

รายงานฉบับสมบูรณ์

การเรียนรู้แบบจำลองเชิงกราฟสำหรับกลุ่มข้อมูลอนุกรมเวลา: งานประยุกต์กับข้อมูล fMRI

Learning multiple graphical models for time series: Application to fMRI time series

รหัสโครงการ RES-57-354-21-062

จิตโกมุท สงศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

รายงานฉบับสมบูรณ์

การเรียนรู้แบบจำลองเชิงกราฟสำหรับกลุ่มข้อมูลอนุกรมเวลา: งานประยุกต์กับข้อมูล fMRI

Learning multiple graphical models for time series: Application to fMRI time series

รายนามคณะผู้วิจัย

จิตโกมุท ส่งศิริ

หน่วยงาน: ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

วันเริ่มต้นโครงการ: 1 กันยายน 2557

วันสิ้นสุดโครงการ: 31 ธันวาคม 2559

กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับการสนับสนุนจาก
ศูนย์ประสานงานนักเรียนทุนรัฐบาลทางด้านวิทยาศาสตร์และเทคโนโลยี
กระทรวงวิทยาศาสตร์และเทคโนโลยี
และสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

ปี พ.ศ. ที่ได้รับทุน: 2557

บทคัดย่อ

งานวิจัยนี้พิจารณาปัญหาการเรียนรู้แบบจำลองเชิงกราฟของข้อมูลอนุกรมเวลาหลายอนุกรมเวลา ที่แต่ละอนุกรมเวลานั้นมาจากระบบที่ต่างกันไป สำหรับปัญหาการหาแบบจำลองเชิงกราฟสำหรับข้อมูลอนุกรมเวลาเดียวนั้น สามารถแสดงให้เห็นได้ว่า อยู่ในรูปแบบปัญหาแลชโซเชิงกราฟโดยที่มีตัวแปรเป็นพารามิเตอร์ในแบบจำลองเคลื่อนที่อัตโนมัติโดยตนเอง ในงานวิจัยนี้ เราจึงพัฒนาปัญหาดังกล่าวไปสู่การประมาณหลายแบบจำลองเชิงกราฟโดยใช้รูปแบบปัญหาแลชโซเชิงกราฟเชื่อมโยง ที่มีหลักการคือ แบบจำลองเชิงกราฟหลายแบบจำลองนั้นควรจะมีรูปแบบของเส้นเชื่อมบางส่วนที่เหมือนกัน และในขณะเดียวกันบางส่วนของเส้นเชื่อมก็ควรจะต่างกัน รูปแบบปัญหาการประมาณที่ได้นั้นจะอยู่ในรูปของปัญหาการหาค่าสูงสุดของฟังก์ชันล็อกเป็นไปได้อีกกับฟังก์ชันลงโทษแบบนอร์ม-1 ซึ่งเป็นปัญหาคอนเวกซ์ ปัญหาที่นำเสนอขึ้นนั้นสามารถนำไปประยุกต์กับการหาเครือข่ายเชื่อมโยงการทำงานของสมองได้ โดยเราต้องการหาเครือข่ายดังกล่าวสำหรับกลุ่มคนไข้สองกลุ่ม อันได้แก่ คนไข้ปกติ และคนไข้ที่มีโรคที่เกี่ยวกับการผิดปกติทางสมอง เรามีสมมติฐานว่าเครือข่ายการทำงานของสมองของสองกลุ่มคนไข้ควรมีบางส่วนที่เหมือนกัน และก็มีบางส่วนที่ต่างกันอันเกิดจากสภาวะผิดปกติในสมอง ในงานวิจัยนี้จะพัฒนาขั้นตอนวิธีที่มีประสิทธิภาพสำหรับการแก้ปัญหาค่าประมาณแบบจำลองเชิงกราฟขนาดใหญ่ และแสดงให้เห็นถึงประสิทธิภาพของวิธีที่นำเสนอขึ้นผ่านทางทดลองบนข้อมูลที่สังเคราะห์ขึ้นและข้อมูลเอฟเอ็มอาร์ไอ ที่ได้จากการทดลองจริง

คำสำคัญ: แบบจำลองเชิงกราฟ ข้อมูลอนุกรมเวลา ปัญหาแลชโซเชิงกราฟ เอฟเอ็มอาร์ไอ

Abstract

This work considers a problem of learning multiple graphical models for a group of time series, each of which belongs to a certain class. The problem of estimating a single graphical model of time series can be formulated as a graphical lasso problem whose variables are coefficient matrices in AR or ARMA processes. In this work, we propose a fused graphical lasso framework for estimating multiple graphical models that share a pattern of partial nonzero edges, while also allows other nonzero edges of each graphical model to be distinct. The estimation formulation takes the form of maximizing a loglikelihood function with an ℓ_1 -type penalty function, which is a convex problem. The proposed method can be applied to a modeling of brain networks using fMRI data, where we wish to estimate brain function connectivities for several groups of patients, *e.g.*, normal patients and patients with brain disorder. The two brain networks are expected to share some common characteristics but also partially distinct due to patient's conditions. We will develop an efficient algorithm for solving the estimation problem in large scale and demonstrate the efficiency and effectiveness of the proposed method on randomly generated and fMRI data sets.

Keywords: graphical models, time series, graphical lasso, fMRI

สารบัญ

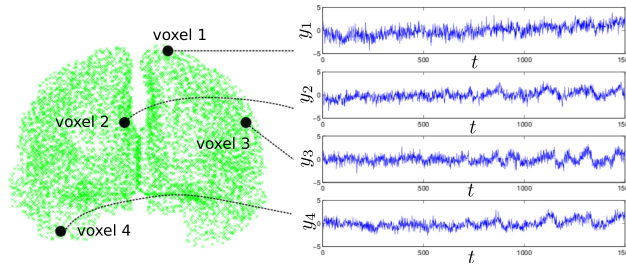
1	บทนำ	8
2	วัตถุประสงค์ของโครงการ	11
3	การออกแบบการวิจัย	11
4	ขอบเขตของการวิจัย	12
5	ระเบียบวิธีการวิจัย	12
5.1	รูปแบบปัญหาที่นำเสนอ	12
5.2	การจัดให้อยู่ในรูปแบบปัญหา Group fused lasso	13
5.3	ขั้นตอนวิธีเชิงเลข	15
6	ผลการวิจัย	17
7	สรุปผล	26
8	ปัญหาในการทำวิจัย และการแก้ไขปัญหา	26

สารบัญรูป

1	ตัวอย่างของสัญญาณ fMRI ที่วัดจาก 4 จุดในสมอง ซึ่งแทนด้วยตัวแปร y_1, y_2, y_3, y_4 . ในกรณีทั่วไป เราจะสมมติให้ m เป็นจำนวน voxels ของสัญญาณที่วัดได้ และพิจารณา $y(t) = (y_1(t), \dots, y_m(t))$ ว่าเป็นอนุกรมเวลาหลายตัวแปร	8
2	ตัวอย่างแบบจำลองเชิงกราฟเพื่ออธิบายโครงสร้างรูปแบบการเชื่อมโยงของสมอง [PLS13]	9
3	พารามิเตอร์ในแบบจำลอง AR ที่ถูกประมาณจากการแก้ปัญหา (6) เมื่อ λ_2 หรือ λ_1 มีค่าใหญ่สัมพัทธ์เมื่อเทียบกับอีกค่าหนึ่ง สัญลักษณ์สีเหลี่ยมสีแดงคือค่าในแนวทแยงมุมของเมทริกซ์ AR และสีเหลี่ยมสีน้ำเงิน คือค่าที่ไม่ได้อยู่ในแนวทแยงมุมของเมทริกซ์ AR	14
4	แสดงค่า True Positive Rate (TPR) กับค่า False Positive Rate (FPR) โดยที่ TPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณได้ถูกต้อง และ FPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณผิด สัญลักษณ์สีเหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL วงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ และสัญลักษณ์กากบาทสีแดงคือผลการแบบจำลอง CGL	19
5	แสดงค่าความคลาดเคลื่อนของแบบจำลองเทียบกับจำนวนสมาชิกที่ไม่เป็นศูนย์ของพารามิเตอร์ AR สัญลักษณ์สีเหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ สัญลักษณ์กากบาทสีแดงคือผลการแบบจำลอง CGL	19
6	รูปแบบศูนย์ของพารามิเตอร์ AR ที่ประมาณได้จากแบบจำลอง Group Fused Lasso (แถวกลาง) และ Group Lasso (แถวล่าง) เมื่อเปรียบเทียบกับรูปแบบศูนย์ที่แท้จริง (แถวบนสุด)	20
7	แสดงค่า True Positive Rate (TPR) กับค่า False Positive Rate (FPR) โดยที่ TPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณได้ถูกต้อง และ FPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณผิด สัญลักษณ์สีเหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL สัญลักษณ์กากบาทสีแดงคือผลจากการใช้แบบจำลอง CGL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ	21
8	แสดงค่าความคลาดเคลื่อนของแบบจำลองเทียบกับจำนวนสมาชิกที่ไม่เป็นศูนย์ของพารามิเตอร์ AR สัญลักษณ์สีเหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL สัญลักษณ์กากบาทสีแดงคือผลจากการใช้แบบจำลอง CGL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL และผลถูกแสดงที่ค่า λ_2 ต่างๆ	21
9	ค่าความคลาดเคลื่อนของผลตอบและค่าฟังก์ชันต้นทุนของปัญหา (6) เทียบกับดัชนีการวนซ้ำของวิธี ADMM	22
10	เมทริกซ์แสดงค่าความเข้มของพารามิเตอร์ในแบบจำลอง AR ทั้งสามชนิด (ซ้าย) แบบจำลองสมองบุคคลพักนิ่ง (กลาง) แบบจำลองเมื่อบุคคลเห็นภาพและประโยคที่สอดคล้องกัน (ขวา) แบบจำลองเมื่อบุคคลเห็นภาพและประโยคที่ไม่สอดคล้องกัน จุดสีน้ำเงินคือจุดที่มีร่วมกันในทั้งสามแบบจำลอง และจุดสีแดงคือจุดที่ต่างกันสามแบบจำลอง	24
11	รูปแบบการเชื่อมโยงการทำงานของสมองด้วย GC causality ของข้อมูลสมอง 3 ชนิด	25

1 บทนำ

โครงการวิจัยนี้จะศึกษา การหาแบบจำลอง หรือการหาเอกลักษณ์ของระบบ (system identification) สำหรับระบบพลวัตการทำงานของสมอง โดยใช้ข้อมูลที่เกี่ยวข้องกับสัญญาณทางสมองของมนุษย์เช่น ข้อมูลอนุกรมเวลาที่วัดจากเทคนิค fMRI (functional magnetic resonance imaging) หรือที่เรียกกันว่า fMRI time series เป็นที่ทราบกันดีว่า ระดับความเข้มจากภาพสแกน fMRI ณ จุดหนึ่งๆ นั้นจะแปรผันกับระดับความเข้มข้นของออกซิเจนในเลือดที่อยู่ในบริเวณนั้นของสมอง ดังนั้น สัญญาณ fMRI ที่นำมาศึกษานั้นจึงเป็นตัวบ่งชี้ว่า ณ เวลาหนึ่งๆ สมองมีการถูกกระตุ้นและมีกิจกรรม (functional activity) เกิดขึ้นในบริเวณต่างๆ มีการกระจายตัวมาน้อยเพียงใด จากตัวอย่างในรูปแบบที่ 1 แบบจำลองพลวัตของการทำงานของสมอง (ซึ่งแทนด้วย



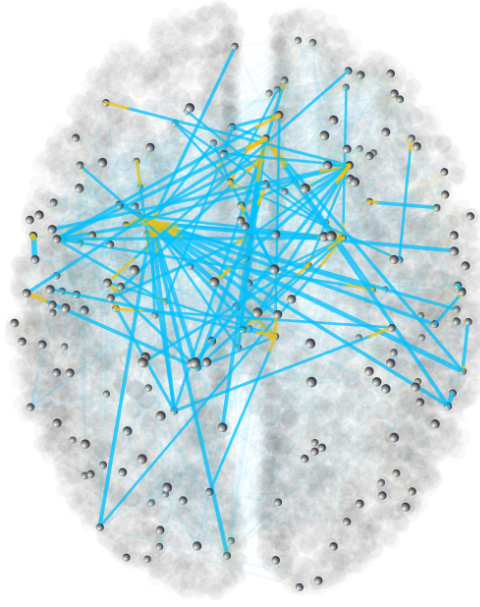
รูปที่ 1: ตัวอย่างของสัญญาณ fMRI ที่วัดจาก 4 จุดในสมอง ซึ่งแทนด้วยตัวแปร y_1, y_2, y_3, y_4 . ในกรณีทั่วไป เราจะสมมติให้ m เป็นจำนวน voxels ของสัญญาณที่วัดได้ และพิจารณา $y(t) = (y_1(t), \dots, y_m(t))$ ว่าเป็นอนุกรมเวลาหลายตัวแปร

ตัวแปร $y(t)$ จึงมีประโยชน์โดยตรงสำหรับผู้เชี่ยวชาญทางด้านสมอง (neuroscientists) เพื่อที่จะได้เข้าใจระบบการทำงานที่ซับซ้อน และนำไปเชื่อมโยงกับการศึกษาและวิเคราะห์บุคคลภายใต้สถานะที่ต่างๆ กันได้อีกด้วย เช่น การแยกแยะคนไข้ที่มีสถานะไม่ปกติ เช่น ซึมเศร้า จิตเภท (schizophrenia) อัลไซเมอร์ หรือโรคพาร์กินสัน ออกจากคนไข้ที่มีสุขภาพปกติ โดยสมมติฐานที่สำคัญคือ คนที่เป็นโรคดังกล่าวนั้นจะมีสถานะของสมองที่ผิดปกติไป [CRT+09, HLS+10, RPSTMM13, dSCS+11]

ตัวอย่างแบบจำลองที่ใช้บรรยายพลวัตของกิจกรรมการทำงานในสมอง เช่น แบบจำลองสมการอนุพันธ์ไม่เชิงเส้น (nonlinear differential equations) ซึ่งรู้จักกันในนามว่าแบบจำลองเหตุและผลแบบพลวัต (dynamic causal modeling-DCM) แบบจำลองปริภูมิสถานะ [RSCM11] แบบจำลองอัตถะถดถอยหลายตัวแปร (multivariate autoregressive models) หรือแบบจำลองสมการโครงสร้าง (structural equation modeling หรือ SEM)[GCR11, CGS+11] เป็นตัวอย่างแบบจำลองพลวัตที่ได้รับความสนใจในปัจจุบันนี้ ซึ่งแบบจำลองที่กล่าวมานั้น สามารถดูได้จากบทความสรุปงานวิจัยที่ผ่านมาใน [SR12]

