ $h(k)$ ที่สอดคล้องกับ k มากๆ เริ่มมีค่าเข้าใกล้ศูนย์ และผลการประมาณจากทั้งสองวิธีก็ได้ใกล้เคียงกัน เมื่อเราเริ่มให้ M สูงขึ้น นั่นคือ $M = 100$ จะได้ว่าเมทริกซ์ A ในรูปแบบปัญหาซึ่งแต่เดิมเป็นเมทริกซ์ผอม (เมื่อ M น้อย) จะกลายเป็นเมทริกซ์เกือบจัตุรัส (กล่าวคือ เราเริ่มมีจำนวนข้อมูลที่ใช้ในการประมาณน้อยลง เทียบกับจำนวนตัวแปรที่ต้องการประมาณ) ดังนั้น จะเห็นว่าผลการประมาณด้วยวิธีกำลังสองต่ำสุด (LS) มีค่าความคลาดเคลื่อนในการประมาณสูงขึ้นอย่างเห็นได้ชัด และค่าประมาณของ $h(k)$ เมื่อ k มากๆ มีค่าไม่เป็นศูนย์ อันเกิดจากปัญหา overfitting ซึ่งตัวแบบจำลองพยายามจะไปปรับตัวเองให้ใกล้กับสัญญาณรบกวนรบกวนมากเกินไป เมื่อเทียบกับผลการประมาณจากวิธี 1-นอร์ม จะเห็นว่า ค่าประมาณของ $h(k)$ มีค่าศูนย์เป็นจำนวนมาก ที่ k มากๆ ซึ่งสอดคล้องกับลักษณะผลตอบอิมพัลส์จริงในทางทฤษฎี กล่าวโดยสรุปว่า สำหรับตัวอย่างนี้ได้แสดงให้เห็นว่า การเลือกใช้ค่าอันดับของแบบจำลอง FIR ที่สูง (เนื่องจากเราไม่ทราบค่าใดจึงจะสูงพอ) บวกกับการใช้เทคนิคการประมาณแบบ 1-นอร์ม ซึ่งทำให้แบบจำลองที่ได้เป็นแบบจำลองแบบบาง จะทำให้ผลการประมาณโดยรวมดีขึ้นจากการใช้เทคนิคพื้นฐานเช่น วิธีกำลังสองต่ำสุด

6. บทสรุป

ในบทความนี้ได้นำเสนอเบื้องต้น ถึงปัญหาการหาเอกลักษณ์ของระบบซับซ้อน หรือที่มีจำนวนตัวแปรมาก ปัญหาที่น่าสนใจคือเราต้องการใช้แบบจำลองที่ง่าย หรือมีพารามิเตอร์อยู่ในแบบจำลองมีค่าเป็นศูนย์อยู่จำนวนมาก เราเรียกแบบจำลองชนิดนี้ว่า แบบจำลองแบบบาง (sparse models) เราได้นำเสนอนิยามของแบบจำลองแบบบางในงานวิจัยที่ผ่านมา ซึ่งต่างกันไปตามแต่ลักษณะหรือชนิดของแบบจำลอง นิยามความแบบบางระบบนั้น สามารถอธิบายทางคณิตศาสตร์ได้ในรูปของเงื่อนไขสมการของตัวแปรพารามิเตอร์ ซึ่งอาจจะเป็นสมการเชิงเส้น หรือไม่เชิงเส้นก็ได้แล้วแต่ ดังนั้น ปัญหาการ

(a) $M = 20$ (b) $M = 50$ (c) $M = 100$

รูปที่ 3 ผลการประมาณแบบจำลอง FIR ซึ่งเทียบผลตอบหิมพัลส์ที่ประมาณได้ (สีเหลี่ยมสีแดง) กับ ผลตอบหิมพัลส์ทางทฤษฎี (วงกลมสีน้ำเงิน) ซึ่งคำนวณจาก (22) เมื่อเปลี่ยนค่า M (อันดับของแบบจำลอง FIR) จาก 20 ถึง 100 สำหรับกราฟด้านครึ่งบนเป็นผลจากการใช้เทคนิคกำลังสองต่ำสุด และกราฟด้านครึ่งล่างเป็นผลจากการใช้เทคนิคการประมาณแบบ 1-นอร์ม

ประมาณแบบจำลอง ซึ่งก็คือปัญหาการหาค่าพารามิเตอร์ของแบบจำลอง จึงต้องคำนึงถึงเงื่อนไขความเบาบางของระบบดังกล่าวด้วย ในการนี้ เทคนิคเชิงฮิวริสติกที่นิยมใช้ คือการเพิ่มฟังก์ชันลงโทษในฟังก์ชันจุดประสงค์ของปัญหาประมาณด้วยเทอมที่เกี่ยวข้องกับ 1-นอร์มของตัวแปรพารามิเตอร์ที่ต้องการประมาณ เนื่องจากคุณสมบัติของ 1-นอร์ม ที่จะทำให้ผลเฉลยของปัญหาการประมาณมีค่าศูนย์เป็นจำนวนมาก ปัญหาการประมาณที่มักจะกลายเป็นปัญหาการหาค่าเหมาะสมที่สุดที่มีผลตอบมีศูนย์จำนวนมากนั้น ได้ถูกเรียกว่า ปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบาง (sparse optimization)

