



รายงานวิจัยฉบับสมบูรณ์

การขยายเพิ่มของอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน

Further Extension of Hypercircle Inequality to Inaccurate Data

โดย กรรณิการ์ ขำพืงสน

ธันวาคม 2558

รายงานวิจัยฉบับสมบูรณ์

การขยายเพิ่มของอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน

Further Extension of Hypercircle Inequality to Inaccurate Data

ผู้วิจัย

กรรณิการ์ ขำพึงสน

สังกัด

คณะวิทยาศาสตร์ มหาวิทยาลัยพะเยา

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

และมหาวิทยาลัยพะเยา

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว.ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

ผู้วิจัยใคร่ขอขอบพระคุณ สำนักงานกองทุนสนับสนุนการวิจัย (สกว.) และมหาวิทยาลัยพะเยา ที่ให้การสนับสนุนงานวิจัยในครั้งนี้ผ่านทางทุนพัฒนาศักยภาพในการทำงานวิจัยของอาจารย์รุ่นใหม่ ประจำปีงบประมาณ 2556

ขอขอบพระคุณ ศาสตราจารย์ ดร. สุเทพ สอนใต้ ที่ให้คำปรึกษาและคำแนะนำ ที่เป็นประโยชน์ต่องานวิจัยนี้

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร. ไฉพจน์ งามสอาด ผู้ช่วยศาสตราจารย์ ดร. สายยัน ปันมา และ ดร. นราธิป สงมณี ที่ให้ความช่วยเหลือแก่ผู้วิจัย

ท้ายที่สุดขอกราบขอบพระคุณ คุณพ่อและคุณแม่ น้องๆ และเพื่อนๆ ผู้ให้กำลังใจที่ดีเสมอมา

กรรณิการ์ ขำพึงสน

บทคัดย่อ

รหัสโครงการ: TRG5680030

ชื่อโครงการ: การขยายเพิ่มของอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน

ชื่อนักวิจัย: กรรณิการ์ ขำพึงสน

E-mail Address : kkannika13@hotmail.com

ระยะเวลาโครงการ: 2 ปี

การสร้างแบบจำลองทางคณิตศาสตร์เพื่ออธิบายปรากฏการณ์ทางวิทยาศาสตร์วิธีหนึ่งคือ นำข้อมูลจากการทดลองมาหาความสัมพันธ์ระหว่างตัวแปรที่ปรากฏในการทดลอง แล้วเขียนฟังก์ชันหรือสมการทางคณิตศาสตร์ เพื่อเป็นตัวแทนใช้ในการอธิบายปรากฏการณ์ อสมการไฮเพอร์เซอร์เคิล (Hypersphere inequality) ถือเป็นอีกเครื่องมือหนึ่งที่ถูกนำมาประยุกต์ใช้ในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐานและเป็นที่รู้จักเป็นอย่างดีในการศึกษาแขนงนี้ แต่เนื่องจากข้อจำกัดของอสมการไฮเพอร์เซอร์เคิลที่สามารถใช้กับการศึกษานี้ได้ในกรณีที่ทราบค่าของข้อมูลอย่างแน่นอน ต่อมาผู้วิจัยได้ศึกษาอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนสองแนวทาง คือ การขยายอสมการดังกล่าวไปยังข้อมูลที่มีค่าคลาดเคลื่อนทุกตัว และขยายอสมการดังกล่าวไปยังข้อมูลที่มีค่าคลาดเคลื่อนเพียงบางตัวเท่านั้น ดังนั้นในงานวิจัยนี้ผู้วิจัยได้นำองค์ความรู้จากการขยายทั้งสองรูปแบบมาประยุกต์กับปัญหาการประมาณค่าของฟังก์ชันในปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต นอกจากนี้ในงานวิจัยนี้ยังได้เสนอการแปลงสมการดังกล่าวไปสู่เซตเชิงตั้งฉาก (orthonormal set) ซึ่งเป็นเครื่องมือที่มีประโยชน์มากในปริภูมิฮิลเบิร์ต แต่อย่างไรก็ตามอสมการดังกล่าวก็ยังมีข้อจำกัดในการประยุกต์หลายด้าน อาทิเช่น ปริภูมิที่จะนำไปประยุกต์ใช้ต้องเป็นปริภูมิฮิลเบิร์ต ดังนั้นผู้วิจัยจึงเสนอแนวทางสำหรับงานวิจัยในอนาคตคือ ขยายอสมการดังกล่าวไปยังปริภูมิบานาค

คำหลัก : อสมการไฮเพอร์เซอร์เคิล ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต การหาค่าเหมาะสมที่สุด

ฟังก์ชันนูล ข้อมูลที่มีค่าคลาดเคลื่อน

Abstract

Project Code : TRG5680030

Project Title : Further Extension of Hypercircle Inequality to Inaccurate Data

Investigator : Kannika Khompurngson

E-mail Address : kkannika13@hotmail.com

Project Period : 2 years

A foundational concept in learning problem is to construct a functional representation from given data. Among learning methods, Hypercircle inequality (Hi) has been applied to kernel-based machine learning when data is known exactly. Recently, we have extended Hi to data error in two ways: First, we have extended it to circumstance for which all data is known within error. Second, we have extended it to partially-corrupted data. That is, data set contains both accurate and inaccurate data. In this research, we continue our study of this subject by using the material from both previous work to estimate the unknown vectors in Hilbert space from knowledge of both its norm and linear observations of it, known within error. Furthermore, we proposed the transformation of hypercircle inequality for partially-corrupted data to orthonormal bases and also specialized the new result to the learning problem in Reproducing kernel Hilbert space (RKHS). In future work, we propose to extend Hi to Banach space.

Keywords : Hypercircle inequality, Reproducing kernel Hilbert space, convex optimization and noisy data

สารบัญเรื่อง

หัวข้อ	หน้า
บทที่ 1 ที่มาและความสำคัญของปัญหา	8
บทที่ 2 ทบทวนวรรณกรรม	11
บทที่ 3 วัตถุประสงค์ของโครงการ	17
บทที่ 4 ผลการดำเนินงาน	18
บทที่ 5 สรุปและวิจารณ์ผลการวิจัย และข้อเสนอแนะสำหรับงานวิจัยในอนาคต	29
เอกสารอ้างอิง	31
Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.	33
บทที่ 6 ภาคผนวก	34

สารบัญภาพ

รูป		หน้า
รูปที่ 4.1	เซตของข้อมูลที่ต้องการประมาณค่าจากข้อมูลที่มีค่าคลาดเคลื่อน	25
รูปที่ 4.2	ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_0)$ ณ σ ต่าง ๆ และค่าจริง $g(t_0) = 0.6767$	26
รูปที่ 4.3	ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-1})$ ณ σ ต่าง ๆ	27
รูปที่ 4.4	ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-2})$ ณ σ ต่าง ๆ	27

บทที่ 1 ที่มาและความสำคัญของปัญหา

การสร้างแบบจำลองทางคณิตศาสตร์เพื่ออธิบายปรากฏการณ์ทางวิทยาศาสตร์วิธีหนึ่งคือ นำข้อมูลจากการทดลองมาหาความสัมพันธ์ระหว่างตัวแปรที่ปรากฏในการทดลอง แล้วเขียนฟังก์ชันหรือสมการทางคณิตศาสตร์ เพื่อเป็นตัวแทนใช้ในการอธิบายปรากฏการณ์นั้น นอกจากนี้เรายังสามารถนำความสัมพันธ์ดังกล่าวมาคาดคะเนผลการทดลองที่ไม่สามารถทำการทดลองจริงได้ ซึ่งอาจเกิดจากอันตรายที่อาจจะเกิดขึ้นกับผู้ทดลอง ตลอดจนอุปกรณ์ที่ใช้ อาจมีราคาแพง จึงทำให้การวิจัยในแขนงนี้มีความสำคัญและมีบทบาทอย่างมากต่อการพัฒนาความเจริญก้าวหน้าทางวิชาการในทางวิทยาศาสตร์และเทคโนโลยี ประกอบกับปัจจุบันคอมพิวเตอร์มีวิวัฒนาการก้าวหน้าอย่างมาก จึงทำให้มีนักคณิตศาสตร์จำนวนมากสนใจและศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน (kernel-based machine learning) ซึ่งมีจุดประสงค์คือหาฟังก์ชันในปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต (Reproducing kernel Hilbert space) [1] เป็นตัวแทนในการอธิบายความสัมพันธ์ ระหว่างตัวแปรที่ปรากฏในข้อมูล

ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H ประกอบด้วยสมาชิกของฟังก์ชัน K จากเซต Λ ไปยังจำนวนจริง $\square$ เมื่อ Λ เป็นเซตย่อยของจำนวนจริง และ K เป็นฟังก์ชัน ของ s และ t เมื่อ $s, t \in \Lambda$ จะถูกเรียกว่า รีโพรดิวซิงเคอร์เนล (Reproducing kernel) [1] ถ้าสำหรับทุกๆ $t \in \Lambda$ และ ทุกๆ ฟังก์ชัน $f \in H$ แล้ว ให้

$$f(t) = \langle f, K_t \rangle$$

เมื่อ $\langle \cdot, \cdot \rangle$ แทนผลคูณภายใน และ K_t แทนฟังก์ชันจากเซต Λ ไปยังจำนวนจริง $\square$ และกำหนดให้ ทุก ๆ $s \in \Lambda$, $K_t(s) = K(s, t)$ เรานิยาม ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต [1] H คือปริภูมิที่ประกอบด้วยสมาชิกของฟังก์ชันจากเซต Λ ไปยังจำนวนจริง $\square$ เมื่อ Λ เป็นเซตย่อยของจำนวนจริง และมีฟังก์ชัน K ของ s และ t เมื่อ ทุกๆ $s, t \in \Lambda$ เป็นรีโพรดิวซิงเคอร์เนลสำหรับปริภูมิฮิลเบิร์ตนั้น ในปี 1950 Aronszajn and Moore [1] ได้ค้นพบทฤษฎีบท กล่าวคือ ฟังก์ชัน K ข้างต้นจะเป็นรีโพรดิวซิงเคอร์เนล ก็ต่อเมื่อ สำหรับทุกๆ $\{t_1, t_2, \dots, t_n\} \subseteq T$ และ เมทริกซ์ $K(t_i, t_j)$ ซึ่งกำหนดโดย

$$K(t_i, t_j) = \begin{bmatrix} K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \vdots & \ddots & \vdots \\ K(t_n, t_1) & \cdots & K(t_n, t_n) \end{bmatrix}$$

เป็นเมทริกซ์กึ่งบวกแน่นอน (positive semi-definite) ซึ่งนับได้ว่าเป็นทฤษฎีบทที่มีประโยชน์มากในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน นอกจากนี้ยังพบว่า ทุกๆ รีโพรดิวซิงเคอร์เนล K จะมี ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H เพียงปริภูมิเดียวที่สมนัยกับรีโพรดิวซิงเคอร์เนล K ตัวอย่างปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H ที่เป็นที่รู้จักและนิยมนำมาศึกษาในปัญหาหลาย ๆ

ปัญหาคือปริภูมิ $H^2(\Delta)$ ประกอบด้วยสมาชิกที่เป็นฟังก์ชันวิเคราะห์ (analytic function) บนจานหนึ่งหน่วย (unit disc) $\Delta := \{z: |z| \leq 1\}$, [7] และนอร์มถูกกำหนดโดย

$$\|f\| = \sup_{0 < r < 1} \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta \right)^{\frac{1}{2}}$$

และรีโพรดิวซิงเคอร์เนลสำหรับปริภูมินี้คือ

$$K(z, \varsigma) = \frac{1}{1 - \bar{\varsigma}z}$$

สำหรับทุก ๆ $\varsigma, z \in \Delta$,

Tikhonov regularization [9,16,17,20] เป็นวิธีหนึ่งที่เป็นที่นิยมและเป็นที่รู้จักอย่างมากในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน กล่าวคือ กำหนดให้ H เป็นปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต และ $S = \{(t_j, d_j): j \in \square_n\} \subseteq T \times \square$ คือเซตที่เป็นตัวแทนของข้อมูล เมื่อ $\square_n = \{1, 2, \dots, n\}$ จากข้อมูลดังกล่าว หากเราต้องการประมาณของค่าข้อมูลที่จุด t_0 โดยวิธี Tikhonov regularization เริ่มต้นโดยกำหนดให้ ℓ เป็น loss ฟังก์ชันจาก $\square^2$ ไปยังจำนวนจริง $\square$ และ L เป็นฟังก์ชันที่วัดค่าคลาดเคลื่อนจาก H ไปยังจำนวนจริง และกำหนดให้

$$L(f) := \frac{1}{n} \sum_{j \in \square_n} \ell(d_j, f(t_j))$$

ให้ $\sigma > 0$ และฟังก์ชันที่เป็นตัวแทนของข้อมูลชุดนี้ถูกเลือกมาจาก

$$L(f_\sigma) + \sigma \|f_\sigma\|^2 = \min \{L(f) + \sigma \|f\|^2: f \in H\}$$

จาก Representer Theorem พบว่า [8,16,17,20]

$$f_\sigma(t) = \sum_{j \in \square_n} c_j K(t_j, t) \quad \text{เมื่อ } t \in T$$

ดังนั้น $f_\sigma(t_0) = \sum_{j \in \square_n} c_j K(t_j, t_0)$ คือค่าประมาณที่ดีที่สุดจากการใช้วิธีการ Tikhonov regularization

อสมการไฮเพอร์เซอร์เคิล (Hypersphere inequality) ถือเป็นอีกเครื่องมือหนึ่งที่ถูกนำมาประยุกต์ใช้ในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐานและเป็นที่รู้จักเป็นอย่างดีในการศึกษาแขนงนี้ แต่เนื่องจากข้อจำกัดของอสมการไฮเพอร์เซอร์เคิลที่สามารถใช้กับการศึกษานี้ได้ในกรณีที่ทราบค่าของข้อมูลอย่างแน่นอน [5, 11, 15] ซึ่งความเป็นจริงแล้ว ข้อมูลโดยไปมักจะมีความคลาดเคลื่อนรวมอยู่ด้วยเสมอ ดังนั้นจากข้อจำกัดนี้ ในปี 2010 ดร. กรรณิการ์ ขำพึงสน และ Professor Charles A. Micchelli ซึ่งเป็นศาสตราจารย์คณิตศาสตร์ชาวอเมริกัน ที่มีชื่อเสียงมากในการศึกษาแขนงทฤษฎีการประมาณค่า (Approximation Theory) และมีผลงานวิจัยที่ได้ตีพิมพ์เผยแพร่ในวารสารทางวิชาการระดับชาติที่มี impact factor สูง ได้ศึกษาอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน [12] รวมทั้งเสนอการวิธีการใหม่ในการศึกษาการประมาณค่าของฟังก์ชันจากข้อมูลที่มี

ค่าคลาดเคลื่อนโดยใช้องค์ความรู้จากอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน และในปีเดียวกัน ดร. บริบูรณ์ เนาวประทีป และ ดร. กรรณิการ ขำพึงสน ได้ศึกษาปัญหาการจำลองแบบเชิงตัวเลขของการประมาณค่าของฟังก์ชันจากข้อมูลที่มีค่าคลาดเคลื่อนโดยใช้องค์ความรู้จากอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนในปริภูมิรีโพรดิวซ์เคอร์เนลฮิลเบิร์ต H [13] กล่าวคือกำหนดให้ H เป็นปริภูมิรีโพรดิวซ์เคอร์เนลฮิลเบิร์ต และ $S = \{(t_j, d_j) : j \in \square_n\} \subseteq \Lambda \times \square$ เมื่อ $\square_n = \{1, 2, \dots, n\}$ โดยที่เซต S เป็นตัวแทนของข้อมูล และจากข้อมูลดังกล่าว หากต้องการประมาณค่าของข้อมูลที่จุด t_0 ดังนั้น กำหนดให้ I เป็นฟังก์ชันจาก H ไปยัง $\square$ และ

$$If = (f(t_j) : j \in \square_n)$$

ให้ $\varepsilon > 0$ และ $\delta > 0$ และ $I(t_0, \varepsilon, \delta) = \{f(t_0) : \|d - If\| \leq \varepsilon, \|f\| \leq \delta\}$ เมื่อ $\square$ คือนอร์มใน $\square$

จะเห็นได้ชัดเจนว่าตัวประมาณค่าที่ดีที่สุดในการประมาณค่าของฟังก์ชันที่จุด t_0 คือจุดกึ่งกลางของช่วง $I(t_0, \varepsilon, \delta)$ เมื่อเราทราบค่าของฟังก์ชันบางค่า ซึ่งได้ถูกแสดงในเซต S และ เราเรียกวิธีการนี้ว่า ขั้นตอนวิธีค่ากึ่งกลาง (midpoint algorithm) องค์ความรู้จากอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนเป็นเครื่องมือสำคัญที่ใช้หาจุดกึ่งกลางของ $I(t_0, \varepsilon, \delta)$ นอกจากนี้ยังทราบว่าตัวประมาณค่าที่ดีที่สุดหรือค่ากึ่งกลางอยู่ในรูปของ

$$f(t_0) = \sum_{j \in \square_n} c_j K(t_j, t_0)$$

ซึ่งจะเห็นได้ว่ามีรูปแบบเดียวกับ วิธี Tikhonov regularization แต่สัมประสิทธิ์นั้นแตกต่างกันในกรณีทั่วไป นอกจากนี้ยังพบว่าในการทดลองดังกล่าวแสดงให้เห็นว่าขั้นตอนวิธีค่ากึ่งกลาง (midpoint algorithm) ซึ่งได้นำองค์ความรู้จากอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนมาใช้นั้น ให้ผลการทดลองที่ดีกว่าวิธี Tikhonov regularization ซึ่งเป็นที่นิยมในปัจจุบัน