ภาพสแกนสมองด้วยวิธี fMRI นั้น จะมีความละเอียดเชิงพื้นที่ (spatial resolution) ค่อนข้างสูง (ขึ้นกับค่าความเข้มของสนามแม่เหล็กในเครื่องวัด fMRI) เมื่อเทียบกับความละเอียดเชิงเวลา (temporal resolution) กล่าวคือ ภาพสแกนหนึ่งๆ จะมีข้อมูลอยู่เป็นหลัก 10,000-100,100 voxels (หน่วยจุดๆ หนึ่งๆ ในปริมาตรสมอง) และภาพสแกนเหล่านี้จะเก็บได้ประมาณหลักร้อยวินาทีในทางเวลา นั้นหมายความว่า จำนวนตัวแปรในแบบจำลองที่ต้องการหา (ซึ่งจะแปรผันตรงกับ m) มีมากเมื่อเทียบกับจำนวนชุดข้อมูลที่ใช้ในการประมาณ สิ่งนี้จะก่อให้เกิดปัญหาในการขั้นตอนการประมาณแบบจำลอง ที่เรียกว่าปัญหา overfitting กล่าวคือเมื่อประมาณด้วยแบบจำลองซับซ้อนแล้ว แบบจำลองพยายามจะเรียนรู้พลวัตของสัญญาณรบกวนที่อยู่ในข้อมูลที่วัดได้ แทนที่จะเรียนรู้พลวัตของระบบที่แท้จริง ด้วยเหตุนี้ แบบจำลองที่เราสนใจจะมีโครงสร้างที่ไม่ซับซ้อน กล่าวคือ เรามุ่งหาแบบจำลองที่มีจำนวนพารามิเตอร์ในแบบจำลองมีไม่มากนัก (parsimonious models) แต่ยังสามารถอธิบายพลวัตของระบบสมองได้ดีระดับหนึ่ง การใช้แนวทางดังกล่าว มีข้อดีคือ 1) ในขั้นตอนการหาพารามิเตอร์ของแบบจำลองจะสามารถหลีกเลี่ยงปัญหา overfitting ได้ และ 2) โครงสร้างของแบบจำลองที่ไม่ซับซ้อนนั้น จะช่วยแสดงถึงความสัมพันธ์ที่เด่นชัดของกลุ่มตัวแปรในระบบ ได้ดีกว่าการ

ใช้แบบจำลองที่ซับซ้อน ข้อดีหลัง จะเป็นจุดสำคัญที่งานวิจัยในปัจจุบันได้ให้ความสนใจ ยกตัวอย่าง เช่น การสนใจว่า กิจกรรมในสมองส่วน A จะมีความสัมพันธ์กับกิจกรรมในสมองส่วน B อย่างไร ซึ่งหัวข้อนี้ เรียกว่าการหาแบบจำลองเหตุและผล (causal modeling) หรือ การหาเครือข่ายเชื่อมโยงของสมอง (brain connectivity network) ซึ่งจะเป็นจุดประสงค์หลักของโครงการวิจัยนี้



รูปที่ 2: ตัวอย่างแบบจำลองเชิงกราฟเพื่ออธิบายโครงสร้างรูปแบบการเชื่อมโยงของสมอง [PLS13]

เครือข่ายเชื่อมโยงของสมอง คือการอธิบายปฏิสัมพันธ์ของสมองในแต่ละส่วนผ่านจากข้อมูล fMRI และแสดงผลสัมพันธ์กันนั้นในรูปแบบแบบจำลองเชิงกราฟ (graphical model) ในกราฟหนึ่งๆ ตามรูป 2 นั้นจะประกอบไปด้วยเซตของโหนดซึ่งแทนตัวแปรในระบบ และเซตของเส้นเชื่อม (edges) ซึ่งเส้นเชื่อมระหว่างโหนด i และโหนด j จะอธิบายว่าตัวแปร i และตัวแปร j มีความสัมพันธ์กัน สำหรับเครือข่ายสมองนั้น เรามักจะแทนโหนดหนึ่งๆ ด้วยส่วนของสมองที่สนใจ (regions of interest หรือ ROI) และสำหรับการที่แต่ละส่วนของสมองมีการทำงานสัมพันธ์กันหรือไม่นั้น มีการใช้นิยามทางคณิตศาสตร์และทางสถิติที่แตกต่างกันไป ในงานวิจัยนี้เราจะเลือกการใช้เหตุและผลแบบเกรนเจอร์ (Granger causality) โดยมีหลักการคือ ถ้าหากสัญญาณจากตัวแปร i มีผลต่อสัญญาณจากตัวแปร j แล้วนั้น การทำนายตัวแปร j ที่ใช้ตัวแปร i ด้วยนั้นย่อมทำให้ผลการทำนายมีความแปรปรวนที่ต่ำลง [Lüt05] ในการอธิบายความสัมพันธ์ของสมองด้วยนิยามดังกล่าวนี้ สามารถดูบทความงานด้านการหาเครือข่ายเชื่อมโยงของสมอง ได้จากบทความวิจัยสรุปเช่น [RHH⁺10, SMSK⁺11, VSRDF11, Smi12] และสำหรับงานที่ได้พิจารณาข้อมูล fMRI ว่าเป็นสัญญาณทางเวลาด้วย และได้ใช้หลักการเช่นเหตุและผลแบบเกรนเจอร์ [VSBSVH⁺06, DLJ⁺09, TBS⁺12] กล่าวคือ ถ้าหากเราใช้แบบจำลองอัตโนมัติโดยตนเองหลายตัวแปร อันดับ p (multivariate autoregressive models):

$$y(t) = A_1y(t-1) + A_2y(t-2) + \dots + A_p y(t-p) + u(t) \quad (1)$$

ที่มีเมทริกซ์สัมประสิทธิ์เป็น A_1, A_2, \dots, A_p และ $A_k \in \mathbb{R}^{n \times n}$ แล้วนั้น จะได้ว่า สมองส่วน j ไม่ได้ส่งผลกระทบต่อส่วน i ถ้าหาก

$$(A_k)_{ij} = 0 \quad (2)$$

สำหรับ $k = 1, 2, \dots, p$ ดังนั้น จะเห็นได้ว่า ปัญหาการหาแบบจำลองพลวัตของการทำงานของสมอง เพื่อที่จะนำไปอธิบายความสัมพันธ์ของสมองแต่ละส่วนในรูปแบบเครือข่ายเชื่อมโยงนั้น จะเกี่ยวข้องกับเงื่อนไขความเป็นศูนย์ในฟังก์ชันของพารามิเตอร์ในแบบจำลอง ดังนั้น ในปัญหาการประมาณแบบจำลองซึ่งมักจะอยู่ในรูปปัญหาการหาค่าเหมาะสมที่สุด (optimization problems) นั้น จึงอยู่ในรูปปัญหาที่มีผลตอบมีค่าเป็นศูนย์อยู่

จำนวนมาก ปัญหาในประเภทนี้ เราเรียกว่า การหาเอกลักษณ์ของระบบแบบเบาบาง (sparse system identification) หรือปัญหาการหาค่าเหมาะสุดแบบเบาบาง (sparse optimization) สำหรับตัวอย่างปัญหา งานวิจัยเหล่านี้สามารถดูได้จากตัวอย่างใน [Son13a, PLS13, Son13b]

ในงาน [Son13a] ได้มีการนำเสนอปัญหาการประมาณแบบจำลอง AR ที่มีเมทริกซ์สัมประสิทธิ์มีค่า ศูนย์เป็นจำนวนมาก และอ่านรูปแบบศูนย์ของเมทริกซ์เหล่านั้น เพื่อนำไปสรุปเป็นโครงสร้างความสัมพันธ์แบบเกรนเจอร์ของกลุ่มตัวแปรในอนุกรมเวลา (Granger graphical model of time series) รูปแบบปัญหาดังกล่าวคือ ปัญหาการหาค่าต่ำสุดกำลังสอง (linear least squares) บวกกับเทอมที่ทำให้เป็นปกติ (regularization) เพื่อบังคับให้ผลตอบนั้นมีค่าเป็นศูนย์เป็นจำนวนมาก ปัญหาที่กล่าวนี้สามารถบรรยายได้เป็น

$$\min_A \frac{1}{2} \|Y - AH\|_2^2 + \lambda \sum_{i \neq j} \|[(A_1)_{ij} (A_2)_{ij} \cdots (A_p)_{ij}]\|_2 \quad (3)$$

ที่มีตัวแปรเป็น $A = (A_1, \dots, A_p)$, and $A_k \in \mathbf{R}^{q \times q}$, $k = 1, \dots, p$ เมทริกซ์ Y และ H ประกอบไปด้วยข้อมูล $y(t)$ ที่เวลาต่างๆ กัน ปัญหาดังกล่าวเป็นกรณีเฉพาะของปัญหาในรูปแบบทั่วไป ที่มีรูปแบบเป็น

$$\text{minimize}_x f(x) + \lambda \|x\|_1 \quad (4)$$

โดยที่ $x \in \mathbf{R}^n$ จะเป็นตัวแปรที่ต้องการหา และบ่งบอกถึงพารามิเตอร์ของแบบจำลอง f จะเป็นฟังก์ชันที่บ่งชี้ความคลาดเคลื่อนระหว่างสัญญาณออกจากแบบจำลองกับข้อมูลจริง และเทอม $\|x\|_1$ จะเป็นฟังก์ชันลงโทษที่จะทำให้ผลตอบของปัญหามีค่าเป็นศูนย์เป็นจำนวนมาก ซึ่งรูปแบบของศูนย์ใน x จะสะท้อนถึงโครงสร้างความสัมพันธ์ระหว่างตัวแปรในระบบของแบบจำลองที่ทำได้ ส่วน λ ที่เป็นค่าสเกลาร์บวก จะเป็นค่าที่เราเลือก เพื่อจะชั่งน้ำหนักว่า จะให้ความสำคัญกับความคลาดเคลื่อนของแบบจำลอง หรือลักษณะของผลตอบที่อยากให้เป็นศูนย์ มากกว่ากัน ในปัญหารูปทั่วไป (4) ที่เราจะเรียกว่าปัญหาการหาค่าเหมาะสุดแบบเบาบางนั้น (sparse optimization) มีงานวิจัยในด้านนี้พบได้ใน ตัวอย่างเช่น [HNK08, LALR09, ABVN11, RM11, Son13a, PLS13, SV10, ALW13] การแก้ปัญหารูป (4) เป็นที่รู้จักกันในอีกนามว่า ปัญหาการหาค่าต่ำสุดแบบ ℓ_1 หรือปัญหาแลซโซ (lasso) เมื่อ f เป็นฟังก์ชันกำลังสอง [HTF09, S3] ซึ่งความยากของปัญหาคือการใช้เทคนิคการหาค่าต่ำสุดเพื่อแก้ปัญหารูป (4) ที่มีขนาดใหญ่ ดังที่พบในงานประยุกต์ด้านข้อมูลทางชีวภาพ เช่น ข้อมูล fMRI เนื่องจาก x จะมีอันดับเป็นตั้งแต่หลักหมื่นไปจนถึงหลักล้านเลยก็ว่าได้ สำหรับการเลือกรูปแบบปัญหาเหล่านี้ ในการหาโครงสร้างความสัมพันธ์ของการทำงานของสมองจากข้อมูล fMRI นั้น พบได้ใน [VSBVH+06, GCR11, NA11, PLS13] ที่เรียนรู้โครงสร้างความสัมพันธ์แบบเกรนเจอร์ดังปัญหา (7) โดยมี f เป็นฟังก์ชันกำลังสอง (quadratic loss) และ x แทนเมทริกซ์สัมประสิทธิ์ในแบบจำลอง AR หรือ [Son13a]

การหาโครงสร้างความสัมพันธ์ของการทำงานของสมองนั้น สามารถทำได้จากข้อมูล fMRI ที่เก็บมาจากรสภาวะบุคคลต่างๆ กัน เช่น การให้คนไข้นอนนิ่งๆ และหลับตา (resting state) หรือการกระตุ้นให้คนไข้มองภาพหลายๆ อย่าง เช่น ตัวหนังสือ รูปภาพสิ่งของ (task-related state) ภายใต้สภาวะที่ต่างกันนี้ ย่อมมีสมมติฐานว่า การทำงานของสมองย่อมต่างกัน ดังนั้นเครือข่ายเชื่อมโยงของสมองย่อมออกมาเป็นรูปแบบเครือข่ายที่ต่างกันด้วย เครือข่ายเชื่อมโยงของสมองที่ได้จากคนไข้ในสภาวะนี้จะเรียกว่าเป็น เครือข่ายโหมดปริยาย (default mode network) [HP10, GSMD09, GKRM03] ซึ่งมักจะถูกนำมาใช้เปรียบเทียบกับเครือข่ายเชื่อมโยงของสมองที่ได้ภายใต้สภาวะที่สมองโดนกระตุ้นว่าจะมีรูปแบบเครือข่ายเปลี่ยนไปอย่างไรบ้าง ด้วยเหตุนี้ ปัญหาการหาเครือข่ายเชื่อมโยงของสมอง จึงมีผลประโยชน์อีกแง่หนึ่ง คือ เราสามารถใช้รูปแบบเครือข่ายที่ประมาณได้มาเป็นตัวแยกแยะ (classifier) สภาวะของคนไข้ได้ เช่น การแยกแยะคนไข้ปกติ กับคนไข้ที่มีสภาวะผิดปกติทางสมอง [CRT+09, HLS+10, RPSTMM13, dSCS+11] ¹

¹อย่างไรก็ตาม ต้องขอล่าไว้ในที่นี้ว่า ถ้าหากพิจารณาการแยกแยะสภาวะคนไข้เป็นปัญหาหลักแล้ว ยังมีเทคนิคทางด้านการแยกแยะ (classification) โดยตรงที่ถูกใช้ในด้านการเรียนรู้ด้วยเครื่อง (machine learning) อีกเป็นจำนวนมาก ซึ่งเทคนิคพวกนี้อาจจะไม่ต้องอาศัยแบบจำลองพลวัตของสัญญาณ fMRI เช่น เทคนิค support vector machine (SVM), logistic regression, Gaussian Naive Bayes และอื่นๆ ซึ่งสามารถอ่านทบทวนได้ใน [PMB09]

หากสมมติว่ามีสถานะสมองของบุคคล 2 สถานะที่ต่างกันให้เป็นกลุ่ม A และ B ตามลำดับ การหาเครือข่ายเชื่อมโยงสมอง *หนึ่งเครือข่าย* จากข้อมูลสมองของทั้งสองกลุ่มรวมกัน นั้นย่อมไม่ใช่แนวทางที่ดี เนื่องจากสถานะการทำงานของสมองจากบุคคลสองคนนั้นควรจะต่างกัน ในทางกลับกัน หากเราหาเครือข่ายเชื่อมโยงสมองที่ประมาณจากกลุ่ม A และหาอีกหนึ่งเครือข่ายที่ประมาณจากกลุ่ม B แยกจากกัน วิธีนี้ก็อาจจะยังไม่ใช่วิธีดีนัก เนื่องจากการหาเครือข่ายเชื่อมโยงแยกกันในลักษณะนี้ ไม่ได้ใช้ข้อสมมติฐานที่ว่า ในสถานะสมองของบุคคลสองกลุ่มก็ควรจะมี *บางส่วนของ* เครือข่ายเชื่อมโยงที่ตรงกัน (common underlying network) ทั้งสองสถานะ ดังนั้นเราจึงต้องการพิจารณาปัญหาการประมาณเครือข่ายเชื่อมโยง 2 เครือข่ายจากบุคคลทั้งสองสถานะไปพร้อมๆ กัน โดยมีเงื่อนไขว่า เครือข่ายทั้งสองนั้นจะมีรูปแบบบางส่วนของที่สอดคล้องกัน (กล่าวคือ มีเส้นเชื่อมจากสองโหนดบางคู่ที่เหมือนกัน) แต่ในขณะเดียวกัน ก็มีบางส่วนของเครือข่ายที่ต่างกันด้วย