จากนั้น บทความนี้ได้บรรยายถึงรูปแบบปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบาง ที่เกี่ยวข้องกับการหาเอกลักษณ์ของระบบ รูปแบบปัญหาเหล่านี้แบ่งได้ออกเป็นสองกลุ่มใหญ่ๆ คือ รูปแบบปัญหาที่ใช้ในการประมาณระบบเบาบางโดยตรง

กับรูปแบบปัญหาที่ใช้ความเบาบางของผลเฉลยของปัญหา เพื่อตอบสนองของจุดประสงค์ในการหาเอกลักษณ์ ตัวอย่างของรูปแบบปัญหาหลัง คือ การหาแบบจำลองอันดับต่ำ และการหาแบบจำลองของระบบเชิงเส้นแปรผันตามเวลา

เนื่องจากปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบางนั้น จัดอยู่ในรูปปัญหาไม่เรียบ (nonsmooth problems) และเรามักจะสนใจปัญหาที่มีขนาดใหญ่ เพราะว่าตัวแปรพารามิเตอร์นั้นมาจากแบบจำลองสำหรับระบบซับซ้อน ในส่วนสุดท้ายของบทความ จึงได้ยกตัวอย่างของขั้นตอนวิธีเชิงเลขสำหรับแก้ปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบางที่เป็นปัญหาคอนเวกซ์ ขั้นตอนวิธีส่วนใหญ่ มีลักษณะร่วมกันคือ ขั้นตอนการปรับตัวแปรจะใช้เพียงแค่เกรเดียนต์ของเทอมในฟังก์ชันจุดประสงค์ ที่หาอนุพันธ์ได้เท่านั้น (เราเรียกวิธีเหล่านี้ว่า วิธีอันดับหนึ่ง หรือ first-order methods) การใช้เพียงแค่เกรเดียนต์ มีเหตุผลจากการต้องการหลีกเลี่ยงการคำนวณเมทริกซ์ฮessianอันดับสอง เพราะจะต้องอาศัยทรัพยากรสูงในแต่ละขั้นตอนการปรับ (iteration) ขั้นตอนวิธีที่นำเสนอในบทความนี้ อยู่ในกลุ่มของวิธีแบบพร็อกซิมอล (proximal methods) ที่มีอัตราการลู่เข้าที่น่าพอใจในทางปฏิบัติ และเหมาะกับการคำนวณแบบกระจาย ได้อย่างมีประสิทธิภาพ

กล่าวโดยสรุป เราพบว่า รูปแบบปัญหาการประมาณแบบเบาบางนั้น มีงานประยุกต์ในการหาเอกลักษณ์ของระบบอย่างมาก เนื่องจาก หลายๆ ระบบที่สำคัญในปัจจุบัน เช่น ระบบทางชีวภาพ มักจะเกี่ยวข้องกับตัวแปรจำนวนมาก จึงมีแนวโน้มที่ต้องการหาแบบจำลองอย่างง่ายสำหรับระบบซับซ้อนเช่นนี้ นอกจากนี้ ในด้านงานวิจัยอื่นๆ เช่น การประมวลสัญญาณ (signal processing) การเรียนรู้ด้วยเครื่อง (machine learning) การเรียนรู้เชิงสถิติ (statistical learning) ก็พบรูปแบบปัญหาการหาค่าเหมาะสมที่สุดแบบเบาบาง เป็นอย่างมาก ดังนั้น งานวิจัยที่เป็นที่น่าสนใจในปัจจุบันนี้คือการหารูปแบบปัญหาที่คำนึงถึงความเบาบางของตัวแปรที่ต้องการประมาณ ซึ่งความเบาบางของตัวแปรก็มีความหมายแตกต่างกันไปตามงานประยุกต์ที่สนใจ หรืองานวิจัยอีกด้าน คือ การพัฒนาขั้นตอนวิธีเชิงเลข ที่จะนำมาแก้ปัญหาการประมาณดังกล่าว ที่ต้องเป็นขั้นตอนวิธีที่มีคุณสมบัติการลู่เข้าที่เร็ว และต้องรองรับถึงขนาดปัญหาใหญ่เป็นสำคัญด้วย