ด้วยเหตุนี้ ผู้วิจัยจึงให้ความสนใจที่จะขยายและพัฒนาอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน ให้สามารถนำไปใช้ในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน เพื่อใช้ในการแก้ปัญหาได้อย่างกว้างขวางและมีประสิทธิภาพมากยิ่งขึ้น รวมทั้งสร้างทฤษฎีต่างๆ ที่เป็นประโยชน์ในการประยุกต์ใช้ได้กับข้อมูลจริง และต่อยอดงานวิจัยทางด้านวิทยาศาสตร์สาขาอื่นๆ และนอกจากนี้ในงานวิจัยนี้ยังมีความประสงค์จะนำผลการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน โดยใช้วิธีการจากอสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนมาเปรียบเทียบกับวิธีการอื่นที่ถูกนำมาศึกษาในปัญหานี้เช่นกัน ซึ่งจะเป็นประโยชน์ต่อผู้ใช้งานในการพิจารณาเลือกไปใช้ให้เหมาะสมกับข้อมูลมากที่สุด จะเห็นได้ว่างานวิจัยที่เสนอข้อทูลในโครงการนี้มีประโยชน์อย่างมากในสาขาวิชาทางวิทยาศาสตร์อย่างเห็นได้ชัด ดังนั้นผลของงานวิจัยที่ได้ผู้วิจัยคาดว่าจะได้รับการตอบรับให้ตีพิมพ์ในวารสารทางวิชาการระดับนานาชาติ ตามวัตถุประสงค์ที่ผู้ให้ทุนตั้งไว้ได้อย่างแน่นอน และยังสามารถนำไปเผยแพร่ในเวทีวิจัยระดับนานาชาติ อันจะนำไปสู่การพัฒนาทางวิชาการด้านคณิตศาสตร์ไทยในเวทีวิจัยระดับนานาชาติต่อไป

บทที่ 2 ทบทวนวรรณกรรม

ในหัวข้อนี้ เราจะกล่าวถึงความรู้เบื้องต้น นิยาม ความหมาย และสัญลักษณ์ต่าง ๆ ของ อสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน ซึ่งผู้วิจัยให้ความสนใจที่จะขยายและพัฒนา อสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนให้สามารถนำไปใช้ในการศึกษาการเรียนรู้ของ เครื่องโดยใช้เคอร์เนลเป็นฐาน ในการแก้ปัญหาได้กว้างขวางขึ้นและมีประสิทธิภาพมากขึ้น

กำหนดให้ H เป็นปริภูมิฮิลเบิร์ต ซึ่งผลคูณภายใน และนอร์ม (norm) นิยามโดย $\langle \cdot, \cdot \rangle$ และ $\|\cdot\|$ ตามลำดับ กำหนดให้เวกเตอร์ $x_1, \dots, x_n$ เป็นเวกเตอร์อิสระเชิงเส้นใน H และกำหนดให้ Q เป็นตัวดำเนินการเชิงเส้น (linear operator) จาก H ไปยัง $\mathbb{R}^n$ และสำหรับทุก ๆ $x \in H$

$$Qx = (\langle x, x_j \rangle : j \in \{1, \dots, n\})$$

และกำหนดให้ Q^T เป็นตัวดำเนินการผกผันของ Q (adjoint operator) จาก $\mathbb{R}^n$ ไปยัง H และสำหรับทุก ๆ $a \in \mathbb{R}^n$

$$Q^T(a) = \sum_{j \in \{1, \dots, n\}} a_j x_j$$

เรานิยาม Gram เมทริกซ์ของเวกเตอร์ $x_1, \dots, x_n$ โดย

$$G = QQ^T = \begin{bmatrix} \langle x_1, x_1 \rangle & \cdots & \langle x_1, x_n \rangle \\ \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \cdots & \langle x_n, x_n \rangle \end{bmatrix}$$

นอกจากนี้เราพบว่า Gram เมทริกซ์ของเวกเตอร์ $x_1, \dots, x_n$ เป็นเมทริกซ์บวกแน่นอน (positive definite) กำหนดให้ $d = (d_j : j \in \{1, \dots, n\}) \in \mathbb{R}^n$ เรานิยามเซตระนาบเกิน (hyperplan) [5] คือเซตที่มีสมาชิกสอดคล้องกับเงื่อนไข

$$Qx = d \quad (2.1)$$

นอกจากนี้เรายังทราบว่า เซตระนาบเกินไม่เป็นเซตว่าง เนื่องจาก มีสมาชิก $x(d) = Q^T G^{-1} d$ ใน H ที่สอดคล้องเงื่อนไข (2.1) สำหรับทุก ๆ $d = (d_j : j \in \{1, \dots, n\}) \in \mathbb{R}^n$ เสมอ

เรานิยามเซตไฮเพอร์เซอร์เคิล (hypercircle) [5] คือ เซตที่มีสมาชิกอยู่ในเซตระนาบเกิน และบอลหนึ่งหน่วย B ใน H และเราใช้สัญลักษณ์นี้แทนเซตไฮเพอร์เซอร์เคิล

$$H(d) = \{x : Qx = d, \|x\| \leq 1\}$$

เราทราบว่าเซตไฮเพอร์เซอร์เคิล (hypercircle) จะไม่เป็นเซตว่าง สำหรับทุก ๆ $d = (d_j : j \in \{1, \dots, n\}) \in \mathbb{R}^n$ ก็ต่อเมื่อ $(d, G^{-1}d) \leq 1$ เมื่อ $(\cdot, \cdot)$ คือผลคูณภายในแบบยูคลิดใน $\mathbb{R}^n$ และอสมการไฮเพอร์เซอร์เคิล [5,10,14] แสดงให้เห็นว่า ทุก $x \in H(d)$ และ $x_0 \in H$ จะได้ว่า

$$|\langle x(d), x_0 \rangle - \langle x, x_0 \rangle| \leq \text{dist}(x_0, M) \sqrt{1 - \|x(d)\|^2}$$

นั่นคือ $\langle x(d), x_0 \rangle$ คือตัวประมาณค่าที่ดีที่สุดในการประมาณค่า $\langle x, x_0 \rangle$ เมื่อ $x \in H(d)$ หรือกล่าวได้ว่า $\langle x(d), x_0 \rangle$ คือจุดกึ่งกลางของช่วง $I(x_0) = \{\langle x, x_0 \rangle : x \in H(d)\}$ นั่นเอง นอกจากนี้เรายังทราบว่า $x(d) = Q^T G^{-1} d$ ซึ่งทำให้เห็นว่าตัวประมาณค่าที่ดีที่สุดนั้นไม่ขึ้นอยู่กับเวกเตอร์ x_0

จากนิยามของไฮเพอร์เซอร์เคิล เห็นได้อย่างชัดเจนว่าสมการไฮเพอร์เซอร์เคิลสามารถใช้กับปัญหาการประมาณค่าของข้อมูล ในกรณีที่ทราบข้อมูลอย่างแน่นอน ($Qx = d$) ต่อมาในปี 2010 ดร. กรรณิการ์ ขำพืงสน และ Professor Charles A. Micchelli ได้กำหนดเซตไฮเพอร์อิลลิปส์ (hyperellipse) [12] ขึ้นมา โดยให้ $E = \{e: e \in \mathbb{R}^n, |e| \leq \varepsilon\}$ ซึ่งเป็นเซตที่เป็นตัวแทนการวัดค่าคลาดเคลื่อนสำหรับข้อมูล เมื่อ $\varepsilon > 0$ และ $\square$ คือนอร์มบน $\mathbb{R}^n$ จากนั้นกำหนดให้เซตไฮเพอร์อิลลิปส์คือเซตที่มีสมาชิกสอดคล้องเงื่อนไข $|d - Qx| \leq \varepsilon$ และ x อยู่ใน B บอลหนึ่งหน่วยใน H และเราใช้สัญลักษณ์ต่อไปนี้แทนเซตของไฮเพอร์อิลลิปส์

$$H(d | E) = \{x: Qx - d \in E, \|x\| \leq 1\}$$

จากนิยามเซตของไฮเพอร์เซอร์เคิล และเซตของไฮเพอร์อิลลิปส์ เราพบว่า

$$H(d | E) = \bigcup_{e \in E} H(d + e)$$

ให้ $x_0 \in H$ จุดประสงค์ของเราคือต้องการตัวประมาณค่าที่ดีที่สุดให้การประมาณค่าของ $\langle x, x_0 \rangle$ เมื่อ $x \in H(d | E)$ ในขณะนี้เราจะเห็นได้ว่าข้อมูลที่เรามีนั้นเป็นข้อมูลที่มีค่าคลาดเคลื่อน ($Qx = d + e$ เมื่อ $e \in E$) ซึ่งการหาจุดกึ่งกลางของช่วง

$$I(x_0, d | E) = \{\langle x, x_0 \rangle: x \in H(d | E)\}$$

นั้นขึ้นอยู่กับหลายปัจจัยอาทิเช่น เวกเตอร์ x_0 รวมถึงค่า ε ซึ่งเป็นค่าความคลาดเคลื่อนของข้อมูล จากความสัมพันธ์ระหว่างเซตไฮเพอร์เซอร์เคิลและเซตของไฮเพอร์อิลลิปส์ ทำให้เราทราบว่า $I(x_0, d | E) = \{\langle x, x_0 \rangle: x \in H(d | E)\}$ เป็นเซตปิดและมีขอบเขตดังนั้นเราจึงได้ว่า

$$I(x_0, d | E) = [m_-(x_0, d | E), m_+(x_0, d | E)]$$

เมื่อ $m_+(x_0, d | E) = \{\langle x, x_0 \rangle: x \in H(d | E)\}$ และ $m_-(x_0, d | E) = \{\langle x, x_0 \rangle: x \in H(d | E)\}$

และเราใช้สัญลักษณ์นี้แทนจุดกึ่งกลางของช่วง $I(x_0, d | E)$

$$m(x_0, d | E) = \frac{m_+(x_0, d | E) + m_-(x_0, d | E)}{2}$$

เริ่มต้นในการศึกษาปัญหานี้เราสามารถแสดงได้ว่ามี e_+ และ e_- ใน E ซึ่งทำให้เวกเตอร์ $x_+(d + e_+)$ และ $x_-(d + e_-)$ ใน $H(d | E)$ ซึ่งทำให้

$$\langle x_+(d + e_+), x_0 \rangle = m_+(x_0, d | E) \text{ และ } \langle x_-(d + e_-), x_0 \rangle = m_-(x_0, d | E)$$

หลังจากที่เราได้ e_+ และ e_- เราพบว่ามี $e_0 \in E$ ซึ่ง $\langle x(d + e_0), x_0 \rangle = m(x_0, d | E)$ นอกจากนี้

เรายังทราบว่า $x(d + e_0) = Q^T G^{-1}(d + \lambda e_+ + (1 - \lambda)e_-)$ สำหรับบาง $\lambda \in [0, 1]$

จากที่กล่าวมาข้างต้นจะเห็นได้ชัดเจนว่าถ้าเราทราบ e_+ และ e_- เราจึงจะสามารถคำนวณค่ากึ่งกลางของช่วงได้ แต่เนื่องจากมิติของปริภูมิอิลลิปส์อาจจะไม่ใช้ปริภูมิจำกัด จึงทำให้การคำนวณหาเวกเตอร์ดังกล่าวนี้ทำได้ยาก ดังนั้นเราจึงได้มีการศึกษาถึงวิธีการคำนวณค่า

$m_+(x_0, d | E) = \{\langle x, x_0 \rangle : x \in H(d | E)\}$ และ $m_-(x_0, d | E) = \{\langle x, x_0 \rangle : x \in H(d | E)\}$ และพบว่าถ้า $H(d) \neq \emptyset$ เราจะได้

$$m_+(x_0, d | E) = \min\{\|x_0 - Q^T c\| + \varepsilon \|c\|_* + (c, d) : c \in \mathbb{R}^n\} \quad (2.2)$$

เมื่อ $\|\cdot\|_*$ คือนอร์มสังยุคของ $\|\cdot\|$ บน $\mathbb{R}^n$ (conjugate norm)

จากนิยามของ $m_-(x_0, d | E) = \{\langle x, x_0 \rangle : x \in H(d | E)\}$ เราพบว่า

$$m_-(x_0, d | E) = \{\langle x, x_0 \rangle : x \in H(d | E)\} = -m_+(x_0, -d | E)$$

ดังนั้นเราสามารถหาค่าสมการ (2.2) เพื่อหา $m_+(x_0, d | E)$ และ $m_-(x_0, d | E)$ จากนั้นจึงคำนวณหาจุดกึ่งกลางของช่วงได้ในลำดับต่อไป

นอกจากนี้เรายังได้ศึกษาในกรณีเฉพาะ $E = \{e : e \in \mathbb{R}^n, \|e\|_2 \leq \varepsilon\}$ ซึ่งเป็นเซตที่ตัวแทนการวัดค่าคาดเคลื่อนสำหรับข้อมูล เมื่อ $\varepsilon > 0$ และ $\|\cdot\|_2$ คือนอร์มยุคลิดบน $\mathbb{R}^n$ และเราใช้สัญลักษณ์นี้แทนเซตไอเพอร์อิลลิปส์ [12]

$$H_2(d | E) = \{x : Qx - d \in E, \|x\| \leq 1\}$$

เราพบว่าได้เราเงื่อนไขที่อ่อนกว่าการแสดงสมการ (8.2) กล่าวคือ ถ้า $H_2(d | E)$ มีสมาชิกมากกว่าหนึ่ง

$x_0 \notin M$ และ $\frac{x_0}{\|x_0\|} \in H_2(d | E)$ แล้ว

$$m_+(x_0, d | E) = \min\{\|x_0 - Q^T c\| + \varepsilon \|c\|_2 + (c, d) : c \in \mathbb{R}^n\}$$

การพิสูจน์ในทฤษฎีบทนี้เราได้ประยุกต์ทฤษฎีบทของ Von Neumann Minimax [2,3,6] ซึ่งเป็นทฤษฎีบทที่มีประโยชน์และมีชื่อเสียงมากมาช่วยในการแสดงสมการ (2.2) และเนื่องจาก $c \rightarrow \|x_0 - Q^T c\| + \varepsilon \|c\|_2 + (c, d)$ เป็นฟังก์ชันคอนเวกซ์โดยแท้ (strictly convex function) ดังนั้นเราจึงทราบว่าจะมี $c^* \in \mathbb{R}^n$ เพียงตัวเดียวเท่านั้นที่ทำให้

$$m_+(x_0, d | E) = \|x_0 - Q^T c^*\| + \varepsilon \|c^*\|_2 + (c^*, d)$$

และเมื่อเราหาความชันของฟังก์ชัน (gradient) ดังกล่าวจึงทำให้ได้ว่าเวกเตอร์ c^* สอดคล้องกับสมการ

$$-Q\left(\frac{x_0 - Q^T c^*}{\|x_0 - Q^T c^*\|}\right) + \varepsilon \frac{c^*}{\|c^*\|_2} + d = 0$$

จากนั้นเรากำหนดการวิธีการทำซ้ำ

$$c^{k+1} = (\rho_k G + \varepsilon \tau_k I)^{-1} (\rho_k Qx_0 - \rho_k \tau_k d) \quad (2.3)$$

เมื่อ $\rho_k = \|c^k\|_2$ และ $\tau_k = \|x_0 - Q^T c^k\|$

จากที่ได้กล่าวมาแล้วนั้น ในปี 2010 ดร.บริบูรณ์ เนาวประทีป และ ดร.กรรณิการ ขำพึงสน ได้นำสมการไอเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน มาศึกษาการประมาณค่าของฟังก์ชันในปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต จากข้อมูลที่มีค่าคลาดเคลื่อน [13,19] และทำการจำลองผลโดยวิธีเชิงตัวเลขเพื่อหาวิธีการจากการศึกษาสมการไอเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนมาเปรียบเทียบกับวิธี Tikhonov regularization ซึ่งในการจำลองผลโดยวิธีเชิงตัวเลขดังกล่าว ได้พิจารณาเคอร์เนลต่างๆ ดังนี้

Szegő kernel: ซึ่งกำหนดโดย ทุก ๆ $s, t \in (-1, 1)$

$$K(s, t) = \frac{1}{1 - st}$$

Polynomial kernel: ซึ่งกำหนดโดย ทุก ๆ $s, t \in \mathbb{R}^2$

$$K(s, t) = (1 + \langle s, t \rangle)^4$$

ในการหาค่ากึ่งกลางของช่วงเราใช้โปรแกรม fminunc ใน optimization toolbox ใน Matlab 7.3.0 ในการคำนวณหาเวกเตอร์ c^* รวมทั้งใช้วิธีการทำซ้ำในสมการ (2.3) จากประสบการณ์การทดลองเชิงตัวเลขในการหาเวกเตอร์ c^* พบว่าวิธีการทำซ้ำในสมการ (2.3) ลู่เข้าหา c^* แต่อย่างไรก็ตามงานวิจัยที่ผ่านมาก็ยังไม่สามารถพิสูจน์ได้ว่าวิธีการนี้ลู่เข้า จึงทำให้เห็นว่ามันยังคงเป็นอีกปัญหาหนึ่งที่น่าสนใจของการพัฒนาการศึกษาสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อน และมีประโยชน์อย่างมากสำหรับผู้ที่จะนำไปประยุกต์ใช้ในการจำลองผลโดยใช้วิธีเชิงตัวเลข โดยใช้เคอร์เนลทั้งสองแบบอย่างใดก็ตาม จากที่กล่าวมาข้างต้นนั้น การศึกษาการประมาณค่าของฟังก์ชันจากข้อมูลที่มีค่าคลาดเคลื่อน เราพบว่าวิธีการจากสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนให้ผลการทดลองที่ดีกว่าวิธี Tikhonov regularization

เนื่องจากข้อมูลในปัจจุบันมีความหลากหลายมาก ทั้งรวมถึงข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน ต่อมาในปี 2014 ดร.บริบูรณ์ เนาวประทีป และ ดร.กรรณิการ ขำพึงสน ได้ขยายสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน โดยกำหนดให้ $I \subseteq X$ โดย m แทนจำนวนสมาชิกในเซต I และ $J = N_n \setminus I$ โดย $n - m$ แทนจำนวนสมาชิกในเซต J สำหรับทุก ๆ $e \in \mathbb{R}^n$ เราใช้สัญลักษณ์

$$e_I = (e_i : i \in I) \in \mathbb{R}^m \text{ และ } e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$$

สำหรับทุก $d \in \mathbb{R}^n$ และให้ $\|\cdot\| : \mathbb{R}^{n-m} \rightarrow \mathbb{R}$ เป็นนอร์มบน $\mathbb{R}^{n-m}$ กำหนดให้

$$\mathcal{E} = \{e : e \in \mathbb{R}^n, e_I = 0, \|e_J\| \leq \varepsilon\} \text{ เมื่อ } \varepsilon \text{ คือจำนวนจริงบวก}$$