ในโครงการวิจัยนี้ เราจะพิจารณาใช้หลักการ *โครงสร้างเชิงสัมพันธ์แบบเกรนเจอร์* ที่นิยามจากรูปแบบศูนย์ในเมทริกซ์ AR ดัง (2) ในการสร้างเครือข่ายเชื่อมโยงของสมอง 2 แบบจำลองเชิงกราฟพร้อมกันที่ได้ จะอธิบายว่าข้อมูลอนุกรมเวลาจากคนไข้สองสถานะนั้น จะมีรูปแบบกราฟที่เหมือนกันในบางส่วน และต่างกันในส่วนอย่างไร โดยทั้งนี้ สิ่งที่เป็นตัวแปรของปัญหานั้น จะเป็นพารามิเตอร์ของแบบจำลองพลวัต (และในที่นี้ เราจะพิจารณาแบบจำลองเชิงเส้น ที่มีสัญญาณเข้าเป็นกระบวนการสุ่ม) ทั้งนี้ ขอกล่าวว่า ประโยชน์ของโครงการวิจัยนี้ไม่ได้จำกัดที่งานประยุกต์บนข้อมูล fMRI เท่านั้น หากแต่รวมถึงข้อมูลอนุกรมเวลาจากระบบอื่นๆ ได้อีกด้วย เช่น ข้อมูลทางการเงิน ข้อมูลยีนส์ ข้อมูล EEG หรือข้อมูลทางชีวภาพอื่นๆ เป็นต้น งานวิจัยนี้จะเน้นไปทางด้านการจัดรูปแบบปัญหาการหาค่าต่ำสุด (optimization formulation) และการประยุกต์และพัฒนาขั้นตอนวิธี (algorithm) เพื่อใช้ในการแก้ปัญหาดังกล่าวที่มีขนาดใหญ่ สำหรับรายละเอียดทางเทคนิคนั้น จะกล่าวต่อไปในส่วนที่ 5

2 วัตถุประสงค์ของโครงการ

1. พัฒนาแบบจำลองทางคณิตศาสตร์เชิงเส้นสำหรับข้อมูลกลุ่มอนุกรมเวลา ที่อธิบายโครงสร้างความสัมพันธ์ของตัวแปรได้ และมีโครงสร้างความสัมพันธ์ของแต่ละกลุ่มที่เหมือนกันและต่างกันบางส่วน โดยจะเน้นงานวิจัยไปที่ระบบการทำงานของสมองที่วิเคราะห์ผ่านทางข้อมูล fMRI จากบุคคลภายใต้สถานะที่ต่างกันหลายสถานะ
2. ประยุกต์และพัฒนาเทคนิคเชิงเลข (numerical technique) ที่มาใช้ในการแก้ปัญหาการประมาณแบบจำลองพลวัตของสมอง ที่สามารถรองรับขนาดของปัญหาระดับใหญ่ได้

3 การออกแบบการวิจัย

การออกแบบวิจัยเป็นไปตามขั้นตอนดังนี้

1. ศึกษาถึงรูปแบบปัญหาการประมาณโครงสร้างความสัมพันธ์ที่เกี่ยวข้อง ที่คำนึงถึงความแตกต่างของรูปแบบโครงสร้าง และใช้การประมาณแบบ l_1
2. นำเสนอรูปแบบปัญหาที่จะประยุกต์ใช้กับสมการพลวัตที่อธิบายการทำงานของสมอง
3. ทดลองการใช้รูปแบบปัญหาที่นำเสนอขึ้น กับข้อมูลที่สังเคราะห์ขึ้น เพื่อทดสอบสมมติฐานว่าปัญหาที่เสนอนั้น สามารถนำไปเรียนรู้โครงสร้างความสัมพันธ์ของกลุ่มอนุกรมเวลาได้จริง
4. ศึกษาขั้นตอนเชิงเลขที่สามารถประยุกต์ใช้กับปัญหาที่นำเสนอได้ โดยขั้นตอนเชิงเลขนี้ควรมีประสิทธิภาพกับการแก้ปัญหาขนาดใหญ่
5. ทดลองการใช้รูปแบบปัญหา และการแก้ปัญหาด้วยขั้นตอนเชิงเลขที่มีประสิทธิภาพ กับข้อมูลสมอง fMRI

4 ขอบเขตของการวิจัย

โครงการวิจัยมีขอบเขตดังนี้

1. การพัฒนาแบบจำลองที่อธิบายการทำงานของสมองจะใช้แบบจำลองแบบเชิงเส้นเท่านั้น เนื่องจากจะนำไปสู่ปัญหาการประมาณที่ไม่ซับซ้อนจนเกินไป
2. โครงการวิจัยพิจารณาการประมาณแบบจำลองข้อมูลสมองจาก 2 สภาวะ แต่ไม่ได้ตั้งให้การแยกแยะสภาวะของสมองเป็นจุดประสงค์หลักของงานวิจัย
3. วิธีเชิงเลขที่ใช้ในการแก้ปัญหาการประมาณแบบจำลอง ถึงแม้ว่าจะเหมาะสมกับปัญหาที่มีจำนวนตัวแปรมาก แต่ก็ยังมีข้อจำกัดทางด้านหน่วยความจำ ซึ่งขึ้นอยู่กับทรัพยากรของเครื่องคำนวณ ดังนั้นในโครงการวิจัยนี้ จึงได้แก้ปัญหามาจากข้อมูลสมองที่มีจำนวนตัวแปรลดลงมาให้เหมาะกับระดับทรัพยากรคอมพิวเตอร์ PC ที่มี

5 ระเบียบวิธีการวิจัย

5.1 รูปแบบปัญหาที่นำเสนอ

ปัญหาการประมาณเครือข่ายเชื่อมโยง 2 เครือข่ายจากบุคคลทั้งสองสภาวะไปพร้อมๆ กัน โดยมีเงื่อนไขว่าเครือข่ายทั้งสองนั้นจะมีรูปแบบบางส่วนที่สอดคล้องกัน (กล่าวคือ มีเส้นเชื่อมจากสองโหนดบางคู่ที่เหมือนกัน) แต่ในขณะเดียวกัน ก็มีบางส่วนของเครือข่ายที่ต่างกันด้วย ปัญหาในลักษณะนี้สามารถจัดให้อยู่ในรูปแบบของ *ปัญหาแลชโซเชื่อมกัน* (fused lasso) [TSR+05, YPS+15, MCH+12] ซึ่งมีรูปแบบปัญหาเป็น

$$\underset{x,y}{\text{minimize}} f(x) + f(y) + \lambda_1(\|x\|_1 + \|y\|_1) + \lambda_2\|x - y\|_1 \quad (5)$$

โดยที่ $x \in \mathbb{R}^n$ และ $y \in \mathbb{R}^n$ แทนพารามิเตอร์ในแบบจำลองเชิงกราฟของกลุ่มสภาวะ A และ B ตามลำดับ ซึ่งรูปแบบของศูนย์ใน x และ y จะบอกถึงรูปแบบของเส้นเชื่อมในเครือข่ายของกลุ่ม A และ B ตามลำดับ ค่า λ_1, λ_2 นั้นเป็นค่าสเกลาร์บวกที่กำหนดเอง เราจะเห็นว่าเทอมที่เพิ่มมาใน (5) คือเทอม $\|x - y\|_1$ นั้นจะเป็นการลงโทษให้ x และ y มีรูปแบบศูนย์ที่ตรงกันในบางสมาชิกผ่านการเลือกค่า λ_2 ให้มีค่าใหญ่พอ ส่วนรูปแบบของศูนย์ใน x และ y ที่ตำแหน่งสมาชิกอื่นที่อาจจะไม่เหมือนกันนั้นถูกกำหนดจากเทอมลงโทษ $\|x\|_1$ และ $\|y\|_1$ ผ่านการควบคุมค่า λ_1 นั้นเอง ดังนั้นเมื่อ $\lambda_2 = 0$ ปัญหา (5) จึงลดรูปเป็นปัญหาแลชโซปกติ (4) ที่เป็นปัญหาการประมาณสองปัญหาที่แยกจากกันได้ นั้นเอง สำหรับงานวิจัยในกลุ่มนี้ที่ประยุกต์กับข้อมูล fMRI มิงานใน [YPS+15] ที่พิจารณาปัญหาแลชโซเชิงกราฟเชื่อมกัน (fused multiple graphical lasso) กล่าวคือ เป็นปัญหาการประมาณแบบจำลองเชิงกราฟของตัวแปรสุ่มเกาส์เซียน (ใช้สมมติฐานว่าข้อมูล fMRI เป็นตัวแปรสุ่ม) หรือการประมาณเมทริกซ์ความแปรปรวนร่วม (Σ_1, Σ_2) ของข้อมูลคนไข้ที่เป็นอัลไซเมอร์ และคนไข้ปกติ ตามลำดับ โดยที่ x, y แทนค่าสมาชิกใน $\Sigma_1^{-1}, \Sigma_2^{-1}$ (และบ่งบอกความเป็นอิสระต่อกันแบบมีเงื่อนไข) ดังนั้น รูปแบบของศูนย์ใน x, y จึงแทนรูปแบบทอพอโลยีของแบบจำลองเชิงกราฟจากกลุ่มสองสภาวะ ปัญหาที่ความท้าทายอยู่ที่เทคนิคการหาค่าต่ำสุดมาแก้ เนื่องจากเป็นอันดับของปัญหาเพิ่มขึ้นเพราะเราประมาณแบบจำลองเชิงกราฟสองอันพร้อมๆ กัน

ในโครงการวิจัยนี้ เรานำเสนอรูปแบบปัญหาการประมาณแบบจำลอง AR จำนวน K แบบจำลองพร้อมๆ กัน โดยที่ทั้ง K แบบจำลองนั้น จะมีโครงสร้างความสัมพันธ์แบบเกรนเจอร์บางส่วนร่วมกัน ปัญหาดังกล่าวสามารถบรรยายได้ดังนี้

$$\underset{A^{(1)}, \dots, A^{(K)}}{\text{minimize}} \sum_{k=1}^K \frac{1}{2} \|Y^{(k)} - A^{(k)} H^{(k)}\|_2^2 + \lambda_1 \sum_{i \neq j} \sum_{k=1}^K \|B_{ij}^{(k)}\|_2 + \lambda_2 \sum_{i \neq j} \sum_{k=1}^{K-1} \|B_{ij}^{(k+1)} - B_{ij}^{(k)}\|_2 \quad (6)$$

โดยที่ตัวแปรต่างๆ มีนิยามดังนี้

- ▶ $A^{(k)} = (A_1^{(k)}, \dots, A_p^{(k)})$ คือเมทริกซ์สัมประสิทธิ์ของแบบจำลอง AR ตัวที่ K
- ▶ $B_{ij}^{(k)} = [(A_1^{(k)})_{ij} \quad (A_2^{(k)})_{ij} \quad \dots \quad (A_p^{(k)})_{ij}]^T \in \mathbb{R}^p$
- ▶ $Y^{(k)}$ และ $H^{(k)}$ ประกอบไปด้วยข้อมูลทางเวลาจากอนุกรมเวลากลุ่มที่ k
- ▶ λ_1 and λ_2 เป็นจำนวนจริงค่าบวกซึ่งเป็นพารามิเตอร์ถ่วงน้ำหนักในเทอมที่ทำให้ปกติ (regularization term)

เมื่อพิจารณาฟังก์ชันต้นทุนนั้น เทอมสุดท้ายเป็นผลบวกของ 2-นอร์ม ซึ่งเป็นฟังก์ชันลงโทษแบบ ℓ_1 ชนิดหนึ่ง และเป็นฟังก์ชันของความต่างระหว่างสัมประสิทธิ์ในพารามิเตอร์ AR แนวทางแยงมุมจากแบบจำลอง 2 แบบจำลองที่ติดกัน ถ้ากำหนดให้ λ_2 มีค่ามากระดับหนึ่ง พารามิเตอร์ $(A^{(1)})_{ij}, (A^{(2)})_{ij}, \dots, (A^{(K)})_{ij}$ จะมีค่าที่เท่ากัน และนั่นหมายความว่า โครงสร้างความสัมพันธ์แบบเกรนเจอร์ของกลุ่ม K แบบจำลองนั้นจะมีรูปแบบเหมือนกัน ถ้าหากกำหนดให้ λ_1 มีค่ามากระดับหนึ่ง สมาชิกต่างๆ ใน $(A^{(k)})_{ij}$ (สำหรับบาง k) จะมีค่าเป็นศูนย์ ซึ่งบ่งชี้ว่า เราได้แบบจำลอง AR แบบเบาบาง (sparse)

เมื่อกำหนดให้ $\lambda_2 = 0$ ปัญหา (6) จะลดรูปลงมาเป็นปัญหาการประมาณแบบจำลอง AR ทั้งหมด K แบบจำลองที่ประมาณได้แยกเป็นอิสระต่อกัน

$$\underset{A^{(1)}, \dots, A^{(K)}}{\text{minimize}} \sum_{k=1}^K \frac{1}{2} \|Y^{(k)} - A^{(k)} H^{(k)}\|_2^2 + \lambda_1 \sum_{i \neq j} \sum_{k=1}^K \|B_{ij}^{(k)}\|_2 \quad (7)$$

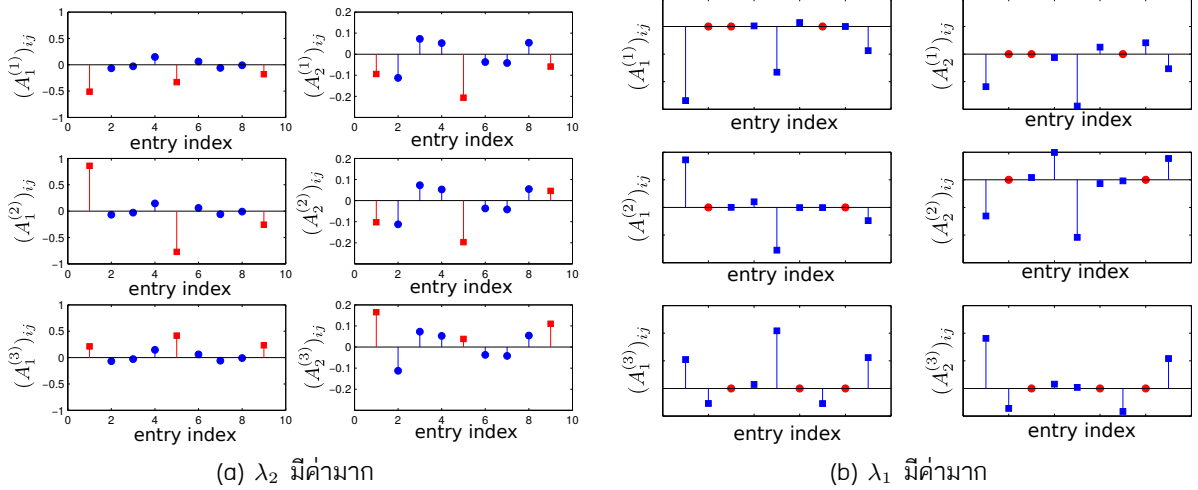
ดังนั้น การใช้รูปแบบปัญหา (6) จึงทำให้เราสามารถได้กลุ่มแบบจำลอง AR แบบเบาบาง (parsimonious model) และในขณะเดียวกัน กลุ่มแบบจำลองเหล่านั้นก็จะมีการสร้างโครงสร้างความสัมพันธ์แบบเกรนเจอร์ที่มีบางส่วนที่เหมือนกันได้