บรรณานุกรม

- [1] A. Lozano, N. Abe, Y. Liu, and S. Rosset, "Grouped graphical granger modeling for gene expression regulatory networks discovery," *Bioinformatics*, vol. 25, pp. 110–118, 2009.
- [2] K. E. Stephan and A. Roebroeck, "A short history of causal modeling of fMRI data," *NeuroImage*, 2012.
- [3] S. Ryali, K. Supekar, T. Chen, and V. Menon, "Multivariate dynamical systems models for estimating causal interactions in fMRI," *Neuroimage*, vol. 54, no. 2, pp. 807–823, 2011.
- [4] P. Valdés-Sosa, J. Bornot-Sánchez, M. Vega-Hernández, L. Melie-García, A. Lage-Castellanos, and E. Canales-Rodríguez, "Granger causality on spatial manifolds: Applications to neuroimaging," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. Wiley, 2006.
- [5] A. Pongrattarakul, P. Lerdkultanon, and J. Songsiri, "Sparse system identification for discovering brain connectivity from fMRI time series," in *Proceedings of SICE Annual Conference*, 2013, pp. 949–954.

- [6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [7] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the Lasso," *Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed. Springer, 2009.
- [10] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [11] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [12] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [13] E. J. Candès and J. K. Romberg, *ℓ_1 -magic: Recovery of Sparse Signals via Convex Programming*, Caltech, 2005.
- [14] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [15] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.
- [16] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 936–952, 2011.
- [17] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming Series A*, vol. 103, pp. 127–152, 2005.

- [18] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," *SIAM Journal on Optimization*, 2008, submitted for publication.
- [19] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009. [Online]. Available: <http://www.public.asu.edu/~jye02/Software/SLEP>
- [20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [21] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014. [Online]. Available: <http://dx.doi.org/10.1561/24000000003>
- [22] A. Aswani, P. Bickel, and C. Tomlin, "Statistics for sparse, high-dimensional, and nonparametric system identification," in *IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 2133–2138.
- [23] G. Weber, O. Defterli, S. Gök, and E. Kropat, "Modeling, inference and optimization of regulatory networks based on time series data," *European Journal of Operational Research*, vol. 211, no. 1, pp. 1–14, 2011.
- [24] A. Julius, M. Zavlanos, S. Boyd, and G. Pappas, "Genetic network identification using convex programming," *IET Systems Biology*, vol. 3, no. 3, pp. 155–166, 2009.
- [25] M. Zavlanos, A. A. Julius, S. Boyd, and G. Pappas, "Inferring stable genetic networks from steady-state data," *Automatica*, vol. 47, no. 6, pp. 1113–1122, 2011.
- [26] W. Shi, Q. Ling, and G. Wu, "Sparsity-enhanced linear time-invariant MIMO system identification," in *Control and Decision Conference (CCDC), 2011 Chinese*. IEEE, 2011, pp. 2026–2029.
- [27] A. Chiuso and G. Pillonetto, "A Bayesian approach to sparse dynamic network identification," *Automatica*, vol. 48, no. 8, pp. 1553–1565, 2012.
- [28] Q. Ling, W. Shi, G. Wu, and Z. Tian, "Sparse FIR estimation of low-order systems," in *Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2012, pp. 321–324.
- [29] T. Baran, D. Wei, and A. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1605–1617, 2010.
- [30] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

- [31] S. Schuler, C. Ebenbauer, and F. Allgöwer, " ℓ_0 -system gain and ℓ_1 -optimal control," in *Proceedings of the 18th IFAC World Congress*, 2011, pp. 9230–9235.
- [32] B. Sanandaji, T. Vincent, M. Wakin, R. Tóth, and K. Poolla, "Compressive system identification of LTI and LTV ARX models," in *Proceeding of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. IEEE, 2011, pp. 791–798.
- [33] R. Tóth, H. Hjalmarsson, and C. Rojas, "Sparse estimation of rational dynamical models," in *16th IFAC Symposium on System Identification*. IFAC, 2012, pp. 983–988.
- [34] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds. Cambridge University Press, 2009.
- [35] E. Avventi, A. Lindquist, and B. Wahlberg, "ARMA identification of graphical models," *IEEE Transactions on Automatic Control*, vol. 58, no. 5, pp. 1167–1178, 2013.
- [36] R. Dahlhaus, "Graphical interaction models for multivariate time series," *Metrika*, vol. 51, no. 2, pp. 157–172, 2000.
- [37] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2005.
- [38] A. Shojaie and G. Michailidis, "Discovering graphical Granger causality using the truncating lasso penalty," *Bioinformatics*, vol. 26, no. 18, pp. 517–523, 2010.
- [39] G. Chen, D. R. Glen, Z. S. Saad, J. P. Hamilton, M. E. Thomason, I. H. Gotlib, and R. W. Cox, "Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis," *Computers in biology and medicine*, pp. 1142–1155, 2011.
- [40] S. Haufe, G. Nolte, and N. Kräemer, "Sparse causal discovery in multivariate time series," *Proceedings of JMLR Workshop and Conference*, vol. 6, pp. 97–106, 2008.
- [41] W. Tang, S. Bressler, C. M. Sylvester, S. L. Gordon, and M. Corbetta, "Measuring Granger causality between cortical regions from voxelwise fmri bold signals with lasso," *PLoS Computational Biology*, vol. 8, no. 5, p. e1002513, 2012.
- [42] Y. Liu and M. T. Bahadori, "A survey on Granger causality: A computational view," University of Southern California, Tech. Rep., 2012.