เซตพาเซียวไฮเพอร์อิลลิปส์ (partial hyperellipse) คือเซตที่มีสมาชิกสอดคล้องเงื่อนไข x อยู่ใน B บอลหนึ่งหน่วยใน H และ เราใช้สัญลักษณ์ต่อไปนี้แทนเซตของไฮเพอร์อิลลิปส์ $Qx - d \in \mathcal{E}$

$$H(d | \mathcal{E}) = \{x : Qx - d \in \mathcal{E}, \|x\| \leq 1\}$$

[13] จากนิยามข้างต้นจะได้ว่า สำหรับทุก ๆ $x \in H(d | \mathcal{E})$

$$(Qx - d)_I = 0 \text{ และ } \|(Qx - d)_J\| \leq \varepsilon$$

นอกจากนี้เรายังพบว่า

$$H(d | \mathcal{E}) = H(d_I) \cap H(d_J | \mathcal{E}_J)$$

เมื่อเรากำหนดให้ $H(d_I) = \{x : Q_I x = d_I, \|x\| \leq 1\}$ เมื่อ $Q_I x = (\langle x, x_j \rangle : j \in I)$ และ

$H(d_J | \mathcal{E}) = \{x : Q_J x - d_J \in \mathcal{E}_J, \|x\| \leq 1\}$ เมื่อ $\mathcal{E}_J = \{e : e \in \mathbb{R}^{n-m}, \|e\| \leq \varepsilon\}$ และ

$$Q_J x = (\langle x, x_j \rangle : j \in J)$$

ให้ $x_0 \in H$ จุดประสงค์ของเราคือต้องการตัวประมาณค่าที่ดีที่สุดให้การประมาณค่าของ $\langle x, x_0 \rangle$ เมื่อ $x \in H(d | \mathcal{E})$ ในขณะนี้เราจะเห็นได้ว่าข้อมูลที่เรามีนั้นเป็นข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน ($\| (Qx - d)_J \| \leq \varepsilon$) ซึ่งการหาจุดกึ่งกลางของช่วง

$$I(x_0, d | \mathcal{E}) = \{ \langle x, x_0 \rangle : x \in H(d | \mathcal{E}) \}$$

จากความสัมพันธ์ระหว่างเซตไฮเพอร์เซอร์เคิลและเซตของไฮเพอร์อิลลิปส์ ทำให้เราทราบว่า

$$I(x_0, d | \mathcal{E}) = \{ \langle x, x_0 \rangle : x \in H(d | \mathcal{E}) \}$$

เป็นเซตปิดและมีขอบเขตดังนั้นเราจึงได้ว่า

$$I(x_0, d | \mathcal{E}) = [m_-(x_0, d | \mathcal{E}), m_+(x_0, d | \mathcal{E})]$$

เมื่อ $m_+(x_0, d | \mathcal{E}) = \max\{ \langle x, x_0 \rangle : x \in H(d | \mathcal{E}) \}$ และ $m_-(x_0, d | \mathcal{E}) = \min\{ \langle x, x_0 \rangle : x \in H(d | \mathcal{E}) \}$

และเราใช้สัญลักษณ์นี้แทนจุดกึ่งกลางของช่วง $I(x_0, d | \mathcal{E})$

$$m(x_0, d | \mathcal{E}) = \frac{m_+(x_0, d | \mathcal{E}) + m_-(x_0, d | \mathcal{E})}{2}$$

ในทำนองเดียวกันในการพิสูจน์จาก [11] เราสามารถแสดงได้ว่ามี e_+ และ e_- ใน $\mathcal{E}$ ซึ่งทำให้เวกเตอร์ $x_+(d + e_+)$ และ $x_-(d + e_-)$ ใน $H(d | \mathcal{E})$ ซึ่งทำให้

$$\langle x_+(d + e_+), x_0 \rangle = m_+(x_0, d | \mathcal{E}) \text{ และ } \langle x_-(d + e_-), x_0 \rangle = m_-(x_0, d | \mathcal{E})$$

นอกจากนี้ในงานวิจัยนี้เรายังศึกษาการหาวิธีการหาค่าจุดปลายของช่วงดังกล่าวอีกวิธีหนึ่งดังต่อไปนี้

ถ้า $H(d | \mathcal{E})$ มีสมาชิกมากกว่าหนึ่งเราจะได้ว่า

$$m_+(x_0, d | \mathcal{E}) = \min\{ \|x_0 - Q^T c\| + \varepsilon \|c_J\|_* + (c, d) : c \in \square^n \} \quad \text{เมื่อ}$$

$\|\square\|_*$ คือนอร์มสังยุคของ $\|\square\|$ บน $\square^{n-m}$ (conjugate norm)

ถ้า $x_0 \notin M_I := \{Q_I^T(a) : a \in \square^m\}$ และ $H(d_I)$ มีสมาชิกมากกว่าหนึ่ง แล้วจะได้ว่า

$$c^* = \arg \min\{ \|x_0 - Q^T c\| + \varepsilon \|c_J\|_* + (c, d) : c \in \square^n \} \text{ และ } c_J^* = 0 \text{ ก็ต่อเมื่อ}$$

$$\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} \in H(d | \mathcal{E}) \text{ เมื่อ } \frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} = \min\{ \|x_0 - Q_I^T(a)\| + (a, d_I) : a \in \square^m \}$$

$$\text{และ } Q_I^T(a) = \sum_{j \in I} a_j x_j \text{ นอกจากนี้ในกรณีนี้ค่า } m_+(x_0, d | \mathcal{E}) = \|x_0 - Q_I^T a^*\| + (a^*, d_I)$$

ถ้า $x_0 \notin M_J := \{Q_J^T(b) : b \in \square^{n-m}\}$ และ $H(d_J | \mathcal{E})$ มีสมาชิกมากกว่าหนึ่ง แล้วจะได้ว่า

$$c^* = \arg \min\{ \|x_0 - Q^T c\| + \varepsilon \|c_J\|_* + (c, d) : c \in \square^n \} \text{ และ } c_I^* = 0$$

$$\text{ก็ต่อเมื่อ } \frac{x_0 - Q_J^T(b^*)}{\|x_0 - Q_J^T(b^*)\|} \in H(d | \mathcal{E}) \text{ เมื่อ}$$

$$\frac{x_0 - Q_J^T(b^*)}{\|x_0 - Q_J^T(b^*)\|} = \min\{ \|x_0 - Q_J^T(b)\| + \varepsilon \|b\|_* + (b, d_J) : b \in \square^{n-m} \}$$

และ $Q_J^T(a) = \sum_{j \in J} a_j x_j$ นอกจากนี้ในกรณีนี้ค่า

$$m_+(x_0, d | \mathcal{E}) = \|x_0 - Q_J^T(b^*)\| + \varepsilon \|b^*\|_* + (b^*, d_J)$$

ในการวิจัยในครั้งนี้ยังคงแสดงได้ว่าตัวประมาณค่าที่ดีที่สุดในการประมาณค่าของ $\langle x, x_0 \rangle$ เมื่อ $x \in H(d | \mathcal{E})$ ยังคงอยู่ในรูปของผลรวมเชิงเส้นของเวกเตอร์ในเซต X โดยเราพบว่า $x(d + e_0) = Q^T G^{-1}(d + \lambda e_+ + (1 - \lambda)e_-)$ สำหรับบาง $\lambda \in [0, 1]$

บทที่ 3 วัตถุประสงค์ของโครงการ

1. ขยายองค์ความรู้ของอสมการไฮเพอร์เซอร์เคลสู่ข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน
2. ประยุกต์องค์ความรู้ใหม่เกี่ยวกับปัญหาการประมาณค่าของฟังก์ชันจากข้อมูลจริงที่มีค่าคลาดเคลื่อนบางส่วน โดยทำการจำลองผลทางคณิตศาสตร์โดยวิธีเชิงตัวเลข

บทที่ 4 ผลการดำเนินงาน

การทำงานวิจัยเป็นไปตามแผนที่วางไว้ โดยได้รวบรวมและศึกษาสมการไฮเพอร์เซอร์เคิล สำหรับข้อมูลที่มีค่าคลาดเคลื่อน รวมถึงขยายผลการตั้งกล่าวให้ใช้ได้กับข้อมูลที่หลากหลายมากขึ้น ซึ่งเป็นไปตามจุดประสงค์ของการทำวิจัยในครั้งนี้ ซึ่งผลการดำเนินงานได้นำเสนอไว้ดังต่อไปนี้

จากจุดประสงค์ข้อหนึ่งของโครงการคือ ขยายองค์ความรู้ของสมการไฮเพอร์เซอร์เคิลสู่ข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน โดยมีจุดเริ่มต้นโดยการที่ผู้วิจัยจึงได้ทำการศึกษาสมการไฮเพอร์เซอร์เคิล สำหรับข้อมูลที่มีค่าคลาดเคลื่อนบางส่วนโดยค่าคลาดเคลื่อนอยู่ใน $\square$ และ $\square^2$ กำหนดให้ H เป็นปริภูมิฮิลเบิร์ต ซึ่งผลคูณภายใน และนอร์ม (norm) นิยามโดย $\langle \square, \square \rangle$ และ $\|\square\|$ ตามลำดับ

กำหนดให้ $X = \{x_1, \dots, x_n\}$ เป็นเซตย่อยอิสระเชิงเส้นใน H และ $I = \{1, 2, \dots, n-1\}$ และ $J = \{n\}$ สำหรับทุก $e \in R^n$ เราใช้สัญลักษณ์

$$e_I = (e_i : i \in I) \in R^{n-1}$$

สำหรับทุก $d \in R^n$ และให้ $|\cdot| : R \rightarrow R$ เป็นสับมูลติไพลิน R กำหนดให้

$$\mathcal{E} = \{e : e \in \square^n, e_I = 0, |e_n| \leq \varepsilon\} \text{ เมื่อ } \varepsilon \text{ คือจำนวนจริงบวก}$$

เซตพาเซียวไฮเพอร์อิลลิปส์ (partial hyperellipse) คือเซตที่มีสมาชิกสอดคล้องเงื่อนไข x อยู่ใน B บอลหนึ่งหน่วยใน H และ เราใช้สัญลักษณ์ต่อไปนี้แทนเซตของไฮเพอร์อิลลิปส์ $Qx - d \in \mathcal{E}$

$$H(d | \mathcal{E}) = \{x : Qx - d \in \mathcal{E}, \|x\| \leq 1\}$$

จากการศึกษาก่อนหน้านี้พบว่า $H(d | \mathcal{E}) \neq \emptyset$ ก็ต่อเมื่อ

$$\min_{e \in \mathcal{E}} (d + e, G^{-1}(d + e)) \leq 1 \quad (4.1)$$

ในวิจัยนี้ผู้วิจัยได้ศึกษารูปแบบของคำตอบ (1) โดยใช้เริ่มต้นจากใช้กระบวนการ Gram-Schmidt เปลี่ยนเซต X เป็นเซตเชิงตั้งฉาก (orthonormalized) นอกจากนี้พบว่าสำหรับทุก $x \in H(d)$ เงื่อนไข

$$Qx = d \quad \text{สมนัยกับ} \quad Rx = d^*$$

เมื่อตัวดำเนินการเชิงเส้น $R : H \rightarrow \square^n$ โดยกำหนดโดย $R(x) := (\langle x, x_j^* \rangle : j \in \square_n)$ และเวกเตอร์ $d^* \in \square^n$ ถูกกำหนดโดย

$$d_1^* = \frac{d_1}{\sqrt{|G(x_1)|}}$$

สำหรับ $k = 2, \dots, n$

$$d_k^* = \frac{1}{\sqrt{|G(x_1, x_2, \dots, x_{k-1})| |G(x_1, x_2, \dots, x_k)|}} \begin{vmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \dots & \langle x_k, x_1 \rangle \\ \vdots & \vdots & & \vdots \\ \langle x_1, x_{k-1} \rangle & \langle x_2, x_{k-1} \rangle & \dots & \langle x_k, x_{k-1} \rangle \\ d_1 & d_2 & \dots & d_k \end{vmatrix}$$

จากสมการข้างต้นผู้วิจัยได้กำหนดเมทริกซ์ A ซึ่งเป็นตัวแทนสัมประสิทธิ์ของ d_k^* สำหรับทุก $k = 1, 2, \dots, n$ ซึ่งกำหนดโดย

$$A = \begin{bmatrix} \beta_{11} & 0 & \dots & 0 \\ \beta_{21} & \beta_{22} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{bmatrix}$$

นอกจากนี้ยังพบว่า $\beta_{kk} = \sqrt{\frac{|G(x_1, \dots, x_{k-1})|}{|G(x_1, \dots, x_k)|}}$ สำหรับทุก ๆ $k = 1, 2, \dots, n$ ดังนั้นเราสามารถสรุปได้ว่า

$$H(d) = H^*(d^*) = \{x : x \in B, Rx = d^* = Ad\}$$

และ

$$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) = \{x : Qx - d \in \mathcal{E}(A), \|x\| \leq 1\}$$

เมื่อ $\mathcal{E}(A) = \{Ae : e \in \mathcal{E}\}$ นั่นคือสำหรับทุก ๆ $x \in H(d | \mathcal{E})$

$$Qx = d + e \text{ จะสมนัยกับ } Rx = d^* + Ae$$

หรือสอดคล้อง $R_I(x) = d_I^*$ และ $R_n(x) - d_n^* = \beta_{nn}e_n$

นอกจากนี้ สำหรับทุก ๆ $x(d+e) \in H(d | \mathcal{E})$ จะได้ว่า

$$x(d+e) = x^*(d^* + Ae) = x^*(d_I^*) + x^*(d_n + \beta_{nn}e_n)$$

และได้ว่า

$$\begin{aligned} \|x(d+e)\|^2 &= \|x^*(d^* + Ae)\|^2 = \|x^*(d_I^*)\|^2 + \|x^*(d_n + \beta_{nn}e_n)\|^2 \\ &= (d_I^*, d_I^*) + (d_n + \beta_{nn}e_n)^2 \end{aligned}$$

ดังนั้นเราจึงสามารถสรุปได้ว่า

$$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) \neq \emptyset \text{ ก็ต่อเมื่อ}$$

$$\min_{|e| \leq \varepsilon} (d_n^* + \beta_{nn}e)^2 \leq 1 - (d_I^*, d_I^*) \quad (4.2)$$

ในงานวิจัยนี้ผู้วิจัยได้ศึกษาลักษณะของคำตอบ (1)

$$\min_{|e| \leq \varepsilon} (d_n^* + \beta_{nn}e)^2 = \begin{cases} 0, & \left| \frac{d_n^*}{\beta} \right| \leq \varepsilon \\ d_n^* + \beta_{nn}\varepsilon, & \left| \frac{d_n^*}{\beta} \right| > \varepsilon \end{cases}$$

ให้ $x_0 \in H$ เราต้องการประมาณค่า $\langle x, x_0 \rangle$ ที่ดีที่สุด เมื่อ $Q_I x = d_I + e$ และ $|Q_n x - d_n| \leq \varepsilon$
จากงานวิจัยก่อนหน้านี้ผู้วิจัยพบว่า

ถ้า $x_0 \notin M_n := \{Q_n^T(b) : b \in \square\}$ และ $H(d_n | E)$ มีสมาชิกมากกว่าหนึ่ง แล้วจะได้ว่า

$$c^* = \arg \min \{\|x_0 - Q^T c\| + \varepsilon \|c_n\| + (c, d) : c \in \square^n\} \text{ และ } c_n^* = 0$$

ก็ต่อเมื่อ

$$\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} \in H(d | E)$$

$$\text{เมื่อ } \frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} = \min \{\|x_0 - Q_I^T(a)\| + (a, d_I) : a \in \square^m\} \text{ และ } Q_I^T(a) = \sum_{j \in I} a_j x_j$$

นอกจากนี้ในกรณีนี้ค่า $m_+(x_0, d | E) = \|x_0 - Q_I^T a^*\| + (a^*, d_I)$

ถ้า $\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} \notin H(d | E)$ จะได้ว่า จะมีเวกเตอร์ $c^* \in \square^n$ เพียงตัวเดียวที่ทำให้

$$-Q \left(\frac{x_0 - Q^T c^*}{\|x_0 - Q^T c^*\|} \right) + \varepsilon w^* + d$$

$$\text{เมื่อเวกเตอร์ } w^* = \begin{cases} 0, & i \in I \\ \text{sgn}(c_n^*), & i = n \end{cases}$$

ดังนั้น

$$x_+(d) = \arg \max \{\langle x, x_0 \rangle : H(d_I) \cap H(d_n \pm \varepsilon)\} \quad (4.3)$$

ในลำดับต่อไปผู้วิจัยกำหนดให้ $I \subseteq X$ โดย $n-2$ แทนจำนวนสมาชิกในเซต I และ $J = N_n \setminus I$ โดย 2 แทนจำนวนสมาชิกในเซต J สำหรับทุก ๆ $e \in R^n$ เราใช้สัญลักษณ์

$$e_I = (e_i : i \in I) \in R^{n-2} \text{ และ } e_J = (e_i : i \in J) \in R^2$$

สำหรับทุก $d \in R^n$ และให้ $\|\cdot\|_2 : R^2 \rightarrow R$ เป็นยูคลิดเนียนนอร์มบน R^2 กำหนดให้

$$E = \{e : e \in \square^n, e_I = 0, \|e_J\|_2 \leq \varepsilon\} \text{ เมื่อ } \varepsilon \text{ คือจำนวนจริงบวก}$$

เซตพาเซียวไฮเพอร์อิลลิปส์ (partial hyperellipse) คือเซตที่มีสมาชิกสอดคล้องเงื่อนไข x อยู่ใน B บอลหนึ่งหน่วยใน H และ เราใช้สัญลักษณ์ต่อไปนี้แทนเซตของไฮเพอร์อิลลิปส์ $Qx - d \in E$

$$H(d | E) = \{x : Qx - d \in E, \|x\| \leq 1\}$$

จากการศึกษาก่อนหน้านี้พบว่า $H(d | E) \neq \emptyset$ ก็ต่อเมื่อ

$$\min_{e \in E} (d + e, G^{-1}(d + e)) \leq 1$$

ในวิจัยนี้ผู้วิจัยได้ศึกษารูปแบบของคำตอบ (1) โดยใช้เริ่มต้นจากใช้กระบวนการ Gram-Schmidt เปลี่ยนเซต X เป็นเซตเชิงตั้งฉาก (orthonormalized) นอกจากนี้พบว่าสำหรับทุก ๆ $x \in H(d)$ เงื่อนไข

$$Qx = d \quad \text{สมนัยกับ} \quad Rx = d^*$$

เมื่อตัวดำเนินการเชิงเส้น $R : H \rightarrow \square^n$ โดยกำหนดโดย $R(x) := (\langle x, x_j^* \rangle : j \in \square_n)$ และเวกเตอร์