เราสามารถยกตัวอย่างง่ายเพื่ออธิบายการประยุกต์ใช้รูปแบบปัญหา (6) ที่มีค่าพารามิเตอร์ $q = 3, p = 2$ และ $K = 3$ กล่าวคือ เราประมาณแบบจำลอง AR ที่มี 3 ตัวแปรและมี lag order เท่ากับ 2 ในตัวอย่างแรกนั้น เรากำหนดให้ λ_2 มีค่าใหญ่เมื่อเทียบกับ λ_1 ซึ่งหมายความว่า เรามีจุดประสงค์ที่จะประมาณให้แบบจำลองทุกอันมีค่าเมทริกซ์ AR ที่ใกล้เคียงกันทั้ง 3 แบบจำลอง รูป 3 (a) แสดงค่า $(A^{(k)})_{ij}$ จำนวน q^2 สมาชิก จากรูปอธิบายได้ว่าสำหรับแต่ละ A_k (แต่ละคอลัมน์ในรูป) ทั้งสามแบบจำลองมีค่าเมทริกซ์ AR ในแนวทแยงมุมที่เท่าๆกัน (สังเกตจากวงกลมน้ำเงิน) หากเราใช้ค่า λ_2 ที่ใหญ่ระดับหนึ่ง อย่างไรก็ตาม ค่าเมทริกซ์ AR ในแนวทแยงมุมของทั้งสามแบบจำลองสามารถมีค่าต่างกันได้ ในกรณีที่สอง หากค่า λ_1 มีค่าใหญ่พอเมื่อเทียบกับ λ_2 (ซึ่งหมายความว่า เราต้องการหาแบบจำลองเบาบาง หรือเมทริกซ์ AR ควรจะมีศูนย์จำนวนมาก) จากรูป 3 (b) แสดงว่าสำหรับแต่ละ A_k (แต่ละคอลัมน์ในรูป) ทั้งสามแบบจำลองสามารถมีค่าเมทริกซ์ AR ที่ต่างกันได้ หากแต่ว่าในแต่ละแบบจำลอง (แต่ละแถวในรูป) ค่าเมทริกซ์ AR จะมีศูนย์อยู่หลายตำแหน่ง และตำแหน่งที่เป็นศูนย์นั้นจะเกิดขึ้นที่ตำแหน่งเดียวกันในแต่ time-lag coefficient เนื่องจากเราใช้รูปแบบปัญหาที่ทำให้เกิดรูปแบบศูนย์ที่ตรงกันของ $(A_1^{(k)})_{ij}, \dots, (A_p^{(k)})_{ij}$

5.2 การจัดให้อยู่ในรูปแบบปัญหา Group fused lasso

ในส่วนี้เราสามารถแสดงให้เห็นได้ว่า ปัญหา (6) ที่นำเสนอ นั้น สามารถจัดให้อยู่ในรูปแบบปัญหาเวคเตอร์ได้ และทำให้ปัญหานี้้อยู่ในรูปแบบปัญหาที่เรียกว่า *group fused Lasso* [ABD13] โดยการกำหนดตัวแปร

$$x = (B_{11}^{(1)}, \dots, B_{11}^{(K)}, B_{12}^{(1)}, \dots, B_{12}^{(K)}, \dots, B_{qq}^{(1)}, \dots, B_{qq}^{(K)}) \quad (8)$$



รูปที่ 3: พารามิเตอร์ในแบบจำลอง AR ที่ถูกประมาณจากการแก้ปัญหา (6) เมื่อ λ_2 หรือ λ_1 มีค่าใหญ่สัมพันธ์เมื่อเทียบกับอีกค่าหนึ่ง สัญลักษณ์สีเหลี่ยมสีแดงคือค่าในแนวทแยงมุมของเมทริกซ์ AR และสีเหลี่ยมสีน้ำเงิน คือค่าที่ไม่ได้อยู่ในแนวทแยงมุมของเมทริกซ์ AR

กล่าวคือ เรากำหนดให้ x เป็นตัวแปรที่เป็นพารามิเตอร์ในเมทริกซ์ AR ทั้งหมด และ x มีสมาชิกทั้งหมด n ตัว โดยที่ $n = q^2 pK$ เราสังเกตว่าเทอมที่เป็นฟังก์ชันลงโทษในปัญหา (6) จะลงโทษเฉพาะค่าเมทริกซ์ AR ในแนวที่ไม่ใช่ทแยงมุมเท่านั้น ด้วยข้อกำหนดดังกล่าว เราจึงนิยามการแปลง P $P \in \mathbb{R}^{(q^2-q) \times q^2}$ ซึ่งแปลงจากค่าเมทริกซ์ AR ในแนวที่ไม่ใช่ทแยงมุมทุกตัวไปยังเวกเตอร์ กล่าวคือ เราสามารถถอดสมาชิกในแนวที่ไม่ใช่ทแยงมุมของเมทริกซ์ $X \in \mathbb{R}^{q \times q}$ ได้ด้วยการแปลง $P \cdot \text{vec}(X)$ ตัวอย่างเช่น สำหรับ $X = [x_{ij}] \in \mathbb{R}^{2 \times 2}$ จะได้ว่า

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad P \cdot \text{vec}(X) = \begin{bmatrix} x_{21} \\ x_{12} \end{bmatrix}.$$

หากเรากำหนดให้ $\mathcal{P} = P \otimes I_{pK}$ จะได้ว่า $\mathcal{P}x$ จะเลือกเฉพาะสมาชิกในแนวไม่ใช่เส้นทแยงมุม กล่าวคือ

$$\mathcal{P}x = \left(B_{12}^{(1)}, \dots, B_{12}^{(K)}, \dots, B_{1q}^{(1)}, \dots, B_{1q}^{(K)}, \dots, B_{q,q-1}^{(1)}, \dots, B_{q,q-1}^{(K)} \right) \in \mathbb{R}^{(q^2-q)pK}. \quad (9)$$

เราสามารถมอง $\mathcal{P}x$ ออกเป็นท่อนได้ โดยแต่ละท่อนมีขนาดที่ขึ้นกับปัญหาที่เราต้องการพิจารณา ในการนี้เราใช้สัญลักษณ์ $z = (z_1, z_2, \dots, z_L)$ และพิจารณา *sum of 2-norm*: $\|z\|_{2,1} = \sum_{k=1}^L \|z_k\|_2$ โดยที่ขนาดของแต่ละท่อน z_k ขึ้นกับบริบทของปัญหาที่พิจารณาอยู่

- ▶ หากเราแบ่ง $\mathcal{P}x$ ออกเป็นท่อนที่มีขนาดเท่ากับ p จะได้ว่าเทอมที่สองของฟังก์ชันจุดประสงค์ใน (6) สามารถเขียนได้เป็น $\lambda_1 \|\mathcal{P}x\|_{2,1}$.
- ▶ เทอมสุดท้ายของฟังก์ชันจุดประสงค์ใน (6) สามารถเขียนได้เป็น $\lambda_2 \|Dx\|_{2,1}$ โดยที่ $D = P \otimes D$ และ D เป็น forward difference matrix:

$$D = \begin{bmatrix} -I_p & I_p & & & & \\ & -I_p & I_p & & & \\ & & & \ddots & \ddots & \\ & & & & -I_p & I_p \end{bmatrix}.$$

ดังนั้นจากนิยามของ x ที่เรากำหนดใน in (8) จะได้ว่าปัญหา (6) นั้นสมมูลกับ

$$\underset{x}{\text{minimize}} \quad (1/2) \|Gx - b\|_2^2 + \lambda_1 \|\mathcal{P}x\|_{2,1} + \lambda_2 \|Dx\|_{2,1} \quad (10)$$

(เราขอละรายละเอียดการแสดงความสมมูลของสองปัญหา) โดยมีตัวแปรคือ $x \in \mathbb{R}^n$ เมทริกซ์ $G \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\mathcal{P} \in \mathbb{R}^{s \times n}$ and $\mathcal{D} \in \mathbb{R}^{r \times n}$ เป็นพารามิเตอร์ของปัญหา เราอธิบายได้ว่า G และ b ประกอบไปด้วยค่าของข้อมูล $y^{(1)}(t), \dots, y^{(K)}(t)$ for $t = 1, 2, \dots, N$ สำหรับอันดับของ G, b, \mathcal{P} และ \mathcal{D} นั้นจะสัมพันธ์กับอันดับของตัวแปรใน (6) ด้วยความสัมพันธ์ $n = q^2 pK, m = qNK, s = (q^2 - q)pK$ และ $r = (q^2 - q)p(K - 1)$ เราจะเรียกปัญหา (10) ว่าเป็นรูปแบบปัญหา *group fused lasso* และมีกรณีเฉพาะดังนี้

- ▶ เมื่อ $\lambda_2 = 0$ and $\lambda_1 > 0$, ปัญหา (10) จะลดรูปเป็น

$$\underset{x}{\text{minimize}} \quad (1/2)\|Gx - b\|_2^2 + \lambda_1\|\mathcal{P}x\|_{2,1} \quad (11)$$

และเราจะเรียกปัญหานี้ว่า *group lasso* [YL06, ZRY09].

- เมื่อเทอม sum of 2-norm คำนวนจาก $\mathcal{P}x$ ที่แบ่งเป็นท่อนมีมิชขนาด p ปัญหานี้คือการประมาณ K แบบจำลอง (7) อย่างเป็นอิสระต่อกัน และเมื่อเราเพิ่มค่า λ_1 เราจะได้แบบจำลองที่เบาบางขึ้น
- เมื่อเทอม sum of 2-norm คำนวนจาก $\mathcal{P}x$ ที่แบ่งเป็นท่อนมีมิชขนาด pK ปัญหานี้คือการประมาณ K แบบจำลองที่บังคับให้แบบจำลองทุกตัวนั้นมีโครงสร้าง Granger causality ที่เหมือนกัน เราจะเรียกปัญหานี้ว่าเป็นปัญหา *common group lasso*

ดังนั้น จะเห็นว่า การแบ่ง $\mathcal{P}x$ ออกเป็นท่อนที่มีขนาดต่างกัน จะทำให้ค่าในฟังก์ชันลงโทษ $\|\mathcal{P}x\|_{2,1}$ มีค่าต่างกัน ที่นำไปสู่รูปแบบปัญหาที่ต่างกันด้วย

- ▶ หาก $\lambda_1 = 0$ และ $\lambda_2 > 0$ จะได้ว่าปัญหา (10) กลายเป็น

$$\underset{x}{\text{minimize}} \quad (1/2)\|Gx - b\|_2^2 + \lambda_2\|\mathcal{D}x\|_{2,1} \quad (12)$$

ซึ่งอยู่ในรูปแบบปัญหาที่เรียกว่า *total variation regularized problem* [WBAW12] ที่มีงานประยุกต์ เช่น image reconstruction [BT09, OV14] หรือการประมาณแบบจำลองที่มีค่าพารามิเตอร์เปลี่ยนแปลงแบบ piece-wise constant [OLB10] รูปแบบปัญหาดังกล่าวมีการใช้ฟังก์ชันลงโทษบนความต่างของตัวแปรสองตัวที่ติดกัน (เทอม $\|\mathcal{D}x\|_{2,1}$)

- ▶ รูปแบบปัญหา *group fused lasso* เคยมีการพิจารณาใน [ABD13] กับปัญหา image denoising เราพบว่าปัญหาใน [ABD13] นั้นสมมูลกับ (10) เมื่อ $\mathcal{P} = I$ และ $\mathcal{D} = D$, ซึ่งก็คือปัญหาเมื่อ $q = 1$ หรือกล่าวคือการประมาณแบบจำลอง AR แบบสเกลาร์

5.3 ขั้นตอนวิธีเชิงเลข

ในส่วนนี้จะอธิบายถึงขั้นตอนวิธีเชิงเลขในการแก้ปัญหา (10) ที่รองรับจำนวนแปรที่มากได้ ด้วยวิธีแบบคอนเวกซ์ ค่าฟังก์ชันในปัญหามีเทอม $\|\cdot\|_{2,1}$ ซึ่งไม่สามารถหาอนุพันธ์ได้ งานวิจัยนี้จึงมุ่งหาวิธีที่เหมาะสมกับปัญหาแบบ non-smooth ในการนี้ Alternating direction method of multiplier (ADMM) ซึ่งเป็นเทคนิค Douglas-Rachford splitting ที่ประยุกต์กับปัญหาสมมูล [BPC+10, PB14] ได้นำมาพิจารณาเนื่องด้วยข้อดีหลายประการ ตัวอย่างเช่น 1) การคำนวณในแต่ละรอบไม่ซับซ้อนซึ่งเหมาะกับปัญหาที่มีตัวแปรมาก และ 2) ความเร็วการลู่เข้าที่ดีในทางปฏิบัติ

การประยุกต์ ADMM นั้นจำเป็นต้องอาศัยเทคนิคการแบ่งฟังก์ชันจุดประสงค์ออกเป็นหลายเทอมอย่างเหมาะสม เราจึงกำหนดให้

$$f(x) = (1/2)\|Gx - b\|_2^2, \quad g(x) = \lambda_1\|x\|_{2,1}, \quad h(x) = \lambda_2\|x\|_{2,1}$$

จากนั้น เราสามารถจัดรูป (10) ให้อยู่ในรูปปัญหาทั่วไปที่ ADMM จะประยุกต์ได้

$$\begin{aligned} & \text{minimize} && f(x_1) + g(x_2) + h(x_3) \\ & \text{subject to} && \begin{bmatrix} \mathcal{P} \\ \mathcal{D} \end{bmatrix} x_1 = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \end{aligned} \quad (13)$$

ที่มีตัวแปรคือ $x_1 \in \mathbf{R}^n, x_2 \in \mathbf{R}^s$ and $x_3 \in \mathbf{R}^r$ จากนั้น เรากำหนดให้ augmented Lagrangian ซึ่งเป็นฟังก์ชัน Lagrangian บวกกับฟังก์ชันลงโทษแบบกำลังสอง ดังนี้

$$L_\rho = f(x_1) + g(x_2) + h(x_3) + (\rho/2)\|\mathcal{P}x_1 - x_2 + z_1/\rho\|_2^2 + (\rho/2)\|\mathcal{D}x_1 - x_3 + z_2/\rho\|_2^2, \quad (14)$$