- [43] D. Materassi, G. Innocenti, and L. Giarré, "Reduced complexity models in the identification of dynamical networks: links with sparsification problems," in *Proceedings of the 48th IEEE Conference on Decision and Control*. IEEE, 2009, pp. 4796–4801.
- [44] A. Zymnis, S. Boyd, and D. Gorinevsky, "Mixed linear system estimation and identification," *Signal Processing*, vol. 90, no. 3, pp. 966–971, 2010.
- [45] V. Kekatos, D. Angelosante, and G. Giannakis, "Sparsity-aware estimation of nonlinear volterra kernels," in *The 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2009, pp. 129–132.
- [46] V. Kekatos and G. Giannakis, "Sparse volterra and polynomial regression models: Recoverability and estimation," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5907–5920, 2011.
- [47] K. Shi and P. Shi, "Adaptive sparse volterra system identification with ℓ_0 -norm penalty," *Signal Processing*, vol. 91, no. 10, pp. 2432–2436, 2011.
- [48] N. Hsu, H. Hung, and Y. Chang, "Subset selection for vector autoregressive processes using lasso," *Computational Statistics and Data Analysis*, vol. 52, no. 7, pp. 3645–3657, 2008.
- [49] 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)*, 2013, pp. 3198–3202.
- [50] C. Rojas, B. Wahlberg, and H. Hjalmarsson, "A sparse estimation technique for general model structures," in *European Control Conference (ECC)*, 2013, pp. 2410–2414.
- [51] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, Available: www.stanford.edu/~boyd/cvxbook.
- [52] J. Songsiri and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," *Journal of Machine Learning Research*, vol. 11, pp. 2671–2705, 2010.
- [53] O. Banerjee, A. d'Aspremont, and L. El Ghaoui, "Sparse covariance selection via robust maximum likelihood estimation," ArXiv cs.CE/0506023, July 2005.
- [54] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference*, vol. 6. IEEE, 2001, pp. 4734–4739.

- [55] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proceedings of the American Control Conference*, 2003, pp. 2156–2162.
- [56] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.
- [57] —, "Semidefinite programming methods for system realization and identification," in *Proceedings of the 48th IEEE Conference on Decision and Control*. IEEE, 2009, pp. 4676–4681.
- [58] L. Vandenberghe, "Convex optimization techniques in system identification," in *Proceeding of IFAC Symposium on System Identification*, 2012, pp. 71–76.
- [59] M. Fazel, T. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.
- [60] H. Hjalmarsson, J. Welsh, and C. Rojas, "Identification of box-jenkins models using structured ARX models and nuclear norm relaxation," in *16th IFAC Symposium on System Identification*. IFAC, 2012, pp. 322–327.
- [61] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107–1111, 2010.
- [62] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.
- [63] D. Piga and R. Tóth, "An SDP approach for ℓ_0 -minimization: Application to ARX model segmentation," *Automatica*, pp. 1–8, 2013.
- [64] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, no. 4, pp. 668–677, 2011.
- [65] Y. Nesterov, *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, 2004.
- [66] L. Vandenberghe, *Optimization Methods for Large-Scale Systems*, University of California, Los Angeles, 2012. [Online]. Available: <http://www.seas.ucla.edu/~vandenbe/ee236c.html>

- [67] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1 -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [68] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, pp. 1235–1256, 2009.
- [69] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [70] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds. Cambridge University Press, 2009.
- [71] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, pp. 141–148, 1988.
- [72] A. d'Aspremont, O. Banerjee, and L. El Ghaoui, "First-order methods for sparse covariance selection," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 1, pp. 56–66, 2008.
- [73] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *The 27th Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 40–44.
- [74] R. Tibshirani. The lasso page. [Online]. Available: <http://statweb.stanford.edu/~tibs/lasso.html>
- [75] A. Yang, Z. Zhou, A. Balasubramanian, S. Sastry, and Y. Ma, "Fast ℓ_1 -minimization algorithms for robust face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3234–3246, 2013.
- [76] CVX Research, Inc. (2012, Aug.) CVX: Matlab software for disciplined convex programming, version 2.0. [Online]. Available: <http://cvxr.com/cvx>
- [77] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, Available: http://stanford.edu/~boyd/graph_dcp.html.