$d^* \in \square^n$ ถูกกำหนดโดย

$$d_1^* = \frac{d_1}{\sqrt{|G(x_1)|}}$$

สำหรับ $k = 2, \dots, n$

$$d_k^* = \frac{1}{\sqrt{|G(x_1, x_2, \dots, x_{k-1})| |G(x_1, x_2, \dots, x_k)|}} \begin{vmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \dots & \langle x_k, x_1 \rangle \\ \vdots & \vdots & & \vdots \\ \langle x_1, x_{k-1} \rangle & \langle x_2, x_{k-1} \rangle & \dots & \langle x_k, x_{k-1} \rangle \\ d_1 & d_2 & \dots & d_k \end{vmatrix}$$

จากสมการข้างต้นผู้วิจัยได้กำหนดเมทริกซ์ A ซึ่งเป็นตัวแทนสัมประสิทธิ์ของ d_k^* สำหรับทุก ๆ $k = 1, 2, \dots, n$ ซึ่งกำหนดโดย

$$A = \begin{bmatrix} \beta_{11} & 0 & \dots & 0 \\ \beta_{21} & \beta_{22} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{bmatrix}$$

นอกจากนี้ยังพบว่า $\beta_{kk} = \sqrt{\frac{|G(x_1, \dots, x_{k-1})|}{|G(x_1, \dots, x_k)|}}$ สำหรับทุก ๆ $k = 1, 2, \dots, n$ ดังนั้นเราสามารถสรุปได้ว่า

$$H(d) = H^*(d^*) = \{x : x \in B, Rx = d^* = Ad\}$$

และ

$$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) = \{x : Qx - d \in \mathcal{E}(A), \|x\| \leq 1\}$$

เมื่อ $\mathcal{E}(A) = \{Ae : e \in \mathcal{E}\}$ นั่นคือสำหรับทุก ๆ $x \in H(d | \mathcal{E})$

$$Qx = d + e \quad \text{จะสมนัยกับ} \quad Rx = d^* + Ae$$

หรือสอดคล้อง $R_I(x) = d_I^*$ และ $R_J(x) - d_J^* = Ae_J$ เมื่อ A_J คือ 2×2 เมทริกซ์ซึ่งกำหนดโดย

$$A_J = \begin{bmatrix} \beta_{m+1m+1} & 0 \\ \beta_{nm+1} & \beta_{nn} \end{bmatrix}$$

นอกจากนี้ สำหรับทุก ๆ $x(d+e) \in H(d | \mathcal{E})$ จะได้ว่า

$$x(d+e) = x^*(d^* + Ae) = x^*(d_I^*) + x^*(d_J + A_J e_J)$$

และได้ว่า

$$\begin{aligned} \|x(d+e)\|^2 &= \|x^*(d^* + Ae)\|^2 = \|x^*(d_I^*)\|^2 + \|x^*(d_J + A_J e_J)\|^2 \\ &= (d_I^*, d_I^*) + (d_J + A_J e_J, d_J + A_J e_J) \end{aligned}$$

ดังนั้นเราจึงสามารถสรุปได้ว่า

$$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) \neq \emptyset \quad \text{ก็ต่อเมื่อ}$$

$$\min_{e \in \mathcal{E}} (d_J^* + A_J e_J, d_J^* + A_J e_J) \leq 1 - (d_I^*, d_I^*) \quad (4.4)$$

ในงานวิจัยนี้ผู้วิจัยได้ศึกษาลักษณะของคำตอบ (4.4) โดยอ้างอิงจากงานวิจัย [9] และทำการศึกษาวิธีการเดียวกันกับในกรณีที่มีค่าตลาดเคลื่อนเพียวตัวเดียวจะสามารถหาค่ากลางในกรณีนี้ด้วยเมื่อผู้วิจัยเลือก l^∞ นอร์มเป็นเครื่องมือวัดค่าตลาดเคลื่อน

ต่อมาผู้วิจัยนำองค์ความรู้ข้างต้นมาประยุกต์ใช้การประมาณค่าของฟังก์ชันในปริภูมิรีโพรดิวซิงเคอร์เนล ซึ่งผู้วิจัยได้เลือก $H^2(\Delta)$ คือเซตของ analytic function บนบอลปิดหนึ่งหน่วย และ

Szegő kernel: ซึ่งกำหนดโดย ทุก ๆ $s, t \in (-1, 1)$

$$K(s, t) = \frac{1}{1 - st}$$

โดยกำหนดให้ $T = \{t_1, t_2, \dots, t_n\} \subseteq (0, 1)$ และจะได้ว่า $\{K_{t_i} : i \in \square_n\}$ เป็นเวกเตอร์อิสระเชิงเส้นใน $H^2(\Delta)$ เมื่อ $K_{t_i}(t) = \frac{1}{1 - t_i t}$, $i \in \square_n$ กระบวนการ Gram-Schmidt เปลี่ยนเซต

$\{K_{t_i} : i \in \square_n\}$ เป็นเซตเชิงตั้งฉาก (orthonormalized) โดย

$$K_1^* = \sqrt{1 - t_1^2} K_{t_1}$$

สำหรับ $k = 2, 3, \dots, n$

$$K_k^* = \sqrt{1 - t_k^2} \sum_{l=1}^k (-1)^{k+l} \frac{\prod_{i \in \square_{k-1}} |1 - t_l t_i|}{\prod_{\substack{i \in \square_k \\ i \neq l}} |t_l - t_i|} K_{t_l}$$

สำหรับทุก $d \in \square^n$ จะได้ว่า

$$Qf = (f(t_j) : j \in \square_n) = d \quad \text{สมนัยกับเงื่อนไข} \quad Rf = (\langle f, K_j^* \rangle : j \in \square_n) = d^*$$

เมื่อ $d_1^* = \sqrt{1 - t_1^2} d_1$ และสำหรับ $k = 2, 3, \dots, n$

$$d_k^* = \sqrt{1 - t_k^2} \sum_{l=1}^k (-1)^{k+l} \frac{\prod_{i \in \square_{k-1}} |1 - t_l t_i|}{\prod_{\substack{i \in \square_k \\ i \neq l}} |t_l - t_i|} d_l$$

และเซตพาเซียวไฮเพอร์อิลลิปส์ถูกกำหนดโดย

$$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) = \{f : Qf - d \in \mathcal{E}(A), \|f\| \leq 1\}$$

เมื่อ $\mathcal{E}(A) = \{Ae : e \in \mathcal{E}\}$ นั่นคือสำหรับทุก ๆ $x \in H(d | \mathcal{E})$

$$Qf = d + e \quad \text{จะสมนัยกับ} \quad Rf = d^* + Ae$$

หรือสอดคล้อง $R_1(f) = d_1^*$ และ $R_n(f) - d_n^* = \beta_{nn} e_n$

ลำดับถัดไปจะประมาณค่าที่จุด $t_0 \in (0,1)$ โดยทราบว่า $R_I(f) = d_I^*$ และ $R_n(f) - d_n^* = \beta_{nn} e_n$ ในการหาค่ากลางซึ่งคือค่าที่ดีที่สุดสำหรับการประมาณค่าเราจะใช้องค์ความรู้จาก (4.1) – (4.3)

จากจุดประสงค์ข้อที่สองของโครงการคือประยุกต์องค์ความรู้ใหม่เกี่ยวกับปัญหาการประมาณค่าของฟังก์ชันจากข้อมูลจริงที่มีค่าคลาดเคลื่อนบางส่วน โดยทำการจำลองผลทางคณิตศาสตร์โดยวิธีเชิงตัวเลข ผู้วิจัยจึงได้ทำการศึกษาสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนบางส่วนซึ่งถูกวัดด้วยยูคลิเดียนนอร์ม กำหนดให้ H เป็นปริภูมิฮิลเบิร์ต ซึ่งผลคูณภายใน และนอร์ม (norm) นิยามโดย $\langle \cdot, \cdot \rangle$ และ $\|\cdot\|$ ตามลำดับ กำหนดให้ $X = \{x_1, \dots, x_n\}$ เป็นเซตย่อยอิสระเชิงเส้นใน H และ $I \subseteq X$ โดย m แทนจำนวนสมาชิกในเซต I และ $J = X \setminus I$ โดย $n-m$ แทนจำนวนสมาชิกในเซต J สำหรับทุก ๆ $e \in R^n$ เราใช้สัญลักษณ์

$$e_I = (e_i : i \in I) \in R^m \text{ และ } e_J = (e_i : i \in J) \in R^{n-m}$$

สำหรับทุก $d \in R^n$ และให้ $\|\cdot\|_2 : R^{n-m} \rightarrow R$ เป็นยูคลิเดียนนอร์มบน R^{n-m} กำหนดให้

$$\mathcal{E} = \{e : e \in R^n, e_I = 0, \|e_J\|_2 \leq \varepsilon\} \text{ เมื่อ } \varepsilon \text{ คือจำนวนจริงบวก}$$

เซตพาเซียวไฮเพอร์อิลลิปส์ (partial hyperellipse) คือเซตที่มีสมาชิกสอดคล้องเงื่อนไข x อยู่ใน B บอลหนึ่งหน่วยใน H และ เราใช้สัญลักษณ์ต่อไปนี้แทนเซตของไฮเพอร์อิลลิปส์ $Qx - d \in \mathcal{E}$

$$H(d | \mathcal{E}) = \{x : Qx - d \in \mathcal{E}, \|x\| \leq 1\}$$

[12] จากนิยามข้างต้นจะได้ว่า สำหรับทุก ๆ $x \in H(d | \mathcal{E})$

$$(Qx - d)_I = 0 \text{ และ } \|(Qx - d)_J\|_2 \leq \varepsilon$$

นอกจากนี้เรายังพบว่า

$$H(d | \mathcal{E}) = H(d_I) \cap H(d_J | \mathcal{E}_J)$$

เมื่อเรากำหนดให้ $H(d_I) = \{x : Q_I x = d_I, \|x\| \leq 1\}$ เมื่อ $Q_I x = (\langle x, x_j \rangle : j \in I)$ และ

$H(d_J | \mathcal{E}) = \{x : Q_J x - d_J \in \mathcal{E}_J, \|x\| \leq 1\}$ เมื่อ $\mathcal{E}_J = \{e : e \in R^{n-m}, \|e\|_2 \leq \varepsilon\}$ และ

$$Q_J x = (\langle x, x_j \rangle : j \in J)$$

ให้ $x_0 \in H$ จุดประสงค์ของเราคือต้องการตัวประมาณค่าที่ดีที่สุดให้การประมาณค่าของ $\langle x, x_0 \rangle$ เมื่อ $x \in H(d | \mathcal{E})$ ในขณะนี้เราจะเห็นได้ว่าข้อมูลที่เรามีนั้นเป็นข้อมูลที่มีค่าคลาดเคลื่อนบางส่วน $(\|(Qx - d)_J\|_2 \leq \varepsilon)$ ซึ่งการหาจุดกึ่งกลางของช่วง

$$I(x_0, d | \mathcal{E}) = \{\langle x, x_0 \rangle : x \in H(d | \mathcal{E})\}$$

จากความสัมพันธ์ระหว่างเซตไฮเพอร์เซอร์เคิลและเซตของไฮเพอร์อิลลิปส์ ทำให้เราทราบว่า

$$I(x_0, d | \mathcal{E}) = \{\langle x, x_0 \rangle : x \in H(d | \mathcal{E})\}$$

เป็นเซตปิดและมีขอบเขตดังนั้นเราจึงได้ว่า

$$I(x_0, d | \mathcal{E}) = [m_-(x_0, d | \mathcal{E}), m_+(x_0, d | \mathcal{E})]$$

เมื่อ $m_+(x_0, d | \mathcal{E}) = \max\{\langle x, x_0 \rangle : x \in H(d | \mathcal{E})\}$ และ $m_-(x_0, d | \mathcal{E}) = \min\{\langle x, x_0 \rangle : x \in H(d | \mathcal{E})\}$

และเราใช้สัญลักษณ์นี้แทนจุดกึ่งกลางของช่วง $I(x_0, d | \mathcal{E})$

$$m(x_0, d | \mathcal{E}) = \frac{m_+(x_0, d | \mathcal{E}) + m_-(x_0, d | \mathcal{E})}{2}$$

ในกรณีพบว่าการหาวิธีการหาค่าจุดปลายของช่วงคือ

ถ้า $H(d | \mathcal{E})$ มีสมาชิกมากกว่าหนึ่งเราจะได้ว่า

$$m_+(x_0, d | \mathcal{E}) = \min\{\|x_0 - Q^T c\| + \varepsilon \mid \|c\|_2 \leq 1, c \in \mathbb{R}^n\}$$

ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H ประกอบด้วยสมาชิกของฟังก์ชัน K จากเซต Λ ไปยังจำนวนจริง $\mathbb{R}$ เมื่อ Λ เป็นเซตย่อยของจำนวนจริง และ K เป็นฟังก์ชันของ s และ t เมื่อ $s, t \in \Lambda$ จะถูกเรียกว่า

รีโพรดิวซิงเคอร์เนล (Reproducing kernel) ถ้าสำหรับทุกๆ $t \in \Lambda$ และ ทุกๆ ฟังก์ชัน $f \in H$ แล้วให้

$$f(t) = \langle f, K_t \rangle$$

เมื่อ $\langle \cdot, \cdot \rangle$ แทนผลคูณภายใน และ K_t แทนฟังก์ชันจากเซต Λ ไปยังจำนวนจริง $\mathbb{R}$ และกำหนดให้ ทุกๆ $s \in \Lambda$, $K_t(s) = K(s, t)$ เรานิยาม **ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H** คือปริภูมิที่ประกอบด้วยสมาชิกของฟังก์ชันจากเซต Λ ไปยังจำนวนจริง $\mathbb{R}$ เมื่อ Λ เป็นเซตย่อยของจำนวนจริง และมีฟังก์ชัน K ของ s และ t เมื่อ ทุกๆ $s, t \in \Lambda$ เป็นรีโพรดิวซิงเคอร์เนลสำหรับปริภูมิฮิลเบิร์ตนั้น ในปี 1950 Aronszajn and Moore ได้ค้นพบทฤษฎีบท กล่าวคือ ฟังก์ชัน K ข้างต้นจะเป็น รีโพรดิวซิงเคอร์เนล ก็ต่อเมื่อ สำหรับทุกๆ $\{t_1, t_2, \dots, t_n\} \subseteq \Lambda$ และ เมทริกซ์ $K(t_i, t_j)$ ซึ่งกำหนดโดย

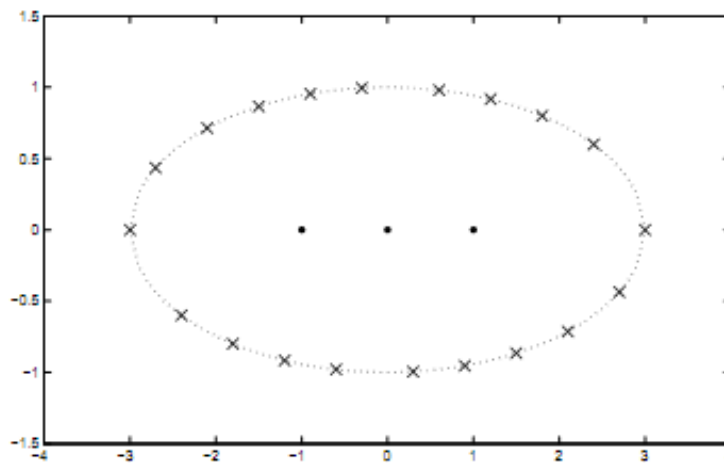
$$K(t_i, t_j) = \begin{bmatrix} K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \vdots & \vdots & \vdots \\ K(t_n, t_1) & \cdots & K(t_n, t_n) \end{bmatrix}$$

เป็นเมทริกซ์กึ่งบวกแน่นอน (positive semi-definite) ซึ่งนับได้ว่าเป็นทฤษฎีบทที่มีประโยชน์มากในการศึกษาการเรียนรู้ของเครื่องโดยใช้เคอร์เนลเป็นฐาน นอกจากนี้ยังพบว่า ทุกๆ รีโพรดิวซิงเคอร์เนล K จะมี ปริภูมิรีโพรดิวซิงเคอร์เนลฮิลเบิร์ต H เพียงปริภูมิเดียวที่สมนัยกับรีโพรดิวซิงเคอร์เนล K

ในลำดับต่อไปผู้วิจัยของเสนอผลการทดลองโดยวิธีเชิงตัวเลขเพื่อนำวิธีการจากการศึกษาการประมาณค่าของฟังก์ชันจากข้อมูลที่มีค่าคลาดเคลื่อนโดยวิธีค่ากึ่งกลาง (Midpoint Algorithm) มาเปรียบเทียบกับวิธี Tikhonov regularization ซึ่งผู้วิจัยได้ศึกษาในปริภูมิโดยใน Gaussian Kernel 2 มิติ Gaussian kernel: $K(t, s) = e^{t-s^2}$, $t, s \in \Lambda = \mathbb{R}^2$ โดยที่ฟังก์ชันแน่นอนตรง กำหนดโดย

$$g(t) = 3.5K_{(1,1)}(t) + 1.75K_{(1,-1)}(t) + 3.25K_{(-1,-1)}(t) - 3.5K_{(-1,1)}(t)$$

จากภาพที่ 1 เซต $T = \{t_j : j \in N_n\}$ เป็นเซตของข้อมูล 20 ตัว และจุดที่ต้องการประมาณค่าสามจุดซึ่งตรงกลางวงรีดังกล่าว



ภาพที่ 4.1 เซตของข้อมูลที่ต้องการประมาณค่าจากข้อมูลที่มีค่าคลาดเคลื่อน

เริ่มต้นโดยกำหนดให้ ℓ เป็น loss ฟังก์ชันจาก $\mathbb{R}^n$ ไปยังจำนวนจริง $\mathbb{R}$ และ L เป็นฟังก์ชันที่วัดค่าคลาดเคลื่อนจาก H ไปยังจำนวนจริง และกำหนดให้

$$L(f) := \frac{1}{n} \sum_{j \in \mathbb{I}_n} \ell(d_j, f(t_j))$$

ให้ $\rho > 0$ และฟังก์ชันที่เป็นตัวแทนของข้อมูลชุดนี้ถูกเลือกมาจาก

$$L(f_\rho) + \rho \|f_\rho\|^2 = \min\{L(f) + \rho \|f\|^2 : f \in H\}$$

จาก Representer Theorem พบว่า

$$f_\rho(t) = \sum_{j \in \mathbb{I}_n} c(\rho)_j K(t_j, t) \quad \text{เมื่อ } t \in T$$