โดยที่ $\rho > 0$ เป็นพารามิเตอร์ลงโทษ (penalty parameter) ที่จะส่งผลต่อความเร็วในการลู่เข้า ขั้นตอนนั้นคือการหาค่าต่ำสุดของ augmented Lagrangian บนตัวแปร x_1 และ (x_2, x_3) สลับกันไป โดยอาศัยข้อดีที่ว่าเทอมที่เกี่ยวข้องกับสามตัวแปรนั้นสามารถแยกจากกันได้ (separable functions in three variables) หากเรากำหนดให้ x_i^+ คือค่าตัวแปรในรอบวนถัดไป เราสามารถอธิบายกฎการปรับตัวแปร [BPC⁺10] ของขั้นตอน ADMM ดังนี้

$$x_1^+ = \operatorname{argmin}_{x_1} \frac{1}{2}\|Gx_1 - b\|_2^2 + \frac{\rho}{2}\|\mathcal{P}x_1 - x_2 + z_1/\rho\|_2^2 + \frac{\rho}{2}\|\mathcal{D}x_1 - x_3 + z_2/\rho\|_2^2, \quad (15)$$

$$x_2^+ = \operatorname{argmin}_{x_2} \lambda_1\|x_2\|_{2,1} + \frac{\rho}{2}\|\mathcal{P}x_1^+ - x_2 + z_1/\rho\|_2^2, \quad (16)$$

$$x_3^+ = \operatorname{argmin}_{x_3} \lambda_2\|x_3\|_{2,1} + \frac{\rho}{2}\|\mathcal{D}x_1^+ - x_3 + z_2/\rho\|_2^2, \quad (17)$$

$$z_1^+ = z_1 + \rho(\mathcal{P}x_1^+ - x_2^+), \quad (18)$$

$$z_2^+ = z_2 + \rho(\mathcal{D}x_1^+ - x_3^+). \quad (19)$$

ในรายละเอียดของการปรับค่า x_1 นั้น จะเห็นว่าเราต้องกำหนดให้อนุพันธ์ของฟังก์ชันเป็นศูนย์ (เทียบกับ x_1) เงื่อนไขที่ได้จะเป็นสมการเชิงเส้น

$$[G^T G + \rho(\mathcal{P}^T \mathcal{P} + \mathcal{D}^T \mathcal{D})] x_1^+ = G^T b + \mathcal{P}^T (\rho x_2 - z_1) + \mathcal{D}^T (\rho x_3 - z_2).$$

เนื่องจาก $G^T G + \rho(\mathcal{P}^T \mathcal{P} + \mathcal{D}^T \mathcal{D}) \succ 0$ เราสามารถทำ Cholesky factorization และเก็บ Cholesky factor (L) ไว้ได้ นอกจากนี้ $\mathcal{P}^T \mathcal{P}$ เป็นเมทริกซ์ทแยงมุม และ $\mathcal{D}^T \mathcal{D}$ เป็น banded matrix และ $G^T G$ เป็นเมทริกซ์เบาบาง (sparse matrix) ดังนั้น การคำนวณ Cholesky factorization จึงทำได้อย่างมีประสิทธิภาพ การปรับค่า x_1 จึงเป็นการแก้สมการเชิงเส้น: $L^T v = c, Lx_1 = v$ โดยที่ $c = G^T b + \mathcal{P}^T (\rho x_2 - z_1) + \mathcal{D}^T (\rho x_3 - z_2)$.

การปรับค่าของ x_2 และ x_3 จะมีรูปแบบที่คล้ายกัน ที่สามารถเขียนเป็นรูปปัญหาทั่วไปคือ

$$\operatorname{minimize}_x \gamma\|x\|_{2,1} + (1/2)\|x - u\|_2^2,$$

สำหรับค่า $\gamma > 0$ และ $u \in \mathbf{R}^n$ หากเรากำหนดให้ u และ x สามารถแบ่งได้ออกเป็น L ท่อน ปัญหาข้างต้นนั้นคือ การหา proximal operator of $f_1(x) = \|x\|_{2,1} = \sum_{k=1}^L \|x_k\|_2$ ที่มีรูปคำตอบทั่วไปคือ

$$\begin{aligned} \operatorname{prox}_{\gamma f_1}(u) &= \operatorname{argmin}_x \gamma\|x\|_{2,1} + (1/2)\|x - u\|_2^2, \\ (\operatorname{prox}_{\gamma f_1}(u))_k &= \max \left\{ 1 - \frac{\gamma}{\|u_k\|_2}, 0 \right\} u_k, \end{aligned}$$

สำหรับ $k = 1, 2, \dots, L$ Proximal operator ของฟังก์ชันผลบวก 2-norm นั้นจะเป็นที่รู้จักว่าเป็น *block soft thresholding operator* [PB14] จากการปรับค่าข้างต้นของ 3 ตัวแปร เราจึงสรุปขั้นตอน ADMM ที่ใช้ปัญหา (10) ดังนี้

ADMM สำหรับปัญหาฟิวซ์แลซโซ่ กำหนดค่าเริ่มต้นของ x_1, x_2, x_3, z_1, z_2 และเลือก $\rho > 0$ กำหนดให้ (x, x^+) คือตัวแปรในรอบปัจจุบันและรอบถัดไป เราจะวนซ้ำการปรับค่าต่อไปนี้

$$\begin{aligned} c &= G^T b + \mathcal{P}^T(\rho x_2 - z_1) + \mathcal{D}^T(\rho x_3 - z_2), \\ x_1^+ &= [G^T G + \rho(\mathcal{P}^T \mathcal{P} + \mathcal{D}^T \mathcal{D})]^{-1} c, \\ x_2^+ &= \text{prox}_{(\lambda_1/\rho)f_1}(\mathcal{P}x_1^+ + z_1/\rho), \\ x_3^+ &= \text{prox}_{(\lambda_2/\rho)f_1}(\mathcal{D}x_1^+ + z_2/\rho), \\ z_1^+ &= z_1 + \rho(\mathcal{P}x_1^+ - x_2^+), \\ z_2^+ &= z_2 + \rho(\mathcal{D}x_1^+ - x_3^+), \end{aligned}$$

จนกระทั่งค่า primal residual, r และ dual residual, s นั้นน้อยกว่าค่าที่ยอมรับได้ค่าหนึ่ง

$$\begin{aligned} \|r\|_2 &= \left\| \begin{bmatrix} \mathcal{P}x_1 - x_2 \\ \mathcal{D}x_1 - x_3 \end{bmatrix} \right\|_2 \leq \epsilon^{\text{pri}}, \\ \|s\|_2 &= \rho \left\| \begin{bmatrix} \mathcal{P}^T(x_2^+ - x_2) \\ \mathcal{D}^T(x_3^+ - x_3) \end{bmatrix} \right\|_2 \leq \epsilon^{\text{dual}}. \end{aligned}$$

ค่า ϵ^{pri} และ ϵ^{dual} สามารถคำนวณได้ตาม [BPC+10] เราจะเห็นว่า x_2 จะเป็น sparse vector จาก block soft-thresholding operator ในขณะที่ x_1 จะเข้าใกล้ x_2 แต่อาจจะไม่ sparse

6 ผลการวิจัย

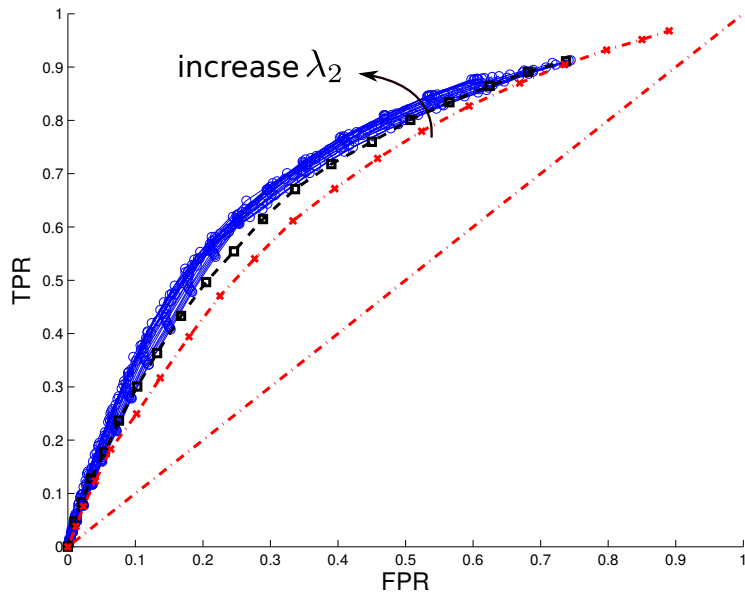
ในส่วนนี้จะแสดงผลการทดลองจากการใช้ 3 วิธีนั้นคือ

- ▶ การใช้รูปแบบปัญหา (10) หรือจะเรียกว่าปัญหา group fused lasso (GFL) ที่แบบจำลองทุกอันจะมีโครงสร้าง Granger causality คล้ายกัน ซึ่งจะสัดส่วนความคล้ายจะควบคุมด้วยพารามิเตอร์ λ_2
- ▶ การใช้รูปแบบปัญหา (11) หรือจะเรียกว่าปัญหา group lasso (GL) ที่แบบจำลองทุกอันจะถูกประมาณอย่างเป็นอิสระต่อกัน (โดยการกำหนดให้ $\mathcal{P}x$ แบ่งเป็นท่อนที่มีขนาด p)
- ▶ การใช้รูปแบบปัญหา (11) หรือจะเรียกว่าปัญหา common group lasso (CGL) ที่บังคับให้แบบจำลองทุกอันมีโครงสร้าง Granger causality เหมือนกัน (โดยการกำหนดให้ $\mathcal{P}x$ แบ่งเป็นท่อนที่มีขนาด pK)

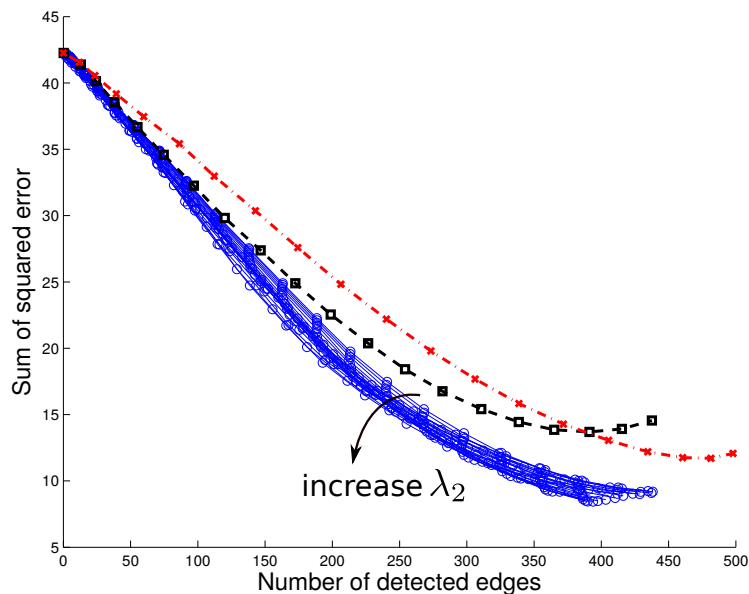
การทดลองที่ 1 คือการทดสอบว่า รูปแบบปัญหา GFL จะสามารถทำให้ประมาณแบบจำลอง AR หลายแบบจำลอง ได้ดีกว่าการใช้รูปแบบปัญหา GL หรือ CGL (7) หรือไม่ ถ้าหากมีสมมติฐานที่ว่าข้อมูลอนุกรมเวลาจากหลายๆ กลุ่มนั้นมีโครงสร้างความสัมพันธ์แบบเกรนเจอร์ที่คล้ายกัน

ในการนี้ เราจึงสร้างแบบจำลอง AR อันดับ 2 ที่มี 10 ตัวแปร จำนวน 3 แบบจำลอง ($q = 10, p = 2, K = 3$) โดยทั้ง 3 แบบจำลองจะมีค่าพารามิเตอร์บางค่าที่เท่ากัน จากแบบจำลองเหล่านี้ เราจะสร้างข้อมูลอนุกรมเวลา AR ขึ้นมาทั้งหมด 100 ชุด และแต่ละชุดมีข้อมูลทางเวลา 30 จุด ซึ่งเป็นค่าที่ต่ำระดับหนึ่งเมื่อเทียบกับจำนวนตัวแปรพารามิเตอร์ที่ต้องประมาณทั้งหมด ทั้ง 3 แบบจำลองนั้นมีจำนวนตำแหน่งของ AR coefficients ที่ไม่เป็นศูนย์ ต่างกันด้วยสัดส่วน 40% จากนั้นค่าของ λ_1 และ λ_2 จะถูกเลือกในสเกลเชิงเส้นเป็นจำนวน M คู่ ในแต่ละค่า (λ_1, λ_2) เราจะแก้ปัญหา (6) และอ่านรูปแบบโครงสร้างความสัมพันธ์แบบเกรนเจอร์ จากรูปแบบศูนย์ของค่าเมทริกซ์สัมประสิทธิ์ AR จาก K แบบจำลองที่ประมาณได้ จากนั้น เรายุบรวมเปรียบเทียบรูปแบบโครงสร้างที่ประมาณได้ กับรูปแบบโครงสร้างที่แท้จริง และคำนวณค่าความผิดพลาดที่เกิดขึ้น ผลการทดลองในส่วนนี้ สามารถสรุปประเด็นสำคัญได้ดังนี้

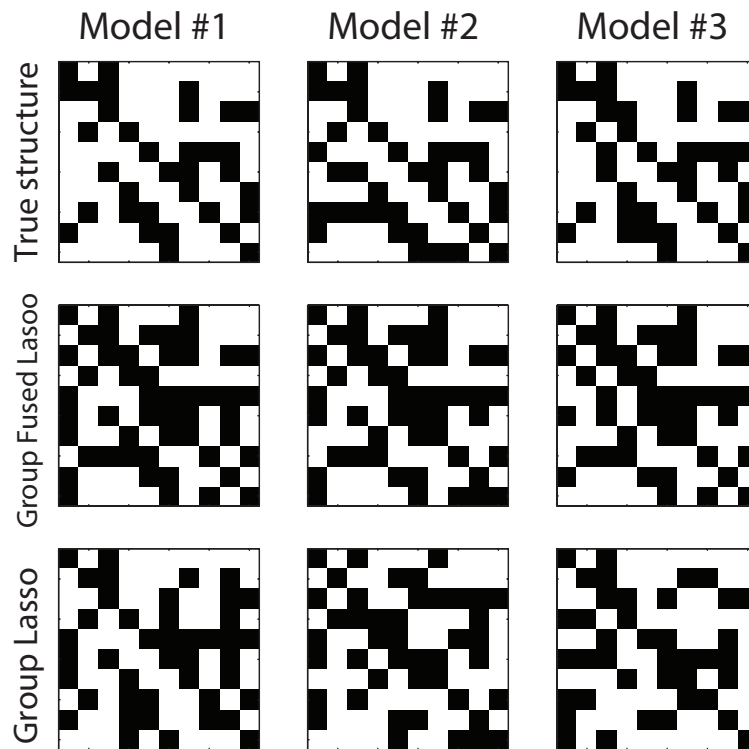
- ▶ รูป 4 แสดงถึงค่าอัตราการทำนายค่าที่ไม่เท่ากับศูนย์ที่ถูกต้อง (True Positive Rate-TPR) เทียบกับอัตราการทำนายค่าที่ไม่เท่ากับศูนย์ที่ผิด (False Positive Rate-FPR) ซึ่งค่าที่แสดงนี้ถูกเฉลี่ยจากการทดลองบนข้อมูล 100 ชุด จากรูปจะเห็นว่าที่ค่า FPR ที่กำหนดค่าหนึ่ง กราฟจากแบบจำลอง GFL (แนวทางที่เสนอขึ้นใหม่) จะอยู่เหนือกราฟที่ได้จากแบบจำลอง GL และ CGL นั้นหมายความว่าแนวทางที่เราเสนอขึ้นนั้น ทำให้การประมาณรูปแบบความสัมพันธ์แบบเกรนเจอร์ทำได้ดีขึ้น โดยการเพิ่ม λ_2 เนื่องจากเราให้ความสำคัญกับเทอมที่บังคับให้ทุก K แบบจำลองมีโครงสร้างที่คล้ายกันมากขึ้น เมื่อให้ $\lambda_2 = 0$ ผลลัพธ์ของการใช้ GFL และคู่เข้าสู่ผลเดียวกับผลของการใช้ GL นอกจากนี้ การเปรียบเทียบผลการใช้ GL และ CGL จะเห็นว่า การใช้ GL ให้ผลที่ดีกว่าเนื่องจากแบบจำลองจริงไม่ได้มีโครงสร้าง Granger causality ที่เหมือนกันหมด ดังนั้นการประมาณแบบที่ไม่ได้บังคับให้โครงสร้าง Granger causality ของทุกแบบจำลองให้เหมือนกัน จึงได้ผลที่ดีกว่า
- ▶ รูป 5 อธิบายถึงข้อดี ข้อเสีย (trade-off) ของการให้ความสำคัญระหว่าง ความแม่นยำของแบบจำลอง กับความซับซ้อนของแบบจำลอง เราจะเห็นว่าความคลาดเคลื่อนของแบบจำลองที่ได้จะเพิ่มมากขึ้น ถ้าหากรูปแบบโครงสร้างเกรนเจอร์ที่ประมาณได้นั้นหนาแน่นมากเกินไป ซึ่งมีสาเหตุหลักมาจากปัญหา overfitting เราจะเห็นว่าแนวทางที่ใช้ GFL นั้นจะให้ค่าความคลาดเคลื่อนแบบจำลองที่ต่ำกว่า เมื่อเทียบกับการใช้แบบจำลอง GL และ CGL เมื่อเพิ่มค่า λ_2
- ▶ เมทริกซ์ในสีดำขาวในรูป 6 แสดงถึงรูปแบบศูนย์จากพารามิเตอร์ AR ที่ประมาณได้ สีขาวแทนตำแหน่งที่เป็นศูนย์และสีดำแทนตำแหน่งที่ไม่เท่ากับศูนย์ เราเลือกรูปแบบศูนย์หนึ่งๆ จากทั้งแนวทางการใช้แบบจำลอง Group Fused Lasso และ Group Lasso ที่สอดคล้องกับค่า FPR พอๆ กันประมาณ 0.21–0.23 เราจะเห็นจากผลการทดลองว่ารูปแบบโครงสร้างเกรนเจอร์ที่ประมาณได้จากแบบจำลอง group fused lasso นั้นจะมีความเหมือนกันทั้ง 3 แบบจำลองมากกว่า การใช้แบบจำลอง Group Lasso (โดยการเลือก λ_2 ให้มีค่ามากพอ) ซึ่งจากแถวสุดท้ายจะเห็นว่าทั้ง 3 แบบจำลองที่ประมาณได้จะมีโครงสร้างที่ค่อนข้างต่างกัน



รูปที่ 4: แสดงค่า True Positive Rate (TPR) กับค่า False Positive Rate (FPR) โดยที่ TPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณได้ถูกต้อง และ FPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณผิด สัญลักษณ์สี่เหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL วงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ และสัญลักษณ์กากบาทสีแดงคือผลการแบบจำลอง CGL



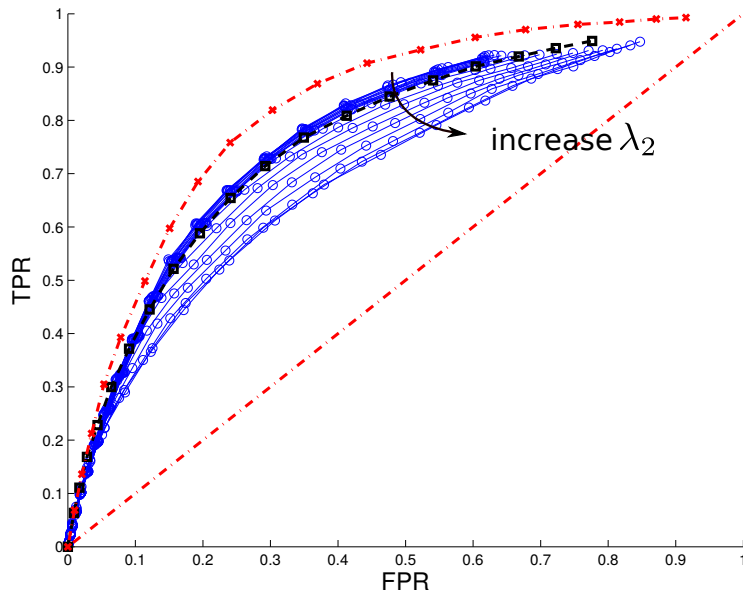
รูปที่ 5: แสดงค่าความคลาดเคลื่อนของแบบจำลองเทียบกับจำนวนสมาชิกที่ไม่เป็นศูนย์ของพารามิเตอร์ AR สัญลักษณ์สี่เหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ สัญลักษณ์กากบาทสีแดงคือผลการแบบจำลอง CGL



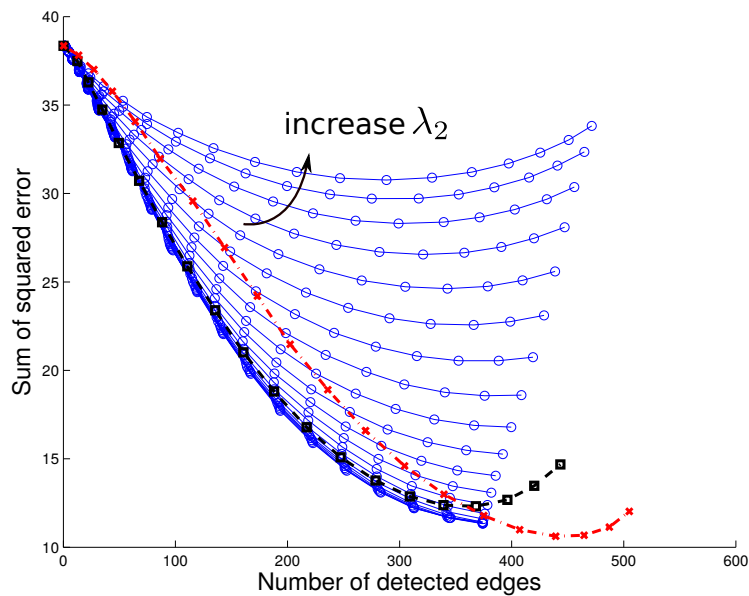
รูปที่ 6: รูปแบบศูนย์ของพารามิเตอร์ AR ที่ประมาณได้จากแบบจำลอง Group Fused Lasso (แถวกลาง) และ Group Lasso (แถวล่าง) เมื่อเปรียบเทียบกับรูปแบบศูนย์ที่แท้จริง (แถบบนสุด)

การทดลองที่ 2 คือการทดลองเหมือนการทดลองที่ 1 เพียงแต่เปลี่ยนสมมติฐานว่า ระบบจริง K ระบบนั้น มีรูปแบบเกรนเจอร์ที่เหมือนกันทั้งหมด การทดลองจึงเปรียบเทียบผลสมรรถนะของ 3 วิธีคือ Group Fused Lasso (GFL), Group Lasso (GL) และ Common Group Lasso (CGL)

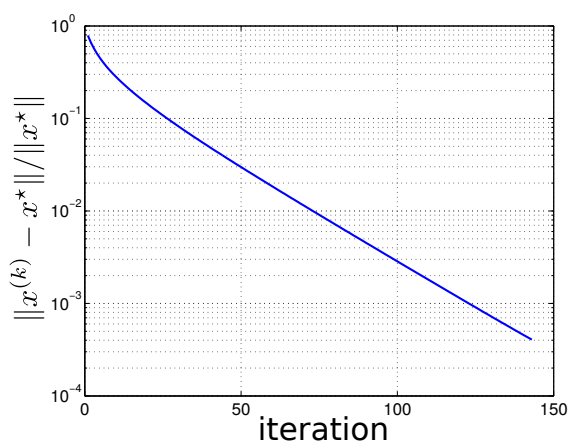
- ▶ รูป 7 แสดงถึงค่าอัตราการทำนายค่าที่ไม่เท่ากับศูนย์ที่ถูกต้อง (True Positive Rate-TPR) เทียบกับ อัตราการทำนายค่าที่ไม่เท่ากับศูนย์ที่ผิด (False Positive Rate-FPR) ซึ่งค่าที่แสดงนี้ถูกเฉลี่ยจากการทดลองบนข้อมูล 100 ชุด จากรูปจะเห็นว่าที่ค่า FPR ที่กำหนดค่าหนึ่ง กราฟจากแบบจำลอง Common Group Lasso นั้นจะอยู่เหนือกราฟที่ได้จากแบบจำลอง Group Lasso และ Group Fused Lasso นั้นหมายความว่า หากระบบจริงมีรูปแบบเกรนเจอร์ที่เหมือนกัน การใช้รูปแบบปัญหา Common Group Lasso นั้นสมเหตุสมผลที่จะทำให้ผลการประมาณรูปแบบเกรนเจอร์ถูกต้องกว่าแนวทางอื่น นอกจากนี้ พบว่าจากกราฟ การใช้ Group Fused Lasso ที่ดีที่สุดนั้น จะได้จากการใช้ค่า λ_2 ที่มากที่สุด นั่นคือเราบังคับให้ทั้ง K แบบจำลองมีรูปแบบศูนย์ที่เหมือนกัน ซึ่งเป็นไปตามสมมติฐานของระบบจริง
- ▶ รูป 8 จะเห็นว่า ในกรณีที่ระบบจริงมีรูปแบบเกรนเจอร์เหมือนกันทั้ง K ระบบ ผลการประมาณด้วย GL กลับให้ผลความคลาดเคลื่อนของแบบจำลองที่ต่ำที่สุด ซึ่งต่ำกว่า CGL (ซึ่งวิธีนี้ให้ผลการประมาณรูปแบบเกรนเจอร์ดีที่สุด) เนื่องจากรูปแบบปัญหา GL เป็นการประมาณ K แบบจำลองอย่างเป็นอิสระต่อกัน จึงยอมทำให้อธิบายข้อมูลได้ดีกว่า การประมาณด้วย CGL



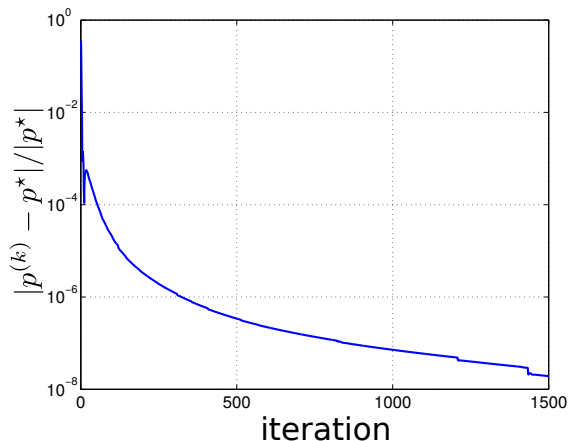
รูปที่ 7: แสดงค่า True Positive Rate (TPR) กับค่า False Positive Rate (FPR) โดยที่ TPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณได้ถูกต้อง และ FPR คือจำนวนสมาชิกที่ไม่เป็นศูนย์ในพารามิเตอร์ AR ที่ถูกประมาณผิด สัญลักษณ์สี่เหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL สัญลักษณ์กากบาทสีแดงคือผลจากการใช้แบบจำลอง CGL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL ที่ค่า λ_2 ต่างๆ



รูปที่ 8: แสดงค่าความคลาดเคลื่อนของแบบจำลองเทียบกับจำนวนสมาชิกที่ไม่เป็นศูนย์ของพารามิเตอร์ AR สัญลักษณ์สี่เหลี่ยมสีดำคือผลจากการใช้แบบจำลอง GL สัญลักษณ์กากบาทสีแดงคือผลจากการใช้แบบจำลอง CGL และวงกลมสีน้ำเงินคือผลที่ได้จากแบบจำลอง GFL และผลถูกแสดงที่ค่า λ_2 ต่างๆ



(a) ค่าความคลาดเคลื่อนสัมพัทธ์ของผลตอบ



(b) ค่าความคลาดเคลื่อนสัมพัทธ์ของค่าฟังก์ชันต้นทุน

รูปที่ 9: ค่าความคลาดเคลื่อนของผลตอบและค่าฟังก์ชันต้นทุนของปัญหา (6) เทียบกับดัชนีการวนซ้ำของวิธี ADMM

การทดลองที่ 3 คือการทดลองเพื่อทดสอบขั้นตอนวิธีเชิงเลข (algorithms) ที่นำมาใช้ในการแก้ปัญหา (6) โดยวิธีที่เลือก จะเป็นวิธีที่ใช้เกรเดียนต์เท่านั้นเพื่อให้เหมาะสมกับขนาดของปัญหาที่ใหญ่ วิธีนี้ที่เลือกคือ Alternating Direction Method of Multiplier (ADMM) ซึ่งในรายละเอียดขอให้ดูในบทความ [Son15] ในการทดลองนี้ เราแก้ปัญหา (6) ซึ่งมีจำนวนตัวแปร 4800 ตัว รูป 9 (a) แสดงค่าความคลาดเคลื่อนสัมพัทธ์ $\|x^{(k)} - x^*\| / \|x^*\|$ โดยที่ $x^{(k)}$ หมายถึงผลตอบของปัญหาเมื่อแปลงให้อยู่ในรูปแบบที่จะใช้กับวิธี ADMM ได้ สัญลักษณ์ k^{th} หมายถึงตัวชี้การวนในขั้นตอนเชิงเลข และ x^* หมายถึงผลตอบเหมาะสมที่สุด (optimal solution) ที่คำนวณจากโปรแกรม CVX [GB07] ซึ่งเป็นชุดคำสั่งที่ใช้บนโปรแกรม MATLAB สำหรับการแก้ปัญหาแบบคอนเวกซ์ และให้ผลตอบที่มีความแม่นยำค่อนข้างสูง จะเห็นได้ว่าวิธี ADMM นั้น สามารถให้ผลตอบที่มีความคลาดเคลื่อนในระดับ 10^{-3} ภายในไม่กี่รอบวนซ้ำ และภายใน 2 – 3 วินาทีเท่านั้น