และ $c(\rho) = (G + \rho I_n)^{-1}$ เมื่อ I_n เป็นเมทริกซ์เอกลักษณ์ ดังนั้น $f_\rho(t_0) = \sum_{j \in \mathbb{I}_n} c(\rho)_j K(t_j, t_0)$ คือ

ค่าประมาณที่ดีที่สุดจากวิธีการ **Tikhonov regularization**

ต่อมาเราจะนำค่าทั้งสองมาใช้ในการพิจารณาวิธีค่ากึ่งกลาง (midpoint algorithm) ให้กำหนดให้ $\varepsilon = \varepsilon_\rho$ และ $\delta = 3\varepsilon_\rho$ จากการกำหนดดังกล่าวจึงได้เซตไฮเพอร์เซอร์เคิล

$$H(d | E(\delta)) = \{x : \|Qx - d\|_2 \leq \varepsilon, \|x\| \leq \delta\}$$

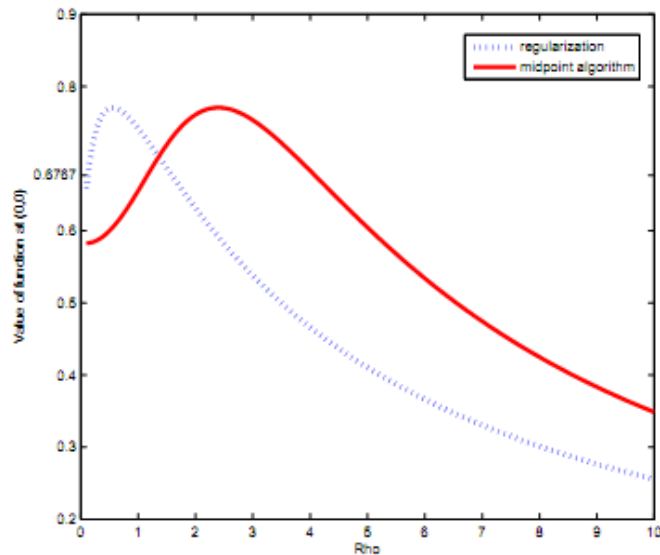
เนื่องจากตัวประมาณค่าที่ดีที่สุดของวิธีค่ากึ่งกลางคือ ค่ากึ่งกลางของช่วง

$$I(x_0, d | E(\delta)) = \{\langle x, x_0 \rangle : x \in H(d | E(\delta))\}$$

ซึ่งสามารถคำนวณหาได้จาก

$$m_+(t_0, \pm d | E(\delta)) = \min_{c \in \mathbb{R}^n} \delta \|K_{t_0} - Q^T c\| + \varepsilon \|c\|_2 \pm (c, d)$$

และได้ค่ากึ่งกลางเท่ากับ $m(t_0, d | E(\delta)) = \frac{1}{2}(m_+(t_0, d | E(\delta)) - m_+(t_0, -d | E(\delta)))$

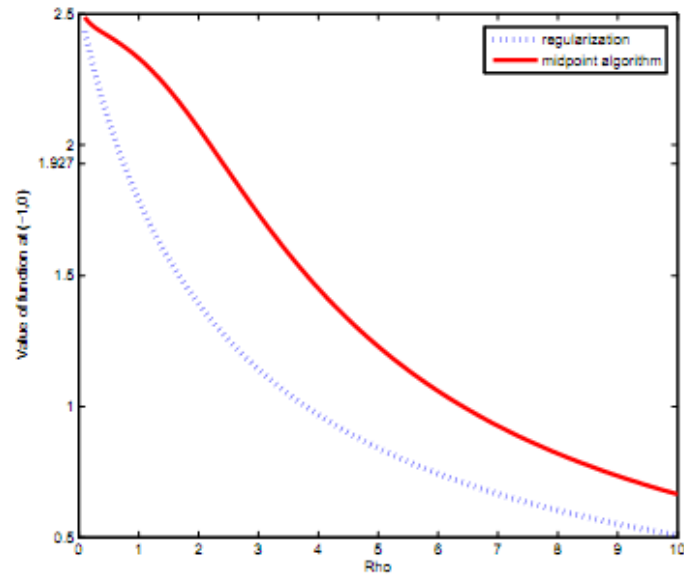


ภาพที่ 4.2 ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_0)$ ณ σ ต่าง ๆ และค่าจริง $g(t_0) = 0.6767$

ลำดับถัดไปจะประมาณค่าที่จุด t_{-1} โดยพิจารณาเซตเซตพาเซียวไฮเพอร์อิลฟ์ คือ

$$H(d | \delta E_2) = \{x : \|Qx - d\|_2 \leq \varepsilon, \|x\| \leq \delta, f(t_0) = f_\sigma(t_0)\}$$

ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-1})$ เมื่อ $f \in H(d | \delta E_2)$ นั่นคือข้อมูลที่มีค่าคลาดเคลื่อนต่างกันได้แสดงดังภาพ 3 ในขณะที่ค่าจริงของฟังก์ชันคือ 1.927

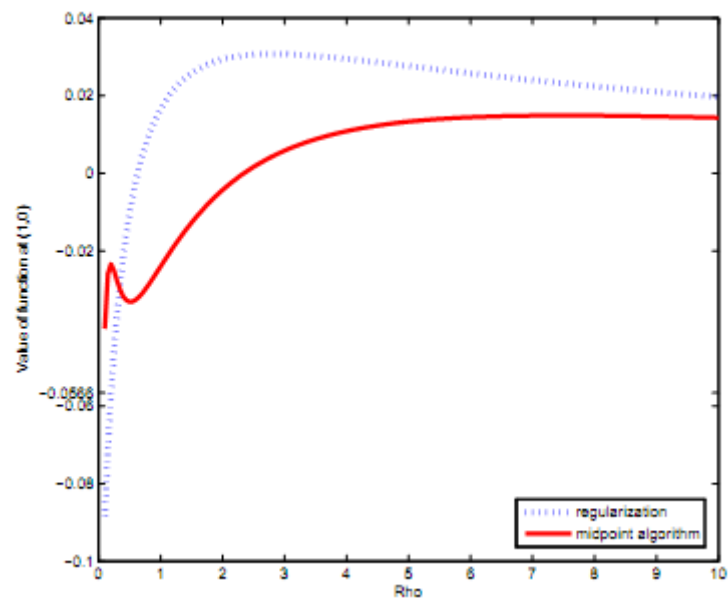


ภาพที่ 4.3 ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-1})$ ณ σ ต่าง ๆ

ในทำนองเดียวกันการประมาณค่าที่จุด t_{-2} โดยพิจารณาเซตเซตพาเซียวไฮเพอร์อิลฟ์ซ์ คือ

$$H(d | \delta E_2) = \{x : \|Qx - d\|_2 \leq \varepsilon, \|x\| \leq \delta, f(t_0) = f_\sigma(t_0), f(t_{-1}) = f_\sigma(t_{-1})\}$$

ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-2})$ เมื่อ $f \in H(d | \delta E_2)$ นั่นคือข้อมูลที่มีค่าคลาดเคลื่อนต่างกันได้แสดงดังภาพ 4 ในขณะที่ค่าจริงของฟังก์ชันคือ -0.0566



ภาพที่ 4.4 ผลการประมาณค่าของฟังก์ชันที่จุด $f(t_{-2})$ ณ σ ต่าง ๆ

จากการรายงานดังกล่าวผู้วิจัยสามารถสรุปได้ว่าผู้วิจัยได้พัฒนาสมการไฮเพอร์เซอร์ไปสู่ข้อมูลที่มีความหลากหลายมากขึ้น รวมถึงผลการแสดงเชิงตัวเลขในการนำองค์ความรู้ที่ได้ไปประยุกต์ใช้กับปัญหาการประมาณค่าของฟังก์ชันนั้นมีความใกล้เคียงกับค่าจริง ซึ่งถือเป็นการพัฒนาองค์ความรู้ของสมการไฮเพอร์เซอร์เคิลไปสู่การนำไปใช้กับข้อมูลจริงได้มีประสิทธิภาพมากขึ้น

บทที่ 5 สรุปและวิจารณ์ผลการวิจัย และข้อเสนอแนะสำหรับงานวิจัยในอนาคต

จากผลการวิจัยที่ผู้วิจัยได้นำเสนอในข้างต้นแสดงให้เห็นถึงการพัฒนางานวิจัยของ
อสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนที่สามารถนำไปประยุกต์กับปัญหาการ
ประมาณค่าของฟังก์ชันดังต่อไปนี้

1. ผู้วิจัยได้เสนอวิธีการเปลี่ยนเซตพหุเชิงเส้นไฮเพอร์อิลลิปซไปสู่ฐานที่เป็นเซตเชิงตั้งฉากซึ่งถือเป็นเครื่องมือที่มีประโยชน์มากในปริภูมิฮิลเบิร์ต โดยพบว่า

$$H(d | E) = H^*(d^* | E(A)) = \{x : Rx - d \in E(A), \|x\| \leq 1\}$$

เมื่อ $E(A) = \{Ae : e \in E\}$ และเวกเตอร์ $d^* \in \mathbb{R}^n$ ถูกกำหนดโดย

$$d_1^* = \frac{d_1}{\sqrt{|G(x_1)|}}$$

สำหรับ $k = 2, \dots, n$

$$d_k^* = \frac{1}{\sqrt{|G(x_1, x_2, \dots, x_{k-1})| |G(x_1, x_2, \dots, x_k)|}} \begin{vmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \dots & \langle x_k, x_1 \rangle \\ \vdots & \vdots & & \vdots \\ \langle x_1, x_{k-1} \rangle & \langle x_2, x_{k-1} \rangle & \dots & \langle x_k, x_{k-1} \rangle \\ d_1 & d_2 & \dots & d_k \end{vmatrix}$$

และ $A = \begin{bmatrix} \beta_{11} & 0 & \dots & 0 \\ \beta_{21} & \beta_{22} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{bmatrix}$ นอกจากนี้ยังพบว่า $\beta_{kk} = \sqrt{\frac{|G(x_1, \dots, x_{k-1})|}{|G(x_1, \dots, x_k)|}}$ สำหรับทุก ๆ

$k = 1, 2, \dots, n$

2. โดยผู้วิจัยได้เสนอเงื่อนไขที่เห็นได้ชัดมากขึ้นที่ใช้ในการตรวจสอบ $H(d | E) \neq \emptyset$ โดยเริ่มพิจารณาจากกรณีที่มีค่าคลาดเคลื่อนอยู่ใน $\square$ และ $\square^2$

กรณีที่ 1 ค่าคลาดเคลื่อนอยู่ใน $\square$

$$H(d | E) = H^*(d^* | E(A)) \neq \emptyset \quad \text{ก็ต่อเมื่อ}$$

$$\min_{|e| \leq \varepsilon} (d_n^* + \beta_{nn} e)^2 \leq 1 - (d_I^*, d_I^*)$$

ในงานวิจัยนี้ผู้วิจัยได้ศึกษาลักษณะของคำตอบ (1)

$$\min_{|e| \leq \varepsilon} (d_n^* + \beta_{nn} e)^2 = \begin{cases} 0, & \left| \frac{d_n^*}{\beta} \right| \leq \varepsilon \\ d_n^* + \beta_{nn} \varepsilon, & \left| \frac{d_n^*}{\beta} \right| > \varepsilon \end{cases}$$

กรณีที่ 2 ค่าคลาดเคลื่อนอยู่ใน $\square^2$

$H(d | \mathcal{E}) = H^*(d^* | \mathcal{E}(A)) \neq \emptyset$ ก็ต่อเมื่อ

$$\min_{e \in \mathcal{E}} (d_J^* + A_J e_J, d_J^* + A_J e_J) \leq 1 - (d_I^*, d_I^*) \quad (1)$$

ในงานวิจัยนี้ผู้วิจัยได้ศึกษาลักษณะของคำตอบข้างต้นโดยอ้างอิงจากงานวิจัย [9]

3. ผู้วิจัยได้เสนอวิธีการหาค่าขวสุดของช่วง I สำหรับข้อมูลที่คลาดเคลื่อนทั้งใน $\square$ และ $\square^2$ และใช้ L^∞ นอร์มเป็นเครื่องมือที่ใช้วัดค่าคลาดเคลื่อนดังต่อไปนี้

ถ้า $\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} \notin H(d | \mathcal{E})$ จะได้ว่า จะมีเวกเตอร์ $c^* \in \square^n$ เพียงตัวเดียวที่ทำให้

$$-Q \left(\frac{x_0 - Q^T c^*}{\|x_0 - Q^T c^*\|} \right) + \varepsilon w^* + d$$

เมื่อเวกเตอร์ $w^* = \begin{cases} 0, & i \in I \\ \text{sgn}(c_n^*), i = n \end{cases}$ และ

$$x_+(d) = \arg \max \{ \langle x, x_0 \rangle : H(d_I) \cap H(d_n \pm \varepsilon) \}$$

โดยใช้วิธีการเดียวกันสำหรับค่าคลาดเคลื่อนที่อยู่ใน $\square^2$ จากแนวคิดดังกล่าวเราสามารถขยายไปยังกรณีข้อมูลที่คลาดเคลื่อนทั้งใน $\square^n$

4. ผู้วิจัยได้นำองค์ความรู้ของสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนและสมการไฮเพอร์เซอร์เคิลสำหรับข้อมูลที่มีค่าคลาดเคลื่อนบางส่วนมาประยุกต์ใช้ในการประมาณค่าที่ดีที่สุดในการประมาณค่าของฟังก์ชันที่จุด t_{-1}, t_0, t_1 และพบว่าในการทดลองดังกล่าวแสดงให้เห็นการประมาณค่าโดยใช้วิธีจุดกึ่งกลางนั้นให้ผลการทดลองที่ใกล้เคียงกับค่าจริง

แต่อย่างไรก็ตามข้อมูลที่มีในปัญหาจริงนั้นมีความหลากหลายซึ่งขึ้นอยู่กับหลายปัจจัย ดังนั้น ผู้วิจัยจึงเสนอแนวทางแรกสำหรับงานวิจัยในอนาคตคือการพัฒนาสมการดังกล่าวไปสู่ข้อมูลที่มีความหลากหลายมากขึ้น แนวทางที่สองคือ ขยายสมการดังกล่าวไปยังปริภูมิบานาค

เอกสารอ้างอิง

- [1] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* 686, 1950, pp. 337–404.
- [2] J. P. Aubin, Mathematical Methods of Game and Economic Theory Studies in Mathematics and its applications, Vol 7, North Holland(1982)
- [3] A. V. Balakrishnan, An operator theoretic formulation of a class of control problem and a steepest descent method of solution, *J. SIAM control Ser. A*, 1(1963) pp 109-127
- [4] Dimitri P. Bertsekas, Nonlinear Programming, second edition, Athena Scientific, Belmont, Massachusetts, 1995
- [5] Philip J. Davis, Interpolation and Approximation, *J Dover Publications, New York*. 1975.
- [6] Norman R. Draper, “Ridge analysis” of response surfaces, *Technometrix*, 5(1963), pp 469-479
- [7] Peter Duren, Theory of H_p spaces, *Second editon, Dover Publications, New York*. 2000.
- [8] T. Evgeniou, T. Poggio and M. Point, Regularization networks and support vector machines, *Adances in Computational Mathematics*
- [9] A. Fathi and N. Sharifan, A classic new method to solve quartic equations, *SciencePG*, 2013;2(2):pp24-27
- [10] George E. Forsthe and Gene H. Golub, On the stationary values of a second-degree polynomial on the unit sphere, *SIAM J. Appl. Math.* 13(1965), pp 1050-1068.
- [11] M. Golomb and H. F. Weinberger, Optimal Approximation and Error Bounds, pp 117-190 of “ On Numerical Approximation” R. E. Langer, ed., Madison(1956).
- [12] K. Khompurngson and C. A. Micchelli, Hide, submitted for publication
- [13] K. Khompurngson, B Novapruteep and Y Lenbury. Learning the Value of a Function from Inaccurate Data. *East-West Journal of Mathematics*, Special Volume. (Accepted)
- [14] C.A. Micchelli and A. Pinkus, Variational problem arising from balancing several error criteria, *Rendiconti di Mathematica*. Serie VII, 14 Roma, 1994, pp. 37-86.
- [15] C.A Micchelli and T. J Rivlin, A survey of optimal recovery, *Optimal Estimation in Approximation Theory*. C.A Micchelli, and T. J Rivlin, eds., Plenum Press, 1977, pp. 1-53.
- [16] B. Scholkopf and A. J. Smola, Learning with Kernels, *The MIT Press, Cambridge, MA, USA*. 2002.
- [17] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, *Cambridge University Press*. 2004.
- [18] E. Spjotvoll, A note on a theorem of Forsythe and Golub, *SIAM J. Appl. Math.* 3rd edition. 23, 1972, pp. 307-311.

- [19] I. Steinwart and D. Hush and C. Scovel, An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels, *IEEE, Transactions on Information Theory*. 52, 2006, pp. 4635-4643.
- [20] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [21] G. Wahba, *Spline Models for Observational Data*, *SIAM, Philadelphia, Series in Applied Mathematics*. 59, 1990.

Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

การเสนอผลงานในที่ประชุม

1. เสนอผลงานในรูปแบบ Oral ในประชุมสัมมนา TRF SEMINAR ครั้งที่ 110 หัวข้อเรื่อง
"Hypersphere Inequality for Data Error and Its Application"

การนำผลงานวิจัยไปใช้ประโยชน์

เชิงวิชาการ งานวิจัยเป็นส่วนหนึ่งในการพัฒนาการเรียนการสอนวิชา สัมมนา และวิชาการศึกษา
อิสระ หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัย
พะเยา

ผลงานที่คาดว่าจะตีพิมพ์ในอนาคต

1. "Transformation of Hypersphere Inequality for Data Error to Orthonormal Bases"
2. "Alternative Approximation Method for Learning Multiple Feature"

บทที่ 6 ภาคผนวก

Transformation of Hypercircle Inequality for Data Error to Orthonormal Bases

Kannika Khompurngson^{*1} and Suthep Suantai²

¹Division of Mathematics, School of Science, University of Phayao, Phayao, THAILAND

²Department of Mathematics, Faculty of Science, Chiangmai University, Chaingmai, THAILAND

Email: Kannika Khompurngson^{*} - kannika.kh@up.ac.th; Suthep Suantai - suthep.s@cmu.ac.th;

^{*}Corresponding author

Abstract

The classical hypercircle inequality has been applied to kernel-based learning to determine a function representation when there is known data exactly. Our previous work which was motivated by this limited said data has extended the hypercircle inequality to circumstances for which there is known both accurate and inaccurate data. In this paper, we continue our study of this subject by presenting the transformation of its material to orthonormal bases which are useful tools in Hilbert space and practice. Moreover, we also specialize the new result to the learning problem in Reproducing kernel Hilbert space (RKHS). Specifically, we choose the Hardy space of square integrable function on the unit circle which is well-known in RKHS.