เมื่อเราเพิ่มขนาดของปัญหาที่มีจำนวนตัวแปรเท่ากับ 30,000 สำหรับปัญหาใหญ่ขนาดนี้ ชุดคำสั่งสำเร็จรูปอย่าง CVX จะมีปัญหาเรื่องหน่วยความจำเพื่อเก็บตัวแปร เราจึงรันวิธี ADMM ไป 10,000 รอบวนซ้ำ และสมมติให้ค่าฟังก์ชันหลังจากรอบที่ 10,000 iteration นั้นจะเข้าใกล้ค่าฟังก์ชันเหมาะสมที่สุด (optimal value) และใช้สัญลักษณ์ว่าเป็น p^* รูป 9 (b) แสดงค่า $(p^{(k)} - p^*) / p^*$ เทียบกับเลขวนซ้ำ k โดยที่ $p^{(k)}$ แสดงค่าฟังก์ชันที่รอบวนซ้ำที่ k เราจะเห็นจากรูปภาพว่า ค่าความคลาดเคลื่อนลดลงไปต่ำถึง 10^{-6} ภายในรอบวนซ้ำที่ประมาณ 300 และใช้เวลาเพียง 300 – 400 เท่านั้น (เทียบกับจำนวนตัวแปรในปัญหาที่มีถึง 30,000 ตัวแปร) จึงอาจกล่าวได้ว่า วิธี ADMM ที่เราเลือกมาใช้นั้น มีความเหมาะสมกับรูปแบบปัญหาที่พิจารณาในโครงการวิจัยนี้ เนื่องจากเป็นวิธีที่ใช้หน่วยความจำไม่สูง และสามารถลู่เข้าหาผลตอบเหมาะสมที่สุดได้ในเวลาที่ไม่ยาวนานนัก

การทดลองที่ 4 คือการนำรูปแบบปัญหาที่นำเสนอขึ้นมาประยุกต์กับข้อมูล fMRI (functional magnetic resonance imaging) ข้อมูลดังกล่าวเป็นอนุกรมเวลาแสดงถึงระดับเลือดในสมอง (BOLD) ที่เป็นสิ่งชี้ถึงระดับกิจกรรมการทำงานของสมองในระดับ voxel ในการทดลองนี้ได้ใช้ข้อมูล StarPlus fMRI data [MW, MHN+04] ที่เก็บมาจากการทดลองให้บุคคลหนึ่งๆ ทำกิจกรรมตอบสนอง 3 ชนิดได้แก่ 1) การนั่งพักนิ่งและมองที่จุดหนึ่งๆ บนหน้าจอ 2) บุคคลเห็นรูปภาพและตัวหนังสือตามลำดับ และ 3) บุคคลเห็นตัวหนังสือและภาพตามลำดับ โดยที่ในกรณีสองและสามนั้น บุคคลที่อยู่ในการทดลองจะต้องตอบสนองต่อด้วยการกดปุ่มที่แสดงว่า ตัวหนังสือที่เห็นนั้นบรรยายได้ตรงกับภาพหรือไม่ ข้อมูลมีความถี่สุ่มตัวอย่างทุกๆ 500 msec และเก็บมาจาก 25-30 ส่วนของสมอง (หรือที่มักจะเรียกว่า Regions of Interest, ROIs)

Table 1: Region of Interest

ROI	คำจำกัดความ
CALC	calcarine sulcus
(L/R) DLPFC	dorsolateral prefrontal cortex
(L/R) FEF	frontal eye fields
(L) IFG	inferior frontal gyrus
(L/R) IPL	inferior parietal lobule
(L/R) IPS	intraparietal sulcus
(L/R) IT	inferior temporal gyrus
(L/R) OPER	operculum
(L/R) PPREC	(unavailable)
(L/R) SGA	(unavailable)
(L/R) SPL	superior parietal lobule
(L/R) T	temporal lobe
(L/R) TRIA	(unavailable)
SMA	supplementary motor area

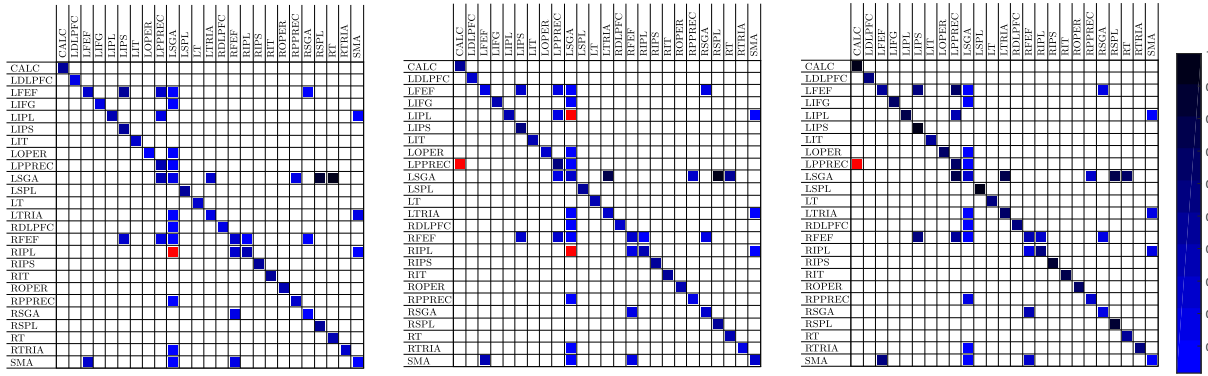
ในบุคคลหนึ่งๆ จะมีการทดลองหลายครั้ง ทำให้มีข้อมูลประมาณ 40 ชุด การทดลองนี้จึงสามารถแบ่งข้อมูลจาก 40 ชุดนี้ ไปเป็น estimation data set (สำหรับหาแบบจำลอง) และแบ่งไปเป็น validation data set (สำหรับทำ cross validation) ได้ การทดลองจะเริ่มจากกำหนดค่า $K = 3$ เพื่อหาแบบจำลองสำหรับข้อมูล 3 ชนิด คือ

1. แบบจำลองสมองขณะบุคคลพักนิ่ง
2. แบบจำลองสมองเมื่อบุคคลเห็นตัวหนังสือบรรยายตรงกับภาพ
3. แบบจำลองสมองเมื่อบุคคลเห็นตัวหนังสือที่ไม่ได้บรรยายตรงกับภาพ

จากนั้น เราจะประมาณแบบจำลองด้วยรูปแบบปัญหา GFL ด้วยการใส่ค่า (λ_1, λ_2) ที่ค่าต่างๆ กัน เป็นจำนวน 20×20 ค่า แต่ละค่าของ (λ_1, λ_2) นั้นจะให้ค่าพารามิเตอร์และรูปแบบการเชื่อมโยงของสมอง (GC structure) ที่ต่างกันไป เราจะนำแบบจำลองนั้นมาคำนวณค่า cross validation measure:

$$J(\lambda_1, \lambda_2) = \sum_{k=1}^K \|Y^{(k)} - \hat{A}^{(k)}(\lambda_1, \lambda_2)H^{(k)}\|_F^2$$

โดยที่ใช้ข้อมูล $(Y^{(k)}, H^{(k)})$ จาก validation data set และค่า (λ_1, λ_2) จะเลือกจากค่าที่ทำให้ J ต่ำสุด



รูปที่ 10: เมทริกซ์แสดงค่าความเข้มของพารามิเตอร์ในแบบจำลอง AR ทั้งสามชนิด (ซ้าย) แบบจำลองสมองบุคคลพักนิ่ง (กลาง) แบบจำลองเมื่อบุคคลเห็นภาพและประโยคที่สอดคล้องกัน (ขวา) แบบจำลองเมื่อบุคคลเห็นภาพและประโยคที่ไม่สอดคล้องกัน จุดสีน้ำเงินคือจุดที่มีร่วมกันในทั้งสามแบบจำลอง และจุดสีแดงคือจุดที่ต่างกัน

ผลการทดลองในรูป 10 คือการแสดงความเชื่อมโยงการทำงานของสมองลงบน adjacency matrix โดยความเข้มของจุดในเมทริกซ์นั้นจะผันตรงกับ

$$\|[(A_1)_{ij} \quad (A_2)_{ij} \quad \dots \quad (A_p)_{ij}]\|_2 \quad (20)$$

ผลการทดลองในรูป 11 แสดงรูปแบบการเชื่อมโยงการทำงานของส่วนต่างๆ ในสมอง โดยความหนาของแถบของเส้นเชื่อมโยงจะแปรผันตรงกับ 20 เช่นกัน (กล่าวคือ ถ้าค่าสัมประสิทธิ์ในแบบจำลองมีค่ามาก ความสัมพันธ์ของส่วนต่างในสมองที่เกี่ยวข้องก็จะตีความว่ามีค่ามาก) ส่วนของสมองที่มีความสัมพันธ์กันอย่างมากและพบในทั้งสามแบบจำลอง ได้แก่

LSGA-RSPL, LSGA-RT, LFEF-LIPS, SMA-LFEF

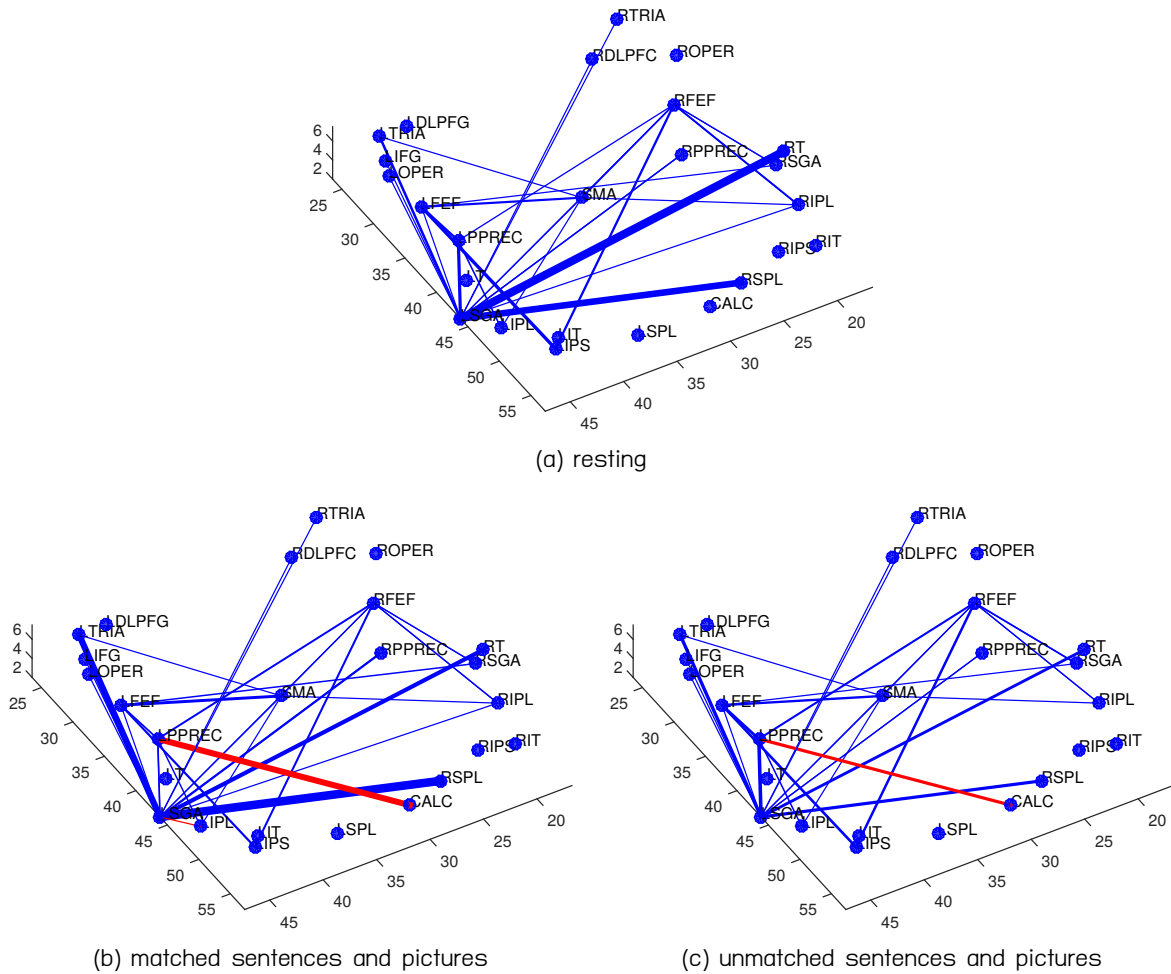
SPL จะเกี่ยวกับระบบ somatosensory การได้ยิน และการเคลื่อนไหว [VG15, pp. 558] T (temporal lobe) จะเกี่ยวข้องกับการประมวลข้อมูลทางสัมผัส เพื่อที่จะเก็บรับรู้ต่อไปในความทรงจำเชิงภาพ การตีความของภาษา หรือความเกี่ยวข้องกับความจำ [Car14, pp.83-85, 155,255] ฟังก์ชันของสมองส่วน SMA นั้นเกี่ยวข้องกับการควบคุมหรือการวางแผนการเคลื่อนไหวที่มาจากความทรงจำ [Car14, pp.117, 180] IPS เกี่ยวข้องกับ perceptual-motor coordination หรือ visuo motor tasks เช่น การขยับตา หรือ การเอื้อมหยิบของ [GF05] และยังเกี่ยวข้องกับการตีความเชิงเลข (numerical values) [Car14, pp.169] FEF จะเกี่ยวกับการรับรู้ตำแหน่งหรือการเคลื่อนไหวของวัตถุผ่านทางภาพ [Car14, pp. 183] หรือการเคลื่อนไหวของตาทั้งสองข้างอย่างรวดเร็วในทิศเดียวกัน (saccadic eye movement) เมื่อมีการรับรู้เชิงภาพ [VG15, pp.531] IPL เป็นสมองส่วนที่เกี่ยวข้องกับการรับรู้อารมณ์ผ่านการแสดงออกทางสีหน้า การตีความของการรับรู้จากสิ่งที่เห็น (mirror neuron) [Car14, pp. 122] CALC เป็นส่วนที่เกี่ยวข้องกับส่วนที่เรียกว่า primary visual cortex [VG15, pp.66] (การประมวลผลจากภาพ)