2010 Mathematics Subject Classification: 46E22, 46C07

Keywords: Hypercircle Inequality, Convex Optimization and Reproducing kernel Hilbert space.

1 Introduction and preliminary results

Let $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ be a set of *linearly independent* vectors in H which is assumed to be Hilbert space over the real number with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Let B be unit ball in H and we denote $\mathbb{N}_n = \{1, 2, \dots, n\}$. For any $d \in \mathbb{R}^n$,

$$\mathcal{H}(d) := \{x : x \in B, Q(x) = d\}$$

is called *hypercircle* where $Q : H \longrightarrow \mathbb{R}^n$ is a linear operator H onto $\mathbb{R}^n$ as

$$Qx = (\langle x, x_j \rangle : j \in \mathbb{N}_n). \quad (1)$$

Consequently, the adjoint map $Q^T : \mathbb{R}^n \rightarrow H$ is given at $a = (a_j : j \in \mathbb{N}_n) \in \mathbb{R}^n$ as

$$Q^T(a) = \sum_{i \in \mathbb{N}_n} a_i x_i.$$

The Gram matrix of the vectors in $\mathcal{X}$ is $G_{\mathcal{X}} = QQ^T = (\langle x_j, x_l \rangle : j, l \in \mathbb{N}_n)$, a symmetric and positive definite matrix. It is well-known that there exist a unique vector $x(d) \in M$ such that

$$x(d) := \arg \min \{ \|x\| : x \in H, Q(x) = d \}, \quad (2)$$

where M is the n -dimensional subspace of H spanned by the vectors in $\mathcal{X}$, see for example [7]. If $H \neq M$ then $\mathcal{H}(d)$ consists of exactly one point if and only if $\|x(d)\| = 1$, [5]. The principal significance of hypercircle inequality ensures that the value $\langle x(d), x_0 \rangle$ is the best estimator to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d)$. Moreover, the vector $x(d) := Q^T(G_{\mathcal{X}}^{-1}d)$ and $\|x(d)\|^2 = (d, G_{\mathcal{X}}^{-1}d)$. Therefore, the hypercircle inequality states that.

If $x \in \mathcal{H}(d)$ and $x_0 \in H$ then

$$|\langle x(d), x_0 \rangle - \langle x, x_0 \rangle| \leq \text{dist}(x_0, M) \sqrt{1 - \|x(d)\|^2}, \quad (3)$$

where $\text{dist}(x_0, M) := \min \{ \|x_0 - y\| : y \in M \}$. Moreover, if $\mathcal{H}(d) \neq \emptyset$ then there is an element $x_{\pm}(d) = \pm \frac{x_0 - Q^T a_{\pm}}{\|x_0 - Q^T a_{\pm}\|}$ for which equality above holds where the vector $a_{\pm} \in \mathbb{R}^n$ is given by the formula

$$a_{\pm} := G_{\mathcal{X}}^{-1}(Qx_0 \mp \frac{\sqrt{\|x_0\|^2 - (Qx_0, G_{\mathcal{X}}^{-1}Qx_0)}}{\sqrt{1 - \|x(d)\|^2}}d). \quad (4)$$

Let the vectors in $\mathcal{X}$ be orthonormalized according to the Gram-Schmidt process yielding $x_1^*, \dots, x_n^*$. Let M^* be the n -dimensional subspace of H spanned by the vectors $x_1^*, \dots, x_n^*$, see [3]. Consequently, the Gram matrix for $x_1^*, \dots, x_n^*$ is the identity matrix. For any $x \in \mathcal{H}(d)$, the condition

$$Qx = d \text{ is equivalent to } Rx = d^*$$

where $R : H \longrightarrow \mathbb{R}^n$ is a linear operator H onto $\mathbb{R}^n$ as $R(x) = (\langle x, x_j^* \rangle : j \in \mathbb{N}_n)$ and $d_1^* = \frac{d_1}{\sqrt{|G(x_1)|}}$

$$d_k^* = \frac{1}{\sqrt{|G(x_1, \dots, x_{k-1})||G(x_1, \dots, x_k)|}} \begin{vmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \cdots & \langle x_k, x_1 \rangle \\ \vdots & \vdots & & \vdots \\ \langle x_1, x_{k-1} \rangle & \langle x_2, x_{k-1} \rangle & \cdots & \langle x_k, x_{k-1} \rangle \\ d_1 & d_2 & \cdots & d_k \end{vmatrix}, k = 2, \dots, n \quad (5)$$

Therefore, we represent the coefficients of d_k^* for all $k = 1, \dots, n$ by the following matrix

$$A = \begin{bmatrix} \beta_{11} & 0 & 0 & \cdots & 0 \\ \beta_{21} & \beta_{22} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \beta_{n1} & \beta_{n2} & \beta_{n3} & \cdots & \beta_{nn} \end{bmatrix},$$

where $\beta_{kk} = \sqrt{\frac{|G(x_1, \dots, x_{k-1})|}{|G(x_1, \dots, x_k)|}}$ for all $k = 2, \dots, n$. Hence, we point out that

$$\mathcal{H}(d) = \mathcal{H}^*(d^*) := \{x : x \in B, Rx = d^* = Ad\}$$

and the vector $x(d) = x^*(d^*) := R^T d^*$ where the adjoint map R^T is given by for each $a \in \mathbb{R}^n$ $R^T(x) = \sum_{j=1}^n a_j x_j^*$.

Recently, an extension of hypercircle inequality to *partially - corrupted* data was proposed by Kannika Khompurongson and Boriboon Novapruteep, [6]. We start with $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). For each $e = (e_1, \dots, e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$ where we denote $J = \mathbb{N}_n \setminus I$. We choose $\|\cdot\|_p : \mathbb{R}^{n-m} \rightarrow \mathbb{R}_+$ is l^p norm on $\mathbb{R}^{n-m}$ and define $\mathbb{E}_p = \{e : e \in \mathbb{R}^n : e_I = 0, \|e_J\|_p \leq \varepsilon\}$, where ε is some positive number and $2 \leq p \leq \infty$. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(d|\mathbb{E}_p) := \{x : x \in H, x \in B, Q(x) - d \in \mathbb{E}_p\}. \quad (6)$$

Similarly, we use the notation $Q_I x = (\langle x, x_j \rangle : j \in I)$ and $Q_J x = (\langle x, x_j \rangle : j \in J)$ respectively. According to the previous work [6], we observe that

$$\mathcal{H}(d|\mathbb{E}_p) = \mathcal{H}(d_I) \cap \mathcal{H}(d_J|E_J), \quad (7)$$

where we denote the *hypercircle* with the constant d_I as $\mathcal{H}(d_I) = \{x : x \in B, Q_I(x) = d_I\}$ and the *hyperellipse* with the constant d_J as $\mathcal{H}(d_J|E_J) = \{x : x \in B, Q_J(x) - d_J \in E_J\}$, where we define $E_J = \{c : c \in \mathbb{R}^{n-m} : \|c\|_p \leq \varepsilon\}$. For each $e \in \mathbb{E}_p$, we have that

$$x(d+e) = Q^T G_{\mathcal{X}}^{-1}(d+e) \text{ and } \|x(d+e)\|^2 = (d+e, G_{\mathcal{X}}^{-1}(d+e)).$$

Now, let us begin with an important result on this set.

Theorem 1.1 $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ if and only if

$$\min\{(d+e, G_{\mathcal{X}}^{-1}(d+e)) : e_I = 0, \|e_J\|_p \leq \varepsilon\} \leq 1 \quad (8)$$

Moreover, if $H \neq M$ then $\mathcal{H}(d|\mathbb{E}_p)$ consists of exactly one point if and only if $\min\{(d+e, G_{\mathcal{X}}^{-1}(d+e)) : e_I = 0, \|e_J\|_p \leq \varepsilon\} = 1$.

The main results address the following way: Given $x_0 \in H$, we want to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d|\mathbb{E}_p)$. That is, data set contains both accurate and inaccurate data. The best estimator is the *midpoint* of this interval

$$I(x_0, d|\mathbb{E}_p) := \{ \langle x, x_0 \rangle : x \in \mathcal{H}(d|\mathbb{E}_p) \}. \quad (9)$$

Next let us recall the duality formula for the right hand end point $m_+(x_0, d|\mathbb{E}_p)$, of the uncertainty interval. We refer the reader to the paper [5] and [6] for more detail information on the theory and proof.

Theorem 1.2 *If $\mathcal{H}(d|\mathbb{E}_p)$ contains more than one element then*

$$m_+(x_0, d|\mathbb{E}_p) = \min \{ \|x_0 - Q^T(c)\| + \varepsilon \|c\|_q + (d, c) : c \in \mathbb{R}^n \}. \quad (10)$$

Therefore, the midpoint is given by

$$m(x_0, d|\mathbb{E}_p) = \frac{m_+(x_0, d|\mathbb{E}_p) - m_+(x_0, -d|\mathbb{E}_p)}{2}. \quad (11)$$

In addition, if $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ is an *orthonormal* set of vector and the Gram matrix becomes the identity matrix. Consequently, we have that $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ if and only if

$$\min \{ (d_J + c, d_J + c) : c \in \mathbb{R}^{n-m}, \|c\|_p \leq \varepsilon \} \leq 1 - \|x(d_J)\|^2.$$

For $p = 2$, we have the following $\mathcal{H}(d|\mathbb{E}_2) \neq \emptyset$ if and only if

$$\min \{ (d_J + c, d_J + c) : c \in \mathbb{R}^{n-m}, \|c\|_2 \leq \varepsilon \} = \Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{d_j^2}{\Lambda - \varepsilon^2} \leq 1 - \|x(d_J)\|^2,$$

where $\mathbb{I} := \{j : d_j = 0, j \in J\}$ and $\Lambda = \varepsilon^2 - \sqrt{\sum_{i \in J} d_i^2}$. Summarizing, if $\Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{d_j^2}{\Lambda - \varepsilon^2} \leq 1 - \|x(d_J)\|^2$ then $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ for all $p \geq 2$, see [4, 6]. Therefore, we observe that if $\mathcal{X}$ is an *orthonormal* then the solution for checking $\mathcal{H}(d|\mathbb{E}_p) \neq \emptyset$ differs from (8) which is useful for practice. For this observation, our goal is to propose a dual problem of learning one feature from *partially – corrupted* data in the orthonormal bases which is appeared in section II. Section III specializes the results of section II to the learning problem in the Hardy space of square integrable function on the unit circle which is well-known in RKHS.

2 Main Results

First, we assumed that $\mathcal{X}_I = \{x_1, \dots, x_m\}$ and $\mathcal{X}_J = \{x_{m+1}, \dots, x_n\}$ respectively. Let the vectors in $\mathcal{X}_I \cup \mathcal{X}_J$ be orthonormalized according to the Gram-Schmidt process yielding $x_1^*, \dots, x_m^*, x_{m+1}^*, \dots, x_n^*$. For our purpose, we define $\mathbb{E}_p(A) := \{Ae : e \in \mathbb{E}_p\}$ and

$$\mathcal{H}^*(d^*|\mathbb{E}_p(A)) := \{x : x \in B, Rx - d^* \in \mathbb{E}_p(A)\}. \quad (12)$$

That is, for each $x \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))$

$$R_I(x) = d_I^* \quad \text{and} \quad R_J(x) - d_J^* = \mathbf{A}e_J,$$

where $\|e_J\|_p \leq \varepsilon$ and the $n - m$ matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \beta_{m+1 \ m+1} & 0 & 0 & \cdots & 0 \\ \beta_{m+2 \ m+1} & \beta_{m+2 \ m+2} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \beta_{n \ m+1} & \beta_{n \ m+2} & \beta_{n \ m+3} & \cdots & \beta_{nn} \end{bmatrix}. \quad (13)$$

Similarly, we point out that for each $x \in \mathcal{H}(d|\mathbb{E}_p)$ the condition

$$Qx = d + e \quad \text{is equivalent to} \quad Rx = A(d + e) = d^* + Ae.$$

Returning to our previous work, the solution of the primal problem in (9) is the *midpoint* of the uncertainty interval $I(x_0, d|\mathbb{E}_p)$ which is obtained by (11). Similarly, the midpoint of $I(x_0, d^*|\mathbb{E}_p(A)) := \{\langle x, x_0 \rangle : x \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))\}$ is the solution of dual problem. We then begin with a general result on such dual problem.

According to the definition of $\mathcal{H}^*(d^*|\mathbb{E}_p(A))$, we obtain that $\mathcal{H}^*(d^*|\mathbb{E}_p(A))$ is a sequentially compact subset of H and the function $x \rightarrow \langle x, x_0 \rangle$ is weakly continuous. Therefore, there exist $x_{\pm} \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))$ such that

$$\langle x_{\pm}, x_0 \rangle = m_{\pm}(x_0, d^*|\mathbb{E}_p(A)). \quad (14)$$

where $m_+(x_0, d^*|\mathbb{E}_p(A)) := \max\{\langle x, x_0 \rangle : x \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))\}$ and $m_-(x_0, d^*|\mathbb{E}_p(A)) := \min\{\langle x, x_0 \rangle : x \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))\}$ respectively. In addition, for each $Ae \in \mathbb{E}_p(A)$ the vector $x^*(d^* + Ae) \in \mathcal{H}^*(d^*|\mathbb{E}_p(A)) \cap M^*$ can be written in the form

$$x^*(d^* + Ae) = x^*(d_I^*) + x^*(d_J^* + \mathbf{A}e) \quad \text{and} \quad \|x^*(d^* + Ae)\|^2 = \|x^*(d_I^*)\|^2 + \|x^*(d_J^* + \mathbf{A}e)\|^2, \quad (15)$$

where $x^*(d_I^*) \in \mathcal{H}^*(d_I^*) := \{x : x \in B, R_I(x) = d_I^*\}$ and $x^*(d_J^* + \mathbf{A}e) \in \mathcal{H}^*(d_J^*|E_p(\mathbf{A})) := \{x : x \in B, R_J(x) - d_J^* \in E_p(\mathbf{A})\}$ when $E_p(\mathbf{A}) = \{\mathbf{A}c : c \in \mathbb{R}^{n-m} : \|c\|_p \leq \varepsilon\}$.

Lemma 2.1 $\mathcal{H}^*(d^*|\mathbb{E}_p(A)) \neq \emptyset$ if and only if

$$\min \left\{ (\mathbf{A}^{-1}d_J^* + \varepsilon\xi, \mathbf{A}^T \mathbf{A}(\mathbf{A}^{-1}d_J^* + \varepsilon\xi)) : \|\xi\|_p \leq 1 \right\} \leq 1 - \|x(d_I^*)\|^2. \quad (16)$$

Proof. Let $x \in \mathcal{H}^*(d^*|\mathbb{E}_p(A))$. Then there is $Ae \in \mathbb{E}_p(A)$ such that $x = x^*(d^* + Ae) = R^T(d^* + Ae) \in M^*$ and $\|x^*(d^* + Ae)\|^2 = \|x^*(d_I^*)\|^2 + \|x^*(d_J^* + \mathbf{A}e)\|^2 \leq 1$. We observe that $\|x^*(d_J^* + \mathbf{A}e)\|^2 = (\mathbf{A}^{-1}d_J^* + \varepsilon\xi, \mathbf{A}^T \mathbf{A}(\mathbf{A}^{-1}d_J^* + \varepsilon\xi))$. Hence, $\min \left\{ (\mathbf{A}^{-1}d_J^* + \varepsilon\xi, \mathbf{A}^T \mathbf{A}(\mathbf{A}^{-1}d_J^* + \varepsilon\xi)) : \|\xi\|_p \leq 1 \right\} \leq 1 - \|x(d_I^*)\|^2$. Conversely, (15) and (16) certainly implies $\mathcal{H}^*(d^*|\mathbb{E}_p(A)) \neq \emptyset$. $\square$

For $p = 2$, we describe the solution of the optimization problem appearing in (16) as presented in [8]. We begin with the following definition.

Definition Let C be an $n \times n$ symmetric matrix and $d \in \mathbb{R}^n$. The *spectrum* of the pair (C, d) is defined to be the set of all real numbers λ for which there exists an $x \in \mathbb{R}^n$ with euclidean norm one such that

$$C(x - d) = \lambda x. \quad (17)$$

Let $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-m}$ be eigenvalue of $\mathbf{A}^T \mathbf{A}$, $\{u^j : j \in \mathbb{N}_{n-m}\}$ be a corresponding orthonormal set of eigenvector, write the vector $\mathbf{A}^{-1}d_j^*$ in the form $\mathbf{A}^{-1}d_j^* = \sum_{j \in \mathbb{N}_{n-m}} \gamma_j u^j$ for some constants $\gamma_j \in \mathbb{R}$ and define the subset $\mathbb{I}$ of $\mathbb{N}_{n-m}$ by $\mathbb{I} := \{j : \lambda_j \gamma_j = 0\}$.

Theorem 2.2 If Λ is the least value in the spectrum of the pair $(\varepsilon^2 \mathbf{A}^T \mathbf{A}, \frac{\mathbf{A}^{-1}d^*}{\varepsilon})$ then $\mathcal{H}^*(d^* | (\mathbb{E}_2(\mathbf{A}))) \neq \emptyset$ if and only if

$$\Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} \leq 1 - (d_I^*, d_I^*).$$

Proof. This result is proved in much the same as the paper [5] and we refer the reader to the paper [4] for proofs of the solution of the optimization problem.

Here is another way of stating theorem 1.1: $\mathcal{H}(d | (\mathbb{E}_p)) \neq \emptyset$ if and only if $\Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} \leq 1 - (d_I^*, d_I^*)$ for all $p \geq 2$, Therefore, we establish the new version of theorem 1.2 with the different hypothesis.

Theorem 2.3 If $\Lambda + \Lambda \sum_{j \notin \mathbb{I}} \frac{\lambda_j |\gamma_j|^2}{\Lambda - \varepsilon^2 \lambda_j} < 1 - (d_I^*, d_I^*)$ then

$$m_+(x_0, d | \mathbb{E}_p) = \min \{ \|x_0 - Q^T(c)\| + \varepsilon \|c\|_q + (d, c) : c \in \mathbb{R}^n \}. \quad (18)$$

3 Example and Application

In this section, let us specialize the recent results to the problem of function estimation in reproducing kernel Hilbert space (*RKHS*). To this end, we let H_K be a *RKHS* of real-valued function on a set $\mathcal{T}$. The real value function $K(t, s)$ of t and s in $\mathcal{T}$ is called a *reproducing kernel* of H if the following property is satisfied for all $t \in \mathcal{T}$ and $f \in H$

$$f(t) = \langle K_t, f \rangle, \quad (19)$$

where K_t is the function defined for any $s \in \mathcal{T}$ as $K_t(s) = K(t, s)$. Moreover, for any kernel K there is unique RKHS with K as its reproducing kernel [1]. Specifically, we choose the Hardy space of square integrable

function on the unit circle with reproducing kernel

$$K(z, \zeta) = \frac{1}{1 - \bar{\zeta}z}, \quad \zeta, z \in \Delta$$

where the unit disc $\Delta := \{z : |z| \leq 1\}$. Specifically, we let $H^2(\Delta)$ be the set of all functions analytic in the unit disc Δ with norm

$$\|f\| = \sup_{0 < r < 1} \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta \right)^{\frac{1}{2}}.$$

Let $T = \{t_j : j \in \mathbb{N}_n\}$ be distinct point (increasing order) in $(-1, 1)$. Consequently, we have a finite set of linearly independent elements $\{K_{t_j} : j \in \mathbb{N}_n\}$ in H where $K_{t_j}(t) := \frac{1}{1 - t_j t}$, $j \in \mathbb{N}_n$ and $t \in \Delta$. Thus, the vectors $\{x_j : j \in \mathbb{N}_n\}$ appearing above are identified with the function $\{K_{t_j} : j \in \mathbb{N}_n\}$. Therefore, the Gram matrix of the $\{K_{t_j} : j \in \mathbb{N}_n\}$ is given by

$$G(t_1, \dots, t_n) := (K(t_i, t_j) : i, j \in \mathbb{N}_n).$$

For this purpose, we recall the Cauchy determinant identity which state that for any $\{t_j : j \in \mathbb{N}_n\}, \{s_j : j \in \mathbb{N}_n\}$ that

$$\det \left(\frac{1}{1 - t_i s_j} \right)_{i, j \in \mathbb{N}_n} = \frac{\prod_{1 \leq j < i \leq n} (t_j - t_i)(s_j - s_i)}{\prod_{i, j \in \mathbb{N}_n} (1 - t_i s_j)}, \quad (20)$$

see for example [3]. From this formula we obtain that

$$\det G(t_1, \dots, t_n) = \frac{\prod_{1 \leq i < j \leq n} (t_i - t_j)^2}{\prod_{i, j \in \mathbb{N}_n} (1 - t_i t_j)}. \quad (21)$$

From (19), the linear operator $Q : H^2(\Delta) \longrightarrow \mathbb{R}^n$ defined for $f \in H^2(\Delta)$ as the following way

$$Qf = \left(\langle f, K_{t_i} \rangle = f(t_i) : i \in \mathbb{N}_n \right).$$

By Gram-Schmidt process and the formula (20) and (21), we obtain the vector K_j^* for any $j \in \mathbb{N}_n$. In particular, the vector K_j^* is given by the formula

$$\begin{aligned} K_1^* &= \sqrt{1 - t_1^2} K_{t_1}, \\ K_k^* &= \sqrt{1 - t_k^2} \sum_{l=1}^k (-1)^{k+l} \frac{\prod_{i \in \mathbb{N}_{k-1}} |1 - t_l t_i|}{\prod_{\substack{i \in \mathbb{N}_k \\ i \neq l}} |t_l - t_i|} K_{t_l}, \quad (k = 2, 3, \dots, n). \end{aligned}$$

For any $d = (d_1, d_2, \dots, d_n) \in \mathbb{R}^n$, we obtain that the condition

$$Qf = \left(f(t_i) : i \in \mathbb{N}_n \right) = d \text{ is equivalent to } Rf = d^*$$

where $R : H^2(\Delta) \longrightarrow \mathbb{R}^n$ is a linear operator $H^2(\Delta)$ onto $\mathbb{R}^n$ as $R(f) = (\langle f, K_j^* \rangle : j \in \mathbb{N}_n)$ and

$$\begin{aligned} d_1^* &= \sqrt{1 - t_1^2} d_1, \\ d_k^* &= \sqrt{1 - t_k^2} \sum_{l=1}^k (-1)^{k+l} \frac{\prod_{i \in \mathbb{N}_{k-1}} |1 - t_l t_i|}{\prod_{\substack{i \in \mathbb{N}_k \\ i \neq l}} |t_l - t_i|} d_l, \quad (k = 2, 3, \dots, n) \end{aligned}$$

For our example, we choose $\mathbb{E}$ as follows $\mathbb{E} := \{e : e \in \mathbb{R}^n, e_I = 0, |e_n| \leq \varepsilon\}$ where $I = \{1, 2, \dots, n-1\}$. The *partial hyperellipse* becomes

$$\mathcal{H}(d|\mathbb{E}) := \{f : f \in H^2(\Delta), \|f\| \leq 1, Q_I(f) = d_I, |(Q(f) - d)_n| \leq \varepsilon\}.$$

Clearly, we have only one inaccurate data and for any $f \in H_K$

$$f(t_j) = d_j \text{ for all } j \in \mathbb{N}_{n-1} \text{ and } f(t_n) = d_n + e \text{ where } |e| \leq \varepsilon.$$

In this case, the corresponding *partial hyperellipse* $\mathcal{H}(d|\mathbb{E})$ is given by

$$\mathcal{H}^*(d^*|(\mathbb{E}(\mathbf{A}))) := \{f : f \in H^2(\Delta), \|f\| \leq 1, R_I(f) = d_I^*, R_n(f) - d_n \in \mathbf{A}(\varepsilon)\}.$$

where $\mathbf{A}(\varepsilon) := \{\beta e : e \in \mathbb{R}, |e| \leq \varepsilon\}$ and $\beta = \sqrt{\frac{\det G(t_1, \dots, t_{n-1})}{\det G(t_1, \dots, t_n)}} = \sqrt{1 - t_n^2} \frac{\prod_{i \in \mathbb{N}_{n-1}} |1 - t_n t_i|}{\prod_{i \in \mathbb{N}_{n-1}} |t_n - t_i|}.$

Alternatively, we have that $\mathcal{H}^*(d^*|(\mathbb{E}(\mathbf{A}))) \neq \emptyset$ if and only if

$$\min \{(d_n^* + \beta e)^2 : |e| \leq \varepsilon\} \leq 1 - (d_I^*, d_I^*).$$

Moreover, we point out that formula

$$\min \{(d_n^* + \beta e)^2 : |e| \leq \varepsilon\} = \begin{cases} 0, & \left| \frac{d_n^*}{\beta} \right| \leq \varepsilon \\ d_n^* + \beta \varepsilon \frac{\hat{e}}{|\hat{e}|}, & \left| \frac{d_n^*}{\beta} \right| > \varepsilon \end{cases}$$

where $\hat{e} = -\frac{d_n^*}{\beta}.$

Given $t_0 = 0$ and $t_0 \notin T := \{t_j : j \in \mathbb{N}_n\}$. Again, the vector x_0 appearing the previous is identified with the function K_{t_0} . We wish to estimate $f(0) = \langle f, K_{t_0} \rangle$ optimally, given that $\|f\|_K \leq 1$, $f(t_j) = d_j$ for all

$j \in \mathbb{N}_m$ and $|(Q(f) - d)_j| \leq \varepsilon$. According to theorem 2.4, we have that

$$\begin{aligned} m_+(K_{t_0}, d|\mathbb{E}_p(A)) &= \langle f_{d_I}^*, K_{t_0} \rangle + m_+(K_{t_0}, \mathbf{d}|\delta\mathbb{E}_p(A)) \\ &= \langle \sum_{i=1}^{n-1} d_i^* K_i^*, K_{t_0} \rangle + m_+(K_{t_0}, \mathbf{d}|\delta\mathbb{E}_p(A)) \\ &= \sum_{i=1}^{n-1} d_i^* D_i^* + m_+(K_{t_0}, \mathbf{d}|\delta\mathbb{E}_p(A)) \end{aligned}$$

where D^* is the vector in $\mathbb{R}^n$ with components given by the formula

$$D_k^* = \begin{cases} \sqrt{1 - t_1^2}, & k = 1 \\ \sqrt{1 - t_k^2} \sum_{l=1}^k (-1)^{k+l} \frac{\prod_{i \in \mathbb{N}_{k-1}} |1 - t_l t_i|}{\prod_{\substack{i \in \mathbb{N}_k \\ i \neq l}} |t_l - t_i|}, & k = 2, 3, \dots, n \end{cases}$$

Therefore, we only need to evaluate the following equation

$$m_+(K_{t_0}, \mathbf{d}|\mathbb{E}(A)) = \min \{ \delta \|K_{t_0} - R^T(c)\| + \beta \varepsilon |c_n| + d_n^* c_n : c \in \mathbb{R}^n \}.$$

To obtain the right hand endpoint, we consider two cases. Firstly, if $f_+^* := \delta \frac{K_{t_0} - R_I^T(D_I^*)}{\sqrt{1 - \sum_{i=1}^{n-1} (D_i^*)^2}} \in \mathcal{H}^*(\mathbf{d}|\delta\mathbb{E}(A))$ then

$$m_+(K_{t_0}, \mathbf{d}|\mathbb{E}(A)) = \sum_{i=1}^{n-1} d_i^* D_i^* + \delta \sqrt{1 - \sum_{i=1}^{n-1} (D_i^*)^2},$$

where $f_+^* := \arg \max \{ \langle x, x_0 \rangle : x \in \mathcal{H}^*(0, \delta) \}$. If $f_+^* \notin \mathcal{H}^*(\mathbf{d}|\delta\mathbb{E}(A))$ then the minimum $c^* \in \mathbb{R}^n$ is unique solution of the nonlinear equation

$$-\delta R\left(\frac{K_{t_0} - R^T c^*}{\|K_{t_0} - R^T c^*\|}\right) + \varepsilon w^* + \mathbf{d} = 0,$$

where w^* is the vector in $\mathbb{R}^n$ with components given by the formula

$$w_i^* = \begin{cases} 0, & i \in I \\ \beta \varepsilon \operatorname{sgn}(c_i^*), & i = n \end{cases}$$

and

$$f_+(\mathbf{d}|\delta\mathbb{E}(A)) := \delta \frac{K_{t_0} - R^T c^*}{\|K_{t_0} - R^T c^*\|}$$

satisfies

$$f_+(\mathbf{d}|\delta\mathbb{E}(A)) = \arg \max \{ \langle x, x_0 \rangle : x \in \mathcal{H}^*(0, \delta) \cap \mathcal{H}(d_n^* + \beta \varepsilon \operatorname{sgn}(c_i^*), \delta) \}.$$

Using (4), the vector $c^* \in \mathbb{R}^n$ is obtained by

$$c_i^* = \begin{cases} D_i^*, & i \in I \\ D_n^* - \frac{\sqrt{1 - \sum_{i=1}^{n-1} (D_i^*)^2}}{\sqrt{1 - (D_n^*)^2}} (d_n^* + \beta \varepsilon \operatorname{sgn}(c_i^*)), & i = n. \end{cases}$$

4 Competing interests

The authors declare that they have no competing interests.

5 Authors' contributions

The authors read and approved the final manuscript.

6 Acknowledgements

The author is supported by the Thailand Research Fund, the commission on Higher Education and University of Phayao under Grant TRG5680030.

References

1. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* 68(1950), 337–404
2. J. P. Aubin, *Mathematical methods of game and economic theory*, Studies in Mathematics and its Applications 7, North Holland, 1982.
3. Philip J. Davis, *Interpolation and approximation*, J Dover Publications, New York, 1975.
4. George E. Forsythe and Gene H. Golub, On the stationary values of a second-degree polynomial on the unit sphere , *SIAM J. Appl. Math.*, 13 (1965), 1050-1068.
5. K. Khompurngson and C. A. Micchelli, Hide, *Jaen Journal on Approximation*, 3 (2011), 87-115.
6. K. Khompurngson, B. Novaprateep, Hypercircle inequaliity for partially-corrupted data, *Annals of Functional Analysis*, 6(2015), no. 1, 95-108.
7. C.A Micchelli and T. J Rivlin, *A survey of optimal recovery in: Optimal estimation in approximation theory*, C.A Micchelli, and T. J Rivlin, eds., Plenum Press, 1977.
8. E. Spjøtvoll, A note on a theorem of Forsthe and Golub, *SIAM J. Appl. Math.*, 23(1972), 307-311.

Alternative Approximation Method for Learning Multiple Feature

Kannika Khompurngson^{*1} and Suthep Suantai²

¹Division of Mathematics, School of Science, University of Phayao, THAILAND

²Department of Mathematics, Faculty of Science, Chiangmai University , THAILAND

Email: Kannika Khompurngson^{*} - kannika.kh@up.ac.th; Suthep Suantai - suthep.s@cmu.ac.th;

^{*}Corresponding author

Abstract

A foundational concept in learning problem is to construct a functional representation from given data. Among learning methods, Hypercircle inequality (Hi) has been applied to kernel-based machine learning when data is known exactly. Recently, we have extended Hi to data error in two ways: First, we have extended it to circumstance for which all data is known within error. Second, we have extended it to partially-corrupted data. That is, data set contains both accurate and inaccurate data. In this paper, we apply the material from both previous work to estimate the unknown vectors in Hilbert space from knowledge of both its norm and linear observations of it, known within error.

Keywords. Hypercircle Inequality, Convex Optimization and Noise data.

1 Introduction, Notations and Preliminary Results

The study of learning problem is a topic of current interest in approximation theory and a foundational concept is to construct a functional representation from given data. Specifically, Hypercircle inequality (Hi) has been applied to kernel-based machine learning [8]. Unfortunately, the material on Hi only applies to circumstance for which data is accurate data. Recently, we have extended Hi to data error in two ways: First, we have extended it to circumstance for which all data is known within error [6, 7]. Second, we have extended it to partially-corrupted data [5]. That is, data set contains both accurate and inaccurate data.

In this paper, we apply the material from both previous work to estimate the unknown vectors in Hilbert space from knowledge of both its norm and linear observations of it, known within error.

Following these introductory remarks are two sub-sections which describe some previous results on the extension of hypercircle inequality for inaccurate data and partially-corrupted data respectively. In section 2 we apply the results in this section to the problem of learning the values of a function in a reproducing kernel Hilbert space by our proposed midpoint algorithm.

1.1 Hypercircle inequality for inaccurate data

Let H be the Hilbert space over the real numbers with inner product $\langle \cdot, \cdot \rangle$ and $\mathcal{X} = \{x_j : j \in \mathbb{N}_n\}$ be a set of *linearly independent* vectors in H where we denote $\mathbb{N}_n = \{1, 2, \dots, n\}$. Consequently, we define $M := \left\{ \sum_{j \in \mathbb{N}_n} a_j x_j : (a_j : j \in \mathbb{N}_n) \in \mathbb{R}^n \right\}$ which is the n -dimensional linear subspace of H spanned by the vectors in $\mathcal{X}$. Let $Q : H \longrightarrow \mathbb{R}^n$ be a linear operator H onto $\mathbb{R}^n$ which is defined for $x \in H$ as

$$Qx = (\langle x, x_j \rangle : j \in \mathbb{N}_n). \quad (1)$$

Consequently, the adjoint map $Q^T : \mathbb{R}^n \longrightarrow H$ is given at $a = (a_j : j \in \mathbb{N}_n) \in \mathbb{R}^n$ as

$$Q^T a = \sum_{j \in \mathbb{N}_n} a_j x_j \quad (2)$$

and the Gram matrix of the vectors in $\mathcal{X}$ is

$$G = QQ^T = (\langle x_j, x_l \rangle : j, l \in \mathbb{N}_n). \quad (3)$$

Moreover, G is positive definite matrix. For each $d \in \mathbb{R}^n$, it is well-known that there exist a unique vector $x(d) \in M$ such that

$$x(d) := Q^T(G^{-1}d) := \arg \min\{\|x\| : x \in \mathcal{P}(d)\}, \quad (4)$$

where we denote the hyperplane as the following $\mathcal{P}(d) := \{x : x \in H, Qx = d\}$. Moreover, we provide the useful equation

$$\min\{\|x\| : x \in H, Qx = d\} = \|x(d)\| = \sqrt{\langle d, G^{-1}d \rangle}.$$

Consequently, let us recall the definition of *hypercircle* as follows

$$\mathcal{H}(d) := \{x : x \in B, Q(x) = d\}.$$

We remark that $\mathcal{H}(d) \neq \emptyset$ if and only if $\|x(d)\| = \sqrt{\langle d, G^{-1}d \rangle} \leq 1$. If $H \neq M$ then $\mathcal{H}(d)$ consists of exactly one point if and only if $\|x(d)\| = 1$. Next, we review basic fact about Hypercircle inequality for data error

and discuss what we need for section 3. We begin with $E = \{e : e \in \mathbb{R}^n, |e| \leq \varepsilon\}$, where $|\cdot| : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is some prescribed norm on $\mathbb{R}^n$ and $\varepsilon > 0$. For each $d \in \mathbb{R}^n$, the definition of the *hyperellipse* is given by

$$\mathcal{H}(d|E) := \{x : x \in B, Q(x) - d \in E\}.$$

Consequently, we have the useful fact

$$\mathcal{H}(d|E) = \bigcup_{e \in E} \mathcal{H}(d + e),$$

where $B = \{x : x \in H, \|x\| \leq 1\}$. Given $x_0 \in H$, we want to estimate $\langle x, x_0 \rangle$ when $x \in \mathcal{H}(d|E)$ and the Hypercircle inequality for data error becomes in the following way.