ความสัมพันธ์ที่แบบจำลองทั้ง 3 แบบจำลองพบด้วยวิธีที่เราแนะนำ และสามารถอธิบายได้จากการทำงานของส่วนสมองดังกล่าวข้างต้น จึงเป็นคู่ LFEF-LIPS ที่มีระดับความสัมพันธ์ GC causality ค่อนข้างสูง และเกี่ยวข้องกับการรับรู้เชิงภาพ นอกจากนี้จะเห็นโดยภาพรวมว่า แบบจำลองที่ถูกเลือกมีความใกล้เคียงกันเป็นอย่างมากในสภาพเชิงโยง แต่โครงสร้างที่ไม่เหมือนกันใน 3 แบบจำลองจะมีเพียง

1. RIPL-LSGA: ไม่ปรากฏในแบบจำลองที่ 3
2. LPPREC-CALC: ไม่ปรากฏในแบบจำลองที่ 1

3. LIPL-LSGA: ปรากฏในแบบจำลองที่ 2 เท่านั้น

เราสังเกตว่าสมองส่วน CALC นั้นเกี่ยวกับการประมวลผลจากภาพ ซึ่งไม่ได้ปรากฏในแบบจำลองที่ 1 (บุคคล อยู่นิ่ง) จึงมีความสมเหตุสมผลที่เจอ activation ของสมองส่วนนี้ในแบบจำลอง 2 และ 3 อย่างไรก็ตาม เนื่องจากความสัมพันธ์การทำงานของสมองที่แท้จริงนั้นยังเป็นสิ่งที่ neuroscientists ยังต้องหาคำตอบอยู่ การจะสรุปว่าส่วนของสมองคู่หนึ่งๆ จะสัมพันธ์หรือทำงานต่างกันอย่างไรมันจำเป็นต้องมีการทดลองบนสมองคนจริง อย่างน้อยแต่ การ validate ผลลัพธ์สภาพเชื่อมโยงทุกคู่ดังข้างต้นในงานวิจัยนี้จึงอยู่นอกเหนือขอบเขตที่วางไว้ ผู้วิจัยจึงขอสรุปผลลัพธ์ข้างต้นว่าเป็นเพียง guidelines ในงานวิจัยอื่น ที่จะศึกษาความสัมพันธ์ของส่วน สมองคู่ใดคู่หนึ่งอย่างเจาะจงต่อไป



รูปที่ 11: รูปแบบการเชื่อมโยงการทำงานของสมองด้วย GC causality ของข้อมูลสมอง 3 ชนิด

7 สรุปผล

ในโครงการวิจัยนี้ ผู้วิจัยได้ศึกษาถึงการหาแบบจำลองเพื่อมาอธิบายพลวัตของข้อมูลกรวมเวลาที่มีคุณลักษณะ ดังนี้

1. ข้อมูลอนุกรมเวลามีหลายกลุ่ม (แต่ละกลุ่มเป็นข้อมูลตัวแปรชนิดเดียวกัน) โดยในแต่ละกลุ่มต่างกันไป ตามตัวแปรภายนอกหรือเงื่อนไขจากสภาวะภายนอก
2. แบบจำลองของอนุกรมเวลาจะสามารถอธิบายถึงโครงสร้างความสัมพันธ์แบบ Granger causality ของตัวแปรในข้อมูลได้
3. แบบจำลองของอนุกรมเวลาในแต่ละกลุ่ม อาจจะมีโครงสร้างความสัมพันธ์บางส่วนที่เหมือนกัน และบางส่วนที่ต่างกัน

รูปแบบการประมาณแบบจำลองที่นำเสนอเป็นปัญหาการหาค่าต่ำสุดแบบกำลังสองน้อยสุดที่มีเทอม regularization ด้วย ℓ_1 นอร์ม และอยู่ในรูปแบบปัญหา fused lasso ปัญหาดังกล่าวทำให้ประมาณแบบจำลอง K แบบจำลองได้พร้อมกัน โดยการเรียนรู้โครงสร้างสัมพันธ์ Granger causality จะทำได้จากการเลือกค่าพารามิเตอร์ลงโทษ (λ_1, λ_2) ที่เหมาะสมด้วยวิธี cross validation รูปแบบที่นำเสนอยังเป็นปัญหาในกรณีทั่วไปที่รวมถึงกรณีเฉพาะสองกรณีอื่นได้แก่ การประมาณแบบจำลอง K แบบจำลองอย่างอิสระต่อกัน และการประมาณ K แบบจำลอง ที่ทุกแบบจำลองต้องมีโครงสร้าง GC causality เหมือนกันทั้งหมด ผู้วิจัยได้เสนอการประยุกต์ขั้นตอนวิธี ADMM เพื่อมาแก้ปัญหาเชิงเลขและแสดงให้เห็นว่าสามารถแก้ปัญหาที่มีจำนวนตัวแปรมากได้ ผลการทดลองพบว่า รูปแบบปัญหาการประมาณแบบจำลอง K แบบจำลองพร้อมกันจะทำได้ดี เมื่อระบบจริง K ระบบมีความคล้ายกัน ในขณะที่หากระบบจริง K ระบบมีโครงสร้างเหมือนกันหมด การประมาณ K แบบจำลองพร้อมกันด้วยรูปแบบปัญหาที่บังคับโครงสร้างให้เหมือนกันทุกแบบจำลอง ก็จะทำให้ผลลัพธ์ที่ดีที่สุดเช่นกัน งานวิจัยนี้ได้แสดงตัวอย่างการนำไปประยุกต์กับการหาแบบจำลองของข้อมูลสมอง fMRI ที่เก็บข้อมูลมาจาก 3 สภาวะ และผลที่พบคือ แบบจำลองของสามสภาวะนั้นมีความคล้ายกันในระดับที่สูง

8 ปัญหาในการทำวิจัย และการแก้ไข้ปัญหา

1. การทำวิจัยช่วงแรกจำเป็นต้องศึกษาถึงองค์ความรู้ที่เกี่ยวข้อง ดังนั้นการมีตำราต่างประเทศใหม่ๆ มาศึกษาก็เป็นเรื่องที่จำเป็นในช่วงแรก แต่ทว่าวิจัยในงวดแรกให้มาเฉพาะหมวดครุภัณฑ์ จึงทำให้ยังไม่ได้ซื้อตำรา และแก้ไข้ปัญหาโดยการอ่านแต่บทความวิจัยใหม่ๆ ไปก่อน
2. การทำวิจัยช่วงท้าย ที่จำเป็นต้องทำการทดลองบนข้อมูลสมองจริงนั้น ผู้วิจัยต้องอาศัยความร่วมมือกับอาจารย์ในประเทศมาเลเซีย และการเก็บข้อมูลดังกล่าว ปกติใช้งบประมาณค่อนข้างสูงเพื่อจ่ายให้กับคนไข้ และเป็นค่าใช้จ่ายในการรันเครื่อง fMRI ดังนั้น ปัจจัยนี้อาจจะส่งผลกระทบต่อระยะเวลาดำเนินการวิจัย ที่อาจจะช้าลง ถ้าหากข้อมูลต้องรอจากอาจารย์ที่มาร่วมกัน การแก้ไข้ปัญหาจึงเป็นการติดต่ออาจารย์คนดังกล่าวเป็นระยะๆ รวมถึง เขียนข้อเสนอโครงการร่วมกันและขอทุนที่อื่นเพิ่มเติม เพื่อนำมาเป็นค่าใช้จ่ายในการทำวิจัยเรื่องนี้

เอกสารอ้างอิง

- [ABD13] C. M. Alaíz, A. Barbero, and J. R. Dorronsoro. Group fused lasso. *Artificial Neural Networks and Machine Learning–ICANN 2013*, pages 66–73, 2013.
- [ABVN11] A. A. Bolstad, B. Van Veen, and R. Nowak. Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing*, 59(6): 2628–2641, 2011.
- [ALW13] E. Avventi, A. Lindquist, and B. Wahlberg. ARMA identification of graphical models. *IEEE Transactions on Automatic Control*, 58(5):1167–1178, 2013.
- [BPC⁺10] 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*, 3(1):1–122, 2010.
- [BT09] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.
- [Car14] Rita Carter. *The Human Brain Book*. Penguin, 2014.
- [CGS⁺11] 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*, pages 1142–1155, 2011.
- [CRT⁺09] G. A. Cechhi, I. Rish, B. Thyreau, B. Thirion, M. Plaze, M. L. Paillere-martinot, C. Martelli, J. L. Martinot, and J. B. Poline. Discriminative network models of schizophrenia. In *Advances in Neural Information Processing Systems*, pages 252–260, 2009.
- [DLJ⁺09] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu. Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, 30(4):1361–1373, 2009.
- [dSCS⁺11] E. L. da Silva, G. Castellano, J. Sato, E. F. Cardoso, and E. Amaro Jr. Emotional face perception in healthy subjects and parkinson’s disease: An effective connectivity study. In *Biomedical Engineering Systems and Technologies*, pages 278–289. Springer, 2011.
- [GB07] M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming (web page and software)*. <http://stanford.edu/~boyd/cvx>, 2007.
- [GCR11] R. Garg, G. A. Cecchi, and A. R. Rao. Full-brain auto-regressive modeling (FARM) using fMRI. *Neuroimage*, 58(2):416–441, 2011.
- [GF05] Christian Grefkes and Gereon R Fink. The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of anatomy*, 207(1):3–17, 2005.
- [GKRM03] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.
- [GSMD09] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, 19(1):72–78, 2009.

- [HLS⁺10] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, and E. Reiman. Learning brain connectivity of alzheimer’s disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.
- [HNK08] S. Haufe, G. Nolte, and N. Kräemer. Sparse causal discovery in multivariate time series. *Proceedings of JMLR Workshop and Conference*, 6:97–106, 2008.
- [HP10] M. P. Van Den Heuvel and H. E. Hulshoff Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010.
- [HTF09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd edition, 2009.
- [LALR09] A.C. Lozano, N. Abe, Y. Liu, and S. Rosset. Grouped graphical granger modeling for gene expression regulatory networks discovery. *Bioinformatics*, 25:110–118, 2009.
- [Lüt05] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer, 2005.
- [MCH⁺12] K. Mohan, M. J. Chung, S. Han, D. Witten, S. Lee, and M. Fazel. Structured learning of gaussian graphical models. In *Advances in Neural Information Processing Systems*, pages 629–637, 2012.
- [MHN⁺04] T.M. Mitchell, R. Hutchinson, R.S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine learning*, 57(1-2):145–175, 2004.
- [MW] T. Mitchell and W. Wang. StarPlus fMRI data. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>.
- [NA11] B. Ng and R. Abugarbieh. Generalized sparse regularization with application to fMRI brain decoding. In *Information Processing in Medical Imaging*, pages 612–623. Springer, 2011.
- [OLB10] H. Ohlsson, L. Ljung, and S. Boyd. Segmentation of ARX-models using sum-of-norms regularization. *Automatica*, 46(6):1107–1111, 2010.
- [OV14] D. O’Connor and L. Vandenberghe. Primal-dual decomposition by operator splitting and applications to image deblurring. *SIAM Journal on Imaging Sciences*, 7(3):1724–1754, 2014.
- [PB14] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.
- [PLS13] A. Pongrattarakul, P. Lerdkultanon, and J. Songsiri. Sparse system identification for discovering brain connectivity from fMRI time series. In *Proceedings of SICE Annual Conference*, pages 949–954, 2013.
- [PMB09] F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209, 2009.
- [RHH⁺10] J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fmri. *Neuroimage*, 49(2):1545–1558, 2010.
- [RM11] J. C. Rajapakse and P. A. Mundra. Stability of building gene regulatory networks with sparse autoregressive models. *BMC Bioinformatics*, 12(13):1–10, 2011.

- [RPSTMM13] M. Rosa, L. Portugal, J. Shawe-Taylor, and J. Mourao-Miranda. Sparse network-based models for patient classification using fmri. In *The 3rd International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 66–69. IEEE, 2013.
- [RSCM11] S. Ryali, K. Supekar, T. Chen, and V. Menon. Multivariate dynamical systems models for estimating causal interactions in fMRI. *Neuroimage*, 54(2):807–823, 2011.
- [Smi12] S. M. Smith. The future of fmri connectivity. *Neuroimage*, 62(2):1257–1266, 2012.
- [SMSK⁺11] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich. Network modelling methods for FMRI. *Neuroimage*, 54(2):875–891, 2011.
- [Son13a] J. Songsiri. Sparse autoregressive model estimation for learning Granger causality in time series. In *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3198–3202, 2013.
- [Son13b] J. Songsiri. Sparse optimization problems in system identification. *Engineering Journal (Thai)*, 5(1):51–75, 2013.
- [Son15] J. Songsiri. Learning Multiple Granger Graphical Models via Group Fused Lasso. In *Proceedings of the 10th Asian Control Conference (ASCC)*, 2015.
- [SR12] K. E. Stephan and A. Roebroeck. A Short history of causal modeling of fMRI data. *NeuroImage*, 62:856–863, 2012.
- [SV10] J. Songsiri and L. Vandenberghe. Topology selection in graphical models of autoregressive processes. *Journal of Machine Learning Research*, 11:2671–2705, 2010.
- [TBS⁺12] 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*, 8(5):1–14, 2012.
- [TSR⁺05] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- [VG15] Todd Vanderah and Douglas J. Gould. *Nolte’s The Human Brain: An Introduction to Its Functional Anatomy*. Elsevier Health Sciences, 2015.
- [VSBSVH⁺06] P.A. Valdés-Sosa, J.M. Bornot-Sánchez, M. Vega-Hernández, L. Melie-García, A. Lage-Castellanos, and E. Canales-Rodríguez. Granger causality on spatial manifolds: Applications to neuroimaging. In B. Schelter, M. Winterhalder, and J. Timmer, editors, *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Wiley, 2006.
- [VSRDF11] P. A. Valdés-Sosa, A. Roebroeck, J. Daunizeau, and K. Friston. Effective connectivity: Influence, causality and biophysical modeling. *Neuroimage*, 58(2): 339–361, 2011.
- [WBAW12] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang. An admm algorithm for a class of total variation regularized estimation problems. In *the 16th IFAC Symposium on System Identification*, pages 83–88, 2012.

- [YL06] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 68(1):49–67, 2006.
- [YPS+15] S. Yang, Z. Pan, X. Shen, P. Wonka, and J. Ye. Fused multiple graphical lasso. *SIAM Journal on Optimization*, 25(2):916–943, 2015.
- [ZRY09] P. Zhao, G. Rocha, and B. Yu. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics*, 37(6A): 3468–3497, 2009.

ประวัติผู้วิจัย

จิตโกมุท ส่งศิริ ปัจจุบันดำรงตำแหน่ง ผู้ช่วยศาสตราจารย์ ดร. ที่ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และสังกัดห้องปฏิบัติการระบบควบคุม งานวิจัยที่สนใจคือการหาแบบจำลองของระบบ โครงสร้างความสัมพันธ์ของการทำงานของสมอง การพยากรณ์ความเข้มแสงอาทิตย์ และงานประยุกต์อื่นๆ ของปัญหาการหาค่าต่ำสุดแบบคอนเวกซ์ สำหรับรายละเอียดอื่นๆ สามารถดูได้ที่ <http://jitkomut.eng.chula.ac.th>