If $x_0 \in H$ and $\mathcal{H}(d|E) \neq \emptyset$ then there is a $e_0 \in E$ such that for any $x \in \mathcal{H}(d|E)$

$$|\langle x(d + e_0), x_0 \rangle - \langle x, x_0 \rangle| \leq \frac{1}{2} \left(m_+(x_0, d|E) - m_-(x_0, d|E) \right),$$

where $x(d + e_0) = Q^T(G^{-1}(d + e_0)) \in \mathcal{H}(d|E)$, $m_+(x_0, d|E) := \max \{ \langle x, x_0 \rangle : x \in \mathcal{H}(d|E) \}$ and $m_-(x_0, d|E) := \min \{ \langle x, x_0 \rangle : x \in \mathcal{H}(d|E) \} = -m_+(x_0, -d|E)$ respectively.

In particular, the *Hypercircle inequality* is given by the following.

If $x(d) \in \mathcal{H}(d)$ and $x_0 \in H$ then for any $x \in \mathcal{H}(d)$

$$|\langle x(d), x_0 \rangle - \langle x, x_0 \rangle| \leq \text{dist}(x_0, M) \sqrt{1 - \|x(d)\|^2},$$

where $\text{dist}(x_0, M) := \min \{ \|x_0 - y\| : y \in M \}$.

According to Hypercircle inequality, the best estimator for $x \in \mathcal{H}(d)$ is $x(d)$ which is independent of x_0 . In case of inaccurate data, the best estimator *still* has the form of Representer Theorem (2) but the choice of the coefficients are generally depended on the vector x_0 . In Geometrically speaking, the best estimator $\langle x(d + e_0), x_0 \rangle$ is the midpoint of the interval of uncertainty which is defined by $I(x_0, d|E) := \{ \langle x, x_0 \rangle : x \in \mathcal{H}(d|E) \}$. To find the best predictor, we provided the useful duality formula for the right hand endpoint of the uncertainty interval. We then define the convex function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ defined for $c \in \mathbb{R}^n$

$$V(c) := \|x_0 - Q^T(c)\| + \varepsilon |c|_* + (d, c),$$

where $|\cdot|_* : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is conjugate norm of $|\cdot|$ which is used to measure data error and $(\cdot, \cdot)$ is Euclidean inner product on $\mathbb{R}^n$. The result state as the following. *If $\mathcal{H}(d) \neq \emptyset$ then*

$$m_+(x_0, d|E) = \min \{V(c) : c \in \mathbb{R}^n\}. \quad (5)$$

Moreover, $0 = \arg \min \{V(c) : c \in \mathbb{R}^n\}$ if and only if $\frac{x_0}{\|x_0\|} \in \mathcal{H}(d|E)$.

The detailed proof will appear in [3, 6, 10]. Alternatively, to find the best predictor, we only need evaluate the two numbers $m_+(x_0, \pm d|E)$ and then compute $\frac{1}{2}(m_+(x_0, d|E) - m_+(x_0, -d|E))$. In the special case that data error is measured with square loss, the duality formula in (5) becomes

$$m_+(x_0, d|E) = \min \{ \|x_0 - Q^T(c)\| + \varepsilon |c|_2 + (d, c) : c \in \mathbb{R}^n \}. \quad (6)$$

We provided a possible iterative method to solve the minimum vector c^* proceeds in the following manner, [7]. Let us introduce two positive constants given $\rho_k := |c^k|_2$ and $\tau_k := \|x_0 - Q^T c^k\|$. Next, we choose an initial vector $c^0 \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (\rho_k G + \varepsilon \tau_k I)^{-1} (\rho_k Q x_0 - \rho_k \tau_k d). \quad (7)$$

Our computation experience indicates that this iteration converges if the vector Qx_0 and d are linearly independent in $\mathbb{R}^n$. However, this has not been proved.

1.2 Hypercircle inequality for partially-corrupted data

Let $I \subseteq \mathbb{N}_n$ which contains m elements ($m < n$). Consequently, we use the notations $\mathcal{X}_I = \{x_i : i \in I\} \subset \mathcal{X}$ and $\mathcal{X}_J = \{x_i : i \in J\} \subset \mathcal{X}$, where we denote $J = \mathbb{N}_n \setminus I$. For each $e = (e_1, \dots, e_n) \in \mathbb{R}^n$, we also use the notations $e_I = (e_i : i \in I) \in \mathbb{R}^m$ and $e_J = (e_i : i \in J) \in \mathbb{R}^{n-m}$ respectively. We choose $||| \cdot ||| : \mathbb{R}^{n-m} \rightarrow \mathbb{R}_+$ is some prescribed norm on $\mathbb{R}^{n-m}$ and define $\mathbb{E} = \{e : e \in \mathbb{R}^n : e_I = 0, |||e_J||| \leq \varepsilon\}$, where ε is some positive number. For each $d \in \mathbb{R}^n$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(d|\mathbb{E}) := \{x : x \in H, \|x\| \leq 1, Q(x) - d \in \mathbb{E}\}. \quad (8)$$

In this case, we also provided the existence of the best estimator which *still* has the form of linear combination of vectors in $\mathcal{X}$ and the results follows by the same method as in [6]. Again, we provided the useful duality formula for the right hand endpoint of the uncertainty interval and the midpoint is given by $\frac{1}{2}(m_+(x_0, d|\mathbb{E}) - m_+(x_0, -d|\mathbb{E}))$. The result state as the following.

If $\mathcal{H}(d|\mathbb{E})$ contains more than one element then

$$m_+(x_0, d|\mathbb{E}) = \min \{ \|x_0 - Q^T(c)\| + \varepsilon \|c_J\|_* + (d, c) : c \in \mathbb{R}^n \}, \quad (9)$$

where $\|\cdot\|_* : \mathbb{R}^{n-m} \rightarrow \mathbb{R}_+$ is conjugate norm of $\|\cdot\|$ which is used to measure data error. Moreover, if either $x_0 \notin M$ or $\|\cdot\|_*$ is strictly convex then the right hand side of equation (9) has a unique solution.

Moreover, we provided the necessary and sufficient condition on $\mathcal{H}(d|\mathbb{E})$ such that the minimum vector c^* achieves with $c_J^* = 0$ which is useful for practice. To this end, let us define the convex function $\mathbb{V} : \mathbb{R}^n \rightarrow \mathbb{R}$ defined for $c \in \mathbb{R}^n$

$$\mathbb{V}(c) := \|x_0 - Q^T(c)\| + \varepsilon \|c_J\|_* + (d, c).$$

If $x_0 \notin M_I := \{Q_I^T(a) : a \in \mathbb{R}^m\}$ and $\mathcal{H}(d_I)$ contain more than one point then $c^* = \arg \min \{\mathbb{V}(c) : c \in \mathbb{R}^n\}$ with $c_J^* = 0$ if and only if

$$\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} \in \mathcal{H}(d|\mathbb{E})$$

where the vector

$$\frac{x_0 - Q_I^T(a^*)}{\|x_0 - Q_I^T(a^*)\|} := \arg \min \{\langle x, x_0 \rangle : x \in \mathcal{H}(d_I)\}.$$

The detailed proof will appear in [5]. As we already said, we going to apply the results in this section in learning problem. We then provide a possible iterative method to solve the minimum vector c^* on the right hand side in equation (9) when data error is measured with square loss. We choose an initial vector $c^0 \neq 0$ and then successively define $c^k, k \in \mathbb{N}$, by the formula

$$c^{k+1} = (G + \tau^k D^k)^{-1} (Qx_0 - \tau^k d) \quad (10)$$

where $\tau^k := \|x_0 - Q^T c^k\|$ and the matrix D^k is an $n \times n$ diagonal matrix and we define the elements on diagonal by

$$d_{ii}^k = \begin{cases} 0, & \text{if } i \in I \\ \frac{\varepsilon}{\rho_J^k}, & \text{if } i \in J \end{cases} \quad (11)$$

where $\rho_J^k := \|c_J^k\|_2$.

2 Main Results

In this section, we shall apply the available material from both previous work to the problem of learning the values of a function in a reproducing kernel Hilbert space(RKHS). They have an origin in the theory of reproducing kernel in the classical paper of Golomb and Weinberger [1, 4]. Specifically, we choose the gaussian kernel on $\mathbb{R}^2$, that is

$$K(x, y) := e^{-|x-y|_2^2}, \quad x, y \in \mathbb{R}^2.$$

In our example, we choose the value of $T = \{t_j : j \in \mathbb{N}_{20}\}$ on a ellipse curve surrounding the origin. Consequently, we have a finite set of linearly independent elements $\{K_{t_j} : j \in \mathbb{N}_{20}\}$ in H where $K_{t_j}(t) := K(t_j, t), j \in \mathbb{N}_n$ and $t \in \mathbb{R}^2$. Therefore, the vectors $\{x_j : j \in \mathbb{N}_{20}\}$ appearing in previous section are identified with the function $\{K_{t_j} : j \in \mathbb{N}_{20}\}$. These vectors determine a linear operator $Q : H \longrightarrow \mathbb{R}^n$ defined for $f \in H$ as

$$Qf = \left(\langle f, K_{t_j} \rangle = f(t_j) : j \in \mathbb{N}_{20} \right).$$

Moreover, the Gram matrix of the $\{K_{t_j} : j \in \mathbb{N}_{20}\}$ is given by

$$G := (K(t_i, t_j) : i, j \in \mathbb{N}_{20}).$$

Next, we choose the exact function is given by the formula

$$g(t) = 3.5K_{(1,1)}(t) + 1.75K_{(1,-1)}(t) + 3.25K_{(-1,-1)}(t) - 3.5K_{(-1,1)}(t), \quad t \in \mathbb{R}^2$$

and compute the vector $d := (g(t_j) : j \in \mathbb{N}_{20})$. Given $t_{-2}, t_{-1}, t_0 \in \mathbb{R}^2$, we want to estimate $f(t_{-2}), f(t_{-1}), f(t_0)$ knowing that $\|f\|_K \leq \delta$ and the data error vector $e := (d_j - f(t_j)) : j \in \mathbb{N}_{20}$ has Euclidean norm $\leq \varepsilon$ and δ are prescribed. However, with no effort at all the observations we made so far extend to the case that the unit ball B is replaced by δB where δ is any positive number. In addition, we shall compare the midpoint estimator discussed in previous section to the regularization estimator which is the stand method for learning problem. The computational steps are organized in the following steps:

Step 1. Given $\rho > 0$, we define the quadratic functional R_ρ for $f \in H$ as

$$R_\rho(f) := \|d - Qf\|_2^2 + \rho \|f\|_K^2.$$

The Representer Theorem say that the unique function with minimizes R_ρ has the form

$$f_\rho(t) := \sum_{j \in \mathbb{N}_n} c_\rho(j) K(t_j, t), \quad t \in \mathbb{R}^2$$

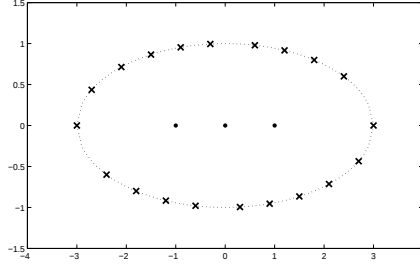


Figure 1. Point on a ellipse curve and estimating points.

and $c_\rho = (G + \rho I)^{-1}d$. Next, we introduce two functions of ρ given by the formula

$$\varepsilon(\rho) := |d - Qf_\rho|_2, \quad \delta(\rho) := \|f_\rho\|_K \quad (12)$$

and choose $f_\rho(t_{-2}), f_\rho(t_{-1}), f_\rho(t_0)$ as regularization estimators.

Step 2. To compare midpoint estimator and estimate $f(t_0)$, we need to identify an *hyperellipse* which contains f_ρ . To this end, we define the hyperellipse as follows $\mathcal{H}(d|\delta(\rho)E) := \{f : \|f\|_K \leq \delta(\rho), Q(f) - d \in E\}$ where $E = \{e : e \in \mathbb{R}^n, |e| \leq \varepsilon(\rho)\}$. Clearly, the regularization estimator f_ρ can be view as an element in hyperellipse

Since hyperellipse $\mathcal{H}(d|\delta(\rho)E)$ consists one point our strategy compare the regularization and midpoint estimator must consider a bigger hyperellipse. Consequently, we then compute both the regularization estimator and midpoint estimator corresponding to this hyperellipse $\mathcal{H}(d|\delta E)$ where $\delta = 3\delta(\rho)$ and compare to the true value of g at t_0 .

As explained earlier, the midpoint algorithm requires us to find numerically the minimum of the function in (5) for d and $-d$. then our midpoint estimator is given by $m(t_0) = \frac{m_+(t_0, d|\delta E) - m_+(t_0, -d|\delta E)}{2}$. For the computation of $m_+(t_0, \pm d|\delta E)$ we use the program `fminunc` in the optimization toolbox of Matlab 7.3.0 and also for comparison sake we use the iteration scheme (7) with an arbitrary chosen non zero initial vector.

Step 3. To estimate $f(t_{-1})$, we define the *partial hyperellipse* as follows

$$\mathcal{H}(d|\delta E) := \{f : \|f\|_K \leq \delta, f(t_0) = f_\rho(t_0), Q(f) - d \in E\}.$$

Clearly, the regularization estimator f_ρ can be view as an element in $\mathcal{H}(d|\delta E)$. To obtain the midpoint,

the algorithm requires us to find numerically the minimum of the function in (9) for $\mathbf{d}$ and $-\mathbf{d}$ where the vector $\mathbf{d} := (d_j : j \in \mathbb{N}_{20} \cup \{0\})$, $d_0 = f_\rho(t_0)$ and $d_j = d_j$ for $j \in \mathbb{N}_{20}$. Our midpoint estimator is given by $m(t_{-1}) = \frac{m_+(t_{-1}, \mathbf{d}|\delta\mathbb{E}) - m_+(t_{-1}, -\mathbf{d}|\delta\mathbb{E})}{2}$. For the computation of $m_+(t_{-1}, \pm\mathbf{d}|\delta\mathbb{E})$ we use the program `fminunc` in the optimization toolbox of Matlab 7.3.0 and also for comparison sake (10) with an arbitrary chosen non zero initial vector.

Step 4. To estimate $f(t_{-2})$, we again define *partial hyperellipse* as follows

$$\mathcal{H}(\mathbf{d}|\delta\mathbb{E}') := \{f : \|f\|_K \leq \delta, f(t_{-1}) = f_\rho(t_{-1}), f(t_0) = f_\rho(t_0), Q(f) - d \in E\}.$$

Clearly, the regularization estimator f_ρ can be view as an element in $\mathcal{H}(\mathbf{d}|\delta\mathbb{E}')$ and our midpoint estimator becomes as the following $m(t_{-2}) = \frac{m_+(t_{-2}, \mathbf{d}|\delta\mathbb{E}') - m_+(t_{-2}, -\mathbf{d}|\delta\mathbb{E}')}{2}$ where the vector $\mathbf{d} := (d_j : j \in \mathbb{N}_{20} \cup \{0, -1\})$ where $\mathbf{d}_{-1} = f_\rho(t_{-1})$, $\mathbf{d}_0 = f_\rho(t_0)$ and $\mathbf{d}_j = d_j$ for $j \in \mathbb{N}_{20}$.

The results of the computation are displayed below and at least for small values of the regularization parameter, but away from zero, the midpoint algorithm is better than regularization.

3 Acknowledgements

The author is supported by the Thailand Research Fund, the commission on Higher Education and University of Phayao under Grant TRG5680030.

References

1. Aronszajn, N: Theory of reproducing kernels. Trans. Amer. Math. Soc. 68, 337–404(1950)
2. Aubin, J, P: Mathematical methods of game and economic theory. Studies in Mathematics and its Applications 7, North Holland (1982)
3. Davis, P, J: Interpolation and approximation. J Dover Publications, New York(1975)
4. Golomb, M, Weinberger, H, F: optimal approximation and error bounds, in: On Numerical Approximation, R.E. Langer(ed). University of Wisconsin Press, Madison,117-190(1959)
5. Khompurngson, K, Novaprateep, B: Hypercircle inequality for partially-corrupted data, Annals of Functional Analysis, (in press).

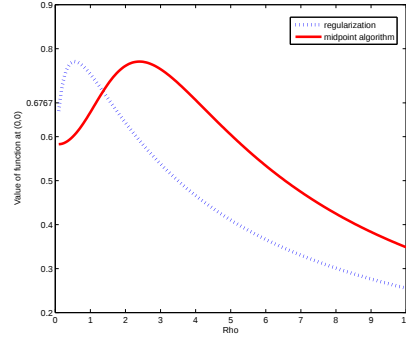


Figure 2. The result of regularization and midpoint algorithm at the point t_0 and exact value is $g(t_0) = 0.6767$

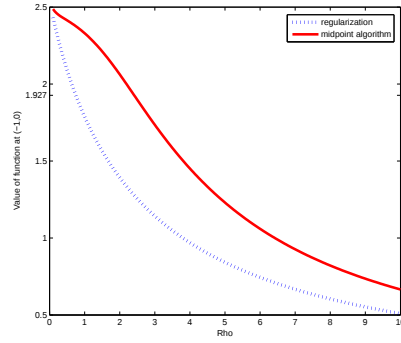


Figure 3. The result of regularization and midpoint algorithm at the point t_{-1} and exact value is $g(t_{-1}) = 1.927$

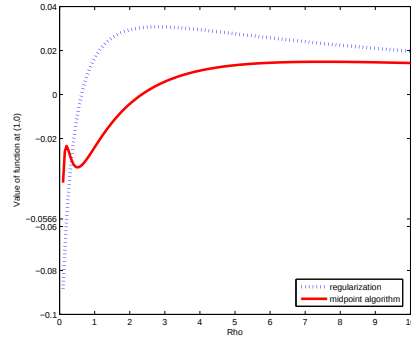


Figure 4. The result of regularization and midpoint algorithm at the point t_{-2} and exact value is $g(t_{-2}) = -0.0566$

6. Khompurngson, K, Micchelli, C, A: Hide.Jaen Journal on Approximation, 3, 87-115(2011)
7. Khompurngson, K ,Novaprteep, B, Lenbury, Y: Learning the value of a function from inaccurate data. East-Weast Journal of Mathematics Special Volume, 5(2011)
8. Micchelli, C,A, Rivlin T, J: A survey of optimal recovery in: Optimal estimation in approximation theory. C.A Micchelli, and T. J Rivlin, eds., Plenum Press(1977)
9. Roberts, A, W, Varberg, D, E: Convex Functions. Academic Press, New York and London(1973)
10. Royden, H, L:Real analysis. Macmillan Publishing Company, 3rd edition, New York(1988)